



**UNIVERSIDADE ESTADUAL DE CAMPINAS**  
**Faculdade de Engenharia Civil, Arquitetura e Urbanismo**

**DOUGLAS LOPES DE SOUZA**

**PROCESSAMENTO DE LINGUAGEM NATURAL DE NORMAS  
TÉCNICAS DA CONSTRUÇÃO CIVIL: ESTUDO DA NBR  
15.575**

**Campinas**  
**2024**

**Douglas Lopes de Souza**

**PROCESSAMENTO DE LINGUAGEM NATURAL DE NORMAS  
TÉCNICAS DA CONSTRUÇÃO CIVIL: ESTUDO DA NBR  
15.575**

Tese de Doutorado apresentada a Faculdade de Engenharia Civil, Arquitetura e Urbanismo da Unicamp, para obtenção do título de Doutor em Arquitetura, Tecnologia e Cidade, na área de Arquitetura, Tecnologia e Cidade.

**Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Regina Coeli Ruschel**

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA TESE  
DEFENDIDA PELO ALUNO DOUGLAS LOPES DE SOUZA E  
ORIENTADA PELA PROF<sup>A</sup> DR<sup>A</sup> REGINA COELI RUSCHEL.

---

**Campinas  
2024**

Ficha catalográfica  
Universidade Estadual de Campinas  
Biblioteca da Área de Engenharia e Arquitetura  
Rose Meire da Silva - CRB 8/5974

So89p Souza, Douglas Lopes de, 1981-  
Processamento de linguagem natural de normas técnicas da construção civil : estudo da NBR 15.575 / Douglas Lopes de Souza. – Campinas, SP : [s.n.], 2024.

Orientador: Regina Coeli Ruschel.  
Tese (doutorado) – Universidade Estadual de Campinas, Faculdade de Engenharia Civil, Arquitetura e Urbanismo.

1. Processamento de linguagem natural (Computação). 2. Extração de informações. 3. Regulamentos. 4. Construção civil - Normas - Brasil. I. Ruschel, Regina Coeli, 1958-. II. Universidade Estadual de Campinas. Faculdade de Engenharia Civil, Arquitetura e Urbanismo. III. Título.

Informações Complementares

**Título em outro idioma:** Natural language processing of technical standard for the construction sector : the study of the NBR 15.575

**Palavras-chave em inglês:**

Natural language processing (Computer science)

Information extraction

Regulations

Civil construction - Standards - Brazil

**Área de concentração:** Arquitetura, Tecnologia e Cidade

**Titulação:** Doutor em Arquitetura, Tecnologia e Cidade

**Banca examinadora:**

Regina Coeli Ruschel [Orientador]

Cláudio Vicente Mitidieri Filho

Carlos Torres Formoso

Dayana Bastos Costa

Sérgio Leal Ferreira

**Data de defesa:** 14-03-2024

**Programa de Pós-Graduação:** Arquitetura, Tecnologia e Cidade

**Identificação e informações acadêmicas do(a) aluno(a)**

- ORCID do autor: <https://orcid.org/0000-0001-5422-0533>

- Currículo Lattes do autor: <http://lattes.cnpq.br/5650168725069185>

**UNIVERSIDADE ESTADUAL DE CAMPINAS  
FACULDADE DE ENGENHARIA CIVIL, ARQUITETURA E  
URBANISMO**

**PROCESSAMENTO DE LINGUAGEM NATURAL DE NORMAS  
TÉCNICAS DA CONSTRUÇÃO CIVIL: ESTUDO DA NBR  
15.575**

**Douglas Lopes de Souza**

**Tese de Doutorado aprovada pela Banca Examinadora, constituída por:**

Prof. Dr<sup>a</sup> Regina Coeli Ruschel  
**Presidente e Orientadora FAUFEC / Unicamp**

Prof. Dr. Claudio Vicente Mitidieri Filho  
**Instituto de Pesquisas Tecnológicas**

Prof. Dr. Carlos Torres Formoso  
**Universidade Federal do Rio Grande do Sul**

Prof<sup>a</sup>. Dr<sup>a</sup> Dayana Bastos Costa  
**Universidade Federal da Bahia**

Prof. Dr. Sérgio Leal Ferreira  
**Universidade de São Paulo**

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

Campinas, 14 de março de 2024.

À minha esposa Michele

À minha filha Valentina

Ao meu filho Antônio

**DEDICO**

Ao meu pai Moacil

À minha mãe Ruth

**OFEREÇO**

## Agradecimentos

À minha orientadora, Prof. Regina Coeli Ruschel por ter me guiado durante todo o percurso da pesquisa e me incentivado a desbravar uma área tão estranha à nossa formação.

Aos professores Romis Ribeiro de Faissol Attux, Ricardo dos Santos Ferreira e Julio Cesar Soares dos Reis pelas orientações motivantes e reveladoras no mundo da computação.

À Universidade Estadual de Campinas por ter oferecido gratuitamente os recursos fundamentais para que esta pesquisa pudesse ter sido concluída.

À Universidade Federal de Viçosa pelo incentivo e contribuições para que minha capacitação fosse executada sem prejuízos à pesquisa.

Aos amigos, Vinicius, Thiago, Leandro e Leonardo pelos momentos de fôlego.

Aos meus pais, Moacil Alves de Souza e Ruth Lopes de Souza pelos diversos incentivos ao longo de minha caminhada e pelo exemplo que têm sido na minha vida e na vida de minha família.

Às minhas irmãs, Adriana e Nádia pelos seus exemplos em quem sempre me espelho e aos momentos fraternos.

À minha esposa, Michele Micheleti de Mello, que suportou com muita perseverança todos os momentos difíceis desse período. Sem sua ajuda essa pesquisa não teria acontecido.

Aos meus filhos, Valentina e Antônio por terem suportado minha ausência e por serem meus guias mirins na minha vida.

## RESUMO

A ABNT NBR 15.575 é um dos principais mecanismos disponíveis para promover a qualidade das habitações no Brasil e vem sendo implementada desde 2013. Atualmente a indústria da construção civil nacional é mediada por uma grande quantidade de regulamentos, dificultando o alinhamento entre eles e favorecendo a ocorrência de conflitos na sua aplicação. Acrescido a isso, os regulamentos são documentos complexos que exigem a interpretação por especialistas porque apresentam diversos parâmetros que devem ser verificados durante todo o ciclo de vida da edificação. Os regulamentos são preparados em linguagem natural exigindo uma análise manual, propensa a erros e demandando um alto nível de capacitação para a programação das regras em plataformas dedicadas à avaliação de conformidade automatizada. As pesquisas nacionais relacionadas ao tema se atêm a tentativas isoladas de conversão manual das diretrizes objetivas da norma em regras para plataformas especializadas de avaliação de conformidade. Esta pesquisa tem como objetivo a criação de um método de extração de informações baseado em regras para a tradução automática da Norma de Desempenho (ND) em formato aberto semiestruturado capaz de ser computado ou incorporado em plataformas de verificação automática de regras. A solução classifica-se na área de Processamento de Linguagem Natural da Inteligência Artificial. A pesquisa adota o método *Design Science Research* a fim de orientar a produção de uma sequência de contribuições. O resultado compreende um algoritmo baseado em regras de extração de informações gerando um modelo semântico de representação da ND. O algoritmo é composto por diversas técnicas de Processamento de Linguagem Natural baseados em padrões linguísticos do português. O modelo semântico da Parte 1 da Norma de Desempenho é expresso em linguagem formal, semiestruturada em *Extensible Markup Language*, que pode ser processada em aplicações relacionadas à verificação de conformidade e traduzidas para *Information Delivery Specification*. A acurácia do algoritmo indica a necessidade de incorporação de novos métodos baseados em inteligência artificial, destacando a técnica de *Retrieval-Augmented Generation* em grandes modelos de linguagem. A pesquisa descartou a adoção de uma ontologia visando o aumento do desempenho dos algoritmos de análise e mapeamento textual dada a inexistência de uma ontologia para o contexto da construção civil em português ou a não adequação da adoção de ontologias genéricas ou específicas na língua inglesa. Desta forma, foi apontada com precisão uma lacuna brasileira. A contribuição e ineditismo da pesquisa está na formalização algorítmica de PLN auditável de requisitos em português em linguagem neutra e aberta.

**Palavras-chave:** processamento de linguagem natural, extração de informações, modelo semântico de representação, NBR 15.575, verificação automática de conformidade.

## ABSTRACT

ABNT NBR 15.575 is one of the main mechanisms available to **encoraje** the quality of housing in Brazil and has been implemented since 2013. Currently, the national construction industry is mediated by a large number of regulations, complicating the alignment between them and causing possible conflicts in their application. In addition, regulations are complex documents that require interpretation by experts because they present various parameters that must be verified throughout the building's lifecycle. The regulations are prepared in natural language requiring manual analysis, prone to errors, and demanding a high level of expertise for programming the rules on platforms dedicated to automated compliance checking. National research related to the theme is limited to isolated attempts at manual conversion of the standard's objective guidelines into rules for specialized conformity checking platforms. This research aims to create a rule-based information extraction method for the automatic translation of the Performance Standard (PS) into an open, semi-structured format that can be computed or incorporated into automatic rule verification platforms. The solution is classified in the field of Natural Language Processing of Artificial Intelligence. The research adopts the Design Science Research method to guide the production of a sequence of contributions. The result includes a rule-based information extraction algorithm generating a semantic model representation of the PS. The algorithm comprises various Natural Language Processing techniques based on portuguese linguistic patterns. The semantic model of Part 1 of the Performance Standard is expressed in formal, semi-structured language in Extensible Markup Language, which can be processed in applications related to conformity verification and translated into Information Delivery Specification. The accuracy of the algorithm indicates the need for incorporating new methods based on artificial intelligence, highlighting the Retrieval-Augmented Generation technique in large language models. The research dismissed the adoption of an ontology to increase the performance of text analysis and mapping algorithms given the absence of an ontology for the construction context in portuguese or the inadequacy of adopting generic or specific ontologies in english. Thus, a brazilian gap was precisely identified. The contribution and novelty of the research lie in the algorithmic formalization of auditable NLP of requirements in portuguese in neutral and open language.

**Key-words:** natural language processing, information extraction, semantic representation model, NBR 15.575, automatic compliance checking.

## LISTA DE FIGURAS

Figura 1 - Quantidade de normas existentes por área.....	32
Figura 2 - Lista dos principais softwares usados no mercado da construção civil.....	47
Figura 3 - Conteúdo de cada página da Norma de Desempenho (parte 1 a 6).....	56
Figura 4 - Manuais e guias para esclarecimento da norma.....	59
Figura 5 - Pesquisas e aplicações para verificação automática de conformidade.....	62
Figura 6 - Funcionalidades que devem interagir em um sistema automático.....	63
Figura 7 - Verificação no plugin UpCodes sobre acessibilidade em um banheiro.....	64
Figura 8 - Framework de trabalho no KBIM.....	67
Figura 9 - Processo de conversão manual de um regulamento para formato computável no software KBIM.....	68
Figura 10 - Exemplo de transformação de uma sentença em sintaxe lógica.....	69
Figura 11 - Exemplo de conversão de texto em grafo e linguagem lógica.....	75
Figura 12 - Exemplo de termos resultantes de uma EI e a sua representação lógica.....	76
Figura 13 - Regulamento decomposto e marcado em XML.....	78
Figura 14 - Modelo geral da ontologia IC-PRO-Onto.....	81
Figura 15 - Exemplo visual de um mapeamento entre termos de um regulamento e conceitos da ontologia.....	82
Figura 16 - Uso de uma ontologia para validação dos requisitos.....	83
Figura 17 - Representação abstrata do IFC e dois MVDs.....	88
Figura 18 - IDS dentro de um fluxo típico de projeto BIM.....	88
Figura 19 - Interface de importação de regras do Solibri.....	90
Figura 20 - Exemplo de conversão XML para o programa Autodesk Model Checker Configurator.....	91
Figura 21 - Diferença entre formatos abertos e computáveis.....	92
Figura 22 - Graus de capacidade de processamento das normas ISO.....	94
Figura 23 - Exemplo da aplicação de diferentes categorias de Smart standards na indústria de aviões.....	97

Figura 24 - Exemplo de alinhamento entre uma ontologia (OWL) e outro XML.....	98
Figura 25 - Exemplo de marcação XML na Constituição Federal de 1988.....	100
Figura 26 - Fluxo do uso de padrões para enriquecimento das sentenças.....	102
Figura 27 - Exemplo de análise utilizando o método Análise de Constituintes.....	105
Figura 28 - Diferença entre o resultado de um shallow parsing e um deep parsing.....	106
Figura 29 - Representação de uma sentença anotada com sete termos semânticos.....	108
Figura 30 - Árvores de constituintes com atributos linguísticos usada para construção dos padrões.....	108
Figura 31 - Exemplo de rotulação de entidades, relações sintáticas e semânticas.....	111
Figura 32 - Exemplo de rotina de busca executada na ontologia.....	112
Figura 33 - Trecho da ontologia UPO.....	113
Figura 34 - Exemplo de sentença do regulamento convertida para classes de variáveis da regra.....	115
Figura 35 - Maiores classes da ontologia de materiais da DICM (Digital Construction Building Material Ontology) existente na DiCon.....	122
Figura 36 - Trecho da ontologia ifcOWL (IFC4_ADD2_TC1).....	122
Figura 37 - Incorporação do NER em extrações de informação.....	124
Figura 38 - Fluxo de trabalho nos dois paradigmas computacionais.....	126
Figura 39 - Acurácia do modelo ChatGPT 3 e de outros modelos de linguagem.....	127
Figura 40 - Comparação de respostas corretas dadas pelos LLM e seres humanos.....	128
Figura 41 - Comparação do desempenho dos LLMs em responder perguntas de dificuldade simples.....	129
Figura 42 - Exemplo de busca em documento usando LLMs e o plugin Retrieval.....	131
Figura 43 - Discriminação das etapas e seus objetivos aplicados a esta pesquisa.....	135
Figura 44 - Etapa de conscientização e o relacionamento de seus objetivos.....	135
Figura 45 - Etapa de sugestão e o relacionamento de seus objetivos.....	137
Figura 46 - Etapa de desenvolvimento e o relacionamento de seus objetivos.....	138
Figura 47 - Etapa de avaliação e o relacionamento de seus objetivos.....	139

Figura 48 - Etapa de conclusão e o relacionamento de seus objetivos.....	140
Figura 49 - Análise multinível proposta por Duque-Ramos et al. (2011) à esquerda, e Lozano-Tello e Gómez-Perez (2004) à direita.....	142
Figura 50 - Gráfico de radar com os resultados das medidas das subcaracterísticas.....	144
Figura 51 - Valores das características por ontologia.....	145
Figura 52 - Similaridade do cosseno.....	156
Figura 53 - Visualização da similaridade do termo "building" na plataforma tensorflow.org.....	157
Figura 54 - Etapas propostas para a transformação da Norma de Desempenho em linguagem computável e os artefatos produzidos.....	165
Figura 55 - Exemplo de trecho do documento armazenado em formato .txt na plataforma Github.....	173
Figura 56 - Trecho da planilha usada para desenvolvimento de padrões.....	180
Figura 57 - Exemplo de padrões linguísticos e seus respectivos rótulos semânticos.....	188
Figura 58 - Exemplos bem-sucedidos da notação XML gerada pelo algoritmo.....	190
Figura 59 - Gráficos das medidas de classificação do Capítulo 8 separadas por frase.....	193
Figura 60 - Gráficos das medidas de classificação da Parte 1 separadas por frase.....	196
Figura 61 - Boxplot com identificação de dados discrepantes.....	199
Figura 62 - Visualização dos dados com identificação dos outliers (em vermelho).....	200
Figura 63 - Gráfico com as medias agrupadas.....	201
Figura 64- Resultado do método da silhueta.....	203
Figura 65 - Avaliação do resultado dos testes de silhueta para grupos 4 a 8.....	204
Figura 66 - Visualização dos grupos 4 a 8.....	205
Figura 67 - Comparação de agrupamento com 8 e 4 grupos.....	206
Figura 68 - Segmento do resultado da análise gramatical e sintática da frase 18.....	208
Figura 69 - Anotação linguística utilizando o dicionário toponímico.....	209
Figura 70 - Uso do dicionário toponímico em conjunto com um tokenizador.....	209
Figura 71 - Alternativa explorada para incorporação de rótulo personalizado.....	210

Figura 72 - Comparação da acurácia do modelo MedPaLM 2 com outros modelos após o ajuste fino.....	215
Figura 73 - Modelo de funcionamento de um RAG comparado com um LLM isolado...	219
Figura 74 - Interface de construção de um IDS pelo IDS Editor.....	228
Figura 75 - Trecho da tela do IDS Editor representando os atributos a serem convertidos.....	229
Figura 76 - Arquivo XML formatado no esquema IDS pelo programa IDSEditor.....	229
Figura 77 - Enriquecimento do arquivo XML com a uri do elemento "estrutura de concreto".....	230
Figura 78 - Tradução do IFC via plataforma Crowdin em novembro de 2023 e fevereiro de 2024.....	260
Figura 79 - Atividade de tradução do IFC para português entre Agosto de 2023 e Fevereiro de 2024.....	261
Figura 80 - Exemplo da rotulação feita pelo SpaCy.....	263
Figura 81 - Comparação entre atributos dos modelos da biblioteca Spacy.....	278

## LISTA DE QUADROS

Quadro 1 - Quadro resumo das dificuldades enfrentadas pelas avaliações de conformidade .....	24
Quadro 2 - Comparação da sintaxe de os modelos baseados em regras e em lógica .....	74
Quadro 3 - Texto rotulado com o protocolo RASE .....	79
Quadro 4 - Grupos de normas organizados pela urgência de digitalização em formato capaz de ser lido por máquina .....	95
Quadro 5 - Exemplo de anotação baseado em padrão gramatical e ontologia .....	113
Quadro 6 - Exemplo de elementos de informação semântica extraídos e convertidos para representação lógica .....	114
Quadro 7 - Composição das variáveis de qualidade de produto .....	116
Quadro 8 - Exemplo de variável de recurso .....	116
Quadro 9 - Rótulos semânticos adotados em pesquisas relacionadas com o enriquecimento semântico de sentenças .....	118
Quadro 10 - Comparativo entre ferramentas de PLN .....	124
Quadro 11 - Exemplo de dados obtidos pelo plugin Retrieval .....	130
Quadro 12 - Principais atributos acessados na OMW .....	153
Quadro 13 - Exploração de synsets da Open Multilingual WordNet .....	152
Quadro 14 - Experimento simplificado para busca de padrões - código em Python .....	163
Quadro 15 - Estrutura de diretrizes da Norma de Desempenho .....	167
Quadro 16 - Trecho da Norma de Desempenho que exemplifica a variabilidade de sua estrutura .....	169
Quadro 17 - Exemplo de uma frase da Parte 1 da Norma em formato original e sua versão de único parágrafo .....	172
Quadro 18 - Exemplo com perda significativa de sentido na remoção de stopwords .....	175
Quadro 19 - Exemplo de frase com pouca perda de sentido na remoção de <i>stopwords</i> .....	175
Quadro 20 - Casos em que a supressão de stop words afetou o resultado .....	176

Quadro 21 - Trecho da formatação ConLL-U usada para extração de padrões.....	177
Quadro 22 - Comparação entre rotulagem em cada modelo de linguagem.....	178
Quadro 23 - Trecho da planilha de elementos relacionados às referências de normas e componentes do edifício.....	181
Quadro 24 - Lista de rótulos e padrões aplicados na Parte 1 da Norma de Desempenho .....	182
Quadro 25 - Verbos modais em português e inglês.....	183
Quadro 26 - Pseudocódigo do algoritmo para a correspondência de padrões .....	184
Quadro 27 - Problema inicial da busca por padrões .....	185
Quadro 28 - Exemplos de padrões compostos por expressões regulares .....	186
Quadro 29 - Trecho da planilha com resultado da extração do Capítulo 8 da Norma de Desempenho.....	187
Quadro 30 - Árvore de elementos XML obtida na extração.....	189
Quadro 31 - Pseudocódigo do método de conversão dos termos extraídos em estrutura XML .....	189
Quadro 32 - Matriz de confusão para qualificação do resultado .....	191
Quadro 33 - Exemplo de avaliação das sentenças extraídas do Capítulo 8 .....	191
Quadro 34 - Cálculo das métricas de qualidade .....	192
Quadro 35 - Exemplos de avaliação realizada sobre textos da Parte 1.....	195
Quadro 36 - Grupo 2, frases com menor capacidade de identificação pelos padrões....	207
Quadro 37 - Comparação das abordagens para extração de informações de um trecho da Norma de Desempenho.....	213
Quadro 38 - Código que acessa o framework LangChain para executar uma extração de informações usando LLM.....	221
Quadro 39 - Quadro comparativo do desempenho de ajuste fino e uso do método RAG .....	222
Quadro 40 - Exemplo de frase e seus termos semânticos definidos por Zhang e El-Gohary (2016) .....	223

Quadro 41 - Trecho do JGJ 120-2012 extraído com <i>framework</i> LangChain .....	224
Quadro 42 - Trecho do ICB 2009, extraído com o <i>framework</i> LangChain.....	224
Quadro 43 - experimento inicial para processamento de trecho da Norma de Desempenho .....	262
Quadro 44 - Código para rotulação usando SpaCy .....	263
Quadro 45 - Comparação da acurácia dos métodos de cada modelo de língua portuguesa da biblioteca SpaCy .....	279

## LISTA DE TABELAS

Tabela 1 - Dados originais de avaliação .....	143
Tabela 2 - Outras medidas elementares usadas para discriminação dos resultados.....	146
Tabela 3 - Análise de similaridade de conceitos da OMW .....	154
Tabela 4 - Resultado de similaridade entre três conceitos e alguns termos da Norma..	158
Tabela 5 - Média das métricas do Capítulo 8 da Norma de Desempenho.....	194
Tabela 6 - Média das métricas da Parte 1 .....	197
Tabela 7 - Métricas de qualidade das frases discrepantes número 14, 18 e 30. ....	199
Tabela 8 - Descrição estatística do conjunto de dados .....	199
Tabela 9 - Quadro resumo dos valores dos outliers .....	200

## LISTA DE ABREVIATURA E SIGLAS

ABNT – Associação Brasileira de Normas Técnicas

AEC – Arquitetura, Engenharia e Construção

ANSI - *American National Standards Institute*

ASBEA – Associação Brasileira dos Escritórios de Arquitetura

ARC – *Automated Rule Checking*

BCF – *BIM Collaboration Format*

BCL – *Building Component Lybray*

BDE – *Building Data Exchange*

BIM – Building Information Model / Modeling

BMP – Windows Bitmap

CAD – *Computer Aided Design*

CAU – Conselho de Arquitetura e Urbanismo

CBIC – Câmara Brasileira da Indústria da Construção

COMAT – Comissão de Materiais, Tecnologia, Qualidade e Produtividade

CSV – *Comma-separated values*

CTECH - Comitê Nacional de Desenvolvimento Tecnológico da Habitação

DSR – *Design Science Research*

DXF – *Drawing Exchange Format*

EIR – Requisitos de Troca de Informação

GATE - *General Architecture for Text Engineering*

gbXML – Green Building XML

HC – Cláusulas de Horn

HIS – Habitação de Interesse Social

HTML – *HyperText Markup Language*

IBC – *International Building Code*

ICC - *International Codes Council*

IDF - *EnergyPlus input data file*

IFC - *Industry Foundation Classes*

ISO - *International Organization for Standardization*

JSON - *JavaScript Object Notation*

LD - *Lógica de Descrições*

LEED - *Leadership in Energy and Environmental Design*

LOD - *Level of Development / Level of Detail*

LOI - *Level of Information*

LPO - *Lógica de Primeira Ordem*

MVCNN - *Multi-View Convolutional Neural Network*

MVD - *Model View Definitions*

ND - *Norma de Desempenho*

NIST - *National Institute of Standards and Technology*

OWL - *Web Ontology Language*

PBQP-H - *Programa Brasileiro de Qualidade e Produtividade do Habitat*

PDF - *Portable Document Format*

PIR - *Requisitos de Informação do Projeto*

PLN - *Processamento de Linguagem Natural*

POS - *Part-of-Speech*

RASE - *Requirement, Applicability, Selection e Exception*

RDF - *Resource Description Framework*

SASE - *Standard Analysis, Synthesis and Expression*

SiAC - *Sistema de Avaliação da Conformidade de Serviços e Obras*

SPARQL - *SPARQL Protocol and RDF Query Language*

STEP - *Standard for the Exchange of Product Data*

URI - *Uniform Resource Identifier*

W3C - *World Wide Web Consortium*

XML - *Extended Markup Language*

# SUMÁRIO

<b>1. INTRODUÇÃO .....</b>	<b>22</b>
1.1. PROBLEMÁTICA.....	24
1.1.1. <i>Tarefa manual</i> .....	25
1.1.2. <i>Ambiguidade dos regulamentos e os limites de codificação</i> .....	28
1.1.3. <i>Grande quantidade de regulamentos</i> .....	32
1.1.4. <i>Diferentes formatos de arquivos e estruturas de dados</i> .....	35
1.1.5. <i>Expressão semântica do regulamento no esquema IFC</i> .....	37
1.2. NOVAS ABORDAGEM PARA O PROCESSAMENTO DE REGULAMENTAÇÕES .....	40
1.3. QUESTÕES DE PESQUISA .....	42
1.4. OBJETIVOS 42	
1.4.1. <i>Objetivos específicos</i> .....	43
1.5. HIPÓTESE 43	
1.6. JUSTIFICATIVA 43	
1.6.1. <i>Adoção do BIM no cenário nacional e a baixa digitalização da construção</i> .....	45
1.6.2. <i>Promoção da qualidade da habitação com a adoção da norma NBR 15.575</i> .....	48
1.7. ESTRUTURA DA TESE.....	51
<b>2. REVISÃO BIBLIOGRÁFICA .....</b>	<b>53</b>
2.1. NBR 15.575:2021 .....	55
2.1.1. <i>Dificuldades da incorporação da Norma de Desempenho</i> .....	58
2.2. SISTEMA DE CHECAGEM AUTOMÁTICA COMO RESPOSTA .....	60
2.3. PLATAFORMAS PÚBLICAS PARA VERIFICAÇÃO AUTOMÁTICA DE CONFORMIDADE .....	64
2.4. MODELOS DE REPRESENTAÇÃO COMPUTÁVEL PARA REGULAMENTOS .....	70
2.4.1. <i>Modelos regrados</i> .....	70
2.4.2. <i>Modelos lógicos</i> .....	73
2.4.3. <i>Modelos semânticos</i> .....	78
2.4.4. <i>Modelos ontológicos</i> .....	80
2.5. REPRESENTAÇÃO DE REGULAMENTOS EM BIM .....	87
2.6. DADOS LEGÍVEIS POR MÁQUINA E A ISO SMART .....	92
2.6.1. <i>Linguagem de marcação XML</i> .....	96
2.7. MÉTODOS DE EXTRAÇÃO DE INFORMAÇÕES .....	101
2.7.1. <i>O Processamento de Linguagem Natural</i> .....	101
2.7.2. <i>Procedimentos de extração de informação baseada em regras</i> .....	101

2.7.3. Ontologias abertas relacionadas com a construção civil.....	120
2.7.4. Ferramentas utilizadas para extração e mapeamento semântico baseados em regras .....	123
2.7.5. Procedimentos de extração de informações baseados em aprendizado de máquina .....	125
<b>3. MÉTODO .....</b>	<b>133</b>
3.1. SOBRE O DESIGN SCIENCE RESEARCH .....	133
3.1.1. O delineamento da DSR adotado.....	134
<b>4. CONSCIENTIZAÇÃO .....</b>	<b>141</b>
4.1. ESTUDO COMPARATIVO DE ONTOLOGIAS EXISTENTES .....	141
4.2. RESULTADO DA SELEÇÃO .....	150
4.3. EXPLORAÇÃO DA ONTOLOGIA WORDNET PARA DESAMBIGUAÇÃO DOS TERMOS DA NORMA DE DESEMPENHO .....	152
4.4. EXPLORAÇÃO DA ANÁLISE DE SIMILARIDADE DE COSSENO PARA DESAMBIGUAÇÃO SEMÂNTICA.....	155
4.5. AVALIAÇÃO DA ABORDAGEM BASEADA EM ONTOLOGIA .....	159
<b>5. SUGESTÃO .....</b>	<b>161</b>
5.1. SUGESTÃO DE FERRAMENTAS .....	162
5.2. PREPARAÇÃO DO TEXTO E ELABORAÇÃO DE PADRÕES .....	163
5.3. CONVERSÃO EM LINGUAGEM COMPUTÁVEL .....	164
5.4. SUGESTÃO DE PROCEDIMENTOS PARA EXTRAÇÃO DE INFORMAÇÕES.....	164
<b>6. DESENVOLVIMENTO .....</b>	<b>167</b>
6.1. SELEÇÃO E ANÁLISE DO TRECHO DA NORMA .....	167
6.2. PRÉ-PROCESSAMENTO.....	170
6.2.1. Conversão do documento em PDF para TXT.....	170
6.2.2. Preparação do texto .....	171
6.2.3. Remoção de legendas e capitulação para segmentação do texto .....	173
6.2.4. Remoção de stop words e pontuação.....	174
6.3. EXECUÇÃO DA EXTRAÇÃO DE INFORMAÇÕES.....	177
6.3.1. Rotulagem de atributos linguísticos .....	177
6.3.2. Consolidação de padrões de extração .....	179
6.3.3. Avaliação da incorporação de dicionário toponímico.....	183
6.3.4. Extração dos termos e revisão dos padrões.....	183
6.4. GERAÇÃO DE LINGUAGEM COMPUTÁVEL .....	188
<b>7. AVALIAÇÃO .....</b>	<b>191</b>
7.1. APLICAÇÃO DE MÉTRICAS NO RESULTADO DO CAPÍTULO 8 DA NORMA DE DESEMPENHO .....	191

7.2. APLICAÇÃO DE MÉTRICAS NO RESULTADO DA PARTE 1 DA NORMA .....	194
7.3. AVALIAÇÃO DA ORDEM DE BUSCA .....	197
7.4. ANÁLISE DOS RESULTADOS .....	198
7.4.1. Agrupamento de sentenças de mesma característica pelo método de clusterização .....	202
7.5. COMPARATIVO LLM .....	212
7.5.1. Análise da abordagem baseada em grandes modelos de linguagem (LLMs).....	212
7.5.2. Síntese do resultado da comparação com LLM .....	213
7.5.3. Ajuste fino dos grandes modelos de linguagem .....	214
7.5.4. Uso de uma base de conhecimento, a aplicação do framework LangChain.....	219
7.6. CAPACIDADE DE ENRIQUECIMENTO DO MODELO COM ATRIBUTOS IFC.....	225
7.6.1. Construção de IDss .....	227
<b>8. DISCUSSÃO.....</b>	<b>231</b>
<b>9. CONCLUSÃO .....</b>	<b>237</b>
<b>REFERENCIAS .....</b>	<b>241</b>
<b>APÊNDICE A – ESTÁGIO DE TRADUÇÃO DO IFC PARA PORTUGUÊS (PT-BR).....</b>	<b>260</b>
<b>APÊNDICE B - EXPERIMENTOS INICIAIS PARA AVALIAÇÃO DAS FERRAMENTAS DE PLN .....</b>	<b>262</b>
<b>APÊNDICE C - NBR 15575-1, CAPÍTULO 8 – SEGURANÇA CONTRA INCÊNDIO.....</b>	<b>266</b>
<b>APÊNDICE D –TRECHO PREPARADO PARA PROCESSAMENTO (CAPÍTULO 8 DA NORMA DE DESEMPENHO).....</b>	<b>268</b>
<b>APÊNDICE E – TRECHO PREPARADO DA PARTE 1 DA NORMA DE DESEMPENHO .....</b>	<b>270</b>
<b>APÊNDICE F – LISTA DE ONTOLOGIAS AVALIADAS .....</b>	<b>276</b>
<b>APÊNDICE G – COMPARAÇÃO DOS MODELOS DE LINGUAGEM DA BIBLIOTECA SPACY .....</b>	<b>278</b>
<b>APÊNDICE H – PADRÕES IDENTIFICADOS NO TEXTO DO CAPÍTULO 8 DA NORMA.....</b>	<b>280</b>
<b>APÊNDICE I - PROBLEMA EXTRAÇÃO DE PADRÕES.....</b>	<b>284</b>
<b>APÊNDICE J- RESULTADO DE EXTRAÇÃO DE INFORMAÇÕES - CAPÍTULO 8 DA NORMA DE DESEMPENHO .....</b>	<b>286</b>
<b>APÊNDICE K - RESULTADO DE EXTRAÇÃO DE INFORMAÇÕES - PARTE 1 DA NORMA DE DESEMPENHO .....</b>	<b>288</b>
<b>APÊNDICE L – COMPARAÇÃO SOBRE ALTERAÇÃO NA ORDEM DOS PADRÕES.....</b>	<b>293</b>
<b>APÊNDICE M - TRECHO DO REGULAMENTO JGJ 120-2012.....</b>	<b>303</b>
<b>APÊNDICE N – TESTE DE EXTRAÇÃO DE INFORMAÇÕES VIA LLM.....</b>	<b>304</b>
<b>APÊNDICE O – TESTE DE IMPLEMENTAÇÃO DA TÉCNICA RAG (LANGCHAIN) .....</b>	<b>317</b>

## 1. INTRODUÇÃO

A contínua evolução da indústria da construção civil, com a incorporação de novos sistemas digitais de projeto e processos de fabricação de bens sociais e materiais, tem ampliado a demanda por novos sistemas de apoio. Um edifício construído até o final do século passado era projetado e produzido a partir de modelos e desenhos físicos com ajuda da abstração matemática dos cálculos desenvolvidos para testar e comprovar a sua possibilidade de construção. Entretanto, no início desse século, com a evolução da tecnologia da informação e comunicação, o ciclo de vida das edificações incorpora paulatinamente novos sistemas e demandas de desempenho possíveis de serem simulados e verificados em ambiente computacional. Os sistemas atuais ampliam a capacidade projetiva de novas geometrias, de produtos da construção e de outros sistemas de suporte para coordenar e gerir o processo em toda a cadeia de produção.

O aumento acelerado da complexidade do projeto, controle e operação do edifício têm exigido a incorporação de novas tecnologias, mas também exige a adequação dessas tecnologias e processos a rotinas e testes de validação<sup>1</sup> desses produtos. No Brasil, essa avaliação é dada pela conformidade de produtos e processos a normas específicas, elaboradas pela Associação Brasileira de Normas Técnicas (ABNT) em conjunto com a iniciativa pública e privada. Esse foro de normalização ocorre em comissões específicas (Comitês Técnicos) que debatem e elaboram regulamentos visando a manutenção da qualidade, desempenho e repetibilidade de processos e serviços.

Há mais de vinte anos, estudos analisaram o cenário de padronização para inovação e para desenvolvimento econômico (Shin; Kim; Hwang, 2015). As normas

---

<sup>1</sup> Enquanto a verificação trata de avaliar se critérios os objetivos pré-estabelecidos, como normas e regulamentos, foram atendidos, a validação trata de conferir se o modelo garante a expectativa do projeto, em níveis menos objetivos de análise (Kwasniewski; Bojanowski, 2017)

definem um vocabulário comum e características essenciais de um produto ou serviço, fomentando a melhoria da prática num ecossistema para garantir melhores resultados.

Embora os estudos sobre os impactos das normas sobre a inovação tenham, em certa medida, visões conflitantes, eles geralmente concordam que as políticas de padronização têm efeitos positivos sobre a política nacional de inovação para o crescimento econômico (Shin; Kim; Hwang, 2015, p. 155, tradução nossa).

As normas nacionais são criadas por equipes técnicas de assessoramento às ABNT por meio de grupos de trabalho com especialistas. O texto final que é publicado é redigido de forma a ser compreendido da melhor forma possível por técnicos daquele domínio do conhecimento. A fim de serem compreendidas, as normas se estruturam com tabelas, listas, desenhos e, principalmente, texto em linguagem natural.

Os projetos, de média ou alta complexidade, dependem de uma elaboração contínua de informações destinadas a fins distintos como para obra, para sistemas, para licitação, para aprovação em agências e instituições, para produção e para documentação. São processos que dependem de um fluxo constante de informações elaboradas pelos projetistas e fornecedores ao longo do ciclo de projeto, construção e operação e que estão condicionadas por normas específicas.

A avaliação de conformidade é uma tarefa essencial na garantia de fornecimento de um produto adequado às expectativas do cliente, mas não é uma tarefa trivial. A construção civil é regulada por padrões, normas e regras fragmentadas, e até conflitantes, que serão usadas na etapa de verificação de conformidade (*compliance checking*) do projeto (Nawari, 2018).

Dessa forma, a avaliação de conformidade é uma tarefa de análise contínua que não ocorre apenas em uma fase inicial do ciclo de vida da construção de uma edificação, mas ocorre em momentos decisivos ao longo de todo o ciclo, como por exemplo na fase de demolição em que o projeto de demolição deve garantir segurança e ocupantes de edificações vizinhas e do ambiente construído ao redor (Dimyadi; Amor, 2013).

De um modo geral, o processo de verificação automática de regras envolve a operação de um *software* que não modifica o projeto do edifício, mas acessa as

configurações ou parâmetros dos objetos do modelo, suas restrições e condições validando o resultado em comparação com regras e normas existentes. O resultado dessa análise são afirmações textuais do tipo “*pass*”, “*fail*” e “*warning*” ou “*unknown*” (Eastman *et al.*, 2009).

## 1.1. Problemática

No caso da codificação de regulamentos para programas de verificação de conformidade, os problemas surgem a partir da tradução de diretrizes textuais para processos de análises objetivas de projeto. Atualmente, no Brasil não existem regulamentos disponíveis em formato digital marcado ou em formato semiestruturado, portanto, não é possível existir um sistema de verificação automática de conformidade sem a intervenção humana para codificar suas regras.

Os problemas que serão apresentados a seguir se organizam em torno da correta inserção de informações no processo, seja pela correta conversão dos regulamentos em formato computável ou a correta preparação do modelo com informações adequadas e consistentes para sua verificação. A análise de bibliografias correlatas permitiu organizar sete categorias de problemas relacionados à implementação de sistemas de avaliação de conformidade, resumidas no Quadro 1 e detalhadas nos tópicos a seguir.

**Quadro 1 - Quadro resumo das dificuldades enfrentadas pelas avaliações de conformidade**

Problema	Descrição
Tarefa manual	É um processo dependente da interpretação e codificação especializada em <i>software</i> . Onera as empresas pelo tempo de conversão dos textos, análises e espera em validação pelos órgãos públicos.
Ambiguidade, subjetividade e dificuldade de formalização	A língua escrita é naturalmente ambígua e os esforços dispendidos para criar diretrizes mais objetivas não atingem mais do que 50% de capacidade de transposição para regras em programas de verificação como Solibri.
Grande quantidade de documentos	Em 2021 foram computadas 2.300 normas publicadas, que afetam a construção civil.
Diferentes formatos de arquivos e estruturas de dados	Os dados de análise são armazenados em planilhas, arquivos vetoriais como DXF, CSV e em formatos nativos de BIM.
Expressão semântica do regulamento no esquema IFC	Ainda que as regras e diretrizes sejam traduzidas para a estrutura de dados dos programas de verificação, o IFC não possui capacidade semântica suficiente para representar todas os parâmetros dos regulamentos.

Fonte: o autor.

### 1.1.1. Tarefa manual

O primeiro problema surge com a tradução do regulamento em um formato estruturado capaz de ser processado consistentemente por um programa. Na atualidade, esta tarefa é amplamente dependente da ação humana, manual e parcialmente automatizada pelos programas disponíveis. O caso relatado por Sawyer (2012), mencionado nas pesquisas de diversos autores (Dimyadi; Amor, 2013; Hjelseth, 2015; Zhang; El-Gohary, 2017b; Soliman-Junior; Formoso; Tzortzopoulos, 2019), evidencia o problema da avaliação manual de projeto.

Sawyer (2012) apresenta o problema da verificação manual pela experiência de implementação do projeto AutoCodes criado em 2011 pela Fiatech (*Fully Integrated and Automated Technologies for Construction*). Ao tentar implementar a prova de conceito de seu projeto, a Fiatech solicitou a seus funcionários que fizessem análise de conformidade dos acessos de uma loja de varejo a ser implementada em todo país. O projeto continha 137 folhas e a avaliação seria feita por critérios “relativamente objetivos”. O processo consistia em fazer a verificação manual de documentos 2D e comparar com o desempenho do AutoCodes em avaliar o projeto em 3D. Porém, a prova de conceito não pôde ser conduzida porque a variação das análises feitas pelos funcionários teve alta variabilidade. Havia análises que encontraram 43 problemas enquanto outras apenas 1 e a empresa compreendeu que a análise manual era completamente dependente da capacidade humana em interpretar as diretrizes, regulamentos e sua relação nos desenhos impressos.

A verificação manual é um trabalho extensivo e cuidadoso de profissionais capacitados em interpretação do projeto e com conhecimento dos regulamentos (Eid; Abounaga; Mahmoud, 2019). Muitas vezes esse trabalho é realizado sobre documentos textuais e desenhos impressos ou em formato digitais fechado, ou sem capacidade de armazenar informações não-geométricas e metadados como arquivos do tipo *Computer Aided Design* (CAD) e *Portable Document Format* (PDF). Nesse caso, o projetista ou analista necessita fazer a conversão e interpretação das informações de um documento para um

programa (ou planilha). Isso gera uma demanda maior sobre os projetistas à medida que a quantidade de normas é crescente.

Um objetivo comum a todas as pesquisas relacionadas ao desenvolvimento de alternativas à verificação manual de regras é o aumento da produtividade orientando a ação humana no processo. Nestes casos a verificação manual é trabalhosa, possui alto custo e propensa a erros (Dimiyadi; Amor, 2013).

Segundo Macit (2014) e Solihin (2004), a verificação de conformidade ainda precisa lidar com ambiguidades e inconsistências das avaliações manuais gerando atrasos no processo geral da construção. Segundo os autores, a tarefa manual ainda pode avaliar incorretamente os projetos concluindo em menor desempenho do edifício e em erros que terão alto custo para serem corrigidos.

Fenves *et al.* (1995), Beach *et al.* (2015), Zhang e El-Gohary (2011, 2012) apontam para o fato de que os regulamentos (*regulations*) são complexos e exigem conhecimento técnico profundo do domínio de operação dessas normas. Segundo os referidos autores, essa complexidade do texto das normas é uma barreira para sua automação devido à necessidade de muito tempo para desenvolver uma tradução em rotinas de busca e verificação automatizada. De acordo com Zhang e El-Gohary (2011, 2012), somente com um método de extração automática dos requisitos normativos, seria possível alcançar a automação completa do processo.

Sobhkhiz *et al.* (2021) explicam que existem diversos problemas na verificação automática de conformidade dos projetos à normas e demais regulamentos. Primeiramente, o processo precisa acompanhar a atualização e surgimento de diversos regulamentos determinados por diferentes agências e em escalas diferentes. Como os regulamentos não são preparados para análise automática, o seu formato de apresentação exige alta capacitação técnica do profissional para a programação dos regulamentos nos sistemas de checagem de projeto. Essa especificidade, e o formato fechado de rotinas de checagem nas soluções disponíveis, dificultam o escalonamento desta prática, uma vez que novos usuários não são familiarizados com a tradução das regras. Ao mesmo tempo,

limita qualquer modificação e atualização dos regulamentos por usuários menos experientes.

A avaliação de conformidade é uma tarefa complexa que envolve diferentes dimensões de problemas que dificultam sua automatização. Como a elaboração de edifícios é uma tarefa que agrega diferentes profissionais que elaboram múltiplos estudos sobre a edificação, o resultado é de alta complexidade. O setor da AEC possui programas específicos que auxiliam o processo de automatização de análises e simulações, mas que não atendem todas as exigências das normas pela dificuldade em transformá-las em regras verificáveis. Dessa forma, os profissionais e consultores operam com processos de análise dependentes da intervenção humana, na maioria de suas etapas.

A inserção manual de informações sobre os regulamentos nos programas de verificação de regras é uma tarefa tediosa, lenta e que demanda um processo de conversão automática das regras (Lee *et al.*, 2016), é ineficiente e de interoperabilidade inadequada (Gallaher *et al.*, 2004). Esses casos podem ser exemplificados com a inserção de informações em diferentes sistemas e programas, utilizando dados obtidos em documentos físicos em arquivos eletrônicos de formato não interoperável.

Ainda que os documentos sejam acessíveis, eles precisam ser compartilhados entre as diversas partes interessadas que, por sua vez, armazenam as informações de diferentes formas. Por vezes, as informações não estão centralizadas sob posse de um único agente que pode não disponibilizar a informação necessária a tempo. Com isso, ocorrem atrasos na capacidade produtiva, necessidade de maiores equipes de projeto para gerenciar as informações e a necessidade de mitigação de problemas decorrentes de informações ausentes, incorretas ou incompletas. No relatório elaborado pelo *National Institute of Standards and Technology* (NIST), Gallaher *et al.* (2004) conseguiram analisar os processos de setenta empresas de projeto e construção de instalações comerciais, institucionais e industriais. Os relatórios mostraram que, em 2002, 40% dos custos envolvendo interoperabilidade inadequada entre arquitetos e engenheiros, durante o ciclo de vida da construção, estão relacionados com a entrada manual incorreta de dados.

### 1.1.2. Ambiguidade dos regulamentos e os limites de codificação

Diversos autores têm uma visão congruente sobre a ambiguidade dos textos de regulamentos. Para Li *et al.* (2021) os processos automáticos são dificultados pela ambiguidade da linguagem natural dificultando a interpretação por sistemas simples de extração de informações. He *et al.* (2009) reforçam a ideia de que a verificação manual é suscetível a erros, uma vez que a regulamentação é registrada em documentos textuais, redigidos e formatados pela linguagem natural, que possui expressões de natureza ambígua e inexata.

Tonso e Nardelli (2015) analisam a aplicação da Modelagem da Informação da Construção (*Building Information Modeling* - BIM) para verificação de desempenho térmico com base na Norma de Desempenho ABNT NBR 155757 (ABNT, 2013). Os autores retratam a dificuldade de inserção de tantos parâmetros manualmente e indicam que o uso do BIM deveria entrar nesse contexto, visando a criação de bancos de dados e informações, os quais gerariam agilidade ao processo de análise e ofereceriam projetos com mais informações, consistência e menos erros.

Segundo Srinivasan e Nawari (2022), existe falta de consistência na interpretação de disposições regulatórias impedindo uniformidade de interpretação adequada porque são documentos escritos e lidos por pessoas. Um exemplo dessa ausência de qualidade semântica é registrado por Li *et al.* (2021):

Um desafio chave (...) é quão certo os termos qualitativos da linguagem natural devem ser interpretados sem ambiguidade, e.g. o documento *Building Code of New Zealand (NZBC) Acceptable Solution D1/AS1* parágrafo 1.1.1 define que [...] deve ser prevista uma via acessível que permita o acesso directo à entrada principal do edifício, sempre que tal seja possível (Li *et al.*, 2021, p. 229, grifo próprio, tradução nossa).

No trecho acima não é claro como interpretar objetivamente “*to give direct access*” e “*where practical*”, dando margem à interpretação das equipes de projeto e dos agentes reguladores e fiscalizadores. Outro exemplo de dificuldade na interpretação de regulamentos é dado pelo trecho a seguir que representa a invalidação de regras em

determinadas circunstâncias ou exceções às regras: “*Except in ‘household units’ or Where permitted by Paragraph 1.3.2, a single isolated step shall not be permitted but the change of level shall be constructed as a ramp (...)*” (Li *et al.*, 2021).

Li *et al.* (2020) abordam o problema da subjetividade dos termos adotando métodos que analisam o termo qualitativo “linha de visão direta”. Eles não alteram os termos da norma nem tentam aprimorar a interpretação destes, mas desenvolvem meios de modelar as qualidades subjetivas do texto como capacidades de visibilidade e movimento nos espaços da edificação. Portanto, acabam por propor soluções baseadas em Isovistas (Benedikt, 1979), já utilizadas em outros paradigmas de análise qualitativa de projeto, como a metodologia Space Syntax (<https://www.spacesyntax.net/>).

As iniciativas das pesquisas nacionais e internacionais para parametrização das regras demonstram a incapacidade de computação completa dos regulamentos em plataformas convencionais.

Silva (2017) e Silva e Arantes (2017) analisam a viabilidade e obstáculos da automatização para verificação dos requisitos da Norma de Desempenho, utilizando o *software* Solibri Model Checker para modelos em BIM. Eles afirmam ser necessária a padronização das diretrizes de modelagem e adoção dos sistemas de classificação da informação da construção que precisam atender aos requisitos da norma. Outro obstáculo é a construção da parametrização das regras preliminares e regras principais que deveriam ser testadas em múltiplos modelos para o aperfeiçoamento da proposta de parametrização. O estudo feito por Silva e Arantes (2017) mostrou uma redução de tempo de conferência de projeto em 60% e apontou 22% de não conformidades a mais do que as encontradas manualmente, o que comprova a melhora em qualidade da verificação computacional. Ademais, apontaram que apenas um terço dos requisitos da norma são passíveis de checagem automática. Como desafio encontrado está a subjetividade de 42% dos requisitos, situação que dificulta a criação de um sistema que dispense a interação humana. Além de reduzir a chance de erros, o uso de um sistema digital torna possível uma verificação sistematizada discriminando os resultados da avaliação para cada

conjunto de informações ou modelos inseridos. Essa avaliação periódica e controlada dos resultados permite ajuste do sistema e maior controle do processo, inclusive a retomada de análises anteriores (Hjelseth, 2009).

Mainardi Neto e Santos (2015), ao analisarem a capacidade de verificação de regras pelo *software* Solibri observaram que 48% das regras da Instrução de projeto básico de arquitetura de estações de Metrô e Monotrilho IP-9.00.00.00/3B2-001 puderam ser verificadas de modo objetivo. Mainardi Neto e Santos (2015) concluem que é necessária a revisão da redação dos documentos regulatórios direcionando-os para sua objetivação.

Mendonça, Manzione e Hjelseth (2020) desenvolveram uma análise e conversão da NBR9050 utilizando a metodologia RASE (*Requirement, Applicability, Selection e Exception*) e o *software* Solibri. Como resultado mostraram que essa norma exigiu a reordenação de 65,9% dos textos e apenas 9,5% foram capazes de ser traduzidos diretamente para a metodologia RASE. Como resultado, 28,9% dos termos da NBR9050 não são capazes de ser configurados no Solibri.

Já İlal e Günaydin (2017) desenvolveram um método de representação do *İzmir Municipality Housing and Zoning Code* (IMHZCode) e identificaram que 92% das regras eram formalizadas ou semiformalizadas. As regras formalizadas eram textos explícitos, objetivos, e poderiam ser “claramente representadas” em formato computável. As regras foram representadas em uma única etapa utilizando o método que foi desenvolvido pelos autores. Como resultado, foi possível representarem 79% das 258 regras em regras quantitativas e comparativas capazes de serem processadas pelo computador.

Kehl e Isatto (2015) desenvolveram um estudo de caso que relatou a experiência de uma empresa na análise de conformidade em que 363 requisitos identificados para seus projetos eram registrados em uma planilha eletrônica e, posteriormente, traduzidas para a sintaxe existente no software. Dessa forma, foi necessária a assimilação de todas as regras do Solibri Model Checker para ser possível atribuí-las à análise, ou seja, demandou a contratação de um profissional capacitado e rigoroso na análise. Porém, mesmo com a sistematização dos parâmetros, os autores verificaram que dos 54 requisitos da produção

identificados, somente 13 (25%) desses eram passíveis de serem traduzidos para uma linguagem adequada para a aplicação e, destes, somente 8 puderam efetivamente ser transformados em regras.

Isso ocorreu devido à incompatibilidade semântica dos dados de entrada com o limite de customização das regras existentes no software. Foi possível observar que o tipo de regra tinha grande influência sobre a possibilidade de tradução. Aqueles requisitos que tratavam características dimensionais foram facilmente traduzidos em regras. Já os requisitos não-dimensionais mostraram-se bem mais difíceis de serem traduzidos em regras a serem aplicadas pelo Solibri (Khel; Isatto, 2015, p. 8, tradução nossa).

Khel e Isatto (2015) ainda mostram a limitação do Solibri em oferecer diferentes conjuntos de regras para além das preexistentes por ser um programa *hard-coded* de código fechado com relativa personalização por meio de *scripts*. A implementação de um sistema de verificação automática de conformidade, como o Solibri, tem um custo elevado porque depende da presença de um especialista com formação e experiência na área e no domínio do regulamento, para poder aplicar e traduzir os termos do regulamento em critérios objetivos para o projeto.

(...) o desenvolvimento de um sistema ARC [*Automated Rule Checking*] requer especialistas treinados que entendam os requisitos normativos e os processos de projeto, e estejam equipados com habilidades técnicas de desenvolvimento. Um desenvolvimento bem-sucedido requer a participação e o engajamento de partes que, por natureza, são fragmentadas e distanciadas. Isso pode induzir investimentos iniciais substanciais que podem não ser viáveis dado o orçamento já apertado com o qual os projetos de projeto e construção geralmente lidam (...) (Sobhkhiz et al., 2021, p. 5, tradução nossa).

Os problemas descritos acima são compartilhados por diversos autores como Dimyadi e Amor (2013), Zhang e El-Gohary (2015), Beach *et al.* (2015), Preidel e Borrmann (2018), Srinivasan e Nawari (2022), Fuchs e Amor (2021), Li *et al.*, (2021) e Solibhkhiz *et al.* (2021).

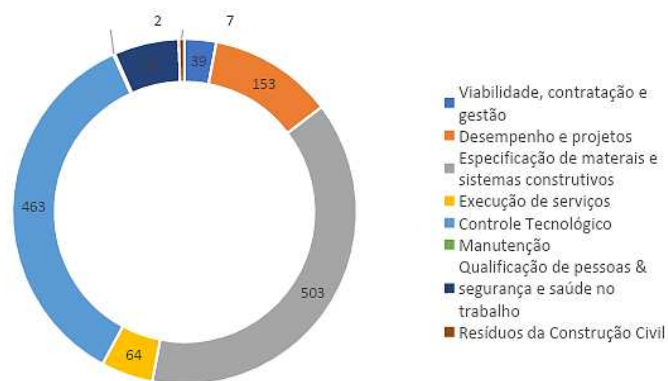
### 1.1.3. Grande quantidade de regulamentos

Mesmo que a prática de verificação seja automatizada e seus instrumentos estejam alinhados em processos mais simplificados, ainda existe uma quantidade elevada de normas a serem verificadas. A quantidade de normas exigidas para construção de edifícios foi quantificada pela Câmara Brasileira da Indústria da Construção (2022) e problematizada em diferentes artigos internacionais.

Zhang e El-Gohary (2016) evidenciam esse problema no estudo do *International Building Code* (IBC) de 2006 que contém 329 seções e, cada uma delas, se divide em dezenas de cláusulas que contém uma variedade de exigências para construção de edifícios nos Estados Unidos. Diferente do caso norte-americano, no Brasil existe a vantagem de que as normas técnicas são elaboradas por comissões especiais de um único órgão, reduzindo a necessidade de acompanhar diversas agências ou associações quando ocorre a atualização ou implantação de novas regras.

Segundo a Comissão de Materiais, Tecnologia e Produtividade da Câmara Brasileira da Indústria da Construção (COMAT/CBIC (2021) em 2021 existiam publicadas 2300 normas que afetam o setor da construção civil, das quais 1319 normas se aplicam às edificações e, destas, 154 estão relacionadas a ‘Desempenho e Projetos’ de edificações (CBIC, 2022) (Figura 1).

Figura 1 - Quantidade de normas existentes por área



Fonte: adaptado de CBIC (2022).

Acrescenta-se a este cenário quantitativo o grande número de normas que são alteradas a cada ano. Entre janeiro e outubro de 2021 foram alteradas 203 normas técnicas relacionadas com o setor da construção civil, 181 foram publicadas e 22 canceladas. Além delas, outras 200 normas relacionadas ao setor estavam em consulta pública nacional e outras 3 foram validadas tecnicamente, o que resulta em um processamento de 606 normas no ano de 2021.

Por outro lado, a quantidade de formatos e semânticas existentes nas cláusulas pode variar até mesmo entre capítulos do mesmo regulamento. Existem formatos de tabelas, quadros, linguagem natural e nem sempre organizados segundo uma estrutura coerente ou desejável para o analista. No caso brasileiro, a NBR 15575, por exemplo, determina eficiência acústica dos itens e projetos, mas esse tema é registrado por sistemas construtivos nos diferentes capítulos ou partes da Norma (parte 2 – sistemas estruturais, parte 3 – sistemas de pisos, parte 4 – sistemas de vedações verticais internas e externas, parte 5 – sistemas de coberturas, parte 6 – sistemas hidrossanitários). A diferença de formatação e semântica entre regulamentos, aliada à quantidade de textos e critérios e à complexidade da informação descrita, impactam no processo manual de verificação gerando alto custo de contratação devido ao tempo necessário para validação dos projetos por equipes de especialistas (Zhang; El-Gohary, 2016).

A alta produção de normas e outros regulamentos como códigos de obras, planos diretores e padrões de projeto e construção desenvolvidos pelas incorporadoras e construtoras, ampliam de modo compulsório a complexidade das edificações exigindo que seus produtos e processos atendam à critérios estabelecidos pelos documentos. Em algumas empresas, a verificação de conformidade é destinada ao setor de qualidade do produto e demanda uso de uma ampla gama de programas e processos de simulação e verificação. Então, a longo prazo, e considerando a grande quantidade de atualizações, espera-se que anualmente esses modelos tenham necessidade de ser atualizados por especialistas da área. Em um cenário de longo prazo esses custos se tornam inviáveis (Zhang; El-Goharay, 2016).

Ainda que fosse possível incorporar as diretrizes dos regulamentos em um programa, de modo rápido e consistente, os documentos sofrem atualizações e demanda incorporação de regras e ajustes subjacentes à regra. Estes ajustes podem ser parte de suposições feitas pelo programador e necessárias para plena codificação nos programas de verificação (Sobhkhiz *et al.*, 2021).

No entanto, o desafio não se limita apenas aos desenvolvimentos iniciais; manter sistemas ARC [*Automated Rule Checking*] também pode se tornar um desafio. Dado os crescentes avanços tecnológicos na indústria AEC, um produto ARC pode se tornar obsoleto cedo demais. Por exemplo, uma mudança nos formatos de dados em que um sistema ARC se baseia pode impor redesevolvimentos significativos no design de todo o sistema. Além disso, o desenvolvimento de um sistema ARC geralmente requer certas suposições com base nas regras de destino. Esses pressupostos podem não estar de acordo com outras regras ou mesmo entrar em conflito com elas. Consequentemente, expandir o sistema para incluir outras regras não é feito facilmente e o sistema se torna não escalável. Outra consequência pode ser que o sistema ARC pode ser totalmente inútil para uma região diferente com pressupostos regulatórios diferentes, e o sistema se torna não generalizável (Sobhkhiz *et al.*, 2021, p. 5, tradução nossa).

Aydin e Yaman (2019) explicam que as normas existentes na indústria AEC são documentos escritos e autorizados para serem compreendidos e implementados por especialistas legais e profissionais da construção civil. Como esses documentos não são definidos seguindo uma lógica formal, “apenas especialistas podem interpretar esses documentos e traduzi-los em notações formais dentro de aplicações de *software*”. Fuchs e Amor (2021) ainda apontam para o fato de que devido à complexidade dos regulamentos e sua terminologia de domínio específico, é difícil garantir a qualidade e consistência das traduções desses documentos em *software* específicos. A solução anunciada por Yurchyshyna *et al.* (2007) seria a transformação dos regulamentos em formato digital, estruturadas e compreendidas por máquinas e representadas em formato padrão interoperável.

#### 1.1.4. Diferentes formatos de arquivos e estruturas de dados

Tradicionalmente, a verificação de conformidade de projetos de edificações aos regulamentos ocorre de modo visual por meio de conhecimento de profissionais técnicos ou *experts* do setor. São avaliados desenhos bidimensionais impressos ou em formatos digitais não estruturados (p.ex.: PDF) e planilhas com dados tabulados que reúnem todas as informações necessárias sobre os materiais e as dimensionais das soluções projetadas. Esse processo envolve diferentes entidades ou agências regulatórias que seguem diretrizes urbanísticas ou normativas locais, estaduais e federais para certificar ou validar as soluções projetadas.

Muitas vezes esse processo demanda diversas revisões causadas por informações ausentes, inadequadas ou incorretas, mas também são necessárias revisões a cada correção de erros apontados pelos órgãos regulatórios. Apesar desses erros serem minimizados pela incorporação de sistemas de Tecnologia da Informação que permitem maior transparência e orientação no processo de submissão de arquivos e dados digitais, a não utilização de práticas e tecnologias mais adequadas impede a ampliação da eficiência desse processo. Com a incorporação do *Building Information Model* (BIM) os modelos ficam aptos a serem coordenados pelas equipes de projeto antes da verificação de sua conformidade às leis e regras às quais o projeto é submetido. Além dessa etapa de coordenação, os modelos de informação são desenvolvidos como um banco de dados do projeto. Dessa forma, é possível gerar arquivos com dados estruturados ou semiestruturados para serem lidos por *softwares*, plataformas ou sistemas dedicados à verificação de conformidade com os regulamentos. Atualmente, essa verificação pode ocorrer de duas formas, ou pela análise das próprias equipes do escritório ou por subcontratação de uma empresa de certificação que verifica, valida e, dependendo do regulamento, certifica a empresa pelo atendimento completo ou parcial à normas nacionais (NBR) ou internacionais (ISO).

Porém, mesmo com a adoção do BIM, as práticas de análise de conformidade, principalmente no âmbito do desempenho energético, ainda são desenvolvidas em diferentes tecnologias, nem sempre aptas à troca irrestrita de informações entre programas e plataformas.

Okamoto (2015) desenvolveu, durante os anos de 2013 e 2014, quatro levantamentos com atores da indústria da construção civil, a saber, incorporadoras e construtoras, projetistas, fabricantes e auditores ou consultores de Gestão da Qualidade. Naquele momento os projetos dessas empresas ainda estavam baseados em arquivos CAD, e não em BIM, e no cenário geral da pesquisa foi identificado haver pouco conhecimento e incorporação de normas técnicas nos processos de produção dos produtos. A pesquisa indicou que as empresas utilizavam planilhas eletrônicas para verificação da conformidade e compatibilização de projeto na etapa de concepção e projeto executivo.

Já em 2020, mesmo com ampla difusão e adoção do BIM para cenários de análise do desempenho energético das edificações, a *Building Data Exchange* (BDE, 2020) verificou que 70% de profissionais que desenvolvem simulações energéticas, usam planilhas para organizar seus modelos de dados. 50% usam BIM, 42% gbXML e 38% DXF e 25% dos profissionais usam *Building Component Library* (BCL). Os formatos diferentes menos usados são variados como: Comma-separated Values (CSV), Windows Bitmap (BMP), Intermediate Data Format (IDF), Open Studio, Python entre outros.

Na pesquisa desenvolvida pelo BDE, 90% dos entrevistados relataram que os dados coletados ficam catalogados em planilhas e que apenas 35% dos entrevistados compartilham seus resultados com agências municipais ou autoridades de classificação do edifício. Ainda questionados sobre os fluxos que poderiam ser melhorados com uso de modelos de dados melhores, cerca de 60% dos entrevistados indicaram que haveria melhoria na inserção de dados de desempenho dos equipamentos, a transferência de informação de projeto para análise de conformidade, o controle de qualidade e depuração

do modelo. Cerca de 58% dos entrevistados indicaram que a inserção de especificações do edifício seria melhorada caso houvesse melhores modelos de dados.

Na mesma pesquisa, os representantes das empresas indicaram que arquivos em formatos de planilhas, BIM, gbXML, DXF e IFC são considerados, nessa ordem, os modelos de dados mais úteis. Um dos fluxos indicados pelas respostas expressa essa complexidade em análise de conformidade em cenário de certificação LEED:

Descreva brevemente como você faz uso dos modelos de dados. (...) Geometria da construção: BIM & gbXML; Também usamos modelos de dados personalizados (via planilha e codificação python) para traduzir informações como modelos térmicos (ganhos internos e cronogramas), propriedades do sistema de cargas HVAC, etc. de planilhas de > de projeto > software de modelagem; Também usamos modelos de dados personalizados para ir de resultados de simulação > relatórios (relatórios de clientes, relatórios de conformidade, documentação LEED/90.1) (BDE, 2020, p. 8, tradução nossa).

Então, dentre outras conclusões, esse relatório indica a necessidade de padronização de modelos de dados, uma vez que é necessária a utilização de diferentes formatos e sistemas para análise.

### **1.1.5. Expressão semântica do regulamento no esquema IFC**

Ainda que seja possível adotar um único modelo de dados para troca de arquivos entre plataformas, dificilmente o modelo contempla todos os parâmetros de projeto exigidos pelos regulamentos. O IFC é a estrutura de dados para interoperabilidade entre projetos BIM mais adotada e completa no estado atual da arte. IFC *schema* é um esquema de dados, padronizado, e dado pela ISO 16739-1:2018. Ele é uma estrutura geral ampla e não específica, para troca de dados de elementos construtivos. Havendo necessidade de elementos novos ou específicos, é necessário vincular esses elementos à estrutura do IFC que possui parâmetros para receber elementos personalizados.

Um limite de representação de conceitos que está relacionado ao esquema de dados IFC é que ele contempla parâmetros apenas para a língua inglesa. Existe uma

iniciativa<sup>2</sup> liderada pela BuildingSMART para tradução do IFC em outras línguas, e que ganhou expressivo desenvolvimento entre 2022 e 2024 (ver Apêndice A), mas ainda não foi concluída.

Venugopal, Eastman e Teizer (2015) indicam que o IFC é insuficiente para a interoperabilidade devido à falta de clareza semântica ao mapear entidades e suas relações. Ainda que o IFC possua a classe *IfcRelationship*, ela descreve as relações entre elementos apenas em cinco categorias: conectividade, associação, decomposição, definição e atribuição. Apesar de tentar reunir todos os elementos possíveis em uma estrutura única de anotação, a interoperabilidade com IFC ainda não é completa porque ele é ausente de semântica para relacionar entidades e suas relações. Este problema é mencionado por diversos autores como Noardo *et al.* (2021), Sobkhiz *et al.* (2021), Ying and Lee (2021) e Costin e Pauwels (2022).

A sintaxe é mais fácil de entender, mas é a semântica que especifica o significado ou o contexto da informação. Como exemplo, um modelo de troca pode transportar apenas os dados básicos de geometria sólida e material do modelo de construção trocado. As rotinas de exportação nesse nível são simples e as trocas são genéricas. Nesse caso, para qualquer uso além de uma simples verificação de choque de geometria, o software importador precisaria interpretar a geometria e associar o significado usando representações internas dos objetos recebidos em termos de seus próprios objetos nativos (Venugopal *et al.*, 2012, p. 418, tradução nossa).

Ainda que o esquema possua capacidade de receber novas anotações de parâmetros, que as informações geométricas sejam enriquecidas com sentido, ainda existe a dificuldade na incorporação das diretrizes para uso em *software*. Eastman (2014) indicou que a partir de 2008 haveria um grande desenvolvimento da área de verificação de conformidade. Mas os programas existentes que são capazes de automatizar o processo, ainda dependem de uma prévia interpretação e transcrição do texto com uma sintaxe e representação predeterminada pelas empresas.

---

<sup>2</sup> <https://buildingsmart.crowdin.com/u/projects/2>

Nardelli e Oliveira (2014), por exemplo, enfrentam o problema inicial da verificação automática de conformidade criando uma biblioteca parametrizada de componentes com os atributos necessários para realização de simulações. O primeiro problema identificado foram os elementos que constituem as famílias fornecidas para o programa Autodesk Revit<sup>3</sup> não possuem parâmetros bem definidos e completos, pela falta de informações sobre os componentes e materiais dos elementos presentes nas famílias. Os autores declararam que somente com dados com informações necessárias e corretas será possível fazer uma análise da Norma de Desempenho e estimar o custo total por unidade, que é um critério decisivo na construção de Habitações de Interesse Social (HIS).

Como mencionado em İlal e Günaydin (2017) um dos problemas mais presentes nas pesquisas sobre ACC nos últimos anos estão relacionados ao tratamento e garantia de fidelidade das informações nos processos BIM. Atualmente existem aplicações para validação dos modelos que tentam garantir sua consistência com as definições de visualização do modelo. As referências levantadas por essa pesquisa indicam que o maior desafio que tem sido enfrentado pelas pesquisas relacionadas ao tema é a tentativa de traduzir representações textuais ou digitais dos regulamentos para padrões computáveis, capazes de serem interpretados por máquinas. Esse foco de pesquisa tem sido abordado há mais de 20 anos em estudos como em Drogemuller, Woodbury e Crawford (2000).

Os regulamentos são construídos com textos em formatos não estruturados, em linguagem natural. Uma sentença escrita em linguagem natural é um sistema estruturado de comunicação entre seus componentes constituintes como caracteres, palavras, sentenças e textos. Porém, esta estrutura gera significado quando seus constituintes são articulados por meio de uma sintaxe dentro de um contexto. A combinação entre sintaxe e contexto é que permite que seja extraído um sentido de cada sentença, também chamado de semântica.

---

<sup>3</sup> <https://www.autodesk.com.br/products/revit/overview>

A formulação de modelos para verificação de regras requer sintaxe e semântica clara, auditável e replicável para representar e racionalizar sobre a construção de requisitos e disposições dos regulamentos em formatos computáveis (Nawari, 2018). Nos casos de uso de plataformas orientadas a objetos como o BIM, ainda existe a incompletude, ausência ou inconsistência da modelagem. Esse caso depende de uma etapa anterior, de validação do modelo, para garantir que todas as informações necessárias foram inseridas de modo correto nos modelos (Amor; Dimyadi, 2021).

Portanto, para executar uma avaliação de conformidade adequada é necessário garantir que a modelagem ocorra segundo critérios previamente estabelecidos, evitando a inclusão de objetos arbitrários e atributos que não possuam relação com o conjunto de regulamentos a serem verificados (Amor; Dimyadi, 2021).

## **1.2. Novas abordagem para o processamento de regulamentações**

Nos últimos anos o Processamento de Linguagem Natural (PLN) tem sido abordado para ações de tradução, conversão de dados, análise do sentimento, conteúdo e recuperação de informação. O processamento de linguagem natural atua sobre dois tipos de dados, os dados textuais e dados orais. No caso de regulamentos, os dados disponíveis estão em formato estrutura (tabelas), mas principalmente em sentenças organizadas em capítulos. Um texto, em formato .docx, .pdf ou qualquer outro utilizado para produção de conteúdo pelo usuário final, é um dado do tipo não estruturado e incapaz de ser lido por máquinas. Portanto, trata-se um conteúdo textual que é apresentado em linguagem natural e orientado apenas para fins de leitura por seres humanos.

Um texto em linguagem natural não é capaz de fornecer informações ao ser incorporado em processos e rotinas computacionais. Para permitir que dados sejam extraídos do texto, ele deve ser convertido e interpretado pelo computador por meio de rotinas (algoritmos) de processamento de linguagem natural. Segundo Zhang e El-Gohary (2016) é possível desenvolver análises de um texto de regulamento a partir de sua

estrutura implícita utilizando PNL para ampliar a capacidade de interpretação e extração de dados não estruturados.

Um trabalho típico que use PLN objetiva encontrar expressões regulares no texto, rotular termos da frase, desenvolver análises sintáticas e morfológicas, realizar traduções, perceber sentimento e outros padrões em textos e, mais recentemente, desenvolver automaticamente textos e códigos computacionais. Em um processo convencional de PLN o texto é decomposto em partes e analisado pela sintaxe que é determinada pelos seus elementos constituintes (palavras).

Um dos subcampos do PLN é a Extração de Informações que procura extrair informações específicas de fontes textuais, não estruturadas e organizá-las ou preencher modelos com essas informações. Um exemplo trivial de aplicação é a localização de datas e nomes dentro de textos, para reinseri-los em formulários. A Extração de Informações não é um campo novo, as referências de pesquisas na área da ciência da computação na década de 1990. Entretanto, tem recebido atenção nas pesquisas sobre PLN, principalmente as que procuram incorporar novos métodos de Inteligência Artificial em sistemas de verificação automática de conformidade como Drogemuller, Woodbury e Crawford (2000), Zhang e El-Gohary (2011, 2015, 2016), Hjelseth (2012) e Fuchs e Amor (2021).

O processamento de linguagem natural já foi utilizado em diversas pesquisas sobre extração de informações de documentos regulatórios. As pesquisas normalmente seguem processos similares que contemplam adoção de análises da estrutura da frase (Zhang; El-Gohary, 2011, 2012, 2016, 2017b; Lee; Yi; Son, 2019; Wang; Issa, 2020; Wu *et al.*, 2022). Outras pesquisas também incorporam uma base de dados semânticos (ontologia) para aumentar a eficácia do processo de extração de informações dos textos regulatórios para construção de regras de verificação (El-Diraby; Lima; Feis, 2005; Zhang, El-Gohary, 2011, 2012, 2016, Xu; Cai, 2020). A Extração de Informações é um subcampo do Processamento de Linguagem Natural que trata de extrair conceitos como pessoas, lugares e datas de um grande volume de textos. Uma vez que textos são uma fonte de dados não estruturada, a

Extração de Informações busca padrões textuais necessários para revelar algum conhecimento sobre o conteúdo. Para tanto são usadas técnicas como rotulagem de palavras, reconhecimento de entidades, busca por expressões regulares formadas por determinada sequência de caracteres alfanuméricos e análises sintáticas baseadas em heurísticas, mapeamento em uma base de dados semânticos ou aplicação de técnicas baseadas em aprendizado de máquina.

### 1.3. Questões de Pesquisa

A problemática apresentada sobre a verificação de conformidade do projeto aos regulamentos da AEC levanta questões relacionadas aos métodos automatizados e à sua relação com a NBR 15.575 que é foco de aplicação desta pesquisa:

- Quais formatos de arquivos e estruturas de dados permitem incorporar dados semânticos aos sistemas de checagem automática de conformidade de regras?
- Quais são os métodos de processamento de linguagem natural que permitem obter dados semânticos e semiestruturados de regulamentos em língua portuguesa?
- Quais são os métodos de extração de informações, baseados em ontologia, que permitem obter dados semiestruturados a partir de textos regulatórios?
- Quais ontologias disponíveis têm potencial de aplicação no domínio da NBR 15575?
- Quais são os métodos mais adequados para converter os dados extraídos em linguagem computável para sistemas de checagem de adequação?

### 1.4. Objetivos

Dessa forma, **esta pesquisa tem como objetivo** a criação de um método de Extração de Informações da NBR 15.575 pelo Processamento de Linguagem Natural para tradução automática em um formato computável e capaz de ser incorporado em ferramentas de verificação automática de conformidade.

### 1.4.1. Objetivos específicos

Para atingir esse objetivo são esperados os seguintes produtos:

- Um algoritmo para reconhecimento de padrões semânticos no texto da Norma de Desempenho.
- Padrões textuais que representam as diretrizes da Norma de Desempenho.
- Um conjunto de regras de extração de informação da Norma de Desempenho para sua posterior conversão em linguagem computável.
- Um arquivo, em formato computável, que valide o método, representando um trecho da Norma de Desempenho.

## 1.5. Hipótese

A incorporação de uma análise semântica, baseada em ontologia, a um sistema de extração de informações da Norma NBR 15.57:2021, permite que seja desenvolvido um método de Processamento de Linguagem Natural com desempenho similar aos desenvolvidos em outras pesquisas internacionais.

## 1.6. Justificativa

A CBIC, por meio do projeto “Construção 2030”, iniciado em 2018 em parceria com o SENAI, supôs um cenário futuro da indústria da construção considerando estágios anteriores de inovação do setor AEC no Brasil. O projeto indicou um cenário pessimista caso não sejam realizadas determinadas modificações nas políticas nacionais de tecnologia e inovação:

Em 2030, o setor produtivo nacional não é capaz de incorporar soluções inovadoras nem nos processos construtivos e tampouco nos produtos finais. Processos arcaicos sem automação ou algum tipo de industrialização ainda têm taxas de desperdício comparáveis a 2018 (CBIC, 2022, p. 33).

Para a CBIC (2022), esse cenário dependeria de melhoria e adequações na política econômica nacional, mas também de políticas de inovação sustentável, automação e industrialização da construção. Nesse cenário é necessário que a inovação aconteça pela adequação dos produtos à normatização nacional e internacional, sejam eles produtos para construção ou o produto entregue ao usuário final que utiliza os produtos e processos desenvolvidos para essa cadeia produtiva.

Essa inovação, lenta na indústria da construção civil, pode ser direcionada segundo os modelos e práticas desenvolvidas por outras indústrias do setor manufatureiro como sugerem Succar e Poirier (2020):

A adoção de soluções tecnológicas avançadas está acelerando em toda a Indústria da Construção e nos setores mais amplos responsáveis por projetar, entregar, gerenciar e manter o ambiente construído. Isso oferece maiores oportunidades para lidar com o desempenho caracteristicamente ruim do setor e a produtividade atrasada. No entanto, não é possível obter plenamente os benefícios da inovação tecnológica sem inovações paralelas nos processos e nas políticas. Ao abordar os desafios de gerenciamento de informações do ciclo de vida, em particular, há muitas lições a serem aprendidas com os setores de Manufatura e Software e benefícios significativos a serem colhidos com a adaptação de seus métodos de produção *Lean, Agile* e integrativos (Succar; Poirier, 2020, p. 1, tradução nossa).

O relatório McKinsey de 2015 qualifica o setor da construção civil mundial como altamente dependente de políticas públicas e baixa digitalização e com crescimento negativo entre 2005 e 2014 (Manyika *et al.*, 2015). Esse relatório acompanha a última análise do Fórum Econômico Mundial realizada em 2016 sobre o setor da construção civil.

O setor de Engenharia e Construção (E&C) tem sido mais lento para adotar e se adaptar às novas tecnologias do que outros setores globais. Embora a inovação tenha ocorrido em certa medida no nível da empresa ou da empresa, a produtividade geral no setor permaneceu quase estável nos últimos 50 anos. Como uma indústria baseada em modelos de compras competitivos, a E&C historicamente adotou uma abordagem conservadora para o design e a entrega de produtos, levando a silos no gerenciamento de projetos e a uma indústria um tanto fragmentada (World Economic Forum, 2016, p. 5, tradução nossa).

No relatório elaborado pela Comissão Alemã para Tecnologias Elétricas, Eletrônicas e de Tecnologia de Informação (*Deutschen Kommission Elektrotechnik, DKE*), a criação de normas digitais (*digital standards*) permitirá apoio de especialistas, conversão e expansão da infraestrutura de TI, estabelecimento de processos ágeis, desenvolvimento de modelos de financiamento digitais, desenvolvimento de novas normas para padronização digital ampliando a força semântica do conteúdo e definindo padrões para desenvolvimento de conteúdo, interpretação e troca de informação normativa.

O programa de digitalização dos regulamentos promovido pela DKE (2020a) previu o uso de conteúdos de gestão baseados em XML devido a ser um formato independente e pode ser armazenado de forma modularizada. O formato é flexível porque é capaz de ser enriquecido posteriormente com mais metadados quando novas variáveis e padrões surgirem. Outro produto destacado pela DKE é o uso das normas em formato compatível para máquinas (*machine-compatible*) em ambientes de gêmeos digitais DKE (2020a). A Comissão alemã entende que o desenvolvimento de produtos exige contínua simulação e validação dos resultados em tempo real e a digitalização permite que as regras estejam explicitadas nos sistemas.

Assim que um regulamento é disponibilizado em formato digital estruturado, é possível que seja incorporado em um sistema de gerenciamento de regulamentos para que os usuários sejam notificados em tempo real sobre eventuais modificações do texto, acelerando o tempo de adaptação de uma nova regra de projeto.

### **1.6.1. Adoção do BIM no cenário nacional e a baixa digitalização da construção**

Um dos problemas mencionados por Succar e Poirier (2020), que impede a adoção completa de tecnologias no setor da construção, é a existência de um fluxo de informação sem integração, requerido pela manufatura industrializada. A digitalização do processo de projeto e construção é facilitada com a adoção de ferramentas adequadas do BIM que conseguem centralizar e compartilhar dados da construção a todos os atores envolvidos.

No modelo de adoção BIM de Succar e Kassem (2015) é possível analisar a capacidade BIM das empresas a partir de padrões e regras de modelagem, colaboração e integração. Uma vez que o BIM é um modelo de informações da edificação, as regras de modelagem indicam quais níveis de informação e quais os dados devem ser incorporados ao modelo para seus usos futuros. As informações inseridas no modelo segundo os Requisitos de Informação do Projeto (PIR) e os Requisitos de Troca de Informação (EIR), orientados pela ABNT/NBR 19650, devem considerar as normas nacionais que definem critérios de desempenho de execução de itens da construção. Portanto, os requisitos de informação devem estar organizados segundo critérios preestabelecidos e que habilitem o modelo a uma verificação de conformidade com múltiplas normas nacionais.

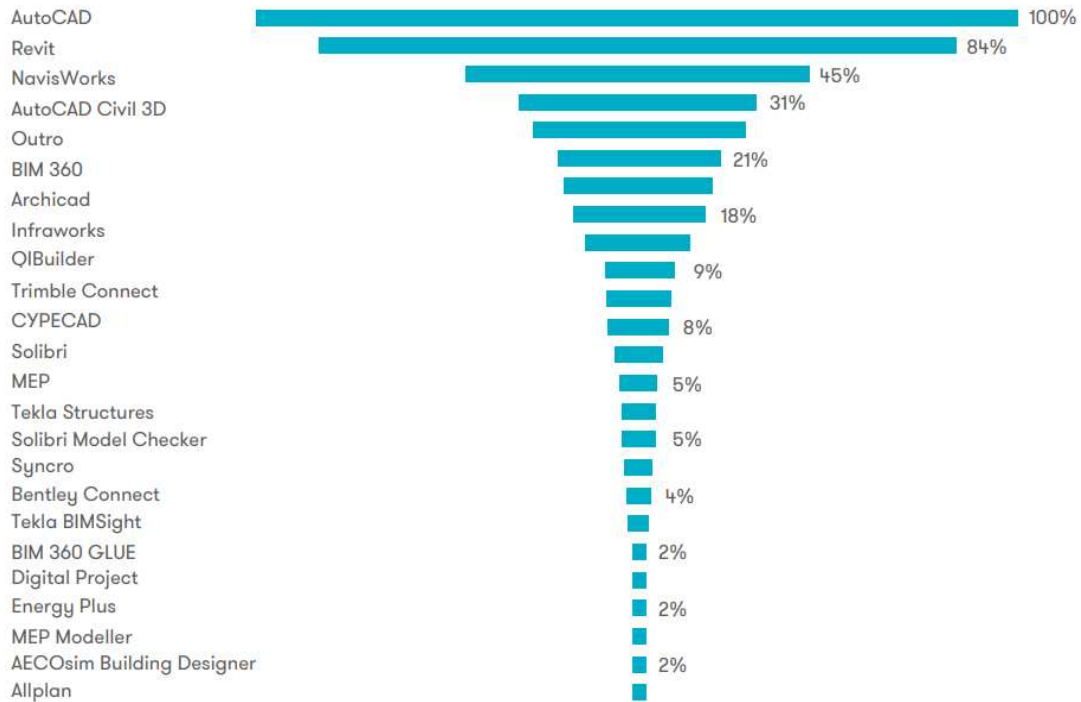
Em outro modelo de maturidade que avalia componentes de maturidade em macroescala, Succar e Kassem (2015) também deixam implícita a participação de regulamentos nas métricas de desempenho do projeto. O modelo sugere cinco níveis de maturidade das métricas de desempenho que vão desde a ausência de referências de desempenho (*benchmarks*) até a otimização contínua das referências de desempenho para refletir tecnologias emergentes, fluxos de trabalho e protocolos. Ao longo da escala, o atendimento a regulamentos da indústria, de associações ou de corpos regulatórios e sua posterior divulgação a *stakeholders* e membros da equipe é um critério que qualifica a maturidade macro do país e de organizações que adotam o BIM.

Apesar do Decreto Nacional 10.306/2020 instituir a obrigatoriedade na adoção do BIM para projetos da iniciativa pública, a pesquisa promovida pela Sienge e a Grant Thornton (Grant Thornton, 2020) mostrou baixa capacitação das empresas em participar de licitações. Apenas 24% das empresas afirmaram possuir equipes capazes de participar minimamente de processos licitatórios, outros 24% afirmaram estar em fase de preparação e 20% terceirizam os processos quando participam de projetos licitados.

A baixa maturidade BIM nas empresas também é registrada com a baixa adoção de sistemas para validação de projetos.

A pesquisa desenvolvida em 2022 pela Grant Thornton em parceria com a Sienge e a ABDI (Grant Thornton, 2022) mostrou que apenas 5% das empresas entrevistadas de uma amostra de 478 empresas adotam o *software* Solibri usado em verificação de conformidade de projetos (Figura 2).

**Figura 2 - Lista dos principais *softwares* usados no mercado da construção civil**



Fonte: adaptado de Grant Thornton (2022).

A baixa adoção de outros usos BIM e a baixa adoção de programas suplementares para análise e planejamento de projeto como Green Building, Solibri, Energy Plus, Syncro evidencia um estágio ainda inicial de maturidade BIM e a baixa digitalização da dos processos da indústria da construção civil nacional.

No cenário de maior digitalização da construção o modelo BIM permitiria a verificação do desempenho do edifício em estágios iniciais e não somente ao final do projeto como uma validação (Eastman, 2009). Para que isso aconteça é necessário que o processo de verificação seja automatizado para trazer celeridade às verificações. A automação de um sistema de verificação de regras depende de dois fatores que antecedem

a análise: depende da digitalização do regulamento em um formato capaz de ser processado pelo *software* e depende de um modelo de informações da construção com informações adequadas (Eastman *et al.*, 2014; Hakim; Garret Jr., 1992).

Na análise que Sobhkhiz *et al.* (2021) desenvolve sobre diversos métodos usados para ACC, eles observam que além dos motivos principais para utilização de ACC como o aumento da produtividade por meio da aceleração da verificação e diminuição de erros, a prática de ACC explicita a relação entre regulamentos. Com a interação de regras, diretrizes, códigos municipais, de desempenho e boas práticas de manufatura, é possível perceber quais as relações de dependência e concorrência entre as diretrizes. Essa análise permite ao escritório ampliar suas estratégias para enriquecer os modelos e processos de seus produtos além de informar as entidades regulatórias sobre inconsistências dos regulamentos.

De acordo com a ISO 19650-1, há quatro tipos de requisitos de informação e dois modelos de informação de ativos: Requisitos de Informação do Projeto (PIR), Requisitos de Informação do Ativo (AIR), Requisitos de Troca de Informação (EIR), Modelo de Informação do Ativo (AIM) e Modelo de Informação do Projeto (PIM). No caso de verificação de conformidade à NBR 15.575, com exceção do PIR, todos os requisitos de informação são necessários para verificação de conformidade porque tratam de exigir informações sobre produtos e prestadores de serviços (AIR), informações necessárias na troca de informações entre equipes (EIR), informações sobre uso, manutenção, gerenciamento e equipamentos instalados no edifício (AIR) e informações para projeto, ou seja, as informações relativas a dimensões, quantitativos e outros dados técnicos (PIM).

### **1.6.2. Promoção da qualidade da habitação com a adoção da norma NBR 15.575**

O Decreto 10.306/2022 obrigou o processo de digitalização com BIM no setor público e, de modo similar, ocorre com a adoção NBR 15.575 que exige a comprovação de oferecimento de um projeto com diferentes níveis de desempenho.

A Norma de Desempenho NBR 15.575 se coloca no cenário habitacional como o dispositivo que regulamenta a manutenção de dignidade da habitação de baixa renda. A promoção da qualidade é dada pela menção da Norma de Desempenho em dispositivos legais para aquisição de financiamento dessas habitações, como a sua referência no Sistema de Avaliação da Conformidade de Serviços e Obras (SiAC) (Brasil, 2021a).

O principal mecanismo para incorporação de rotinas de projeto que atendessem à norma de Desempenho ocorre com a menção da Norma no Programa Brasileiro de Qualidade e Produtividade do Habitat (PBQP-H). O programa, criado em 1998 pela Portaria nº 134, de 18 de dezembro de 1998, do antigo Ministério do Planejamento e Orçamento e tem, dentre outros objetivos, o fomento a instrumentos e mecanismos que garantam qualidade de projetos, obras, materiais, componentes e sistemas construtivos, o aperfeiçoamento da elaboração e difusão de normas técnicas, códigos de práticas e códigos de edificações e a promoção da articulação internacional.

Também em 2020, foi reconstituído o Comitê Nacional de Desenvolvimento Tecnológico da Habitação (CTECH), criado em criado pela Portaria nº 05, de 16 de fevereiro de 1998 e reconstituído pelo Decreto nº 10.325, de 22 de abril de 2020. O Comitê é composto por diversas entidades de representação civil e outros órgãos do governo para, sobretudo, propor mecanismos de ampliação de acesso à população de baixa renda às moradias, de avaliação técnica, de garantia de conformidade de produtos em toda cadeia logística de fabricação, importação e distribuição e criar ou implementar mecanismos de certificação de sistemas de gestão da qualidade.

Além da criação dos dois órgãos, a adesão das empresas no PBQP-H as obriga ao registro e certificação dentro desses sistemas para poder executar empreendimentos habitacionais com recursos públicos federais. Em qualquer um dos sistemas, para uma empresa se cadastrar para ter acesso a licitações e recursos federais é obrigatória a validação de seus produtos à Norma NBR 15.575.

Segundo Andery e Barbosa (2018), grandes empresas estavam paulatinamente implementando processos para atendimento à NBR15.575 enquanto as pequenas

empresas tiveram que executar rapidamente a adequação de seus processos para sua certificação. As empresas são obrigadas à certificação junto ao SiAC, que avalia a conformidade a sistemas de qualidade e que se tornou o catalizador do processo induzindo as empresas à adequação. A incorporação de uma nova estrutura de processos e a adequação à Norma também induziu ao desenvolvimento de mais análises de viabilidade técnica, de riscos dos empreendimentos e análises prévias de soluções como drenagem, desempenho térmico e acústico, aprimorando os projetos que passaram a incorporar os estudos e simulações iniciais como dados de projeto (Andery; Barbosa, 2018).

Apesar de as normas serem de uso voluntário em atividades não governamentais ou de caráter privado, a manutenção da qualidade é obrigatória, tem força de lei uma vez que a produção da habitação no Brasil é amparada pelo Código de Defesa do Consumidor, a saber:

Art 18. § 6º São impróprios ao uso e consumo:  
(...)

II - os produtos deteriorados, alterados, adulterados, avariados, falsificados, corrompidos, fraudados, nocivos à vida ou à saúde, perigosos ou, ainda, aqueles em desacordo com as **normas regulamentares de fabricação**, distribuição ou apresentação;

(...)

Art. 20. § 2º São impróprios os serviços que se mostrem inadequados para os fins que razoavelmente deles se esperam, bem como aqueles que **não atendam as normas regulamentares de prestabilidade**.

(...) Art. 39. É vedado ao fornecedor de produtos ou serviços, dentre outras práticas abusivas:

(...)

VIII - colocar, no mercado de consumo, qualquer produto ou serviço **em desacordo com as normas expedidas pelos órgãos oficiais competentes ou, se normas específicas não existirem, pela Associação Brasileira de Normas Técnicas** ou outra entidade credenciada pelo Conselho Nacional de Metrologia, Normalização e Qualidade Industrial (Conmetro) (Brasil, 1990, grifo próprio).

Desde a primeira iniciativa de criação de um sistema de verificação de conformidade nos anos 60 é feita a tentativa de objetivação de normas. A tendência de objetivação da norma e de critérios de medição ou avaliação dos parâmetros de

desempenho vem ocorrendo junto às Comissões de Estudo da Norma de Desempenho (ABNT CE-002:136.001). As reuniões que ocorreram para a atualização da Norma (versão de 2013 para 2021), mostraram a preocupação não só dos membros da comissão como dos representantes do mercado em definir critérios objetivos para análise (CBIC, 2019a, 2019b, 2019c).

Dadas essas questões operacionais, o cenário de digitalização da construção civil e a maturidade BIM no país se faz necessário desenvolver práticas e modelos universais que acelerem a digitalização, a ampliação da capacidade das ferramentas e a consequente adoção de padrões de qualidade pelas empresas de projeto.

## 1.7. Estrutura da tese

Este trabalho adota a estrutura de pesquisa apresentada pelo método *Design Science Research* e, portando, organizamos os capítulos segundo o delineamento do método, como apresentado a seguir.

No **Capítulo 1** foi apresentado o cenário geral sobre avaliação de conformidade e seus problemas. Deste cenário foram elaboradas as questões de pesquisa, o objetivo principal, a hipótese que será testada e a justificativa desta pesquisa.

No **Capítulo 2** é apresentada a revisão bibliográfica sobre características da Norma de Desempenho e sua dificuldade de incorporação, sistemas de checagem automática, as plataformas para verificação automática de conformidades e modelos de representação computáveis de regulamentos. Também apresentamos uma análise de representação de regulamentos em BIM, a iniciativa ISO Smart para criação de regulamentos ISO computáveis e, por fim, conceitos e etapas típicas de um processo de extração de informações do processamento de linguagem natural.

No **Capítulo 3** é apresentado o método selecionado para a pesquisa, o *Design Science Research*, suas etapas, objetivos e produtos esperados que, em conjunto, formalizam os produtos esperados com esta pesquisa.

No **Capítulo 4** é apresentada a etapa de conscientização adentrando a problemática da dependência do processamento de linguagem natural com ontologias. Avaliamos se ontologias existentes relacionadas à construção civil teriam a capacidade de oferecer precisão na tradução automática da Norma de Desempenho.

No **Capítulo 5** é apresentada a etapa de sugestão que propõem a abordagem da pesquisa, os paradigmas relacionados e ferramentas para desenvolvimento. Aqui também são detalhadas as etapas de construção do produto desta pesquisa.

No **Capítulo 6** é apresentada a etapa de desenvolvimento que está organizada em etapas bem definidas e orientadas por produtos. Neste capítulo são desenvolvidos os métodos e procedimentos, assim como são apresentados os ajustes necessários para a execução do método proposto.

No **Capítulo 7** é apresentada a etapa de avaliação sobre os dados produzidos no capítulo anterior, de desenvolvimento. Os dados foram tabulados e organizados em categorias a partir da aplicação de métricas de qualificação de classificações de dados. Ao final foi conduzida uma análise de agrupamento dos dados permitindo maior interpretabilidade do resultado e foi feito um segundo experimento para validação do método com relação a outros métodos notáveis.

No **Capítulo 8** é apresentada a discussão sobre o método aplicado e a síntese dos produtos da pesquisa. Também é avaliada a capacidade em generalizações do método em outras aplicações.

No **Capítulo 9** é apresentada a conclusão que evidencia a contribuição da pesquisa, seus produtos e seu ineditismo.

## 2. REVISÃO BIBLIOGRÁFICA

Existem diferentes termos que expressam o mesmo sentido na avaliação de conformidade do projeto a regulamentos, normas, códigos e padrões. Com a finalidade de estabelecer um alinhamento conceitual é necessário apresentar e organizar as terminologias oficiais com as abordagens utilizadas no setor da construção civil.

Segundo a ABNT (1994), **regulamento técnico** é:

[...] documento que estabelece características de um produto ou processo a ele relacionadas e métodos de produção, incluindo as cláusulas administrativas aplicáveis, com as quais a conformidade é obrigatória. Este documento pode também incluir ou tratar exclusivamente de requisitos de terminologia, símbolos, embalagens, marcação, rotulagem, e como eles se aplicam a um produto, processo ou método de produção (ABNT ISO/IEC Guia 59:1994, p. 5, grifo próprio).

Já **norma**, é entendida como

[...] documento que estabelece características de um produto ou processo a ele relacionadas e métodos de produção, incluindo as cláusulas administrativas aplicáveis, com as quais a conformidade é obrigatória. Este documento pode também incluir ou tratar exclusivamente de requisitos de terminologia, símbolos, embalagens, marcação, rotulagem, e como eles se aplicam a um produto, processo ou método de produção (ABNT ISO/IEC Guia 59:1994, p. 5, grifo próprio).

[...] documento estabelecido por consenso e aprovado por um organismo reconhecido, que fornece regras, diretrizes ou características mínimas para atividades ou para seus resultados, visando à obtenção de um grau ótimo de ordenação em um dado contexto (ABNT, 2022, grifo próprio).

Apesar de toda generalização reduzir a quantidade de informações específicas, ambos conceitos de norma e regulamento são entendidos como documentos que definem atributos, características de produtos e regras para execução de processos. Apesar da obrigatoriedade de conformidade para normas ser declarada em sua definição como ação facultativa, elas são usadas para harmonização de regulamentos díspares em mercados internacionais. Uma avaliação de conformidade de produtos e processos estrangeiros verifica se os produtos e processos usados em outros para determinado fim possuem equivalência à regulamentos nacionais e, na ausência dos regulamentos, verifica se existe a equivalência às normas nacionais.

[...] a normalização (e o conseqüente uso de normas técnicas) é utilizada como um instrumento que torna possível a globalização dos mercados e o estabelecimento de novas bases para a competitividade das empresas (Brasil, 2009, p. 7).

As maneiras empregadas para verificar o atendimento aos requisitos especificados pelo comprador são atividades de avaliação da conformidade. Mas os requisitos precisam estar previamente estabelecidos, e normalmente estão, sob forma de uma norma ou de outro documento equivalente, como por exemplo: um regulamento técnico ou uma especificação (Brasil, 2009, p. 39).

Os regulamentos mencionados nesta pesquisa podem ser urbanos (código de obras, planos diretores), normativas nacionais (ABNT), padrões de projeto de empresas, organizações (p. ex.: Green Building Council e certificação LEED), normas internacionais (p. ex.: ISO), regras de concessionárias, instruções técnicas, regimentos de condomínios, entre outros. Uma vez que a pesquisa trata de processamento de texto para extração de seu conteúdo visando sua transformação para outro formato, não será feita a diferenciação do texto de um regulamento de um texto de norma, pois ambos possuem as mesmas características estruturais. Desta forma, quando a pesquisa menciona regulamento, o faz em caráter amplo e se refere a todos documentos que definem regras e requisitos a fim de estabelecer a conformidade de produtos e processos a seus requisitos.

Sobre o termo “avaliação de conformidade” nesta pesquisa não será feita diferenciação dos termos “avaliação” e “verificação”. Também não serão diferenciados os termos “verificação de regras” (*code checking*) e “verificação de conformidade” (*compliance checking*) ou “verificação de conformidade a regras” (*code compliance checking*) devido à percepção que diversos autores utilizam estes termos para tratar da **verificação da conformidade do projeto aos regulamentos**.

O termo adotado nesse trabalho será **avaliação de conformidade**, que:

“envolve atividades que buscam demonstrar o atendimento, aos requisitos especificados por produtos, processos, sistemas, pessoas ou organismos. A avaliação da conformidade inclui atividades que proporcionam diversos meios de assegurar que os produtos, processos, sistemas, pessoas ou organismos atendem aos requisitos exigidos em especificações tais como normas, guias ou outros documentos normativos internacionais, regionais ou nacionais” (ABNT, 2005, p. 5).

## 2.1. NBR 15.575:2021

Segundo Allen e Sriram (2000), a existência de um padrão define uma base de confiança entre cientistas, engenheiros, fornecedores de bens e serviços, seus consumidores, governo e o setor privado. Os padrões (*standards*) são críticos para segurança nacional e para o avanço econômico porque facilitam a exportação mesmo com a existência de barreiras técnicas para muitos produtos comercializados internacionalmente.

No caso nacional, assim como em países europeus e na Costa do Pacífico, os padrões regulatórios são centralizados por agências acreditadas pelo Governo. Entretanto, nos Estados Unidos são elaborados por mais de 400 desenvolvedores do setor privado como a American National Standards Institute (ANSI) e o próprio Governo Norte-americano (Allen; Sriram, 2000).

A Norma de Desempenho, publicada em 2013, teve início em 2008 com um projeto financiado pela Caixa Econômica Federal para “avaliação de sistemas construtivos inovadores” para edifício de até cinco pavimentos (Kern; Silva; Kazmierczak, 2014). Naquele momento houve uma grande demanda na alteração do texto original, permitindo que a norma entrasse em vigor apenas em 2013. A Norma de Desempenho teve sua primeira revisão iniciada em 2018 e concluída em 2021 estabelecendo ajustes, principalmente na avaliação de desempenho térmico dos edifícios.

A NBR 15575:2021 se constitui como um instrumento normativo que está estruturado em seis partes, a saber:

- 1) Requisitos gerais, que estabelecem os requisitos e critérios de desempenho aplicáveis às habitações;
- 2) Requisitos para os sistemas estruturais,
- 3) Requisitos para os sistemas de piso,
- 4) Requisitos para os sistemas de vedações verticais internas e externas,
- 5) Requisitos para os sistemas de coberturas,

6) Requisitos para os sistemas hidrossanitários.

Na Figura 3, é demonstrada a estrutura da norma distribuídas nas partes: capa (A), contracapa (B), índice (C), prefácio (D), introdução (E), escopo (F), referências normativas (G) e termos e definições (H). Estes elementos estão representados em azul na Figura 3. Os itens em cinza são os requisitos específicos (I) e representam 74% do conteúdo da norma, que correspondem aproximadamente a 300 páginas. Os requisitos são compostos por textos, esquemas conceituais, desenhos técnicos, figuras, tabelas e gráficos.

**Figura 3 - Conteúdo de cada página da Norma de Desempenho (parte 1 a 6)**

P1	A	B	C	C	C	C	C	C	C	D	F	E	F	G	G	G	G	H	H	H	H	H
	H	H	H	H	H	H	H	H	I	I	I	I	I	I	I	I	I	I	I	I	I	I
	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
P2	A	B	C	C	C	D	D	E	F	G	H	H	I	I	I	I	I	I	I	I	I	I
	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
P3	A	B	C	C	C	C	C	C	D	F	E	F	G	G	H	H	H	I	I	I	I	I
	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
P4	A	B	C	C	C	C	C	C	C	D	F	E	F	G	G	H	I	I	I	I	I	I
	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
P5	A	B	C	C	C	C	C	C	C	D	F	E	F	G	G	H	H	H	I	I	I	I
	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
P6	A	B	C	C	C	C	D	F	E	F	G	G	G	H	H	I	I	I	I	I	I	I
	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I

Fonte: o autor.

Os requisitos estão organizados em seis capítulos divididos em três grandes grupos (ABNT, 2021): (i) segurança englobando desempenho estrutural; segurança contra o fogo; segurança no uso e na operação; (ii) habitabilidade compreendendo estanqueidade; desempenho térmico; desempenho acústico; desempenho lumínico; saúde, higiene e qualidade do ar; funcionalidade e acessibilidade; conforto tátil e antropodinâmico e (iii) sustentabilidade abrangendo durabilidade; manutenibilidade; impacto ambiental. Todos os requisitos estão apresentados na primeira parte da norma e seguem o seguinte padrão de redação: descrição do requisito, critério e método de Avaliação.

O escopo dos requisitos da Norma afeta os fornecedores, projetistas, construtoras, incorporadoras e usuários. Os fornecedores, devem comprovar o desempenho de seus produtos se baseando em normas nacionais ou internacionais. Os projetistas devem especificar a vida útil de projeto, materiais, produtos e processos que atendam o mínimo recomendado pela Norma. As construtoras e incorporadoras devem providenciar aos projetistas informações necessárias sobre riscos do ambiente onde será instalado o edifício e devem elaborar o “manual do proprietário” e manual das áreas comuns e de uso, operação e manutenção da edificação. O usuário é responsável por efetuar manutenções para manter garantia e desempenho da edificação.

A NBR 15.575, como principal referência para o desempenho da habitação, fortalece a percepção e função da edificação como um sistema de soluções que estão em interação. Sua própria estrutura é complexa, dividida em seis requisitos organizados pelos sistemas e que estão articulados pela interrelação do desempenho de seus componentes.

Ao longo de todos os capítulos diversos conceitos que são declarados na norma estão referenciados em outras normas, o que dificulta sua interpretação para fins de conformidade. O conceito de vida útil, por exemplo, é um parâmetro da norma que se refere em Normas externas:

[...] pela comprovação da durabilidade dos elementos e componentes dos sistemas, bem como de sua correta utilização, conforme as Normas a elas associadas que tratam da especificação dos elementos e componentes, sua aplicação e métodos de ensaios específicos, como ABNT NBR 5649, ABNT NBR 6136, ABNT NBR 8491, ABNT NBR 9457, ABNT NBR 10834, ABNT NBR 11173, ABNT NBR 13281, ABNT NBR 13438, ABNT NBR 13858-2, ABNT NBR 15210-1, ABNT NBR 15319, ABNT NBR 6565; ABNT NBR 7398; ABNT NBR 7400; ABNT NBR 9781; ABNT NBR 13528 ABNT NBR 8094; ABNT NBR 8096 e outras Normas Brasileiras específicas, conforme o caso (ABNT, 2021, p. 61).

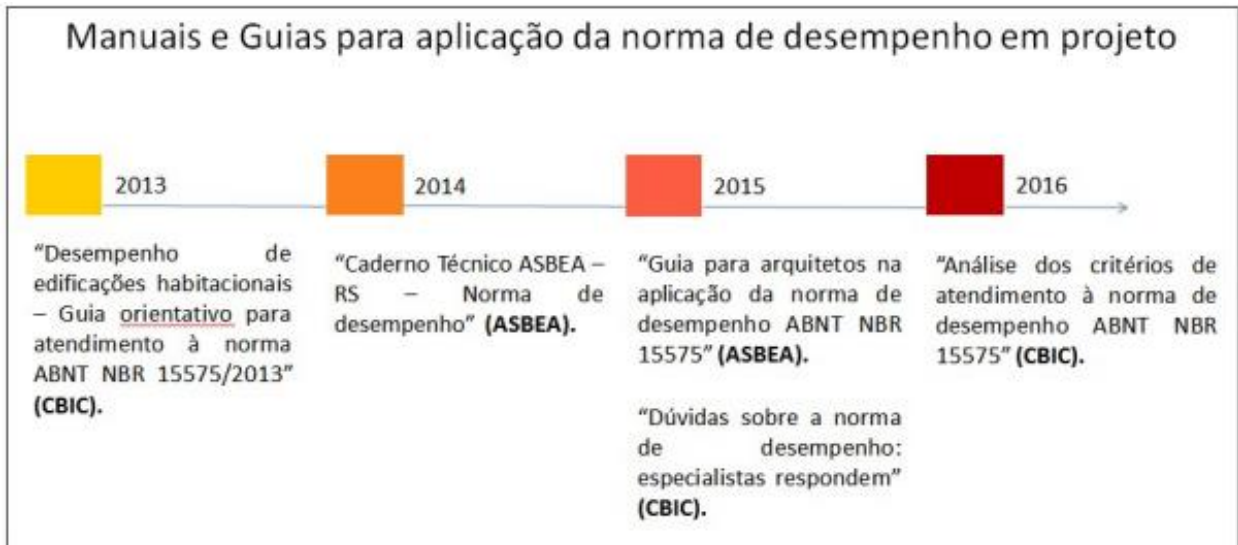
Esta complexidade é representada pela quantidade de trabalhos nacionais que se propuseram a estudá-la. Nunes *et al.* (2021) mostraram que foram publicados 1054 trabalhos acadêmicos relacionados a impactos da Norma de Desempenho entre 2008 e 2018. Alguns dos resultados evidenciados pelas pesquisas destacam-se: 1) a norma é complexa porque possui grande quantidade de critérios de avaliação (155 critérios), 2) existe dificuldade na sua interpretação e 3) a Norma não é replicada em regulamentos urbanísticos apesar de terem natureza e objetivos similares (o que também é verificado em Ribeiro; Fossati, 2020). Em outras pesquisas com datas mais recentes levantadas por Nunes *et al.* (2021) verificou-se que a adoção da Norma gerou maior preocupação das empresas em fases de projeto e planejamento integrando diferentes setores da empresa (suprimentos, projeto e qualidade) com maior ênfase em compatibilização antecipada dos projetos para verificar falhas antes da execução.

### **2.1.1. Dificuldades da incorporação da Norma de Desempenho**

Okamoto (2015) mostra que no surgimento da Norma as equipes de projeto ficaram sobrecarregadas e passaram a adotar sistemas que pudessem apoiá-las em validações dos seus projetos. Mesmo que as entidades relacionadas ao setor construtivo (CBIC, ASBEA e CAU) estivessem mobilizadas para elaboração de manuais e guias para esclarecer dúvidas e rerepresentar a norma com formatos mais simplificados, Pereira (2021) percebe baixa adesão das empresas à adequação da norma. Durante os primeiros anos de surgimento da Norma houve tentativas de entidades em esclarecer sua adoção (Figura 4).

Segundo a autora a dificuldade de adoção é causada pela abrangência, subjetividade e dificuldade de compreensão da Norma pelas empresas.

**Figura 4 - Manuais e guias para esclarecimento da norma**



Fonte: Pereira (2021).

Nunes *et al.* (2021) realizaram uma ampla revisão da literatura sobre a adequação à norma de Desempenho evidenciando o impacto no meio acadêmico. Eles identificam diversos autores que, compreendendo a complexidade e baixa capacitação dos profissionais, sistematizaram soluções para orientar a adequação à norma como planilhas de checagem, catálogo de materiais e normas secundárias e documentos que auxiliam a implantação da norma. Nesse sentido é necessária a adoção de métodos que ofereçam apoio às decisões dos projetistas em diversas fases do projeto. Ferramentas de suporte que considerem a manutenção da qualidade do projeto permitem que os recursos humanos sejam deslocados para atividades intelectuais sobre a experiência do usuário minimizando a dedicação a tarefas repetitivas (Sobhkhiz *et al.*, 2021).

## 2.2. Sistema de checagem automática como resposta

Os estudos relacionados com automação da avaliação de conformidade são relatados há 50 anos (Amor; Dimyadi, 2021). Em um primeiro momento eles se relacionaram com verificação de projetos bidimensionais, mas hoje as pesquisas estão direcionadas à avaliação de projetos em formato BIM dada a quantidade de informação disponível em ambiente digital do modelo e em esquemas de dados passíveis de serem analisados em profundidade.

O estudo sobre *Automated Compliance Checking* tem início com Fenves (1966) considerado o pioneiro na criação de sistemas para verificação automática de regras. Ele desenvolve um processo de análise manual apoiado por uma estrutura tabular de verificação de regulamentos sobre projeto de estruturas. A tentativa do autor e dos demais que abordaram um método de avaliação automática se baseiam na variabilidade de resultados e interpretações na verificação manual.

Succar e Poirier (2020) definem automação como uma tarefa ou atividade executada por um ator máquina a partir da sua iniciação por um ator humano, por exemplo, executar um *script* para impressão 3D de um objeto. Nesse caso estaríamos executando uma tarefa **automatizada**. Já uma rotina **automática** seria realizada segundo um conjunto de atividades encadeadas com uma ação pré-programada para sua realização segundo determinadas condições, por exemplo, fechamento de portas corta-fogo em caso de acionamento de um sistema de incêndio. O oposto ao sentido de automação seria a ação manual, executada por um agente humano sem assistência de um ator máquina. Entre os dois cenários existe a tarefa assistida que é executada por um ser humano com a assistência de uma máquina como por exemplo, a verificação de um modelo BIM a partir de um *software* de verificação de interferências (Solibri ou Navisworks).

Ao longo do tempo foram feitas diversas tentativas de implementação de ferramentas e plataformas de análise da conformidade de projetos. Amor e Dimyadi

(2021) apresentam a linha do tempo dos sistemas automáticos de checagem (Figura 5) e nela podemos perceber a intensificação das pesquisas e ferramentas relacionadas à avaliação de conformidade com o surgimento da estrutura de dados IFC. Dessas iniciativas, o Solibri Model Checker<sup>4</sup> detém elevada parcela do mercado, mas o sistema CORENET e-PlanCheck do governo de Singapura, que ainda está em uso e recebendo atualizações, mostra robustez e tem sido amplamente mencionado em pesquisas de avaliação de conformidade no setor de AEC. Uma descrição sucinta de cada um dos sistemas é apresentada na Seção 2.3.

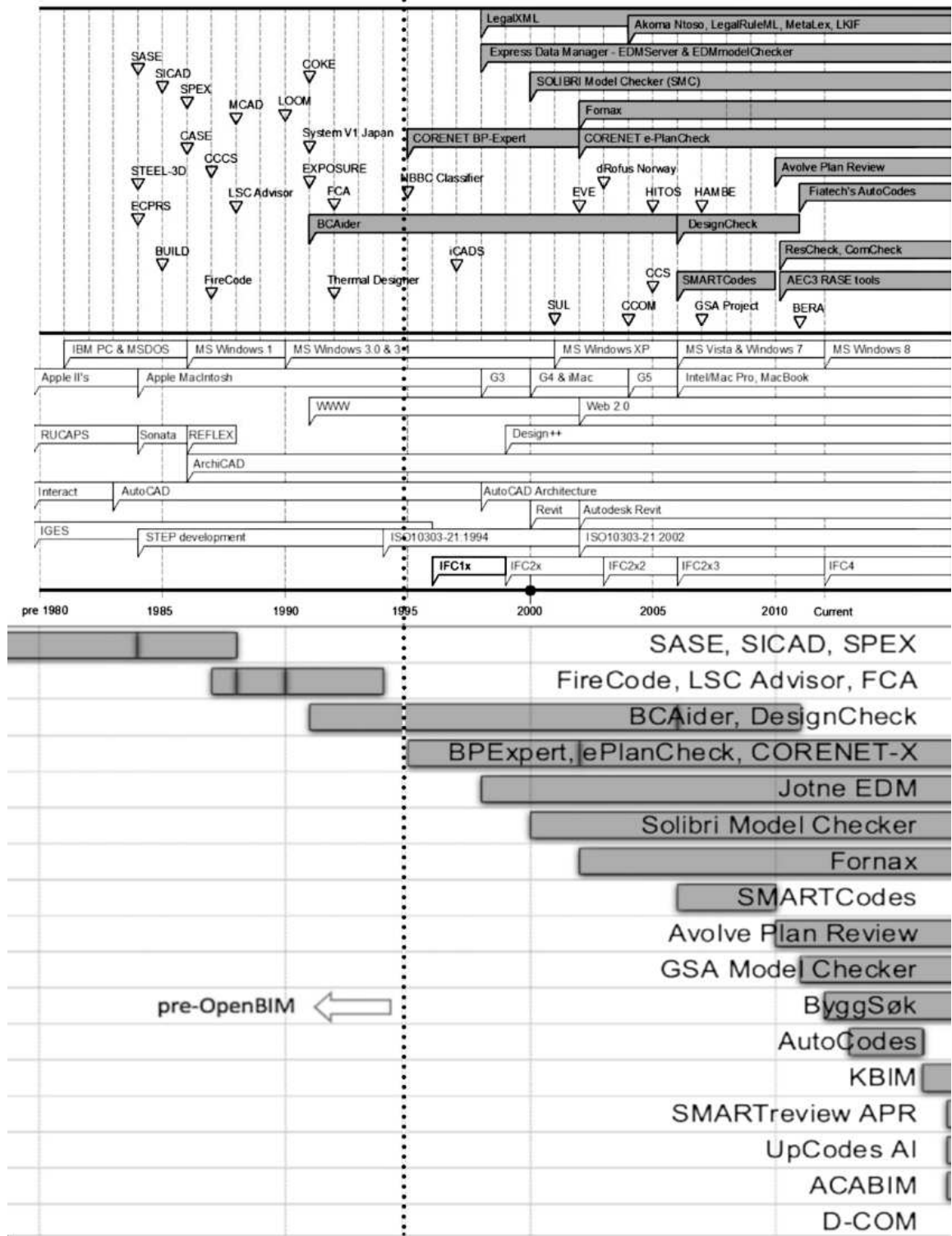
Ao considerar as tentativas de criação de métodos automáticos, pode-se afirmar que eles ocorrem de modo pleno a partir da década de 1990, quando surgem os primeiros questionamentos sobre a possibilidade de computação dos códigos regulatórios (Nawari, 2018). Durante esse período, o surgimento do IFC permitiu que muitas pesquisas fossem direcionadas para aumentar a interoperabilidade entre sistemas e adequação dos regulamentos em direção a um formato comum. O sistema CORENET, por exemplo, passa a utilizar o IFC em 1998, três anos após ser idealizado pelo Governo de Singapura, que estava participando ativamente para a criação do formato IFC.

Segundo Eastman *et al.* (2014), uma iniciativa para sistemas *Automated Compliance Checking* seria a incorporação de *plug-ins* em ferramentas BIM onde o “*plug-in* extrai dados de requisitos locais de bases de dados mantidas por serviços de autoridades locais”. Nesse modelo, idealizando esta proposta para o Brasil, a base de dados seria central, mantida pela ABNT, por exemplo, governo ou prefeituras. Outro modelo, amplamente utilizado hoje, é a criação de programas que analisam arquivos de formato neutro (IFC) com as regras formuladas manualmente e possivelmente compartilhadas entre técnicos que as desenvolveram (Ex.: Solibri Model Checker e UpCodes).

---

<sup>4</sup> <https://www.solibri.com/>

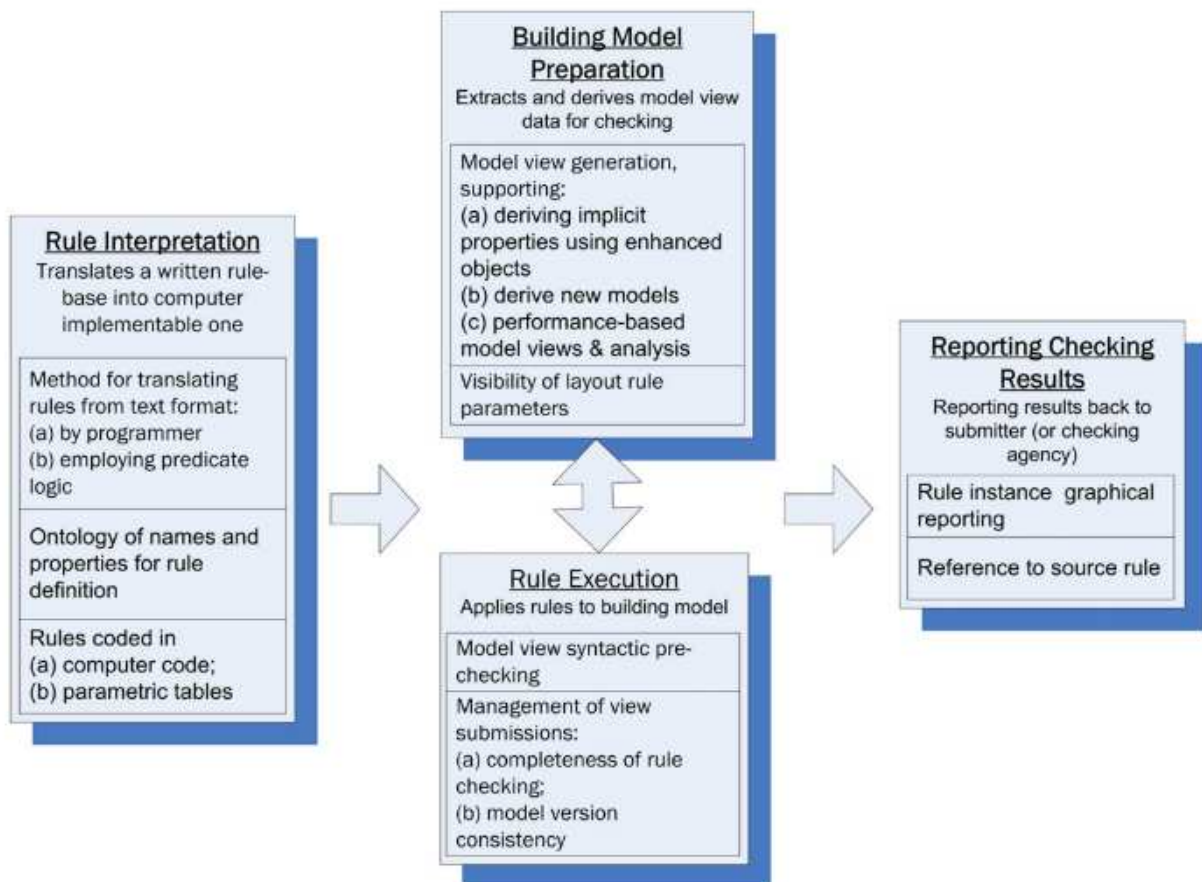
Figura 5 - Pesquisas e aplicações para verificação automática de conformidade



Fonte: adaptado de Dimyadi e Amor (2013) e Amor e Dimyadi (2021).

Hakim e Garrett Jr. (1992) propuseram uma das primeiras definições de um “*automated standards processing environment*” que deveria possuir 4 componentes: 1) o modelo do regulamento em formato de hipertexto, 2) um modelo semântico do regulamento para modelagem de seu conteúdo semântico, 3) um ambiente para desenvolvimento desses dois itens anteriores por ‘pesquisadores conhecedores’ do regulamento e 4) um sistema de processamento do regulamento. De algum modo esses componentes hoje estão presentes nos programas que analisam conformidade de regras, com exceção do modelo semântico que é específico para cada domínio do conhecimento, mas tem sido utilizado em uma parcela relativamente recente de pesquisas (Amor; Dimyadi, 2021).

**Figura 6 - Funcionalidades que devem interagir em um sistema automático**



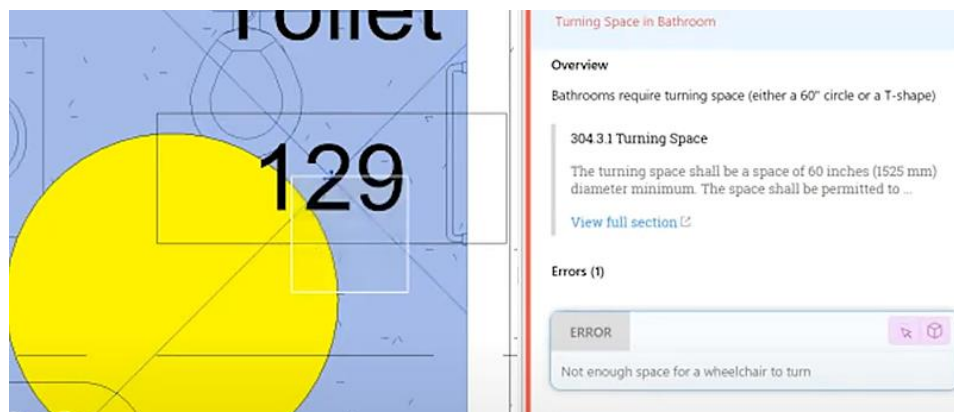
Fonte: Eastman *et al.* (2009).

### 2.3. Plataformas públicas para verificação automática de conformidade

Na Figura é apresentado o modelo de Hakim e Garrett Jr. (1992) com componentes criados por Eastman *et al.* (2009). O modelo apresenta simplificada um sistema de checagem automática que executa a verificação e emite relatórios contendo as regras interpretadas, que são inseridas no sistema de execução de regras, que analisa um modelo preparado para essa ação. Ao final são gerados relatórios de conformidade.

O UpCodes<sup>5</sup>, assim como MasterSpec (Nova Zelândia), é um “*office master specification system*” ou um sistema de gerenciamento de especificação de padrões. Um sistema dessa categoria funciona pelo navegador com planos de inscrições pagas pelos seus usuários. Ele se constitui como um serviço que centraliza, em uma plataforma única, diferentes regulamentos aplicáveis à construção civil e mantém o escritório atualizado sobre as suas alterações. O UpCodes possui funções de gestão de regulamentos como monitoramento de atualizações, comparativo entre normas, centralização de normas a acesso público, *hiperlinks* e um sistema que avalia conformidade de critérios quantitativos a partir de regras previamente estabelecidas pela plataforma, mas são “caixas pretas” porque não são explicitadas para os usuários (Grubber, 1993).

**Figura 7 - Verificação no *plugin* UpCodes sobre acessibilidade em um banheiro**



Fonte: adaptado de UpCodes (2019).

5 <https://up.codes/>

A verificação de um projeto no sistema UpCodes (Figura 7) gera um relatório de saída em PDF ou arquivo de planilha do Excel (\*.XLSX). Ele possui uma funcionalidade expandida na opção por *plugin* para Autodesk Revit.

O sistema também oferece serviços de elaboração de especificações técnicas (*specification writing*) que, em conjunto com caderno de encargos<sup>6</sup>, determinam de forma precisa os materiais e procedimentos para execução das atividades e componentes da construção.

A iniciativa **Digital Codes do International Codes Council (ICC)** dos Estados Unidos cria uma plataforma (*ICC Digital Codes Premium Complete*) que permite inscrição a um serviço de aplicativo *off-line* com acesso a todas os regulamentos e fica apto a realizar classificação de textos, histórico de interpretação dos regulamentos pelos respectivos comitês, compartilhamento, busca de termos e o usuário passa a ser informado quando ocorrer atualização dos regulamentos. Além dessa plataforma, a *General Code*, também vinculada ao ICC possui o serviço eCode360<sup>7</sup> que realiza serviços similares, com adição de uma atividade de busca mais detalhada de regras. De forma similar o serviço eCode360 estão as iniciativas UpCodes (descrita anteriormente) e o MasterSpec, ambos nomeados como sistemas de gerenciamento de regulamentos e não como sistemas de verificação automática de conformidade.

Uma alternativa a esses sistemas de gerenciamento de regulamentos seria abertura dos regulamentos em formato aberto e codificado (estruturado), para incorporação em diferentes programas e plataformas. Nesse caso, acompanhando a indicação de Eastman *et al.* (2014), os regulamentos seriam centralizados, atualizados e ajustados pela devida agência regulatória criando o que é chamado de “*single source of truth*”. No caso brasileiro, a regulamentação normativa fica restrita ao domínio da ABNT e poderia incorporar

---

<sup>6</sup> A diferença entre as especificações técnicas e cadernos de encargos é que esse último trata de contratação de serviços e obras, sua execução e controle. Já as especificações técnicas tratam de como executar determinados procedimentos e uso de materiais.

<sup>7</sup> <https://www.generalcode.com/subscription-services/>

algum sistema de gerenciamento em sua plataforma de acesso, como é hoje oferecido pelo ICC.

O **KBIM (Korea BIM)** é uma plataforma de verificação de conformidade em uso pelo Governo Sul Coreano desde 2013. Foi construído pela Inno-lab, empresa sul-coreana que desenvolve aplicações para BIM. A plataforma possui diversas ferramentas incorporadas como KBIM Asses, Asses-Lite, KBim Logic, KBim Collaboration, entre outras. Diferente de outras plataformas, o KBim agrega aprendizagem de máquina e algoritmo genético para otimização da implantação e uso do solo pela edificação.

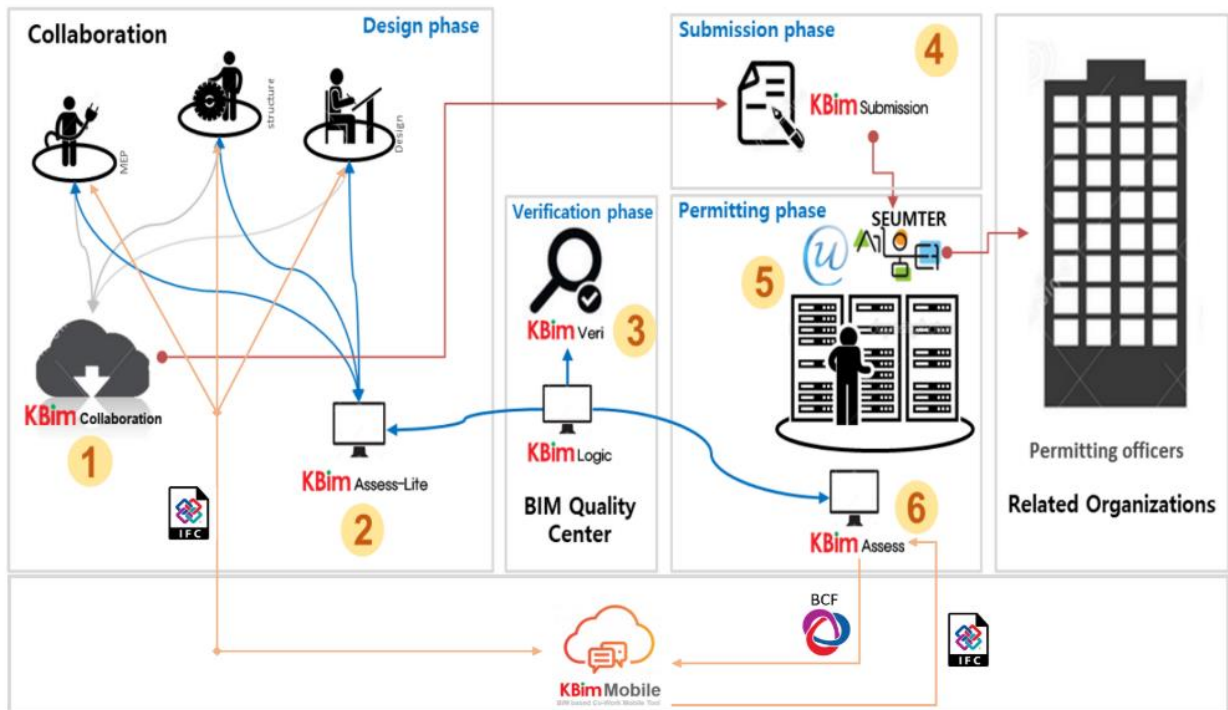
O sistema coreano de conformidade com as regras de construção é mais avançado e automatizado do que os sistemas de outros países e, uma vez implementado, pode melhorar os setores de AEC em muitos aspectos (Rehman et al., 2021, p. 10, tradução nossa).

O KBim Assess-Lite é um verificador de conformidade para profissionais validarem seus modelos antes da submissão enquanto o KBim Assess é utilizado apenas pelo corpo técnico do SEUMTER (agência pública sul-coreana relacionada à construção civil). O KBim Assess é utilizado pela SEUMTER para emissão de relatórios e histórico do serviço, enquanto na versão Lite, que opera em modo *standalone*, localmente, executa a pré-análise pelo projetista (Rehman *et al.*, 2021).

Na Figura 8 é demonstrado o processo de trabalho e aprovação de projetos por meio da plataforma. O fluxo se resume a uma elaboração colaborativa de projetos por meio da ferramenta KBim Collaboration (1) que integra todos os projetos. Em seguida é feita uma análise de validação e verificação prévia do modelo no KBim Assess-Lite (2).

Na etapa de verificação do modelo ocorre a avaliação do conteúdo do modelo apoiado por uma Multi-View Convolutional Neural Network (MVCNN) (Koo *et al.*, 2021). Esse algoritmo compara os elementos e define categorias possíveis para aqueles que não estejam corretamente modelados ou classificados (p. ex.: o algoritmo sugere que uma geometria plana não identificada na cobertura e modelada em formato genérico seja considerada uma laje).

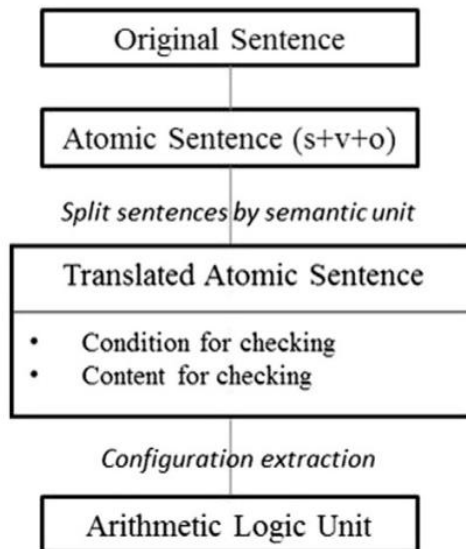
Figura 8 - *Framework* de trabalho no KBIM



Fonte: Rehman *et al.* (2021).

A verificação de conformidade é baseada nas regras traduzidas na plataforma KBim Logic que é um sistema lógico de gerenciamento baseado em regras que possui os regulamentos aplicáveis convertidos em um formato KBim Code. O KBim Logic contempla um banco de dados lógico baseado em regras dos regulamentos pertinentes, uma ferramenta de criação de códigos no formato KBim Code e um banco de dados com regras já formatadas em KBim Code (Choi; Kim, 2017). Este banco de dados com regulamentos é construído de forma manual (Lee *et al.*, 2016) (Figura 9). Uma vez que o arquivo esteja validado e verificado, é submetido para verificação pela SEUMTER usando o KBim Assess.

**Figura 9 - Processo de conversão manual de um regulamento para formato computável no software KBIM**



Fonte: Lee *et al.* (2016)

O KBIM, portanto, se coloca como uma “*single source of truth*” (SSOT), que na Tecnologia da Informação é usado para definir sistemas ou locais confiáveis para armazenamento de dados. Em BIM, uma SSOT é o conceito que fundamenta o *Common Data Environment* (CDE) que deve ser considerado como autoridade definitiva, padronizada e confiável de fonte de dados porque estabelece mais consistência e integração entre sistemas<sup>8</sup>.

Park e Lee (2016) trazem uma descrição do sistema de representação das regras utilizadas e construídas pelo programa KBIM. Inicialmente as regras são decompostas em frases nominais que carregam consigo informações dos objetos e propriedades para uma primeira classificação da regra. Em seguida, as frases verbais definem a estrutura da regra. Os procedimentos seguintes são desenvolvidos de modo similar aos trabalhos de Zhang e El-Gohary (2011, 2012, 2016). As sentenças são convertidas para um formato mínimo (atômico) de sujeito + verbo+ objeto. Ao que tudo indica na descrição das etapas é

<sup>8</sup> <https://www.solibri.com/news/supercharge-bim-workflows-with-single-source-of-truth>

utilizado um sistema que processa o texto por meio dos métodos de processamento de linguagem natural. Usando este método os autores executam a extração e análise das sentenças a fim de obter sua representação atômica e ao final são transformadas em uma estrutura de decisão tipo IF-THEN (Figura 10). Mas não é informado se existe intervenção humana no processo ou se a conversão é automática.

Ao executar a avaliação de conformidade o KBim Assess-Lite executa a regra por meio de uma “*parsing library*” definida no programa que avalia o arquivo IFC (Park; Lee, 2016). Esse banco de dados compreende um conjunto de regras executáveis pelo computador em formato JSON e XML (Park; Lee, 2016) (Figura 10).

**Figura 10 - Exemplo de transformação de uma sentença em sintaxe lógica**

Senten ce		[Enforcement Decree of Building Act, article 64] A project owner of a building with six or more floors and a total floor area of 2,000 square meters or more shall have an elevator installed therein. In such cases, the size and structure of elevators shall be prescribed by Ordinance of the Ministry of Land, Infrastructure and Transport. [22]
AS		A building with six or more floors and a total floor area of 2,000 square meters or more shall have an elevator installed.
TAS	Condition	TAS1. A building has six or more floors TAS2. A building has a total floor area of 2,000 square meters or more
	Content	TAS3. A building has an elevator
ALU	Condition	ALU 1. getBuildingStoriesCount() >= 6 ALU 2. getGrossFloorArea() >= 2000
	Content	ALU3. isExist(Elevator) = TRUE
KBim Code		Check (BA_64_1) { IF (getBuildingStoriesCount() >= 6 AND getGrossFloorArea() >= 2000) THEN isExist(Elevator) = TRUE ENDIF }

Fonte: Park e Lee (2016).

No estado atual da arte das tecnologias e ferramentas analisadas anteriormente, apenas a plataforma KBim incorpora algum sistema de transformação automática das regras em uma sintaxe capaz de ser processada pela plataforma. O Solibri Model Checker, amplamente conhecido no setor da avaliação de conformidade em AEC depende de uma inserção e programação manual. Já a Corenet possui uma versão do regulamento formalizado previamente inserido na plataforma, mas que não pode ser acessado pelos usuários.

## 2.4. Modelos de representação computável para regulamentos

Existem diversas pesquisas que abordaram diferentes métodos para implementar modelos de regras. Na sua pesquisa de doutorado, Macit (2014) analisa os modelos utilizados para representação e avaliação de regras dos regulamentos. Ela elenca os modelos SASE (*Standards Analysis, Synthesis and Expression*) (Fenves *et al.*, 1987), modelos baseados em regras, baseados em lógica, orientados por objetivos, modelos semânticos e baseados em ontologia.

### 2.4.1. Modelos regrados

Dada a complexidade e variedade dos assuntos que são tratados nessa pesquisa é preciso fazer uma distinção de termos quando nos referimos a processos de verificação de conformidade. Diversas pesquisas mencionam abordagens “baseadas em regras”, mas deve ser feito um esclarecimento conceitual do termo. O **processo** baseado em regras é um método com etapas e comandos claros que devem ser seguidos, como em uma programação de etapas. Mas neste capítulo adotamos o termo “modelo regrado” como definição de um modelo de informações que é representado **por regras**”. Ele deve ser compreendido como um objeto constituído por determinada linguagem de programação para ser processado ou armazenado em alguma plataforma. O modelo representado por regras produz uma tradução do regulamento em sentenças condicionais com estruturas

de decisão do tipo “IF-THEN”. Neste contexto, destacamos uma afirmação de Eastman *et al.* (2014):

[...] a codificação interna gera ferramentas que são muito caras para desenvolver e depurar, e que são inflexíveis a mudanças. Em vez disso, **linguagens de definição de regras em alto nível e com propósito específico** emergirão, facilitando o desenvolvimento de regras para verificação de conformidade em edificações. [...] Essas linguagens permitirão que pessoas sem conhecimento de programação de computadores escrevam e editem regras de verificação de um modo mais direto. [...] O desenvolvimento dessas linguagens facilitaria a implementação de ferramentas de avaliação de projeto em uma infinidade de áreas, para diferentes tipos de edifícios, clientes e agências reguladoras (Eastman *et al.*, 2014, p. 301, grifo próprio).

Os requisitos de avaliação são definidos pelo usuário em declarações condicionais (IF - THEN). Trata-se de uma tradução manual dos regulamentos que demanda a elaboração de muitas regras no sistema de processamento. As regras também são codificadas isoladamente dificultando a manutenção futura no sistema. Além disso, como não são facilmente compreensíveis por usuários não especialistas, são dependentes de programadores. Esse modelo de codificação manual é amplamente questionado por autores mais recentes (Seção 1.1.1) devido à chance de erros de codificação e a necessidade de revisões naturais do processo manual de tradução e codificação. A inviabilidade desse método também é apresentada em Hakim e Garrett Jr. (1992). A determinação de regras de modo *hard-coding* é uma prática que dificulta revisão, alteração e manutenção do programa porque a cada alteração é preciso editar e recompilar o código fonte.

Zhang (2015) considera que o regulamento está restrito a um domínio do conhecimento mais controlado ou finito. Desta forma, o modelo baseado em regras exige menos esforço manual para desenvolver e adaptar modificações nas regras.

A plataforma Solibri sofreu uma expansão agregando outras funções além do processo de checagem do modelo. No caso da elaboração de regras, o processo de checagem é feito em regras baseadas em parâmetros e analisa quatro parâmetros possíveis: dimensões, colisões, valores de propriedades e comportamento (Solibri, 2020). No caso da inserção de comportamento, a plataforma Solibri permite personalização de

regras de decisão (IF-THEN) entre elementos, por exemplo: “*IF a space is an electrical room THEN no plumbing system that is of type cold water domestic, hot water domestic, or sanitary should run above that space*” (Solibri, 2020).

Ainda sobre a plataforma Solibri, dada a quantidade de regulamentos em cada país (no Brasil são mais de 1.300 normas da ABNT relacionadas a desempenho da construção civil) seria inviável agregar em sua plataforma milhares de *rulesets* de cada regulamento. Além das dificuldades legais e autoria dessas codificações, seria necessário gerenciar um banco de dados de regras mundial. Mesmo que todas essas situações fossem garantidas, as regras ainda deveriam passar por uma auditoria ou certificação para atestar que o código é confiável e estável.

Algumas soluções adotaram regras executadas a partir de linguagem de programação visual para minimizar a curva de aprendizado para não especialistas (Kim *et al.*, 2019; Ghannad *et al.*, 2018; Preidel; Bormann, 2015). Para Preidel e Bormann (2015) o processo de abrir o método de análise para programação visual transforma o processo em “*White-box*”, ou seja, revelando o modo como os algoritmos interagem.

Pensando em desenvolver um modelo automático de verificação de conformidade, Hjelseth (2009) faz uma crítica ao modelo formado por regras, típico das aplicações recentes, como o Solibri. Para o autor, esse modelo de operação não considera que, dada a complexidade dos regulamentos, por vezes ocorrem sobreposições e conflitos na adequação de duas regras de regulamentos distintos. Se um sistema de verificação for composto por modelos regrados, ao ser aplicado para diferentes regulamentos ao mesmo tempo, os modelos seriam povoados por milhares de erros, ou seja, teríamos um “*Building Information Mess*” (Hjelseth, 2009).

Uma variação do modelo composto por regras é o modelo de objetos apresentado por Macit (2014). Nesta pesquisa ele não é diferenciado como uma categoria a parte porque trata do mesmo tipo de representação do modelo regrado. A diferença entre os dois modelos é que no baseado em objetos as regras são codificadas no próprio objeto. Os objetos e componentes mencionados pela norma são detalhados com características

formais e vínculos funcionais (ex.: uma forma, não simétrica, em formato de caixa e função de viga). Hakim e Garrett Jr. (1992) explicam que neste modelo são configuradas condições de existência para cada grupo de objetos, por exemplo:

```

Alloable-floor-area-req't
  if no-of-stories = 1
  then total-floor-area <= 29.000

```

No interpretador desse modelo a busca é feita pela análise das condições internas de cada objeto ou classe de objeto ao invés das regras invocarem os elementos que podem ser associados a elas.

### 2.4.2. Modelos lógicos

Dentre todos os modelos lógicos, o mais comum para expressar os aspectos linguísticos utiliza a lógica predicada que formaliza os objetos (ou modela as expressões) por meio de predicados, conectivos lógicos e quantificadores. A lógica predicada, também chamada de lógica de primeira ordem, fornece a validação objetiva de sentenças verificando a veracidade da formulação lógica. A diferença da lógica predicada para a lógica proposicional é que a primeira possui sintaxe que representa quantificadores (todo, algum, nenhum) enquanto a segunda não possui capacidade para essa representação. Mais diferenças e a sintaxe desse tipo de lógica serão tratadas mais adiante.

A formulação de axiomas lógicos representa as relações e elementos das regras do documento e elimina a redundância, ambiguidade e possíveis contradições encontradas no texto. Cada axioma estabelece o encadeamento de cláusulas formadas por fórmulas atômicas, expressões lógicas compostas pela atomização dos termos da expressão original, que expressam o sentido da sentença original. A Lógica de Primeira Ordem permite que sejam realizados “cálculos de prova” que verificam a validade das expressões lógicas fazendo com que, ao contrário dos outros métodos, a interpretabilidade e desambiguidade dos axiomas possam ser verificadas (Macit, 2014). Por outro lado, modelos baseados em lógica dependem da capacidade do desenvolvedor em elaborar os

axiomas e restringe a expressão de todas as expressões textuais aos operadores lógicos existentes nesse tipo de formalização.

Para implementação de expressões lógicas (cláusulas) um programa de verificação de conformidade deve incorporar algum motor de processamento desta linguagem. Uma alternativa à adoção de um motor de processamento é a criação um mecanismo de interação entre dois programas; o primeiro deve ser o programa gráfico que armazena e processa o modelo da edificação e o outro programa deve ser orientado para processamento de bases de dados. A interação do usuário pode ser feita por meio de buscas pelo usuário que analisam as regras em formato de axiomas lógicos.

A diferença entre o modelo baseado em regras e o baseado em lógica pode ser exemplificado pelas expressões extraídas de Macit (2014) apresentadas no Quadro 2.

**Quadro 2 - Comparação da sintaxe de os modelos baseados em regras e em lógica**

Baseado em regras	Baseado em Lógica
<b>IF</b> Related space of a door is an entrance <b>THEN</b> Check the width of the door is equal to 1.00 meters.	$\forall x(\text{Door}(x) \wedge (\text{RelatedSpaceUsage}(x)=\text{Entrance})) \rightarrow \exists(x)(\text{Width}(x) \geq 1)$

Fonte: o autor.

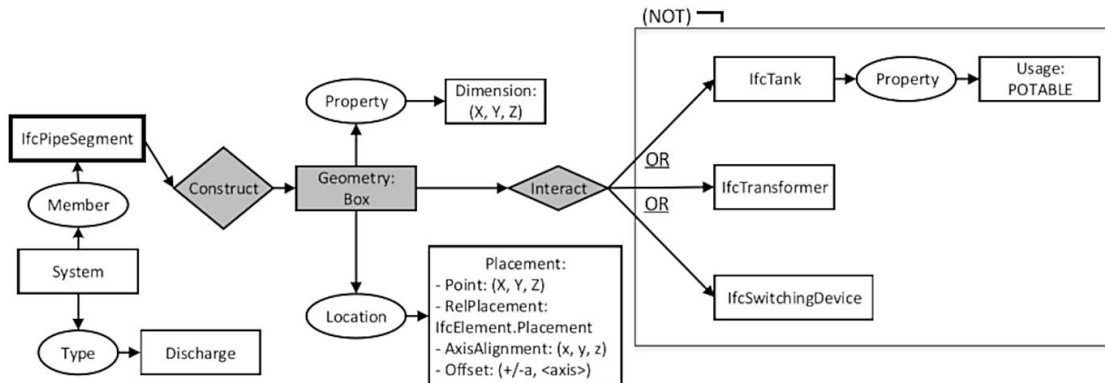
Fica claro que o modelo baseado em regras é mais expressivo porque permite maior compreensão pelo ser humano e é mais simples de ser codificado em qualquer linguagem de programação. Por outro lado, esse tipo de declaração baseada em regras exigirá a formulação de regras para cada diretriz do documento o que implica em um número muito alto de regras para serem processadas, porque muitas vezes cada diretriz deverá ser separada em sentenças menores (Figura 11). Uma alteração na legislação exigiria a modificação em cada sentença programada. Outro limitador do modelo baseado em regras é que a sua codificação em aplicações deve ser desenvolvida segundo a arquitetura e linguagem permitida para aquela aplicação. Esse processo de *hard-coding* (programação fixada no *software*) é extremamente dependente do profissional programador. Esse processo de programação dificulta a manutenção futura das regras porque a existência de uma atualização da legislação demandaria uma reconstrução do

desenvolvedor, uma reconstrução de todas as regras impactadas pela atualização e a posterior recompilação do código.

**Figura 11 - Exemplo de conversão de texto em grafo e linguagem lógica**

(42) 3.2.2 Design Criteria (Singapore, 2013)

f) The discharge pipe shall not be located in places where it can cause health and safety hazards such as locating the discharge pipe above any potable water storage tank and electrical transformer/ switchgear.



Representation in FOL

$$\begin{aligned} & \exists b \forall a (\forall g (IfcPipeSegment(a) \wedge System(b) \wedge Discharge(g) \wedge Member(b, a) \wedge Type(b, g)) \\ & \wedge (\forall c \forall p \forall d (Box(c) \wedge Placement(p) \wedge Dimension(d) \wedge Location(c, p) \wedge Property(c, p)) \\ & \wedge Construct(a, c)) \wedge \neg (\exists w \exists t \exists m (Interact(c, (\forall u (IfcTank(w) \\ & \wedge Usage(u, POTABLE) \wedge Property(w, u)) \vee IfcTransformer(t) \vee IfcSwitchingDevice(m)))))) \end{aligned}$$

Fonte: adaptado de Solihin e Eastman (2016).

Uma das primeiras referências que utilizou modelos baseados em lógica para avaliação de conformidade foi Kerrigan e Law (2003). Os autores desenvolveram um sistema baseado em Lógica de Primeira Ordem construído sobre regulamentos em formato XML. Os textos foram separados em seções onde foram feitas marcações com conceitos gerados de modo semiautomático por meio de mineração de texto e recuperação de informação. Ao final foi possível extrair do arquivo XML os conceitos necessários para serem acessados e verificados em buscas de conteúdo por meio de regras de inferência lógica no programa. Kerrigan e Law (2003) relatam que o sistema de Lógica de Primeira Ordem (LPO) não possui capacidade expressiva suficiente para representar toda a

complexidade do texto. Para os autores a LPO enfrenta limites na ACC com regras de textura aberta (dependem de um julgamento ou interpretação), modalidade deôntica (permissão, obrigação ou proibição) ou condicionais subjuntivas (uma condição hipotética, mas não necessariamente real).

Enquanto o modelo baseado em regras se dedica a construir uma estrutura fixa e codificada (*hard-coded*) determinada pelo desenvolvedor do programa, o modelo baseado em representação lógica se preocupa em não criar regras de ação, mas em formalizar aquele domínio do conhecimento e seus pressupostos com a maior capacidade semântica possível. O uso de metalinguagem para representar regras de modo simplificado, reduz sua complexidade e permite sua expressão em argumentos lógicos (Hjelseth; Nisbet, 2010). A incorporação de uma metalinguagem lógica permite a descrição de relações entre tópicos, determinação de quantificadores para cada condição e definição de critérios-verdade (Figura 12).

**Figura 12 - Exemplo de termos resultantes de uma EI e a sua representação lógica**

Information tuple extracted from text sentences	Subject	airspace
	Subject restriction	relation (between, insulation, roof_sheathing)
	Compliance checking attribute	NA
	Deontic operator indicator	obligation
	Quantitative relation	provide
	Comparative relation	greater_than_or_equal
	Quantity value	1
	Quantity unit/reference	inch
	Quantity restriction	NA
Horn clause logic representation		$\forall (a, i, r, s) ((\text{airspace}(a) \wedge \text{insulation}(i) \wedge \text{roof\_sheathing}(r) \wedge \text{between}(a, i, r) \wedge \text{has}(a, s)) \supset O(\text{greater\_than\_or\_equal}(s, \text{quantity}(1, \text{inch})))$
<p>Note: Universal quantifier ('<math>\forall</math>' or 'for all') asserts that the sentence is true for all instances of a variable; Conjunction '<math>\wedge</math>': '<math>A \wedge B</math>' indicates that 'A' is true and 'B' is true; Implication '<math>\supset</math>': '<math>A \supset B</math>' indicates that 'A' implies 'B' (if 'A' is true then 'B' is true); Obligation operator (<math>O</math>): <math>O A</math> indicates that 'A' is obligated.</p>		

Fonte: Zhang (2015).

No caso linguagens baseadas em lógicas simbólicas como HC, LPO e LD, o racionalizador consegue fazer inferências a partir das regras de modo independente da aplicação. Nesse caso seria uma solução mais flexível e aberta ao contrário de uma opção *hard-coded*.

Zhang (2015) sugere que as instâncias extraídas do texto e em formatos de tupla sejam transformadas em representações de base lógica para dedução lógica e uma racionalização das expressões sobre critérios de conformidade. Clausula de Horn é o tipo de expressão usada pelos autores Zhang e El-Gohary (2015) para representar de forma lógica (i.e., computável) de uma diretriz do regulamento. Em outras pesquisas como Zhang e El-Gohary (2011, 2016) é usada a Lógica de Primeira Ordem (LPO), um superconjunto lógico que engloba as Cláusulas de Horn (CH). Apesar de possuir mais expressividade, a LPO tem custo computacional maior. Outra opção amplamente aceita em representação ontológica é a Lógica de Descrições (LD), um subconjunto da LPO. Os dois subconjuntos CH e LD possuem notação mais simples e melhor desempenho computacional, portanto são os mais usados na instanciação dessas formulações.

O método de representação em linguagem de representação lógica com Cláusulas de Horn usado por Zhang e El-Gohary (2015), implementado em B-Prolog, é um racionalizador independente de outros programas. É possível utilizar a linguagem Prolog em outras implementações como o SWI Prolog<sup>9</sup>. Essa implementação possui conexões com Python, C e Java de modo a permitir maior interação entre as linguagens e facilidade de codificação. Comparativamente, Prolog é uma linguagem lógica de alto poder semântico utilizada para registrar regras gerais, conhecimento e pressupostos em estruturas de dados com alta eficiência. As linguagens baseadas em objetos (classes) são altamente conhecidas para desenvolvimento de aplicações direcionadas ao usuário, mas voltadas a construção de sistemas de conhecimento (Wielemaker, 2009).

---

<sup>9</sup> <https://www.swi-prolog.org/>. Atualmente existem iniciativas de utilização de Prolog com Python e Web Semântica (RDF) que necessitam análise para verificar sua viabilidade para esse projeto.

Körner *et al.* (2022) mostram diversas vantagens do Prolog como mecanismo de inferência eficiente, analisadores, estruturas de dados imutáveis, linguagem e semântica limpa. Mas também apresenta alguns problemas como a pouca portabilidade, são diferentes das linguagens mais usuais, os ambientes de desenvolvimento de programas atuais (*Integrated Development Environment - IDE*) possuem limitações com essa representação, ela depende de maior curva de aprendizado, entre outras.

### 2.4.3. Modelos semânticos

A diferença entre este modelo e os demais é o seu produto. Assim como a modelagem semântica, o modelo semântico cria uma representação de dados estruturados ou semiestruturados explicitando o sentido dos dados. Ao selecionar os dados de forma estruturada, eles podem ser identificados com rótulos que os classificam em um formato que possa ser compreendido tanto pelos sistemas computacionais como por seres humanos. A criação de modelos semânticos é entendida como uma abordagem de representação que trata da incorporação de novas camadas de informação nos dados, por meio do enriquecimento semântico. Os produtos destes modelos são usualmente representados por linguagens de marcação como XML, JSON e HTML (Figura 13).

**Figura 13 - Regulamento decomposto e marcado em XML**

```
<regulation id="40.cfr.279" name="Standards For The Management Of Used Oil" type="US Federal">
...
<regElement id="40.cfr.279.B" name="Subpart B">
...
< regElement id="40.cfr.279.12" name="Prohibitions">
< regElement id="40.cfr.279.12.a" name="Surface Impoundment prohibition">
<regText>
<paragraph>
Used oil shall not be managed in surface impoundments or waste piles...
</paragraph>
</regText>
</regElement>
<regElement id="40.cfr.279.12.b" name="Use as a dust suppressant">
...
</regElement>
...
</regElement>
...
</regulation>
```

Fonte: adaptado de Kerrigan e Law (2003).

Macit (2014) apresenta o Projeto SMARTcodes como exemplo de modelo semântico. Neste projeto os elementos textuais dos regulamentos são classificados a fim de criar versões “inteligentes” dos regulamentos. A execução seria feita por meio de um protocolo e um programa que geraria representações rotuladas dos regulamentos conforme descrito a seguir:

- 1) Os usuários identificam nos regulamentos as diretrizes que a serem analisadas nos projetos.
- 2) Os usuários rotulam trechos das sentenças utilizando o protocolo RASE que aplica os rótulos “*Requirement*”, “*Applicability*”, “*Selection*” e “*Exception*” (Quadro 3).
- 3) O texto rotulado é processado pelo programa SMARTcodes Builder que checa a validade da estrutura proposta e gera um arquivo XML.

**Quadro 3 - Texto rotulado com o protocolo RASE**

<p><b>ICC IECC 2006 502.5 Moisture control</b>  All &lt;applicability&gt;framed&lt;/applicability&gt; &lt;selection&gt;walls, floors&lt;/selection&gt; and &lt;selection&gt;ceilings&lt;/selection&gt; &lt;exception&gt;not ventilated&lt;/exception&gt; to allow moisture to escape shall be provided with an &lt;requirement&gt;approved vapor retarder&lt;/requirement&gt; having &lt;requirement&gt;a permeance rating of 1 perm&lt;/requirement&gt; (<math>5.7 \times 10^{-11}</math> kg/Pa s m<sup>2</sup>) or less, when tested in accordance with the desiccant method using Procedure A of ASTM E 96. The vapor retarder shall be &lt;requirement&gt;installed on the warm-in-winter side&lt;/requirement&gt; of the insulation. Exceptions: &lt;exception&gt;Buildings located in Climate Zones 1 through 3&lt;/exception&gt; as indicated in Figure 301.1 and Table 301.1. In construction where &lt;exception&gt;moisture&lt;/exception&gt; or its &lt;exception&gt;freezing&lt;/exception&gt; will not damage the materials. Where other approved means to avoid &lt;exception&gt;condensation&lt;/exception&gt; in unventilated framed wall, floor, roof and ceiling cavities.</p>
--

Fonte: Macit (2014).

A necessidade de enriquecimento semântico do modelo como forma de permitir um sistema de verificação automática de conformidade também foi mencionada por Solihin *et al.* (2004). O autor também apresenta a plataforma FORNAX, que foi utilizada pelo sistema de checagem de projetos pelo governo de Singapura (CORENET) para criar o enriquecimento dos arquivos IFC a serem verificados na plataforma. Esse processo

desenvolve uma rotulagem semântica por meio de adição de metadados ao texto (Clarke; Harley, 2014).

Após a análise de outras pesquisas como Zhang e El-Gohary (2021), é possível compreender que o modelo semântico de representação pode ser considerado um meta-modelo que subsidia outras representações. Zhang e El-Gohary (2021) também executam a rotulagem semântica, ou enriquecimento das sentenças, com categorias de elementos como “CR-Relação Comparativa”, “SU-unidade sintática”, “QU-unidade quantitativa” etc. Em seguida utilizam as sentenças rotuladas para treinamento da rede neural de um modelo de aprendizagem profunda.

Para İlal e Günaydin (2017), os modelos de representação baseados em regras, lógica ou semântica são representações “*hard-coded*”, ou seja, representações em um código fixo, muitas vezes compilado no *software*, que exige um alto custo para criação e manutenção. Esse não seria o caso da construção de ontologias ou de uma modelagem semântica que represente ou apoie o processo de extração de informações.

O modelo semântico pode ser construído com diferentes métodos, desde a informação manual, a o reconhecimento de padrões textuais e a rotulagem por meio de mapeamento dos termos em conceitos externos de um banco de dados.

O banco de dados pode ser uma taxonomia que define a hierarquia entre elementos, ou pode ser relacional explicitando relações de existência entre elementos e suas propriedades. Neste caso é consolidada um modelo semântico baseado em ontologia.

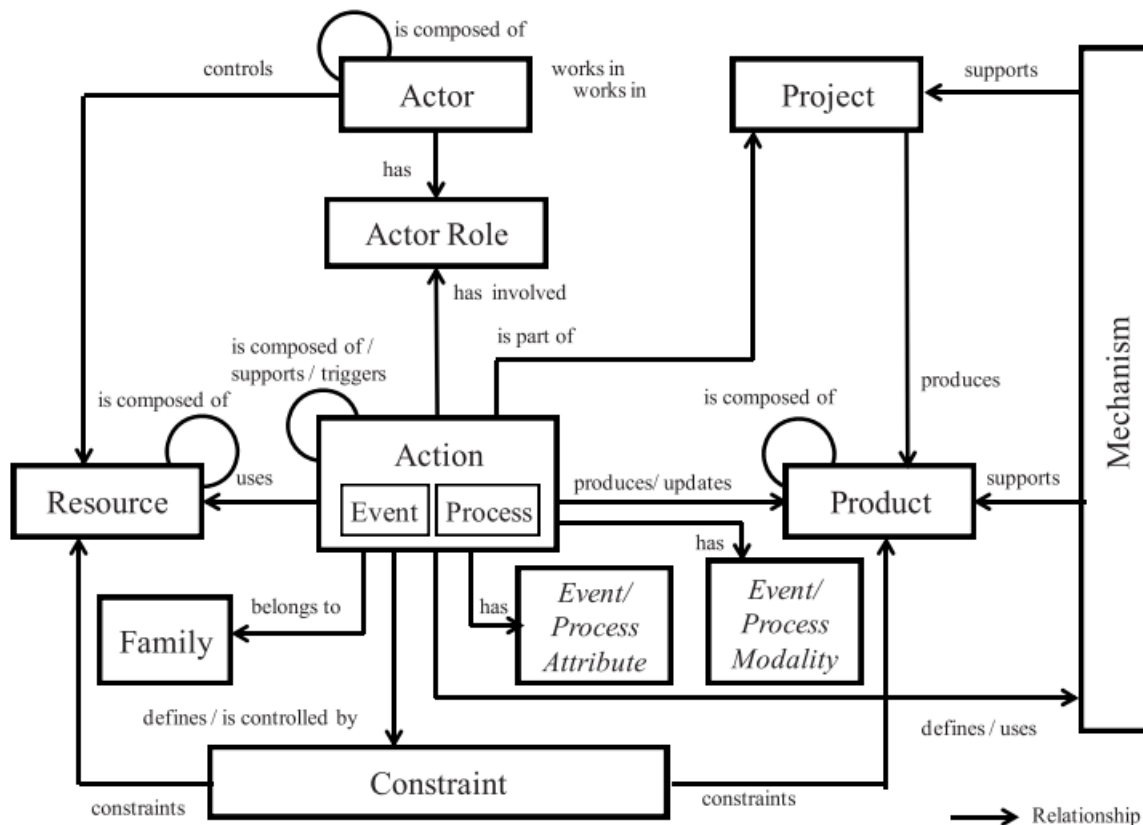
#### **2.4.4. Modelos ontológicos**

Ontologia pode ser compreendida como uma especificação explícita de um vocabulário que represente um domínio de conhecimento e que formaliza nesta representação classes, relações entre as classes, instâncias e outros objetos ou atributos (Gruber, 1993). Portanto, é uma representação, através de linguagem estruturada, dos

conceitos, propriedades e relações de um determinado modelo abstrato de um contexto. No setor AEC, estas representações detalhadas podem descrever o conhecimento de processos e subprocessos relacionados à construção civil.

O uso de ontologias como base de dados de consulta permite que sejam utilizadas informações formalizadas e precisas baseadas na sua formalidade (Almeida, 2021). A formalidade de uma ontologia se refere à construção sistemática da base de dados em formato estruturado e, em alguns casos, formatado em estrutura lógica descritiva que representa de modo explícito e objetivo as relações entre elementos (Figura 14, Figura 15).

**Figura 14 - Modelo geral da ontologia IC-PRO-Onto**



Fonte: El-Gohary e El-Diraby (2010).

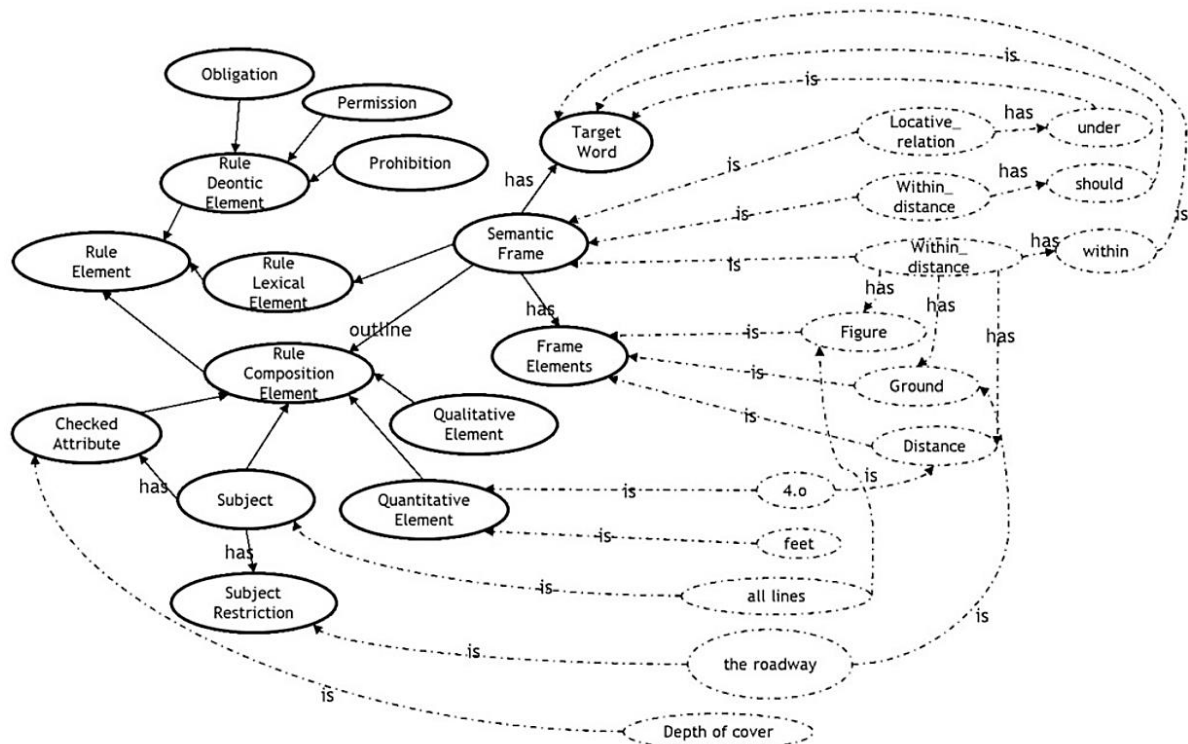
Uma ontologia construída com os métodos convencionados para engenharia de ontologias garante validação dos axiomas que estruturam as relações da ontologia e as

inferências a partir dos dados. A formalização favorece a precisão dos dados evitando ambiguidade e que os resultados das buscas de informação sempre obtenham o mesmo resultado (consistência). Ontologias são formalizadas em arquivos RDF, OWL e em alguns casos XML.

O RDF é um modelo de dados criado pelo *World Wide Web Consortium (W3C)* para representação de modelos de informação. Esses modelos são grupos de conceitos organizados em tuplas formadas pelos campos recurso, propriedade e valor do elemento representado. É um modelo comum também usado representação de ontologias e é expresso em sintaxe XML do tipo `<rdf: RDF xmlns: rdf = "recurso"></rdf : RDF>`.

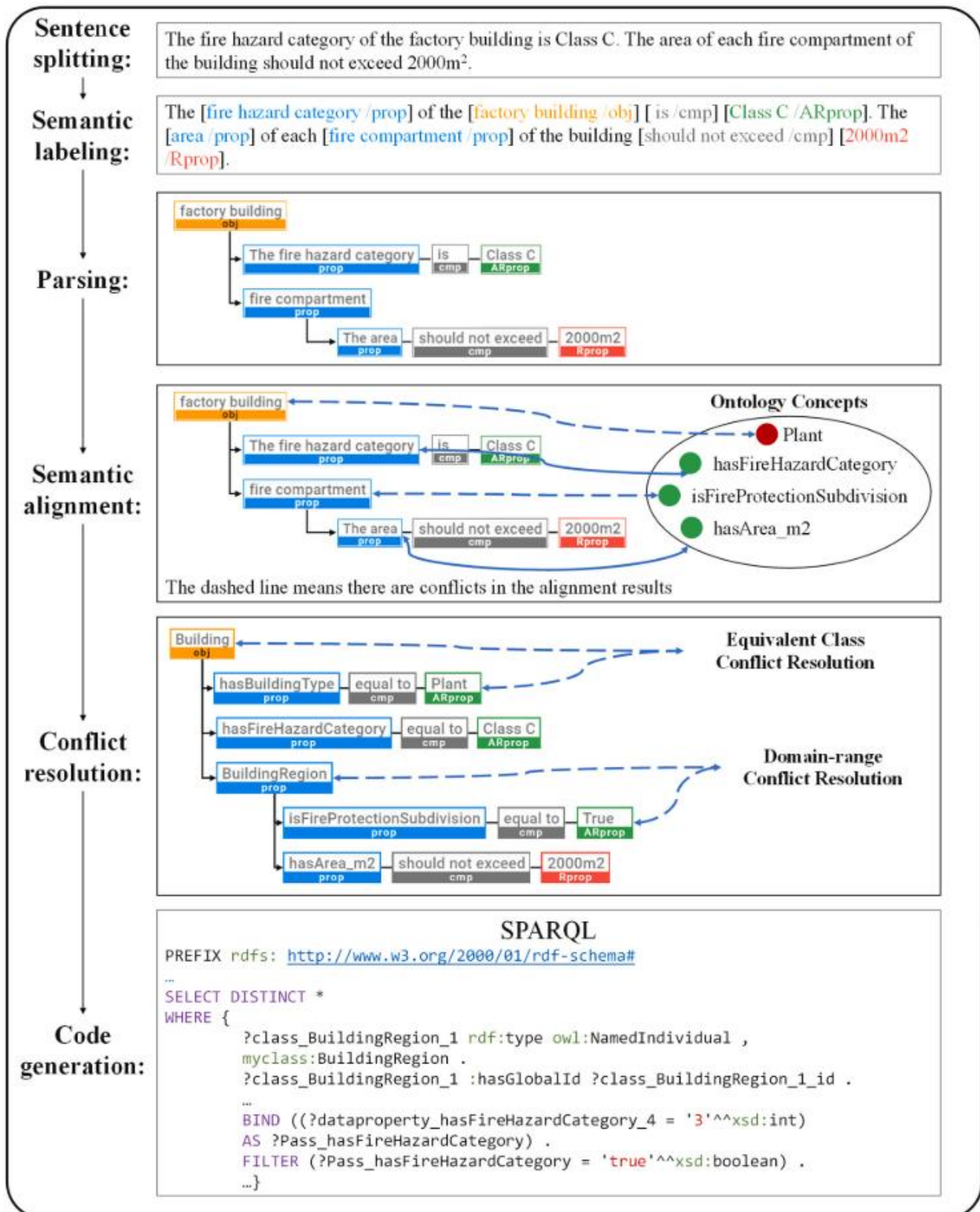
Já o formato OWL é preferido para construção de ontologias por conseguir representar axiomas (regras existenciais) na ontologia e que poderão ser usados para executar inferências e validação da ontologia.

**Figura 15 - Exemplo visual de um mapeamento entre termos de um regulamento e conceitos da ontologia**



Fonte: Xu e Cai (2019)

Figura 16 - Uso de uma ontologia para validação dos requisitos

Fonte: Zheng *et al.* (2022)

Ontologias de domínio são distintas de ontologias gerais que contém menos textos específicos, mas têm maior capacidade de generalização. As ontologias de domínio são uma representação de um determinado conhecimento, portanto deve ser construída, analisada e utilizada a partir de critérios claros e bem definidos que relacionam os conceitos segundo os pressupostos assumidos para a ontologia.

Zhong *et al.* (2012) construíram um processo de verificação completamente baseado no processamento dos axiomas OWL representados em uma ontologia (CQIEOntology). A abordagem de modelagem das regras em formato axiomático em expressões lógicas do tipo SWRL também desenvolve uma proposta de representação de regulamento por esta abordagem ontológica (Menegotto, 2023). Os autores citados consolidam o regulamento de forma relacional, semântico, capaz de ser lido por máquinas, mas depende de intervenção humana para registrar todas as diretrizes neste formato.

Uma das primeiras tentativas em desenvolver a avaliação de conformidade a partir de ontologias parte de sua incorporação em um sistema de busca (*queries*) utilizando SPARQL (Figura ). Esta abordagem foi utilizada a partir de 2008 por diversos autores mencionados por Pauwels, Zhang e Lee (2017).

É possível existir diferentes termos e expressões se referindo ao mesmo objeto ao logo do texto e isso pode dificultar a análise de múltiplos regulamentos que citam o mesmo termo, mas com sentidos diferentes. Uma das vantagens do uso de ontologias é a eliminação desse tipo de inconsistência no regulamento. A ontologia surge como meio de mapear e garantir que diferentes termos estejam relacionados ao mesmo objeto e as relações que ele estabelece com outros termos (Hjelseth, 2009). A essa ação, dá-se o nome de desambiguação do sentido da palavra (*word sense disambiguation*) que é a ação de entender o sentido do termo no contexto em que está inserido. Por exemplo, a palavra “banco” pode representar uma instituição financeira ou um mobiliário urbano. Nesse caso, ambas as palavras “banco” são homônimas e a desambiguação da frase pode ser verificada pelas palavras e expressões próximas para definir o sentido da palavra

ambígua. No exemplo acima, se o texto remeter a termos financeiros é mais provável que a frase se refira a “banco” como instituição financeira.

A modelagem semântica, i.e., a construção de ontologias, permite incorporar relações entre elementos a fim de elaborar um referencial. Utilizando uma ontologia os sistemas podem executar uma ação de busca para comparar os termos e indicar o sentido esperado para aquele elemento.

Alguns dos trabalhos mais recentes que incorporam Processamento de Linguagem Natural (PLN) e ontologias são as pesquisas desenvolvidas por Cheng *et al.* (2009), El-Gohary e El-Diraby (2010), Zhong *et al.* (2012), Zhang e El-Gohary (2011, 2012, 2014, 2015, 2016, 2017a, 2017b), Zheng *et al.* (2022) e Ren *et al.* (2022). Nesses trabalhos a incorporação de ontologias ocorre pelo registro de relações entre elementos de determinado domínio do conhecimento e de modo formalizado. O mapa de relações de uma ontologia consegue registrar as relações entre elementos e, portanto, a semântica inerente daquele domínio (Cheng *et al.*, 2009).

Segundo Cheng *et al.* (2009), a construção de uma ontologia específica é um meio de automatizar o processo de integração de dados e sua adoção pode facilitar o acesso aos regulamentos governamentais. Eles desenvolveram um sistema que mapeia conceitos dos regulamentos a outros conceitos similares em taxonomias do setor da construção civil (IFC, OmniClass e CIMsteel). Esse sistema não é detalhado, mas indica a possibilidade de auxiliar a compreensão, usabilidade e alinhamento dos conceitos dos regulamentos locais às taxonomias universais.

Em Zhang e El-Gohary (2016) a incorporação de ontologia permitiu a identificação de atributos semânticos do texto porque mapeava o domínio específico do conhecimento usado no regulamento. Com uso de um modelo semântico e técnicas PLN foi possível analisar sentenças complexas que seriam muito complexas para outros métodos de extração de informação. Em 2011 os autores perceberam que a incorporação da ontologia ampliou o desempenho da Extração de Informações de modo significativo quando

comparado com a extração baseada apenas em atributos sintáticos das sentenças (Zhang; El-Gohary, 2011).

O modelo de representação baseado em ontologia é experimentado em diversas áreas do conhecimento. Embley *et al.* (1998) desenvolvem um modelo de extração de informação e representação do conhecimento que se aproxima do modelo desenvolvido por Zhang e El-Gohary (2011, 2016). A partir do uso de uma ontologia para permitir o mapeamento de entidades e suas relações, Embley *et al.* (1998) extraem “a informação de interesse de um texto não estruturado e a reformula em um documento estruturado”. Os autores explicam que um processo de Extração de Informações baseado em ontologia é resiliente a mudanças no formato dos documentos fonte uma vez que a intervenção humana ocorre somente na etapa de definição da ontologia.

O trabalho desenvolvido por Ren *et al.* (2022) é um dos mais recentes na aplicação de ontologia para extração de fatos legais de documentos regulatórios chineses. Na pesquisa eles explicam que a ontologia torna o domínio do conhecimento explícito, e tem se comprovado como o método mais adequado para extração de informação de textos de domínio específico como regulamentos.

É importante notar no trabalho de Ren *et al.* (2022) que mesmo abordando o problema por meio da incorporação de um sistema baseado em Redes Neurais Profundas (*Deep Neural Networks*), o resultado para extração de informações baseado em regras se mostrou mais eficaz 99,7% contra 91,43% usando Redes Neurais Profundas. Com resultados similares Zhan e El-Gohary (2021) aplicam Redes Neurais profundas e *Transfer Learning* atingindo desempenho de 94% (F1-measure). Comparativamente, o caso de aplicação de ontologia, mas baseado em regras de Extração de Informações em Zhang e El-Gohary (2011) atingiu um desempenho semelhante, 94,3%. O que pode indicar é que a diferença entre os desempenhos se reflete na escalabilidade das soluções e não na qualidade das informações extraídas.

## 2.5. Representação de regulamentos em BIM

Para garantir a validade das regras e dos dados do modelo antes de seguir para a avaliação de conformidade, é necessário incluir o processo de checagem do modelo (*model checking*) que é uma etapa mais ampla que agrega outras etapas além da etapa de verificação de conformidade a regras. A checagem do modelo idealmente ocorre em fases de coordenação do projeto e contempla as etapas de validação do modelo, a checagem de interferências e a verificação de conformidade de regras.

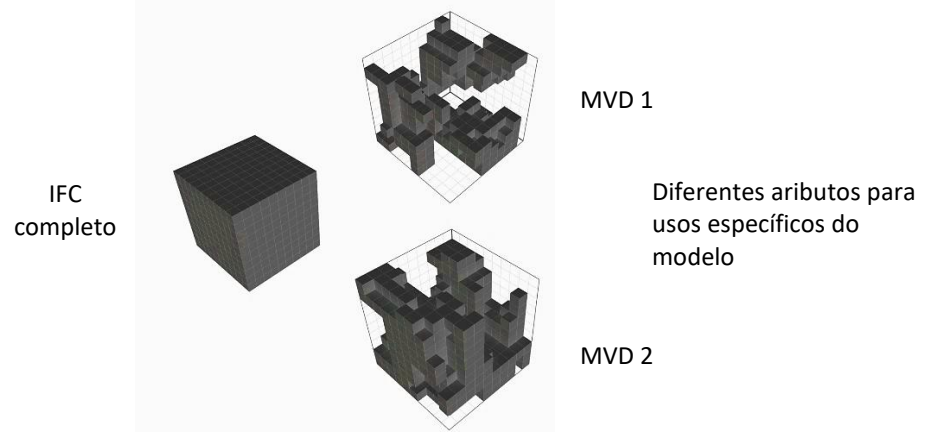
Para que o modelo possa ser verificado, a sua construção demanda informações e documentos que orientam quais os dados que devem estar contidos no modelo para a correta troca de informações entre as equipes. Entre os modelos de informação existentes, destacam-se o *Model View Definition* (MVD) e a especificação de entrega de informações (*Information Delivery Specification* - IDS). Outros requisitos de informação são apresentados pela NBR ISO 19650-1 e demandam diferentes informações relacionadas, desde protocolos de informação e dados de componentes às metas e indicadores da organização. São eles: Requisitos de Informação da Organização (OIR), Requisitos de Informação do Ativo (AIR), Requisitos de Informação do Projeto (PIR), os Requisitos de Troca de Informação (EIR) e o Plano de Entrega de Informação (MIDP).

O IDM é um documento complexo que mapeia as trocas de informações entre equipes e se apresenta como um documento de consulta que orienta quais informações serão trocadas entre os agentes durante a fase de projeto.

O MVD é um modelo que seleciona, dentre todos as entidades e propriedades do IFC quais as que serão demandadas em determinado uso. Estes modelos permitem que as equipes de projeto mantenham apenas as informações necessárias para o uso daquele modelo (Figura 17). As pesquisas relacionadas à verificação de conformidade apresentavam o MVD como meio de garantir consistência das informações e as demandas de informação do modelo. Recentemente o MVD tem sido substituído por outro formato

de arquivo e informações, o IDS, devido à baixa interoperabilidade do MVD em diferentes programas (BuildingSMART International, 2022).

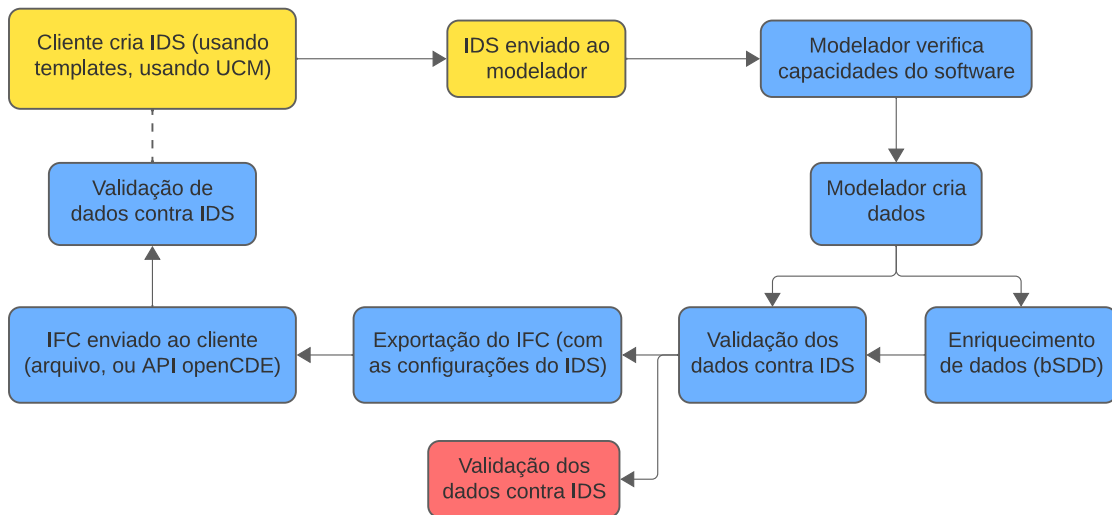
**Figura 17 - Representação abstrata do IFC e dois MVDs**



Fonte: o autor.

Já o IDS constitui um documento em formato interoperável (XML) que unifica os requisitos de projeto, vinculados os termos às informações constantes no IFC como classes e propriedades. Porém, não é do escopo de um arquivo IDS representar os requisitos em forma de regra para ser validada, mas indicar quais propriedades e elementos IFC devem constar no modelo para que seja feita a análise de conformidade do projeto (BuildingSMART International, 2022) (Figura 18).

**Figura 18 - IDS dentro de um fluxo típico de projeto BIM**



Fonte: adaptado de BuildingSMART International (2020).

Como demonstrado na Seção 2.3, as soluções disponíveis no mercado exigem uma representação restrita dos regulamentos para serem processados pelos programas. As avaliações são obtidas parcialmente em soluções *black-box* porque os métodos de verificação não são explicitados pelos programas e permitem apenas uma personalização controlada, com poucos parâmetros. Ao externalizar as sentenças lógicas, por meio de linguagens lógicas, garantem-se a liberdade de implementação em diferentes aplicações.

Apesar de haver poucas aplicações que utilizam esse tipo de representação, a aplicabilidade desse tipo de representação lógica pode ser feita por meio de um racionalizador<sup>10</sup> embutido no programa (Hjelseth; Li, 2021).

A opção contrária à essa capacidade de implementação genérica, seria criar uma representação em linguagem específica, mais popular, como Javascript ou Python, mas são linguagem restritas a aqueles programas que as implementam para a representação de regras, como a plataforma KBim. As linguagens textuais são dinâmicas e atualizadas frequentemente, mas elas dificultam o desenvolvimento de soluções que podem ser implementados em outras linguagens. Desta forma, havendo uma opção de representação mais genérica, as regras permitiriam otimização da aplicação ao reduzir o tempo e custo computacional da implementação.

Algumas aplicações BIM como Solibri permitem inserção de dados em formato semiestruturado aberto como XML, mas nesse caso o processamento do regulamento não é feito pela racionalização do arquivo XML. Apesar do XML poder incorporar regras em suas marcações por meio de esquemas que traduzem a lógica das regras, no Solibri o XML é um formato de entrada de dados e que depende da aplicação para racionalizar e validar as regras.

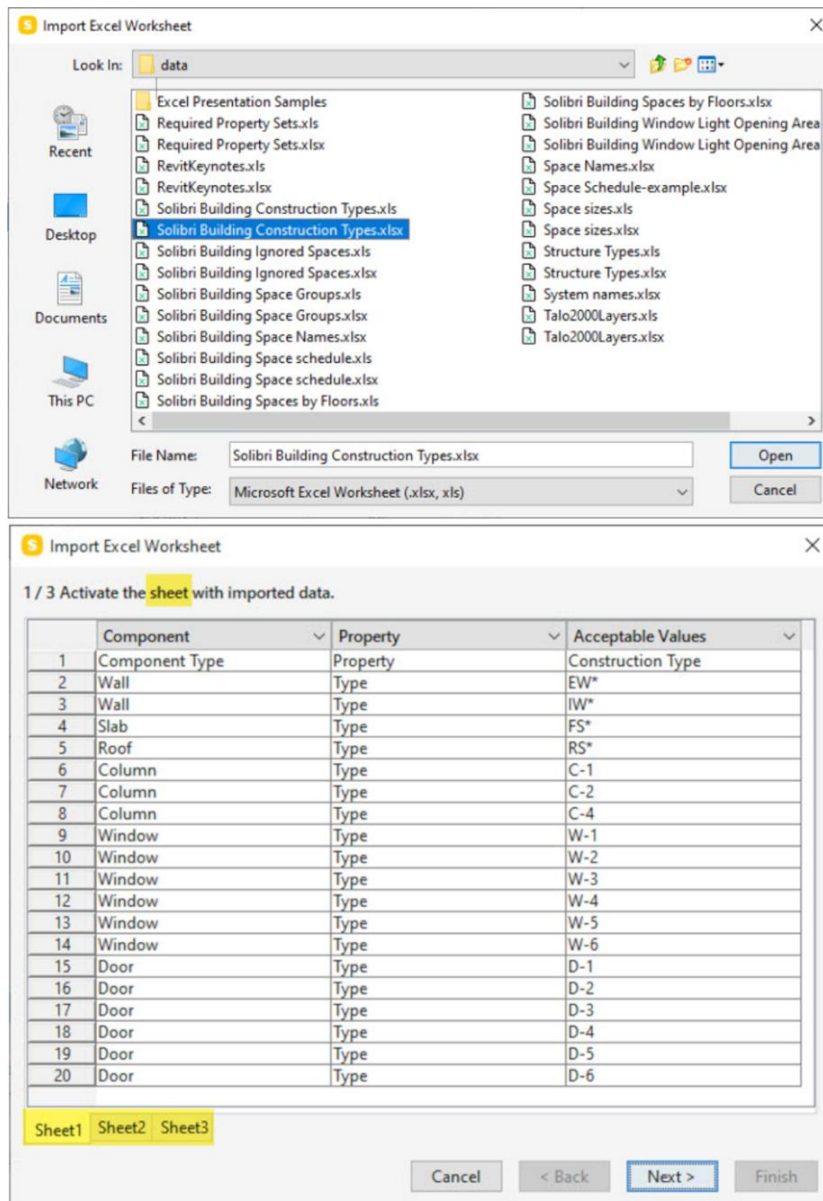
Com relação às ferramentas disponíveis no mercado, o Solibri Model Checker organiza os componentes de um requisito em quatro categorias: componente, propriedade, operador e valor. Componente se refere ao objeto IFC que será avaliado;

---

<sup>10</sup> <http://www.cs.man.ac.uk/~sattler/reasoners.html>

propriedade se refere às dimensões geométricas; operador se refere a um termo comparativo (menor, igual, etc.); e valor se refere a um número (Figura 19). O Solibri permite que as regras sejam formatadas segundo estes rótulos e em formato XML e XLSX, mas os termos que nomeiam os rótulos devem ser uniformizados em todo o projeto para evitar erros de correspondência entre o item a ser verificado pela regra e o componente da construção que será avaliado.

**Figura 19 - Interface de importação de regras do Solibri**



Fonte: adaptado de Solibri (2020).

O Autodesk Revit Model Checker, outra ferramenta de geração de regras para análise de conformidade também aceita formatos de regras em linguagem XML (Figura 20). No esquema definido pelo programa estão presentes os seguintes elementos XML:

- Descrição do documento e cabeçalho do arquivo XML: autor, versão, ID, data, LastModified, etc.
- Configurações de checagem : Critérios de análise predefinidos (*fail When no Matching Elements are found, fail When Matching Elements are found, CountOnly, CountAndList*).
- Operadores: *And, Or, Exclude*.
- Categorias: Elementos do modelo relacionados a atributos IFC (*Room, Host, Family, PhaseCreated*, etc.).
- Property: Atributos relacionados aos dados do modelo, similar ao atributo Property do IFC (*Name, Elevation, Is Defined*, etc.).

### Figura 20 - Exemplo de conversão XML para o programa Autodesk Model Checker Configurator

```
<Check CheckName="To meet the thermal insulation requirement, concrete walls that having a width less than 200 mm shall have a thermal resistance no less than 0.1 m2K/W" ResultCondition="FailMatchingElements">
  <Filter Category="TypeOrInstance" Condition="Equal" Property="Is Element Type" Unit="Default" UnitClass="None" Value="0" />
  <Filter Category="Category" Condition="Included" Property="OST_Walls" Value="True" />
  <Filter Category="Parameter" Condition="Wildcard" Property="Structural Material" Unit="Default" UnitClass="None" Value="Concrete" />
  <Filter Category="Parameter" Condition="LessThan" Property="Width" Unit="Default" UnitClass="Length" Value="0.6561679790026247" />
  <Filter Category="Parameter" Condition="LessThan" Property="Thermal Resistance (R)" Unit="Default" UnitClass="None" Value="0.1" />
</Check>
```

Fonte: adaptado de Zhou *et al.* (2022)

Em ambos os casos, o arquivo com diretrizes projetuais é apresentado em linguagem de marcação (XML), mas a partir de esquemas diferentes. Enquanto o Solibri adota os rótulos “componente, propriedade, operador e valor”, o programa da Autodesk exige os rótulos Critérios de análise, operador, CategoriaIFC e PropriedadeIFC.

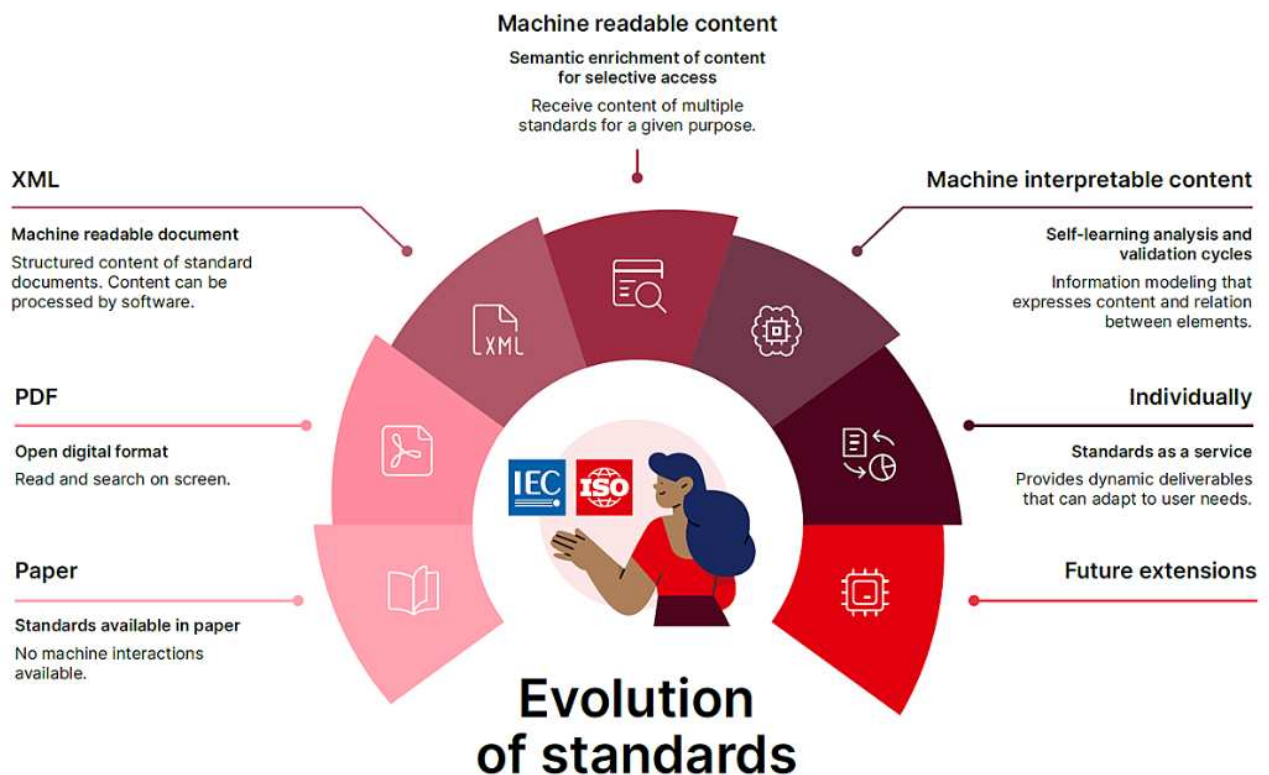
Observam-se dois estudos brasileiros que visam a representação da Norma de Desempenho em BIM para a verificação automática em projeto. Silva (2017) desenvolveu uma solução para empreendimentos habitacionais do PMCMV por meio da plataforma Solibri Model Checker. Silva Júnior e Mitidieri Filho (2018) propuseram incorporar a

verificação automática na própria ferramenta de modelagem BIM. Ambas as soluções, por aplicativo de verificação ou de modelagem BIM, codificaram aproximadamente um terço dos requisitos de projeto da norma pela checagem de forma automatizada.

## 2.6. Dados legíveis por máquina e a ISO Smart

Um texto escrito em formato .doc, .odt ou .pdf pode ser intercambiável entre diferentes programas e sistemas operacionais, mas seu conteúdo não é capaz de ser processado diretamente por uma aplicação que dependa da informação contida no arquivo. Essa característica indica que o documento não está em formato computável ou capaz de ser lido por máquinas (Figura 21).

Figura 21 - Diferença entre formatos abertos e computáveis



Fonte: International Electrotechnical Commission (2022b)

A iniciativa ISO Smart Standards deseja incorporar regulamentos lidos por máquinas no ciclo de vida do produto e de serviços para acelerar o desenvolvimento a um custo menor e garantir conformidade com regulamentos atualizados. O uso de regulamento capaz de ser lido por máquina garante a desenvolvedores (e projetistas) mais atenção à criação do conteúdo do projeto de modo mais efetivo usando ferramentas digitais que automatizarão o processo por todo ciclo de vida de seu desenvolvimento (IEC, 2022a). A iniciativa da ISO Smart inicialmente contempla diferentes países e possui representação brasileira da ABNT como a liderança ISO das Américas (IEC, 2022b).

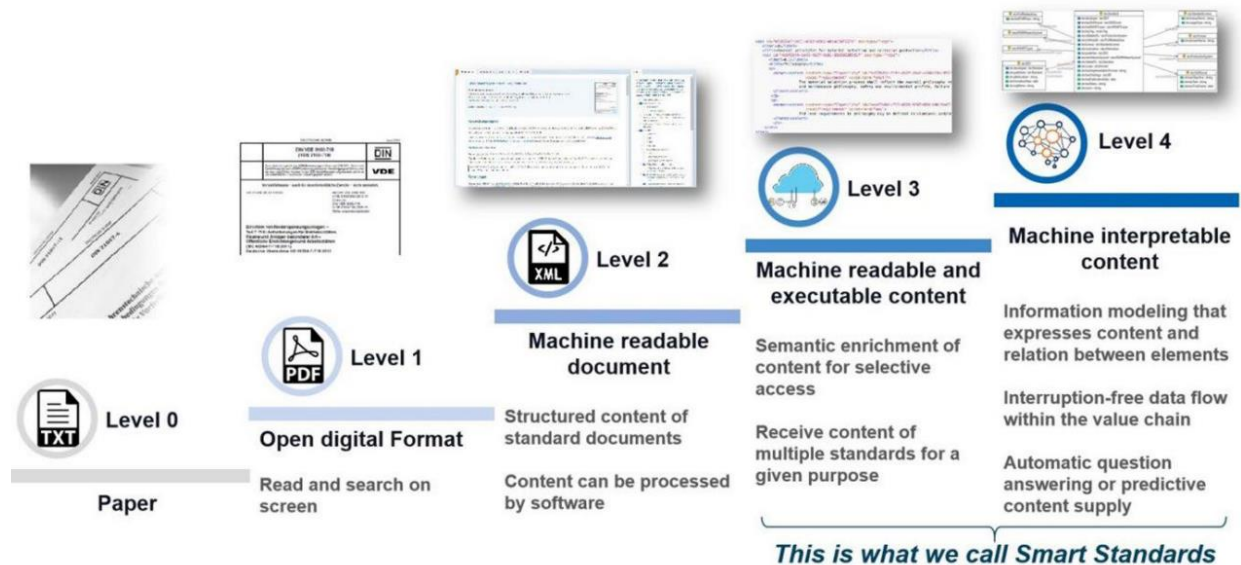
Neste paradigma as normas poderão ser atualizadas com mais frequência e disponível para os usuários relevantes. Uma vez formatadas em padrões universais e computáveis, serão mais bem integradas na vida cotidiana, sendo facilmente incorporadas em computadores, máquinas complexas e dispositivos inteligentes.

Para a ISO, uma norma “SMART” representa um documento padronizado, lido, aplicado e transferido por máquinas. É um documento claro, preciso e sem ambiguidade que possui conteúdo enriquecido semanticamente, ou seja, marcado com rótulos e atributos que definem metadados daquela norma.

Propõem-se cinco classes de regulamentos possíveis de serem usados na indústria (Figura 22). O primeiro nível é o documento em meio físico, em formato de papel. O segundo é o formato digital, capaz de ser **apresentado** e enviado para outros computadores em formato digital, como .pdf. O terceiro nível descreve documentos capazes de serem **lidos** e **processado** por máquina; são documentos digitais, estruturados e com conteúdo capaz de ser processado em algum programa, por exemplo um arquivo .xml. O quarto nível envolve documentos com conteúdo capaz de ser lido por máquinas e possui conteúdo **enriquecido** e referenciado por múltiplas normas e regulamentos. Neste nível será possível executar inferências a partir das informações marcadas no documento. O quinto nível representa um documento capaz de ser **interpretado** porque é modelado em formato de dados que expressam relações entre elementos. Torna-se um

banco de dados relacional, que pode receber buscas inteligentes que analisem e retornem seu conteúdo (Figura 22).

Figura 22 - Graus de capacidade de processamento das normas ISO



Fonte: IEC (2023).

As primeiras reuniões do grupo de trabalho da ISO Smart definiram que algumas das normas relacionadas à construção civil possuíam urgência imediata para conversão formato computável, em prioridade acima de setores como indústria aérea, veículos e biotecnologia (Quadro 4).

Na lista de comitês com urgência em conversão de conteúdo, o comitê técnico TC 59<sup>11</sup> trata de trabalhos relacionados a edifícios e engenharia civil. Dentre os regulamentos produzidos pelo grupo destacam-se: organização da informação nos processos de projeto, fabricação e construção; requisitos geométricos, elementos e componentes construtivos; requisitos de desempenho, funcionais e do usuário relacionados à vida útil, sustentabilidade, acessibilidade e usabilidade; e impactos e aspectos econômicos, ambientais e sociais relacionados ao desenvolvimento sustentável.

<sup>11</sup> <https://www.iso.org/committee/49070.html>

**Quadro 4 - Grupos de normas organizados pela urgência de digitalização em formato capaz de ser lido por máquina**

Noted Immediate industry demands for SMART standards in SAG MRS survey and have submitted a proposal	TC 59 (buildings and civil engineering works)
	TC 37 SC 3, SC 2 (language and terminology)
	TC 211 (geographic information / geomatics)
Noted emergent industry demands for SMART standards in SAG MRS survey and have submitted a proposal	TC 23 SC 5 (tractors and machinery for agriculture and forestry)
	TC 184 SC 4, SC 5 (automation systems and integration)
	TC 171 SC 2 (document file formats)
Noted potential for conversion of existing standards to SMART standard	TC 29 (small tools)
	TC 46 (information and documentation)
	TC 67 (materials, equipment and offshore structures for petroleum, petrochemical and natural gas industries)
	TC 154 (processes, data elements and documents in commerce, industry and administration)
	TC 163 (thermal performance and energy use in the built environment)
Other TCs having submitted a proposal for ISO SMART	TC 205 (Building environment design)
Other TCs noting immediate or emergent demands for SMART standards in SAG MRS survey	TC 17 (steel)
	TC 20 (aircraft and space vehicles)
	TC 22 (road vehicles)
	TC 28 (petroleum and related products)
	TC 38 (textiles)
	TC 68 (financial services)
	TC 193 (natural gas)
	TC 228 (tourism and related services)
	TC 307 (blockchain and distributed ledger technologies)
	JTCA1 (information technology)
	TC 131 (fluid power systems)
	TC 175 (jewellery and precious metals)
	TC 276 (biotechnology)
	CASCO (conformity assessment)

Fonte: adaptado de IEC (2022b).

No projeto da Smart Standards, que ocorre em parceria com outras entidades mundiais como a ABNT, a IEC/ISO propõe um modelo que possa ser lido e interpretado não somente por computadores em escritórios, mas em conceito amplo, ou seja, para processamento em máquinas e sistemas digitais. Exemplo desta nova compreensão é dado pela conversão do termo *"computer interpretable data"* que tem sido convertido para *"machine-interpretable data"* na ISO 10303-232:2002 que trata de Sistemas industriais

automáticos e integração, ampliando o conceito para máquinas que possuem capacidade de processamento computacional.

Em linhas gerais, um regulamento “Smart” seria capaz de interpretar os dados publicados e relacionados com atributos dos elementos de projeto. Apesar da ISO ainda não definir como pode ser feita esta interação, eles avançaram com a proposta de criação de um arquivo XML com diferentes categorias de processamento dos regulamentos.

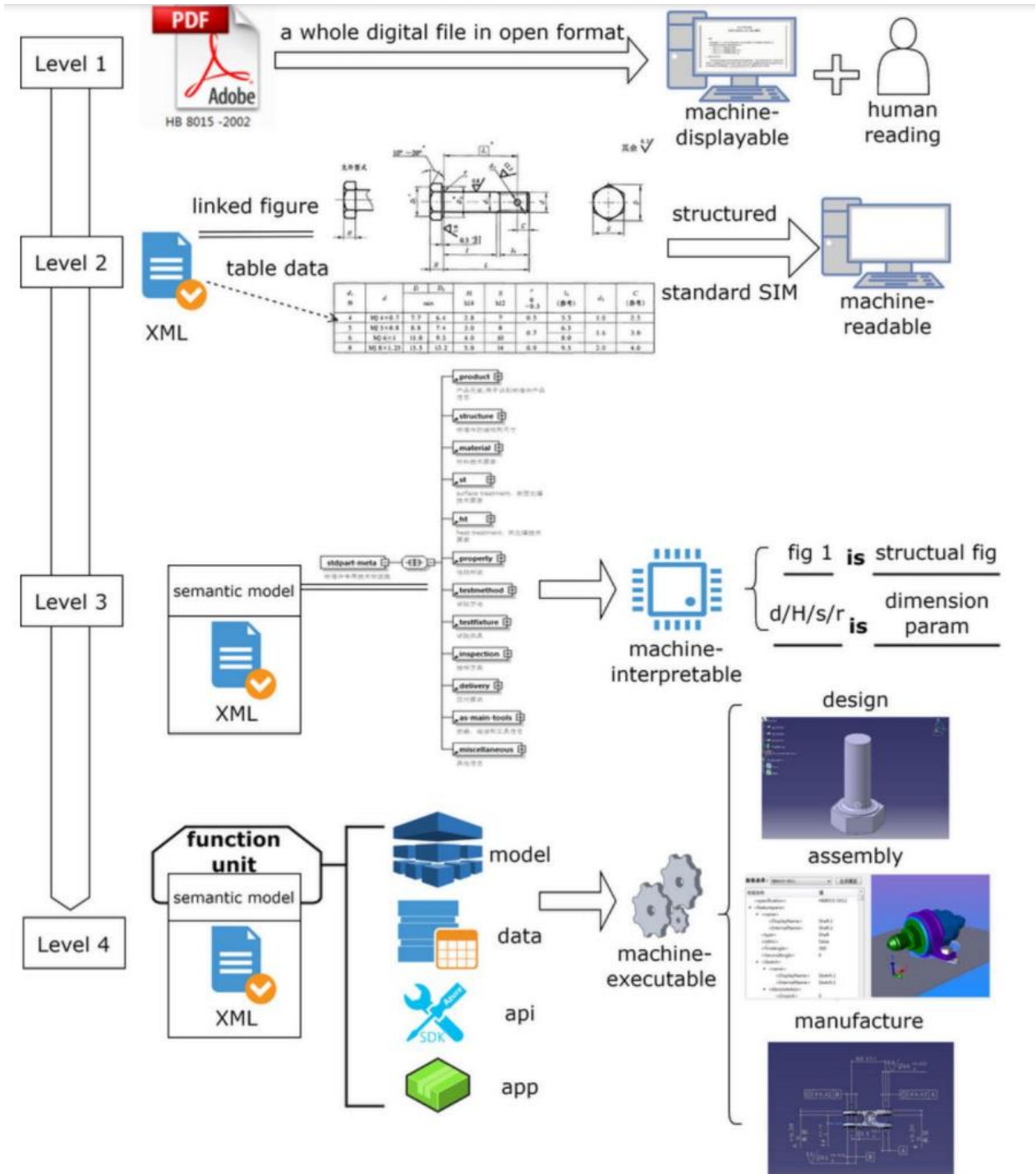
No exemplo da Figura 23, é possível compreender a correspondência entre diferentes níveis de capacidade do regulamento e o tipo de uso. Por exemplo, no nível 3, o arquivo XML já apresenta informação estruturada e anotada de modo que pode ser extraída uma lista de atributos de determinada diretriz e que seja capaz de ser processada por máquina. Ao ser processada, são validados atributos como parâmetros dimensionais. A diferença desta proposta da ISO Smart para o que já ocorre nos programas de verificação automática (como Solibri), é que neles as regras são configuradas pelo usuário manualmente (nível 1 da ISO Smart).

### **2.6.1. Linguagem de marcação XML**

Segundo o relatório NISTIR 7511: 2016 (Cook *et al.*, 2016), um documento capaz de ser lido por máquinas “está estruturado, tipicamente em XML e pode ser consumido por outro programa usando uma lógica de processamento consistente”.

Como apresentado pelo projeto ISO Smart, a XML é uma linguagem de marcação extensível (*Extensible Markup Language*) de propriedades simples, que pode ser estendida com incorporação de esquemas de notação. O projeto indica que a linguagem seja usada a partir do segundo nível de digitalização dos regulamentos, porque a linguagem de marcação é capaz de estruturar um texto agregando rótulos e notações, incorporando semântica aos dados.

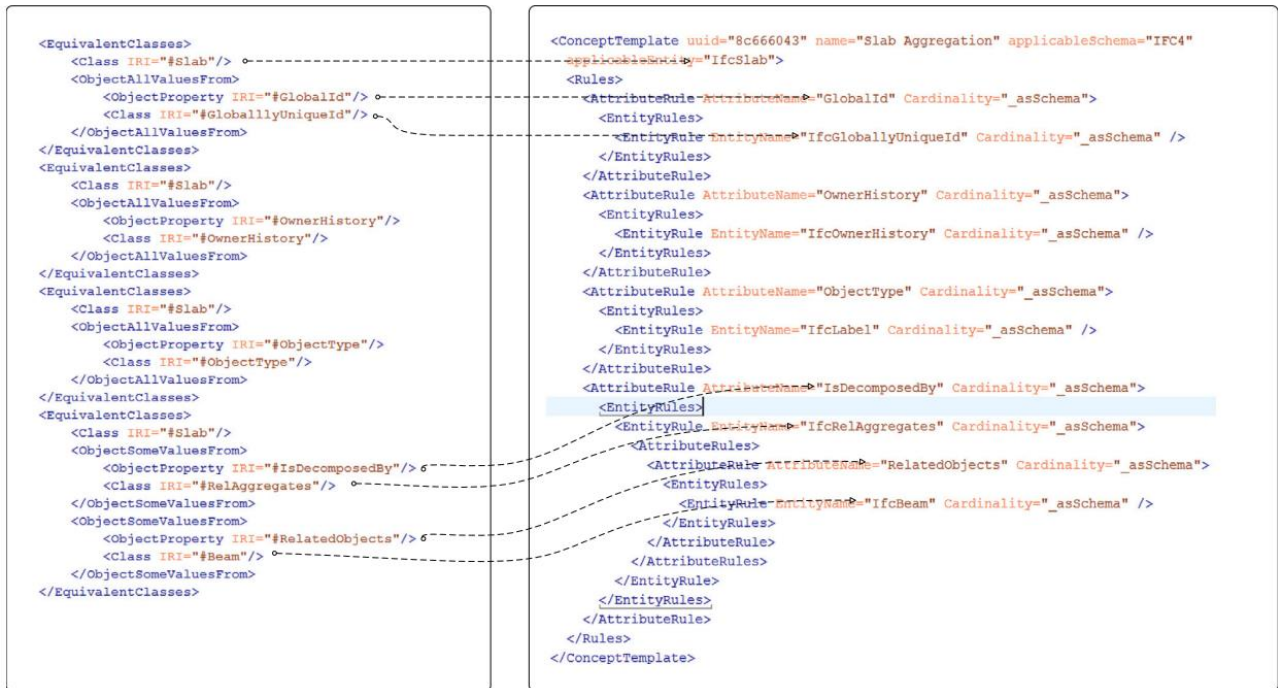
Figura 23 - Exemplo da aplicação de diferentes categorias de Smart standards na indústria de aviões



Fonte: IEC (2023).

Uma linguagem de marcação, por definição, adiciona rótulos (*tags*) e atributos em um documento (Figura 24). A criação de um arquivo XML implica em organizar o texto por elementos e subelementos previamente definidos que serão os rótulos usados para marcação do texto.

Figura 24 - Exemplo de alinhamento entre uma ontologia (OWL) e outro XML



Fonte: Lee, Eastman e Solihin (2016)

Um arquivo XML possui dados estruturados como uma árvore hierárquica de relações. É uma estrutura que explicita os vínculos entre elementos que possuem apenas uma classe pai e zero ou mais elementos filhos. Cada elemento é rotulado com a sintaxe <rótulo >texto</ rótulo>. Além disso cada elemento pode receber atributos predefinidos em seus rótulos <rótulo atributo = 'valor'>texto<rótulo>. Desta forma é possível construir uma estrutura hierárquica que armazena atributos de cada elemento e que estabelece interdependência entre elementos.

O XML ainda é um formato de arquivo capaz de ser incorporado em ferramentas de avaliação de conformidade. O Solibri, por exemplo, além de permitir que as regras

sejam construídas na sua interface principal, possui uma ferramenta de importação de regras definidas pelo usuário e que recebe arquivos em formato XML.

Como a linguagem utilizada para registro dos textos deve incorporar atributos dos termos e das sentenças, o esperado é que essa informação seja construída em um modelo de marcação. Nesse sentido, a linguagem HTML não atende porque possui marcações previamente definidas. Já a *Extensible Markup Language* (XML) fornece a capacidade de um *framework* para representar textos anotados e recursos léxicos.

Como uma linguagem de marcação aberta, o XML permite que sejam criados atributos personalizados o que amplia a capacidade de anotação, mas impede a interoperabilidade com códigos e sistemas que utilizam outras marcações. Devido a esse problema são adotados '*schemas*' ou esquemas de marcação definidos por grupos, agências ou entidades sobre determinado assunto. Nesse sentido existem os esquemas ifcXML para marcação segundo o esquema do IFC entre outros vinculados ao setor da AEC como bcfXML, mvdXML, gbXML e LegalRuleML.

Kerrigan e Law (2003), Li, Raskin e Ramani (2008) e Nirenburg e Raskin (2004) se basearam em uma sintaxe de marcação XML para registrar seus regulamentos. Bouzidi *et al.* (2010) mostram um exemplo de conversão dos arquivos regulatórios em formato XML para formar um corpo técnico de regulamentos e uma posterior conversão em formato RDF (*Resource Description Framework*).

Com a intenção de criar um formato estruturado capaz de ser processado por computadores, o projeto brasileiro **LexML**, de 2009, se baseou na iniciativa italiana LexML-IT para estruturação das informações legislativas e jurídicas nacionais. Em 2008 a iniciativa também foi desenvolvida nos Estados Unidos usando a linguagem de marcação StratML<sup>12</sup>. Em 2019 a Secretaria de Tecnologia da Informação (Prodasen) publicou no portal do projeto LexML todas as normas jurídicas nacionais com a marcação

---

<sup>12</sup> [https://stratml.us/references/HR4174.htm#Title\\_II](https://stratml.us/references/HR4174.htm#Title_II)

'Legislation'<sup>13</sup>. Essa iniciativa permite que os dados da legislação brasileira sejam publicados de modo estruturado com a identificação de seus metadados pelos provedores Web (Lima, 2019) (Figura 25). Esse esquema de marcação baseado em XML já está consolidado com marcações para documentos jurídicos (Artigo, Caput, Paragrafo, Inciso, Alinea, etc), mas ainda não foi orientado para sua aplicação em normas técnicas.

**Figura 25 - Exemplo de marcação XML na Constituição Federal de 1988**

```
<Artigo id="art1">
  <Rotulo>Art. 1º</Rotulo>
  <Caput id="art1_cpt"><p>A República Federativa do Brasil, formada pela união
indissolúvel dos Estados e Municípios e do Distrito Federal, constitui-se em Estado
democrático de direito e tem como fundamentos:</p>
  <Inciso id="art1_cpt_inc1"><Rotulo>I - </Rotulo><p>a soberania;</p></Inciso>
  <Inciso id="art1_cpt_inc2"><Rotulo>II - </Rotulo><p>a cidadania;</p></Inciso>
  <Inciso id="art1_cpt_inc3"><Rotulo>III - </Rotulo><p>a dignidade da pessoa
humana;</p></Inciso>
  <Inciso id="art1_cpt_inc4"><Rotulo>IV - </Rotulo><p>os valores sociais do
trabalho e da livre iniciativa;</p></Inciso>
  <Inciso ID="art1_cpt_inc5"><Rotulo>V - </Rotulo><p>o pluralismo político.</p>
  </Inciso>
</Caput>
  <Paragrafo ID="art1_par1"><Rotulo>Parágrafo único.</Rotulo><p>Todo o poder emana
do povo, que o exerce por meio de representantes eleitos ou diretamente, nos termos
desta Constituição.</p></Paragrafo>
</Artigo>
```

Fonte: Brasil (2008).

Observa-se que no Portal Brasileiro de Dados Abertos existe a recomendação de uso de arquivos em formato não proprietário e estruturado para conversão de regulamentos. O formato não proprietário (livre) e estruturado possibilita seu uso irrestrito e automatizado através da Web e em formato amplamente conhecido (Brasil, 2016). Dessa forma entre os formatos aceitos e recomendados pelo Padrões de Interoperabilidade de Governo Eletrônico – ePING<sup>14</sup> estão os formatos XML, RDF, CSV, TXT, JSON, OWL, SKOS, SPARQL e LexML.

<sup>13</sup> <https://schema.org/Legislation>

<sup>14</sup> <https://eping.governoeletronico.gov.br/>

## 2.7. Métodos de extração de informações

### 2.7.1. O Processamento de Linguagem Natural

O Processamento de Linguagem Natural (PLN) trata de desenvolver procedimentos de análise e reconhecimento de padrões textuais por meio de análise computacional do texto. As abordagens atuais para o PLN, como os demais campos da Inteligência Artificial (IA), podem estar orientadas a processos baseados em regras ou em aprendizagem de máquina. No campo da linguística computacional os sistemas de PLN tratam de construção de regras para executar determinadas ações e em PLN isso requer a participação de pessoas com certo grau de conhecimento no domínio do conhecimento (Vajjala *et al.*, 2020).

Atualmente os paradigmas de extração de informações, pertencentes ao PLN, estão divididos entre dois métodos computacionais, o baseado em regras e o baseado em aprendizagem de máquina.

### 2.7.2. Procedimentos de extração de informação baseada em regras

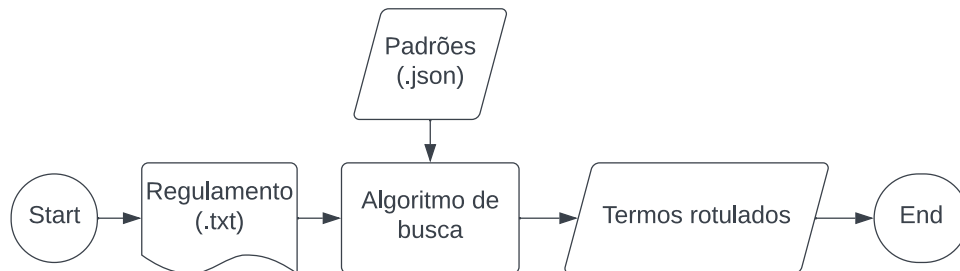
O processo de extração de informações é uma atividade do processamento de linguagem natural é executado por meio de conversão do texto em *tokens*, que são unidades básicas de uma sentença. A partir de então são aplicadas diferentes técnicas como reconhecimento de caracteres, padrões textuais, reconhecimentos de entidades e vetorização de palavras. Esse processamento serve à diferentes propósitos, mas neste trabalho ficará evidente o uso para a extração de informações relevantes do texto.

A rotulagem do texto com termos semânticos permite que sejam identificados elementos textuais dentro de algum contexto de uso. A rotulagem incorpora uma camada de informação que deve ser de entendimento compartilhado no domínio do conhecimento para que o modelo gerado seja interoperável com outros (Alexopoulos, 2020). O processo de conversão de padrões em rótulos semânticos também é chamado de

processo de anotação das sentenças (Xu; Cai, 2021). Em linhas gerais, ele pode ser separado em duas etapas, a de construção de padrões linguísticos e a etapa de enriquecimento semântico dos termos extraídos.

Diversas pesquisas apresentam o método de extração de informações (Drogemuller; Woodbury; Crawfor, 2000; Wimalasuriya; Dou, 2010; Wyner; Peters, 2011; Zhang; El-Gohary, 2013, 2014, 2015, 2016, 2017<sup>a</sup>, 2017<sup>b</sup>; Choi; Nguyen; Kerr, 2021; Zhong *et al.*, 2022, Xu; Cai, 2021 e Zhou *et al.*, 2022) se diferenciam em alguns procedimentos e conceitos terminológicos. Ao analisar as pesquisas destes autores é possível compreender o processo geral da a rotulagem e como ela tem sido abordada em diferentes pesquisas de extração de informação (Figura 26Figura 26).

**Figura 26- Fluxo do uso de padrões para enriquecimento das sentenças**



Fonte: o autor.

A primeira etapa trata da incorporação do texto em um algoritmo de busca que busca por padrões de combinação de atributos linguísticos para selecionar os trechos da sentença que correspondem a estes padrões. A segunda etapa é a rotulagem baseada em cada conjunto de padrões criado. Nas pesquisas analisadas, há diferentes fases de rotulagem. Xu e Cai (2021), por exemplo, utiliza consultas à ontologia e atributos linguísticos para identificar os trechos da frase que serão rotulados (classificados) em uma estrutura conceitual que representa o regulamento. As duas etapas serão descritas com mais detalhes a seguir.

### 2.7.2.1. Etapa 1 - Preparação do texto e análise linguística da sentença

O primeiro procedimento de PLN para extração de informações baseada em regras trata do pré-processamento do texto por meio da transformação do conteúdo em um modelo universal sem formatação. A transformação inicial do texto normalmente ocorre com a extração do texto com menor índice de perda possível, mantendo a quantidade e qualidade dos dados textuais. Podem ser removidas tabulações, espaçamentos, hifenização e demais formatações textuais para resumir o texto a um padrão comum. Como resultado, os dados textuais são mantidos no formato \*.TXT ou organizados de forma tabular em formato \*.CSV.

Após o pré-processamento segue a *tokenização*. A *tokenização* se refere à segmentação das sentenças em *tokens* que são os elementos constituintes da sentença, i.e., palavras, símbolos, valores ou pontuações. A separação de todos os elementos textuais é feita em formato estruturado de listas. As listas são uma estrutura de dados que reúne elementos armazenados em ordem específica e indexados para ser futuramente analisados. Uma vez que as sentenças estão segmentadas, é possível utilizar o método de Expressões Regulares (RegEx), para identificar padrões nos caracteres visando a localização, eliminação e caracterização dos *tokens*. Um exemplo de aplicação de expressões regulares é a identificação de datas e endereços eletrônicos (\*\*/\*/\*\*\*\*\* e http://\*\*\*.\*\*\*) que possuem composição discrepante de demais elementos textuais.

Na *tokenização* podem ser descartados elementos de pontuação ou outros padrões textuais que não possuem relevância para a análise. Ainda é possível preparar o texto removendo palavras desnecessárias como conectivos, artigos, preposições e conjunções. Estes elementos são chamados de *stop words*<sup>15</sup> e em alguns casos não possuem função semântica nos termos encontrados no texto. Ao contrário da língua inglesa que adota com frequência a remoção de *stop words*, a língua portuguesa possui diferenças sutis, mas

---

<sup>15</sup> Apesar de algumas pesquisas trazerem o termo “palavras de parada”, nesta pesquisa adotamos os termos em inglês *stop words* porque é o termo mais usual e reconhecido nas pesquisas envolvendo PLN.

fundamentais nesta etapa como artigos e preposições (a, o, que, de, do, etc.). Por exemplo, o conceito “vida útil de projeto” passaria a ser representado na frase como “vida útil projeto” e pode sugerir a existência de dois termos “vida útil” e “projeto”. No caso da língua inglesa esse efeito é inexistente porque o termo “*project lifespan*” é um termo independente de outras palavras.

Uma vez que o texto foi preparado, é iniciada a análise linguística a fim de revelar padrões gramaticais, morfológicos e sintáticos da sentença. O resultado da análise gramatical é revelar as funções gramaticais de cada *token*. Já a análise sintática pretende expor as estruturas sintáticas da sentença, i.e., as hierarquias e regras sintáticas definidas pela ordenação dos termos na sentença.

Para iniciar o processo da análise gramatical são elaboradas a) a análise gramatical que identifica as funções gramaticais (ex.: substantivo, verbo, adjetivo) de cada palavra na frase, e b) classificação dos termos de acordo com sua função sintática (ex.: advérbio, locução adjetiva, raiz, sujeito, objeto direto) na frase. O método utilizado para iniciar a análise gramatical é a rotulagem *part-of-speech* (POS), que classifica cada componente da frase de acordo com sua função gramatical como substantivo, verbo e adjetivo. A ação da rotulagem é diferente em cada ferramenta, mas o que elas têm em comum no estado atual da arte é que são baseadas em modelos de linguagem ou corpus previamente rotulados e que servem de bases de dados para treinamento de modelos de aprendizado de máquina incorporados nas ferramentas. Portanto, a rotulagem é um processo estatístico probabilístico e não determinístico.

Com base na rotulagem dos *tokens*, algumas pesquisas utilizam análises sintáticas (*parsing*) complementares com o método de Análise de Constituintes (Figura 27).



**Figura 28 - Diferença entre o resultado de um *shallow parsing* e um *deep parsing***

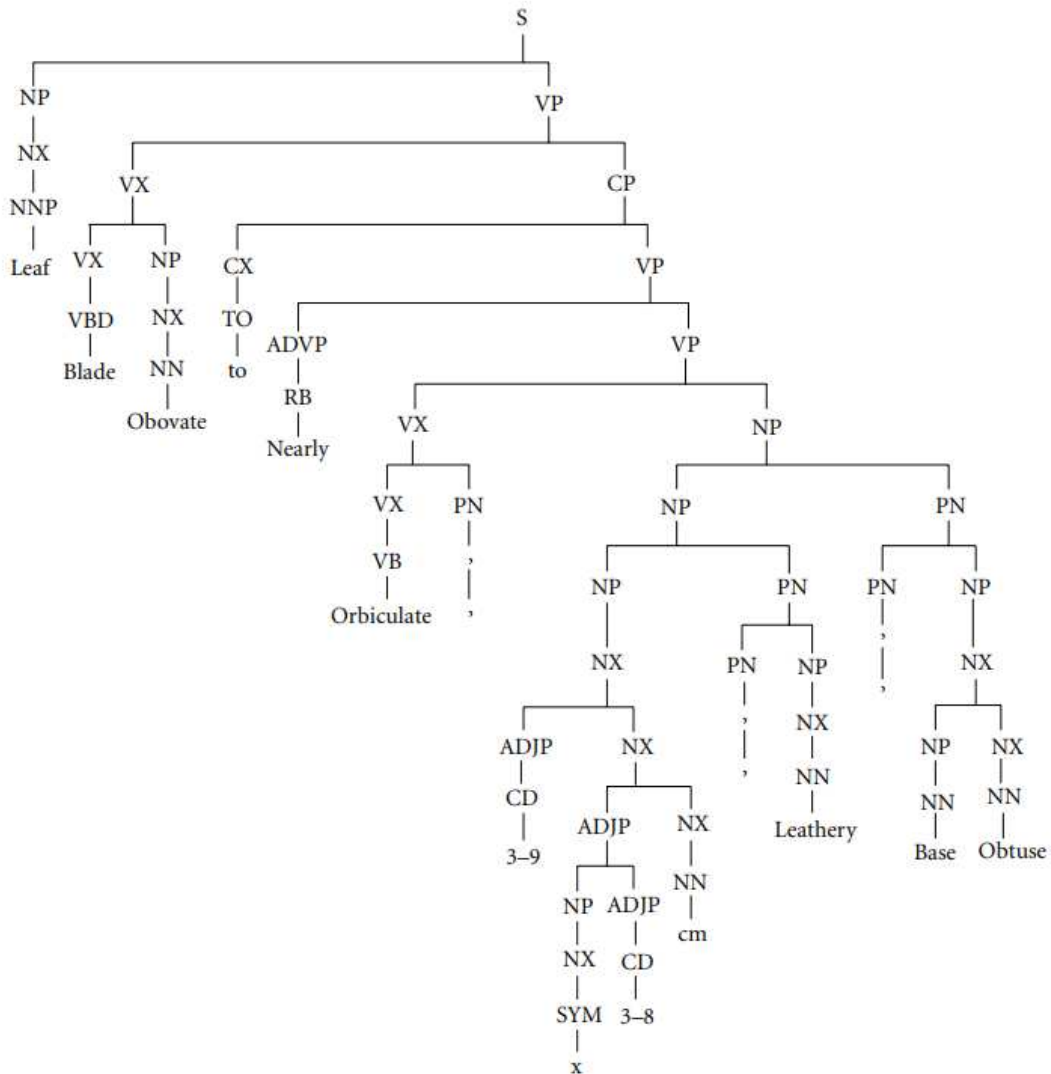
Original text:

leaf blade obovate to nearly orbiculate, 3-9 × 3-8 cm, leathery, base obtuse.

Shallow parsing result :

[NP leaf blade] [VP obovate] [VP to nearly orbiculate], [NP 3-9 × 3-8 cm], [NP leathery], [NP base obtuse].

Deep parsing result:



Fonte: Thessen, Cui e Mozzherin (2012).

A característica deste processo é a combinação de termos da sentença para formar os tipos de sentenças (*sintagmas*) que, articuladas de forma correta, representam frases com mesmas categorias de conteúdo. Por exemplo, na expressão:

QP → JJR IN CD

Onde:

QP – Frase quantificadora

JJR – Adjetivo comparativo

IN – Preposição

CD – Numeral, cardinal

A expressão QP → JJR IN CD deve ser lida como: “quando existir um JJR com IN com CD, existirá um CD”. Esse padrão textual é composto por rótulos gramaticais (preposição, numeral, adjetivo), mas também pelo constituinte “frase quantificadora” (QP) que é um constituinte da sentença (um dos nós das árvores de constituintes). Seguindo esta regra, todas as sentenças que possuírem correspondência com este padrão gramatical, deverão ser classificadas como “frase quantificadora”.

Esse processo se baseia na teoria da composicionalidade da linguagem iniciada com a gramática gerativa de Chomsky (1956). A teoria interpreta a língua como um conjunto de enunciados gramaticais cuja análise e reordenação pode produzir um número infinito de novas sentenças (Goddard; Schalley, 2010). Segundo a teoria, uma vez que existe um limitado conjunto de estruturas sintáticas que compõem um texto, é possível mapear os padrões sintáticos das sentenças.

A teoria da composicionalidade é apresentada como Gramática Livre de Contexto (*Context Free Grammar*) e Análise da Estrutura da Frase (*Phrase Structure Grammar*) em algumas das pesquisas, e trata fundamentalmente de uma definição de padrões gramaticais que representam os constituintes da frase. Padrões preestabelecidos, como o padrão  $S \rightarrow NP VP$  ou  $NP \rightarrow Det N$ , representam uma raiz ou *sintagma* (S) da árvore composta por uma Frase Nominal (NP) e uma Frase Verbal (VP) no primeiro exemplo, e uma Frase Nominal (NP) composta por determinante (Det) e um substantivo (N) no segundo exemplo.

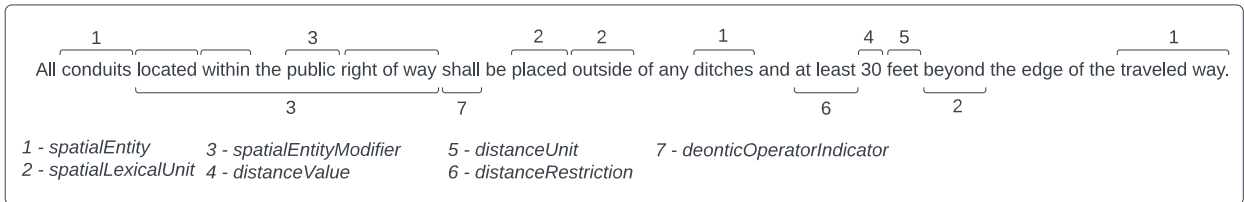
Os padrões de elementos constituintes e funções gramaticais são usados para definir os trechos da sentença que possuem importância na representação do

regulamento. Em Zhou e El-Gohary (2021), por exemplo, o uso de uma análise de constituintes pelo padrão “(JJR IN):cr + QuantityValue → cr.ComparativeRelation” classificam os termos como uma relação comparativa quando existirem termos na frase com a função gramatical adjetivo + preposição (ex.: *less than*).

Este tipo de estratégia é utilizada em outras pesquisas como Zhang e El-Gohary (2011, 2012, 2015 e 2017b), Xu e Cai (2019), Choi, Nguyen e Kerr Jr. (2021) e Zhou *et al.* (2022).

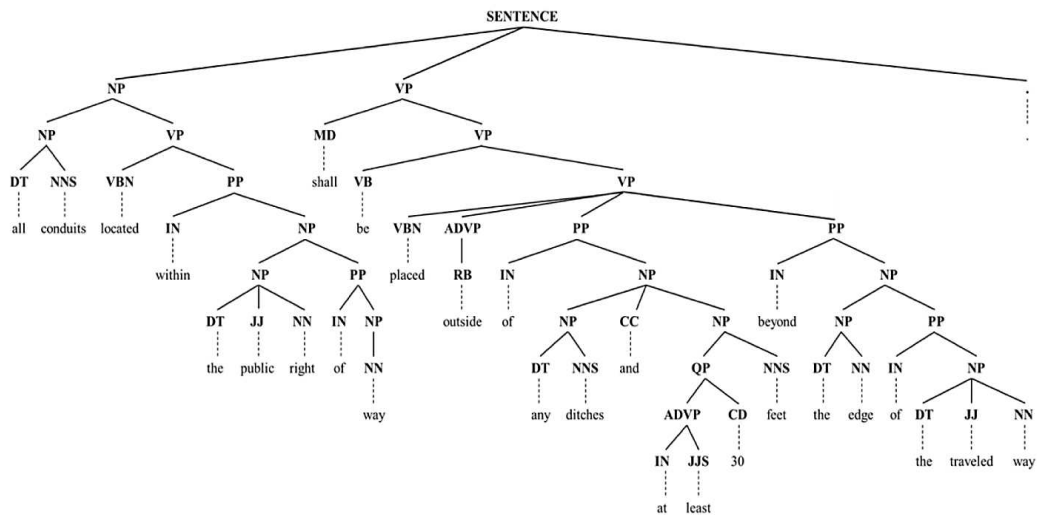
Xu e Cai (2021) deixam evidente o uso destes padrões linguísticos. Na Figura 29 foram inseridos diversos rótulos que identificam todos os termos da sentença a partir de suas composições linguísticas (Figura 30).

**Figura 29 - Representação de uma sentença anotada com sete termos semânticos**



Fonte: adaptado de Xu e Cai (2021).

**Figura 30 - Árvores de constituintes com atributos linguísticos usada para construção dos padrões**



Fonte: Xu e Cai (2021).

Na Figura 29 foram inseridos sete rótulos que incorporam uma camada de sentido a mais para a sentença. De todos os rótulos indicados na frase, apenas o “*spatialEntity*” não foi gerado pela análise de constituintes e é indicado pela etapa de análise semântica (apresentada mais adiante).

O uso de análises de constituintes (ou gramáticas livres de contexto) pode criar categorias textuais que representam grupos de requisitos dos regulamentos. Comparado com o método de análise rasa (*shallow parsing*), a análise de constituintes reduz a quantidade de padrões que seriam mapeados nos regulamentos (Xu; Cai, 2021).

O problema deste tipo de anotação é que o padrão que define os *sintagmas* é dado *a priori* e não existe um modelo universal que classifique uma constituinte como uma frase verbal ou nominal. Esta definição deve ser feita pelo usuário que propõe combinações entre “folhas” para representar as categorias de elementos constituintes. No caso da língua inglesa, existe um corpus que define estas estruturas<sup>16</sup>, mas esta pesquisa não identificou material similar para língua portuguesa.

Uma vez que os rótulos linguísticos estão identificados é possível ampliar a identificação dos elementos textuais por meio da incorporação de uma base externa de dados, de caráter semântico.

Ao final desta etapa, os padrões de composição frasal podem ser extraídos e incorporados em uma análise semântica que vincula os termos destacados a uma base de dados semânticos (i.e., ontologia) para referenciá-los aos seus conceitos (Zhang; El-Gohary, 2011, 2012, 2015, 2017b, Choi; Nguyen; Kerr Jr., 2021). Na análise semântica o resultado é a troca dos termos explícitos do texto pelos seus conceitos abstratos dados de uma ontologia. O correto mapeamento desses conceitos e das estruturas sintáticas garante maior fidelidade (ou precisão) do padrão de estrutura frasal a ser buscado pelo algoritmo de extração de informações.

---

<sup>16</sup> <https://catalog.ldc.upenn.edu/docs/LDC95T7/cl93.html>

### 2.7.2.2. Etapa 2 - Análise semântica

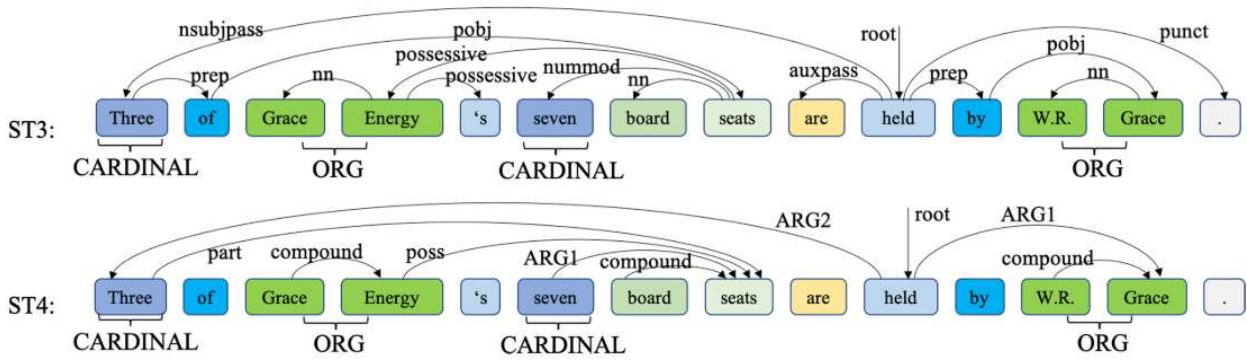
Além de explorar a estrutura profunda da sentença, as pesquisas que se baseiam em ontologias para extração de informação (*Ontology-Based Information Extraction, OBIE*), utilizam o mapeamento ontológico para constituição de padrões frasais. A incorporação de uma ontologia é a principal inovação trazida para melhoria do desempenho nos trabalhos de Zhang e El-Gohary (2016, 2017a) e ecoado em outras pesquisas como em Zhong *et al.* (2022) e Xiu e Cai (2021).

As pesquisas que se relacionam com uma OBIE desenvolvem uma representação abstrata de tuplas de termos semânticos que, ao final do processo, serão convertidos em linguagem lógica nas análises de conformidade.

Zhang (2015) propõe que seja usada a estrutura < Substantivo, Atributo e Valor> (*Subject, Attribute, Value*) como estrutura de tupla para sintetizar a representação da frase a ser extraída. Cada um destes três componentes da tupla foram chamados de “elemento de informação semântica” e serão vinculados aos elementos do tipo *conceito* e *relação* já existentes na ontologia. Outros termos como operadores deôntico (verbos modais como *dever*, *precisar* e *poder*) são em menor número, e podem ser referenciados a uma lista de termos de mesma característica (um *gazetteer*). Dessa forma, palavras como “*courts*”, “*net area*” e “*not less than*” são informações de entrada que alimentam a tupla com os termos “*court*”, “*net\_area*” e “*greater\_than\_or\_equal*”. Estes termos estão diretamente associados a conceitos e propriedades (por exemplo, *owl:DatatypeProperty*) de relações entre componentes da ontologia.

Na Figura 31 a sentença ST3 representa a rotulagem de Entidades Nomeadas (Org e Cardinal) e os arcos são representações da dependência sintática entre termos enquanto. Já na sentença ST4 ocorre a adição da representação semântica (ARG1, part, compound, poss etc.) baseada em combinações de padrões da estrutura sintática da frase.

Figura 31 - Exemplo de rotulação de entidades, relações sintáticas e semânticas



Fonte: adaptado de Wang *et al.* (2021).

As pesquisas que se dedicam ao uso de métodos OBIE desenvolvem suas próprias ontologias. Nenhuma das pesquisas apresentadas mostrou uso de ontologias abertas, capazes de serem usadas por terceiros.

Zhang (2015) cria uma ontologia para dois capítulos do *International Building Codes* (IBC). O autor utilizou conceitos da IC-PRO-Onto (El-Gohary; El-Diraby, 2010) e do IFC unificados em uma ontologia. O autor identificou os conceitos principais do domínio (*General Service Administration - GSA*) PBS-P100 - *Facilities Standards for the Public Building Services* (2015) e normas relacionadas a edificações (*International Building Codes - IBC*).

A ontologia criada para a extração de informações foi composta de 360 conceitos organizados em uma taxonomia. Os conceitos foram estabelecidos pelo autor ao analisar o regulamento, já os termos foram coletados em sentenças por meio de busca com padrões organizados na análise de constituintes (*deep parsing*). O mapeamento dos termos na ontologia é feito segundo uma regra simples de busca no banco de dados (Figura 32).

Figura 32 - Exemplo de rotina de busca executada na ontologia

```

Phase: quantitative_relation_and_deontic_operator_indicator_extraction
Input: Token Lookup MD VB VBN neg VBZ TO VBP VBD
Options: Control = appelt
Rule: quantitative_relation_extraction
// IE Rule #1 for quantitative relation:
// If "MD + VB" is matched, extract the text matched with "vb" as an instance for "quantitative relation".
(
  {{MD}} {{VB}}:QRel
):QuantitativeRelation
-->
:QuantitativeRelation
{
  gate.AnnotationSet matchedQRel=(gate.AnnotationSet) bindings.get("QRel");
  Annotation TheQuantitativeRelation=matchedQRel.iterator().next();
  gate.AnnotationSet matchedAnns= (gate.AnnotationSet)
  bindings.get("QuantitativeRelation");
  gate.FeatureMap newFeatures= Factory.newFeatureMap();
  newFeatures.put("QuantitativeRelation",TheQuantitativeRelation);
  newFeatures.put("rule","QuantitativeRelation");
  annotations.add(matchedAnns.firstNode(),matchedAnns.lastNode(),"QuantitativeRelation", newFeatures);
}

```

Fonte: adaptado de Zhang (2015).

Na figura acima, a rotina procura pelo padrão MD (verbo modal) junto do padrão VB (verbo base) e, se for encontrado, o elemento correspondente é extraído e rotulado como *QuantitativeRelation*. Outro exemplo de regra mencionado por Zhang e El-Gohary (2013) e repetido em Zhang (2015), Zhou e El-Gohary (2017) e Zhong *et al.* (2022) é a busca por um determinado termo no texto. Caso este termo esteja presente nos subconceitos (subclasses ou instâncias) da classe “*buildingelement*” da ontologia, o termo do requisito deve ser reconhecido como sujeito da sentença. A construção dos padrões é manual, iterativa e validada pelos resultados de teste da extração.

Xu e Cai (2021), de modo semelhante, executam uma regra formatada em sintaxe JAPE<sup>17</sup> (Quadro ), onde JJ, JJR, JJS são tipos de adjetivos e VP, PP e SBAR são frase verbal, proposicional e expressão subordinada. Na regra apresentada no Quadro 5, a combinação de adjetivo com uma *Spatial Entity* (conceito da ontologia) e uma frase verbal a partir de uma outra *Spatial Entity*, é anotado como *spatialEntityModifier*.

<sup>17</sup> Sintaxe usada pelo programa GATE.

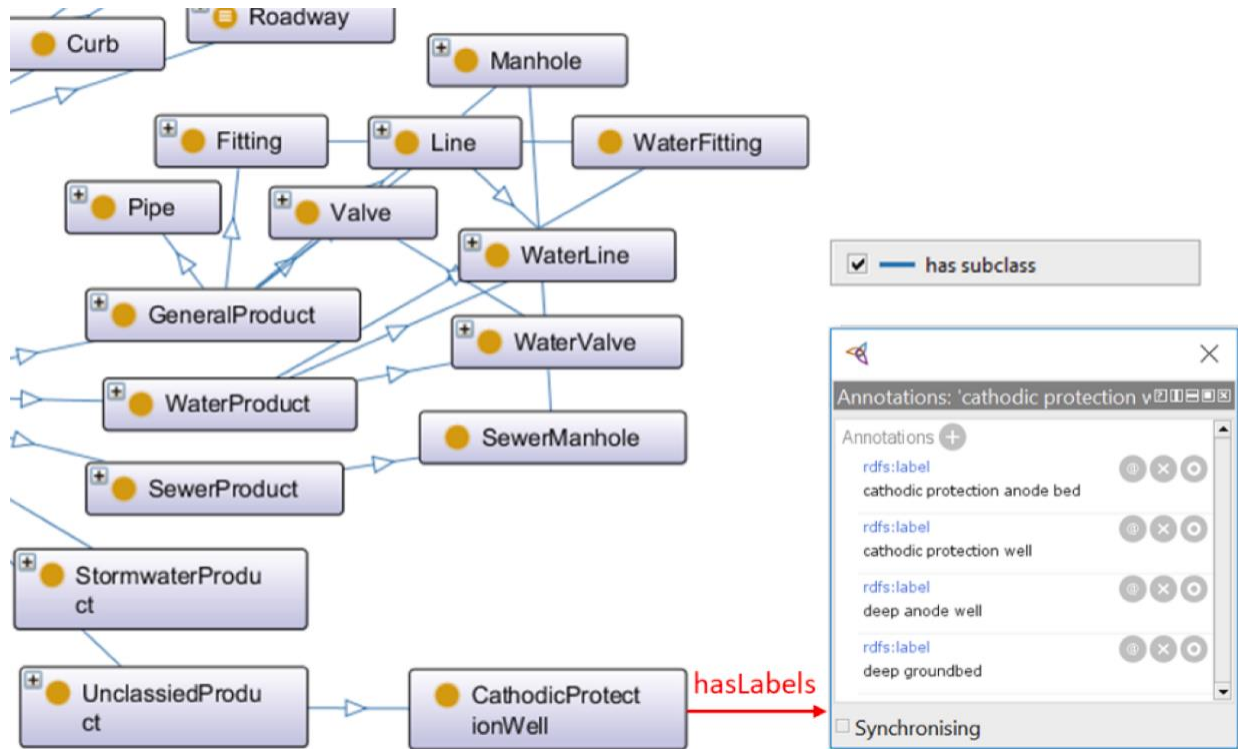
### Quadro 5 - Exemplo de anotação baseado em padrão gramatical e ontologia

```
((({JJ}|{JJR}|{JJS}):sem)*{spatialEntity}({{VP}|{PP}|{SBAR}}): sem?)
→:sem.spatialEntityModifier
```

Fonte: adaptado de Xu e Cai (2021).

Os algoritmos de mapeamento do texto na base de dados semânticos (ontologia) não são apresentados nas pesquisas analisadas. Este processo está implícito na ferramenta usada pelos autores (Plataforma GATE, apresentada no próximo capítulo), mas pôde ser extraído da interpretação do trabalho de Xu e Cai (2021). Os autores desenvolveram duas ontologias (chamadas de UPO e SO) para melhorar o resultado da extração de informações.

Figura 33 - Trecho da ontologia UPO



Fonte: adaptado de Xu e Cai (2021).

Em Xu e Cai (2021), a busca por termos na ontologia identificava se os termos encontrados no texto do requisito possuíam correspondência com os rótulos dos elementos da ontologia. Na Figura 33 é possível verificar que os autores adotaram a estratégia incomum de instanciar objetos como rótulos de cada elemento da ontologia. Na

Figura 33 o elemento *CatholicProtectionWell* possui a propriedade “*hasLables*” que é povoada com uma lista de componentes construtivos da mesma família. Quando os autores efetuam uma busca no texto do requisito e encontram o termo, por exemplo, o termo “*deep anode well*” pertencente aos rótulos do elemento *CatholicProtectionWell*, esse termo é identificado como pertencente à ontologia e recebe a anotação *spatialEntity* na sentença (ver Figura 29).

De modo similar a Xu e Cai (2021), o Quadro 6 mostra os diferentes produtos obtidos em todo processo de extração de informações de Zhang (2015) a partir da sentença “*A minimum of 1 inch of airspace shall be provided between the insulation and the roof sheathing*”. No Quadro 6 estão representados os conceitos que compõem das tuplas (*Subject, Subject restriction, Compliance checking attribute etc.*) e suas correspondências obtidas em um dos requisitos do regulamento (*airspace, relation, NA, obligation etc.*).

**Quadro 6 - Exemplo de elementos de informação semântica extraídos e convertidos para representação lógica**

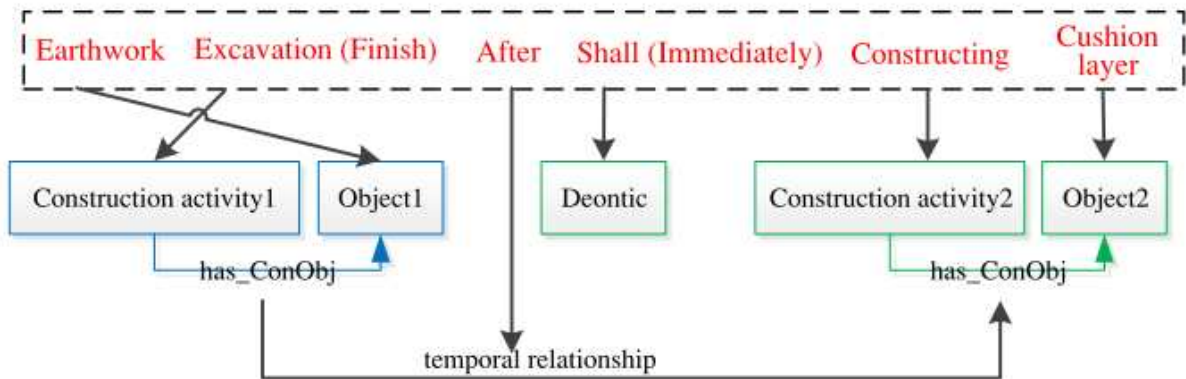
Information tuple extracted from text sentences	Subject	airspace
	Subject restriction	relation (between, insulation, roof_sheathing)
	Compliance checking attribute	NA
	Deontic operator indicator	obligation
	Quantitative relation	provide
	Comparative relation	greater_than_or_equal
	Quantity value	1
	Quantity unit/reference	inch
	Quantity restriction	NA
Horn clause logic representation		$\forall (a, i, r, s) ((\text{airspace}(a) \wedge \text{insulation}(i) \wedge \text{roof\_sheathing}(r) \wedge \text{between}(a, i, r) \wedge \text{has}(a, s)) \supset O(\text{greater\_than\_or\_equal}(s, \text{quantity}(1, \text{inch})))$

Fonte: adaptado de Zhang (2015).

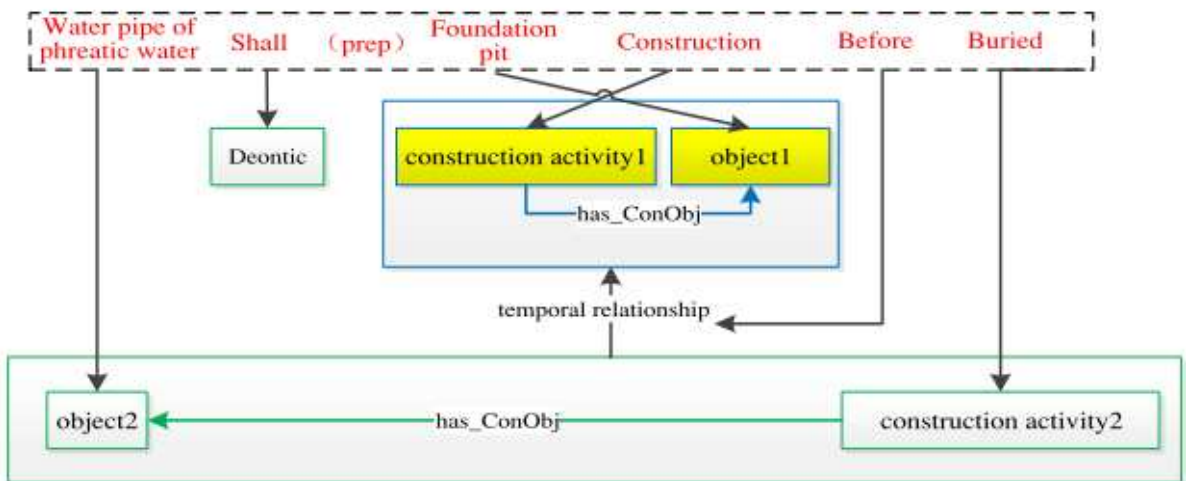
Ao final, o autor desenvolve manualmente uma representação lógica estruturada para cada categoria de sentença identificada no texto. A representação dos dados em lógica de primeira ordem (LPO) foi feita manualmente pelo autor.

Outro exemplo de extração de informações baseado em padrões é apresentado em Zhong *et al.* (2022) (Figura 34). As sentenças apresentadas na Figura 34 correspondem às diretrizes: 1) “After finishing the earthwork excavation, the cushion layer construction shall be conducted immediately” e 2) “The water pipe of phreatic water shall be buried before the foundation pit construction”.

Figura 34 - Exemplo de sentença do regulamento convertida para classes de variáveis da regra



Regulation text: After finishing the earthwork excavation, the cushion layer construction shall be conducted immediately



Regulation text: The water pipe of phreatic water shall be buried before the foundation pit construction.

Fonte: Zhong *et al.* (2022).

Para representação de variáveis da qualidade do produto, Zhong *et al.* (2022) atingem três tipos de variáveis (Quadro 7). As variáveis numéricas medem o objeto a partir de um ou mais valores, por exemplo, “*the protective layer thickness of the main reinforcement of the precast underground dia- phragm wall shall not be less than 30 mm*”. As variáveis de atributos determinam uma relação entre parâmetros de elementos da construção, por exemplo, “*the width of the crown beam should not be less than the thickness of the under- ground continuous wall*”. Já as variáveis de texto são dependências mencionam outras regras, índices ou códigos como, por exemplo, “*the concrete strength grade of the pile body must not be less than C25*”.

**Quadro 7 - Composição das variáveis de qualidade de produto**

Variável de produto	Padrão
numéricas	<object><attribute><deontic><comparison><value>
atributos	<object1><attribute1><deontic><comparison> <object2><attribute2>
dependência	<construction component><deontic><adopt/select> <specific resource>

Fonte: adaptado de Zhong *et al.* (2022).

Além dos grupos de variáveis de processo e produto, a categoria de variáveis de recurso é similar ao tipo “variável de atributo” da categoria Variáveis de produto, com pouca diferença entre elas. Enquanto uma Variável de Produto (atributo) é dada pela relação entre dois ou mais parâmetros dos objetos como largura e altura, uma Variável de Recurso mostra dependência entre composições de elementos construtivos (Quadro 8). Um exemplo é a sentença “*the anchorage body should adopt cement paste or mortar*”, que pode ser expressa pelo padrão.

**Quadro 8 - Exemplo de variável de recurso**

Variável	Padrão
Recurso	<construction component><deontic><adopt/select> <specific resource>

Fonte: adaptado de Zhong *et al.* (2022).

Zhong *et al.* (2022) identificaram que o padrão acima sempre possui os verbos *adopt/select* nas sentenças que declaram essas variáveis.

Finalmente, o processo baseado em regras se fundamenta em uma construção do menor número possível de padrões textuais capazes de representar os textos normativos. Zhang e El-Gohary (2016), por exemplo, precisaram usar 46 padrões para representar todo documento, mas aliados à análise de constituintes e a uma análise profunda da estrutura frasal conseguiram otimizar o número de padrões encontrados em apenas 22 ao incorporar o método de análise de constituintes com a fase de análise semântica.

A estrutura de tupla <objeto, relação, valor> também é base da estrutura de uma ontologia <sujeito, relação, objeto> e facilita o mapeamento de conceitos na ontologia em algumas sentenças simples. Esta representação rígida, eficaz e de fácil captura em sentenças quantitativas, captura apenas regras de baixa complexidade e interpretabilidade. As regras que expressam condições e relações existenciais são complexas de manter a relação entre os elementos, como “*between the insulation and the roof sheathing*” (Zhang, 2015).

Os rótulos semânticos usados para classificar os elementos da sentença variam em cada abordagem. A maioria das pesquisas relacionadas à construção civil acompanha os contextos de criação de suas ontologias, normalmente orientadas para representar componentes e processos da construção (Quadro 9).

Apesar de serem formulados poucos padrões, ao considerarmos a capacidade computacional de busca e seleção deste número de padrões e escalarmos a análise para trechos maiores seria possível ampliar o número de padrões. O trabalho de Zhang e El-Gohary (2016), por exemplo, não analisa todo o regulamento, mas apenas um trecho. Zhong *et al.* (2022), por outro lado, desenvolve uma análise de cinco regulamentos chineses relativos à inspeção de qualidade da construção: *Technical specification for retaining and protection of building foundation excavations* (1999), *Code for acceptance of construction quality of building foundation* (2002), *Technical building code for monitoring of building excavation engineering* (2009), *Pile foundation engineering manual* (2012) e *Technical specification for retaining and protection of building foundation excavations* (2012).

**Quadro 9 - Rótulos semânticos adotados em pesquisas relacionadas com o enriquecimento semântico de sentenças**

Autores	Rótulos
Zhang; El-Gohary (vários)	Subject Subject Restriction Compliance Checking Attribute Deontic Operator Indicator Quantity Relation Comparative Relation Quantity Value Quantity Unit/Reference Quantity Restriction
Xu; Cai (2021)	<i>Superclasses</i> Trajector Trajector attribute Spatial indicator Landmark Landmark attribute Deontic operator indicator Negation indicator  <i>Subclasses</i> Negation indicator Spatial entity Spatial entity modifier Spatial lexical unit Distance restriction Distance value Distance unit Deontic operator indicator
Zhong <i>et al.</i> (2022)	Construction componente Attribute Deontic Compare Adopt/Select Material type # Value #
Zhou <i>et al.</i> (2022)	Prop (objeto da ontologia) Obj (elemento filho do objeto prop) Sobj (elemento filho de Obj) Cmp (relação comparativa) Rprop (exigência) ARprop (exigência condicional) Robj (elemento filho de Rprop e ARprop)

Fonte: o autor.

Ao analisarmos trechos da última norma, a *Technical specification for retaining and protection of building foundation excavations* (China, 2012), é possível reconhecer uma estrutura clara e com diretrizes organizadas em tópicos e sentenças com frases curtas,

referentes à processos de construção e discriminação de valores verificáveis. Algumas frases são complexas, compostas por mais de duas cláusulas como no exemplo abaixo.

The material of dynamic replacement pier should be adopted with with coarse particle materials with hard texture and stable performance such as well-graded block stone, gravel and slag, and the content of particle with the size greater than 300mm should not be greater than 30% of the full weight (China, 2012, p. 8).

Zhong *et al.* (2022) classificam as do sentenças do MCPRC como complexas porque são compostas por correferências e omissões de termos que precisam ser considerados na análise para uma correta extração de informações. A sentença a seguir é um exemplo de termos omissos em que a “espessura de camada de proteção do concreto” está subtendida, e portando, omissa, nas cláusulas seguintes.

[...] the thickness of the concrete protective layer of the main pile shall not be less than 50 mm; the precast pile shall not be less than 45 mm, and the prestressed tubular pile shall not be less than 35 mm; the grouting pile in corrosive environment shall not be less than 55 mm (China, 2012).

Para conseguir compreender o contexto da extração dos itens como “*precast pile*” e “*precast tubular pile*” é utilizada a ontologia que já possui um mapa das relações dos conceitos carregando consigo o relacionamento entre estes termos ao termo “*concrete protective layer*”. Ao explicitar essa relação entre conceitos, a ontologia permite orientar a conexão entre termos dentro da frase.

### 2.7.2.3. Avaliação do método baseado regras

De modo geral, a precisão dos modelos baseados em regras com suporte de uma ontologia se mostra eficaz e promissora para a extração das entidades, atingindo valores de precisão acima de 94% (Zhang; Ma; Broyd, 2023). Entretanto, pesquisadores da área avaliam que a criação de padrões de informação semântica é uma fase trabalhosa e limitada ao conhecimento do domínio dos especialistas que serão dedicados a esta tarefa. Outro limite atingido foi definir apenas três grupos de critérios de regulamentos para serem convertidos em padrões. Segundo os autores, regulamentos são documentos

complexos e possuem sentenças não mapeadas pelos grupos de padrões propostos. Além disso, a criação de uma ontologia para o domínio de aplicação foi trabalhosa e consumiu muito tempo.

No caso nacional, sabendo que a redação da norma é dada por um comitê encarregado pela ABNT e que não existe protocolo de redação dos textos, é colocado um limite ao método baseado em regras. O espaço de soluções possíveis para este caso é elevado porque deverá considerar não somente padrões para outras normas, mas estilos distintos de redação. Por outro lado, na análise empírica dos textos normativos, o cenário de soluções é menor porque os textos de regulamentos obedecem a uma escrita técnica sucinta e com sentenças similares em sua estrutura.

Xu e Cai (2021) identificam que este método possui limites em regulamentos de outros domínios porque os termos da tupla são restritos aos requisitos obtidos no documento analisado pelos autores (dois capítulos do IBC). Ainda que o método consiga mapear elementos de um requisito com estruturas hierárquicas complexas, as implicações lógicas derivadas do texto não seriam bem capturadas (Xu; Cai, 2021).

A inexistência de uma ontologia aberta, rica de classes e propriedades e em língua portuguesa é o primeiro limite que se impõe sobre a adoção de uma ontologia para extração de informações automatizado de regulamentos nacionais. Porém, é possível que as ontologias abertas do domínio da construção civil sejam capazes de ser incorporadas no método de extração de textos.

### **2.7.3. Ontologias abertas relacionadas com a construção civil**

A construção sistemática de uma ontologia é uma atividade complexa que demanda muito tempo para seu desenvolvimento e validação. Diferentes autores se utilizam de alternativas para evitar a construção completa de uma nova ontologia dada a complexidade do trabalho. Zhong *et al.* (2022), por exemplo, criam uma ontologia a partir de uma taxonomia já existente do regulamento nacional chinês GB 50202-2002.

Até o momento não foi identificada uma ontologia brasileira que pudesse ser utilizada para desenvolver o mapeamento semântico dos regulamentos relacionados à construção civil. A única participação brasileira identificada foi na criação de uma ontologia para construção civil com a participação do professor Celson Pantoja Lima (Lima *et al.*, 2003). Em trabalho desenvolvido com outros autores como El-Diraby, desenvolveram a ontologia e-COGNOS em um consórcio entre três universidades europeias. Porém, a ontologia é baseada nos termos do *British Standard Glossary of Buildings and Civil Engineering*, portanto, de língua inglesa.

Outro exemplo de construção de ontologia no Brasil foi o trabalho realizado por Li, Raskin e Ramani (2008) que desenvolveram uma ontologia para sistemas de busca em engenharia se baseando nas orientações de Nirenburg e Raskin (2004). Foram criadas as taxonomias e estabelecidas interrelações entre conceitos e taxonomias que resultou em uma ontologia com conceitos, relações e predicados. Os autores desenvolveram as relações em planilhas e posteriormente importaram no *software* Protégé para construção das relações e geração da base de dados semânticos. Foi possível construir 2.629 conceitos, 13 tipos de relações e 10 taxonomias. Ao todo os autores calcularam o uso de 500 h de trabalho dedicado pelos pesquisadores e alunos de graduação.

Apesar da ausência de conteúdo nacional, existe uma iniciativa internacional para reunião e alinhamento de ontologias relacionadas à construção civil. O *Linked Building Data*<sup>18</sup> é uma iniciativa vinculada ao *World Wide Web Consortium* (W3C) que mantém a ontologia *Digital Construction Ontologies* (DiCon)<sup>19</sup>. Essa iniciativa é uma plataforma online, colaborativa, aberta que permite a interoperabilidade entre sistemas no domínio da construção e renovação de edifícios. A DiCon continuamente é alinhada a diversas ontologias como a DICM e ifcOWL (Figura 35 e Figura 36Figura ) que já haviam sido desenvolvidas e validadas por pesquisas anteriores. A DiCon está organizada em

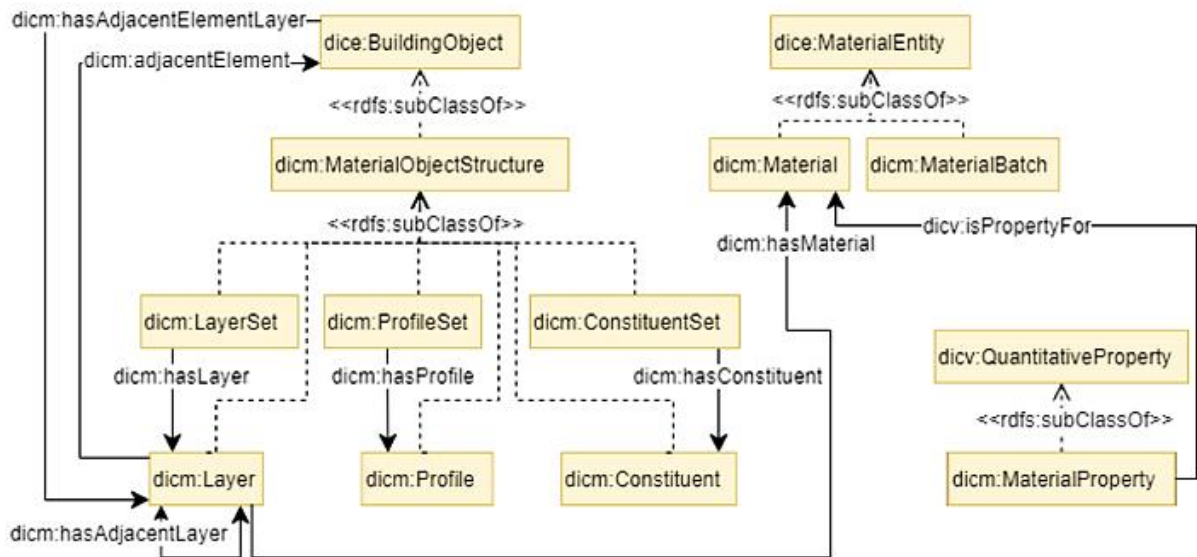
---

<sup>18</sup> <https://www.w3.org/community/lbd/>

<sup>19</sup> <https://digitalconstruction.github.io/v/0.5/>

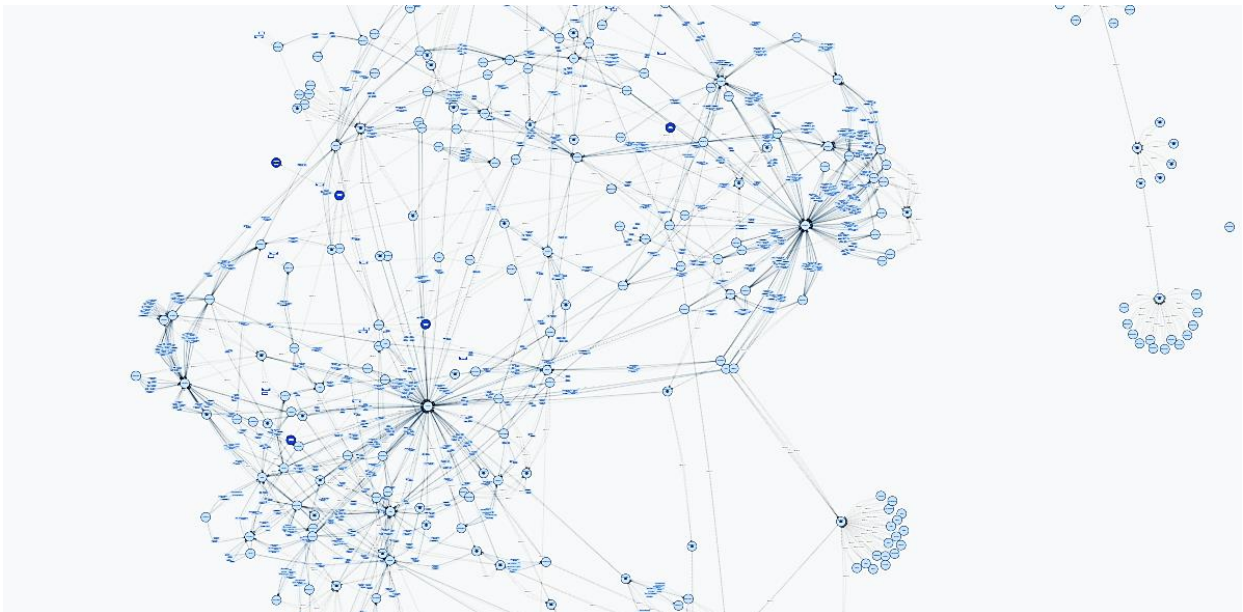
diversos módulos e o alinhamento da ontologia às demais é fruto de um esforço da comunidade *Linked Building Data* para uniformizar termos e relações entre ontologias segundo os mesmos conceitos, relações e propriedades entre ontologias.

**Figura 35 - Maiores classes da ontologia de materiais da DICM (Digital Construction Building Material Ontology) existente na DiCon**



Fonte: Karlapudi e Valluru (2021).

**Figura 36 -Trecho da ontologia ifcOWL (IFC4\_ADD2\_TC1)**



Fonte: o autor.

#### 2.7.4. Ferramentas utilizadas para extração e mapeamento semântico baseados em regras

Os procedimentos apresentados na Seção 2.7.2 seguem processos semelhantes de preparação do texto, extração de padrões linguísticos e mapeamento dos termos em uma ontologia. Este processo é influenciado pela ferramenta que adotam. Por outro lado, é possível serem utilizadas outras técnicas, como o uso de linguagem de programação neutra, desvinculadas de uma ferramenta.

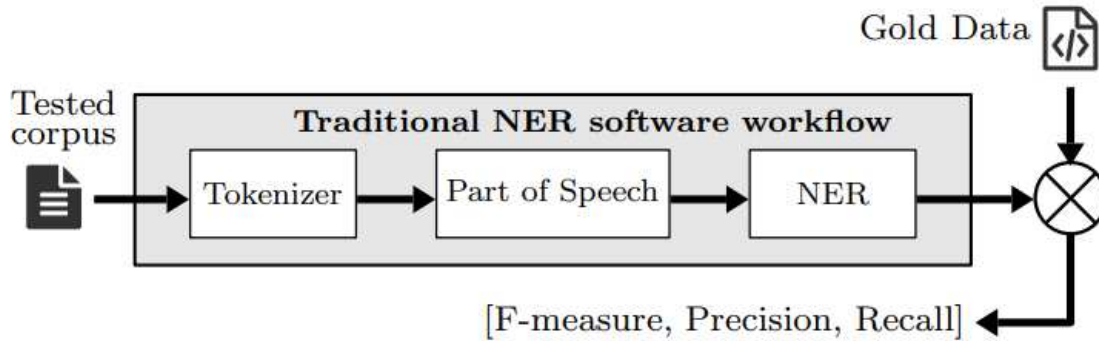
Apesar da linguagem de programação com a biblioteca (NLTK), dedicada ao processamento de linguagem, ter sido mencionada em diversas pesquisas, os trabalhos relacionados à extração de informações de textos regulatórios se restringem ao uso de processamento em bases de dados (linguagem SQL e *software* Protegé) e ao *software* *General Architecture for Text Engineering* (GATE). A maioria das pesquisas relacionadas à OBIE utiliza o *software* GATE e suas ferramentas<sup>20</sup> para o processamento e mapeamento semântico (Zhang; El-Gohary, 2016, 2017; Zhang, 2015, Zhou; El-Gohary, 2017; Xu; Cai, 2021; Choi; Nguyen; Kerr Jr., 2021).

O cenário de aplicações para PLN não é amplo e existem ferramentas dominantes para o processamento por código e capazes de ser incorporado em futuras aplicações. As mais difundidas são as bibliotecas de PLN spaCy, NLTK acessadas por meio da linguagem Python. Ainda é mencionado o acesso à plataforma GATE por meio de JavaScript; a ferramenta openNLP baseada em linguagem Java, e StanfordNLP (Stanza). Todos esses recursos foram avaliados por Schmitt *et al.* (2019) pelo seu desempenho em atividades de Reconhecimento de Entidades Nomeadas (*Named Entity Recognition* - NER). O NER é acionado em algumas aplicações envolvendo o processo de Extração de informações (Figura 37).

---

<sup>20</sup> <https://gate.ac.uk/>

Figura 37 - Incorporação do NER em extrações de informação



Fonte: Schmitt *et al.* (2019).

As ferramentas avaliadas com suas respectivas versões estão relacionadas no Quadro 10.

Quadro 10 - Comparativo entre ferramentas de PLN

Name	Prog.	Vers.	Algo	Classifiers
Stanford NLP	Java	3.9.2	CRF	CoNLL, MUC6-7, ACE
OpenNLP	Java	1.9	Max entropy	OntoNotes
SpaCy	Python	2.0.16	Neural (2.0)	Stanford
NLTK	Python	3.4	Max entropy	NER
GATE	Java	8.5.1	JAPE	

Fonte: adaptado de Schmitt *et al.* (2019).

Foi constatado que a ferramenta StanfordNLP (atual Stanza) teve desempenho médio muito superior aos demais (75,97%, desvio padrão ( $\sigma$ ) = 5.08) quando treinado por ambos os *corpora* do estudo (CoNLL 2003 e GMB). Pela média dos resultados os modelos seguintes tiveram desempenho próximo, SpaCy(57%), NLTK (55.4%), Gate (53,38) e OpenNLP(40,8%). O programa GATE manteve menor desvio padrão (0.16), seguido por OpenNLP (1.91), SpaCy (2,67), StanfordNLP (5,08) e NLTK (6.96%).

O estudo de Schmitt *et al.* (2019) avalia as ferramentas em diversos *corpora* em inglês, porém das ferramentas analisadas, apenas o NLTK, SpaCy e OpenNLP possuem treinamento em *corpora* de português brasileiro. O StanfordNLP é treinado com *corpora* europeu. O *software* GATE declara que é possível incorporar um corpus na língua

portuguesa, mas não tem uma solução implementada previamente no modelo. Da mesma forma ocorre com o StanfordNLP e OpenNLP.

A última etapa da etapa de extração de informações é a interpretação semântica que trata de analisar os resultados da extração de informações a fim de elaborar uma metalinguagem semântica (Goddard, Schalley, 2010). Existem diferentes abordagens na representação semântica e o que tem sido percebida nos artigos relacionados é a adoção de uma linguagem de representação lógica baseada na Lógica de Primeira Ordem (LPO). Com a tradução dos padrões em lógica simbólica ou em uma sintaxe com marcações como arquivos XML é possível converter as informações extraídas para um formato computável.

### **2.7.5. Procedimentos de extração de informações baseados em aprendizado de máquina**

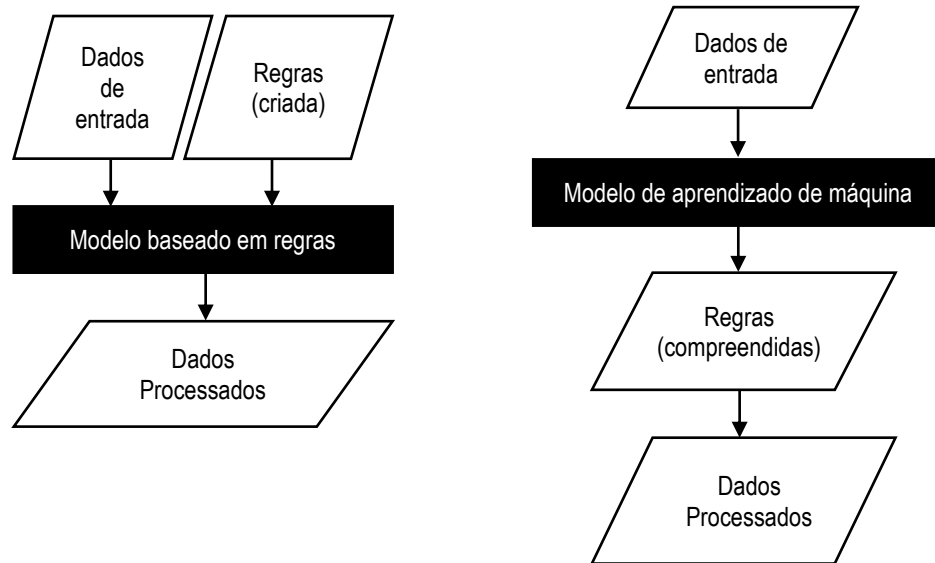
Alguns dos motivos considerados para adoção de sistemas baseados em regras são: 1) é necessário um corpus menor para reconhecimento de padrões ou rotulagem linguística; 2) é mais simples para humanos de determinado domínio gerar regras somente pela observação de um conjunto restrito de dados (Thanaki, 2017).

Modelos baseados em aprendizado de máquina dependem de grande quantidade de dados de treinamento e com dados relacionados ao problema em questão para conseguirem alcançar bons resultados. Por outro lado, modelos baseados em regras são altamente dependentes de especialistas capazes de criar as regras de funcionamento dos modelos e avaliar se o resultado está adequado ao problema que se pretende resolver.

Já um processo baseado em aprendizagem de máquina possui menor dependência do especialista do domínio porque as regras que o modelo usa para processar o texto são descobertas pelo próprio modelo por meio do aprendizado estatístico dos dados. Uma grande quantidade de dados revela padrões implícitos nos textos que são incorporados pelos modelos para 'compreender' o domínio de sua ação. Uma vez treinado neste banco

de dados, o modelo é capaz de executar suas tarefas baseando-se na experiência que adquiriu (Figura 38).

**Figura 38 - Fluxo de trabalho nos dois paradigmas computacionais**



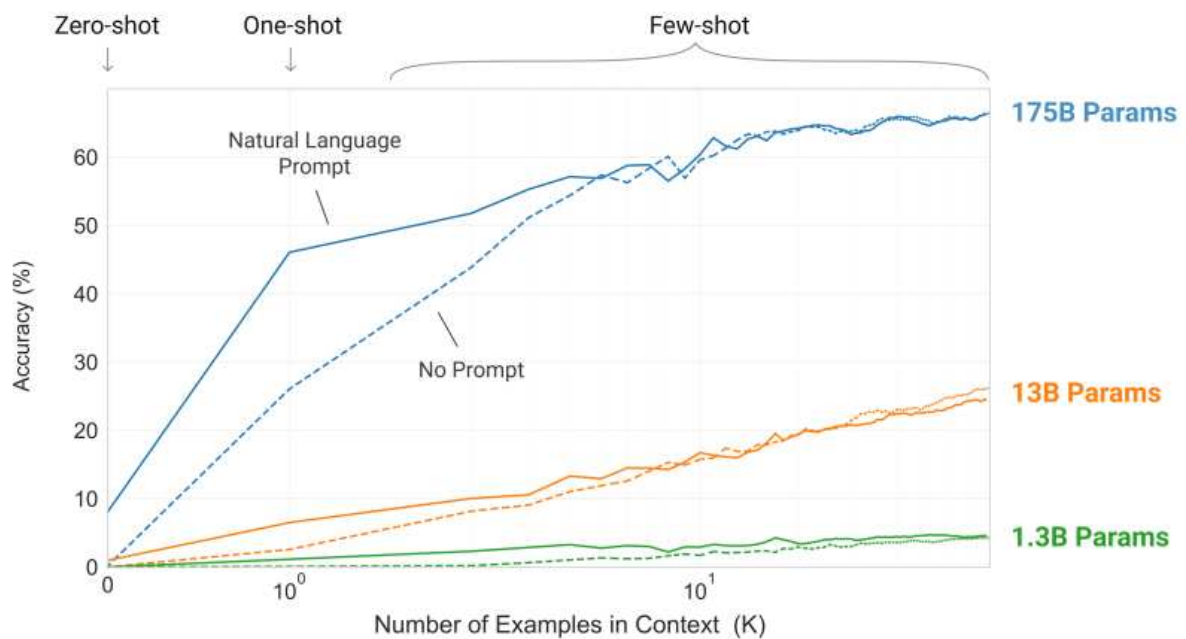
Fonte: o autor.

Os exemplos apresentados até o momento nesta pesquisa abordam aplicações baseadas em regras, que podem recorrer a consultas a bases de conhecimento, como ontologias e dicionários, para executar suas operações. Por outro lado, o aprendizado de máquina é um subcampo da Inteligência Artificial que opera através da criação de modelos que representam dados do mundo real. Como resultado, tais modelos podem classificar ou prever resultados com base nos padrões identificados nos dados.

Em pesquisas de verificação e extração de informações na área da construção civil, o processamento de linguagem natural é empregado juntamente com técnicas de aprendizado de máquina para diversas finalidades, como análise de similaridade entre termos do regulamento e do modelo BIM (Zhou; El-Gohary, 2021), extração de termos semânticos de ontologias (Ren *et al.*, 2022) e geração de texto rotulado (Zhang; El-Gohary, 2022).

A partir de 2023 a IA tem tido a atenção desviada para modelos generativos baseados em grandes modelos de linguagem (*Large Language Models - LLM*). Modelos de linguagem (*Language Models - LM*) existem há poucos anos (Dai; Le, 2015) e têm sido utilizados para diversas tarefas em PLN porque são treinados em corpus rotulados usando redes profundas. Alguns dos modelos são o BERT e RoBERTa. Já o termo “Large” é originário do artigo que anunciou o ChatGPT3 em 2020 (Brown *et al.*, 2020). O modelo RoBERTa Large, por exemplo, é um modelo com 355 milhões de parâmetros, enquanto o ChatGPT 3 Large possui 760 milhões. A partir de então, o termo *Large Language Model* se refere a estes modelos que hoje atingem 1 trilhão de parâmetros de treinamento.

**Figura 39 - Acurácia do modelo ChatGPT 3 e de outros modelos de linguagem**

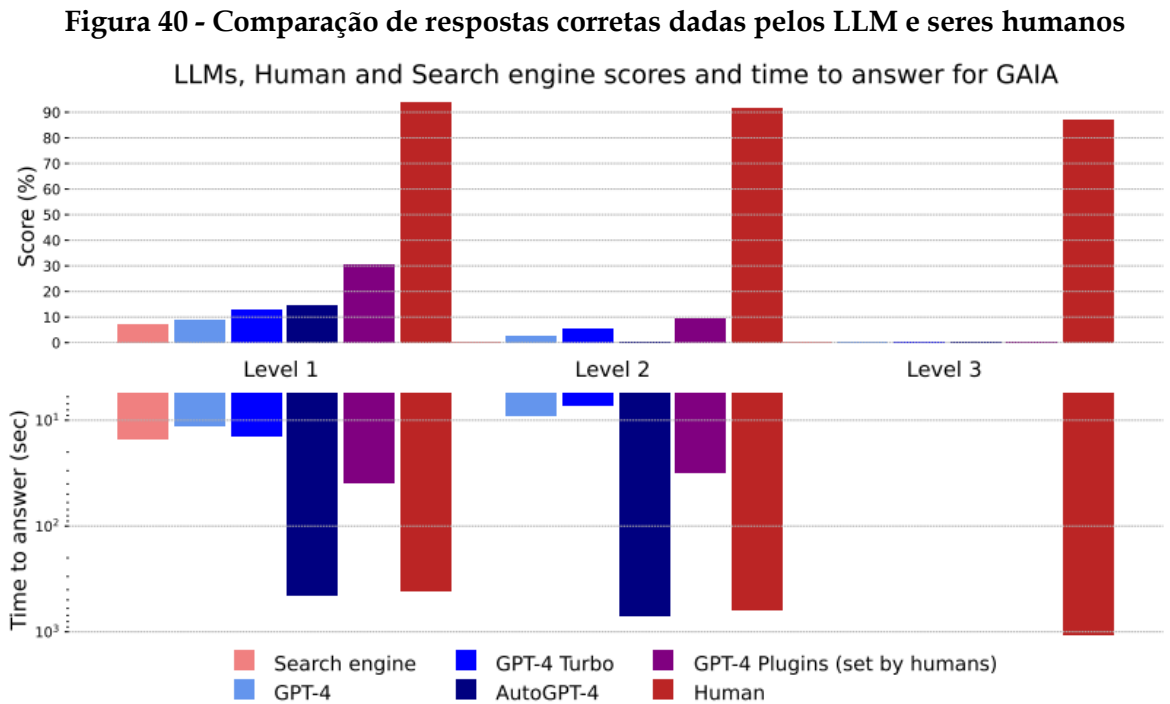


Fonte: Brown *et al.* (2020).

O gráfico da Figura 39 mostra o modelo ChatGPT comparado a outras versões do mesmo modelo com menos parâmetros de treinamento. A diferença entre os pontos de inflexão “zero-shot”, “one-shot” e “few-shot” se refere à quantidade de demonstrações dadas ao modelo para executar uma tarefa. “Zero-shot”, por exemplo, é fornecido apenas uma instrução e nenhuma demonstração de como executar é oferecida. Como pode ser

visto no eixo das ordenadas, a acurácia do modelo atingiu 60%. Hoje o modelo aberto disponível é a versão 3.5, mas já está disponível a versão paga ChatGPT 4 que possui melhor capacidade em diversas métricas.

No estado da arte atual, os LLMs estão sendo continuamente desenvolvidos e avaliados. O *framework* GAIA usado para avaliação de grandes modelos de linguagem indicou a impossibilidade de um LLM responder a algumas perguntas que o classificariam como uma IA geral. Segundo o framework, uma IA geral deve responder a perguntas complexas que dependem de memorização, sequências de textos longos e com ações encadeadas que dependem da busca de informações na internet (Mialon *et al.*, 2023). O estudo mostrou que mesmo nas perguntas mais simples o ser humano ainda adota um desempenho muito superior (Figura 40).



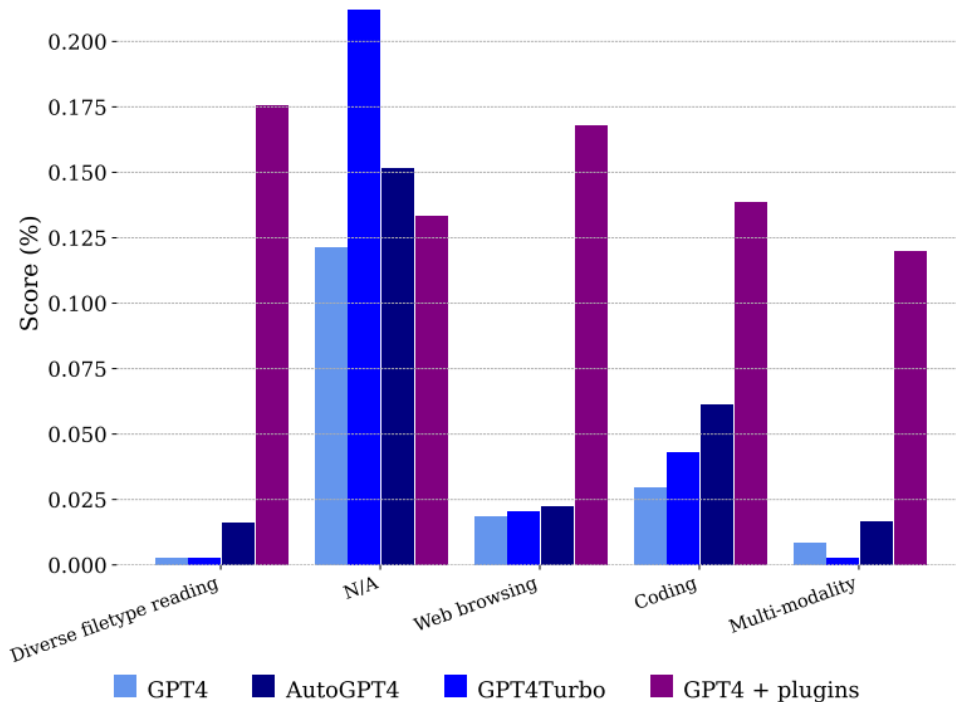
Fonte: Mialon *et al.* (2023).

Apesar da evidência da limitação dos grandes modelos de linguagem, ainda é possível perceber sua ampla adoção e desempenho significativo em diversas áreas, como na computação. A pesquisa realizada pelo site Retool (2023) indicou que 57.6% dos

entrevistados estavam utilizando menos o site StackOverflow<sup>21</sup> devido à adoção de IA como suporte à programação por meio do navegador ChatGPT e a extensão GitHub Copilot. Dos entrevistados, 10.2% deixaram de usar completamente o site StackOverflow como forma de busca por solução de problemas de programação.

O estudo de Mialon *et al.* (2023) ainda mostra a comparação do desempenho dos LLMs com diferentes recursos para executarem as mesmas tarefas. É significativa a diferença do resultado quando um dos modelos adota *plugins* externos que orientam a produção de conteúdo (Figura 41).

**Figura 41 - Comparação do desempenho dos LLMs em responder perguntas de dificuldade simples**



Fonte: Mialon *et al.* (2023).

Os LLMs independentes apresentaram desempenho significativamente inferior aos modelos que utilizaram *plugins* especialistas, apesar da capacidade dos LLMs em

<sup>21</sup> <https://stackoverflow.com/>

responder perguntas contextualizadas e relacionadas a uma busca semântica de informações. Um dos *plugins* disponíveis para uso com o ChatGPT, o Retrieval, visa efetuar buscas semânticas em documentos a partir de buscas feitas em linguagem natural em documentos a partir de metadados definidos pelo usuário, conforme a classe DocumentMetadataFilter apresentada no Quadro 11:

**Quadro 11- Exemplo de dados obtidos pelo plugin Retrieval**


```
class DocumentMetadataFilter(BaseModel):
    document_id: Optional[str] = None
    source: Optional[Source] = None
    source_id: Optional[str] = None
    author: Optional[str] = None
    start_date: Optional[str] = None # any date string format
    end_date: Optional[str] = None # any date string format
```


Fonte: o autor

Desta forma, os metadados podem ser alterados para o contexto do documento ou do objetivo da busca, enquanto o *plugin* alia a busca por rótulos definidos previamente nos documentos com o apoio de um LLM para ampliar o desempenho de busca. O processo baseia-se em efetuar uma busca em uma base de dados vetorial de textos usando uma ou mais sentenças de busca feitas em linguagem natural e filtros de metadados. O desempenho apresentado no exemplo, disponível no site da OpenAI, varia em torno de 80% de similaridade (a métrica de similaridade não foi declarada) (Figura 42).

No estado da arte atual, quando comparados um cenário de automação completa e automatização parcial, onde o ser humano ainda intervém no processo, a diferença de desempenho pode até ser próxima, mas ainda não é possível uma automação completa das atividades com *deep learning* (Mialon *et al.*, 2023). Segundo os autores, os modelos baseados em redes neurais podem falhar completamente e de modo repentino, indicando a dificuldade em serem incorporados nas novas tecnologias.

Figura 42 - Exemplo de busca em documento usando LLMs e o *plugin* Retrieval

 How has thinking on climate change evolved over the years you have access to? Use a table

 Used **United Nations** UNVERIFIED ▼

The thinking on climate change has evolved over the years as follows:

Year	Climate Change Thinking
2018	The Paris Agreement was recognized as an important expression of collective commitment to limit the rise in global temperature to well below 2 degrees Celsius. Countries were urged to turn pledges into national climate action, and cities, regions, territories, and private entities were contributing by setting their own ambitious targets. A summit was planned for 2019.
2019	The world was far off track to reach the Paris Agreement targets, with global temperatures potentially increasing by 3–5°C compared to pre-industrial levels. The United Nations collaborated with the Intergovernmental Panel on Climate Change (IPCC) on a special report outlining pathways to limit the global temperature rise to no more than 1.5°C. A climate summit was held in September 2019 to enhance national commitments on greenhouse gas reductions and strengthen adaptation measures.
2020	The Climate Action Summit in September 2019 resulted in

Fonte: <https://openai.com/blog/chatgpt-plugins>.

As abordagens que utilizam métodos baseados em regras e métodos baseados em aprendizagem de máquina diferem significativamente na capacidade de verificação do método. Enquanto no método baseado em regras todas as etapas estão declaradas, no método baseado em aprendizado de máquina (AM), que utiliza redes neurais ou redes neurais profundas, o método não pode ser auditado. As redes neurais se utilizam de organização de pesos em seus neurônios, os quais podem estar distribuídos em múltiplas camadas contendo dezenas ou centenas de neurônios em cada uma. Essa ponderação dos pesos que acionam as funções internas da rede não é definida pelo programador, mas sim

descoberta pela própria rede para alcançar um desempenho regido pelo algoritmo de otimização probabilística da rede. Enquanto a abordagem baseada em regras define o comportamento e as regras pelas quais o sistema funciona, a abordagem baseada em aprendizado de máquina define suas próprias regras.

Outro ponto importante a ser considerado no uso de grandes modelos de linguagem é que esses modelos são treinados em bases de dados da internet (Touvron *et al.*, 2023), portanto, não produzem bons resultados para ações de domínio especializado sem o ajuste fino (*fine tuning*) do modelo. Além disso, os LLMs são generativos, ou seja, criam conteúdo com base no treinamento realizado com dados externos ao problema. A geração de conteúdo está sujeita a alucinações, ou seja, o modelo pode produzir um conteúdo plausível, mas inverídico. Uma vez gerado o conteúdo, seria necessário validar o mesmo por um especialista para garantir sua veracidade e utilidade.

### 3. MÉTODO

Esta pesquisa adotará como método do *Design Science Research*. Desta forma, inicia-se resgatando os fundamentos do método para, na sequência, apresentar a apropriação dele no desenvolvimento dos objetivos pretendidos.

#### 3.1. Sobre o Design Science Research

O *Design Science Research* (DSR) é o método de pesquisa adequado para projetar e desenvolver artefatos e soluções prescritivas em um ambiente real ou simulado (Hevner; Chatterjee, 2010). No DSR todo o processo pode ser apresentado de forma transparente e mais compreensível, porque o pesquisador explicita conceitos e procedimentos relacionados com seu artefato proposto (Braun *et al.*, 2015).

Um artefato pode ser entendido como o conjunto de elementos que, organizados dentro de um ambiente interno dotado de determinadas características, executa uma determinada função para atingir um objetivo definido em um ambiente exterior a esse artefato (Simon, 1996). Os artefatos podem ser divididos em cinco categorias: **constructos**, **modelos**, **métodos** ou **teorias** (Lacerda *et al.*, 2013; Dresch; Lacerda; Antunes Jr, 2015; Hevner; Chatterjee, 2010).

Os **constructos** são vocabulários, símbolos (Hevner; Chatterjee, 2010) ou conceitos (Dresch, Lacerda; Antunes Jr., 2015) de determinado domínio que são capazes de descrever os problemas e especificar soluções. Nesse contexto se enquadram as taxonomias, dicionários e tesouros de determinado domínio do conhecimento.

Os **modelos** são um conjunto de proposições e declarações que expressam relações entre os constructos. Portanto eles são abstrações e representações da realidade que apresentam não apenas as variáveis de determinado sistema, mas as relações que são estabelecidas (Dresch; Lacerda; Antunes Jr., 2015). Nesse contexto se enquadram as ontologias e *frameworks* para definição de processos.

Os **métodos** são conjuntos de passos necessários para executar uma tarefa como algoritmos. São baseados em conceitos e termos (constructos) e representações do problema-solução (modelos). Podem ser utilizados para traduzir um modelo ou representação com objetivo de resolução de um problema (Lacerda *et al.*, 2013).

As **instanciações** são a realização de um artefato em seu ambiente e operacionalizam constructos, modelos e métodos (March; Smith, 1995).

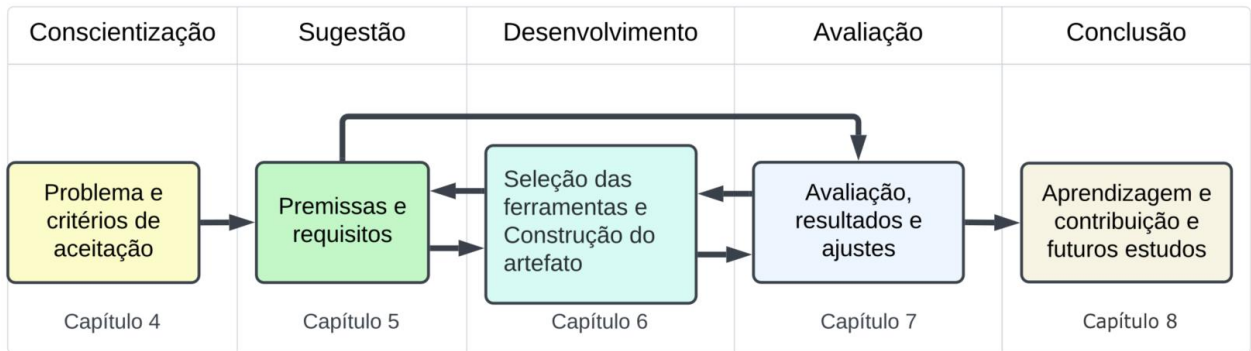
As **teorias ou proposições de projeto** são contribuições teóricas ou modelos genéricos (i.e., generalizações) que podem ser usados para desenvolver soluções de determinada classe de problemas (Dresch; Lacerda; Antunes Jr, 2015).

Nesta pesquisa, que trata da extração de informações para conversão da Norma de Desempenho em formato computável, serão elaborados os seguintes artefatos: 1) um conjunto de regras (**modelo**), de extração de informações e 2) um **método** que converte itens extraídos do texto em formato computável e, por fim, 3) a instanciação deste método com uma versão de parte da Norma de Desempenho em formato capaz de ser processado por máquinas. Os detalhes e a descrição de cada artefato será apresentado no Capítulo 5.

### 3.1.1. O delineamento da DSR adotado

A fim de viabilizar a pesquisa, as ações de pesquisa serão executadas de modo sequencial e com ações iterativas distribuídas nas cinco etapas definidas por Lacerda *et al.* (2013): conscientização, sugestão, desenvolvimento, avaliação e conclusão (Figura 43Figura 4).

**Figura 43 – Discriminação das etapas e seus objetivos aplicados a esta pesquisa**

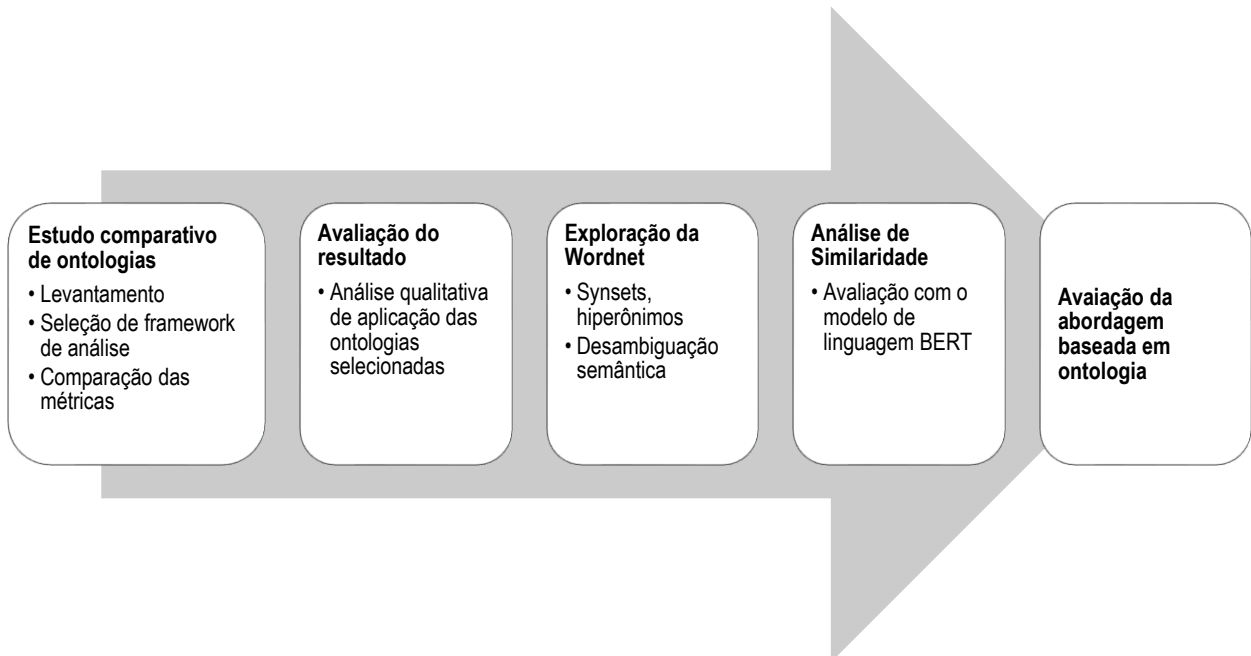


Fonte: o autor.

A seguir, são apresentados a descrição de cada uma das etapas desenvolvidas na pesquisa, um diagrama síntese e os procedimentos adotados para sua execução.

A primeira etapa do DSR é a **conscientização** (Figura 44), que direciona as demais etapas e é a tomada de consciência sobre o problema para identificá-lo e defini-lo com mais detalhamento. É uma etapa que avalia a relevância do artefato e cria o embasamento para sua construção.

**Figura 44 - Etapa de conscientização e o relacionamento de seus objetivos**



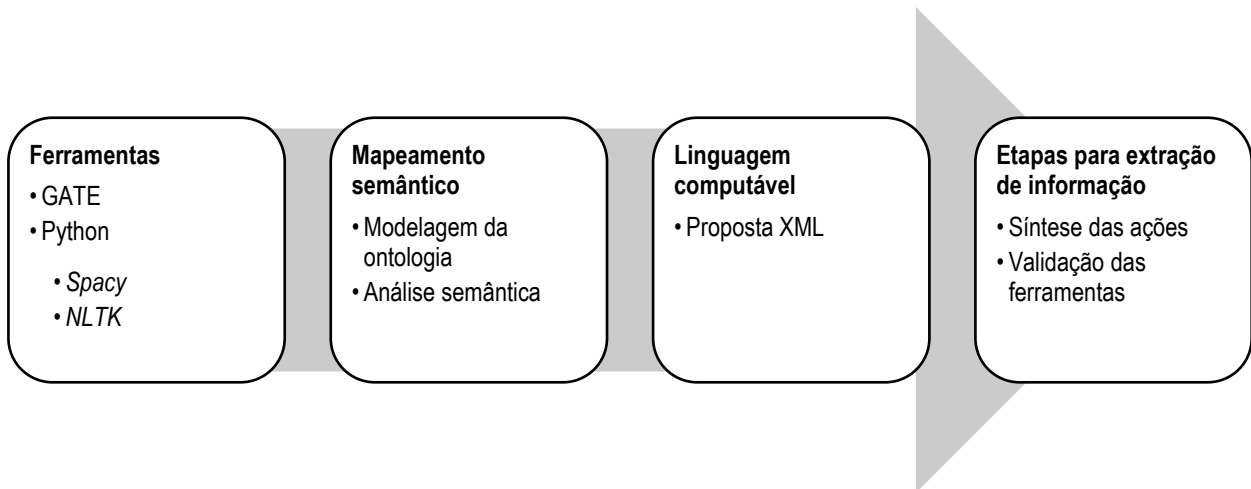
Fonte: o autor.

No Capítulo 4, a conscientização foi conduzida por meio de um levantamento e seleção das ontologias abertas relacionadas à construção civil. A análise das ontologias foi executada com o uso de frameworks consolidados de análise qualitativa que comparam atributos como modularidade e reusabilidade das ontologias. O objetivo desta etapa é a verificação da capacidade das ontologias, e de outros sistemas baseados em desambiguação semântica, em representar o contexto da Norma de Desempenho nas ações de resolução de entidades entre conceitos para extração de informações.

No Capítulo 5, é feita uma **sugestão** preliminar (Figura 45), deduzida a partir do contexto delimitado na etapa de conscientização. O objetivo desta etapa é a elaboração de uma arquitetura para o sistema e que apresenta as relações estruturais entre componentes e as restrições impostas pelo ambiente (Nunamaker; Chen; Purdin, 1990).

Nesta pesquisa elencamos de modo encadeado diferentes etapas e métodos usados para extração de informações. A partir das análises das referências realizada no Capítulo 2, foram selecionadas ferramentas para processamento de textos e para modelagem e processamento de ontologias. Em seguida foram executados experimentos que avaliaram a viabilidade de uso das ferramentas selecionadas a sugestão de uso da linguagem de marcação XML para futura conversão do regulamento. Como resultado desta etapa, apresentamos um fluxograma detalhando as etapas de execução do processo e todos os artefatos produzidos.

Figura 45 - Etapa de sugestão e o relacionamento de seus objetivos

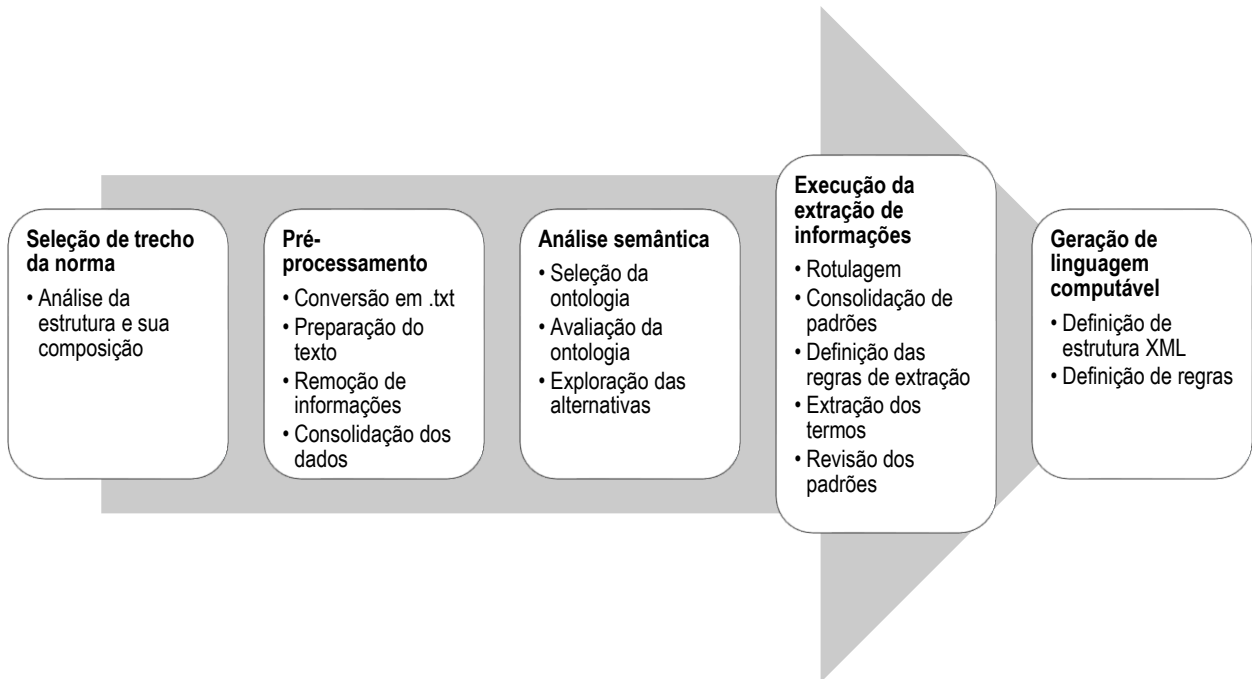


Fonte: o autor.

Já na terceira etapa, de **desenvolvimento**, ocorre a criação dos artefatos (constructos, modelos e métodos). Nesta etapa é determinada a funcionalidade esperada do artefato e sua arquitetura a partir das análises, teorias e exemplos analisados (ambiente externo) na etapa de conscientização (Figura 46). Ao final esta etapa o artefato final é construído em estado funcional com componentes (outros artefatos) e relações. Nesta etapa também são explicitadas suas limitações e considera-se quais soluções são satisfatórias para os problemas em estudo (Dresch; Lacerda; Antunes Jr, 2015).

Desta forma, o Capítulo 6 de Desenvolvimento, inicia com a descrição e argumentação sobre a seleção do trecho da Norma a ser traduzido. São descritas e iniciadas as etapas de processamento textual, a saber, pré-processamento, conversão de arquivos, preparação e consolidação dos dados textuais e remoção de informações desnecessárias.

Figura 46 - Etapa de desenvolvimento e o relacionamento de seus objetivos



Fonte: o autor.

Em seguida o capítulo desenvolve a extração de informações baseada em padrões linguísticos. A construção dos padrões foi uma ação experimental que desenvolveu combinações de padrões avaliando o resultado da extração. Em seguida, a aplicação dos padrões consolidados foi realizada em outro trecho da Norma de Desempenho. O final da etapa foi dado pela construção de um algoritmo para conversão dos dados em formato computável.

Na etapa seguinte, de **avaliação**, o comportamento do artefato é verificado no ambiente para o qual foi projetado e em relação às soluções ou objetivos que se propôs alcançar (Lacerda *et al.*, 2013) (Figura 47). Na avaliação é necessário considerar os critérios de validade científica e medidas de desempenho para comprovar a adequação do artefato (Lacerda *et al.*, 2013). Essa etapa itera com a anterior no caso da não adequação aos critérios estabelecidos previamente. Nesta etapa selecionamos as métricas usuais de qualidade de classificação de dados também relatadas pelas análises da bibliografia na etapa de Conscientização.

Figura 47 - Etapa de avaliação e o relacionamento de seus objetivos



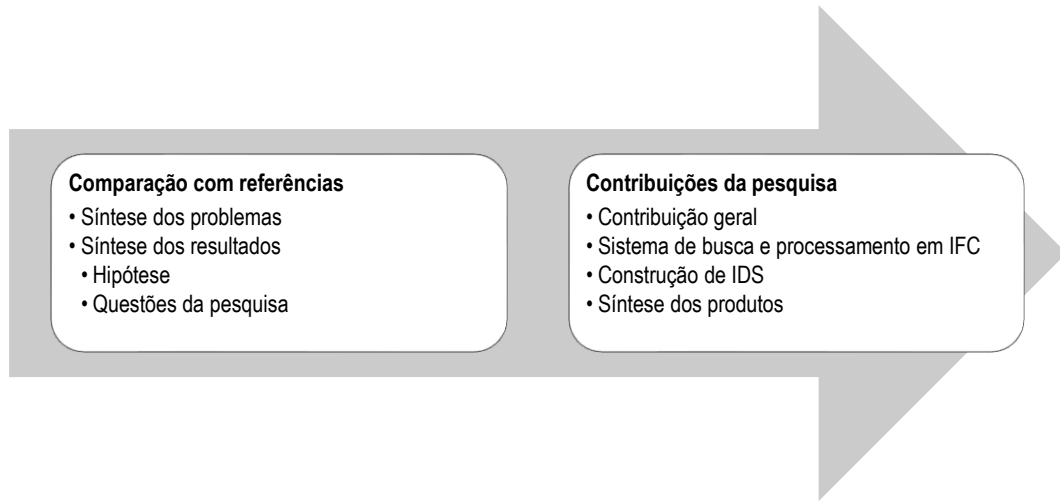
Fonte: o autor.

Foi executada a avaliação do conjunto de dados usado para construção dos padrões da extração de informações e, em seguida, foi avaliado o resultado obtido pela extração dos padrões no trecho usado para validação do resultado. Como forma de aprimorar a análise e interpretação dos resultados foi usado o método de agrupamento (*clusterização*) dos dados, que apresentou duas categorias de frases da Norma que obtiveram pior resultado nas análises.

Ainda ao final do Capítulo 6 de Avaliação, para verificação da confiança do método desenvolvido (baseado em regras), foi desenvolvido um experimento similar com outro paradigma de extração de informações e que tem apresentado destaque nas pesquisas recentes.

Ao final da pesquisa, a última etapa de **conclusão** (Figura 48) sintetizou as decisões e produtos da pesquisa, apontou a limitação do método e direciona futuros estudos como complemento do método.

**Figura 48 - Etapa de conclusão e o relacionamento de seus objetivos**



Fonte: o autor.

Mesmo que o artefato já esteja definido para um caso específico, é necessário avaliar seu desenvolvimento de modo generalista para uma classe de problemas. Então o último capítulo faz uma análise geral do processo generalizando sua aplicação e definindo a sua contribuição para o avanço do conhecimento e da ciência.

## 4. CONSCIENTIZAÇÃO

Conforme mencionado na Seção 2.4.4, a aplicação de uma ontologia, uma base de dados semânticos, visa a correta identificação e classificação dos termos identificados e capturados nos regulamentos. Diante da ausência de uma ontologia específica para a construção civil em língua portuguesa, realizou-se uma análise comparativa de outras ontologias já utilizadas em estudos correlatos, visando à sua possível incorporação nesta pesquisa.

### 4.1. Estudo comparativo de ontologias existentes

As ontologias podem descrever domínios relacionados à Indústria 4.0, como processos de *Design for Manufacturing and Assembly* (DfMA) para construção externa (Kalemi *et al.*, 2020). Além disso, uma ontologia de domínio pode ajudar a alcançar a intercambialidade de dados e modelos dentro e entre comunidades (Olivares-Alarcos *et al.*, 2019). Shin e Issa (2021), Sobhkhiz *et al.* (2021) e Shen *et al.* (2022) utilizam ontologias para validação de modelos para regulamentações.

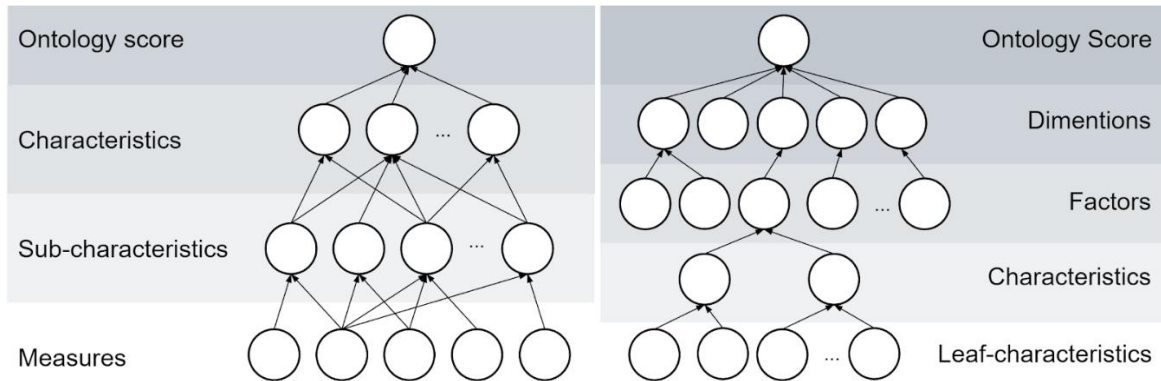
A engenharia de ontologias segue critérios rígidos de codificação, vinculação e validação de informações e a seleção de ontologias depende de uma qualificação prévia para seleção de possíveis candidatos. A análise de ontologias tem sido objeto de pesquisa desde o trabalho de Gruninger e Fox (1995), porém, a qualificação das ontologias é difícil de alcançar (Reiz *et al.*, 2020), especialmente se os referenciais de análise dependerem de uma abordagem subjetiva. Numerosos métodos quantificáveis podem ser usados para analisar uma ontologia, como suas propriedades e atributos de grafos, diversidade de relações, diversidade de relações e assim por diante. Os quadros de qualidade desenvolvidos proporcionam orquestração e significado a estas medidas arbitrárias (Reiz; Sandkhul, 2022).

Inicialmente, levantamos 29 ontologias representadas no Apêndice F. Das 29 ontologias, apenas aquelas que ainda estavam publicamente depositadas e atualizadas

foram selecionadas para análise. Como as regulamentações possuem um amplo espectro de termos e conceitos construtivos, foram selecionados apenas aqueles de escopo geral, retirando as ontologias especializadas. A seleção foi feita a partir da documentação das classes e subclasses que compõem os arquivos.

Aplicamos o *framework* OQuaRE para a comparação e análise das ontologias selecionadas, conforme o modelo multinível (Figura 49). Inicialmente é feito o levantamento das medidas elementares e calculadas as métricas (ANOnto, DITOnto) combinando as medidas elementares e a escala de normalização dos valores (Tabela 1). A normalização dos valores cumpre o objetivo de avaliar os parâmetros na mesma escala de valores, evitando a ênfase de que as características de valores mais elevados desequilibram a avaliação.

**Figura 49 - Análise multinível proposta por Duque-Ramos *et al.* (2011) à esquerda, e Lozanno-Tello e Gómez-Perez (2004) à direita**



Fonte: os autores.

A partir daí, é feita a média aritmética das métricas que afetam uma determinada subcaracterística (Coesão, Redundância e Similaridade) e a média aritmética das subcaracterísticas para definir um valor para as características principais (a Adequação Estrutural e Funcional) (Tabela 1). Se necessário, a pontuação global da ontologia pode ser obtida pela média das características.

**Tabela 1 - Dados originais de avaliação**

Métricas	Valores Originais			Normalização					Valores finais		
	DiCon	BOT	ifcOWL	1	2	3	4	5	DiCon	BOT	ifcOWL
ANOnto	5.04	84.20	2.09	[0–20]%	(20–40]%	(40–60]%	(60–80]%	> 80%	5	5	5
AROnto	1.51	1.10	1.20	[0–20]%	(20–40]%	(40–60]%	(60–80]%	> 80%	5	5	5
CBOnto2	0.97	0.40	1.17	> 12	(8–12]	(6,8]	(3,6]	[1,3]	5	5	5
CROnto	0.09	0.90	1.23	[0–20]%	(20–40]%	(40–60]%	(60–80]%	> 80%	1	5	5
DITOnto	10.00	2.00	10.00	> 8	(6–8]	(4,6]	(2, 4]	[1,2]	1	5	1
INROnto	1.74	0.40	3.88	[0–20]%	(20–40]%	(40–60]%	(60–80]%	> 80%	5	2	5
LCOMOnto2	2.36	1.40	2.96	> 8	(6–8]	(4,6]	(2, 4]	[1,2]	4	5	4
NOCOnto	1.84	1.00	4.09	> 12	(8–12]	(6,8]	(3,6]	[1,3]	5	5	3
NOMOnto	0.24	0.00	2.64	> 8	(6–8]	(4,6]	(2, 4]	<=2	5	5	4
RFCOnto	215.40	0.00	14311.52	> 12	(8–12]	(6–8]	(3–6]	[1–3]	1	5	1
RROnto	0.12	0.00	0.40	[0–20]%	(20–40]%	(40–60]%	(60–80]%	> 80%	1	1	2
TMOnto2	2.00	0.00	2.17	> 8	(6–8]	(4,6]	(2, 4]	(1,2]	5	5	4
WMCOnto2	1.19	1.00	1.47	> 8	(6–8]	(4,6]	(2, 4]	[1,2]	5	5	5

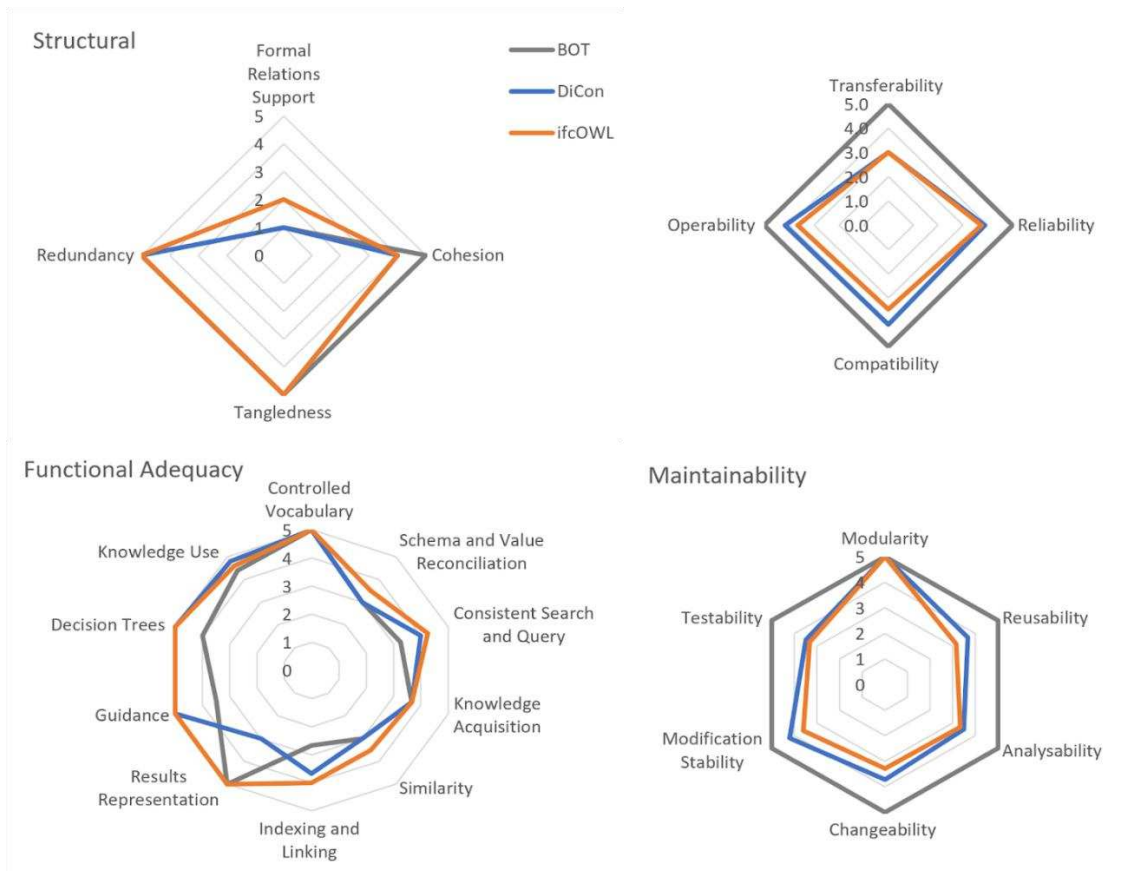
Fonte: Os autores.

Na Tabela 1 as métricas básicas estão representadas na primeira coluna. As outras colunas representam o processo de normalização dos valores em que 1 é o melhor resultado e 5 é o pior resultado.

Com as métricas quantificadas, obtivemos os valores das sub características (Figura 50) e características (Figura 51) a partir da média das métricas associadas.

Embora a ifcOWL ofereça os piores resultados na característica Adequação Funcional, importante para a extração de informações, a sua pior pontuação está na subcaracterística “Representação de Resultados”, que apresenta baixo impacto na extração de informações. Nas demais pontuações, os resultados das ontologias ifcOWL e DiCon são muito próximos, resultando na baixa capacidade do *framework* em discriminá-las para aplicação de extração de informações.

Figura 50 – Gráfico de radar com os resultados das medidas das subcaracterísticas



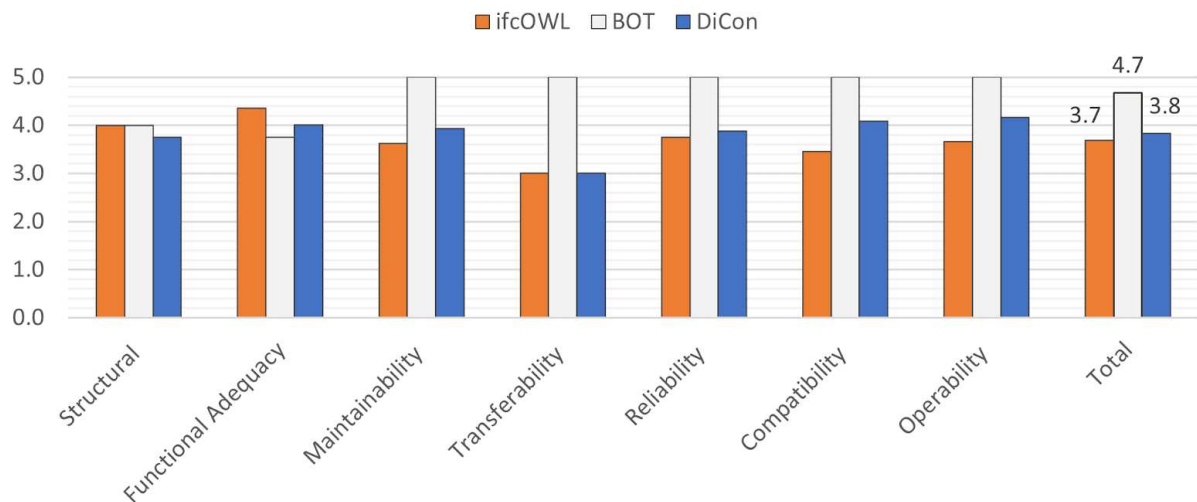
Fonte: os autores.

A BOT se destaca pela capacidade de orientar processos de busca fornecendo um contexto semântico para avaliação dos dados, principalmente por possuir propriedades e axiomas computáveis (subcaracterística “Busca e Consulta Consistentes”). Nos critérios de “Manutenção”, o BOT apresenta os piores resultados que indicam uma ontologia composta por elementos altamente relacionados e poucos discretos para dificultar mudanças em um componente, como na construção de relações com outras ontologias.

A característica “Transferibilidade” é importante para a aplicação da ontologia nos processos de extração de informação porque mede a capacidade da ontologia de se adaptar a diferentes linguagens e níveis de expressão sem aplicar ações além daquelas fornecidas por uma ontologia. Neste caso, a BOT apresenta o pior resultado.

O recurso Indexação e Vinculação avalia a capacidade da ontologia de fornecer informações rápidas para operações de recuperação de informações. Na análise realizada, o BOT atua com melhores pontuações neste item e no item Busca e Consulta Consistentes, que avalia a capacidade do contexto semântico em avaliar quais dados são desejados nas consultas dos usuários.

**Figura 51 -Valores das características por ontologia**



Fonte: os autores.

Embora o ifcOWL ofereça os piores resultados no grupo Adequação Funcional, característica importante para a extração de informações, sua pior pontuação está na subcaracterística “Representação de Resultados”, que apresenta baixo impacto na extração de informações. Nas demais pontuações, os resultados das ontologias ifcOWL e DiCon são muito próximos, resultando na baixa capacidade do *framework* em discriminá-las para aplicação de extração de informações.

O BOT se destaca pela capacidade de orientar processos de busca fornecendo um contexto semântico para avaliação dos dados, principalmente por possuir propriedades e axiomas computáveis (subcaracterística “Busca e Consulta Consistentes”). Nos critérios de “Manutenção”, o BOT apresenta os piores resultados que indicam uma ontologia

composta por elementos altamente relacionados e poucos discretos para dificultar mudanças em um componente, como na construção de relações com outras ontologias.

A característica “Transferibilidade” é importante para a aplicação da ontologia nos processos de extração de informação porque mede a capacidade da ontologia de se adaptar a diferentes linguagens e níveis de expressão sem aplicar ações além daquelas fornecidas por uma ontologia. Neste caso, o BOT apresenta o pior resultado.

O recurso Indexação e Vinculação avalia a capacidade da ontologia de fornecer informações rápidas para operações de recuperação de informações. Na análise realizada, o BOT atua com melhores pontuações neste item e no item Busca e Consulta Consistentes, que avalia a capacidade do contexto semântico em avaliar quais dados são desejados nas consultas dos usuários.

**Tabela 2 - Outras medidas elementares usadas para discriminação dos resultados**

Elemental Measure	DiCon	ifcOWL
<b>classAnnotations</b>	736.00	0.00
<b>axioms</b>	6986.00	20823.00
<b>classes</b>	478.00	1327.00
<b>logicalAxioms</b>	1775.00	13971.00
<b>individuals</b>	142.00	1162.00
<b>classesWithObjectProperties</b>	95.00	722.00
<b>dataProperties</b>	129.00	5.00
<b>objectProperties</b>	363.00	1591.00
<b>objectPropertiesOnClasses</b>	117.00	3498.00
<b>OQual_Axiomclass_ratio</b>	14.62	15.69
<b>OntoQA_Average_Population</b>	0.30	0.88
<b>OntoQA_Cohesion</b>	7.00	10.00
<b>OntoQA_Inheritance_Richness</b>	1.74	3.88

Fonte: os autores.

As medidas acima não existem no framework OQuaRE e evidenciam a discrepância da composição das ontologias. A métrica OQual\_Axiomclass\_ratio avalia o número de axiomas por classe e é uma medida relevante porque os axiomas são as entidades ontológicas que permitem a desambiguação entre os termos. No OntoQA\_Cohesion do *framework* OntoQA, valores mais baixos indicam menos nós

conectados, sugerindo maior facilidade de associação semântica entre elementos da ontologia e uma base externa. A medida *OntoQA\_Inheritance\_Richness* denota ontologias horizontais, com pouca herança (valores mais baixos) das ontologias mais verticais, com maior profundidade.

As ontologias foram avaliadas para uso em desambiguação semântica para aumentar a precisão de possíveis tarefas de extração de informações. Neste cenário, a característica “Adequação Funcional” do OQuaRE recebe maior importância por possuir subcaracterísticas baseadas na capacidade da ontologia de representar e fornecer informações. As subcaracterísticas “Busca e consulta consistente” e “Aquisição de conhecimento”, por exemplo, baseiam-se em medidas do número de classes, atributos, propriedades e anotações e indicam a capacidade de representação semântica da ontologia e o domínio do conhecimento. Quanto mais ricas forem essas características, maior será o nível de granularidade e especificidade da ontologia.

É possível entender o ifcOWL como uma ontologia com alto nível de granularidade composta pela representação de metaconceitos que não serão utilizados por consultas em integrações futuras. Ao mesmo tempo, DiCon oferece um som mais sucinto, modulado e sem excessos. O grande tamanho do ifcOWL levou Pauwels (2020) a criar uma versão simplificada com a descrição apenas das classes vinculadas à árvore *IfcBuildingElement* que contém os produtos de construção mais comuns. Enquanto ifcOWL tem mais de 20.000 axiomas, *Building Element Ontology* tem pouco mais de 1.500 e uma média de 8 axiomas por classe.

As métricas OQuaRE de Adequação Funcional indicam altos valores de propriedades, anotações e relacionamentos por classe, sugerindo que a BOT é mais capaz de representação de conhecimento do que DiCon e ifcOWL. Porém, a análise quantitativa indicou que BOT possui poucas classes e DiCon e ifcOWL possuem valores bem superiores. Quando analisamos o relacionamento de classe do BOT, percebemos que ele seria limitado para uso de desambiguação porque não representaria um espectro granular do domínio.

A ontologia ifcOWL é a maior das três e possui valores muito mais elevados. Porém, uma análise mais aprofundada identificou que é construída a partir de relações entre metacaracterísticas que criam um caminho de busca multiplicado e redundante devido à natureza do esquema IFC. Em sistemas robustos como em Cao *et al.* (2022), a incorporação da ifcOWL é válida porque está alinhada ao formato de arquivo BIM. Porém, sua utilização em um ambiente de ampla industrialização com a incorporação de protocolos como o DFMA, a ifcOWL possui um excesso de entidades ontológicas que não permitem buscas e representações eficientes.

Comparando à DiCon sem normalização (Tabela 1), as diferenças entre as duas ontologias são mais evidentes. A DiCon possui menos entidades de escopo mais generalista, simplificando e acelerando uma busca ou análise detalhada a partir desses pontos. Além dos critérios quantitativos, o critério de seleção por modularidade é relevante nas ontologias mais recentes (SAREF, BIMMER e DiCon). Assim, a melhor alternativa entre as três ontologias é a DiCon, que atualmente é composta por 14 módulos. DiCon ainda possui muitas classes, axiomas e atributos e uma pontuação final comparável ao ifcOWL.

Werbrouck *et al.* (2020) identificam que grafos baseados em ontologias complexas como ifcOWL são difíceis de manter e vincular com outras ontologias devido à necessidade de um maior mapeamento entre conceitos das duas ontologias. A solução para ontologias de maior demanda por representação de entidades é a modulação de vocabulários e ontologias como DiCon.

Corroboramos em entendimento de Reiz e Sandkuhl (2022) de que os *frameworks* atuais não devem ser utilizados para simples análise de resultados; não existe solução mágica para avaliação da qualidade na construção e avaliação de ontologias. Fica explícito nos gráficos apresentados a impossibilidade de diferenciação entre ontologias dado que os meios quantitativos do *framework* sofrem uma normalização. É também evidente que o quadro OQuARE não fornece informação significativa quando as outras medidas do

quadro não o acompanham e não estão contempladas na ferramenta de análise automatizada NEOntometrics.

As análises indicam também que o *framework* OQuaRE utilizado isoladamente não consegue apresentar informações suficientemente úteis para a qualificação de ontologias de forma automatizada. No caso do experimento desenvolvido foi necessário recorrer a outras métricas externas.

Considerando que o grupo de ontologias será selecionado por profissionais que não desejam aprofundar a análise, as métricas e ferramentas baseadas em dimensões elementares da ontologia são úteis para identificar discrepâncias entre os dados. A partir do resultado inicial é possível desenvolver uma segunda etapa de análise direcionada à compreensão dos elementos e maior diferenciação das ontologias permitindo maior assertividade na avaliação.

Desta forma, a análise de ontologias nesta pesquisa organizou informações dispersas sobre ferramentas e os meandros envolvidos em uma avaliação abrangente de ontologias e identificou a necessidade de novas avaliações além das ferramentas apresentadas. Como existem diversos critérios para avaliar uma ontologia para cada tipo de aplicação, um método quantitativo, conforme apresentado neste artigo, limita-se a uma análise preliminar e mais geral da discriminação de ontologias. Para a desambiguação semântica, por exemplo, é necessário avaliar se a ontologia permite fácil manutenção e modularidade para ser preenchida com novas instâncias e propriedades que completariam sua representação de um domínio mais específico. Para tais casos, o método baseado em métricas demonstrou sua utilidade ao facilitar a análise preliminar de ontologias, auxiliando os profissionais a se aprofundarem em um conjunto de dados menor.

As análises subsequentes variam dependendo do contexto da aplicação, mas requerem mais tempo e conhecimento. Para a desambiguação semântica nas atividades de extração de informações, é necessário avaliar se os nomes e propriedades das instâncias e classes da ontologia estão alinhados com o domínio da aplicação. Embora uma análise

precisa deva ser conduzida por especialistas, é desejável que eles trabalhem com um corpo de ontologias previamente selecionadas por um método auxiliar, conforme apresentado neste artigo.

No entanto, a ontologia DiCon 0.5 emergiu como notável, apresentando uma rica composição de axiomas, classes, atributos e relações. Crucialmente, estes elementos estão intrinsecamente distribuídos por toda a ontologia, com um forte foco em entidades reais. Esta característica torna-o mais apto a representar o contexto autêntico e, conseqüentemente, mais adequado à desambiguação semântica. A modularidade do DiCon 0.5 oferece maior flexibilidade para implementação em diversos sistemas e alinhamento com outras ontologias especializadas. A capacidade da ontologia para múltiplos relacionamentos facilita sua integração em uma cadeia de produção 4.0 e 5.0 mais ampla, permitindo conexões perfeitas entre dispositivos, edifícios e linhas de produção.

## **4.2. Resultado da seleção**

Nesta etapa, procedeu-se à revisão da análise dos métodos empregados por artigos de referência nesta pesquisa que executavam o processo de mapeamento mediante buscas na base de dados semânticos. Nestas pesquisas, os elementos textuais do regulamento eram pesquisados na referida base de dados e os conceitos que compõem a ontologia foram utilizados para estabelecer correspondências entre propriedades ou instâncias de classes. Portanto, é essencial que a ontologia seja enriquecida com conceitos instanciados ou com relações (propriedades) declaradas nos textos.

Apesar de estar em uma estrutura de dados mais simples e capaz de otimizar o processo de busca por informações, a DiCon ainda é ausente de classes ou elementos relacionados à construção civil. Um exemplo são as categorias e características relacionadas à condição de proteção e combate a incêndio. Outro exemplo são termos singulares da língua portuguesa e que não são encontrados nas ontologias mencionadas,

como “assobradadas” e “sobretensão”. Termos muito específicos dependeriam de uma ontologia com um nível de granularidade muito alto para poder identificar estes termos. Ontologias específicas com termos próprios do domínio existem na língua inglesa, mas não possuem alinhamento com as demais mais gerais.

Como a ontologia DiCon não possui instanciações suficientes e em português, não poderia ser usada. A ifcOWL possui de relações hierárquicas de alta complexidade relacionando classes de objetos e classes abstratas. A ifcOWL atribui propriedades às relações entre objetos “owl:ObjectProperty”, mas não possui relações de propriedades de dados “owl:DatatypeProperty”. Isso afeta a extração dificultando que atributos e valores de propriedades sejam facilmente mapeados às propriedades dos elementos mencionados no texto, como isolamento e dimensões. Ainda que o mapeamento seja possível, a ontologia não é povoada com instâncias que representam o contexto da norma tornando o trabalho de representação ainda mais complexo e demorado.

Desta forma foi avaliada a possibilidade de criação de uma ontologia própria da norma, mas esta ação gera um problema de generalização do resultado. Uma vez que seria constituída uma base de dados orientada a um documento específico, outros documentos não terão o mesmo desempenho no reconhecimento de seus termos dentro desta base de dados. Portanto, a constituição de uma base de dados dedicada à Norma de Desempenho não é uma solução eficaz porque demandaria uma carga de trabalho excessiva para atingir um resultado restrito a apenas um documento.

Desta forma a pesquisa explora outra base de dados semânticos capaz de identificar os termos das sentenças por similaridade a outros termos de mesmo contexto. Ao identificar que os termos encontrados possuem proximidade de outros termos de mesmo domínio, é possível classificar os termos encontrados. A esta diferenciação e relacionamento entre termos similares, damos o nome de **desambiguação semântica**.

### 4.3. Exploração da ontologia WordNet para desambiguação dos termos da Norma de Desempenho

WordNet é uma base de dados semânticos (ontologia) em inglês e composta por verbos, nomes, adjetivos e advérbios representados por conceitos semânticos chamados *synsets*. Cada *synset* é conectado a outros a ele relacionado por atributos como hierarquia e sentido (ver Quadro 12). A WordNet foi criada a partir de iniciativas dos anos 1980 relacionadas a entender como crianças aprendem novas palavras (Miller; Fellbaum, 2007). A WordNet, assim como outras bases semânticas (como a DBpedia), é utilizada para ações de desambiguação semântica no campo do Processamento Natural de Linguagem e é usada como uma grande base de dados para treinamento para algoritmos de Aprendizagem de Máquina.

A WordNet foi iniciada para a língua inglesa e já possui alinhamento com outras línguas. Atualmente existe um grande conjunto de WordNets formalizadas em diferentes línguas baseadas no projeto original e estão alinhadas com o projeto em língua inglesa. O projeto Open Multilingual WordNet<sup>22</sup> (OMW), utilizado nesta pesquisa, consolidou diversas WordNets em um único pacote multilíngue que expande o acesso à base de dados em outras línguas, como o português. Atualmente a base de dados semânticos pode ser acessada via código com o pacote “omw-1.4” incorporado na biblioteca NLTK.

No caso da língua portuguesa está inserida neste projeto a OpenWN-PT<sup>23</sup> com 54.071 palavras estruturadas e representa aproximadamente 25% das palavras da língua portuguesa<sup>24</sup>.

A busca por palavras por si não retorna resultados significativos, mas é possível acessar as relações e atributos das palavras (Quadro 12). Esta pesquisa desenvolveu uma análise de busca na base semântica da OMW explorando relações entre conceitos

---

<sup>22</sup> <https://omwn.org/omw1.html>

<sup>23</sup> <https://github.com/own-pt/openWordnet-PT>

<sup>24</sup> <https://www.academia.org.br/nossa-lingua/dicionario-da-lingua-portuguesa>

semânticos, ou “*synsets*”. Cada *synset* é uma representação semântica na OMW que relaciona consigo outros conceitos como sinônimos, definições e outros conceitos em hierarquias taxonômicas.

**Quadro 12 - Principais atributos acessados na OMW**

Atributo buscado	Conteúdo retornado
<i>synsets</i>	Lista indexada de conceitos com sentidos similares
<i>synsets(#id).examples()</i>	Exemplo de aplicação do <i>synset</i> #id
<i>synsets(#id).definition()</i>	Definição do termo
<i>synonyms</i>	Lista indexada de sinônimos
<i>hypernyms</i>	Lista indexada de superconceitos (superclasse)
<i>hyponyms</i>	Lista indexada de subconceitos (subclasse ou instância)

Fonte: o autor.

Um dos experimentos desenvolvidos por esta pesquisa tentou analisar correspondência entre os conceitos da base semântica Open Multilingual WordNet e conceitos da construção civil (Quadro 13).

**Quadro 13 - Exploração de *synsets* da Open Multilingual WordNet**

```
wn.synsets("construção", lang='por')

>> [Synset('construction.n.01'), Synset('fabrication.n.04'),
Synset('construction.n.07')]

wn.synset('insulating_material.n.01').definition()

>> "a material that reduces or prevents the transmission of heat or sound or
electricity"

wn.synsets('instalação', lang = 'por')[1]
print(material)

>> synset('facility.n.01')

material. material.hypernym_distances()

>> {(Synset('artifact.n.01'), 1), (Synset('entity.n.01'), 5),
(Synset('facility.n.01'), 0), (Synset('object.n.01'), 3),
(Synset('physical_entity.n.01'), 4), (Synset('whole.n.02'), 2)}
```

Fonte: o autor.

No exemplo acima, foi realizada uma busca pelos *synsets* de classes superiores de maior abstração na taxonomia retornando conceitos como “entidade”, “objeto”, “entidade

física” e “artefato”. No exemplo acima os valores indicados para cada *synset* indicam quão distantes eles estão do nó relacionado ao *synset*. No caso apresentado, o conceito “artefato” é o mais próximo, i.e., similar ao conceito de “instalações”.

A resolução de similaridade entre conceitos na OMW é dada por diferentes métodos que se baseiam entre distâncias entre nós do grafo e profundidade de conceitos nas taxonomias. Ao tentar desenvolver uma busca aplicável ao contexto da construção civil foi realizado um experimento com palavras identificadas na Norma de Desempenho, como “porta” e “isolante”. O experimento procurou similaridade entre os conceitos segundo três métricas de similaridade incorporadas no pacote “own-1.4”. O resultado apresentado na Tabela 3 foi obtido iterando os *synsets* de “porta” com os de “isolante” para revelar os que possuem maior similaridade.

**Tabela 3 - Análise de similaridade de conceitos da OMW**

Nº	Synsets		Métricas de similaridade		
	porta	isolante	Wu-Palmer	Path	Leacock Chodorow
1	'car_door.n.01'	'insulator.n.01'	0.24	0.07	1.0
2	'car_door.n.01'	'insulating_material.n.01'	0.56	0.11	1.44
3	'door.n.01'	'insulator.n.01'	0.25	0.08	1.07
4	'door.n.01'	'insulating_material.n.01'	0.59	0.12	1.56
5	'door.n.05'	'insulator.n.01'	0.27	0.08	1.15
6	'door.n.05'	'insulating_material.n.01'	0.62	0.14	1.69
7	'doorway.n.01'	'insulator.n.01'	0.27	0.08	1.15
8	'doorway.n.01'	'insulating_material.n.01'	0.62	0.14	1.69
9	'exit.n.01'	'insulator.n.01'	0.31	0.1	1.34
10	'exit.n.01'	'insulating_material.n.01'	0.71	0.2	2.03
11	'gate.n.01'	'insulator.n.01'	0.25	0.08	1.07
12	'gate.n.01'	'insulating_material.n.01'	0.59	0.12	1.56
13	'interface.n.04'	'insulator.n.01'	0.24	0.07	1.0
14	'interface.n.04'	'insulating_material.n.01'	0.56	0.11	1.44

Fonte: o autor.

A tabela mostra que o valor máximo de similaridade é de 71%, que é um valor baixo esperado para o contexto da Norma de Desempenho. Considerando que uma ontologia deve conseguir mapear com precisão dois termos relacionados no mesmo domínio, a busca na Open Multilingual WordNet não atinge resultado satisfatório porque agrega significados não relacionados com a construção civil, dificultando a busca. Além

disso, quando relaciona *synsets* próximos de seu uso no domínio da construção, como a linha 10 da Tabela 3, o valor de acurácia ainda não é alto. No caso de utilização da OMW como uma ontologia possível para desambiguação semântica no domínio da construção civil estes dois conceitos deveriam estar próximos, seja porque “isolante” estaria como uma propriedade do objeto “porta” ou estaria relacionado a “porta” em até dois níveis de relacionamento.

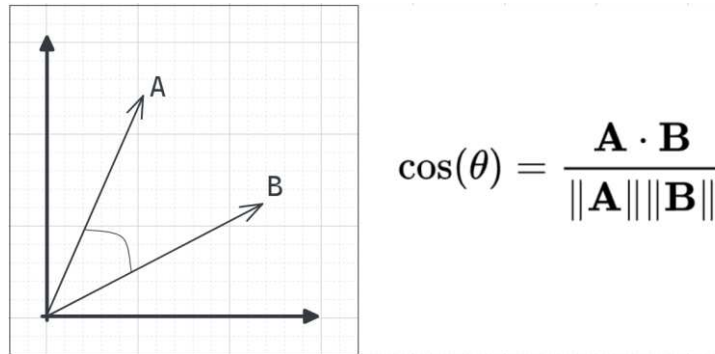
Outro problema encontrado é que a quantidade de elementos instanciados na OMW em língua portuguesa não é suficiente para representar o domínio da Norma de Desempenho. Um termo recorrente como “vida-útil”, que é composto, não possui reconhecimento. Buscas sobre sinônimos, hiperônimos e Tópicos de domínio, por exemplo, retornam resultados vazios.

#### **4.4. Exploração da análise de similaridade de cosseno para desambiguação semântica**

A abordagem de exploração da *WordNet* indicou a possibilidade de relacionar as palavras a termos-chave da construção civil para validar estes termos como elementos construtivos ou componentes do edifício. Porém, com o resultado inexpressivo e insuficiente obtido para caracterização automática dos termos, este trabalho explorou novas métricas de similaridade para validar a abordagem de desambiguação por similaridade.

Ao invés de métricas de similaridade baseadas em profundidade da taxonomia ou quantidade de nós, foi explorada a análise de similaridade de cosseno (Figura 52). A similaridade de cosseno é o cálculo da distância entre vetores que é dada pelo produto de dois vetores  $\bar{A} * \bar{B}$  normalizada pelo produto do comprimento dos vetores  $\|A\| * \|B\|$ .

Figura 52 - Similaridade do cosseno



Fonte: O autor.

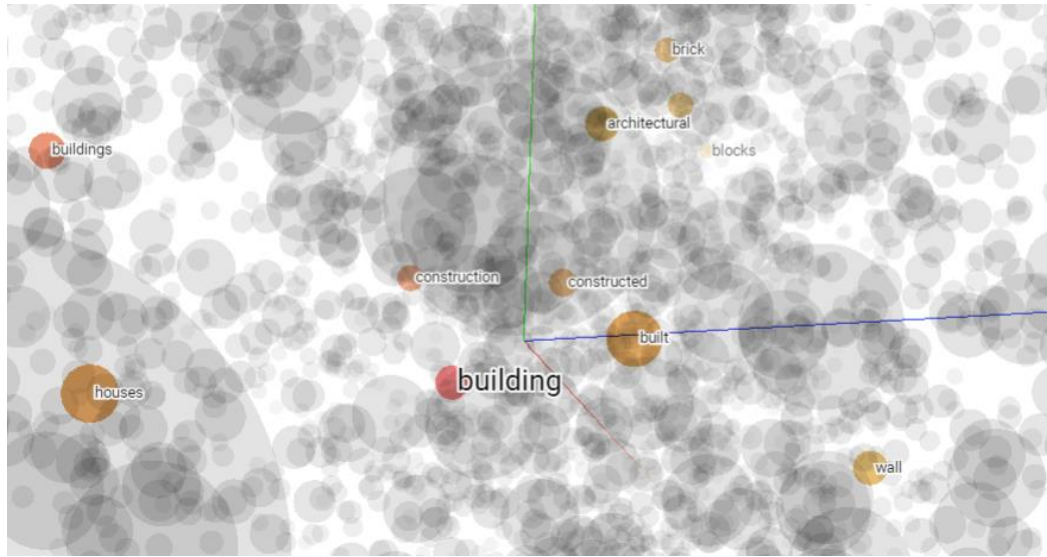
Segundo o cálculo acima, vetores próximos, com valor de  $\theta$  próximo de  $0^\circ$  indica alta similaridade ( $\cos 0^\circ = 1$ ), mesmo que tenham comprimentos distintos. Para que o método de similaridade seja utilizado é necessário recorrer a outro modelo de representação dos dados de linguagem. É utilizado o método de *word embeddings* que vetoriza as palavras no seu contexto de uso. A Figura 53 é um exemplo de uma representação vetorial de palavra.

A imagem é abstração, em 3 dimensões, de um espaço vetorial de duzentas dimensões que representam a palavra “*building*” em um corpus de palavras vetorizadas. Os círculos em amarelo são as palavras mais similares à “*building*” no espaço de 200ª dimensão. A imagem mostra outros círculos cinzas, mas que não estão próximos do vetor “*building*”<sup>25</sup> e atingem um coeficiente de similaridade próximo de 90%.

---

<sup>25</sup> A impressão de proximidade (ou similaridade) é dada porque a representação tridimensional é gerada pelo método *Principal Component Analysis* (PCA) que condensa os atributos originais para o espaço visível de três dimensões.

Figura 53 - Visualização da similaridade do termo "*building*"<sup>26</sup> na plataforma tensorflow.org



Fonte: O autor.

Para tentar garantir qualidade na representação vetorial das palavras, foi utilizado o modelo de aprendizado de máquina *Bidirectional Encoder Representations from Transformer* (BERT) que utiliza aprendizado profundo (*deep learning*). Este modelo foi desenvolvido em 2018, é utilizado pela empresa Google para buscas e testes de NLP. Ele possui um *corpus* textual de *tokens* vetorizados que permite a busca com maior acurácia. O modelo foi implementado em um código simples usando a versão de mais atributos e em português (*bert-large-portuguese-cased*). O modelo está incorporada na biblioteca *Transformers*<sup>27</sup> e o cálculo de similaridade de cosseno foi feito com uso da biblioteca *sklearn*. A Tabela 4 mostra parte do resultado obtido ao calcular o coeficiente para as vinte e uma sentenças do Capítulo 8 da Norma.

<sup>26</sup> Conferir <https://projector.tensorflow.org/>.

<sup>27</sup> <https://huggingface.co/docs/transformers/index>

**Tabela 4 - Resultado de similaridade entre três conceitos e alguns termos da Norma**

Conceito	Termos	Similaridade
construção civil	projeto estrutural	0,818
	sistema construtivo	0,811
	edificações habitacionais	0,808
	aço	0,752
	concreto	0,745
	ABNT	0,704
	isolamento	0,659
	sobretensões	0,658
	condição	0,646
	comprovação	0,604
	rotas	0,581
construção	edificação	0,926
	concreto	0,890
	aço	0,846
	selos	0,821
	estabelecido	0,819
	isolamento	0,815
	portas	0,805
	incêndio	0,791
	que	0,755
	ABNT	0,724
	comprovação	0,654
curtos-circuitos	0,589	
habitação	edifícios	0,949
	revestimento	0,900
	isolamento	0,880
	disposto	0,881
	concreto	0,874
	aço	0,810
	projeto	0,731
	rotas	0,690
	face interna	0,651
	ABNT	0,634
	descargas atmosféricas	0,544

Fonte: O autor.

O resultado obtido mostrou valores variando de 0.949 a 0.517. A abordagem baseada em um modelo de linguagem como BERT para análise de similaridade foi consistente, repetindo o mesmo valor para múltiplas aparições do mesmo termo. Na análise completa foi possível observar que termos específicos foram corretamente relacionados com os conceitos-chave “construção civil”, “construção” e “habitação”. Termos como estanqueidade, isolamento e sobretensões foram corretamente classificados como próximos dos conceitos-chave. Porém, a classificação teve alta variação considerando termos genéricos como “análise”, “condição”, “estabelecido” e “que” com

níveis de similaridade iguais ou superiores a termos que deveriam estar mais próximos. Termos como “acabamento”, “face interna”, “descargas atmosféricas” e “curtos-circuitos” e “alarme” são termos que seriam desejados em uma extração de informações, mas pelo valor de similaridade seriam descartados.

A exploração da análise de similaridade baseada em cosseno utilizando o modelo BERT obteve resultados precisos, mas com baixa acurácia na identificação de termos relevantes. Desta forma os resultados da Tabela 4 mostram o método de desambiguação deve ser descartado como uma alternativa ao uso da ontologia porque não mostrou capacidade de desambiguação de termos de uma sentença típica de um regulamento.

#### **4.5. Avaliação da abordagem baseada em ontologia**

A desambiguação executada em algumas das referências deste trabalho utilizava o método baseado em ontologias, mas para tanto ficou claro que deve existir uma ontologia especialista do domínio capaz de cobrir mais documentos. Nas pesquisas analisadas, poucas ontologias de fato possuíam o caráter geral que permitiria que outros documentos relacionados pudessem ser processados e mapeados na ontologia. Em alguns casos analisados as ontologias possuíam alto grau de especificidade, restringindo seu uso para um domínio de aplicação muito específico. Desta forma, as ontologias existentes e analisadas (Seção 5.3.1) não possuem a capacidade de generalização do processo demandada.

Realizamos uma tentativa de produção de uma ontologia baseada no sistema de classificação da ISO 12006-2, porém em um processo *botton-up* de construção da ontologia a partir dos termos e atributos definidos no texto. A tentativa então foi interrompida porque foi percebido que não haveria tempo para a coleta da informação de uma ontologia dedicada à Norma de Desempenho. Mas, ainda que houvesse tempo hábil, a construção *bottom-up* iria reproduzir apenas os termos e atributos dos componentes existentes no texto da norma criando um constructo que possuiria a paridade completa

de seus termos com o documento. Ao criar uma ontologia a partir dos termos da norma, essa ontologia seria dedicada à Norma e perderia sua capacidade de generalização.

Após as tentativas de adaptar uma ontologia específica como ifcOWL e DiCon para desambiguação mostrarem a impossibilidade devido à ausência de instanciação de elementos nos conceitos ou propriedades das ontologias, a *WordNet* 3.0 se apresentou como uma base de dados não específica que permitiria a desambiguação não apenas para a Norma de Desempenho, mas para outros regulamentos da construção. Mas seu limite linguístico para o português (pt-BR) direcionou o uso para a *Open Multilingual WordNet*, que possui incorporada a OpenWN-PT em língua portuguesa. As análises indicaram que, no contexto da construção civil, a desambiguação e o mapeamento de conceitos de regulamentos não possuíam resultados satisfatórios, impedindo o uso de uma ontologia geral, não específica, para estas ações.

Desta forma a pesquisa se consolidou como um método de extração de informações que não é baseado em ontologias ou outra base de dados semânticos. **A pesquisa então optou pela abordagem baseada em regras com a construção de padrões e análises gramaticais das sentenças da Norma para extrair termos importantes para a tentativa de construção de uma versão computável da Norma de Desempenho.**

## 5. SUGESTÃO

Neste capítulo foi feita a proposição da sequência de ações planejadas para desenvolvimento dos artefatos da pesquisa. Ele inicia com a apresentação da estrutura de ações propostas e uma breve descrição de cada ação indicando variáveis e objetivos de cada ação. Ao final são propostos os artefatos resultantes das ações.

A pesquisa usou a versão 2021 da Norma, que possui seis partes. Dada a complexidade do processo em desenvolver regras e padrões para extração de informações, foi feita uma seleção de um trecho da norma para essa construção do algoritmo.

O trecho selecionado foi avaliado segundo a composição de sua estrutura da previsão de padrões e métodos disponíveis para seu processamento usando PLN.

A seleção do trecho considerou as classificações de tipos de requisitos feitas por Solihin e Eastman (2015), Zhang e El-Gohary (2017), Soliman-Junior, Formoso e Tzortzopoulos (2019) e Zheng *et al.* (2022). Estes autores desenvolveram classificações de regulamentos como regulamentos relacionados a critérios quantitativos, geométricos, espaciais, de maior ou menor grau de dificuldade de representação geométrica, subjetividade ou condicionamentos existenciais. Porém, como nesta pesquisa foi feita a tentativa de conversão do conteúdo completo da norma representada pelo trecho selecionado para a extração, será feita a análise dos requisitos como estão apresentados, independente das características de suas diretrizes.

Para seleção do método de extração de informações, foram consideradas as análises de Ren *et al.* (2022) e as pesquisas de Zhang e El-Gohary (2011, 2021) mencionadas na Seção 2.4.4. Os autores mostram o potencial da utilização de redes neurais, mas indicaram que o processo de extração de informações baseado em ontologia (OBIE) ainda oferecia melhor desempenho.

Este trabalho considerou que a extração de informações da norma deve garantir que os termos extraídos são termos existentes da construção civil e que foram classificados

corretamente em categorias reais. Desta forma, um resultado esperado neste cenário pode apresentar mais falsos negativos do que falsos positivos. Um falso positivo pode indicar conceitos e termos que não existem e que não deveriam ser considerados pelo usuário criando necessidade de retrabalho. Já os falsos negativos criariam um sistema incompleto de termos, mas não apresentaria termos errados ao usuário.

A partir das comparações e considerações apresentadas, **este trabalho se orienta pela construção de um sistema baseado em regras de extração de informação. O método é baseado no reconhecimento de padrões gramaticais e sintáticos do texto normativo.**

## 5.1. Sugestão de ferramentas

Essa pesquisa inicialmente faz uso da linguagem de programação Python 3.10.12 e da biblioteca SpaCY 3.7.2 para processamento do texto e sua preparação para mapeamento na ontologia e posterior conversão em arquivo XML. No caso do mapeamento semântico as pesquisas analisadas até o momento utilizam a ferramenta GATE, usada para processamento de linguagem natural, desenvolvida e mantida na Universidade de Sheffield. O *software* GATE (*General Architecture for Text Engineering*) é baseado na inserção de rotinas em linguagem JavaScript. O *software* GATE é tradicionalmente aceito pela comunidade acadêmica, mantido pela Universidade de Sheffield (Reino Unido), possui um grupo ativo de usuários, mas possui limites no processamento da língua portuguesa. A lista de ferramentas disponíveis na plataforma GATE Cloud contempla a língua portuguesa apenas para serviços de criação de rótulos baseados na Universal Dependencies (UD)<sup>28</sup>. Depois de algumas tentativas para avaliação da plataforma não foi possível rotular uma sentença do texto de língua portuguesa.

O *software* GATE também possui a versão *Embedding* capaz de ser invocada em códigos baseados em JavaScript, mas as bibliotecas NLTK e SpaCy têm maior capacidade

---

<sup>28</sup> A UD é um framework que reúne diferentes corpora linguísticos rotulados com linguísticos. No caso da língua portuguesa (pt-BR) os bancos da UD ainda são limitados para aplicações especialistas.

de ser incorporada em aplicações com pouca conversão. Já o uso de uma interface de programação (*Application Programming Interface*, API) da plataforma GATE pode criar problemas de interoperabilidade entre ferramentas. Ao utilizar uma linguagem aberta de programação ainda é possível construir um código com variáveis e valores que serão mais facilmente incorporados em outros fluxos de projeto. O processo foi desenvolvido na IDE Google Colab para permitir acesso externo, backup e comparação de alterações. A plataforma é gratuita, e por processar todos os dados na nuvem, não demanda uso recursos computacionais locais.

Foi feita a opção pela adoção da biblioteca SpaCy que possui menos funcionalidades e personalização em comparação à NLTK, mas é uma ferramenta amplamente aceita no mercado porque responde com acurácia, consistência e velocidade as demandas de rotina.

## 5.2. Preparação do texto e elaboração de padrões

O processo de extração da informação tem início com os procedimentos de preparação e consolidação do texto, definição dos grupos de padrões linguísticos e regras de extração tratados Sessão 2.7.2.

Para seleção e validação das técnicas de extração foi feito um experimento simplificado feito sobre a Norma de Desempenho usando apenas os rótulos sintáticos que identificaram o fluxo de informações para o desenvolvimento do processo (Quadro 14).

**Quadro 14 - Experimento simplificado para busca de padrões - código em Python**

```
# Definindo o padrão de busca apenas com base nos rótulos sintáticos.
pattern = [{'POS': 'NOUN'},
           {'POS': 'ADP'}]
from spacy.matcher import Matcher
#Matcher class object
matcher = Matcher(nlp.vocab)
matcher.add("matching_1", patterns = [pattern])
result = matcher(doc, as_spans=True)
print(result)
>> [Método de, cações em, fase de, projeto de, número de, unidade do, andar com,
recintos da, resultados dos, entrada de, paredes de, crítica do...]
```

Fonte: o autor.

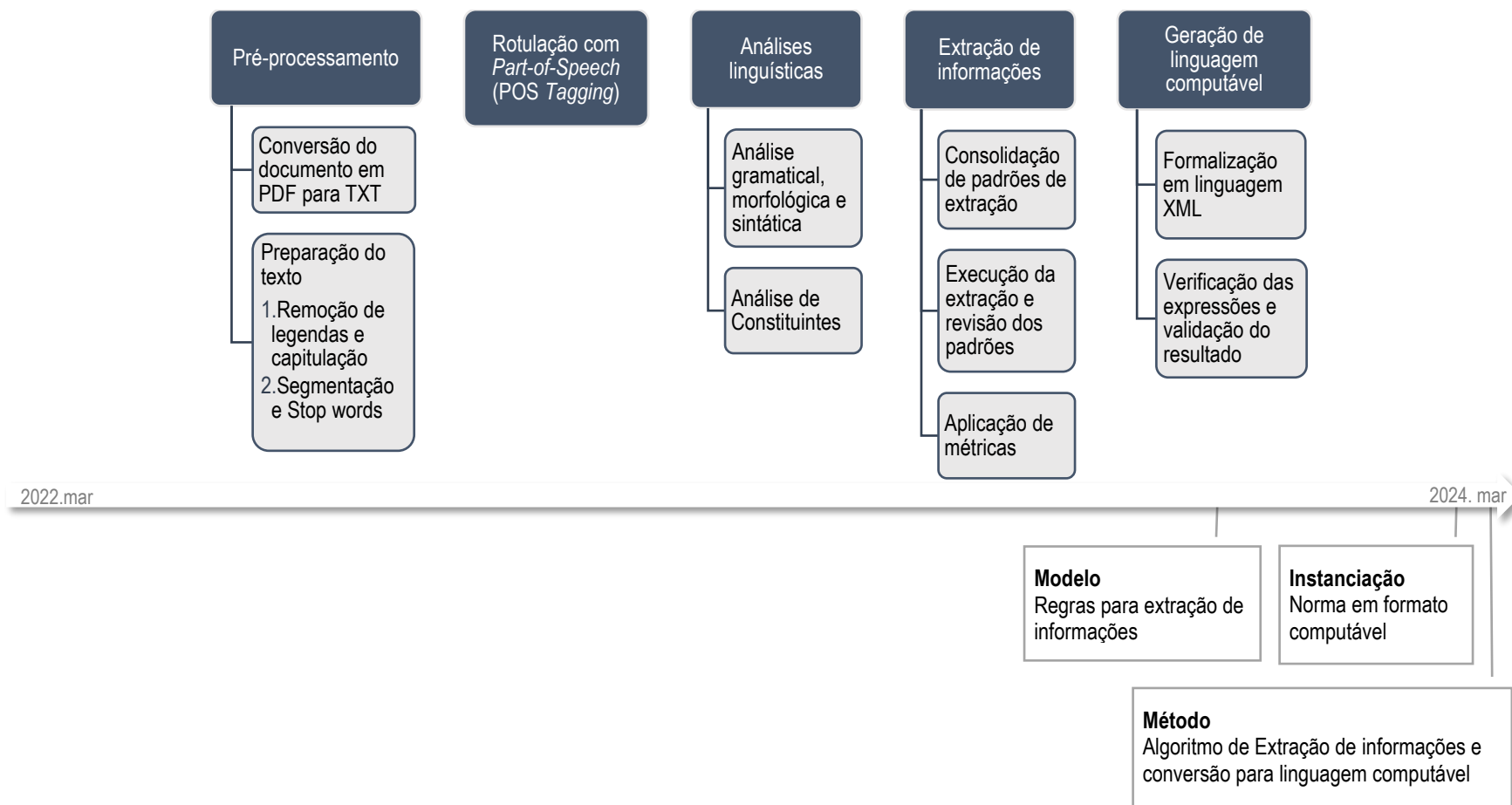
### **5.3. Conversão em linguagem computável**

A pesquisa foi direcionada para a conversão da Norma de Desempenho em linguagem de marcação XML, considerando que é o formato proposto pelo projeto ISO Smart para novos regulamentos e que oferece maior potencial de adesão das ferramentas disponíveis no mercado. Outras ferramentas de verificação automática já desenvolveram soluções que utilizavam a mesma linguagem (Autodesk Model Checker, Solibri, SmartCODES, FORNAX-CorenetX). Além disso as pesquisas anteriores mostraram grande adesão deste formato que pode ser facilmente modificado por meio de esquemas ou novos atributos.

### **5.4. Sugestão de procedimentos para extração de informações**

Com base nos experimentos de validação dos códigos e análise das referências apresentadas no Capítulo 3, foram feitas tentativas de implementação do processo de extração em linguagem Python. A sequência das ações e os artefatos resultantes estão apresentados na Figura 54.

Figura 54 - Etapas propostas para a transformação da Norma de Desempenho em linguagem computável e os artefatos produzidos



Fonte: o autor.

As tentativas já realizadas e que serão usadas como critério de comparação dos artefatos construídos são as medidas-F1 obtidas por Zhang (2015) com valores médios entre 91.1% e 95.6%, Zhang e El-Gohary (2017) com valor de 92.8% e Zheng et al. (2022) que possui uma média de 83.4%.

Como resultados, são esperados os seguintes artefatos:

1. **Modelo:** Conjunto de regras para extração das informações.
2. **Método:** Algoritmo de Extração de Informações e conversão para linguagem computável.

Com o **artefato 1** a pesquisa consegue atingir o objetivo específico da criação de um conjunto de regras de extração de informação da Norma de Desempenho. Juntamente com o **artefato 2**, a pesquisa responde aos objetivos específicos de desenvolvimento de padrões textuais e um algoritmo para reconhecimento de padrões visando extração de informações.

Ao construir o **artefato 2** a pesquisa também executa o objetivo proposto que é a criação de um método de PLN baseado em regras para tradução automática da NBR 15.575. A instanciação do algoritmo conclui o último objetivo específico e valida a execução do objeto da pesquisa ao fornecer um arquivo de um trecho da Norma de Desempenho, capaz de ser computado ou incorporado em ferramentas de verificação automática de conformidade.

## 6. DESENVOLVIMENTO

### 6.1. Seleção e análise do trecho da norma

A seleção do trecho da ND a ser traduzido para linguagem estruturada selecionada baseou-se no capítulo que possuísse amplas características presentes nos demais capítulos, como informações quantitativas, existenciais (que se referem dependência de outros componentes ou condições), menções à normas, tabelas, imagens e tabelas. Selecionamos o Capítulo 8 – Segurança contra incêndio (Apêndice D) da Norma de Desempenho para explorar a opção de automatizar o fluxo de conversão em formato textual com a incorporação de um arquivo PDF e uma segmentação automática do texto.

De modo geral, os textos da norma de Desempenho se estruturam segundo o modelo do Quadro 15, com até quatro níveis de hierarquia de informação.

**Quadro 15 - Estrutura de diretrizes da Norma de Desempenho**

<b>Nível 1</b>	<b>Nível 2</b>	<b>Nível 3</b>	<b>Nível 4</b>
X.1 – Generalidades (explicitação dos requisitos do assunto e objetivos dos requisitos)	X.n – Requisito n (descrição do objetivo do requisito)	X.n.1- Critérios	X.n.1.m (descrição do requisito)
		X.n.2 – Métodos de Avaliação (descrição do método de avaliação)	

Fonte: o autor.

Inicialmente foi feita a tentativa de capturar esta hierarquia rotulando cada sentença segundo o capítulo ou número do capítulo a que pertence, mas foram percebidas ausências do padrão de nomeação dos tópicos evidenciando que a nomeação de tópicos não possui conteúdo padronizado e sofre alterações na redação ao longo da norma. Alguns tópicos são nomeados com termos simples e amplos como “8.7.2 Métodos de avaliação”, e outros são escritos usando termos longos e específicos, conforme o tema relacionado àquele requisito, p. ex. “8.2.2 Métodos de avaliação da segurança relativa ao princípio do incêndio”.

Não haveria prejuízo para leitura e interpretação humana se os tópicos fossem nomeados sem descrição do assunto de modo a permitir a extração de informações por meio do uso do método de Expressões Regulares (Regex) com padrões:

```
\b\d+\.\d+\.1\sCritério\b
```

Onde:

**\b** - representa a fronteira da palavra que deve ser buscada. Inicia no primeiro **\b** e termina no **\b** final da expressão.

**\d+** -busca qualquer dígito composto por 1 ou mais algarismos. Isso se aplica ao primeiro índice relativo ao capítulo, mas também ao segundo índice referente ao grupo de requisitos. Aqui é possível extrair todos os números de capítulos, compostos por um ou mais números (3, 8, 12, 140, 3945, etc.).

**“\1\sCritério”** indica que o elemento Critério deve possuir final 1 seguido de um espaço e a palavra Critério.

Desta forma, o padrão Regex conseguiria extrair textos do tipo "8.7.1 Critério", "6.3.1 Critério", "8.2.1 Critério" e "5.5.1 Critério" que poderiam ser incorporados automaticamente como índices das sentenças.

Apesar desta possibilidade de implementação, a pesquisa adotou a extração manual do texto e sua conversão para uma representação mais simples que contém não apenas os tópicos “Critérios”, mas também os “Métodos de Avaliação” do requisito. A análise inicial indicou a manutenção dos itens do tópico “Método de Avaliação” para processamento. Foi considerado que os textos referentes a este tópico possuem caráter normativo de requisito para conformidade do processo e do projeto. Assim sendo, é de interesse converter para um formato capaz de ser lido por máquina.

A ausência de um padrão uniforme dificultou a extração por palavras-chave como “Critério”, “Requisito”, valores de índices como “8.2” uma vez que seria necessário definir um padrão de busca comum aos valores de tópico, os termos que classificam o tópico (Métodos de avaliação, Requisito e Critério) e eliminasse as palavras restantes até a quebra de linha. O Quadro 16 apresenta exemplos de variabilidade da representação

que dificulta a extração. Em algumas partes do texto o critério é subdividido em listas rotuladas (a, b, c, d...). Também estão presentes sentenças longas, compostas por mais de 85 palavras dificultando a compreensão automática das relações entre os termos.

**Quadro 16 - Trecho da Norma de Desempenho que exemplifica a variabilidade de sua estrutura**

Texto
<p>“11.4.7.2 Características do modelo de referência No modelo de referência, devem-se adotar paredes e pisos, de APP e APT, referentes a um elemento de vedação com 100 mm de espessura, composto por um material com propriedades térmicas de acordo a Tabela 7”.</p> <p>“18.1 Generalidades 18.1.1 Técnicas de avaliação do impacto ambiental resultante das atividades da cadeia produtiva da construção ainda são objeto de pesquisa e, no atual estado da arte, não é possível estabelecer critérios e métodos de avaliação relacionados à expressão desse impacto”.</p> <p>“O projeto deve prever para as áreas comuns e, quando contratado, também para as áreas privativas, as adaptações que normalmente referem-se a:</p> <ul style="list-style-type: none"> <li>a) acessos e instalações;</li> <li>b) substituição de escadas por rampas;</li> <li>c) limitação de declividades e de espaços a percorrer;</li> <li>d) largura de corredores e portas;</li> <li>e) alturas de peças sanitárias;</li> <li>f) disponibilidade de alças e barras de apoio.” <p>“Para edifícios ou conjuntos habitacionais com local de implantação definido, os projetos de arquitetura, da estrutura, das fundações, contenções e outras eventuais obras geotécnicas devem ser desenvolvidos com base nas características do local da obra (topográficas, geológicas etc.), avaliando-se convenientemente os riscos de deslizamentos, enchentes, erosões, vibrações transmitidas por vias férreas, vibrações transmitidas por trabalhos de terraplenagem e compactação do solo, ocorrência de subsidência do solo, presença de crateras em camadas profundas, presença de solos expansíveis ou colapsíveis, presença de camadas profundas deformáveis e outros”.</p> </li></ul>

Fonte: ABNT (2021).

A fim de automatizar o processo, seria necessário dedicar mais esforços para a construção de padrões mais complexos que pudessem contemplar estes casos. Desta forma, foram removidas manualmente as legendas, tabelas, capítulos e números de páginas do trecho selecionado (Capítulo 8 da Norma) mantendo apenas os requisitos de cunho textual. Como mencionado na Seção 2.7.1, as informações textuais não são estruturadas e de difícil conversão em formato computável. Já listas e tabelas estão em formato estruturado ou semiestruturado, portanto não há necessidade de conversão automatizada de seus dados. Apesar de alguns quadros da Norma possuírem conteúdo textual, e não estarem organizados de forma estruturada, a pesquisa optou pela redução

do escopo selecionando apenas dados textuais que apresentam maior dificuldade para conversão.

## 6.2. Pré-processamento

Um pré-processamento padrão de texto envolve a preparação do texto para ser incorporado no código. Um texto comum, extraído de um documento formatado em linguagem natural, precisa ser consolidado em sentenças livres de simbologias, hifens, imagens e formatação. Ao contrário do pré-processamento em língua inglesa, o pré-processamento convencional que remove *stop words* e pontuações, deve ser avaliado com ressalvas na língua portuguesa. Na língua portuguesa algumas expressões perdem sentido quando removidas palavras de conexão como artigos e preposições.

Estas questões foram consideradas, verificadas e apresentadas nas etapas descritas a seguir.

### 6.2.1. Conversão do documento em PDF para TXT

O texto da Norma, que foi usado na pesquisa, foi disponibilizado em formato PDF, portanto foi necessário convertê-lo em sentenças .txt e .csv. Nesta pesquisa foram exploradas ambas as opções, mas para fins de armazenamento no repositório na plataforma de dados abertos Zenodo<sup>29</sup>. A pesquisa optou por utilizar arquivos .txt que podem ser lidos mais facilmente diretamente na plataforma.

Inicialmente foi adotado um programa em linguagem python para fazer a conversão. Foram usadas expressões regulares (Regex) para seleção e segmentação das sentenças. O resultado foi satisfatório, mostra potencial para extração completa de textos, mas com necessidade de ajustes e correções como remoção de espaços. A pesquisa abandonou este procedimento entendendo que no caso de uma aplicação do método

---

<sup>29</sup> <https://zenodo.org/records/11315651>

desenvolvido pela pesquisa o texto normativo seria disponibilizado em formato textual (\*.txt, \*.docx, ou similares).

Foram exploradas diferentes ferramentas para extração do texto a partir de um documento PDF. Dentre elas, a ferramenta Adobe Acrobat<sup>30</sup> obteve maior precisão na extração de sentenças em comparação com o programa criado a partir da biblioteca PyPDF2 na linguagem Python em comparação com outras ferramentas baseadas em OCR (*Optical Character Recognition*).

O texto precisou ser revisado manualmente para remover excesso de espaços vazios e parágrafos duplos. A correção automática de problemas de formatação como estes é possível e de certa forma trivial. Por outro lado, entendendo que em uma implementação oficial da transformação de regulamentos os documentos serão fornecidos em formato textual, não foi dedicado tempo para criação de um programa para correção do texto e optou-se pelo ajuste manual. É importante observar que a extração do arquivo de uma fonte em formato PDF é desnecessária caso o arquivo seja fornecido em formato textual (.rtf, .docx, .odt, .txt).

### **6.2.2. Preparação do texto**

Com o texto em formato .txt, foram eliminadas manualmente as legendas, tabelas, capítulos e números de páginas do trecho selecionado, mantendo apenas os requisitos de natureza textual. Ao contrário de tabelas e índices, as informações textuais não são dados estruturados e são de difícil conversão em formato computacional. As listas e tabelas estão em formato estruturado ou semiestruturado, logo não há necessidade de conversão automatizada de seus dados. Para os casos de tabelas elaboradas para representar critérios quantitativos, a pesquisa sugere a possibilidade de utilizar diversas ferramentas disponíveis e abertas que podem realizar a extração e conversão dos dados para o formato de planilhas (.csv, .xlsx). Embora alguns quadros de dados na Norma contenham

---

<sup>30</sup> <https://www.adobe.com/br/acrobat/online.html>

conteúdo textual e não estejam organizados de forma estruturada, a pesquisa optou por restringir o escopo, selecionando apenas dados textuais que apresentam maior dificuldade para conversão. A preparação do texto envolveu a correção de palavras, a exclusão de espaços, a reescrita de palavras com translineação (troca de linha com hífen), a quebra de linhas e a aglutinação de sentenças com múltiplas linhas e tópicos (Quadro 17).

**Quadro 17 - Exemplo de uma frase da Parte 1 da Norma em formato original e sua versão de único parágrafo**

Formato original	<p>O arquivo climático utilizado deve fornecer valores mensais de temperatura média do solo, em graus Celsius (°C), para todos os meses do ano, além dos seguintes valores horários representativos das 8760 h do ano meteorológico típico:</p> <ul style="list-style-type: none"> <li>a) temperatura de bulbo seco, expressa em graus Celsius (°C);</li> <li>b) temperatura do ponto de orvalho, expressa em graus Celsius (°C);</li> <li>c) umidade relativa, expressa em porcentagem (%);</li> <li>d) pressão atmosférica, expressa em Pascal (Pa);</li> <li>e) intensidade de radiação horizontal de onda longa, expressa em watts-hora por metro quadrado (Wh/m<sup>2</sup>);</li> <li>f) radiação horizontal global, expressa em watts-hora por metro quadrado (Wh/m<sup>2</sup>);</li> <li>g) radiação normal direta, expressa em watts-hora por metro quadrado (Wh/m<sup>2</sup>);</li> <li>h) radiação horizontal difusa, expressa em watts-hora por metro quadrado (Wh/m<sup>2</sup>); i) direção do vento, expressa em graus (°), considerando o sentido horário a partir da direção Norte;</li> <li>j) velocidade do vento, expressa em metros por segundo (m/s)”</li> </ul>
Traduzido para um único parágrafo	<p>O arquivo climático utilizado deve fornecer valores mensais de temperatura média do solo, em graus Celsius (°C), para todos os meses do ano, além dos seguintes valores horários representativos das 8760 h do ano meteorológico típico: a) temperatura de bulbo seco, expressa em graus Celsius (°C); b) temperatura do ponto de orvalho, expressa em graus Celsius (°C); c) umidade relativa, expressa em porcentagem (%); d) pressão atmosférica, expressa em Pascal (Pa); e) intensidade de radiação horizontal de onda longa, expressa em watts-hora por metro quadrado (Wh/m<sup>2</sup>); f) radiação horizontal global, expressa em watts-hora por metro quadrado (Wh/m<sup>2</sup>); g) radiação normal direta, expressa em watts-hora por metro quadrado (Wh/m<sup>2</sup>); h) radiação horizontal difusa, expressa em watts-hora por metro quadrado (Wh/m<sup>2</sup>); i) direção do vento, expressa em graus (°), considerando o sentido horário a partir da direção Norte; j) velocidade do vento, expressa em metros por segundo (m/s).</p>

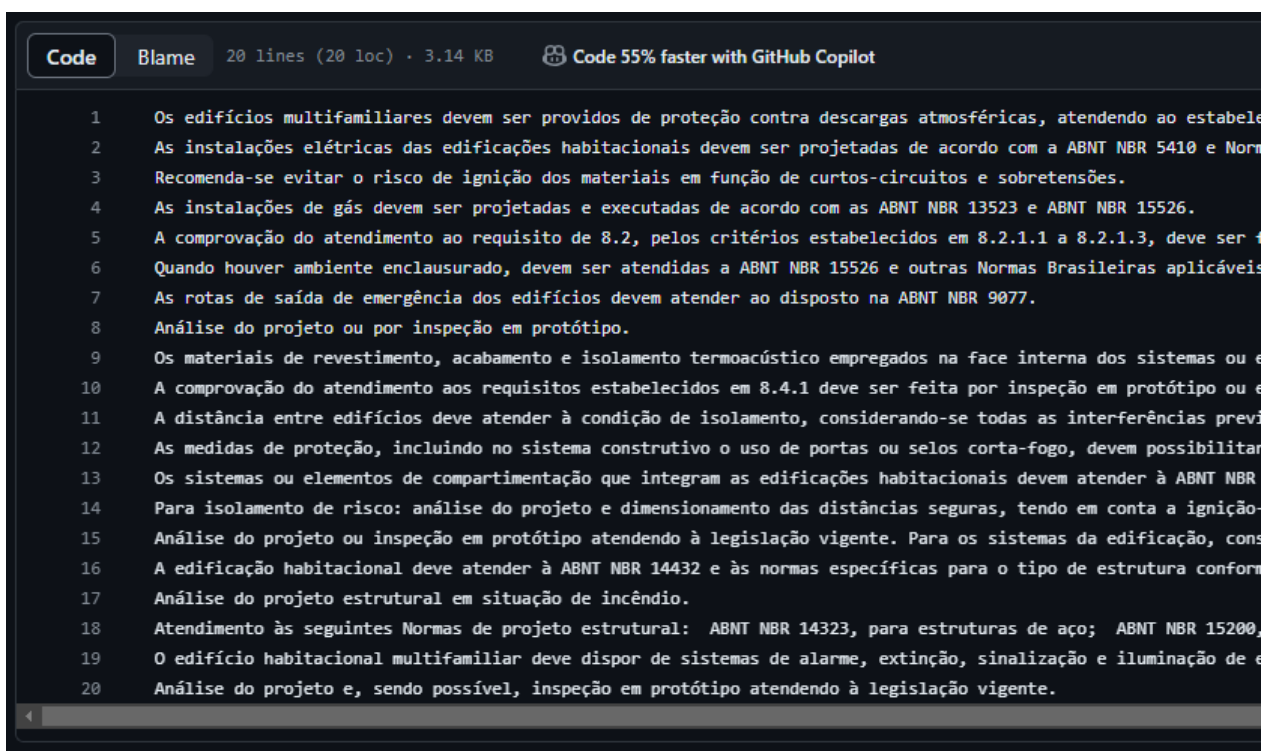
Fonte: o autor.

As sentenças da Norma possuem independência interpretativa, ou seja, o texto de cada requisito reúne, na mesma frase, todos os elementos que devem ser analisados pelos interessados. Dessa forma, as sentenças puderam ser segmentadas a partir do ponto final de cada uma. Esse processo foi automatizado utilizando código Python. O texto final pode ser visualizado no Apêndice E. Os segmentos selecionados foram consolidados como

arquivo .txt e armazenados em um repositório do GitHub para acelerar e simplificar o processo de carregamento dos dados no programa.

Os segmentos selecionados foram consolidados como arquivo .txt e armazenados em um repositório do GitHub para acelerar e simplificar o processo de carregamento dos dados no programa (Figura 55).

**Figura 55 - Exemplo de trecho do documento armazenado em formato .txt na plataforma Github**



```
Code Blame 20 lines (20 loc) · 3.14 KB Code 55% faster with GitHub Copilot
1 Os edifícios multifamiliares devem ser providos de proteção contra descargas atmosféricas, atendendo ao estabelecido
2 As instalações elétricas das edificações habitacionais devem ser projetadas de acordo com a ABNT NBR 5410 e Norm
3 Recomenda-se evitar o risco de ignição dos materiais em função de curtos-circuitos e sobretensões.
4 As instalações de gás devem ser projetadas e executadas de acordo com as ABNT NBR 13523 e ABNT NBR 15526.
5 A comprovação do atendimento ao requisito de 8.2, pelos critérios estabelecidos em 8.2.1.1 a 8.2.1.3, deve ser feita
6 Quando houver ambiente enclausurado, devem ser atendidas a ABNT NBR 15526 e outras Normas Brasileiras aplicáveis
7 As rotas de saída de emergência dos edifícios devem atender ao disposto na ABNT NBR 9077.
8 Análise do projeto ou por inspeção em protótipo.
9 Os materiais de revestimento, acabamento e isolamento termoacústico empregados na face interna dos sistemas ou e
10 A comprovação do atendimento aos requisitos estabelecidos em 8.4.1 deve ser feita por inspeção em protótipo ou e
11 A distância entre edifícios deve atender à condição de isolamento, considerando-se todas as interferências previ
12 As medidas de proteção, incluindo no sistema construtivo o uso de portas ou selos corta-fogo, devem possibilitar
13 Os sistemas ou elementos de compartimentação que integram as edificações habitacionais devem atender à ABNT NBR
14 Para isolamento de risco: análise do projeto e dimensionamento das distâncias seguras, tendo em conta a ignição-
15 Análise do projeto ou inspeção em protótipo atendendo à legislação vigente. Para os sistemas da edificação, cons
16 A edificação habitacional deve atender à ABNT NBR 14432 e às normas específicas para o tipo de estrutura conform
17 Análise do projeto estrutural em situação de incêndio.
18 Atendimento às seguintes Normas de projeto estrutural: ABNT NBR 14323, para estruturas de aço; ABNT NBR 15200,
19 O edifício habitacional multifamiliar deve dispor de sistemas de alarme, extinção, sinalização e iluminação de e
20 Análise do projeto e, sendo possível, inspeção em protótipo atendendo à legislação vigente.
```

Fonte: o autor.

### 6.2.3. Remoção de legendas e capitulação para segmentação do texto

Uma questão que surgiu nos itens da norma, que dificulta a extração automática, é a ausência de um padrão rígido de organização de texto. Alguns dos critérios são precedidos por números seguidos de títulos ou subtítulos da descrição do critério (Quadro 16).

O procedimento de extração do Capítulo 8 implicou na supressão manual das descrições contidas no tópico "Generalidades" e dos objetivos especificados no "Requisito", mantendo-se os demais textos referentes aos tópicos "Critérios" e "Métodos de Avaliação".

Predominantemente, as sentenças da Norma são verbais, seguindo uma estrutura frasal tradicional, com clara presença de sujeito, verbo e predicado. O produto desta fase foi a elaboração do arquivo documentado no Apêndice D.

Após remover os textos desnecessários, o texto foi segmentado em sentenças individuais. Dado que o ponto "." é empregado em simbologias e numerais, isto inviabiliza sua utilização como marcador de segmentação, optou-se pela alternativa de dividir o texto em linhas de planilha, assegurando a individualização e indexação das sentenças. O arquivo foi então convertido e armazenado no formato .CSV para subsequente processamento por meio da linguagem de programação Python. Em etapas subsequentes, foram exploradas alternativas de segmentação em formato .TXT via caractere oculto "\n", responsável pela quebra de linha no arquivo .TXT.

#### **6.2.4. Remoção de *stop words* e pontuação**

A remoção de *stop words* deve ser considerada em casos que as *palavras* não afetam o sentido das palavras (ver Seção 2.7.2). A lista de *stop words* pode ser customizada para preservar determinadas palavras-chave, tais como "ou", "e", "de" e "ser", as quais representam expressões recorrentes como "devem ser", "de forma a", "sistemas ou elementos". O exemplo do Quadro 18 e Quadro 19 evidenciam a problemática inerente à exclusão indiferenciada de *stop words* durante a extração de termos da Norma.

**Quadro 18 - Exemplo com perda significativa de sentido na remoção de *stopwords***

<b>Original</b>	Os edifícios multifamiliares devem ser providos de proteção contra descargas atmosféricas atendendo ao estabelecido na ABNT NBR 5419 e demais Normas Brasileiras aplicáveis nos casos previstos na legislação vigente.
<b>Filtrada</b>	edifícios multifamiliares devem providos proteção contra descargas atmosféricas atendendo estabelecido ABNT NBR 5419 demais Normas Brasileiras aplicáveis casos previstos legislação vigente.
<b>Personalizada</b>	edifícios multifamiliares devem ser providos de proteção contra descargas atmosféricas atendendo ao estabelecido ABNT NBR 5419 e demais Normas Brasileiras aplicáveis casos previstos legislação vigente.

Fonte: o autor.

**Quadro 19- Exemplo de frase com pouca perda de sentido na remoção de *stopwords***

<b>Original</b>	As instalações elétricas das edificações habitacionais devem ser projetadas de acordo com a ABNT NBR 5410 e Normas Brasileiras aplicáveis
<b>Filtrada</b>	instalações elétricas edificações habitacionais devem projetadas acordo ABNT NBR 5410 Normas Brasileiras aplicáveis
<b>Personalizada</b>	instalações elétricas das edificações habitacionais devem ser projetadas de acordo ABNT NBR 5410 e Normas Brasileiras aplicáveis

Fonte: o autor.

Durante o desenvolvimento, observou-se que, no contexto do texto normativo e em documentos regulatórios relacionados à indústria da construção civil, há termos que requerem a inclusão de *stop words*, como "da", "de" e "do", uma vez que essas preposições determinam termos compostos. Por exemplo, a expressão "ponto de orvalho", estruturada com os rótulos gramaticais <NOUN, ADP, NOUN>, deve ser extraída como um único termo e não como termos independentes. O mesmo princípio se aplica a outras construções semelhantes, como "temperatura de bulbo seco", "vida útil de projeto" e "metros por segundo".

Para determinar a abordagem adequada em relação à preservação ou exclusão de *stop words*, foram conduzidas análises de remoção dessas palavras e avaliou-se o impacto dessa remoção na rotulação sintática das sentenças. Considerando que a exclusão completa de *stop words* é desaconselhada devido ao risco de comprometer o reconhecimento de palavras compostas, optou-se por utilizar a lista de *stop words* em português da biblioteca NLTK, com exceção das palavras "e", "por", "ou", "de", "ser", "seja", "dos", "das", "do", "da", "em" e "ao". As sentenças foram classificadas conforme apresentado no Quadro 20 e avaliadas com base no resultado obtido em uma análise de

dependências. O resultado da análise não indicou alteração entre as sentenças originais e as com manutenção de *stop words* personalizadas (Quadro 20).

**Quadro 20 - Casos em que a supressão de *stop words* afetou o resultado**

1	<b>Original</b>	('Normas', 'PROPN', 'conj'), ('Brasileiras', ' <b>PROPN</b> ', 'flat:name'), ('aplicáveis', 'ADJ', 'amod')]
	<b>Filtrada</b>	('Normas', 'PROPN', ' <b>appos</b> '), ('Brasileiras', ' <b>ADJ</b> ', 'flat:name'), ('aplicáveis', 'ADJ', 'amod')]
	<b>Personalizada</b>	('Normas', 'PROPN', 'conj'), ('Brasileiras', ' <b>PROPN</b> ', 'flat:name'), ('aplicáveis', 'ADJ', 'amod')]
2	<b>Original</b>	[(('As', 'DET', 'det'), ('instalações', 'NOUN', 'nsubj:pass'), ('de', 'ADP', 'case'), ('gás', ' <b>NOUN</b> ', 'nmod'), ('devem', 'VERB', 'ROOT'), ('ser', 'AUX', 'aux:pass'), ('projetadas', 'VERB', 'xcomp'), ('e', 'CCONJ', 'cc'), ('executadas', 'VERB', 'conj'))]
	<b>Filtrada</b>	[(('instalações', 'NOUN', 'nsubj'), ('gás', ' <b>ADJ</b> ', 'amod'), ('devem', 'VERB', 'ROOT'), ('projetadas', 'VERB', 'xcomp'), ('executadas', 'VERB', 'acl'))]
	<b>Personalizada</b>	[(('instalações', 'NOUN', 'nsubj:pass'), ('de', 'ADP', 'case'), ('gás', ' <b>NOUN</b> ', 'nmod'), ('devem', 'VERB', 'ROOT'), ('ser', 'AUX', 'aux:pass'), ('projetadas', 'VERB', 'xcomp'), ('e', 'CCONJ', 'cc'), ('executadas', 'VERB', 'conj'))]
3	<b>Original</b>	[(('As', 'DET', 'det', medidas), ('medidas', 'NOUN', 'nsubj', devem), ('de', 'ADP', 'case', proteção), ('proteção', 'NOUN', 'nmod', medidas))]
	<b>Filtrada</b>	[(('medidas', 'NOUN', 'nsubj', proteção), ('proteção', 'NOUN', 'ROOT', proteção))]
	<b>Personalizada</b>	[(('medidas', 'NOUN', 'nsubj', devem), ('de', 'ADP', 'case', proteção), ('proteção', 'NOUN', 'nmod', medidas))]

Fonte: o autor.

Após analisar os resultados, constatou-se que a lista personalizada não apresenta contribuição significativa, pois não altera de maneira relevante os rótulos linguísticos. Portanto, para simplificar o processamento e o código utilizado, optou-se por manter as *stop words* na análise das sentenças para a construção dos padrões de extração de informação. É importante notar que este processo difere das pesquisas internacionais, baseadas em regulamentos de língua inglesa.

Um procedimento análogo foi executado para a remoção de pontuação, tendo em vista que apenas alguns símbolos poderiam ser eliminados sem comprometer o sentido e a informação do texto, como o ponto final, que indica a separação de valores, capítulos e é parte das unidades de medida. A análise da exclusão da pontuação, considerando os atributos linguísticos das sentenças com e sem pontuação, resultou em uma conclusão similar à remoção de *stop words*.

### 6.3. Execução da extração de informações

#### 6.3.1. Rotulagem de atributos linguísticos

A atribuição de rótulos aos termos do texto pré-processado da Norma (Seção 6.2), juntamente com seus atributos linguísticos, foi realizada com base nos atributos gramaticais, morfológicos e semânticos (Seção 2.7.2). A análise gramatical, também conhecida como rotulagem de parte do discurso (*Part-of-Speech* ou *POS Tagging*), foi repetida nesta etapa, ampliando-se os rótulos dos *tokens* e organizando o conteúdo no formato de dados ConLL-U (Buchholz; Marsi, 2006), que é um padrão para análise de dados textuais (Quadro 21). Isso possibilitou a análise das relações entre os elementos da frase considerando cinco parâmetros: o lema ou forma raiz da palavra, a função gramatical da palavra na frase, a função sintática da palavra, o pai sintático da palavra (ou seja, a palavra à qual ela se refere na frase) e os atributos morfológicos da palavra.

Quadro 21 - Trecho da formatação ConLL-U usada para extração de padrões

Token	Lemma	POS	Sintática	Head	Morph
edifícios	edifício	NOUN	nsubj:pass	devem	Gender=Masc Number=Plur
multifamiliares	multifamiliar	ADJ	amod	edifícios	Gender=Masc Number=Plur
devem	dever	VERB	ROOT	devem	Mood=Ind Number=Plur Person=3  Tense=Pres VerbForm=Fin
ser	ser	AUX	aux:pass	providos	VerbForm=Inf
providos	provir	VERB	xcomp	devem	Gender=Masc Number=Plur  VerbForm=Part Voice=Pass
de	de	ADP	case	proteção	
proteção	proteção	NOUN	obj	providos	Gender=Fem Number=Sing
contra	contra	ADP	case	descargas	
descargas	descarga	NOUN	nmod	proteção	Gender=Fem Number=Plur
atmosféricas	atmosférico	ADJ	amod	descargas	Gender=Fem Number=Plur

Fonte: o autor.

A anotação foi realizada com o modelo de linguagem integrado à biblioteca *Spacy*, seguindo um método baseado em um modelo linguístico pré-treinado, portanto, probabilístico. Como em todo processo de classificação, a rotulagem está sujeita a erros em sentenças complexas ou com palavras desconhecidas. Para mitigar esse problema, foi tentado extrair valores de confiança do resultado da análise do programa. Entretanto, desde a versão 3.0, essa métrica não retorna resultados pelo método disponível na

biblioteca (pans.attrs["scores"])). Assim, foi realizada uma análise aleatória de algumas sentenças para validar a rotulagem feita pela biblioteca *Spacy*. Após avaliação de diversas sentenças foi percebido que as sentenças recebiam rótulos diferentes conforme o modelo usado. O Quadro 22 mostra um exemplo de frase com rótulos linguísticos para cada um dos modelos, pequeno (SM), médio (MD) e grande (LG). A coluna da direita mostra a classificação correta e deixa evidente que nenhum dos três modelos possui garantia de rotulagem correta.

**Quadro 22 - Comparação entre rotulagem em cada modelo de linguagem**

Token	SM		MD		LG		Correto
<i>text</i>	<i>pos</i>	<i>dep</i>	<i>pos</i>	<i>dep</i>	<i>pos</i>	<i>dep</i>	
Os	DET	det	DET	det	DET	det	
edifícios	NOUN	nsubj:pass	NOUN	nsubj:pass	NOUN	nsubj:pass	
multifamiliares	ADJ	amod	ADJ	amod	ADJ	amod	
devem	VERB	ROOT	VERB	ROOT	VERB	ROOT	
ser	AUX	aux:pass	AUX	aux:pass	AUX	aux:pass	
providos	VERB	xcomp	VERB	xcomp	VERB	xcomp	
de	ADP	case	ADP	case	ADP	case	
proteção	NOUN	obl	NOUN	obl	NOUN	obj	
contra	ADP	case	ADP	case	ADP	case	
descargas	NOUN	nmod	NOUN	nmod	NOUN	nmod	
atmosféricas	ADJ	amod	ADJ	amod	ADJ	amod	
,	PUNCT	punct	PUNCT	punct	PUNCT	punct	
atendendo	VERB	advcl	VERB	advcl	VERB	advcl	
ao	ADP	case	ADP	case	ADP	case	
estabelecido	NOUN	obl	NOUN	obl	VERB	obl	ADJ
na	ADP	case	ADP	case	ADP	case	
ABNT	NOUN	obl	PROPN	obl	PROPN	obl	PROPN
NBR	PROPN	obj	PROPN	obj	PROPN	appos	
5419	NUM	nummod	NUM	dep	NUM	nummod	
e	CCONJ	cc	CCONJ	cc	CCONJ	cc	
demais	DET	conj	DET	det	DET	det	
Normas	PROPN	ROOT	NOUN	dep	NOUN	conj	NOUN
Brasileiras	PROPN	flat:name	ADJ	flat:name	PROPN	flat:name	ADJ
aplicáveis	ADJ	amod	ADJ	amod	ADJ	amod	
,	PUNCT	punct	PUNCT	punct	PUNCT	punct	
nos	ADP	case	ADP	case	ADP	case	
casos	NOUN	nmod	NOUN	nmod	NOUN	nmod	
previstos	VERB	acl	VERB	acl	VERB	acl	
na	ADP	case	ADP	case	ADP	case	
legislação	NOUN	obl	NOUN	obl	NOUN	obl	
vigente	ADJ	amod	ADJ	amod	ADJ	amod	
.	PUNCT	punct	PUNCT	punct	PUNCT	punct	

Fonte: o autor.

O verbo estabelecer, por exemplo, possui flexão no particípio passado do verbo e isso lhe permite funcionar como adjetivo na frase, i.e., “atendendo [ao que foi] estabelecido”. Na frase completa, sem ocultação dos termos “ao que foi”, “foi” adota a função de verbo copulativo e transforma a frase em frase predicativa e o verbo “estabelecido”, um adjetivo.

Os modelos de linguagem desempenham um papel crucial na etapa de análise e rotulagem na biblioteca *SpaCy*, e possuem características distintas comparadas no Apêndice G. Contrariamente à lógica de que modelos maiores, compostos com mais atributos e dimensões, seriam superiores, constatou-se que o modelo de médio tamanho obteve melhores resultados na comparação. É importante ressaltar que, apesar da variação nos resultados, observou-se que os modelos mantêm consistência na análise, ou seja, se a mesma frase for analisada em outros códigos ou fluxos diferentes do projeto, seus rótulos linguísticos serão os mesmos.

### **6.3.2. Consolidação de padrões de extração**

A construção de padrões com base em atributos linguísticos foi realizada manualmente, utilizando os dados gerados na etapa anterior. A planilha em formato ConLL-U possibilitou a identificação de grupos de padrões em diferentes categorias de elementos (Figura 56).

Figura 56 - Trecho da planilha usada para desenvolvimento de padrões

das	de o	ADP	case	edificações	Definite=Def  Gender=Fem  Nu		
edificações	edificação	NOUN	nmod	instalações	Gender=Fem  Number=Plur		Refer. à Normas
habitacionais	habitacional	ADJ	amod	edificações	Gender=Fem  Number=Plur		
devem	dever	VERB	ROOT	devem	Mood=Ind  Number=Plur  Perso		Exigências
ser	ser	AUX	aux:pass	projetadas	VerbForm=Inf		
projetadas	projetar	VERB	xcomp	devem	Gender=Fem  Number=Plur  Ve		Pontuação
de	de	ADP	case	acordo			
acordo	acordo	NOUN	obl	projetadas	Gender=Masc  Number=Sing		Componente
com	com	ADP	case	ABNT			
a	o	DET	det	ABNT	Definite=Def  Gender=Fem  Nu		Atributo do Elemento
ABNT	ABNT	PROPN	nmod	acordo	Gender=Fem  Number=Sing		
NBR	NBR	ADP	appos	ABNT			Método de avaliação
5410	5410	NUM	nummod	aplicáveis	NumType=Card		
e	e	CCONJ	cc	Normas			
Normas	Normas	PROPN	conj	5410	Gender=Fem  Number=Plur		
Brasileiras	Brasileiras	PROPN	flat:name	Normas	Number=Sing		
aplicáveis	aplicável	ADJ	amod	ABNT	Gender=Fem  Number=Plur		
.	.	PUNCT	punct	projetadas			
Recomenda-se	Recomenda-se	VERB	ROOT	Recomenda-se	Case=Acc  Mood=Ind  Number=		
evitar	evitar	VERB	xcomp	Recomenda-se	VerbForm=Inf		
o	o	DET	det	risco	Definite=Def  Gender=Masc  Ni		
risco	risco	NOUN	obj	evitar	Gender=Masc  Number=Sing		
de	de	ADP	case	ignição			
ignição	ignição	NOUN	nmod	risco	Gender=Fem  Number=Sing		
dos	de o	ADP	case	materiais	Definite=Def  Gender=Masc  Ni		
materiais	material	NOUN	nmod	ignição	Gender=Masc  Number=Plur		
em	em	ADP	case	função			
função	função	NOUN	obl	evitar	Gender=Fem  Number=Sing		
de	de	ADP	case	curtos-circuitos			
curtos-circuitos	curtos-circuito	NOUN	nmod	função	Gender=Masc  Number=Plur		
e	e	CCONJ	cc	sobretensões			
sobretensões	sobretensão	NOUN	conj	curtos-circuito:	Gender=Fem  Number=Plur		
.	.	PUNCT	punct	Recomenda-se			
As	o	DET	det	instalações	Definite=Def  Gender=Fem  Nu		
instalações	instalação	NOUN	nsubj:pass	devem	Gender=Fem  Number=Plur		
de	de	ADP	case	gás			
gás	gás	NOUN	nmod	instalações	Gender=Masc  Number=Sing		
devem	dever	VERB	ROOT	devem	Mood=Ind  Number=Plur  Perso		
ser	ser	AUX	aux:pass	projetadas	VerbForm=Inf		
projetadas	projetar	VERB	xcomp	devem	Gender=Fem  Number=Plur  Ve		
e	e	CCONJ	cc	executadas			
executadas	executar	VERB	conj	projetadas	Gender=Fem  Number=Plur  Ve		
de	de	ADP	case	acordo			

Fonte: o autor.

No início, foram coletados padrões para cada *token*, os quais, agrupados de acordo com as categorias de elementos, possibilitaram o reconhecimento de 16 padrões baseados em combinações gramaticais. Uma parte desses padrões está apresentada no Quadro 23, enquanto a lista completa está disponível no Apêndice H. Esses padrões foram fundamentados nos rótulos semânticos identificados em outras pesquisas mencionadas neste trabalho (**Erro! Fonte de referência não encontrada.**).

**Quadro 23 – Trecho da planilha de elementos relacionados às referências de normas e componentes do edifício**

Termos CAP8	Atributos linguísticos		
	Padrão POS	Dependência	Morfologia
REFERÊNCIAS			
ABNT NBR xxxx	{PROPN}{PROPN}{NUM}	{obl}{appos}{nummod}	{Gender=Fem Number=Sing}{NumType=Card}
ABNT NBR xxxxx	{PROPN}{ADP}{NUM}	{nmod}{appos}{nummod}	{Gender=Fem Number=Sing}{NumType=Card}
ABNT NBR xxxxx	{PROPN}{ADP}{NUM}	{nmod}{appos}{nummod}	{Gender=Fem Number=Sing}{NumType=Card}
ABNT NBR xxxxx	{PROPN}{ADP}{NUM}	{conj}{appos}{flat:name}	{Gender=Fem Number=Sing}{NumType=Card}
requisito de 8.2	{NOUN}{ADP}{NUM}	{nmod}{case}{nmod}	{Gender=Masc Number=Sing}{NumType=Card}
ABNT NBR xxxxx	{PROPN}{ADP}{NUM}	{nsubj}{case}{obl}	{Gender=Fem Number=Sing}{NumType=Card}
ABNT NBR xxxxx-x	{PROPN}{ADP}{NUM}	{obl}{case}{obl}	{Gender=Fem Number=Sing}{NumType=Card}
ABNT NBR xxxxx-x	{PROPN}{ADP}{NUM}	{conj}{ROOT}{nummod}	{Gender=Fem Number=Sing}{NumType=Card}
ABNT NBR xxxxx	{PROPN}{ADP}{NUM}	{obl}{appos}{nummod}	{Gender=Fem Number=Sing}{NumType=Card}
ABNT NBR xxxxx	{PROPN}{ADP}{NUM}	{nmod}{case}{ROOT}	{Gender=Fem Number=Sing}{NumType=Card}
COMPONENTE			
edifício multifamiliar	{NOUN}{ADJ}	{nsubj:pass}{amod}	
instalação elétrico	{NOUN}{ADJ}	{nsubj:pass}{amod}	
edificação habitacional	{NOUN}{ADJ}	{nmod}{amod}	
instalação de gás	{NOUN}{ADP}{NOUN}	{nsubj:pass}{case}{nmod}	
ambiente	{NOUN}	{obj}	
saída de emergência	{NOUN}{ADP}{NOUN}	{nmod}{case}{nmod}	
estrutura de aço	{NOUN}{ADP}{NOUN}	{nmod}{case}{nmod}	
estrutura de concreto	{NOUN}{ADP}{NOUN}	{nmod}{case}{nmod}	

Fonte: o autor.

Observa-se no quadro acima que o termo "requisito de 8.2" compartilha os mesmos atributos (rótulos) linguísticos de outro termo como "ABNT NBR xx.xxx", diferenciando-se apenas pelo atributo morfológico. Diante disso, o desenvolvimento passou a considerar uma composição de padrões não apenas pelos atributos gramaticais, como foi feito em todas as pesquisas relacionadas neste trabalho (Thessen; Cui; Mozzherin, 2012; Zhang, 2015; Zhang; El-Gohary, 2011 a 2017; Zhou; El-Gohary, 2017b; Xu; Cai, 2019 e 2021; Choi; Nguyen; Kerr, 2021 e Zhou *et al.*, 2022). A incorporação dos três grupos de atributos linguísticos (gramatical, sintático e morfológico) permitiu uma identificação mais precisa dos termos.

Na sequência, os padrões foram organizados como elementos de dicionários, sendo que cada dicionário é identificado por um rótulo específico. Desse modo, na implementação em código, realizou-se a correspondência do texto com o padrão e seu

respectivo rótulo. Como pode ser observado no Quadro 24, não foram identificadas duplicidades de padrões.

**Quadro 24 - Lista de rótulos e padrões aplicados na Parte 1 da Norma de Desempenho**

Rótulo	Padrão baseado em atributos gramaticais, sintáticos e morfológicos
REFERENCIA	<pre> [{"LOWER": "normas"}, {"POS": "ADJ"}], [{"LOWER": "normas"}, {"LOWER": "específicas"}], [{"LOWER": "normas"}, {"LOWER": "brasileiras"}], [{"LOWER": "legislação"}, {"LOWER": "vigente"}], [{"LOWER": "abnt"}, {"LOWER": "nbr"}, {"POS": "NUM"}], [{"POS": "PROPN", "MORPH": {"IS_SUPERSET": ["Gender=Fem"]}], {"POS": "PROPN"}, {"POS": "NUM"}], [{"POS": "PROPN", "MORPH": {"IS_SUPERSET": ["Gender=Fem"]}], {"POS": "ADP"}, {"POS": "NUM"}], [{"POS": "PROPN", "MORPH": {"IS_SUPERSET": ["Gender=Fem"]}], {"POS": "ADP"}, {"POS": "NUM"}], [{"POS": "NOUN", "MORPH": {"IS_SUPERSET": ["Gender=Fem"]}], {"POS": "PROPN"}, {"POS": "ADJ"}], [{"POS": "PROPN", "DEP": "obj"}, {"POS": "PROPN"}] </pre>
EXIGENCIA	<pre> [{"POS": "VERB"}, {"POS": "AUX"}, {"POS": "VERB"}, {"POS": "CCONJ"}, {"POS": "VERB"}], [{"POS": "VERB"}, {"POS": "AUX"}, {"POS": "CCONJ"}, {"POS": "VERB"}], [{"POS": "VERB"}, {"POS": "ADV", "OP": "?"}, {"POS": "AUX"}, {"POS": "VERB"}], [{"POS": "VERB"}, {"POS": "VERB"}, {"POS": "ADP"}], [{"LOWER": "não", "OP": "?"}, {"POS": "VERB"}, {"POS": "VERB"}], [{"LOWER": "não", "OP": "?"}, {"POS": "VERB"}, {"POS": "VERB"}], [{"LOWER": "não", "OP": "?"}, {"POS": "VERB"}, {"POS": "AUX"}, {"POS": "ADJ"}] </pre>
COMPONENTE	<pre> [{"POS": "NOUN"}, {"POS": "ADP"}, {"POS": "NOUN"}], [{"POS": "NOUN"}, {"POS": "ADP"}, {"POS": "NOUN"}, {"POS": "ADJ"}], [{"POS": "NOUN"}, {"POS": "ADP"}, {"POS": "ADJ"}], [{"POS": "NOUN", "DEP": "nsubj"}, {"POS": "ADJ"}, {"POS": "ADJ"}], [{"POS": "NOUN"}, {"POS": "ADP"}, {"POS": "NOUN"}, {"POS": "ADP"}, {"POS": "NOUN"}], #modif [{"POS": "NOUN"}, {"POS": "ADP"}, {"POS": "NOUN"}, {"POS": "ADP"}, {"POS": "NOUN"}, {"POS": "VERB"}, {"POS": "PUNCT"}], [{"POS": "NOUN", "DEP": "nsubj"}], [{"POS": "NOUN"}, {"POS": "ADJ"}] </pre>
PARAMETRO	<pre> [{"POS": "NOUN"}, {"POS": "ADV"}, {"POS": "ADP"}, {"POS": "NOUN"}], [{"POS": "NOUN", "DEP": "obj"}, {"POS": "ADP"}, {"POS": "NOUN"}] </pre>
VALOR	<pre> [{"POS": "NUM"}, {"POS": "NOUN"}], [{"POS": "NUM"}, {"POS": "SYM"}] </pre>

Fonte: o autor.

Segundo os padrões acima, quando existir uma estrutura de texto com os atributos gramaticais <substantivo, advérbio e substantivo>, será incorporado a este conjunto de texto o rótulo “Componente”, indicando que identificará naquele texto um componente construtivo do edifício. Embora o diagrama apresentado na Figura 26 indique que os padrões são incorporados como dados externos, neste experimento os padrões foram inseridos dentro do código. Contudo, vislumbrando a possibilidade de auditoria e construção dos padrões por especialistas externos, o código foi construído considerando

uma entrada de dados armazenados em um repositório externo e facilmente incorporados ao fluxo do projeto por meio de um arquivo JSON.

### 6.3.3. Avaliação da incorporação de dicionário toponímico

Assim como ocorre com a ontologia, o processo de extração de informação pode se beneficiar de um dicionário toponímico (*gazetteer*) para buscar termos que se repetem ou são importantes para a busca. No entanto, a inserção de um dicionário para buscar termos modais mostrou-se complexa devido à quantidade e ao comportamento desses termos na língua portuguesa (Quadro 25). Ao contrário dos termos em inglês, os verbos modais em português podem ser aplicados em regulamentos com diferentes flexões e combinações de palavras. Por exemplo, a expressão "é desejável que", deveria ser reconhecida como um elemento modal, assim como as expressões "como desejado" e "deseja-se". Portanto, entendemos que não é viável elaborar uma lista simplificada de termos para compor este dicionário e a pesquisa deve ser conduzida sem o uso de um dicionário toponímico.

**Quadro 25 - Verbos modais em português e inglês**

português		inglês
querer	chegar	<i>will</i>
dever	ousar	<i>shall</i>
poder	atrever	<i>could</i>
conseguir	pôr-se	<i>may</i>
pretender	desejar	<i>might</i>
tentar	necessitar	<i>must</i>
buscar	haver	<i>would</i>
parecer	saber	<i>should</i>
crer	ter	<i>can</i>

Fonte: o autor.

### 6.3.4. Extração dos termos e revisão dos padrões

Após preparar o texto em formato .txt e identificar os padrões linguísticos, inicia-se a correspondência pelos padrões. O código completo do algoritmo está escrito em Python 3.10.12 e SpaCy 3.7.2. Para a execução da busca foi usado o método *Matcher* da biblioteca *SpaCy* (Quadro 26).

**Quadro 26 - Pseudocódigo do algoritmo para a correspondência de padrões**

```

INÍCIO
Texto <- texto já processado
RESULTADOS = 0
lista_padroes <- Conjunto[i] com sub_conjunto[j](patterns) de padrões[k] agrupados
pelo rótulo do conjunto (label)
[
  {"rotulo": "REFERENCIA", "pattern": [...]},
  {"rotulo": "EXIGENCIA", "pattern": [...]},
  {"rotulo": "COMPONENTE", "pattern": [...]},
  {"rotulo": "PARAMETRO", "pattern": [...]},
  {"rotulo": "VALOR", "pattern": [...]}
]
PARA cada sentença no texto:
  Executa tokenização e rotulagem automática
  Aciona a busca dos padrões
  PARA cada conjunto [i] da lista_padroes:
    Armazena o rótulo de [i]
    Aciona motor de busca
    PARA cada padrão[k] existente em [j]:
      Adiciona ao motor de busca os padrões[k] em ordem de tamanho e seus
respectivos rótulos
    PARA cada padrão [k]:
      Motor de busca procura os tokens correspondentes ao padrão[k]
      SE algum todos tokens foram processados
      ATUALIZA adicionando os tokens encontrados pelo padrão ao conjunto de
tokens já processados
      Converte os tokens do padrão em um objeto-span
      Adicionar os tokens com seu respectivo rótulo a RESULTADOS
  RETORNA RESULTADOS
FIM

```

Fonte: o autor.

A estrutura de código acima é a lógica final, fruto da exploração e otimização da busca dos padrões. A construção desta lógica, assim como dos padrões foi um processo iterativo. Foram feitas diversas tentativas para ajustar os padrões apresentados no Quadro 24, abandonando alguns e selecionando novos atributos linguísticos que não foram percebidos na primeira análise (Figura 56). Mesmo com as modificações, a quantidade de itens e atributos alterados foi pequena.

Ao executar o algoritmo, notou-se que o método de busca analisa os padrões sequencialmente, por exemplo, buscando os padrões de rótulo “REFERENCIA” e, somente após analisar todos os padrões desse rótulo, inicia a busca pelos padrões de

“EXIGENCIA”. Além disso, a busca estava sendo executada em todo texto para cada padrão, resultando na repetição de palavras em diferentes padrões (Quadro 27).

**Quadro 27 - Problema inicial da busca por padrões**

<p><b>Frase:</b> Os edifícios multifamiliares devem ser providos de proteção contra descargas atmosféricas, atendendo ao estabelecido na ABNT NBR 5419 e demais Normas Brasileiras aplicáveis, nos casos previstos na legislação vigente.</p>
<p>edifícios multifamiliares (COMPONENTE)  devem (EXIGENCIA)  devem ser providos (EXIGENCIA)  providos (EXIGENCIA)  providos de proteção (EXIGENCIA)  proteção contra descargas (COMPONENTE)  descargas atmosféricas (COMPONENTE)  atendendo (EXIGENCIA)  atendendo ao estabelecido (EXIGENCIA)  ABNT NBR 5419 (REFERENCIA)  Normas Brasileiras (COMPONENTE)  previstos (EXIGENCIA)  previstos na legislação (EXIGENCIA)  legislação vigente (COMPONENTE)</p>

Fonte: o autor.

Foi observado que a busca pela correspondência de padrões ocorre mesmo que outros padrões já tenham encontrado as expressões corretas. A solução encontrada foi inserir um método de busca que considera inicialmente os padrões com maior número de atributos. Desta forma, expressões mais complexas que dependem de uma sequência de termos, como “proteção contra descargas” seriam identificadas antes de identificar “proteção”.

A fim de evitar correspondências da mesma palavra em diferentes padrões, o código desenvolvido incorporou uma regra que executava a busca apenas nos *tokens* que não haviam ainda sido classificados por algum padrão.

Analisando um conjunto inicial de frases foi verificado que havia diferenças de rotulagem para a classificação dos termos “Normas Brasileiras” e “ABNT NBR”. Como a expressão possui pouca variação nos textos, optou-se por uma busca explícita do termo utilizando expressões regulares (*regex*) aliada à buscas por padrões linguísticos (Quadro 28).

**Quadro 28 - Exemplos de padrões compostos por expressões regulares**

```

[{"LOWER": "normas"}, {"POS": "ADJ"}],
[{"LOWER": "normas"}, {"LOWER": "específicas"}],
[{"LOWER": "normas"}, {"LOWER": "brasileiras"}],
[{"LOWER": "legislação"}, {"LOWER": "vigente"}],
[{"LOWER": "abnt"}, {"LOWER": "nbr"}, {"POS": "NUM"}],
[{"POS": "PROPN", "MORPH": {"IS_SUPERSET": ["Gender=Fem"]}}, {"POS": "PROPN"}, {"POS": "NUM"}],
[{"POS": "PROPN", "MORPH": {"IS_SUPERSET": ["Gender=Fem"]}}, {"POS": "ADP"}, {"POS": "NUM"}],
[{"POS": "PROPN", "MORPH": {"IS_SUPERSET": ["Gender=Fem"]}}, {"POS": "ADP"}, {"POS": "NUM"}],
[{"POS": "NOUN", "MORPH": {"IS_SUPERSET": ["Gender=Fem"]}}, {"POS": "PROPN"}, {"POS": "ADJ"}],
[{"POS": "PROPN", "DEP": "obj"}, {"POS": "PROPN"}]

```

Fonte: o autor.

Após a classificação de todos os termos, foi identificado um problema relacionado à precisão da classificação. Foram observadas rotulagens incorretas, especialmente em relação aos termos "COMPONENTE" e "PARAMETRO". Esse cenário era esperado devido à similaridade morfológica entre ambos e à semelhança dos parâmetros. Devido à implementação de uma função de busca sequencial no código, que prioriza a correspondência com padrões classificando como "COMPONENTE" antes de verificar se corresponde a "PARAMETRO", a correspondência por termos do tipo "PARAMETRO" tornou-se mais desafiadora.

Nesta etapa, o Capítulo da Norma de Desempenho foi utilizado para construir o código, formular e ajustar os padrões. Ao todo, foram processadas 20 frases, resultando na marcação de 94 termos ou objetos semânticos (Zhang, 2015). Esta fase foi concluída com um total de 25 padrões linguísticos. A análise deste resultado observou que o trecho escolhido para a construção dos padrões não foi suficiente para registrar a variedade de expressões do texto. Logo uma análise prévia da aplicação na Parte 1 permitiu ampliar a lista de padrões e ajustar alguns para serem mais generalistas e capturarem padrões próximos.

Um exemplo de ajuste necessário foi a alteração dos padrões para capturar as expressões “deve ser” e “deve ainda ser”. As expressões de proibição como “não podem apresentar” e “não podem prejudicar” não existiam nos textos do Capítulo 8 da Norma,

obrigando a incorporação dos padrões [{"LOWER": "não", "OP": "?"}, {"POS": "VERB"}, {"POS": "VERB"}].

O resultado da rotulagem foi armazenado em planilha (Quadro 29) para avaliação do resultado que será apresentado no Capítulo 7.

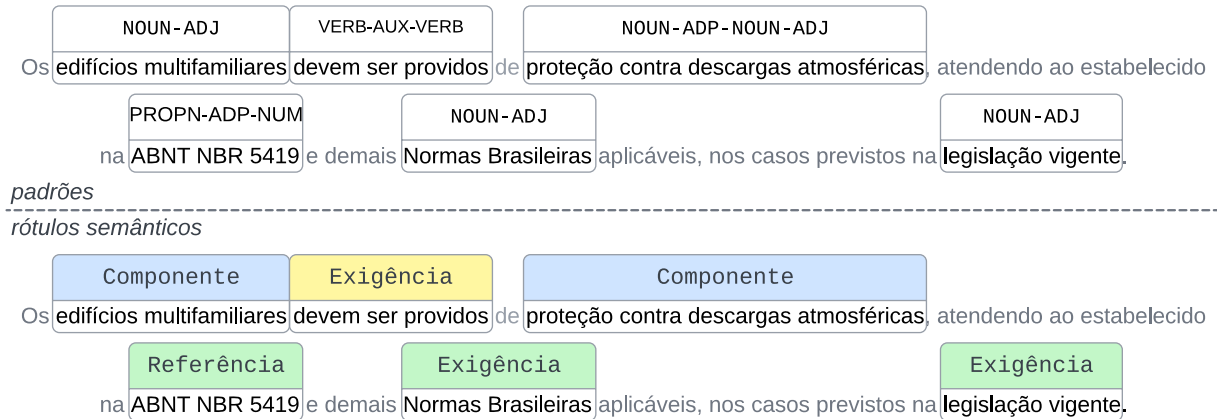
**Quadro 29 - Trecho da planilha com resultado da extração do Capítulo 8 da Norma de Desempenho**

<b>Frase</b>	<b>Objeto Semântico</b>	<b>Texto</b>	<b>Rótulo</b>
Frase 1	Objeto Semântico 1	ABNT NBR 5419	REFERENCIA
Frase 1	Objeto Semântico 2	Normas Brasileiras	REFERENCIA
Frase 1	Objeto Semântico 3	legislação vigente	REFERENCIA
Frase 1	Objeto Semântico 4	proteção contra descargas atmosféricas	COMPONENTE
Frase 1	Objeto Semântico 5	edifícios multifamiliares	COMPONENTE
Frase 1	Objeto Semântico 6	devem ser providos	EXIGENCIA
Frase 2	Objeto Semântico 1	ABNT NBR 5410	REFERENCIA
Frase 2	Objeto Semântico 2	Normas Brasileiras	REFERENCIA
Frase 2	Objeto Semântico 3	instalações elétricas	COMPONENTE
Frase 2	Objeto Semântico 4	edificações habitacionais	COMPONENTE
Frase 2	Objeto Semântico 5	devem ser projetadas	EXIGENCIA
Frase 3	Objeto Semântico 1	risco de ignição dos materiais	COMPONENTE
Frase 3	Objeto Semântico 2	função de curtos-circuitos	COMPONENTE
Frase 3	Objeto Semântico 3	Recomenda-se evitar	EXIGENCIA
Frase 4	Objeto Semântico 1	ABNT NBR 13523	REFERENCIA
Frase 4	Objeto Semântico 2	ABNT NBR 15526	REFERENCIA
Frase 4	Objeto Semântico 3	instalações de gás	COMPONENTE
Frase 4	Objeto Semântico 4	devem ser projetadas e executadas	EXIGENCIA
Frase 5	Objeto Semântico 1	comprovação do atendimento ao requisito	COMPONENTE
Frase 5	Objeto Semântico 2	análise do projeto	COMPONENTE
Frase 5	Objeto Semântico 3	inspeção em protótipo	COMPONENTE
Frase 5	Objeto Semântico 4	deve ser feita	EXIGENCIA

Fonte: o autor.

A Figura 57 sintetiza o processo de conversão dos padrões em rótulos pré-definidos. A imagem evidencia quais termos não serão extraídos (em cinza), eliminando os tokens sem informação útil para conversão da sentença.

**Figura 57 - Exemplo de padrões linguísticos e seus respectivos rótulos semânticos**



Fonte: o autor.

## 6.4. Geração de linguagem computável

Nesta etapa foi desenvolvido um código em Python para automatização do processo de conversão dos termos extraídos em uma linguagem computável. Foram desenvolvidos testes de implementação em textos e sentenças isoladas que mostraram a possibilidade de exploração da biblioteca *ElementTree* para conversão do conteúdo extraído em arquivo de marcação XML.

A biblioteca possui incorporada o módulo *xml.etree.ElementTree*<sup>31</sup> para conversão dos dados extraídos em XML e depende da indicação dos elementos e sub elementos XML. O código desenvolvido conseguiu gerar uma estrutura XML com a estrutura apresentada no Quadro 30. O arquivo XML resultante do algoritmo conseguiu preservar os rótulos (ou marcações) e seu conteúdo correspondente, mantendo a hierarquia esperada. O exemplo apresentado na Figura 25 destaca algumas das frases com maior precisão no resultado, evidenciando a capacidade do algoritmo em separar, ordenar e rotular o conteúdo com informações semânticas.

<sup>31</sup> <https://docs.python.org/3/library/xml.etree.elementtree.html>

A estrutura é formada pelos elementos que recebem o texto do regulamento na marcação “FRASE” e os termos semântico se estruturam em subelementos individuais no elemento “TERMO”.

**Quadro 30 - Árvore de elementos XML obtida na extração**

```

<root>
  <FRASE>
    <TERMOS>
      <RÓTULO>
    </RÓTULO>
    </TERMOS>
  </FRASE>
</root>

```

Fonte: o autor.

O código construído para a conversão dos termos em elementos XML (Quadro 31) manteve a ordem em que os padrões extraíram os termos do texto. Desta forma, as marcações (rótulos) dos elementos no arquivo XML (Figura 25) são iniciadas pelo rótulo de <REFERENCIA> seguido de <EXIGENCIA> e por fim <COMPONENTE>.

**Quadro 31 - Pseudocódigo do método de conversão dos termos extraídos em estrutura XML**

```

INICIO
//Armazenamento e estruturação dos dados extraídos
PARA cada sentença:
  EXECUTA a busca de padrões e armazena em RESULTADOS
//Cria variável que armazenará dados do resultado da extração
sentença_convertida <- {FRASE <- sentença, TERMOS <- []}
  PARA cada conjunto de rótulo e objeto-span existentes em RESULTADOS:
    ADICIONA ao campo TERMOS “RÓTULO” e “TEXTO” no dicionário da variável
    sentença_convertida. Também são inseridos e os respectivos conteúdos de rótulo e
    texto
  ADICIONA cada sentença_convertida à lista_resultados
//Cria o arquivo XML
PARA cada item da lista_resultados:
  XML_SubElement <- marcação FRASE e seu conteúdo
  XML_SubElement.FRASE <- marcação TERMOS
  PARA cada item em XML_SubElement.FRASE.TERMOS:
    XML_SubElement.FRASE.TERMOS <-marcação ROTULO e seu conteudo
FIM

```

Fonte: o autor.

**Figura 58 - Exemplos bem-sucedidos da notação XML gerada pelo algoritmo**

```

<FRASE>
  As instalações de gás devem ser projetadas e executadas de acordo com as ABNT NBR 13523 e ABNT
  NBR 15526.
  <TERMOS>
    <REFERENCIA>ABNT NBR 13523</REFERENCIA>
    <REFERENCIA>ABNT NBR 15526</REFERENCIA>
    <EXIGENCIA>devem ser projetadas e executadas</EXIGENCIA>
    <COMPONENTE>instalações de gás</COMPONENTE>
  </TERMOS>
</FRASE>

<FRASE>
  Os materiais de revestimento, acabamento e isolamento termoacústico empregados na face interna
  dos sistemas ou elementos que compõem a edificação devem ter as características de propagação
  de chamas controladas, de forma a atender aos requisitos estabelecidos nas ABNT NBR 15575-3 e
  a ABNT NBR 15575-5 e ABNT NBR 9442.
  <TERMOS>
    <REFERENCIA>ABNT NBR 15575-3</REFERENCIA>
    <REFERENCIA>ABNT NBR 15575-5</REFERENCIA>
    <REFERENCIA>ABNT NBR 9442</REFERENCIA>
    <EXIGENCIA>devem ter</EXIGENCIA>
    <COMPONENTE>características de propagação de chamas controladas,</COMPONENTE>
    <COMPONENTE>materiais de revestimento</COMPONENTE>
    <COMPONENTE>isolamento termoacústico</COMPONENTE>
    <COMPONENTE>face interna</COMPONENTE>
  </TERMOS>
</FRASE>

<FRASE>
  Atendimento às seguintes Normas de projeto estrutural: ABNT NBR 14323, para estruturas de aço;
  ABNT NBR 15200, para estruturas de concreto; para as demais estruturas, aplica-se o Eurocode
  correspondente, em sua última edição.
  <TERMOS>
    <REFERENCIA>ABNT NBR 14323</REFERENCIA>
    <REFERENCIA>ABNT NBR 15200</REFERENCIA>
    <REFERENCIA>Eurocode correspondente</REFERENCIA>
    <COMPONENTE>Normas de projeto estrutural</COMPONENTE>
    <COMPONENTE>Atendimento às seguintes</COMPONENTE>
    <COMPONENTE>estruturas de aço</COMPONENTE>
    <COMPONENTE>estruturas de concreto</COMPONENTE>
  </TERMOS>
</FRASE>

<FRASE>
  O desempenho da edificação está intimamente associado a todos os projetos de implantação e ao
  desempenho das fundações, devendo ser atendidas as disposições das Normas aplicáveis,
  particularmente das ABNT NBR 8044, ABNT NBR 5629, ABNT NBR 11682, ABNT NBR 6122 e ABNT NBR
  12722.
  <TERMOS>
    <REFERENCIA>ABNT NBR 8044</REFERENCIA>
    <REFERENCIA>ABNT NBR 5629</REFERENCIA>
    <REFERENCIA>ABNT NBR 11682</REFERENCIA>
    <REFERENCIA>ABNT NBR 6122</REFERENCIA>
    <REFERENCIA>ABNT NBR 12722</REFERENCIA>
    <REFERENCIA>Normas aplicáveis</REFERENCIA>
    <EXIGENCIA>devendo ser atendidas</EXIGENCIA>
    <COMPONENTE>desempenho da edificação</COMPONENTE>
    <COMPONENTE>projetos de implantação</COMPONENTE>
    <COMPONENTE>desempenho das fundações</COMPONENTE>
  </TERMOS>
</FRASE>

```

Fonte: o autor.

## 7. AVALIAÇÃO

### 7.1. Aplicação de métricas no resultado do Capítulo 8 da Norma de Desempenho

Ao final da extração e rotulagem de padrões, foi executada a avaliação do resultado a partir das métricas de acurácia, precisão, revocação e valor-F1. Os valores foram obtidos a partir da classificação dos dados obtidos segundo uma matriz de confusão. Cada termo da sentença foi classificado quando há verdadeiros positivos, falsos negativos, falsos positivos e falsos negativos segundo o seguinte critério:

**Quadro 32 - Matriz de confusão para qualificação do resultado**

Descrição	Categorias	Valores
Era padrão e foi capturado	Verdadeiros Positivos	#
Era padrão e não foi capturado	Falsos Negativos	#
Não era padrão e foi capturado	Falsos Positivos	#
Não era padrão e não foi capturado	Verdadeiros Negativos	#

Fonte: o autor.

O Quadro 32 exemplifica o processo de classificação dos termos de cada frase do Capítulo 8 da Norma. Inicialmente foi feita uma análise humana das sentenças extraídas em planilha (Quadro 33) identificando em vermelho os itens desejáveis para a extração. Em seguida foram marcados com cinza os termos encontrados pelo código.

**Quadro 33 - Exemplo de avaliação das sentenças extraídas do Capítulo 8**

2	As instalações elétricas das edificações habitacionais devem ser projetadas de acordo com a ABNT NBR 5410 e Normas Brasileiras aplicáveis.	VP	5
		FN	0
		FP	0
		VN	1
14	Para isolamento de risco: análise do projeto e dimensionamento das distâncias seguras, tendo em conta a ignição-piloto por radiação e a convecção através da cobertura.	VP	5
		FN	0
		FP	0
		VN	2
5	A comprovação do atendimento ao requisito de 8.2, pelos critérios estabelecidos em 8.2.1.1 a 8.2.1.3, deve ser feita pela análise do projeto ou por inspeção em protótipo.	VP	4
		FN	3
		FP	0
		VN	3
19	O edifício habitacional multifamiliar deve dispor de sistemas de alarme, extinção, sinalização e iluminação de emergência, conforme proposto nas ABNT NBR 17240, ABNT NBR 13434 (Partes 1, 2 e 3), ABNT NBR 12693, ABNT NBR 13714 e ABNT NBR 10898.	VP	9
		FN	2
		FP	0
		VN	3

Fonte: o autor.

Uma vez que o número de classificações corretas e erradas foi contabilizado, é feito o cálculo das métricas de classificação. O Quadro 34 evidencia as métricas de qualidade, sua interpretação e fórmula de cálculo.

**Quadro 34 - Cálculo das métricas de qualidade**

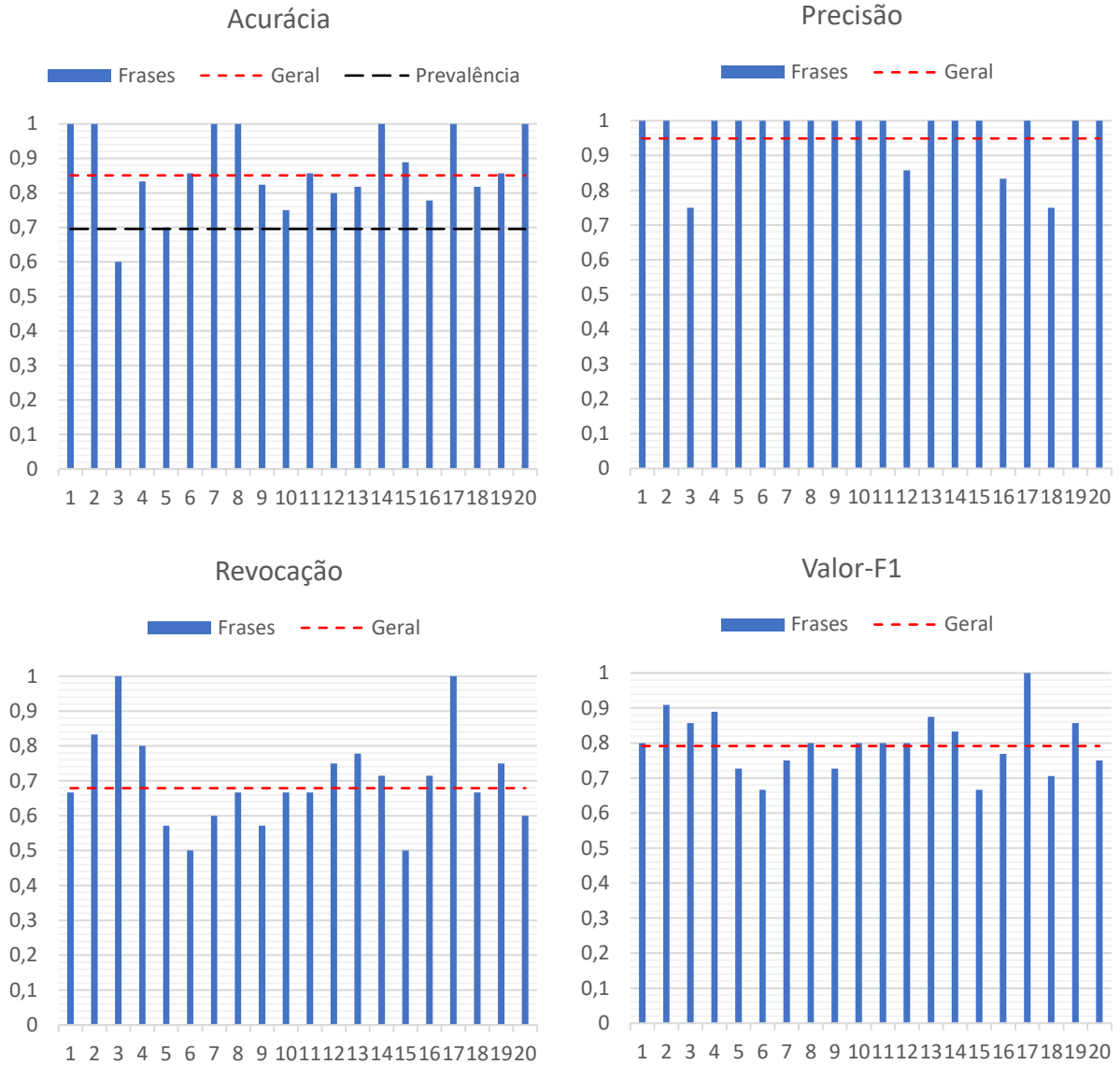
<b>Métrica</b>	<b>Descrição</b>	<b>Fórmula</b>
Acurácia (Acc):	Avalia a quantidade de classificações corretas.	$(VP + VN) / \text{total}$
Precisão (Pre):	Avalia de todos os dados classificados como positivos, quantos são realmente positivos.	$(VP / (VP+FP))$
Revocação (Rev)	(Sensibilidade) Avalia a porcentagem de dados classificados como positivos comparado com a quantidade real de positivos.	$(VP / (VP+FN))$
Valor-F1 (F1)	Medida geral que avalia em um único valor o estado geral de precisão e revocação.	$(2*P*R) / (P+R)$
Prevalência (Prev)	Mostra se as classes para classificação estão desbalanceadas.	$VP+FN / \text{total}$
Especificidade (Esp)	Capacidade do sistema prever corretamente na ausência de casos negativos	$(VN / (VN+FP))$
Eficiência (Efc)	Medida que balanceia medidas opostas, ideal estar mais próximo de 1.	$(Rev+Esp) / 2$

Fonte: o autor.

As métricas foram aplicadas na classificação realizada para o Capítulo 8 da Norma de Desempenho, e foram gerados os gráficos da Figura 59.

Os valores no eixo X (horizontal) estão organizados por sentença, sendo que a linha tracejada em vermelho representa a medida geral. Conforme mostrado na Figura , altos níveis de precisão indicam que todos os termos capturados pelo modelo são os que deveriam ser capturados naquela frase. Por exemplo, as frases 3 e 17 atingiram a totalidade dos termos (100%). A medida de revocação indica quantos dos dados positivos na amostra foram corretamente classificados como positivos. Um baixo número, como ocorre nas frases 5, 6 e 15 indica que o modelo capturou poucos termos. Essa é uma medida relevante para a análise da extração de informações, pois revela a capacidade de captura de termos corretos.

Figura 59 - Gráficos das medidas de classificação do Capítulo 8 separadas por frase



Fonte: o autor.

Desta forma, um modelo que capture todas as palavras de uma frase terá alto valor de revocação. Portanto, é necessário avaliar o modelo também pela métrica da precisão, ou seja, dos termos capturados, quantos realmente deveriam ter sido capturados.

Neste momento, podem ser considerados dois extremos: um modelo capaz de extrair muitos dados errados (alta revocação e baixa precisão) ou um modelo que extrai poucos dados, mas sempre corretos (alta precisão, baixa revocação). Para essa análise

utiliza-se a medida valor-F1, que é uma ponderação entre revocação e precisão. O valor perfeito para a medida valor-F1 é 1, indicando que todos os termos corretos foram extraídos e apenas os corretos.

Neste cenário, a medida de acurácia não é significativa porque só relata a quantidade de termos classificados de modo correto comparado aos totais de termos existentes. A variabilidade das frases em número de palavras e termos corretos não permite que a medida de acurácia traga informação relevante como métrica do modelo.

O valor da média geral separado por métrica está colocado na Tabela 5.

**Tabela 5 - Média das métricas do Capítulo 8 da Norma de Desempenho**

<b>Medida</b>	<b>Valor</b>
Acurácia	0,85
Precisão	0,95
Revocação	0,68
Valor-F1	0,79
Prevalência	0,7
Especificidade	0,9
Eficiência	0,79

Fonte: o autor.

Além das métricas de padrão de acurácia, precisão e revocação, a especificidade é importante nestes dados, porque considera a capacidade do modelo em um cenário de poucos elementos negativos, ou seja, uma frase em que a maioria dos termos deveria ser extraída. A especificidade mede a capacidade do modelo em reconhecer os casos negativos, ou seja, evidencia que o modelo consegue reconhecer os casos que não devem ser extraídos. No caso da extração do Capítulo 8 o modelo de extração atingiu um nível elevado (90%).

## **7.2. Aplicação de métricas no resultado da Parte 1 da Norma**

O código gerado na etapa de desenvolvimento foi aplicado no restante da Parte 1 da Norma de Desempenho para comparação do resultado como um teste cego (Quadro 35). O texto foi segmentado e formalizado em arquivo .txt (Apêndice E). Ao todo foram processadas 53 novas frases e foram obtidos 392 termos rotulados (Apêndice K).

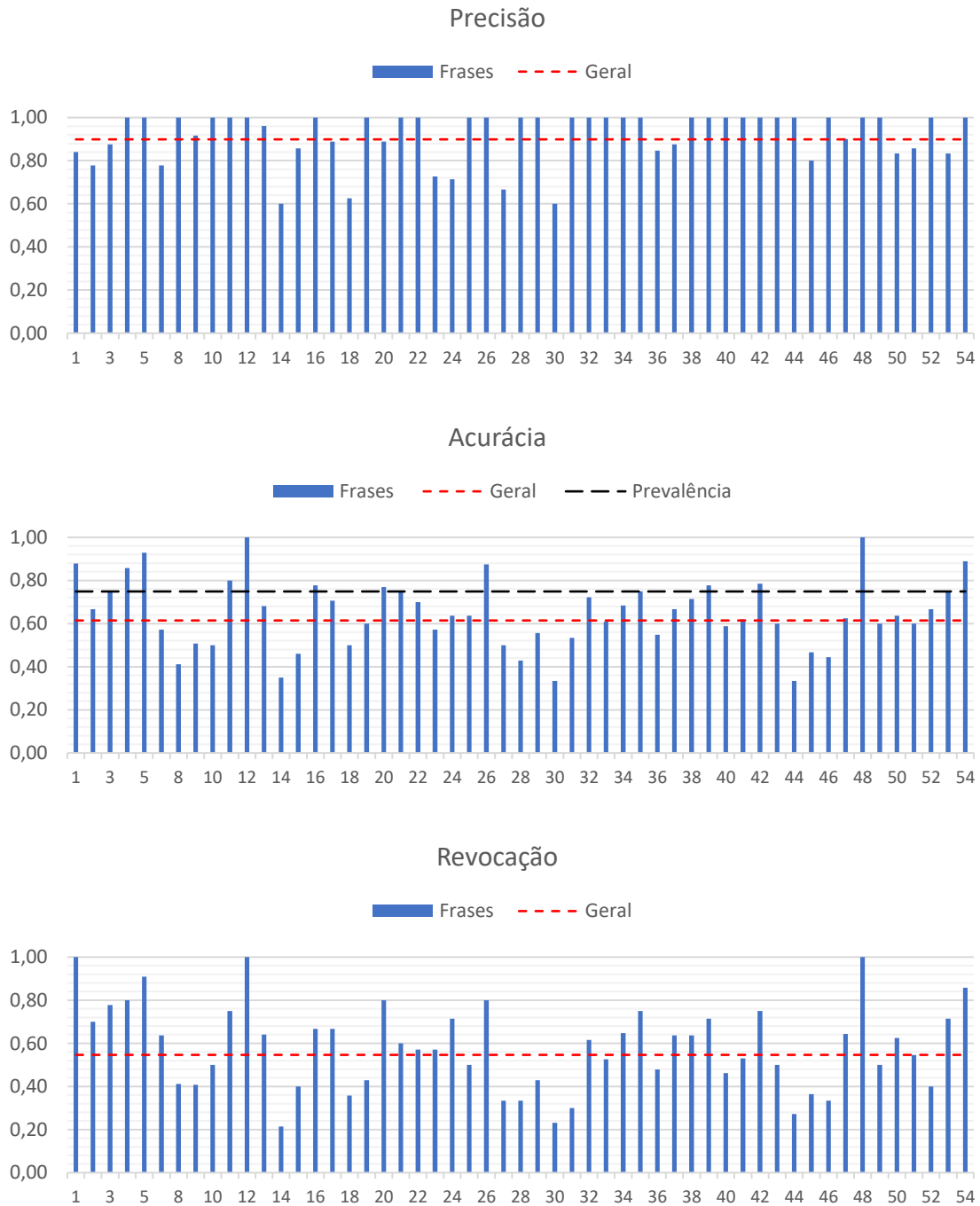
**Quadro 35 - Exemplos de avaliação realizada sobre textos da Parte 1**

1	Para edifícios ou conjuntos habitacionais com local de implantação definido, os projetos de arquitetura, da estrutura, das fundações, contenções e outras eventuais obras geotécnicas devem ser desenvolvidos com base nas características do local da obra (topográficas, geológicas etc.), avaliando-se convenientemente os riscos de deslizamentos, enchentes, erosões, vibrações transmitidas por vias férreas, vibrações transmitidas por trabalhos de terraplenagem e compactação do solo, ocorrência de subsidência do solo, presença de crateras em camadas profundas, presença de solos expansíveis ou colapsíveis, presença de camadas profundas deformáveis e outros.	VP	21
		FN	0
		FP	4
		VN	8
8	Os sistemas não podem apresentar: a) rupturas, instabilidades, tombamentos ou quedas que possam colocar em risco a integridade física dos ocupantes ou de transeuntes nas imediações do imóvel; b) partes expostas cortantes ou perfurantes; c) deformações e defeitos acima dos limites especificados nas ABNT NBR 15575-2 a ABNT NBR 15575-6	VP	7
		FN	10
		FP	0
		VN	0
9	Devem ser previstas no projeto e na execução formas de minimizar, durante o uso da edificação, o risco de: a) queda de pessoas em altura: telhados, áticos, lajes de cobertura e quaisquer partes elevadas da construção; b) acessos não controlados aos locais com riscos de quedas; c) queda de pessoas em função de rupturas das proteções, as quais devem ser ensaiadas conforme ABNT NBR 14718 ou devem possuir memorial de cálculo assinado por profissional responsável que comprove seu desempenho; d) queda de pessoas em função de irregularidades nos pisos, rampas e escadas, conforme a ABNT NBR 15575-3; e) ferimentos provocados por ruptura de subsistemas ou componentes, resultando em partes cortantes ou perfurantes; f) ferimentos ou contusões em função da operação das partes móveis de componentes, como janelas, portas, alçapões e outros; g) ferimentos ou contusões em função da dessolidarização ou da projeção de materiais ou componentes a partir das coberturas e das fachadas, tanques de lavar, pias e lavatórios, com ou sem pedestal, e de componentes ou equipamentos normalmente fixáveis em paredes; h) ferimentos ou contusões em função de explosão resultante de vazamento ou de confinamento de gás combustível.	VP	22
		FN	32
		FP	2
		VN	13
18	O arquivo climático utilizado deve fornecer valores mensais de temperatura média do solo, em graus Celsius (°C), para todos os meses do ano, além dos seguintes valores horários representativos das 8760 h do ano meteorológico típico: a) temperatura de bulbo seco, expressa em graus Celsius (°C); b) temperatura do ponto de orvalho, expressa em graus Celsius (°C); c) umidade relativa, expressa em porcentagem (%); d) pressão atmosférica, expressa em Pascal (Pa); e) intensidade de radiação horizontal de onda longa, expressa em watts-hora por metro quadrado (Wh/m <sup>2</sup> ); f) radiação horizontal global, expressa em watts-hora por metro quadrado (Wh/m <sup>2</sup> ); g) radiação normal direta, expressa em watts-hora por metro quadrado (Wh/m <sup>2</sup> ); h) radiação horizontal difusa, expressa em watts-hora por metro quadrado (Wh/m <sup>2</sup> ); i) direção do vento, expressa em graus (°), considerando o sentido horário a partir da direção Norte; j) velocidade do vento, expressa em metros por segundo (m/s).	VP	10
		FN	18
		FP	6
		VN	14
35	A altura mínima de pé-direito não pode ser inferior a 2,50 m.	VP	3
		FN	1
		FP	0
		VN	0

Fonte: o autor.

Os valores de acurácia, precisão e revocação obtidos para o trecho referente à extração de informações da Parte 1 estão distribuídos segundo os gráficos da Figura 60.

**Figura 60 - Gráficos das medidas de classificação da Parte 1 separadas por frase**



Fonte: o autor.

Tabela 6 - Média das métricas da Parte 1

Medida	Valor
Acurácia	0,61
Precisão	0,90
Revocação	0,55
Valor-F1	0,68
Prevalência	0,75
Especificidade	0,82
Eficiência	0,68

**Fonte: o autor.**

A Tabela 6 evidencia os valores gerais das métricas de qualidade da extração para a Parte 1 da Norma de Desempenho. O baixo nível de revocação indica que o modelo não reconheceu corretamente os padrões em todo o texto da Parte 1, ou seja, muitos termos que deveriam ser extraídos não foram identificados pelo modelo. Frase como a de número 14 e 30 não foram capazes de ter seus termos reconhecidos em níveis aceitáveis indicando necessidade de alteração dos padrões ou de outros parâmetros do modelo de extração.

Assim como os resultados obtidos com o processamento do Capítulo 8 da Norma, aqui o critério de especificidade foi alto, indicando que o modelo não extraiu 82% dos termos que realmente não deveriam ser extraídos.

### 7.3. Avaliação da ordem de busca

Uma vez que os padrões construídos na Seção 6.3.2 obedecem à ordem em que são apresentados, foi desenvolvido um experimento reordenando a sequência que os padrões são buscados para avaliar a sensibilidade do algoritmo para a ordenação dos padrões.

A sequência original de rótulos é 1) REFERÊNCIA, 2) COMPONENTE, 3) PARAMETROS, 4) EXIGÊNCIA e 5) VALOR, e foi alterada para 1) REFERÊNCIA, 2) EXIGÊNCIA, 3) COMPONENTE, 4) PARAMETROS e 5) VALOR.

O teste foi reaplicado sobre o mesmo trecho, Parte 1 da Norma de Desempenho, excluindo o Capítulo 8 da Norma de Desempenho que foi usado para construção dos padrões. A alteração da ordem dos padrões apresentou o seguinte resultado:

1. A quantidade de objetos semânticos identificados foi a mesma (402 elementos);

2. Todos os objetos semânticos, i.e., elementos rotulados, ainda estavam compostos pelos *tokens* originais;
3. A quantidade de rótulos por cada tipo se manteve idêntica.

Estes pontos indicam a robustez do código e dos padrões que conseguiram seguir uma ordem de identificação que garantiu a repetibilidade do processo sem qualquer prejuízo. A ausência de alterações na classificação dos termos também mostra a consistência da rotulagem da biblioteca *SpaCy* que reproduziu a mesma análise das sentenças em cada alteração dos padrões. Não houve variação nas rotulagens e, mesmo com remoção da pontuação a classificação se manteve com alta acurácia. A tabela comparativa das rotulagens semânticas está no Apêndice L.

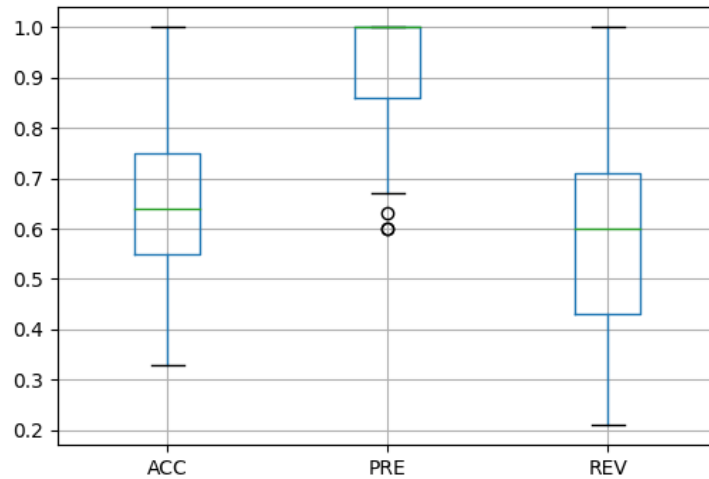
#### **7.4. Análise dos resultados**

Uma vez que o modelo não alcançou um desempenho semelhante ao observado no Capítulo 8 da Norma de Desempenho, surge a hipótese de que as frases que apresentaram baixo valor nas métricas devem ser examinadas, pois seus termos podem não estar representados pelos padrões.

Assim, foi realizada uma análise dos dados para identificar as sentenças que se destacam do conjunto de dados. Inicialmente, buscou-se por dados discrepantes, procurando identificar as sentenças que se diferenciam do restante do conjunto de dados.

A Figura representa um gráfico do tipo *boxplot* que registra a mediana e a distribuição dos dados dentro do 1º e 3º quartis. Este gráfico é usado para identificação de dados discrepantes (*outliers*) cujos atributos estejam fora dos limites do gráfico. A figura revela 3 frases, as de número 14, 18 e 30, que possuem precisão entre 0.6 a 0.63. A identificação desta discrepância mostrou que as frases de número 14 e 30 possuem os mais baixos índices de acurácia e revocação, apontando para a necessidade de avaliação de seus resultados na fase de análise de discrepâncias. As três frases identificadas possuem os menores valores de métricas (Tabela 7).

**Figura 61 - Boxplot com identificação de dados discrepantes**



Fonte: o autor.

**Tabela 7 - Métricas de qualidade das frases discrepantes número 14, 18 e 30.**

	Frase 14	Frase 18	Frase 30
ACC	0,35	0,50	0,33
PRE	0,60	0,63	0,60
VER	0,21	0,36	0,23

Fonte: O autor

**Tabela 8 - Descrição estatística do conjunto de dados**

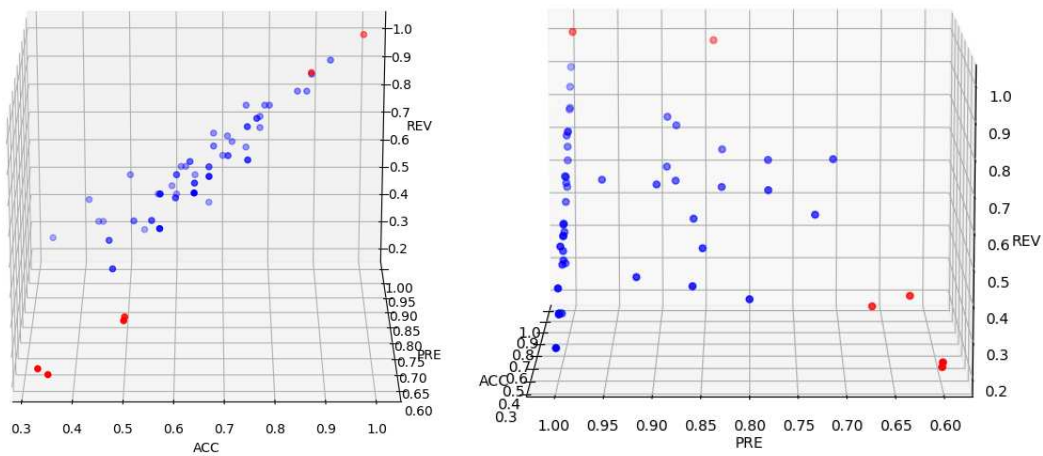
	Média	Desvio Padrão
ACC	0,65	0,16
PRE	0,92	0,12
REV	0,58	0,20

Fonte: o autor.

Apesar do método *boxplot* ser comumente utilizado em análises estatísticas, nesta pesquisa foi considerada a análise do método valor-z para confirmar a descrição dos dados discrepantes. Enquanto o método *boxplot* avalia os outliers que estão fora de determinada porcentagem da principal distribuição dos dados. O método valor-z estabelece seu limite com base nos desvios padrão. Esse método define que serão aceitos todos os dados que estiverem dentro do limite de  $n$  desvios padrão além do valor da média.

Com a aplicação desse método, foi adotado um limite menos rigoroso de valor de dois desvios padrão ( $z=2$ ). Desta forma, a análise é mais restritiva, implementando um limite menor à variabilidade dos resultados, e restringindo um comportamento normal de dados somente para as sentenças que estiverem mais próximas da mediana. Neste caso, o método valor- $z$  é mais restritivo que o método *boxplot* e indica mais outliers nos dados (Figura 62).

**Figura 62 - Visualização dos dados com identificação dos outliers (em vermelho)**



Fonte: o autor.

A análise do valor- $z$  então indicou as frases de número 1, 12, 14, 18, 27, 30, 48, como elementos discrepantes dos dados da amostra (Tabela 9).

**Tabela 9 - Quadro resumo dos valores dos outliers**

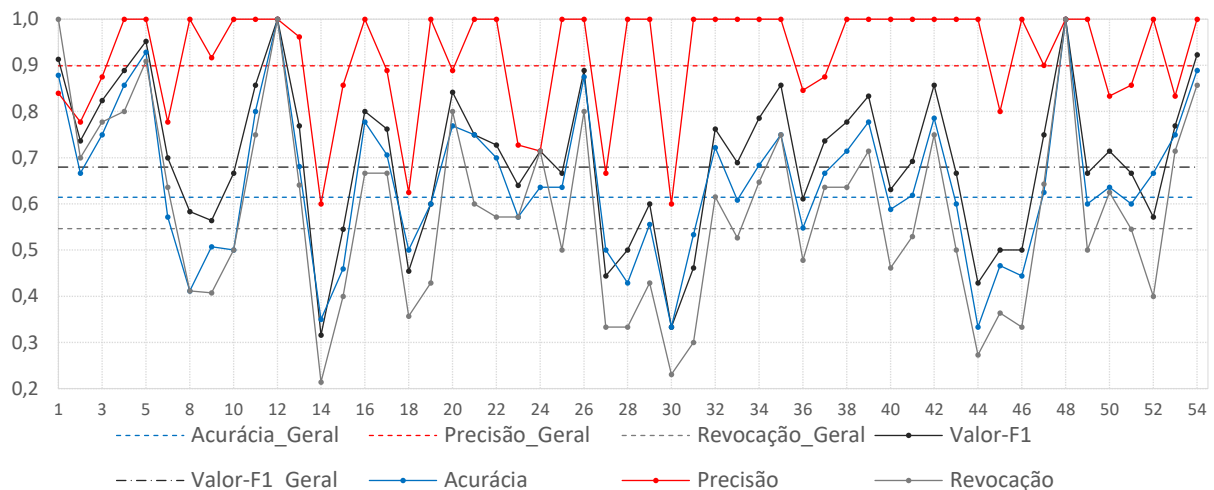
	Frase 1	Frase 12	Frase 14	Frase 18	Frase 27	Frase 30	Frase 48
ACC	0,88	1,00	0,35	0,50	0,50	0,33	1,00
PRE	0,84	1,00	0,60	0,63	0,67	0,60	1,00
REV	1,00	1,00	0,21	0,36	0,63	0,23	1,00

Fonte: o autor.

Os valores em verde representam as frases 1, 12 e 48 que obtiveram altos valores de métricas de qualidade, enquanto as frases 14, 18, 27 e 30 apresentaram resultados mais baixos em acurácia, precisão e revocação. Estas frases serão consideradas mais adiante na análise de discrepâncias.

Ao agrupar as medidas em uma visualização única (Figura ) é possível identificar frases que possuem valores discrepantes pela inversão da correlação entre as medidas Valor-F1, Acurácia e Precisão. O gráfico aponta para as frases já indicadas anteriormente (14, 18, 27 e 30), com acréscimo da frase 52 que tem baixo valor de revocação e, conseqüentemente, baixo valor-F1. A análise da frase 52 indicou que é uma frase curta e obteve baixa acurácia porque duas expressões “destinação não potável” e “Tabela 8” não estavam contempladas pelos padrões, mas, neste caso seriam padrões facilmente inseridos. Ao identificar estes novos padrões foi observado que seriam padrões repetidos em outras sentenças e, portanto, é desejável a incorporação destes padrões para melhoria de qualidade em outras sentenças. No caso da necessidade de identificação de termos como “Tabela X”, é possível construir um padrão baseado em expressões regulares. Por exemplo, ao usar a expressão “Tabela\s\d+” permitiria a identificação de termos como “Tabela 2, Tabela 15 e Tabela 8”.

**Figura 63 - Gráfico com as medias agrupadas**



Fonte: o autor.

A análise das medidas de qualidade se mostrou produtiva, mas ainda incompleta porque não permite que seja reconhecido o comportamento das demais frases. Também não permitiu reconhecer frases com problemas ou qualidades similares.

Desta forma, a fim de ampliar a análise e reconhecer padrões não identificados anteriormente, foi utilizado o método de agrupamento (*clusterização*) do resultado pelas variáveis acurácia, precisão e revocação. O objetivo desta nova análise é identificar grupos de frases de mesma categoria e, possivelmente, mesma estrutura morfológica ou sintática que aponte para correções do processo.

#### 7.4.1. Agrupamento de sentenças de mesma característica pelo método de clusterização

O agrupamento, também conhecido como clusterização, é uma técnica utilizada para identificar grupos ou *clusters* de dados que compartilham características semelhantes. Diversos métodos, como K-Means, DBSCAN e Fuzzy K-Means, são empregados para agrupar dados visando revelar padrões, atributos ou qualidades ocultas nos dados.

No caso dos dados obtidos pela análise das métricas de avaliação, representadas pelas Tabela 5 e Tabela 6 foram selecionados os atributos acurácia, precisão e revocação de cada frase. As demais variáveis, por se tratar de valores derivados de acurácia, precisão e revocação, estão diretamente correlacionadas e poderiam levar a resultados distorcidos.

Como nesta pesquisa a aplicação do agrupamento tem o intuito de ampliar a análise dos resultados, foi escolhido o método K-means que define grupos de modo linear. Foi utilizado o método baseado em aprendizagem de máquina “K-means” da biblioteca *scikit-learn*<sup>32</sup>. O algoritmo opera pela otimização da posição do centróide ( $\mu$ ) que define cada agrupamento de elementos. O processo é regido por uma função objetiva baseada em encontrar a mínima variância entre elementos da amostra usando uma função baseada em erro quadrático médio entre os elementos do grupo e seu centroide (Eq. 1).

$$\sum_{i=1}^m \sum_{k=1}^K \|x_i - \mu_k\|^2 \quad (\text{Equação 1})$$

---

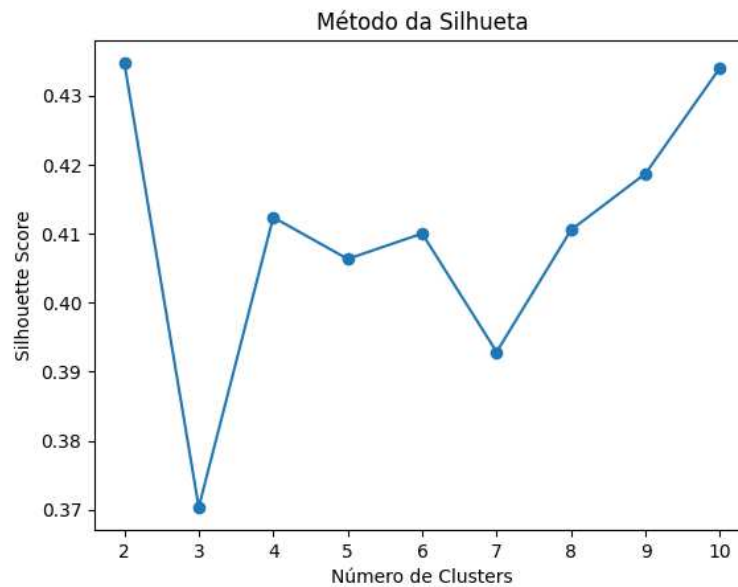
<sup>32</sup> 2.3. Clustering – scikit-learn 1.4.0 documentation

Neste método a quantidade de grupos (centroides) é determinada pelo analista que deve utilizar algum método para selecionar a quantidade ideal de agrupamentos  $k$ .

Existem diferentes métodos para proposição de valores de  $k$ , como o método do cotovelo e o método da silhueta. Nesta pesquisa foi adotado o método da silhueta, criado por Rousseeuw em 1987 (Rousseeuw, 1987), porque é reconhecido por maior capacidade de identificação de grupos em comparação com outros métodos comuns (Arbelaitz *et al*, 2013).

O resultado de aplicação do método da silhueta para agrupamento dos dados obtidos pelas medidas de acurácia, precisão e revocação das sentenças está apresentado na Figura 64.

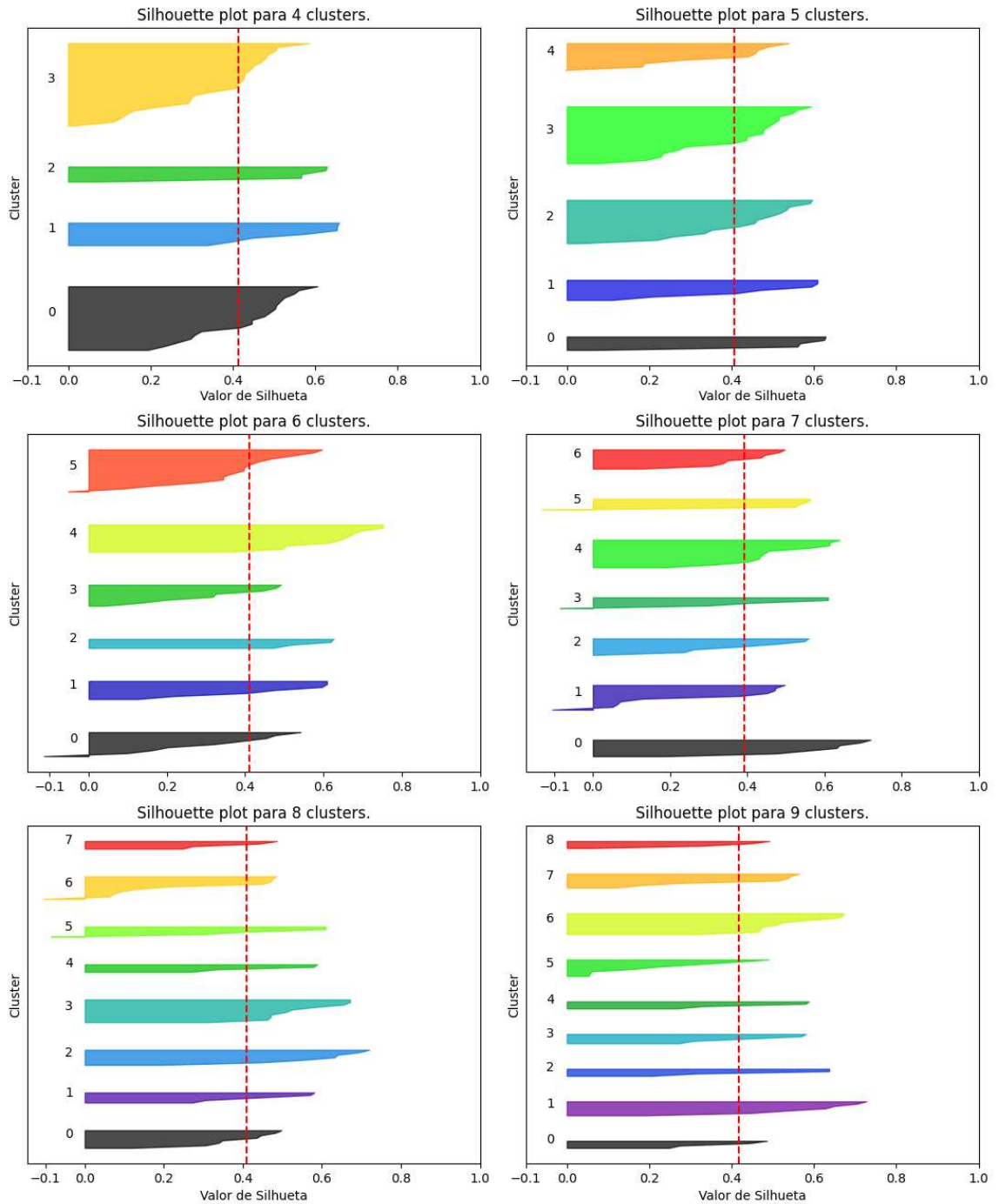
**Figura 64- Resultado do método da silhueta**



Fonte: o autor.

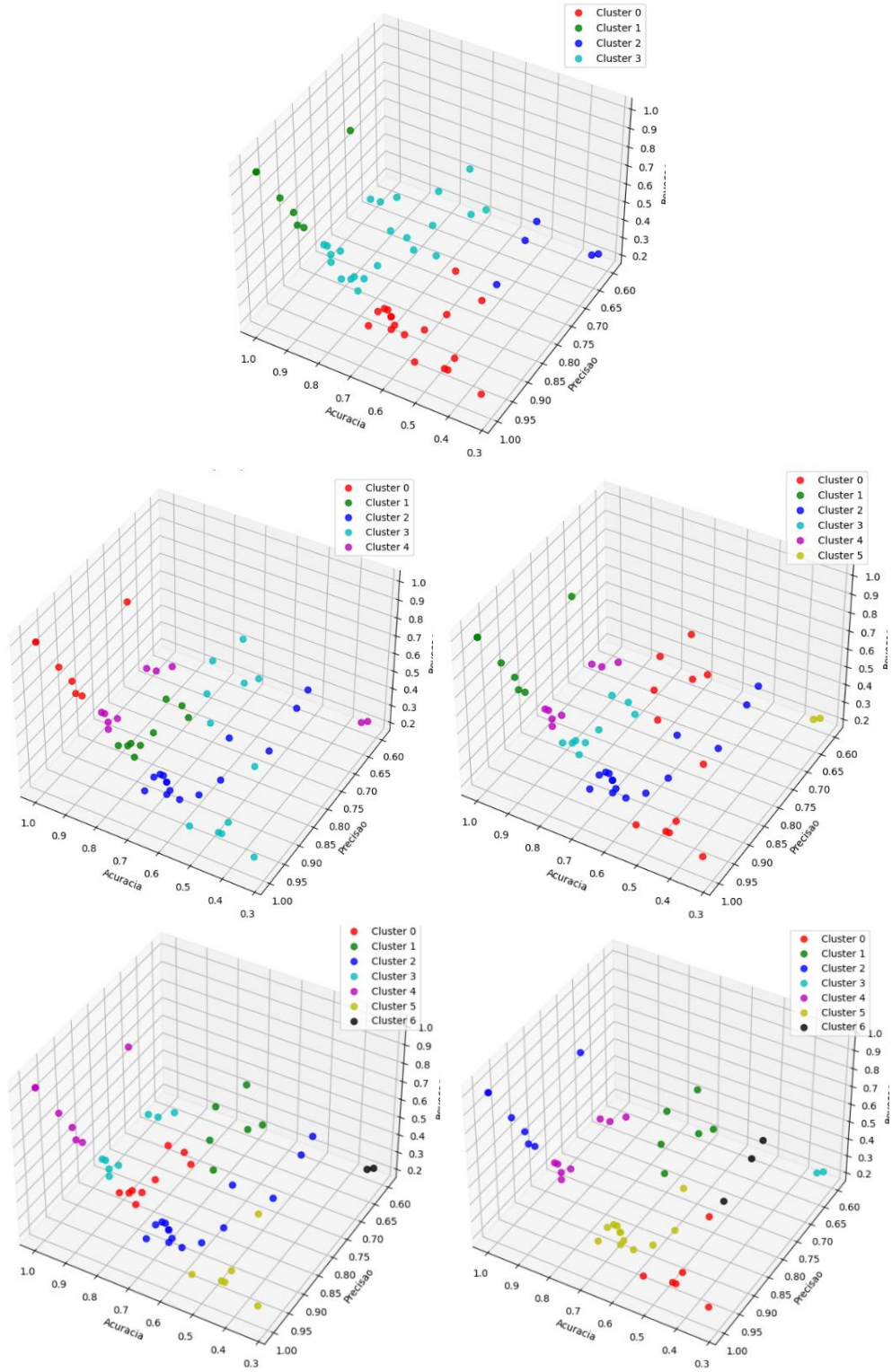
O coeficiente de silhueta aponta uma região próxima de 4-6 clusters com pouca variação. Então os dados com 4 a 8 *clusters* foram interpretados e foi identificada pouca diferença de interpretação do resultado entre os grupos (Figura 65, Figura 66 e Figura 67). Desta forma, fica mais evidente e capaz de ser interpretado os valores de acurácia, precisão e revocação obtidos organizados em 4 grupos.

Figura 65 - Avaliação do resultado dos testes de silhueta para grupos 4 a 8



Fonte: o autor.

Figura 66 - Visualização dos grupos 4 a 8



Fonte: o autor.

Figura 67 - Comparação de agrupamento com 8 e 4 grupos

Acc = Acurácia  
Pre = Precisão  
Rev = Revocação

Frase	Acc	Pre	Rev	Cluster-8	Frase	Acc	Pre	Rev	Cluster-4
10	0.5	1.0	0.5	0	10	0.50	1.00	0.50	0
52	0.7	1.0	0.4	0	28	0.43	1.00	0.33	0
43	0.6	1.0	0.5	0	52	0.67	1.00	0.40	0
25	0.6	1.0	0.5	0	43	0.60	1.00	0.50	0
29	0.6	1.0	0.4	0	44	0.33	1.00	0.27	0
40	0.6	1.0	0.5	0	25	0.64	1.00	0.50	0
33	0.6	1.0	0.5	0	29	0.56	1.00	0.43	0
41	0.6	1.0	0.5	0	46	0.44	1.00	0.33	0
49	0.6	1.0	0.5	0	31	0.53	1.00	0.30	0
19	0.6	1.0	0.4	0	40	0.59	1.00	0.46	0
24	0.6	0.7	0.7	1	33	0.61	1.00	0.53	0
50	0.6	0.8	0.6	1	8	0.41	1.00	0.41	0
37	0.7	0.9	0.6	1	41	0.62	1.00	0.53	0
53	0.8	0.8	0.7	1	49	0.60	1.00	0.50	0
2	0.7	0.8	0.7	1	19	0.60	1.00	0.43	0
51	0.6	0.9	0.5	1	36	0.55	0.85	0.48	0
17	0.7	0.9	0.7	1	15	0.46	0.86	0.40	0
6	0.6	0.8	0.6	1	9	0.51	0.92	0.41	0
47	0.6	0.9	0.6	1	12	1.00	1.00	1.00	1
23	0.6	0.7	0.6	1	4	0.86	1.00	0.80	1
35	0.8	1.0	0.8	2	48	1.00	1.00	1.00	1
11	0.8	1.0	0.8	2	26	0.88	1.00	0.80	1
4	0.9	1.0	0.8	2	54	0.89	1.00	0.86	1
26	0.9	1.0	0.8	2	5	0.93	1.00	0.91	1
20	0.8	0.9	0.8	2	1	0.88	0.84	1.00	1
42	0.8	1.0	0.8	2	45	0.47	0.80	0.36	2
3	0.8	0.9	0.8	2	30	0.33	0.60	0.23	2
39	0.8	1.0	0.7	2	27	0.50	0.67	0.33	2
16	0.8	1.0	0.7	3	14	0.35	0.60	0.21	2
21	0.8	1.0	0.6	3	18	0.50	0.63	0.36	2
22	0.7	1.0	0.6	3	35	0.75	1.00	0.75	3
38	0.7	1.0	0.6	3	11	0.80	1.00	0.75	3
32	0.7	1.0	0.6	3	16	0.78	1.00	0.67	3
34	0.7	1.0	0.6	3	21	0.75	1.00	0.60	3
13	0.7	1.0	0.6	3	24	0.64	0.71	0.71	3
12	1.0	1.0	1.0	4	50	0.64	0.83	0.63	3
48	1.0	1.0	1.0	4	22	0.70	1.00	0.57	3
54	0.9	1.0	0.9	4	37	0.67	0.88	0.64	3
5	0.9	1.0	0.9	4	20	0.77	0.89	0.80	3
1	0.9	0.8	1.0	4	38	0.71	1.00	0.64	3
28	0.4	1.0	0.3	5	42	0.79	1.00	0.75	3
44	0.3	1.0	0.3	5	53	0.75	0.83	0.71	3
46	0.4	1.0	0.3	5	3	0.75	0.88	0.78	3
31	0.5	1.0	0.3	5	2	0.67	0.78	0.70	3
8	0.4	1.0	0.4	5	32	0.72	1.00	0.62	3
30	0.3	0.6	0.2	6	51	0.60	0.86	0.55	3
27	0.5	0.7	0.3	6	17	0.71	0.89	0.67	3
14	0.4	0.6	0.2	6	6	0.57	0.78	0.64	3
18	0.5	0.6	0.4	6	47	0.63	0.90	0.64	3
45	0.5	0.8	0.4	7	34	0.68	1.00	0.65	3
36	0.5	0.8	0.5	7	23	0.57	0.73	0.57	3
15	0.5	0.9	0.4	7	39	0.78	1.00	0.71	3
9	0.5	0.9	0.4	7	13	0.68	0.96	0.64	3

Fonte: o autor.

Na Figura 67, os grupos marcados na coluna da direita revelam quatro grupos de frases com os mesmos resultados das métricas. É importante notar que a identificação de quatro *clusters* corrobora com a análise de *outliers* feita anteriormente. No grupo 2 foram identificadas as frases de menor desempenho e que também foram identificadas pelos métodos *boxplot* e valor-z para avaliar a presença de dados discrepantes. Além das frases identificadas anteriormente o método K-means adicionou uma nova frase à lista de frases de menor desempenho (frase 45) (Quadro 36).

**Quadro 36 - Grupo 2, frases com menor capacidade de identificação pelos padrões**

Nº	ACC	PRE	REV	Frases
14	0,35	0,60	0,21	Devem ser previstos no projeto detalhes que assegurem a estanqueidade de partes do edifício que tenham a possibilidade de ficar em contato com a água gerada na ocupação ou manutenção do imóvel, devendo ser verificada a adequação das vinculações entre instalações de água, esgotos ou águas pluviais e estrutura, pisos e paredes, de forma que as tubulações não venham a ser rompidas ou desencaixadas por deformações impostas.
18	0,50	0,63	0,36	O arquivo climático utilizado deve fornecer valores mensais de temperatura média do solo, em graus Celsius (°C), para todos os meses do ano, além dos seguintes valores horários representativos das 8760 h do ano meteorológico típico: a) temperatura de bulbo seco, expressa em graus Celsius (°C); b) temperatura do ponto de orvalho, expressa em graus Celsius (°C); c) umidade relativa, expressa em porcentagem (%); d) pressão atmosférica, expressa em Pascal (Pa); e) intensidade de radiação horizontal de onda longa, expressa em watts-hora por metro quadrado (Wh/m <sup>2</sup> ); f) radiação horizontal global, expressa em watts-hora por metro quadrado (Wh/m <sup>2</sup> ); g) radiação normal direta, expressa em watts-hora por metro quadrado (Wh/m <sup>2</sup> ); h) radiação horizontal difusa, expressa em watts-hora por metro quadrado (Wh/m <sup>2</sup> ); i) direção do vento, expressa em graus (°), considerando o sentido horário a partir da direção Norte; j) velocidade do vento, expressa em metros por segundo (m/s).
27	0,50	0,67	0,33	O projeto deve especificar o valor teórico para a vida útil de projeto (VUP) para cada um dos sistemas que o compõem, não inferiores aos estabelecidos na Tabela 7, e deve ser elaborado para que os sistemas tenham uma durabilidade potencial compatível com a vida útil de projeto (VUP) a serem considerados nos projetos elaborados a partir da exigibilidade desta parte da ABNT NBR 15575.
30	0,33	0,60	0,23	As especificações relativas à manutenção, uso e operação do edifício e seus sistemas que forem consideradas em projeto para definição da vida útil de projeto (VUP) devem estar também claramente detalhadas na documentação que acompanha o edifício ou subsidia sua construção.
45	0,45	0,80	0,36	Relativamente às instalações hidrossanitárias, devem ser atendidas as disposições da ABNT NBR 15575-6. Os elementos e componentes que contam com normalização específica (portas, janelas, torneiras e outros) devem ainda atender aos requisitos das respectivas normas.

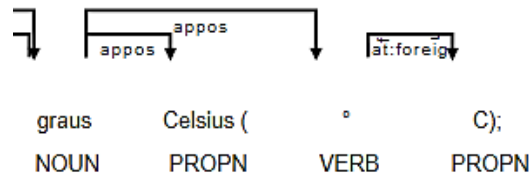
Fonte: o autor.

É importante notar que a frase 45 não foi encontrada pelos métodos *boxplot* e valor-z, mas sua identificação na clusterização permitiu que fosse reconhecido um problema comum a estas frases. As frases com encadeamento elementos como “água, esgotos ou

águas pluviais e estrutura, pisos e paredes” e “portas, janelas, torneiras e outros” induzem a uma baixa pontuação em acurácia porque os padrões que mapeiam os elementos construtivos não estavam preparados para esta sequência de termos.

Já as outras sentenças de baixos resultados indicaram problemas de identificação de elementos da Norma. A frase 18, por exemplo, possui muitos elementos que representam unidades de medida como “°C”, “Wh/m<sup>2</sup>” e “m/s”. A análise gramatical da biblioteca SpaCy gerou um resultado incorreto para estes elementos (Figura 68 e Figura 69).

**Figura 68- Segmento do resultado da análise gramatical e sintática da frase 18**



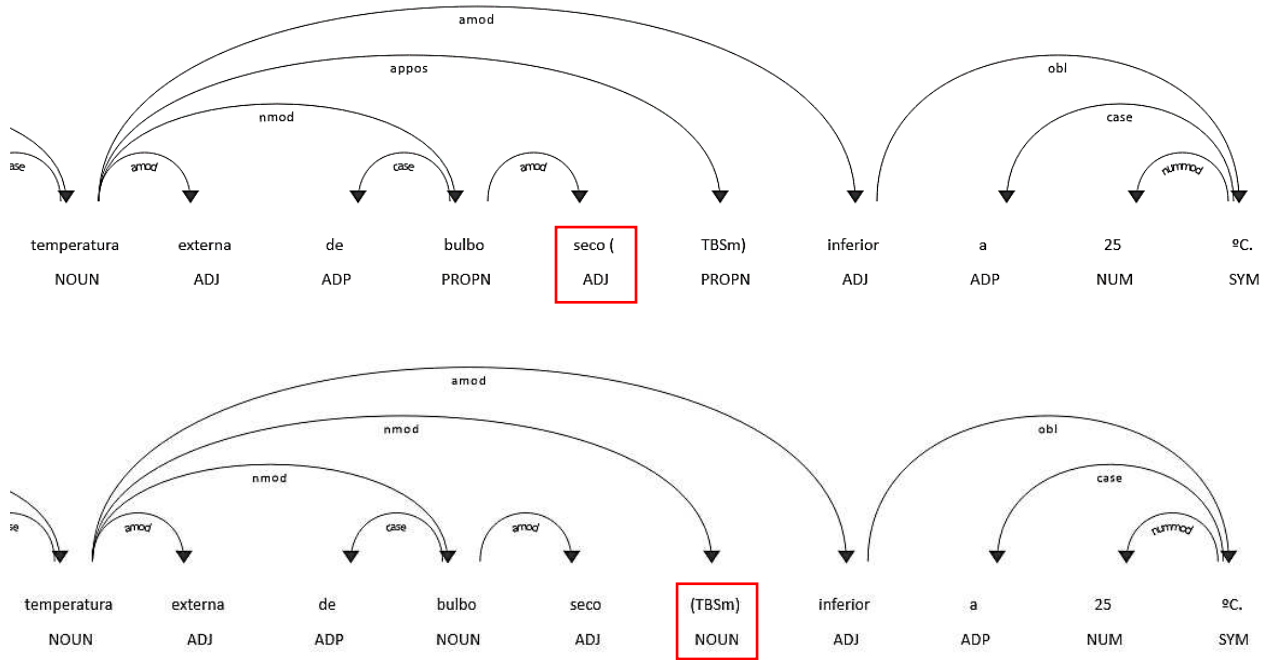
Fonte: o autor.

Neste caso uma solução possível é adotar um dicionário toponímico para auxiliar a rotulagem porque consegue registrar e capturar termos indicados numa lista de itens.

A fim de avaliar a viabilidade de uma solução como a criação de um dicionário toponímico foi desenvolvido um rotulador e tokenizador baseado em um dicionário de unidades de medida.

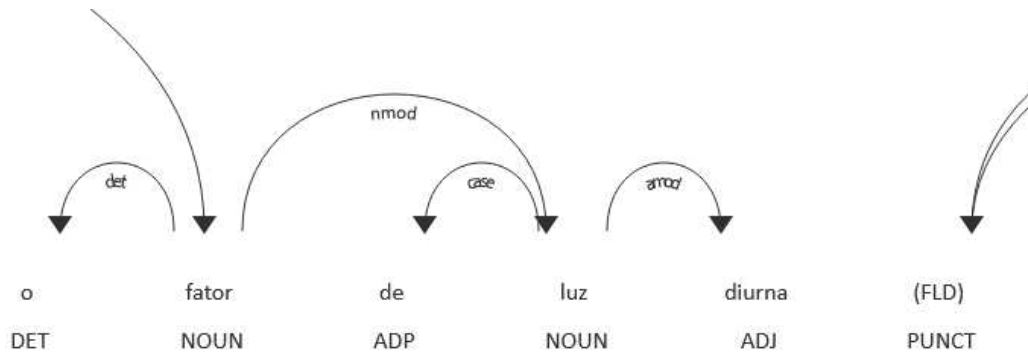
Ao utilizar o dicionário toponímico na sentença representada pela abaixo, o símbolo “°C” foi reconhecido como tal e associado ao numeral 25. Ao utilizar o dicionário o nome “TBSm” também foi identificado corretamente e a frase não sofreu qualquer modificação nas relações de dependência sintática.

Figura 69 - Anotação linguística utilizando o dicionário toponímico



Fonte: o autor.

Figura 70 - Uso do dicionário toponímico em conjunto com um tokenizador



Fonte: o autor.

**Figura 71 - Alternativa explorada para incorporação de rótulo personalizado**

Sentença	Token	W/K
1	Lemma	W/K
	Rótulo POS	PROPN
	UnidMedida	UNM
	Dependencia	flat

Sentença	Token	W/(m <sup>2</sup> .K)
2	Lemma	W/(m <sup>2</sup> .K)
	Rótulo POS	PROPN
	UnidMedida	UNM
	Dependencia	nmod

Fonte: o autor.

Na sentença da Figura 70, houve modificação no rótulo gramatical de “FLD” que passou a receber o rótulo “PUNCT”, definido pelo usuário para representar elementos simbólicos e de pontuação. Já na Figura 71, outro exemplo em que o tokenizador incluiu um novo rótulo (UNM) para identificar unidades de medida dentro do regulamento.

O código desenvolvido para a implementação do tokenizador de unidades de medida pode ser acessado pelo repositório Zenodo<sup>33</sup>.

É importante explicar que a adoção deste tipo de dicionário implica no não pré-processamento da frase para remoção de pontuações como um processo típico de PLN adotaria. Desta forma, esta pesquisa sugere que sejam utilizados tokenizadores que se baseiam em dicionários ou padrões regulares para permitir a identificação mais dos termos.

Também contribuiu para o baixo índice de qualidade de reconhecimento de padrões na frase 18 o fato de que as medidas por extenso, presentes na sentença, não eram necessárias, e já seria suficiente extrair apenas os símbolos referentes às unidades de medida.

---

<sup>33</sup> <https://zenodo.org/records/11315651>

Além disso foi identificado um padrão possível de ser agregado, que contempla termos técnicos como “radiação horizontal difusa” e “durabilidade potencial compatível”. Ambas as expressões são compostas pelo padrão <NOUN, ADJ, ADJ> e estão presentes nas frases do grupo de baixa acurácia. Como não havia este padrão nos textos usados para a construção dos padrões, é natural que a extração tenha sido incompleta para os termos “radiação normal”, “radiação horizontal” e “durabilidade potencial”.

Outro problema identificado nas frases selecionadas pelo método do agrupamento é a incapacidade da biblioteca *SpaCy* em reconhecer uma propriedade de língua portuguesa de converter verbos flexionados no particípio passado em adjetivos. Por exemplo, a expressão “projetos elaborados a partir da exigibilidade desta parte da ABNT” possui o verbo “elaborar” flexionado no particípio passado, que passa a adotar a função gramatical de adjetivo e não mais de verbo na frase.

Na língua portuguesa, a rotulagem por atributos linguísticos é desafiadora devido à flexibilidade das palavras em desempenhar diferentes funções gramaticais dependendo do contexto da sentença. Uma mesma palavra pode atuar como um determinado tipo de classe gramatical em uma sentença e assumir uma função diferente em outra sentença, dependendo da ordem e da interação com os demais termos que compõem o contexto linguístico (Osorio Sanz, 2023).

Desta forma a elaboração de padrões para cada conjunto de sentenças é uma tarefa onerosa em diversas escalas, tanto pela dificuldade de codificação, tempo para codificar todas as sentenças possíveis e necessidade de atualização periódica com surgimento de novas expressões e termos técnicos.

## 7.5. Comparativo LLM

Na Seção 2.7.5 foi apresentada a diferença entre os dois paradigmas atuais sobre extração de informações, o baseado em regras e o baseado em aprendizado de máquina.

A fim de analisar a capacidade do método baseado em regras, esta seção explora a abordagem baseada em aprendizado de máquina, mais especificamente o uso de grandes modelos de linguagem (*Large Language Model* - LLM) para o objetivo que esta pesquisa se propôs.

A partir de 2023, os LLMs têm ganhado evidência devido à sua capacidade de capturar o conhecimento semântico, sintático e pragmático<sup>34</sup> da linguagem e transferir este conhecimento para ações relacionadas à linguagem como a extração de informações e geração de textos. Desta forma, é oportuno experimentar a comparação entre as abordagens inserindo uma comparação entre os resultados obtidos pelo processo baseado em regras, com supervisão de um ser humano e o processo baseado em grandes modelos de linguagem pré-treinados. Esta comparação apresenta uma visão geral das ferramentas disponíveis e a capacidade de análise para a extração e conversão de trechos da Norma.

### 7.5.1. Análise da abordagem baseada em grandes modelos de linguagem (LLMs)

Existem diversos LLMs disponíveis para uso e desenvolvimento de soluções. Dentre todos, destaca-se o ChatGPT, modelo criado pela empresa OpenAI a partir do modelo de linguagem GPT-3, que tem se tornado o principal modelo generativo no atual estado da arte. O modelo tem sido usado em *benchmarks* de outros modelos como o LLaMA 2, BERT, PaLM e BLOOM. Nesta sequência de análise foram executadas diferentes comparações do resultado de extração com o modelo ChatGPT, na versão 3.5. Atualmente o modelo ChatGPT 4 representa o estado da arte dos LLM, mas é um modelo de acesso

---

<sup>34</sup> O estudo de como a linguagem é usada em seu contexto.

restrito. A ferramenta Microsoft Bing anunciou que utiliza o modelo ChatGPT em seus prompts<sup>35</sup>.

Para análise foram usadas as técnicas de *prompt* como personificação e sequenciamento de ações, tanto pela via interface *web* e API. A descrição de cada prompt pode ser observada no Apêndice N.

## 7.5.2. Síntese do resultado da comparação com LLM

**Quadro 37 - Comparação das abordagens para extração de informações de um trecho da Norma de Desempenho**

1	<p><b>Gold Standard:</b> Devem ser previstos nos projetos a prevenção de infiltração da água de chuva e da umidade do solo nas habitações, por meio dos detalhes indicados a seguir: a) condições de implantação dos conjuntos habitacionais, de forma a drenar adequadamente a água de chuva incidente em ruas internas, lotes vizinhos ou mesmo no entorno próximo ao conjunto; b) sistemas que impossibilitem a penetração de líquidos ou umidades de porões e subsolos, jardins contíguos às fachadas e quaisquer paredes em contato com o solo, ou pelo direcionamento das águas, sem prejuízo da utilização do ambiente e dos sistemas correlatos e sem comprometer a segurança estrutural. No caso de haver sistemas de impermeabilização, estes devem seguir a ABNT NBR 9575; c) sistemas que impossibilitem a penetração de líquidos ou umidades em fundações e pisos em contato com o solo; d) ligação entre os diversos elementos da construção (como paredes e estrutura, telhado e paredes, corpo principal e pisos ou calçadas laterais).</p>
2	<p><b>RuleBased:</b> Devem ser previstos nos projetos a prevenção de infiltração da água de chuva e da umidade do solo nas habitações, por meio dos detalhes indicados a seguir: a) condições de implantação dos conjuntos habitacionais, de forma a drenar adequadamente a água de chuva incidente em ruas internas, lotes vizinhos ou mesmo no entorno próximo ao conjunto; b) sistemas que impossibilitem a penetração de líquidos ou umidades de porões e subsolos, jardins contíguos às fachadas e quaisquer paredes em contato com o solo, ou pelo direcionamento das águas, sem prejuízo da utilização do ambiente e dos sistemas correlatos e sem comprometer a segurança estrutural. No caso de haver sistemas de impermeabilização, estes devem seguir a ABNT NBR 9575; c) sistemas que impossibilitem a penetração de líquidos ou umidades em fundações e pisos em contato com o solo; d) ligação entre os diversos elementos da construção (como paredes e estrutura, telhado e paredes, corpo principal e pisos ou calçadas laterais).</p>
3	<p><b>GPT 3.5 via API:</b> Devem ser previstos nos projetos a prevenção de infiltração da água de chuva e da umidade do solo nas habitações, por meio dos detalhes indicados a seguir: a) condições de implantação dos conjuntos habitacionais, de forma a drenar adequadamente a água de chuva incidente em ruas internas, lotes vizinhos ou mesmo no entorno próximo ao conjunto; b) sistemas que impossibilitem a penetração de líquidos ou umidades de porões e subsolos, jardins contíguos às fachadas e quaisquer paredes em contato com o solo, ou pelo direcionamento das águas, sem prejuízo da utilização do ambiente e dos sistemas correlatos e sem comprometer a segurança estrutural. No caso de haver sistemas de impermeabilização, estes devem seguir a ABNT NBR 9575; c) sistemas que impossibilitem a penetração de líquidos ou umidades em fundações e pisos em contato com o solo; d) ligação entre os diversos elementos da construção (como paredes e estrutura, telhado e paredes, corpo principal e pisos ou calçadas laterais).</p>
4	<p><b>Microsoft Bing/ChatGPT4:</b> Devem ser previstos nos projetos a prevenção de infiltração da água de chuva e da umidade do solo nas habitações, por meio dos detalhes indicados a seguir: a) condições de implantação dos conjuntos habitacionais, de forma a drenar adequadamente a água de chuva incidente em ruas internas, lotes vizinhos ou mesmo no entorno próximo ao conjunto; b) sistemas que impossibilitem a penetração de líquidos ou umidades de porões e subsolo, jardins contíguos às fachadas e quaisquer paredes em contato com o solo, ou pelo direcionamento das águas, sem prejuízo da utilização do ambiente e dos sistemas correlatos e sem comprometer a segurança estrutural. No caso de haver sistemas de impermeabilização, {estes devem seguir} a ABNTNBR9575; c) sistemas que impossibilitem a penetração de líquidos ou umidades em fundações e pisos em contato com o solo; d) ligação entre os diversos elementos da construção (como paredes e estrutura, telhado e paredes, corpo principal e pisos ou calçadas laterais).</p>

Fonte: o autor.

<sup>35</sup> [https://blogs.bing.com/search/march\\_2023/Confirmed-the-new-Bing-runs-on-OpenAI%E2%80%99s-GPT-4](https://blogs.bing.com/search/march_2023/Confirmed-the-new-Bing-runs-on-OpenAI%E2%80%99s-GPT-4)

É visível no Quadro 37 que os dois grandes modelos de linguagem (GPT 3.5 e Bing/GPT4) não retornaram a classificação da maioria dos termos. Enquanto o método baseado em regras obteve 23 objetos semânticos, o ChatGPT 3.5 encontrou 15, o que significou que o método baseado em regras obteve desempenho 53% melhor que o LLM para essa frase. Em frases menores o LLM obteve melhor desempenho, mas apresenta dificuldade em deduzir termos que não são comuns. Enquanto o modelo baseado em regras é aberto e auditável, ou seja, as regras podem ser verificadas e ajustadas, ele não seria eficiente para generalização em qualquer texto. Por outro lado, o modelo baseado em regras tem prejuízo na escalabilidade da solução, enquanto o modelo baseado em IA, com ajuste fino por meio de prompts adequados e treinamento em um domínio específico, pode ser inserido em qualquer contexto.

Apesar da sua restrição em escala, o modelo baseado em regras pode manter padrões específicos e controlar ajustes nos padrões. Como os textos de regulamentos possuem escrita similar em relação ao uso de termos e estrutura textual, ele possui boa capacidade de atuação. Já em um cenário de ampla variabilidade de conteúdo dos textos, o modelo baseado em regras é ineficaz porque precisaria de diversas regras e construções linguísticas para cada contexto.

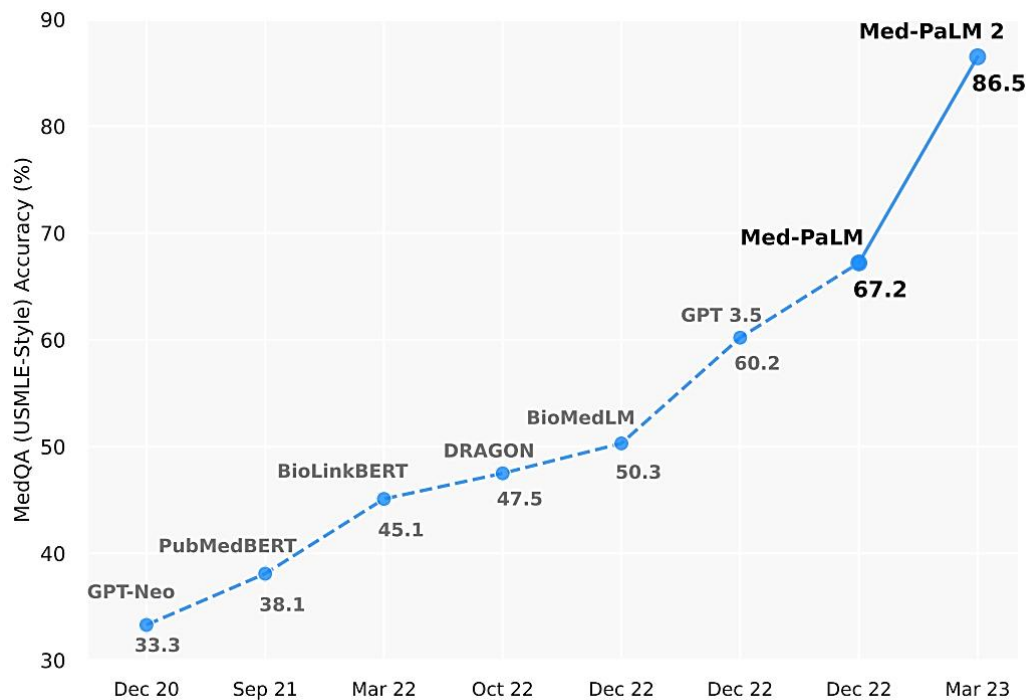
Com relação ao custo de implementação, a abordagem baseada em regras exige pouco recurso computacional e poderia ser implementada em diferentes aplicações. Já o baseado em IA demanda maior poder computacional para treinar e ajustar o modelo no contexto da norma que pretende processar.

### **7.5.3. Ajuste fino dos grandes modelos de linguagem**

No estado da arte atual, o resultado de extração de informações usando modelos de linguagem pode ser melhorado com duas estratégias principais. O ajuste fino (*fine-tuning*) do modelo de linguagem e a incorporação de uma base de conhecimento como suporte à extração.

O modelo Med-PaLM 2, criado pelo Google (Singhal *et al.*, 2023) é uma referência sobre uso de *fine-tuning* para melhoria de acurácia do modelo. Ele foi retreinado a partir de textos médicos alinhados ao domínio de conhecimento médico. É o primeiro LLM a participar de um teste de especialista oficial, atingindo 85% de acurácia nas respostas do exame de *US Medical Licensing Examination*<sup>36</sup> (Singhal *et al.*, 2023). A Figura 72 exemplifica o incremento de ganho de qualidade nos resultados com o ajuste fino dos modelos de linguagem.

Figura 72 - Comparação da acurácia do modelo MedPaLM 2 com outros modelos após o ajuste fino



Fonte: adaptado de Singhal *et al.* (2023).

O uso de LLM para ações muito específicas, como conseguir analisar e relacionar termos incomuns semanticamente, ainda é uma tarefa que não retorna resultados desejados. Apesar dos modelos serem treinados em bilhões de parâmetros e manterem

<sup>36</sup> <https://cloud.google.com/blog/topics/healthcare-life-sciences/sharing-google-med-palm-2-medical-large-language-model>

coerência nas respostas, a precisão dos modelos ainda é um assunto em desenvolvimento. A arquitetura de um LLM é composta por camadas e modelos de grande desempenho, mas dependem da tokenização e transformação dos *tokens* em vetores para identificar e gerar sentenças de texto.

A criação de vetores que representam as palavras é chamada de *embedding* que, em termos gerais, transforma uma palavra em uma matriz de valores que indica quão próxima ela está de determinadas palavras. Desta forma, ao mencionar uma palavra, é possível inferir qual palavra provavelmente seguirá na sequência da frase com base na matriz que a representa. Por exemplo, a palavra "cachorro" é mais frequentemente encontrada em frases que contêm a palavra "gato" do que a palavra "telhado", embora "cachorro" também seja um elemento construtivo de beirais de telhados. Essa ambiguidade de sentido é comum se a palavra não estiver vinculada ao seu contexto. Durante o treinamento dos LLM, as palavras são convertidas em matrizes de vetores com dezenas ou centenas de dimensões, representando a participação e proximidade de cada palavra com outros elementos. Esta proximidade entre termos é dada pelo corpus de treinamento do modelo.

No exemplo acima, se não existirem textos relacionados à beirais de telhados, a palavra "cachorro" sempre será assimilada a um animal. Embora essa representação n-dimensional demande uma capacidade de armazenamento significativa, impedindo o treinamento desses modelos em computadores pessoais, ela garante a associação entre a palavra e seu contexto, mensurável pelo cálculo da similaridade do cosseno (Xu; Xu; Mandic, 2023).

No entanto, os LLM são treinados em bancos de dados gerais, de contexto amplo e não especializado. Portanto, palavras como "cachorro" e "beiral" teriam uma probabilidade menor de serem consideradas semanticamente similares e, conseqüentemente, de serem sugeridas em um preenchimento automático de texto.

Além disso, esses modelos têm dificuldade em lidar com um espaço de busca de *tokens* muito amplo devido à sua capacidade limitada de processar milhares de *tokens* de

cada vez. O conceito de "janela de contexto" (Walter; Bromhead, 2023) explica que os grandes modelos de linguagem tendem a responder de forma ampla a *prompts* redigidos de maneira simples ou sem uma definição clara do contexto devido a essa limitação do espaço de busca.

Portanto, é recomendável que o contexto de ação seja declarado a cada chamada do modelo para que a "janela de contexto" restrinja sua busca por palavras e relacionamentos dentro do contexto desejado. Caso contrário, o modelo tenderá à resposta que usam termos mais gerais, porque essas palavras possuem vínculos mais claros com outras palavras comuns encontradas nos textos de treinamento.

Portanto, para garantir que um LLM tenha capacidade de reconhecer padrões e termos específicos, ele precisa ser treinado com um novo corpus previamente rotulado que represente a presença de determinadas palavras em contextos pouco conhecidos ou que termos não usuais como "porta corta-fogo", seja entendido como um único elemento.

O ajuste fino pode ser feito tanto pela calibragem dos parâmetros de um modelo pré-treinado como um novo treinamento do modelo a partir de um conjunto menor de dados. O novo treinamento permite que o conhecimento já adquirido no treinamento inicial seja direcionado para os novos dados apresentados. Uma das atividades beneficiadas com este ajuste é o Reconhecimento de Entidades Nomeadas (REN), que permite com que determinados termos e expressões sejam identificados no texto como cidades, pessoas ou termos técnicos. Após o treinamento em dados confiáveis, ainda é feito o ajuste dos hiper parâmetros como tamanho da amostra de dados, força de regularização, ajuste do algoritmo de otimização e taxa de aprendizado. O ajuste fino, portanto, procura ajustar o modelo para simular um comportamento ou adaptá-lo a uma tarefa.

O ajuste pode direcionar o resultado para o caminho oposto e ter seu desempenho reduzido por ter um sobreajuste (*overfitting*) aos dados apresentados e perder sua capacidade de generalização em dados não vistos pelo modelo; por esquecer o

conhecimento anterior usado durante o treinamento original e prejudicar o reconhecimento com a ampliação de vieses do modelo original.

Algumas soluções ainda podem ser executadas *a posteriori* com a engenharia de *prompts* e uso de bibliotecas que incorporam algumas técnicas de engenharia de *prompts*. Os LLMs, assim como seres humanos, não são capazes de fornecer respostas específicas sobre questões amplas, por exemplo, “Como redigir uma boa norma para classificação de modelos”. Para reduzir a janela de contexto sugere-se adotar técnicas de engenharia de *prompt* com segmentar a pergunta em questões mais específicas, exemplificar, inserir modelos padrão de resposta e tópicos relacionados aos quais o modelo deve se ater.

O método de engenharia de *prompts*, portanto, desenvolve estruturas de texto que conseguem ampliar a acurácia do resultado de modelos conversacionais generativos como ChatGPT, StableDifusion, Code Llama, entre outros. Alguns dos recursos da engenharia de *prompts* estão relacionados a formar uma sequência de comandos claros e bem definidos (*chain of thought*) para formular a solicitação ao modelo (Zhang *et al.*, 2022; Wei *et al.*, 2022). Após a engenharia de *prompts* e ajuste fino do modelo, ainda é possível efetuar uma melhoria do resultado a partir do oferecimento de recompensas ao modelo quando os resultados são satisfatórios. Com isso, ele aprenderá como responder para gerar resultados com mais qualidade. Em seguida, é possível ainda incrementar o processo adicionando uma camada de aprendizagem por reforço que maximiza a obtenção de recompensas por oferecer resultados melhores.

Uma vez que a estratégia de ajustar o modelo ainda é baseada em adivinhação de conteúdo, existe a chance de que os resultados não possuam acurácia suficiente ou que existam “alucinações” do modelo. “Alucinações” é o nome dado para respostas coerentes, porém inverídicas sobre o assunto<sup>37</sup>. Outra opção que tem ganhado destaque para

---

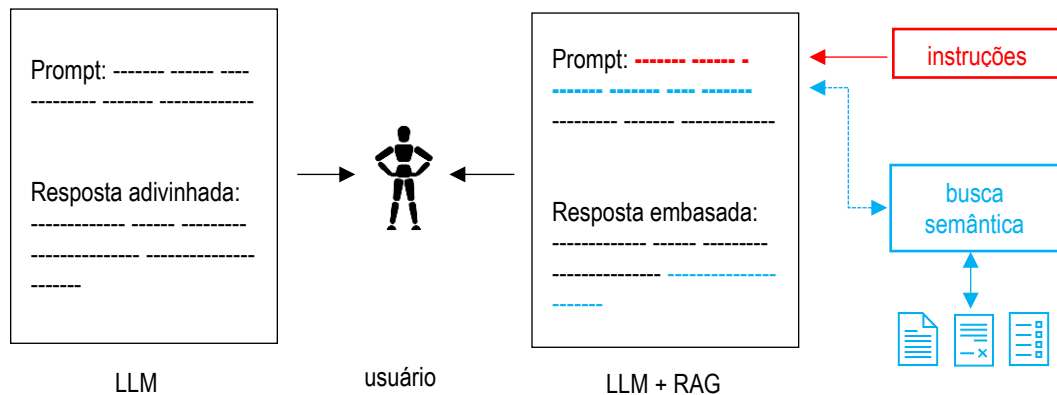
<sup>37</sup> <https://cdn.openai.com/papers/gpt-4-system-card.pdf>

conseguir que LLM consigam analisar e retornar com precisão informações específicas, é a incorporação uma base de conhecimento à busca do LLM.

#### 7.5.4. Uso de uma base de conhecimento, a aplicação do *framework* LangChain

O processo completo de busca em uma base de conhecimento é baseado no método *Retrieval-Augmented Generation* (RAG) (Lewis, 2020), cuja explicação pode ser visualizada na Figura 73.

Figura 73 - Modelo de funcionamento de um RAG comparado com um LLM isolado



Fonte: o autor.

Enquanto um LLM fornece respostas ao usuário com base no treinamento recebido, um LLM associado à técnica RAG implementa um processo orientado à geração de informações a partir de uma base de conhecimento. Em linhas gerais, o modelo RAG opera da seguinte maneira:

1. Vetorização do(s) documento(s), criando *embeddings* para representá-lo vetorialmente;
2. Configuração do LLM e inserção de um texto que sirva como prompt, estabelecendo campos para receber a resposta;
3. Execução de uma busca baseada em similaridade na base de dados, orientando a busca do LLM;

4. Retorno da informação ao usuário com evidências verídicas sobre sua descoberta.

Assim, ao realizar uma busca no modelo de linguagem, é ativada uma função para buscar a similaridade de conteúdo na base de conhecimento. O conteúdo é reduzido e apresentado ao LLM apenas os dados relevantes. Com base nestes dados, o LLM gera uma resposta.

O *framework* LangChain, lançado em 2023, está sendo usado para recuperação de informação a partir de textos e se baseia na técnica RAG. As vantagens desta técnica são que a RAG não demanda uma base de dados anotada, como nos modelos convencionais. Além disso, não demanda o ajuste fino do modelo e, conforme baseado em uma função para extração de dados da base indicada, não gera conteúdo, mas organiza o conteúdo existente, diminuindo ou eliminando as alucinações do modelo de linguagem.

Portanto, a aplicação da técnica RAG assegura precisão dos dados extraídos, quando são possíveis de serem extraídos, pois o LLM assume o papel de “motor de busca e geração de conteúdo” fundamentado nos dados fornecidos. Ademais, em caso de necessidade de aprimoramento do resultado, é possível incluir novos documentos para enriquecer a pesquisa do modelo.

O limite desta técnica é a necessidade de dividir o documento em partes menores e vetorizar o texto. A divisão em partes menores, ou *chunks*, é necessária para atender ao limite de processamento dos LLMs (em torno de 4.000 *tokens*).

O código no Quadro 38 expressa a simplicidade e capacidade do *framework* LangChain em extrair as informações corretas e em formato estruturado de um trecho complexo da Norma de Desempenho. O resultado obtido surpreende pela simplicidade do código fornecido e pela acurácia do resultado. Foram exigidos treze campos de análise e o modelo retornou todos os termos corretamente atingindo o valor máximo na medida valor-F1. Como pode ser observado no código de entrada, o *framework* se baseia na criação de funções e não mais em *prompts* textuais para obter informações do modelo LLM.

**Quadro 38 - Código que acessa o framework LangChain para executar uma extração de informações usando LLM**

```
# Schema
schema = {
  "properties": {
    "componente_edificio": {"type": "string"},
    "existe_menção_à_norma": {"type": "boolean"},
    "valor_exigência": {"type": "integer"},
    "unidade_de_medida": {"type": "string"},
  },
  "required": ["componente_edificio", "existe_menção_à_norma", "valor_exigência", "unidade_de_medida"],
}

# Run chain
llm = ChatOpenAI(temperature=0, model="gpt-3.5-turbo", max_tokens=4000)

# Input
inp = """Em vestíbulos, halls, corredores, instalações sanitárias e despensas, é permitido que o pé-direito seja reduzido ao mínimo de 2,30 m. Nos tetos com vigas, inclinados, abobadados ou, em geral, contendo superfícies salientes na altura piso a piso e/ou o pé-direito mínimo, devem ser mantidos pelo menos 80 % da superfície do teto, permitindo-se na superfície restante que o pé-direito livre possa descer até o mínimo de 2,30 m."""

chain = create_extraction_chain(schema, llm)
output = chain.invoke(inp)

>>>Resultado
[{'componente_edificio': 'vestíbulos, halls, corredores, instalações sanitárias e despensas',
  'existe_menção_à_norma': False,
  'valor_exigência': 2.3,
  'unidade_de_medida': 'm'},
 {'componente_edificio': 'tetos com vigas, inclinados, abobadados ou, em geral, contendo superfícies salientes na altura piso a piso e/ou o pé-direito mínimo',
  'existe_menção_à_norma': False,
  'valor_exigência': 80,
  'unidade_de_medida': '%'},
 {'componente_edificio': 'superfície restante',
  'existe_menção_à_norma': False,
  'valor_exigência': 2.3,
  'unidade_de_medida': 'm'}]
```

Fonte: o autor.

O limite imediato que se apresenta no uso do *framework* LangChain é que o mesmo esquema que estrutura os dados a serem extraídos não pode ser usado para toda a norma dada a variabilidade de estruturas frasais. Por outro lado, caso a sentença se assemelhe a regras que exigem composição ou encadeamento de elementos construtivos e condições

de uso, como as regras existenciais de Zhang (2015), seria formatado um novo esquema de rótulos.

#### 7.5.4.1. Comparação entre técnicas de melhoria de resposta dos LLMs

Gao *et al.* (2023) selecionam nove características para comparação de uso dos métodos de **ajuste fino** e o **baseado em base de conhecimento** para recuperação e geração de informação. A síntese das características pode ser observada no Quadro 39.

**Quadro 39 - Quadro comparativo do desempenho de ajuste fino e uso do método RAG<sup>38</sup>**

Característica	RAG	Ajuste fino
Atualização dos dados	A inserção de novos dados é automática ou simples de ser convertida em vetores.	Necessita novo treinamento do modelo.
Dados externos e problemas éticos e de privacidade	É capaz de usar fontes de dados externos, estruturados ou não. O controle de privacidade e ética é feito no banco de dados apresentado.	Podem ser apresentados para treinamento, mas não é viável se tiver atualização frequente. O controle de privacidade e ética pode ser dificultado pelos dados fornecidos antes do ajuste e pelas alucinações do modelo.
Processamento dos dados	Mínimo pré-processamento e possui ferramentas prontas para vetorização.	Dados devem ser filtrados, analisados e selecionados para garantir qualidade do aprendizado.
Personalização do modelo	Não possui flexibilidade para ser usado para qualquer tipo de uso, deve ser orientado à recuperação de informação.	Pode absorver estilos de escrita e gerar informações para usos diversos dentro do cenário para qual foi re-treinado.
Interpretabilidade	Permite algum rastreamento até os dados originais usados para geração do contexto de busca.	Não é possível rastrear a forma como os dados são obtidos.
Recursos computacionais	Necessita da manutenção e atualização do banco de dados.	Após feito o ajuste, os dados usados para treinamento não são mais necessários, mas é importante mantê-los para atualizações e ajustes finos futuros.
Tempo de resposta (latência)	Maior tempo de resposta para busca em dados.	Baixo tempo de latência porque não envolve a busca, mas geração de dados.
Redução de alucinações	Menos propenso a alucinações porque é baseado em evidências fornecidas ao modelo.	Reduzida em comparação com LLMs sem o ajuste, mas ainda alta quando enfrenta assuntos que não são familiares aos dados de treinamento.

Fonte: Adaptado de Gao *et al.* (2023).

<sup>38</sup> Este quadro está presente na primeira versão do artigo de Gao *et al.* (2023), mas vem sofrendo atualizações e pode ser obtido pelo link: <https://arxiv.org/abs/2312.10997v5>

A comparação entre as duas técnicas de utilização dos LLMs mostra que a geração de conteúdo orientada pela técnica RAG apresenta vantagens significativas, visto que dispensa a necessidade de treinar o modelo em conjuntos extensos de dados rotulados.

O importante neste processo é que o uso deste *framework* põe em questão a aplicação de ontologias como bases de dados semânticos em ações de resolução de entidades. O estado da arte em LLM ainda movimentava o questionamento sobre as ontologias e outros métodos baseados em regras para a extração de informações, sobretudo para tradução de textos criados em linguagem natural para outros formatos.

O estado da arte do uso de LLM para extração de informações tem apresentado incorporações de RAG para extração de informações e geração de prompts. Desta forma, esta pesquisa executou algumas rotinas de programas para gerar os esquemas de dados no framework LangChain e poder comparar com os resultados obtidos até agora.

Os testes foram realizados com o modelo gpt-3.5-turbo-0125 da OpenAI. Inicialmente o teste tentou reproduzir a lógica de extração de informações de Zhang e El-Gohary (2016) que estabelecem nove critérios que podem ser extraídos em termos de regulamentos (Quadro 40).

**Quadro 40 - Exemplo de frase e seus termos semânticos definidos por Zhang e El-Gohary (2016)**

"The minimum net area of ventilation openings shall not be less than 1 ft2 for each 150 ft2 of crawl space area"
"subject": "ventilation openings",
"subject_restriction": "N/A",
"compliance_checking_attribute": "net area",
"deontic_operator_indicator": "shall not be",
"comparative_relation": "less than",
"quantity_value": "1",
"quantity_unit_reference": "ft2",
"quantity_restriction": "for each 150 ft2 of crawl space area"

Fonte: adaptado de Zhang e El-Gohary (2016).

A extração foi comparada para extração dos rótulos semânticos utilizados por Zhang e El-Gohary (2016). Foram selecionadas as sentenças do regulamento JGJ 120-2012 (Apêndice M) e algumas sentenças do Capítulo 19 do ICB 2009, usados pelos autores para

calibragem de padrões e posterior validação em outros regulamentos. Os Quadro 41 e Quadro 42 representam a análise realizada utilizando o modelo gpt-3.5-turbo-0125 (Apêndice N).

**Quadro 41 - Trecho do JGJ 120-2012 extraído com *framework* LangChain**

Frase				
<i>After full compacting times are completed, the loose soil on site surface layer shall be compacted and rolled with low energy full tamping in the tamping principle of lapping 1/5 ~ 1/3 of hammer diameter, and the site elevation after dynamic consolidation shall be measured.</i>				
Termos semânticos	A	B	B (cont.)	C
<i>comparative relation</i>	N/A	N/A	N/A	<i>measured</i>
<i>compliance checking attribute</i>	<i>compactness</i>	<i>soil compaction</i>	<i>elevation measurement</i>	<i>site elevation</i>
<i>deontic operator indicator</i>	<i>shall be</i>	<i>shall be</i>	<i>shall be</i>	<i>shall be</i>
<i>quantity restriction</i>	N/A	-	<i>measured</i>	<i>after dynamic consolidation</i>
<i>quantity unit reference</i>	N/A	<i>full tamping</i>	N/A	-
<i>quantity value</i>	N/A	<i>low energy</i>	N/A	N/A
<i>subject</i>	<i>loose soil</i>	<i>loose soil</i>	<i>site elevation</i>	<i>loose soil</i>
<i>subject restriction</i>	<i>on site surface layer</i>	<i>on site surface layer</i>	<i>after dynamic consolidation</i>	<i>on site surface layer</i>

Fonte: o autor.

**Quadro 42 - Trecho do ICB 2009, extraído com o *framework* LangChain**

Frase			
<i>When the total volume of a given class of concrete is less than 50 cubic yards (38 m3), strength tests are not required when evidence of satisfactory strength is submitted to and approved by the building official.</i>			
Termos semânticos	A	B	C
<i>comparative relation</i>	<i>less than</i>	<i>less than</i>	<i>less than</i>
<i>compliance checking attribute</i>	N/A	<i>strength tests</i>	<i>strength tests</i>
<i>deontic operator indicator</i>	<i>are not required</i>	<i>are not required</i>	<i>are not required</i>
<i>quantity restriction</i>	-	<i>38 m3</i>	<i>38 m3</i>
<i>quantity unit reference</i>	<i>cubic yards</i>	<i>cubic yards</i>	<i>cubic yards</i>
<i>quantity value</i>	<i>50</i>	<i>50</i>	<i>50</i>
<i>subject</i>	<i>total volume</i>	<i>total volume</i>	<i>total volume</i>
<i>subject restriction</i>	<i>of a given class of concrete</i>	<i>of a given class of concrete</i>	<i>of a given class of concrete</i>

Fonte: o autor.

O Quadro 41 e o Quadro 42 mostram exemplos de extração de informações baseadas nos rótulos semânticos desenvolvidos por Zhang e El-Gohary (2016). Diferente da geração de informações feitas nos LLM, o uso do *framework* LangChain mostra redução, mas não eliminação das alucinações do modelo. No Quadro 41 foi inserida a expressão “*soil compaction*” que pode ser deduzida a partir da frase fornecida, mas não existe objetivamente no texto original.

Apesar dos resultados apresentados, a técnica RAG emerge como um novo paradigma para conversão de regulamentos em formato computável, ampliando as abordagens de extração de informações para além das fundamentadas em regras, tanto as que estão orientadas por padrões linguísticos quanto as baseadas em ontologia.

Considerando os avanços que os LLMs têm recebido nos últimos dois anos a acessibilidade crescente de APIs e ferramentas de desenvolvimento, incluindo novos *frameworks* e *plugins*, eles abrem um campo ainda não explorado na indústria da construção civil, sobretudo no contexto brasileiro. Enquanto nas pesquisas analisadas todas as informações exigiam a conversão para o idioma português, a incorporação de LLMs aos processos BIM oferece uma rota mais direta para o desenvolvimento de novas ferramentas de acesso a modelos, bases de dados textuais e regulamentos pertinentes aos projetos.

## **7.6. Capacidade de enriquecimento do modelo com atributos IFC**

A versão digital e computável de um regulamento passa a ser oferecida como uma base de dados semânticos que permite ações como treinamento de modelos, relacionamento entre normas, extração de informações e análises de texto, bem como busca semântica do texto.

O arquivo em formato XML é expansível e pode incorporar novos rótulos e atributos. Isso possibilita que um sistema de busca e filtragem de conteúdo procure por determinados elementos, componentes e funções na construção de habitações e extraia informações relacionadas a esses itens.

Além disso, é possível que os sistemas de busca retornem metadados e alguns campos marcados, como capítulos, páginas, tópicos, tabelas e números das sentenças em que os itens estão presentes. A filtragem por funções ou componentes permite que verificações de itens de projeto sejam feitas por diferentes agentes e profissionais. Como

o arquivo e o sistema de busca são computacionalmente simples, podem ser adicionados como *plugins* para os modeladores.

As ferramentas atuais que desenvolvem avaliação de conformidade também permitem que sejam incorporadas regras em formato XML, devidamente estruturadas segundo os rótulos exigidos pelo programa. Neste sentido, a existência de um arquivo semanticamente rotulado, como o que foi produzido por esta pesquisa, permite que seus termos sejam mapeados em outros esquemas e formatos de dados, uma vez que a linguagem de marcação XML é neutra e extensível.

Além disso, a versão em XML permite o desenvolvimento e implementação de uma solução aberta e não proprietária, como a integração do IfcOpenShell<sup>39</sup> para analisar modelos IFC em linguagem Python em aplicações abertas e gratuitas como o Blender BIM<sup>40</sup>.

Iniciativas deste tipo contribuem para desenvolvimento da iniciativa OpenBIM<sup>41</sup>, de programação de código aberto e, conseqüentemente, incentivam novas ferramentas para o setor da construção civil. Uma vez que rotinas de código aberto são absorvidas pelo mercado, amplia-se a independência dos projetistas de plataformas privadas de modelagem e avaliação de conformidade.

Como regulamento digital, e em formato computável, a Norma pode atingir o nível 3 de digitalização de normas, conforme sugerido pela iniciativa ISO Smart. Com a capacidade de expansão e modelagem de informação obtida pelo formato XML, ainda é possível desenvolver o esquema para que novas informações sejam incorporadas permitindo que o último nível de processamento de normas seja alcançado.

---

<sup>39</sup> <https://ifcopenshell.org/>

<sup>40</sup> <https://blenderbim.org/>

<sup>41</sup> <https://www.buildingsmart.org/about/openbim/>

### 7.6.1. Construção de IDSs

Neste sentido, se apresenta o *Information Delivery Specification* (IDS) lançado pela BuildingSMART em 2023 (BuildingSMART, 2023?), é baseado em uma linguagem de marcação que formata a estrutura de arquivos se baseando na marcação e categorização de rótulos, normalmente chamados *tags* ou elementos. Um esquema XML é definido pelo Schema Definition (XSD) que descrevem e validam o conteúdo de um XML. Os esquemas são criados para fins específicos obedecendo a algumas regras de estrutura<sup>42</sup> que determinam uma notação própria dos dados, definindo a hierarquia entre os elementos e os atributos pertencentes a cada elemento XML.

O IDS surgiu como *Machine Readable Information Delivery Specifications*, mas foi posteriormente convertido para a nomenclatura *Information Delivery Specification*. Ele foi concebido com um enfoque prático, para ser diretamente incorporado no processo como um componente digital em formato mais próximo ao IFC e às especificações do usuário.

O IDS pretende organizar as informações necessárias e as regras para a entrega do produto conforme contratado, diferenciando-se do IDM (*Information Delivery Manual*) em seus fundamentos e usos. O IDM é um documento mais abrangente, que descreve o fluxo de informação e as exigências entre processos. Por outro lado, o IDS absorveu a função de registrar os requisitos do IFC para a troca de informação entre equipes. O IDS não define o fluxo da informação, mas especifica quais dados devem estar incluídos no modelo e qual o nível de informação exigido. Além disso, o IDS visa a automação do processo e é formalizado em uma estrutura de representação XML, enquanto o IDM é um arquivo amplo contendo diversos tipos de informações, como gráficos, planilhas e listas, e, portanto, deve ser entregue em um formato genérico e não estruturado, como PDF.

Em 2020, o IDS se apresentou como um meio de registrar a informação necessária para colaboração em um formato neutro, capaz de ser lido por máquinas e que aceitasse

---

<sup>42</sup> <https://github.com/buildingSMART/IDS/blob/master/Documentation/ids-metadata.md>

informações para além das categorias IFC. O IDS é um arquivo estruturado em XML que possui uma estrutura de dados predefinida.

Atualmente existem algumas ferramentas disponíveis para criação de arquivos IDS como o IDS Editor<sup>43</sup>, BIMcollab Nexus<sup>44</sup> e IDS converter<sup>45</sup>. Todas as ferramentas são baseadas em um preenchimento manual ou semiautomático (baseado em planilha .xlsx). Como exemplo, a ferramenta IDS Editor, permite preencher uma planilha online para geração do arquivo (Figura 74).

**Figura 74 - Interface de construção de um IDS pelo IDS Editor**

The screenshot displays the IDS Editor interface. At the top, there are seven metadata input fields: IFC schema (Any), Description (This IDS...), Author (name@example.co), Copyright (name@example.co), Creation date (2024-01-16), Purpose (BIM Requirements), and Milestone (As-built). Below these are two specification sections:

**1. Fire requirement specification**

All elements	of entity	lfcWall	+	should have	a material	TIMBER	+
	of predefined type	CLADDING	+		a property	Strength cla	from set
						Structural	of value
							C30

**2. Test specification**

All elements	of entity	lfcWall	+	should have	a predefined type	TEST	+
	of predefined type	TEST	+		a classification	TEST	from system
	of classification	TEST	+				TEST
	of property	TEST	+		a property	TEST	from set
							TEST
	of material	TEST	+		a material	TEST	+

Fonte: Tomczak (2022).

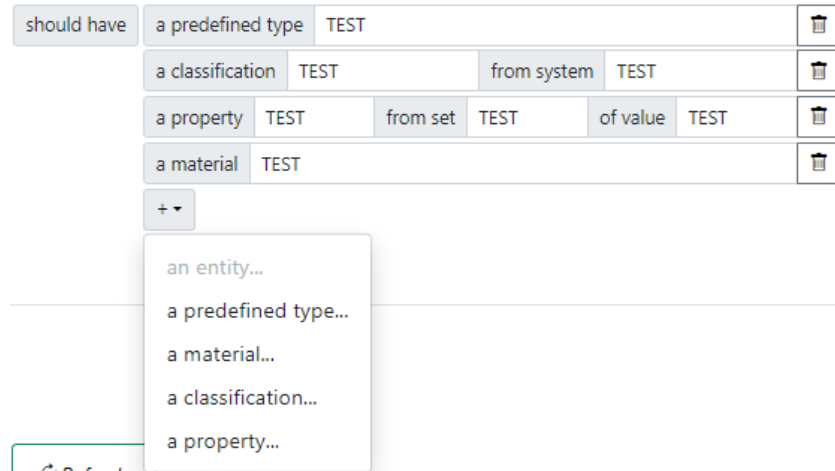
Os campos estão previamente definidos como planilha e com categorias pré-definidas (Figura 75). Desta forma o programa recebe os dados prontos e em uma formatação pré-definida pelo desenvolvedor. Ao final do processo é gerado um arquivo como apresentado pela Figura 76.

<sup>43</sup> <https://artomczak.pythonanywhere.com/create/>

<sup>44</sup> <https://www.bimcollab.com/>

<sup>45</sup> [https://github.com/c4rlosdias/ids\\_converter](https://github.com/c4rlosdias/ids_converter)

Figura 75 - Trecho da tela do IDS Editor representando os atributos a serem convertidos



Fonte: Tomczak (2022).

Figura 76 - Arquivo XML formatado no esquema IDS pelo programa IDSEditor

```
<!-- IDS (INFORMATION DELIVERY SPECIFICATION) CREATED USING IFCOPENSHELL -->
<ids xmlns="http://standards.buildingsmart.org/IDS" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://standards.buildingsmart.org/IDS/IDS.xsd" >
  <specification name="Fire requirement specification (IfcWall)" necessity="required">
    <applicability>
      <entity>
        <name>
          <simpleValue>IfcWall</simpleValue>
        </name>
        <predefinedtype>
          <simpleValue>CLADDING</simpleValue>
        </predefinedtype>
      </entity>
    </applicability>
    <requirements>
      <material location="any">
        <value>
          <simpleValue>TIMBER</simpleValue>
        </value>
      </material>
      <property location="any">
        <propertyset>
          <simpleValue>Structural</simpleValue>
        </propertyset>
        <name>
          <simpleValue>Strength class</simpleValue>
        </name>
        <value>
          <simpleValue>C30</simpleValue>
        </value>
      </property>
    </requirements>
  </specification>
</ids>
```

Fonte: o autor.

A construção de um IDS deve seguir algumas regras estabelecidas pela a BuildingSMART<sup>46</sup>; uma delas é a de incorporar o URI dos elementos mencionados.

<sup>46</sup> <https://github.com/buildingSMART/IDS/blob/master/Documentation/ids-metadata.md>

Neste sentido, para provar a viabilidade do método de conversão da Norma em formato computável foi criado um código<sup>47</sup> em Python que estende a representação do arquivo XML obtido pela extração de informações. A extensão é orientada para a criação de um arquivo IDS enriquecido com um novo atributo XML que recebe o valor URI de seu componente. Como a incorporação da informação URI do componente deve ser automática, o código executa uma busca no dicionário de dados da BuildingSMART, o *Building Smart Data Dictionary*<sup>48</sup>. A busca é feita pela API de acesso ao dicionário e efetua a busca por cada elemento XML rotulado como “COMPONENTE”.

Inicialmente o código verifica se o termo armazenado sob este rótulo está presente no dicionário da BuildingSMART bSDD. Em caso positivo, ele retorna o URI deste termo e incorpora esta informação no arquivo XML como um atributo do elemento “COMPONENTE” (Figura 77). Este enriquecimento com URI é uma recomendação da BuildingSMART<sup>49</sup> que permite que o XML seja convertido mais facilmente para um IDS.

**Figura 77 - Enriquecimento do arquivo XML com a uri do elemento "estrutura de concreto"**

```

▼ <FRASE>
  Atendimento às seguintes Normas de projeto estrutural: ABNT NBR 14323, para estruturas de aço; ABNT NBR 15200, para
  estruturas de concreto; para as demais estruturas, aplica-se o Eurocode correspondente, em sua última edição.
▼ <TERMOS>
  <REFERENCIA> ABNT NBR 14323 </REFERENCIA>
  <REFERENCIA> ABNT NBR 15200 </REFERENCIA>
  <REFERENCIA> Eurocode correspondente </REFERENCIA>
  <COMPONENTE> Normas de projeto estrutural </COMPONENTE>
  <COMPONENTE> Atendimento às seguintes </COMPONENTE>
  <COMPONENTE> estruturas de aço </COMPONENTE>
  <COMPONENTE uri="https://identifier.buildingsmart.org/uri/CD0BIM/CIS001/1.0/class/03.02.000"> estruturas de concreto
  </COMPONENTE>
</TERMOS>
</FRASE>

```

Fonte: o autor.

O sucesso da conversão não é expandido para todos os termos extraídos. Como o bSDD é uma coleção de dicionários, existem apenas dois dicionários disponíveis no banco de dados com termos em português. Portanto ainda é necessário expandir o dicionário incorporando outros elementos com seus respectivos atributos IFC.

<sup>47</sup> <https://zenodo.org/records/11315651>

<sup>48</sup> <https://github.com/buildingSMART/bSDD>

<sup>49</sup> <https://www.buildingsmart.org/what-is-information-delivery-specification-ids/>

## 8. DISCUSSÃO

A Norma de Desempenho é uma obra extensa, composta por diretrizes de temas e estruturas variadas. O método desenvolvido para a extração de informações baseada em regras e baseado em padrões linguísticos, demonstrou habilidade em rotular com precisão os elementos da Norma, superando alguns limites impostos pela natureza do próprio método e aprimorando-se ao longo do processo.

Na primeira etapa do método, realizou-se a conversão do texto no formato PDF para TXT, a fim de remover conteúdos desnecessários para o processamento, como legendas e elementos de apresentação. Por decisão da pesquisa, o texto convertido não foi segmentado em capítulos e não recebeu numeração de páginas. Contudo, a pesquisa demonstrou a viabilidade de extrair índices e tópicos da Norma por meio de expressões regulares e regras específicas.

Assim que foi iniciada a preparação do texto, com a remoção de itens desnecessários, a pesquisa evidenciou uma diferença significativa entre as abordagens baseadas em língua inglesa e que não são aplicáveis, em sua totalidade, na língua portuguesa. Enquanto a remoção de *stopwords* em língua inglesa é comum e necessária para melhoria da rotulagem dos dados em língua inglesa, os termos da construção civil brasileira possuem preposições e artigos necessários para a construção de seus termos. Desta forma, a extração de *stopwords* poderia levar a interpretações equivocadas sobre composição dos termos semânticos.

A rotulagem dos termos linguísticos e sua análise por meio de planilha facilitaram a construção dos padrões linguísticos que compõem a sentença. No entanto, observou-se a necessidade de incorporar atributos sintáticos e morfológicos para garantir uma melhor captura dos termos-chave das sentenças.

O processo de extração demandou revisão dos padrões para priorizar os padrões mais abrangentes, capazes de capturar termos compostos.

Como complemento ao método de extração, a pesquisa testou a incorporação de uma extração de informações baseada em ontologia para melhorar a qualidade dos resultados. Para tanto iniciou desenvolveu a análise de duas ontologias com mais relevância para o cenário da construção civil devido à sua capacidade de representação por múltiplas classes e a capacidade que possuem em ser alinhadas com novas ontologias.

Por outro lado, a técnica de análise semântica (explorada no Capítulo 4) mostrou que o processo de resolução das entidades, presentes tanto nos textos como na ontologia, depende da instanciação destes termos na ontologia, seja como elementos, instâncias, rótulos ou propriedades. Ainda que houvesse estas instanciações seria necessário que fossem traduzidas para a mesma língua do regulamento. Dada a complexidade da construção de uma ontologia ou instanciação de elementos, esta pesquisa optou por não incorporar uma ontologia para resolução de entidades.

Como alternativa ao incremento de acurácia do método de extração de informações, foram exploradas outras abordagens, como cálculo de similaridade entre conceitos e a incorporação de métodos baseados em modelos vetoriais de linguagem. No primeiro caso foi explorada a adoção da ontologia *Open Multilingual WordNet* e o modelo de linguagem BERT para avaliação de similaridade. A pesquisa identificou a limitação destas duas abordagens para melhoraria do resultado por não atingirem um valor significativo de similaridade, tanto com o cálculo baseada em profundidade de grafos usada na *Open Multilingual WordNet* ou pela similaridade de cosseno calculada com o modelo vetorial de linguagem BERT.

Os resultados obtidos com o algoritmo de extração de informações baseado exclusivamente em atributos linguísticos mostraram alta precisão ao identificar corretamente 90% dos termos capturados. Porém, comparado às pesquisas de referência (Zhang; Ma; Broyd, 2023), o modelo proposto por esta pesquisa não superou os métodos baseados em ontologia que alcançaram precisão e revocação acima de 94%.

Enquanto no Capítulo 8 da Norma de Desempenho o método proposto obteve 68% de revocação e 95% de precisão, o valor F1 foi de 79%. Isso indica que os padrões

classificaram corretamente os termos válidos, mas de todos os termos existentes nas frases, o método capturou 68% dos termos. Já na aplicação do método na Parte 1 da Norma, o método proposto também obteve alta precisão (90%), ainda confirmando ser capaz de identificar os itens corretamente. Contudo, a revocação do segundo resultado indicou que apenas 55% dos termos existentes foram capturados, e consequentemente, a valor F-1 foi de 68%.

O valor de especificidade confirmou que o modelo baseado em regras evitou que fossem capturados 82% de termos comuns das frases, desnecessários ao modelo XML, reforçando a qualidade do modelo em encontrar apenas os termos que são significantes para o usuário. A diferença entre os valores de 55% de e 82% devem-se ao fato que os itens capturados compunham palavras compostas como “contusões em função da operação” quando deveriam ser palavras simples “contusões” e “operação”. Desta forma, os termos extraídos não foram pontuados como corretos porque não eram úteis sem um pós-processamento.

Após análise das referências, sobretudo os apontamentos feitos por Ding *et al.* (2022), ficou evidente que pesquisas de referência, como Zhang e El-Gohary (2017), limitaram-se a extrair sentenças com que continuam até um elemento semântico de cada tipo (relação comparativa, atributo a ser verificado, valor, etc). O ICB 2009, por exemplo, é uma norma extensa, e só o Capítulo 19 foi usado como teste para validação do método dos autores, conta com diversas sentenças que fazem menção à termos existenciais e à outras normas. Por opção dos atores estas sentenças não foram consideradas, e restringiram suas análises a sentenças quantitativas ao final do Capítulo 19 do ICB. Essa informação pode ser comprovada no artigo de Zhang e El-Gohary (2016) que analisam o resultado de apenas 24 sentenças<sup>50</sup> do capítulo do ICB, o que representa, aproximadamente, 9% das sentenças existentes naquele capítulo.

---

<sup>50</sup> Os autores declaram em uma tabela que obtiveram 24 objetos na análise. Como o objeto é termo fundamental de cada sentença, é possível inferir que a quantidade máxima de frases analisadas é a mesma quantidade de objetos identificados.

Já esta pesquisa, na tentativa de construir um método para automação da conversão da Norma, abordou o regulamento em sua integralidade, sem a seleção prévia de sentenças ou segmentação de sentenças compostas. A análise de agrupamentos (Seção 6.4.1) revelou que a complexidade das sentenças é um dos principais limites do método e que concluíram em baixos índices nas métricas de qualidade.

O método não apresentou resultado satisfatório em sentenças compostas por índices de itens e sentenças (a, b, c...). Também enfrentou problemas na extração de termos encadeados como “instalações de água, esgoto ou água pluviais e estrutura, pisos e paredes” que não se adequavam aos padrões linguísticos construídos para localizá-los em sentenças comuns. Desta forma, o baixo valor de revocação obtido na aplicação do método na Parte 1 da Norma de Desempenho deve-se, em parte, à variabilidade dos requisitos, que existem mais padrões além dos fornecidos para abranger o conjunto de sentenças e assuntos representados nas diretrizes.

Comparativamente, Zhang e El-Gohary (2016) inicialmente desenvolveram 46 padrões linguísticos para mapear termos dos regulamentos e, posteriormente reduzidos para 22 com a incorporação do método PSG. Os autores testaram o método em testados em nove páginas de regulamento dentro do escopo de requisitos para concreto. Nesta pesquisa, foram desenvolvidos 29 padrões, mas que não foram sintetizados pelo método PSG. O método desenvolvido explorou as possibilidades das ferramentas disponíveis e incorporou na análise não apenas atributos gramaticais, mas também morfológicos e sintáticos. Como resultado os padrões construídos foram testados em 52 sentenças distribuídas em quarenta e oito páginas de diferentes contextos construtivos.

As sentenças complexas, com mais de um atributo por componente de verificação, são comuns e raramente são mencionadas nas pesquisas relacionadas à avaliação de conformidade. As pesquisas nacionais (Mainardi Neto, 2015, Silva, 2017; Soliman-Junior; Formoso; Tzortzopoulos, 2019; Kater; Ruschel, 2020) que se concentram na conversão de regulamentos para modelos de representação das ferramentas de verificação, como o Solibri, acusam baixo índice de conversão dos regulamentos, dada a subjetividade ou

complexidade dos textos e a incapacidade de tradução plena dos textos em sintaxes específicas exigidas pelo programa.

Para validação do resultado no estado da arte do processamento de linguagem natural, a pesquisa incluiu a exploração dos resultados obtidos por métodos baseados em grandes modelos de linguagem (LLM). O resultado apresentado na Seção 7.5.3 indicou que os LLMs, não conseguem atingir o desempenho equivalente com o modelo baseado em regras desenvolvido por esta pesquisa. Os modelos são acessados via *prompt* de comando pela interface do navegador ou por meio de API para automatizar e incorporar o modelo em outras plataformas. com técnicas de descrições, personificação, orientação por etapas e por exemplos (*one-shot*). O resultado mostrou que os modelos não compreendem as requisições e classificações de palavras do texto, obrigado que sejam ajustados com o retreinamento do modelo a partir de um corpus textual próximo do domínio ou que seja usado um método auxiliar para direcionar a análise. Como o ajuste fino depende de um nível de processamento computacional e dados previamente selecionados e rotulados, foi adotada a técnica RAG para permitir que o processamento do LLM seja suportado por dados específicos do domínio.

Desta forma, foi usado *framework* LangChain que incorpora técnicas como RAG e ReAct para orientação do modelo, minimizando os problemas existentes como alucinação do modelo. Os resultados da aplicação desta técnica evidenciaram os limites de uso de LLM para extração de informações de regulamentos mostrando que o modelo, apesar de identificar alguns itens corretamente, não é consistente na maioria das respostas. Como a consistência é um dos principais requisitos para gerar o formato computável, o uso dos modelos, no estado atual da arte, foi descartado.

Ao utilizar os procedimentos descritos nesta pesquisa, foi possível desenvolver um método que amplia o entendimento sobre a tradução de regulamentos para formatos computáveis. Apesar do baixo valor F1 das rotulagens, ficou evidente a possibilidade de automação da extração de termos da norma para sua conversão em um formato não proprietário. Por ser neutro e universal, o modelo computável se alinha à iniciativa

internacional ISO Smart permitindo que seja enriquecido com novos dados e em novas aplicações, como a geração de arquivos IDS a partir da incorporação de atributos IFC.

Ainda que o resultado da extração não tenha conseguido alta valor de medida F1, o processo de extração de informação foi efetivo e consistente, recuperando os dados repetidamente, com repetibilidade do resultado. Desta forma foi possível a conversão dos dados extraídos em elementos de um arquivo XML, marcado com os rótulos previamente definidos pelo método. Ao incorporar um método baseado em linguagem Python, em código aberto, esta pesquisa oferece um material atualizado e que pode ser explorado para a verificação automática de conformidade a regras, uma vez que as pesquisas relacionadas ao tema não possuem dados abertos. Seus códigos, arquivos e ontologias não estão disponíveis, dificultando a reprodutibilidade e o avanço destas pesquisas em outros cenários (Ding *et al.*, 2022; Zhou *et al.*, 2022).

O produto alcançado por esta pesquisa pode ser ampliado para outros regulamentos, mas depende de uma revisão dos padrões linguísticos. Além disso, demanda um novo mapeamento de padrões para ampliar a captura de dados neste novo contexto. Neste cenário o método tem escalabilidade da solução reduzida e deve ser avaliada a inserção de métodos complementares como o treinamento e orientação para os grandes modelos de linguagem. Em ambos os cenários a natureza do procedimento é idêntica, a de interpretar e extrair textos baseados em linguagem natural. Por outro lado, para criar um sistema baseado em conhecimento de maior potencial, seria mais adequada a criação de normas já em formato capaz de ser lido por máquina, como ontologia ou já rotulado. Desta forma, os métodos baseados em processamento de linguagem natural têm mais facilidade de modelagem do problema e interação com os dados existentes.

## 9. CONCLUSÃO

O método de extração de informações baseado em regras é o principal resultado dos procedimentos elaborados por esta pesquisa. Outros produtos desenvolvidos pela pesquisa contribuíram para a construção do método. O primeiro produto foi um conjunto de regras e procedimentos para extração de informação (**modelo**), que foram interativamente construídas ao longo do processo. Estas regras processaram um conjunto de padrões linguísticos, que representam os termos da primeira parte da Norma de Desempenho. Estes dois produtos serviram como dados de entrada para um código, criado em linguagem Python, que foi construído de modo iterativo, ajustando seus procedimentos a partir dos resultados e técnicas disponíveis pelas ferramentas selecionadas. Este código foi consolidado em um repositório aberto<sup>51</sup> e representa o algoritmo de extração de informações baseado em processamento de linguagem natural (**método**). Além de executar a extração de informações, o algoritmo converte as informações extraídas do texto em uma representação computável da Norma de Desempenho (instanciação do método), mas com desempenho limitado. Como resultado desta pesquisa, foi obtido um arquivo XML de todas os requisitos explícitos na Parte 1 da Norma de Desempenho 15.575/2021. O arquivo incorpora rótulos semânticos que categorizam componentes construtivos, as referências normativas e as exigências de cada diretriz estabelecida por este regulamento. Embora a completude dos termos esteja limitada, o arquivo gerado constitui **uma representação digital** possível da Norma de Desempenho e, portanto, tem a capacidade ser enriquecido com dados futuros, relacionados aos componentes identificados.

Adicionalmente, esta pesquisa destacou a importância de uma análise criteriosa sobre a remoção de *stop words* e pontuações, que deve ser adaptada ao contexto específico de aplicação do método proposto. Esta pesquisa identificou a necessidade de preservar

---

<sup>51</sup> <https://zenodo.org/records/11315651>

algumas preposições para garantir a composição de termos compostos do vocabulário da construção civil. Foi evidenciada a ineficácia na transposição direta de métodos aplicados na língua inglesa para o contexto brasileiro (pt-BR). Ao contrário da língua inglesa, é desnecessária e indevida a eliminação de pontuações e *stop words*. Nesta pesquisa ficou evidente que estas remoções não resultaram em melhorias significativas nos resultados, contrariando as expectativas baseadas em procedimentos típicos das pesquisas na língua inglesa. Além disso, esta pesquisa conclui que a remoção de stop words causa imprecisões e perda de acurácia do resultado porque segmenta e altera a composição de termos importantes que estão relacionados à avaliação de conformidade como “porta de correr”, “temperatura de bulbo seco” e “vida útil de projeto”.

A pesquisa apontou para a possibilidade de rápida conversão dos padrões linguísticos para um arquivo JSON (constructo), apto a ser usado por outras pesquisas e em formato aberto. Ao final do desenvolvimento do método de pesquisa, o arquivo que representa o modelo final da Norma, em XML, foi enriquecido evidenciando a possibilidade de incorporação automática de dados da bSDD para formalização de um arquivo IDS.

Além dos testes de extração de informação em LLMs, esta pesquisa apresentou os resultados obtidos por outro paradigma de extração de informações, baseado no *framework* LangChain. A adoção do framework com o modelo gpt-3.5-turbo-0125 permitiu extrair o texto do regulamento com acurácia em poucas frases, de caráter objetivo, e com pouco esforço especializado de programação. Porém, o processo de extração apresentou inconsistência, impedindo sua utilização em programas que desejem incorporar esta técnica para extração de seus regulamentos. Além da baixa repetibilidade dos resultados, foi identificado que o modelo de linguagem gerou resultados que alteraram o texto original. Desta forma, no estado atual da arte, não é possível adotar com confiança um método baseado em LLM.

Neste cenário, os modelos baseados em aprendizado de máquina possuem escalabilidade porque conseguem se adaptar com pouco esforço aos diferentes contextos

de sua aplicação. A aplicação de grandes modelos de linguagem, com a incorporação de técnicas auxiliares como RAG e ReAct mostram potencial de desenvolvimento para tradução de regulamentos em outros formatos. Por outro lado, ainda não podem garantir integridade do resultado. Já os modelos baseados em regras, mesmo desenvolvidos apenas sobre padrões linguísticos e regras declaradas, ainda se mostra com capacidade de reconhecer com precisão os termos exigidos pelas diretrizes. Um dos motivos que ainda mantém estes modelos em evidência é porque são métodos dedicados a um domínio específico e os regulamentos não possuem estrutura linguística complexa, com ampla variação formal.

A contribuição imediata do artefato gerado pela pesquisa é um fluxo de processos funcional, consistente (permite repetibilidade) e de baixo custo computacional que pode ser implementado em sistemas e rotinas para conversão de regulamentos. Além da pesquisa gerar uma representação em formato aberto e marcada, com uma sintaxe clara e facilmente lida por seres humanos, esta representação tem diferentes possibilidades de uso. O sistema desenvolvido facilita a conversão de padrões de busca em sugestões de novas normas, podendo ser integrado em plataformas específicas durante a fase de aceitação de produtos no mercado. Como analisado pela pesquisa, as ferramentas relacionadas com avaliação de conformidade dos projetos adotam em diferentes níveis o formato XML, facilitando sua integração no fluxo de verificação dos projetos. Tal integração permitiria uma avaliação mais acurada por profissionais especializados em modelagem, possibilitando a customização de padrões conforme as necessidades operacionais de diferentes escritórios.

A rotulagem de termos específicos de forma automática contribui para a análise dos documentos ampliando a compreensão sobre normas complexas com a NBR15.575. A rotulagem permite a criação de um arquivo computável que, implementado em um sistema amigável ao usuário, habilita consultas semânticas e a ordenação da norma segundo conceitos-chave de interesse do usuário. Este nível de computação da norma é desejado pela ISO Smart e é possível de ser implementado nos sistemas nacionais. Trata-

se, portanto, em não apenas criar um modelo que permite verificação completa da Norma para sistemas de conformidade, mas em criar um formato computável capaz de fornecer dados e informações significativas para qualquer sistema que interaja com estes regulamentos.

Como contribuição indireta, esta pesquisa aponta para dificuldades em classificação de elementos segundo os padrões linguísticos porque a redação da norma não estabelece um padrão orientado para sistemas digitais. Neste sentido, esta pesquisa sugere que as redações acompanhem protocolos claros que estejam orientados à criação de frases em sequências sintáticas comuns, com os termos encadeados em posições convencionais, facilitando a construção dos padrões gramaticais.

Por outro lado, mesmo superando os modelos atuais de IA, a metodologia empregada, que é baseada em padrões linguísticos, apresenta limitações relacionadas à sua aplicabilidade porque é restrita a contextos específicos. A abordagem baseada em regras é um método dedicado ao contexto de aplicação e regido por um conjunto terminológico limitado e demanda a elaboração de novos padrões para adaptar-se a variações estilísticas ou de domínio de aplicação. Ainda assim, a abordagem revela-se efetiva para textos normativos, caracterizados por estruturas bem definidas, porém com terminologia variável, desafiando sua escalabilidade para novos regulamentos quando suportada apenas por ontologias e modelos baseados em regras.

No contexto atual, a indústria busca solucionar desafios relacionados à precisão e à eficiência dos processos manuais de verificação de conformidade e à dificuldade de converter regulamentos em formatos computacionalmente acessíveis. A iniciativa ISO Smart representa um avanço promissor nessa direção, incentivando a criação de modelos de marcações neutros e universais, enriquecido com atributos específicos de modelos e processos como o que foi gerado pela pesquisa. Este modelo, portanto, aspira a estabelecer-se como um método confiável para profissionais, visando melhorar a gestão de normas e padrões em diversas áreas, por meio da promoção de maior eficiência e confiabilidade nos processos de verificação e conformidade.



## REFERENCIAS

ALEXOPOULOS, Panos. **Semantic modeling for data**. Sebastopol, CA: O'Reilly Media, 2020.

ALLEN, Robert H.; SRIRAM, Ram D. The role of standards in innovation. **Technological Forecasting and Social Change**, v. 64, n. 2-3, p. 171-181, 2000. [https://doi.org/10.1016/S0040-1625\(99\)00104-3](https://doi.org/10.1016/S0040-1625(99)00104-3)

ALMEIDA, Maurício Barcellos. **Ontologia em ciência da informação: tecnologia e aplicações**. Curitiba: CRV, 2021.

AMOR, Robert; DIMYADI, Johannes. The promise of automated compliance checking. **Developments in the Built Environment**, v. 5, 100039, 2021. <https://doi.org/10.1016/j.dibe.2020.100039>

ANDERY, Paulo; BARBOSA, Patrícia. Estudo sobre o impacto do SIAC (Sistema de Avaliação da Conformidade) na implementação da NBR 15.575:2013 em empresas construtoras. In: ENCONTRO NACIONAL DE TECNOLOGIA DO AMBIENTE CONSTRUÍDO, 17., 2018, Foz do Iguaçu. **Anais...** Porto Alegre: ANTAC, 2018.

ARBELAITZ, Olatz *et al.* An extensive comparative study of cluster validity indices. **Pattern Recognition**, v. 46, n. 1, p. 243-256, 2013. <https://doi.org/10.1016/j.patcog.2012.07.021>

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT ISO/IEC Guia 60: avaliação de conformidade - código de boas práticas**. Rio de Janeiro: ABNT, 2005.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR 15575-1: edificações habitacionais - desempenho**. 4.ed. Rio de Janeiro: ABNT, 2013.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR 15575-1: edificações habitacionais: desempenho**. Rio de Janeiro: ABNT, 2021

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **Sobre a normalização**. [abnt.org.br](https://abnt.org.br), 2022. Disponível em: <https://abnt.org.br/normalizacao/sobre-a-normalizacao/>. Acesso em: 06 fev. 2024.

AYDIN, Murat; YAMAN, Hakan. Domain knowledge representation languages and methods for building regulations. In: **Eurasian BIM Forum**. Springer, Cham, p. 101-121, 2019.

BEACH, Thomas H.; REZGUI, Y.; LI, H.; KASIM, T. A rule-based semantic approach for automated regulatory compliance in the construction sector. **Expert Systems with Applications**, v. 42, n. 12, p. 5219-5231, 2015. <https://doi.org/10.1016/j.eswa.2015.02.029>

BENEDIKT, M. L. To take hold of space, isovists and isovist fields. **Environment and Planning B: Urban Analytics and City Science**, v. 6, n. 1, p. 46-65, 1979. <https://doi.org/10.1068/b060047>

BOUZIDI, Khalil *et al.* Towards a semantic-based approach for the creation of technical regulatory documents in building industry. In: **eWork and eBusiness in Architecture, Engineering and Construction**, p. 417, 2010

BRASIL. **Lei 8.078, de 11 de setembro de 1990. Dispõe sobre a proteção do consumidor e dá outras providências.** Brasília, 1990. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/leis/l8078compilado.htm](https://www.planalto.gov.br/ccivil_03/leis/l8078compilado.htm). Acesso em: 04 fev. 2024.

BRASIL. Rede de Informação Legislativa e Jurídica. **LexML Brasil: parte 3 - LexML Schema.** Brasília, 2008. Disponível em: <https://projeto.lexml.gov.br/documentacao/Parte-3-XML-Schema.pdf>. Acesso em: 11 nov. 2022.

BRASIL. Ministério do Desenvolvimento, Indústria e Comércio Exterior. Conselho Nacional de Metrologia, Normalização e Qualidade Industrial (Conmetro). **Resolução nº 04, de 30 de abril de 2009.** Dispõe sobre aprovação da estratégia brasileira de normalização. Brasília: Conmetro, 2009. Disponível em: [http://www.inmetro.gov.br/qualidade/comites/pdf/Resolucao4\\_conmetro.pdf](http://www.inmetro.gov.br/qualidade/comites/pdf/Resolucao4_conmetro.pdf). Acesso em: 18 ago. 2022.

BRASIL. Ministério do Planejamento, Orçamento e Gestão. Secretaria de Logística e Tecnologia da Informação. **Cartilha técnica para publicação de dados abertos no Brasil v1.0.** Brasília, 2016. Disponível em: [http://www.governoaberto.sp.gov.br/wp-content/uploads/2016/02/Cartilha\\_Tecnica.pdf](http://www.governoaberto.sp.gov.br/wp-content/uploads/2016/02/Cartilha_Tecnica.pdf). Acesso em: 11 nov. 2022.

BRASIL. Ministério de Desenvolvimento Regional. **Portaria nº 577, de 30 de março de 2021.** Altera os anexos à Portaria n. 75, de 14 de janeiro de 2021, que restabelece o Sistema de Avaliação da Conformidade de Empresas de Serviços e Obras da Construção Civil (SiAC). Brasília, 2021a. Disponível em: <https://www.in.gov.br/en/web/dou/-/portaria-n-577-de-30-de-marco-de-2021-311657738>. Acesso em: 05 fev. 2024.

BRAUN, Richard *et al.* Proposal for requirements driven design science research. In: DONNELLAN, Brian; HELFERT, Markus; KENNEALLY, Jim; VANDERMEER, Debra; ROTHENBERGER, Marcus; WINTER, Robert (Eds.). **New horizons in design science: broadening the research agenda.** Cham, Switzerland: Springer, 2015. p. 135-151. (Lecture Notes in Computer Science, 9073). [https://doi.org/10.1007/978-3-319-18714-3\\_9](https://doi.org/10.1007/978-3-319-18714-3_9)

BROWN, Tom *et al.* Language models are few-shot learners. In: CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS, 34., 2020, Vancouver, Canada. **Proceedings...** Vancouver, Canada, 2020. p. 1877-1901. (Advances in Neural Information Processing Systems, 33).

BUCHHOLZ, Sabine; MARSÍ, Erwin. CoNLL-X shared task on multilingual dependency parsing. In: CONFERENCE ON COMPUTATIONAL NATURAL LANGUAGE LEARNING (CoNLL-X), 10., 2006, New York. **Proceedings...** p. 149-164. New York: Association for Computational Linguistics, 2006.

BDE. Use of data models within building performance simulation sector. **bde.ibpsa.us**, 2020. Disponível em: <https://bde.ibpsa.us/bde-surveys/>. Acesso em: 10 out. 2022.

BUILDINGSMART INTERNATIONAL. **OpenBIM principles**. 2020. Disponível em: [https://f3h3w7a5.rocketcdn.me/wp-content/uploads/2020/07/20200702\\_bSDD-trifold-flyer.pdf](https://f3h3w7a5.rocketcdn.me/wp-content/uploads/2020/07/20200702_bSDD-trifold-flyer.pdf). Acesso em: 08 jan. 2023.

BUILDINGSMART INTERNATIONAL. **BIM Collaboration Format (BCF): an introduction**. Disponível em: <https://technical.buildingsmart.org/standards/bcf/>. Acesso em: 08 out. 2022.

BUILDINGSMART INTERNATIONAL. **Information delivery specification IDS**. Fev. 2023. Disponível em: <https://technical.buildingsmart.org/projects/information-delivery-specification-ids/>. Acesso em: 10 dez. 2023.

CÂMARA BRASILEIRA DA INDÚSTRIA DA CONSTRUÇÃO. **Reunião do GT da Comissão de Estudo de Revisão da Norma de Desempenho (ABNT/CE-002:136.001) [Bloco1]**. 13 nov. 2019a. Disponível em: <https://www.youtube.com/watch?v=-sILb-uSS0I>. Acesso em: 15 out. 2022.

CÂMARA BRASILEIRA DA INDÚSTRIA DA CONSTRUÇÃO. **Reunião do GT da Comissão de Estudo de Revisão da Norma de Desempenho (ABNT/CE-002:136.001) [Bloco2]**. 13 nov. 2019b. Disponível em: <https://www.youtube.com/watch?v=qBhBALCYvq4>. Acesso em: 15 out. 2022.

CÂMARA BRASILEIRA DA INDÚSTRIA DA CONSTRUÇÃO. **Reunião do GT da Comissão de Estudo de Revisão da Norma de Desempenho (ABNT/CE-002:136.001) [Bloco3]**. 13 nov. 2019c. Disponível em: <https://www.youtube.com/watch?v=RkhQ1tTCFeo>. Acesso em: 15 out. 2022.

CÂMARA BRASILEIRA DA INDÚSTRIA DA CONSTRUÇÃO. Comissão de Materiais, Tecnologia, Qualidade e Produtividade (Comat). **ABNT alterou 203 normas afetas ao setor da construção em 2021**. 23 nov. 2021. Disponível em:

<https://cbic.org.br/inovacao/2021/11/23/abnt-alterou-203-normas-afetas-ao-setor-da-construcao-em-2021/>. Acesso em: 06 set. 2022.

CÂMARA BRASILEIRA DA INDÚSTRIA DA CONSTRUÇÃO. **Catálogo de normas técnicas**: edificações. Brasília: CBIC/SINDUSCON, 2022. Disponível em: <https://cbic.org.br/normasdaconstrucao/api/public/api/v1/uploads/2022/07/PwK9UHwkPTjO4ssi.pdf>. Acesso em: 06 set. 2022.

CÂMARA BRASILEIRA DA INDÚSTRIA DA CONSTRUÇÃO. **CBIC debate ações estratégicas para superar as barreiras de avanço do BIM**. 21 nov. 2023. Disponível em: [https://cbic.org.br/cbic-debate-acoes-estrategicas-para-superar-as-barreiras-de-avanco-do-bim/?utm\\_campaign=cbic\\_hoje\\_21112023&utm\\_medium=email&utm\\_source=RD+Station](https://cbic.org.br/cbic-debate-acoes-estrategicas-para-superar-as-barreiras-de-avanco-do-bim/?utm_campaign=cbic_hoje_21112023&utm_medium=email&utm_source=RD+Station). Acesso em: 22 nov. 2023.

CHENG, Chin *et al.* Improving access to and understanding of regulations through taxonomies. **Government Information Quarterly**, v. 26, n. 2, p. 238-245, 2009. <https://doi.org/10.1016/j.giq.2008.12.008>

CHINA, Republic. Ministry of Housing and Urban-Rural Development of the People's. **Technical specification for retaining and protection of building foundation excavations**. Beijing, China: MOHURD, 2012. (JGJ 120-2012).

CHOI, Jungsik; KIM, Inhan. A methodology of building code checking system for building permission based on openBIM. In: INTERNATIONAL SYMPOSIUM ON AUTOMATION AND ROBOTICS IN CONSTRUCTION, 34., 2017, Taipei, Taiwan. **Proceedings...** Edinburgh, UK: IAARC, 2017. p. 945-950.

CHOI, Yongsun; NGUYEN, Minh Duc; KERR JR., Thomas N. Syntactic and semantic information extraction from NPP procedures utilizing natural language processing integrated with rules. **Nuclear Engineering and Technology**, v. 53, n. 3, p. 866-878, 2021. <https://doi.org/10.1016/j.net.2020.08.010>

CHOMSKY, Noam. Three models for the description of language. **IRE Transactions on information theory**, v. 2, n. 3, p. 113-124, 1956. <https://doi.org/10.1109/TIT.1956.1056813>

CLARKE, Michael; HARLEY, Pam. How smart is your content? Using semantic enrichment to improve your user experience and your bottom line. **Science Editor**, v. 37, n. 2, p. 40-45, 2014.

COOK, Melanie *et al.* **NIST IR 7511 Rev. 4**: Security Content Automation Protocol (SCAP) version 1.2 validation program test requirements. Gaithersburg, MR: Computer Security Resource Center (CSRC), National Institute of Standards and Technology (NIST), 2016. <https://doi.org/10.6028/NIST.IR.7511r4>

DAI, Andrew M.; LE, Quoc V. Semi-supervised sequence learning. **Advances in neural information processing systems**, v. 28, 2015.

DEVLIN, Jacob et al. Bert: Pre-training of deep bidirectional transformers for language understanding. **arXiv preprint arXiv:1810.04805**, 2018.

DEUTSCHE KOMMISSION ELEKTROTECHNIK. Digitalization of standards: DKE digital strategy for the further process of transformation. 21 dez. 2020a. Disponível em: <https://www.dke.de/en/standards-and-specifications/standardization-strategy/digitalization-standards-digital-strategy-dke-transformation>. Acesso em: 27 dez. 2023.

DEUTSCHE KOMMISSION ELEKTROTECHNIK. **Standardization 2020 programme**. 2020b. Disponível em: <https://www.dke.de/en/standards-and-specifications/standardization-strategy/normung2020>. Acesso em: 27 dez. 2023.

DIMYADI, Johannes; AMOR, Robert. Automated building code compliance checking – where is it at? In: CIB WORLD BUILDING CONGRESS, 19., 2013, Brisbane, Austrália. **Proceedings...** Brisbane, Austrália: Construction and Society, 2013. p. 172-185.

DING, Yuexiong; MA, Jie; LUO, Xiaowei. Applications of natural language processing in construction. **Automation in Construction**, v. 136, p. 104169, 2022.

DRESCH, Aline; LACERDA, Daniel P., ANTUNES JR, Antônio V. **Design science research**. Springer International Publishing, 2015.

DROGEMULLER, Robin; WOODBURY, Rob; CRAWFORD, John. Extracting representation from structured text: initial steps. In: INTERNATIONAL CONFERENCE ON CONSTRUCTION INFORMATION TECHNOLOGY, 2000, Reykjavik, Islândia. **Proceedings...** Construction Informatics Digital Library, 2000. p. 302-300.

EASTMAN, Chuck. Automated assessment of early concept designs. **Architectural Design**, v. 79, p. 52-57, 2009. <https://doi.org/10.1002/ad.851>

EASTMAN, Charles *et al.* Automatic rule-based checking of building designs. In: **Automation and Construction**, v. 18, n. 8, p. 1011-1033, 2009. <https://doi.org/10.1016/j.autcon.2009.07.002>

EASTMAN, Chuck *et al.* **Manual de BIM**: um guia de modelagem da informação da construção para arquitetos, engenheiros, gerentes, construtores e incorporadores. Porto Alegre: Bookman, 2014.

EID, Asmaa S.; ABOULNAGA, Mohsen M.; MAHMOUD, Ayman H. Future cities for climate action: automated code compliance checking in reference to energy efficiency

building regulations. In: SAYIGH, Ali (Ed.). **Green buildings and renewable energy**. Alemana: Springer, 2019. p. 71-86.

EL-DIRABY, T. A.; LIMA, C.; FEIS, B. Domain taxonomy for construction concepts: toward a formal ontology for construction knowledge. **Journal of Computing in Civil Engineering**, v. 19, n. 4, p. 394-406, 2005.

EL-GOHARY, Nora M.; EL-DIRABY, Tamer E. Domain ontology for processes in infrastructure and construction. **Journal of Construction Engineering and Management**, v. 136, n. 7, p. 730-744, 2010. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0000178](https://doi.org/10.1061/(ASCE)CO.1943-7862.0000178)

EMBLEY, David *et al.* Ontology-based extraction and structuring of information from data-rich unstructured documents. In: INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, 7., 1998, Bethesda, MD. **Proceedings...** New York, NY: Association for Computing Machinery, 1998. p. 52-59.

FENVES, Steven J. Tabular decision logic for structural design. **Journal of the Structural Division**, v. 92, n. 6, p. 473-490, 1966.

FENVES, Steven *et al.* **Introduction to SASE: standards analysis, synthesis, and expression**. Washington: Editorial NBSIR, 1987. (NBSIR, 87-3513).

FENVES, Steven *et al.* Computer representations of design standards and building codes: U.S. perspective. **The International Journal of Construction Information Technology**, v. 3, n. 1, p. 13-34, 1995.

FUCHS, Stefan; AMOR, Robert. Natural language processing for building code interpretation: a systematic literature review. In: INTERNATIONAL CONFERENCE OF CIB W78, 38., 2021, Luxembourg. **Proceedings...** ITC Digital Library, 2021. p. 294-303.

GALLAHER, Michael *et al.* **Cost analysis of inadequate operability in the US capital facilities industry**. Gaithersburg, MD: National Institute of Standards and Technology, 2004.

GAO, Yunfan *et al.* Retrieval-augmented generation for large language models: A survey. **arXiv preprint arXiv:2312.10997**, 2023. GHANNAD, Pedram *et al.* Modularized BIM Data Validation Framework Integrating Visual Programming Language with LegalRuleML. In: Proceedings of the 35th CIB W78 2018 Conference, IT in Design, Construction, and Management. Chicago, 2018. p 85-94. <https://doi.org/10.48550/arXiv.2312.10997>

GODDARD, Cliff; SCHALLEY, Andrea C. Semantic analysis. In: INDURKHYA, Nitin; DAMERAU, Fred J. (Eds.). **Handbook of natural language processing**. 2.ed. Boca Raton, CA: CRC Press, 2010. p. 93-120.

GRANT THORNTON. **Mapeamento de maturidade BIM no Brasil**. Disponível em [https://siengeprod.wpenginepowered.com/wp-content/uploads/2020/11/ebook\\_mapeamento\\_bim\\_no\\_brasil.pdf](https://siengeprod.wpenginepowered.com/wp-content/uploads/2020/11/ebook_mapeamento_bim_no_brasil.pdf). Acesso em: 10/10/2022.

GRANT THORNTON. **Mapeamento de maturidade BIM no Brasil**. Grant Thornton; Sienge; ABDI, 2022. Disponível em: <https://www.grantthornton.com.br/globalassets/1.-member-firms/brazil/pesquisas/fy22/maturidade-bim/maturidade-bim-no-brasil--2022.pdf>. Acesso em: 10 out. 2022.

GRUBER, Thomas R. A translation approach to portable ontology specifications. **Knowledge Acquisition**, v. 5, n. 2, p. 199-220, 1993. <https://doi.org/10.1006/knac.1993.1008>

HAKIM, M. Maher; GARRETT JR., James H. Issues in modelling and processing design standards. In: 1992 COMPUTERS AND BUILDING STANDARDS WORKSHOP, 1992, Montreal, Canada, 1992. **Paper Presented...** Pittsburgh, PA: Carnegie Mellon University, 1992. p. 242-257.

HE, Xiaoyang *et al.* An automatic compliance checking approach for software processes. In: ASIA-PACIFIC SOFTWARE ENGINEERING CONFERENCE, 16., 2009, Batu Ferringhi, Malaysia. **Proceedings...** New York, NY: IEEE, 2009. p. 467-474. <https://doi.org/10.1109/apsec16387.2009>

HEVNER, Alan; CHATTERJEE, Samir. Design research in information systems. In: **Design Research in Information Systems**. Integrated Series in Information Systems, vol 22. Springer, Boston, 2010.

HJELSETH, Eilif. Foundation for development of computable rules. In: DIKBAS, Attila; ERGEN, Esin; GIRITLI, Heyecan. **Managing IT in construction/managing construction for tomorrow**. London: CRC Press, 2009. p. 257-267.

HJELSETH, Eilif; NISBET, Nick. Exploring semantic based model checking. In: **Proceedings of the 2010 27th CIB W78 international conference**. 2010.

HJELSETH, Eilif. Converting performance-based regulations into computable rules in BIM based model checking software. In: GUDNASON, Gudni; SCHERER, Raimar. **eWork and eBusiness in Architecture, Engineering and Construction**. London: Taylor & Francis, 2012. p. 461-469. <http://dx.doi.org/10.1201/b12516-73>

HJELSETH, Eilif. BIM-based model checking (BMC). In: ISSA, Raja R. A.; OLBINA, Svetlana. **Building information modeling: applications and practices**. Reston, VA: ASCE, 2015. p. 33-61. <https://doi.org/10.1061/9780784413982.ch02>

HJELSETH, Eilif; LI, Beidi. Use of “Reasoner” for automated building code compliance assessment. In: INTERNATIONAL CONFERENCE OF CIB W78, 38., 2021, Luxembourg. **Proceedings...** ITC Digital Library, 2021. p. 138-147.

İLAL, Sibel Macit; GÜNAYDIN, H. Murat. Computer representation of building codes for automated compliance checking. **Automation in Construction**, v. 82, p. 43-58, 2017. <https://doi.org/10.1016/j.autcon.2017.06.018>

INTERNATIONAL ELECTROTECHNICAL COMMISSION. **IEC-ISO smart FAQ**. Geneva, Switzerland: IEC, 2022a. Disponível em: [https://webdesk.jsa.or.jp/pdf/dev/md\\_5615.pdf](https://webdesk.jsa.or.jp/pdf/dev/md_5615.pdf). Acesso em: 11 jan. 2023.

INTERNATIONAL ELECTROTECHNICAL COMMISSION. **ISO smart: use cases**. 08 mar. 2022b. Disponível em: <https://www.youtube.com/watch?v=JzXoihlgzJQ&t=2s>. Acesso em: 11 jan. 2023.

INTERNATIONAL ELECTROTECHNICAL COMMISSION. **IEC/ISO smart: unleashing the digital power of international standards**, 2023. Disponível em: <https://www.iso.org/smart>. Acesso em: 12 jan. 2023.

KARLAPUDI, Janakiram; VALLURU, Prathap. **Digital construction materials**. 21 abr. 2021. Disponível em: <https://w3id.org/digitalconstruction/0.5/Materials>. Acesso em: 10 nov. 2022.

KATER, Marcel; RUSCHEL, Regina Coeli. O potencial da verificação automatizada baseada em regras para as medidas de segurança contra incêndio em BIM. **Ambiente Construído**, v. 20, n. 4, p. 423-444, 2020. <https://doi.org/10.1590/s1678-86212020000400481>

KEHL, Caroline; ISATTO, Eduardo Luís. Barreiras e oportunidades para a verificação automática de regras da produção na fase de projeto com uso da tecnologia BIM. In: ENCONTRO DE TECNOLOGIA DE INFORMAÇÃO E COMUNICAÇÃO NA CONSTRUÇÃO, 7., 2015, Recife. **Anais...** Recife: UFPE, 2015.

KERN, A. P.; SILVA, A.; KAZMIERCZAK, C. D. S. O processo de implantação de normas de desempenho na construção: um comparativo entre a Espanha (CTE) e Brasil (NBR 15575 / 2013). *Gestão e Tecnologia de Projetos*, São Paulo, v. 9, n. 1, p. 89-101, 2014. <https://doi.org/10.11606/gtp.v9i1.89989>

KERRIGAN, Shawn; LAW, Kincho H. Logic-based regulation compliance-assistance. In: INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND LAW, 9.,

2003. **Proceedings...** New York, NY: Association for Computing Machinery, 2003. p. 126-135.

KIM, Hayan *et al.* Visual language approach to representing KBimCode-based Korea building code sentences for automated rule checking. **Journal of Computational Design and Engineering**, v. 6, n. 2, p. 143-148, 2019. <https://doi.org/10.1016/j.jcde.2018.08.002>

KOO, Bonsang *et al.* A geometric deep learning approach for checking element-to-entity mappings in infrastructure building information models. **Journal of Computational Design and Engineering**, v. 8, n. 1, p. 239-250, 2021. <https://doi.org/10.1093/jcde/qwaa075>

KÖRNER, Philipp *et al.* Fifty years of prolog and beyond. **Theory and Practice of Logic Programming**, v. 22, n. 6, p. 776-858, 2022. <https://doi.org/10.1017/S1471068422000102>

KWASNIEWSKI, Leslaw; BOJANOWSKI, Cezary. *Principles of verification and validation*. **Journal of Structural Fire Engineering**, v. 6, n. 1, p. 29-40, 2015. <https://doi.org/10.1260/2040-2317.6.1.29>

LACERDA, D. P. et al. Design Science Research: método de pesquisa para a engenharia de produção. **Gestão & produção**, v. 20, p. 741-761, 2013. <https://doi.org/10.1590/S0104-530X2013005000014>

LEE, Yong-Cheol; EASTMAN, Charles M.; SOLIHIN, Wawan. *An ontology-based approach for developing data exchange requirements and model views of building information modeling*. **Advanced Engineering Informatics**, v. 30, n. 3, p. 354-367, 2016.

LEE, J.; YI, J.; SON, J. Development of automatic-extraction model of poisonous clauses in international construction contracts using rule-based NLP. In: **Journal of Computing in Civil Engineering**, v. 33, n. 3, p. 04019003, 2019.

LEE, Hyunsoo *et al.* Translating building legislation into a computer-executable format for evaluating building permit requirements. **Automation in Construction**, v. 71, part 1, p. 49-61, 2016. <https://doi.org/10.1016/j.autcon.2016.04.008>

LEWIS, Patrick *et al.* Retrieval-augmented generation for knowledge-intensive NLP tasks. **Advances in Neural Information Processing Systems**, v. 33, p. 9459-9474, 2020.

LI, Zhanjun; RASKIN, Victor; RAMANI, Karthik. Developing engineering ontology for information retrieval. In: **Journal of Computing and Information science in Engineering**, v. 8, n. 1, 2008.

LI, Beidi; DIMYADI, Johannes; AMOR, Robert; SCHULTZ, Carl. Qualitative and traceable calculations for building codes. In: CIB W78 INFORMATION TECHNOLOGY

FOR CONSTRUCTION CONFERENCE, 37., 2020, São Paulo. **Proceedings...** p. 69-84. <http://dx.doi.org/10.46421/2706-6568.37.2020.paper006>

LI, B. *et al.* Defeasible reasoning for automated building code compliance checking. **ECPPM 2021-eWork and eBusiness in Architecture, Engineering and Construction**, p. 229-236, 2021. <https://doi.org/10.1201/9781003191476>

LIMA, João. **Portal LexML disponibiliza metadados das normas jurídicas federais com marcação Schema.org/Legislation**. 12 fev. 2019. Disponível em: <https://projeto.lexml.gov.br/institucional/noticias/portal-lexml-disponibiliza-metadados-das-normas-juridicas-federais-com-marcacao-schema-org-legislation-1>. Acesso em: 11 nov. 2022.

LIMA, Celson *et al.* The e-COGNOS project, current status and future directions of an ontology-enabled IT solution infrastructure supporting knowledge management in construction. In: CONSTRUCTION RESEARCH CONGRESS, 2003, Honolulu, Hawaii. **Proceedings...** p. 1-8. [https://doi.org/10.1061/40671\(2003\)103](https://doi.org/10.1061/40671(2003)103)

MACIT, Sibel. Computer representation of building codes for automated compliance checking. 2014. Tese de Doutorado. **Izmir Institute of Technology**, Turquia.

MAINARDI NETO, Antônio Ivo de B.; SANTOS, Eduardo Toledo. Verificação de regras em modelos BIM: um estudo de caso sobre projeto de arquitetura de estações metroviárias. In: ENCONTRO BRASILEIRO DE TECNOLOGIA DE INFORMAÇÃO E COMUNICAÇÃO NA CONSTRUÇÃO, 7., 2015, Recife. **Anais...** Recife: UFPE, 2015. <https://doi.org/10.5151/engpro-tic2015-068>

MANYIKA, James *et al.* **Digital America: a tale of the haves and have-mores**. San Francisco, CA: McKinsey Global Institute, 2015. Disponível em: <https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/digital-america-a-tale-of-the-haves-and-have-mores>. Acesso em: 18 ago. 2022.

MARCH, Salvatore T.; SMITH, Gerald F. Design and natural science research on information technology. **Decision support systems**, v. 15, n. 4, p. 251-266, 1995. [https://doi.org/10.1016/0167-9236\(94\)00041-2](https://doi.org/10.1016/0167-9236(94)00041-2)

MENDONÇA, Eduardo A.; MANZIONE, Leonardo; HJELSETH, Eilif. Converting Brazilian accessibility standard for BIM-based code checking using RASE and SMC. In: CIB W78 INFORMATION TECHNOLOGY FOR CONSTRUCTION CONFERENCE, 37., 2020, São Paulo. **Proceedings...** p. 291-307. <http://dx.doi.org/10.46421/2706-6568.37.2020.paper021>

MENEGOTTO, José Luís. BIM ontology. Some aspects of project knowledge: the building, IFCs and STOs. In: INTERNATIONAL CONFERENCE OF THE IBERO-

AMERICAN SOCIETY OF DIGITAL GRAPHICS, 27., 2023, Punta del Este, Uruguay. **Anais...** Punta del Este: Sigradi, 2023. p. 433-446.

MIALON, Grégoire *et al.* GAIA: a benchmark for general AI assistants. **arXiv:2311.12983**, 2023. <https://doi.org/10.48550/arXiv.2311.12983>

MILLER, George A.; FELLBAUM, Christiane. WordNet then and now. **Language Resources and Evaluation**, v. 41, n. 2, p. 209-214, 2007. <https://doi.org/10.1007/s10579-007-9044-6>

NARDELLI, Eduardo Sampaio; OLIVEIRA, João Tales. BIM e desempenho no programa Minha Casa Minha Vida - PMCMV. In: CONGRESO DE LA SOCIEDAD IBEROAMERICANA DE GRÁFICADIGITAL, 17., 2013, Valparaíso, Chile. **Blucher Design Proceedings...** Santa Maria: SiGraDI, 2014. <https://doi.org/10.5151/despro-sigradi2013-0059>

NAWARI, O. Nawari. **Building information modeling: automated code checking and compliance process**. Boca Raton, CA: CRC Press, 2018. <https://doi.org/10.1201/9781351200998>

NGUYEN, Tang-Hung; KIM, Jin-Lee. Building code compliance checking using BIM technology. In: **Proceedings of the 2011 winter simulation conference (WSC)**. IEEE, 2011. p. 3395-3400. <https://doi.org/10.1109/WSC.2011.6148035>

NIRENBURG, Sergei; RASKIN, Victor. **Ontological semantics (language, speech, and communications series)**. Cambridge, MA: MIT Press, 2004.

NIVRE, Joakim. Inductive dependence parsing. In: IDE, Nancy. **Text, speech and language technology**. Dordrecht: Springer, 2006. p. 87-120. (TLTB, 34). [https://doi.org/10.1007/1-4020-4889-0\\_4](https://doi.org/10.1007/1-4020-4889-0_4)

NOARDO, Francesca *et al.* An inspection of IFC models from practice. *Applied Sciences*, v. 11, n. 5, 2232, 2021. <https://doi.org/10.3390/app11052232>

NUNAMAKER JR, Jay F.; CHEN, Minder; PURDIN, Titus D.M. Systems development in information systems research. **Journal of management information systems**, v. 7, n. 3, p. 89-106, 1990. <http://www.jstor.org/stable/40397957>

NUNES, Vitor *et al.* A implantação da norma NBR 15575 e seu impacto no setor de construção civil. In: **PARC Pesquisa em Arquitetura e Construção**. Campinas, v. 12, p. e021010, 2021. <http://dx.doi.org/10.20396/parc.v12i00.8656159>

OKAMOTO, Patrícia Seiko. **Os impactos da norma brasileira de desempenho sobre o processo de projeto de edificações residenciais**. 2015. 160f. Dissertação (Mestrado em

Engenharia de Construção Civil e Urbana) – Escola Politécnica da Universidade de São Paulo, São Paulo, 2015.

OSOEGAWA, Aline Harumi; MITIDIÉRI FILHO, Cláudio Vicente. Análise de projetos visando o atendimento da norma ABNT NBR 15575: 2013. In: ENCONTRO NACIONAL DE TECNOLOGIA DO AMBIENTE CONSTRUÍDO, 18., 2020, Porto Alegre. **Anais...** Porto Alegre, 2020. <https://doi.org/10.46421/entac.v18i.968>

OSORIO SANZ, Clara. **Lectura fácil**: oraciones subordinadas adjetivas explicativas. 2023. 25f. Proyecto fin de Carrera/Trabajo Fin de Grado (Grado en Ingeniería Informática) - Universidad Politécnica de Madrid, 2023.

PARK, Seokyoung; LEE, Jin-Kook. KBimCode-based applications for the representation, definition and evaluation of building permit rules. In: INTERNATIONAL SYMPOSIUM ON AUTOMATION AND ROBOTICS IN CONSTRUCTION, 33., 2016, Auburn, AL. **Proceedings...** Auburn, AL, 2016. p. 635-642.

PAUWELS, Pieter; ZHANG, Sijie; LEE, Yong-Cheol. Semantic web technologies in AEC industry: a literature overview. **Automation in Construction**, v. 73, p. 145-165, 2017. <https://doi.org/10.1016/j.autcon.2016.10.003>

PEREIRA, Niana Franciscatto. **Ferramenta para atendimento no projeto arquitetônico dos requisitos da NBR 15575/4: 2013** – desempenho de edificações habitacionais. 2021. 160f. Dissertação (Mestrado em Arquitetura, Urbanismo e Paisagem) – Universidade Federal de Santa Maria, Santa Maria, 2021.

PREIDEL, Cornelius; BORRMANN, André. Automated code compliance checking based on a visual language and building information modeling. In: INTERNATIONAL SYMPOSIUM ON AUTOMATION AND ROBOTICS IN CONSTRUCTION, 32., 2015, Oulu, Finland. **Proceedings...** Oulu, Finland, 2015.

PREIDEL, Cornelius; BORRMANN, André. BIM-based code compliance checking. In BORRMANN, André *et al.* (Eds.). **Building information modeling: technology foundations and industry practice**. Cham, Switzerland: Springer, 2018. <https://doi.org/10.1007/978-3-319-92862-3>

REHMAN, Saddiq Ur *et al.* Analysis of development of open bim-based automated rule checking system in KBIM project. In: ASIAN FORUM ON GRAPHIC SCIENCE, 13., 2021, Hong Kong. **Proceedings...** Hong Kong: The Hong Kong Polytechnic University, 2021.

REN, Yong *et al.* An ontology-based and deep learning-driven method for extracting legal facts from Chinese legal texts. **Electronics**, v. 11, n. 12, 1821, 2022. <https://doi.org/10.3390/electronics11121821>

RETOOL. **State of AI**. 2023. Disponível em: [https://retool.com/reports/state-of-ai-2023?utm\\_source=bensbites&utm\\_medium=newsletter&utm\\_campaign=daily-digest-nvidia-strikes-again](https://retool.com/reports/state-of-ai-2023?utm_source=bensbites&utm_medium=newsletter&utm_campaign=daily-digest-nvidia-strikes-again). Acesso em: 25 set. 2023.

RIBEIRO, R. M.; FOSSATI, Michele. Critérios de desempenho térmico para edificações habitacionais: análise da norma de desempenho e de código de obras. In: ENCONTRO NACIONAL DE TECNOLOGIA DO AMBIENTE CONSTRUÍDO, 18., 2020, Porto Alegre. **Anais...** Porto Alegre: ANTAC, 2020. <https://doi.org/10.46421/entac.v18i.826>

ROUSSEEUW, Peter J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. **Journal of Computational and Applied Mathematics**, v. 20, p. 53-65, 1987. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)

SAWYER, Tom. **Making the case for automated BIM review**. 28 mai. 2012. Disponível em: <https://www.enr.com/articles/8456-making-the-case-for-automated-bim-review>. Acesso em: 26 out. 2022.

SCHMITT, Xavier *et al.* A replicable comparison study of NER software: StanfordNLP, NLTK, OpenNLP, SpaCy, Gate. In: INTERNATIONAL CONFERENCE ON SOCIAL NETWORKS ANALYSIS, MANAGEMENT AND SECURITY, 6., 2019, Granada, Spain. **Proceedings...** New York, NY: IEEE, 2019. p. 338-343.

SHIN, Dong-Hee; KIM, Hongbum; HWANG, Junseok. Standardization revisited: a critical literature review on standards and innovation. **Computer Standards & Interfaces**, v. 38, p. 152-157, 2015. <https://doi.org/10.1016/j.csi.2014.09.002>

SIMON, H. A. **The Sciences of the Artificial**. 3rd ed. Cambridge: MIT Press, 1996

SILVA, Flávio Paulino de Andrade e. **Verificação automática dos requisitos de projetos da norma de desempenho pela plataforma BIM Solibri Model Checker**. 2017. 161f. Dissertação (Mestrado em Construção Civil) – Universidade Federal de Minas Gerais, Belo Horizonte, 2017.

SILVA, Flávio Paulino de Andrade e; ARANTES, Eduardo Marques. Verificação automática de requisitos de projetos da norma de desempenho NBR 15575 a partir da adequação de regras da plataforma BIM Solibri Model Checker. In: SIMPÓSIO BRASILEIRO DE TECNOLOGIA DA INFORMAÇÃO E COMUNICAÇÃO NA CONSTRUÇÃO, 1.; SIMPÓSIO BRASILEIRO DE GESTÃO E ECONOMIA DA CONSTRUÇÃO, 10., 2017, Fortaleza. **Anais...** Porto Alegre: Antac, 2017.

SILVA JUNIOR, M. A.; MITIDIERI FILHO, C. V. Verificação de critérios de desempenho em projetos de arquitetura com a modelagem BIM. **PARC Pesquisa em Arquitetura e**

**Construção**, v. 9, n. 4, p. 334-343, 1 dez. 2018. DOI: Disponível em: <  
<https://doi.org/10.20396/parc.v9i4.8650453>

SINGHAL, Karan *et al.* Towards expert-level medical question answering with large language models. **arXiv:2305.09617**, 2023. <https://doi.org/10.48550/arXiv.2305.09617>

SOBHKHIZ, Soroush *et al.* Framing and evaluating the best practices of IFC-based automated rule checking: a case study. **Buildings**, v. 11, n. 10, 456, 2021.  
<https://doi.org/10.3390/buildings11100456>

SOLIBRI. **Getting started with Solibri**: introduction to basic functionalities. 2020.  
 Disponível em: [https://solibri-assets.s3.amazonaws.com/downloads/Getting\\_Started\\_with\\_Solibri-EN.pdf](https://solibri-assets.s3.amazonaws.com/downloads/Getting_Started_with_Solibri-EN.pdf). Acesso em: 10 nov. 2022.

SOLIHIN, Wawan. Lessons learned from experience of code-checking implementation in Singapore. In: BUILDINGSMART CONFERENCE, 2004, Singapore. Disponível em: [https://www.researchgate.net/publication/280599027\\_Lessons\\_learned\\_from\\_experience\\_of\\_code-checking\\_implementation\\_in\\_Singapore](https://www.researchgate.net/publication/280599027_Lessons_learned_from_experience_of_code-checking_implementation_in_Singapore). Acesso em: 02 nov. 2022.

SOLIHIN, Wawan; EASTMAN, Charles. Classification of rules for automated BIM rule checking development. **Automation and Construction**, v. 53, p. 69-82, 2015.  
<https://doi.org/10.1016/j.autcon.2015.03.003>

SOLIHIN, Wawan; EASTMAN, Charles. A knowledge representation approach in BIM rule requirement analysis using the conceptual graph. **Journal of Information Technology in Construction**, v. 21, p. 370-401, 2016. <https://www.itcon.org/paper/2016/24>

SOLIMAN-JUNIOR, João; FORMOSO, Carlos T.; TZORTZOPOULOS, Patrícia. A semantic-based framework for automated rule checking in healthcare construction projects. **Canadian Journal of Civil Engineering**, v. 47, n. 2, p. 202-214, 2019.  
<https://doi.org/10.1139/cjce-2018-0460>

SRINIVASAN, Ravi S.; NAWARI, Nawari O. **Systems and methods for automated building code conformance**. Depositante: University of Florida Research Foundation. n. WO 2022/165483 A1. Depósito: 29 jan. 2021. Concessão: 04 ago. 2022.

SUCCAR, Bilal; KASSEM, Mohamad. Macro-BIM adoption: conceptual structures. **Automation in Construction**, v. 57, p. 64-79, 2015.  
<https://doi.org/10.1016/j.autcon.2015.04.018>

SUCCAR, Bilal; POIRIER, Erik. Lifecycle information transformation and exchange for delivering and managing digital and physical assets. **Automation in Construction**, v. 112, 103090, 2020. <https://doi.org/10.1016/j.autcon.2020.103090>

THANAKI, Jalaj. **Python natural language processing**. Birmingham: Packt Publishing Ltd, 2017.

THESEN, Anne E.; CUI, Hong; MOZZHERIN, Dmitry. Applications of natural language processing in biodiversity science. **Advances in Bioinformatics**, v. 2012, ID 391574, 2012. <https://doi.org/10.1155/2012/391574>

TOMCZAK, Arthur. IDS Editor, 2022. Disponível em <https://artomczak.pythonanywhere.com/create/>. Consultado em jan 2024.

TONSO, Laís Guerle; NARDELLI, Eduardo Sampaio. BIM para a análise de desempenho térmico em edificações do Programa Minha Casa Minha Vida. In: CONGRESSO DA SOCIEDADE IBERO-AMERICANA DE GRÁFICA DIGITAL, 19., 2015, Florianópolis. **Anais...** Florianópolis: Sigradi, 2015. <https://doi.org/10.5151/despro-sigradi2015-10305>

TOUVRON, Hugo *et al.* Llama: open and efficient foundation language models. **arXiv:2302.13971**, 2023. <https://doi.org/10.48550/arXiv.2302.13971>

UPCODES. **UpCodes AI prototype overview (not currently available)**. 19 mai. 2019. Disponível em: <https://youtu.be/0wPICZUGKEc>. Acesso em: 20 nov. 2022.

VAJJALA, Sowmya *et al.* **Practical natural language processing: a comprehensive guide to building real-world NLP systems**. Sebastopol: O'Reilly Media, 2020.

VENUGOPAL, Manu; EASTMAN, Charles M.; TEIZER, Jochen. An ontology-based analysis of the industry foundation class schema for building information model exchanges. *Advanced Engineering Informatics*, v. 29, n. 4, p. 940-957, 2015. <https://doi.org/10.1016/j.aei.2015.09.006>

WALTER, Merlin; BROMHEAD, Ben. How to improve your LLM accuracy and performance with PGVector and PostgreSQL®: introduction to embeddings and the role of PGVector. 19 out. 2023. Disponível em: <https://www.instaclustr.com/blog/how-to-improve-your-llm-accuracy-and-performance-with-pgvector-and-postgresql-introduction-to-embeddings-and-the-role-of-pgvector/>. Acesso em: 10 jan. 2024.

WANG, Ning; ISSA, Raja R. A. Natural language generation from building information models for intelligent NLP-based information extraction. In: EG-ICE 2020 WORKSHOP ON INTELLIGENT COMPUTING IN ENGINEERING, 2020, Berlin, Germany. **Proceedings...** Berlin, Germany, 2020. p. 275-284. <http://dx.doi.org/10.14279/depositonce-9977>

WANG, Peng *et al.* Enhanced named entity recognition with semantic dependency. In: PACIFIC RIM INTERNATIONAL CONFERENCE ON ARTIFICIAL

INTELLIGENCE, 18., 2021, Hanoi, Vietnam. **Proceedings...** New York, NY: Springer International Publishing, 2021. p. 287-298. (Lecture Notes in Computer Science, 13032).

WEI, Jason *et al.* Chain-of-thought prompting elicits reasoning in large language models. In: CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS, 36., 2022, New Orleans. **Proceedings...** San Diego, CA: NeuroIPS, 2022. p. 24824-24837.

WIELEMAKER, Jan. **Logic programming for knowledge-intensive interactive applications.** 2009. 224f. Tese (Doutorado em Informática) – Universiteit van Amsterdam, Amsterdam, 2009.

WIMALASURIYA, Daya C.; DOU, Dejing. Ontology-based information extraction: an introduction and a survey of current approaches. **Journal of Information Science**, v. 36, n. 3, p. 306-323, 2010. <https://doi.org/10.1177/0165551509360123>

WORLD ECONOMIC FORUM. Shaping the future of construction: A breakthrough in mindset and technology. In: **World Economic Forum.** 2016. p. 01-61.

WU, Chengke *et al.* Natural language processing for smart construction: current status and future directions. **Automation in Construction**, v. 134, 104059, 2022. <https://doi.org/10.1016/j.autcon.2021.104059>

WYNER, Adam Zachary; PETERS, Wim. On rule extraction from regulations. In: INTERNATIONAL CONFERENCE ON LEGAL KNOWLEDGE AND INFORMATION SYSTEMS, 24., 2011, Vienna, Austria. **Proceedings...** Vienna, Austria, 2011. p. 113-122. (Frontiers in Artificial Intelligence and Applications, 235). <https://doi.org/10.3233/978-1-60750-981-3-113>

XU, Xin; CAI, Hubo. Semantic frame-based information extraction from utility regulatory documents to support compliance checking. In: CIB W78 CONFERENCE: IT IN DESIGN, CONSTRUCTION, AND MANAGEMENT, 35., 2018, Chicago, IL. **Proceedings...** New York, NY: Springer International Publishing, 2019. p. 223-230. [https://doi.org/10.1007/978-3-030-00220-6\\_27](https://doi.org/10.1007/978-3-030-00220-6_27)

XU, Xin; CAI, Hubo. Semantic approach to compliance checking of underground utilities. **Automation in Construction**, v. 109, p. 103006, 2020. <http://dx.doi.org/10.1016/j.autcon.2019.103006>

XU, Xin; CAI, Hubo. Ontology and rule-based natural language processing approach for interpreting textual regulations on underground utility infrastructure. **Advanced Engineering Informatics**, v. 48, 101288, 2021. <https://doi.org/10.1016/j.aei.2021.101288>

XU, Mingxue; XU, Yao Lei; MANDIC, Danilo P. Tensorgpt, efficient compression of the embedding layer in llms based on the tensor-train decomposition. **arXiv:2307.00526**, 2023. <https://doi.org/10.48550/arXiv.2307.00526>

YURCHYSHYNA, Anastasiya *et al.* Towards an ontology-based approach for conformance checking modeling in construction. In: CIB W78 INTERNATIONAL CONFERENCE ON INFORMATION TECHNOLOGY FOR CONSTRUCTION, 24., 2007, Maribor, Slovenia. **Proceedings...** New York. NY: Springer International Publishing, 2007. p 195-202.

ZHANG, Jiansong. Automated code compliance checking in the construction domain using semantic natural language processing and logic-based reasoning. Tese (Doutorado em Engenharia Civil), University of Illinois at Urbana-Champaign, 2015.

ZHANG, Jiansong; EL-GOHARY, Nora M. Automated information extraction from construction-related regulatory documents for automated compliance checking. In: CIB W78-W102 2011: INTERNATIONAL CONFERENCE, 2011, Sophia Antipolis, France. **Proceedings...** New York. NY: Springer International Publishing, 2011. p. 26-28.

ZHANG, Jiansong; EL-GOHARY, Nora M. Extraction of construction regulatory requirements from textual documents using natural language processing techniques. In: INTERNATIONAL CONFERENCE ON COMPUTING IN CIVIL ENGINEERING, 2012, Clearwater Beach, FL. **Proceedings...** Reston, VA: American Society of Civil Engineers, 2012. p. 453-460. <https://doi.org/10.1061/9780784412343.0057>

ZHANG, Jiansong; EL-GOHARY, N. M. Handling sentence complexity in information extraction for automated compliance checking in construction. **Proc., CIB W78 2013**, 2013.

ZHANG, Jiansong; EL-GOHARY, Nora M. Automated reasoning for regulatory compliance checking in the construction domain. In: CONSTRUCTION RESEARCH CONGRESS, 2014, Georgia, GA. **Proceedings...** Reston, VA: American Society of Civil Engineers, 2014. p. 907-916, 2014. <https://doi.org/10.1061/9780784413517.093>

ZHANG, Jiansong; EL-GOHARY, Nora M. Automated information transformation for automated regulatory compliance checking in construction. In: **Journal of Computing in Civil Engineering**, v. 29, n. 4, B4015001, 2015. [http://dx.doi.org/10.1061/\(ASCE\)CP.1943-5487.0000427](http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0000427)

ZHANG, Jiansong; EL-GOHARY, Nora M. Semantic NLP-based information extraction from construction regulatory documents for automated compliance checking. **Journal of Computing in Civil Engineering**, v. 30, n. 2, p. 04015014, 2016. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000346](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000346)

ZHANG, Jiansong; EL-GOHARY, Nora M. Integrating semantic NLP and logic reasoning into a unified system for fully-automated code checking. **Automation in Construction**, v. 73, p. 45-57, 2017a. <https://doi.org/10.1016/j.autcon.2016.08.027>

ZHANG, Jiansong; EL-GOHARY, Nora M. Semantic-based logic representation and reasoning for automated regulatory compliance checking. **Journal of Computing in Civil Engineering**, v. 31, n. 1, 2017b. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000583](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000583)

ZHANG, Ruichuan; EL-GOHARY, Nora. Clustering-based approach for building code computability analysis. **Journal of Computing in Civil Engineering**, v. 35, n. 6, p. 04021021, 2021. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000967](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000967)

ZHANG, Ruichuan; EL-GOHARY, Nora M. Natural language generation and deep learning for intelligent building codes. **Advanced Engineering Informatics**, v. 52, 101557, 2022. <https://doi.org/10.1016/j.aei.2022.101557>

ZHANG, Zhuosheng *et al.* Automatic chain of thought prompting in large language models. **arXiv:2210.03493**, 2022. <https://doi.org/10.48550/arXiv.2210.03493>

ZHANG, Zijing; MA, Ling; BROYD, Tim. Rule capture of automated compliance checking of building requirements: a review. In: FOCUS: 2022 EUROPEAN CONFERENCE IN COMPUTING IN CONSTRUCTION, 2022, Rhodes, Greece. **Proceedings...** St-Niklaas, Flemish Region, Belgium: European Council on Computing in Construction 2023. p. 224-238. (Smart Infrastructure and Construction, 176). <https://doi.org/10.1680/jsmic.23.00005>

ZHENG, Zhe *et al.* Knowledge-informed semantic alignment and rule interpretation for automated compliance checking. **Automation in Construction**, v. 142, p. 104524, 2022. <https://doi.org/10.1016/j.autcon.2022.104524>

ZHONG, B. T. *et al.* Ontology-based semantic modeling of regulation constraint for automated construction quality compliance checking. **Automation in construction**, v. 28, p. 58-70, 2012. <https://doi.org/10.1016/j.autcon.2012.06.006>

ZHONG, Botao *et al.* Automatic Information Extraction from Construction Quality Inspection Regulations: A Knowledge Pattern-Based Ontological Method. **Journal of Construction Engineering and Management**, v. 148, n. 3, p. 04021207, 2022. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0002240](https://doi.org/10.1061/(ASCE)CO.1943-7862.0002240)

ZHOU, Peng; EL-GOHARY, Nora. Ontology-based automated information extraction from building energy conservation codes. **Automation in Construction**, v. 74, p. 103-117, 2017. <https://doi.org/10.1016/j.autcon.2016.09.004>

ZHOU, Peng; EL-GOHARY, Nora M. Semantic information alignment of BIMs to computer-interpretable regulations using ontologies and deep learning. **Advanced Engineering Informatics**, v. 48, 101239, 2021. <https://doi.org/10.1016/j.aei.2020.101239>

ZHOU, Yu-Cheng *et al.* Integrating NLP and context-free grammar for complex rule interpretation towards automated compliance checking. **Computers in Industry**, v. 142, p. 103746, 2022.

## APÊNDICE A – Estágio de tradução do IFC para português (pt-BR)

Um problema na interoperabilidade entre sistemas e regulamentos é que, apesar do regulamento ser codificado na língua nativa do usuário, o programa de modelagem BIM ainda utiliza o padrão IFC em inglês. Nesse caso seria necessário um dicionário que alinhasse os termos da língua nativa para os respectivos termos padronizados em inglês. O Solibri, por exemplo, apenas possui interface em inglês e todos os termos de elementos e atributos são em língua inglesa. Enquanto o IFC for usado na língua inglesa, existirá essa limitação.

No início de 2023 o IFC possuía a iniciativa de tradução em 9 línguas, mas obteve um desenvolvimento acelerado e em fevereiro de 2024 possui oito línguas com tradução completa e mais oito em vias de finalização.

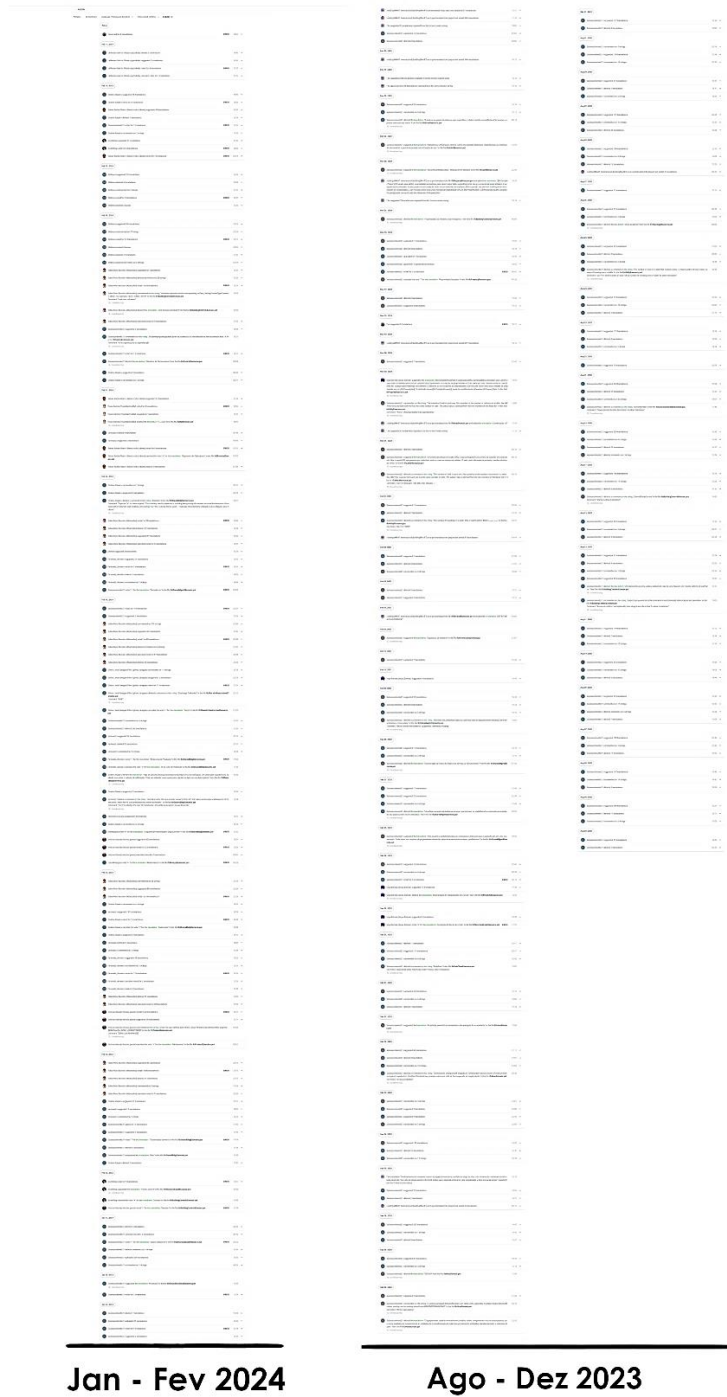
**Figura 78 - Tradução do IFC via plataforma Crowdin em novembro de 2023 e fevereiro de 2024**



Fonte: <https://buildingsmart.crowdin.com/u/projects/2>.

A iniciativa de tradução da língua portuguesa (pt-BR) teve avanço significativo no ano de 2023, mas a versão em português ainda não foi implementada.

Figura 79 - Atividade de tradução do IFC para português entre Agosto de 2023 e Fevereiro de 2024



Fonte: o autor.

## APÊNDICE B - Experimentos iniciais para avaliação das ferramentas de PLN

### Pré-processamento

O texto da Norma é disponível em formato PDF, portanto deve ser convertido para .txt que é o formato utilizado nas etapas iniciais de PLN. Nesse exemplo foi utilizada a biblioteca PyPDF2<sup>52</sup> para Python 3.0.

O Quadro 43 apresenta o exemplo do código de pré-processamento.

#### Quadro 43 - experimento inicial para processamento de trecho da Norma de Desempenho

testes iniciais da inserção do texto em formato *.TXT para processamento
<pre>with open('/content/drive/Othercomputers/Laptop/Pesquisa Doutorado/Dados para Colab/NBR_DOC.txt', 'r') as f:     texto = f.read() f.close()</pre>
saída com diversos erros causados pela conversão do PDF
<pre>11.4.2 Método de avaliaçãoSimulação computacional conforme procedimentos apresentados em 11.2.11.5 Edifi cações em fase de projetoA avaliação deve ser feita para um dia típico de projeto11.4.2 Método de avaliaçãoSimulação computacional conforme procedimentos apresentados em 11.2.11.5 Edifi cações em fase de projetoA avaliação deve ser feita para um dia típico de projeto</pre>
Limpeza de números de capítulos, evitando exclusão de quantidades de itens, usando Expressões Regulares
<pre>texto_limpo = re.sub("([0-9]+.[0-9]+.[0-9]+.[0-9]+) ([0-9]+.[0-9]+.[0-9]+) ([0-9]+[^\,][0-9]+)", "?", texto)</pre>
Remoção de pontuações e sinais no texto
<pre>texto_xpont = re.sub('([@.?!,%:;])', '', texto_limpo) texto_xpont</pre>
Saída
<pre>Método de avaliaçãoSimulação computacional conforme procedimentos apresentados em Edifi cações em fase de projetoA avaliação deve ser feita para um dia típico de projeto</pre>
Tokenização das palavras
<pre>doc = nlp(texto_xpont)</pre>

<sup>52</sup> <https://pypi.org/project/PyPDF2>

```
tokens = [token for token in doc]
print(tokens)
print(type(tokens))
```

#### Saída

```
[ , Método, de, avaliaçãoSimulação, computacional, conforme, procedimentos,
apresentados, em, , Edifi, cações, em, fase, de, projetoA]
```

Fonte: o autor.

## Rotulação

A fase trata de criar etiquetas, rótulos, sintáticos para cada palavra e será usada o método de POS tagging da biblioteca SpaCy.

### Quadro 44 - Código para rotulação usando SpaCy

Criação de todas as features de dependência e POS tagging necessárias

```
for tk in doc[:30]:
    print (tk.text, ':', tk.pos_)
```

Fonte: o autor.

### Figura 80 - Exemplo da rotulação feita pelo SpaCy

```
Método : NOUN
de : ADP
avaliaçãoSimulação : NOUN
computacional : ADJ
conforme : SCONJ
procedimentos : NOUN
apresentados : VERB
em : ADP
: SPACE
Edifi : PROPN
cações : NOUN
em : ADP
fase : NOUN
de : ADP
projetoA : VERB
avaliação : NOUN
deve : VERB
ser : AUX
feita : VERB
```

Fonte: o autor.

A partir da planilha da pesquisa de Osoegawa e Mitidieri Filho (2020), foi feita a seleção manual dos textos originais da parte 1 da norma. A parte 1 é um documento composto de 110 páginas, das quais existem tabelas, quadros e figuras além do texto

organizado em capítulos. Os requisitos da norma são mencionados a partir do quarto capítulo.

O texto extraído de Osoegawa e Mitidieri Filho (2020) foi convertido para arquivo .txt e processado para separação automática das sentenças. O texto possui caracterização de listas numeradas como:

“Os sistemas não podem apresentar:

- a) rupturas, instabilidades, tombamentos ou quedas que possam colocar em risco a integridade física dos ocupantes ou de transeuntes nas imediações do imóvel;
- b) partes expostas cortantes ou perfurantes;
- c) deformações e defeitos acima dos limites especificados nas ABNT NBR 15575-2 a ABNT NBR 15575-6.”

Foi observado que o padrão textual obedece à separação de sentenças por pontos finais resultando na representação abaixo:

“Os sistemas não podem apresentar: a) rupturas, instabilidades, tombamentos ou quedas que possam colocar em risco a integridade física dos ocupantes ou de transeuntes nas imediações do imóvel; b) partes expostas cortantes ou perfurantes; c) deformações e defeitos acima dos limites especificados nas ABNT NBR 15575-2 a ABNT NBR 15575-6.”

Foram selecionadas sentenças simples, com menos cláusulas como nos artigos que se dedicam à representação de requisitos da construção.

Foi usada a IDE Google Colab para programar, gravar e extrair os dados das sentenças. Parte do conteúdo foi inserido em um repositório do Zenodo para permitir acesso remoto não dependente da propriedade do código.

Na análise dos requisitos extraídos por Osoegawa e Mitidieri Filho (2020), na parte 1 da norma, foram extraídos 3108 tokens em 74 requisitos.

Ao preparar o arquivo para anotação de seus constituintes, alguns termos não são rotulados corretamente com unidade do discurso (POS tags):

Dada a frase:

*“Os componentes, equipamentos e dispositivos de manobra devem ser projetados, construídos e montados de forma a evitar que a força necessária para o acionamento não exceda 10 N nem o torque ultrapasse 20 N.m”*

É rotulada da seguinte forma:

[('Os', 'ART'), ('componentes', 'N'), (',', ','), ('equipamentos', 'N'), ('e', 'KC'), ('dispositivos', 'N'), ('de', 'PREP'), ('manobra', 'N'), ('devem', 'VAUX'), ('ser', 'VAUX'), ('projetados', 'PCP'), (',', ','), ('construídos', 'PCP'), ('e', 'KC'), ('montados', 'PCP'), ('de', 'PREP'), ('forma', 'N'), ('a', 'ART'), ('evitar', 'V'), ('que', 'PRO-KS-REL'), ('a', 'ART'), ('força', 'N'), ('necessária', 'ADJ'), ('para', 'PREP'), ('o', 'ART'), ('acionamento', 'None'), ('não', 'ADV'), ('exceda', 'None'), ('10', 'NUM'), ('N', 'PREP | +'), ('nem', 'KC'), ('o', 'ART'), ('torque', 'None'), ('ultrapasse', 'V'), ('20', 'NUM'), ('N.m', 'None)]

As palavras que possuem rótulo “None” não são reconhecidas pelo corpus utilizado (Mac Morpho). Algumas delas precisam ser reconhecidas por uma base de dados externa e pode ser um padrão artificial dado por expressões regulares (sempre que existir duas letras separadas por um ponto = unidade de medida) um gazetteer com uma lista de termos específicos do domínio ou um uma ontologia.

Importante notar que essa rotulação foi feita por um modelo de linguagem usado na biblioteca NLTK, mas na spaCy, eles foram corretamente identificados como NOUN e VERB mostrando a robustez de sua análise em comparação com a desenvolvida pelo modelo MacMorpho.

## APÊNDICE C - NBR 15575-1, Capítulo 8 - Segurança contra incêndio

### 8 Segurança contra incêndio

#### 8.1 Generalidades

Os requisitos desta Norma relativos à segurança contra incêndio são pautados em:

- proteger a vida dos ocupantes das edificações, em caso de incêndio;
- dificultar a propagação do incêndio, reduzindo danos ao meio ambiente e ao patrimônio;
- proporcionar meios de controle e extinção do incêndio;
- dar condições de acesso para as operações do Corpo de Bombeiros.

Os objetivos principais de garantir a resistência ao fogo dos elementos estruturais são:

- possibilitar a saída dos ocupantes da edificação em condições de segurança;
- garantir condições para o emprego de socorro público, onde se permita o acesso operacional de viaturas, equipamentos e seus recursos humanos, com tempo hábil para exercer as atividades de salvamento (pessoas retidas) e combate a incêndio (rescaldo e extinção);
- evitar ou minimizar danos à própria edificação, às outras adjacentes, à infraestrutura pública e ao meio ambiente.

De forma a atender aos requisitos do usuário quanto à segurança (ver 4.2), devem ser atendidos os requisitos estabelecidos na legislação pertinente e na ABNT NBR 14432.

#### 8.2 Requisito – Dificultar o princípio do incêndio

Dificultar a ocorrência de princípio de incêndio por meio de premissas adotadas no projeto e na construção da edificação.

##### 8.2.1 Critérios para dificultar o princípio do incêndio

###### 8.2.1.1 Proteção contra descargas atmosféricas

Os edifícios multifamiliares devem ser providos de proteção contra descargas atmosféricas, atendendo ao estabelecido na ABNT NBR 5419 e demais Normas Brasileiras aplicáveis, nos casos previstos na legislação vigente.

###### 8.2.1.2 Proteção contra risco de ignição nas instalações elétricas

As instalações elétricas das edificações habitacionais devem ser projetadas de acordo com a ABNT NBR 5410 e Normas Brasileiras aplicáveis.

NOTA Recomenda-se evitar o risco de ignição dos materiais em função de curtos-circuitos e sobretensões.

###### 8.2.1.3 Proteção contra risco de vazamentos nas instalações de gás

As instalações de gás devem ser projetadas e executadas de acordo com as ABNT NBR 13523 e ABNT NBR 15526.

##### 8.2.2 Métodos de avaliação da segurança relativa ao princípio do incêndio

A comprovação do atendimento ao requisito de 8.2, pelos critérios estabelecidos em 8.2.1.1 a 8.2.1.3, deve ser feita pela análise do projeto ou por inspeção em protótipo.

###### 8.2.3 Premissas de projeto

Quando houver ambiente enclausurado, devem ser atendidas a ABNT NBR 15526 e outras Normas Brasileiras aplicáveis.

#### 8.3 Requisito – Facilitar a fuga em situação de incêndio

Facilitar a fuga dos usuários em situação de incêndio.

##### 8.3.1 Critério – Rotas de fuga

As rotas de saída de emergência dos edifícios devem atender ao disposto na ABNT NBR 9077.

##### 8.3.2 Métodos de avaliação

Análise do projeto ou por inspeção em protótipo.

#### 8.4 Requisito – Dificultar a inflamação generalizada

Dificultar a ocorrência da inflamação generalizada no ambiente de origem de eventual incêndio.

##### 8.4.1 Critério – Propagação superficial de chamas

Os materiais de revestimento, acabamento e isolamento termoacústico empregados na face interna dos sistemas ou elementos que compõem a edificação devem ter as características de propagação de chamas controladas, de forma a atender aos

requisitos estabelecidos nas ABNT NBR 15575-3 e a ABNT NBR 15575-5 e ABNT NBR 9442.

#### **8.4.2 Métodos de avaliação da segurança à inflamação generalizada de incêndio**

A comprovação do atendimento aos requisitos estabelecidos em 8.4.1 deve ser feita por inspeção em protótipo ou ensaios conforme Normas Brasileiras específicas.

#### **8.5 Requisito – Dificultar a propagação do incêndio**

Dificultar a propagação de incêndio para unidades contíguas ou entre edificações.

Caso não seja possível o atendimento ao critério de isolamento de risco à distância ou proteção

(8.5.1), a edificação não é considerada independente e o dimensionamento das medidas de proteção contra incêndio deve ser feito considerando o conjunto de edificações como uma única unidade.

##### **8.5.1 Critérios**

###### **8.5.1.1 Isolamento de risco à distância**

A distância entre edifícios deve atender à condição de isolamento, considerando-se todas as interferências previstas na legislação vigente.

###### **8.5.1.2 Isolamento de risco por proteção**

As medidas de proteção, incluindo no sistema construtivo o uso de portas ou selos corta-fogo, devem possibilitar que o edifício seja considerado uma unidade independente.

###### **8.5.1.3 Assegurar estanqueidade e isolamento**

Os sistemas ou elementos de compartimentação que integram as edificações habitacionais devem atender à ABNT NBR 14432 e à parte da ABNT NBR 15575 para minimizar a propagação do incêndio, assegurando estanqueidade e isolamento.

##### **8.5.2 Métodos de avaliação**

Para isolamento de risco: análise do projeto e dimensionamento das distâncias seguras, tendo em conta a ignição-piloto por radiação e a convecção através da cobertura.

Análise do projeto ou inspeção em protótipo atendendo à legislação vigente. Para os sistemas da edificação, consultar as demais partes da ABNT NBR 15575.

#### **8.6 Requisito – Segurança estrutural em situação de incêndio**

Minimizar o risco de colapso estrutural da edificação em situação de incêndio.

##### **8.6.1 Critério**

###### **8.6.1.1 Minimizar o risco de colapso estrutural**

A edificação habitacional deve atender à ABNT NBR 14432 e às normas específicas para o tipo de estrutura conforme mencionado em 8.6.2.

##### **8.6.2 Métodos de avaliação**

Análise do projeto estrutural em situação de incêndio.

Atendimento às seguintes Normas de projeto estrutural:

—ABNT NBR 14323, para estruturas de aço;

—ABNT NBR 15200, para estruturas de concreto;

—para as demais estruturas, aplica-se o Eurocode correspondente, em sua última edição.

#### **8.7 Requisito – Sistema de extinção e sinalização de incêndio**

As edificações multifamiliares devem dispor de sistemas de alarme, extinção, sinalização e iluminação

de emergência.

##### **8.7.1 Critério – Equipamentos de extinção, sinalização e iluminação de emergência**

O edifício habitacional multifamiliar deve dispor de sistemas de alarme, extinção, sinalização e iluminação de emergência, conforme proposto nas ABNT NBR 17240, ABNT NBR 13434 (Partes 1, 2 e 3), ABNT NBR 12693, ABNT NBR 13714 e ABNT NBR 10898.

##### **8.7.2 Métodos de avaliação**

Análise do projeto e, sendo possível, inspeção em protótipo atendendo à legislação vigente.

## **APÊNDICE D -Trecho preparado para processamento (Capítulo 8 da Norma de Desempenho)**

1. Os edifícios multifamiliares devem ser providos de proteção contra descargas atmosféricas, atendendo ao estabelecido na ABNT NBR 5419 e demais Normas Brasileiras aplicáveis, nos casos previstos na legislação vigente.
2. As instalações elétricas das edificações habitacionais devem ser projetadas de acordo com a ABNT NBR 5410 e Normas Brasileiras aplicáveis.
3. Recomenda-se evitar o risco de ignição dos materiais em função de curtos-circuitos e sobretensões.
4. As instalações de gás devem ser projetadas e executadas de acordo com as ABNT NBR 13523 e ABNT NBR 15526.
5. A comprovação do atendimento ao requisito de 8.2, pelos critérios estabelecidos em 8.2.1.1 a 8.2.1.3, deve ser feita pela análise do projeto ou por inspeção em protótipo.
6. Quando houver ambiente enclausurado, devem ser atendidas a ABNT NBR 15526 e outras Normas Brasileiras aplicáveis.
7. As rotas de saída de emergência dos edifícios devem atender ao disposto na ABNT NBR 9077.
8. Análise do projeto ou por inspeção em protótipo.
9. Os materiais de revestimento, acabamento e isolamento termoacústico empregados na face interna dos sistemas ou elementos que compõem a edificação devem ter as características de propagação de chamas controladas, de forma a atender aos requisitos estabelecidos nas ABNT NBR 15575-3 e a ABNT NBR 15575-5 e ABNT NBR 9442.
10. A comprovação do atendimento aos requisitos estabelecidos em 8.4.1 deve ser feita por inspeção em protótipo ou ensaios conforme Normas Brasileiras específicas.
11. A distância entre edifícios deve atender à condição de isolamento, considerando-se todas as interferências previstas na legislação vigente.
12. As medidas de proteção, incluindo no sistema construtivo o uso de portas ou selos corta-fogo, devem possibilitar que o edifício seja considerado uma unidade independente.
13. Os sistemas ou elementos de compartimentação que integram as edificações habitacionais devem atender à ABNT NBR 14432 e à parte da ABNT NBR 15575 para minimizar a propagação do incêndio, assegurando estanqueidade e isolamento.
14. Para isolamento de risco: análise do projeto e dimensionamento das distâncias seguras, tendo em conta a ignição-piloto por radiação e a convecção através da cobertura.
15. Análise do projeto ou inspeção em protótipo atendendo à legislação vigente. Para os sistemas da edificação, consultar as demais partes da ABNT NBR 15575.
16. A edificação habitacional deve atender à ABNT NBR 14432 e às normas específicas para o tipo de estrutura conforme mencionado em 8.6.2.

17. Análise do projeto estrutural em situação de incêndio.
18. Atendimento às seguintes Normas de projeto estrutural: ABNT NBR 14323, para estruturas de aço; ABNT NBR 15200, para estruturas de concreto; para as demais estruturas, aplica-se o Eurocode correspondente, em sua última edição.
19. O edifício habitacional multifamiliar deve dispor de sistemas de alarme, extinção, sinalização e iluminação de emergência, conforme proposto nas ABNT NBR 17240, ABNT NBR 13434 (Partes 1, 2 e 3), ABNT NBR 12693, ABNT NBR 13714 e ABNT NBR 10898.
20. Análise do projeto e, sendo possível, inspeção em protótipo atendendo à legislação vigente.

## APÊNDICE E - Trecho preparado da Parte 1 da Norma de Desempenho

1. Para edifícios ou conjuntos habitacionais com local de implantação definido, os projetos de arquitetura, da estrutura, das fundações, contenções e outras eventuais obras geotécnicas devem ser desenvolvidos com base nas características do local da obra (topográficas, geológicas etc.), avaliando-se convenientemente os riscos de deslizamentos, enchentes, erosões, vibrações transmitidas por vias férreas, vibrações transmitidas por trabalhos de terraplenagem e compactação do solo, ocorrência de subsidência do solo, presença de crateras em camadas profundas, presença de solos expansíveis ou colapsíveis, presença de camadas profundas deformáveis e outros.
2. Devem ainda ser considerados riscos de explosões oriundas do confinamento de gases resultantes de aterros sanitários, solos contaminados, proximidade de pedreiras e outros, tomando-se as providências necessárias para que não ocorram prejuízos à segurança e à funcionalidade da obra.
3. Os projetos devem ainda prever as interações entre construções próximas, considerando-se convenientemente as eventuais sobreposições de bulbos de pressão, efeitos de grupo de estacas, rebaixamento do lençol freático e desconfinamento do solo em função do corte do terreno.
4. Tais fenômenos também não podem prejudicar a segurança e a funcionalidade da obra, bem como de edificações vizinhas.
5. O desempenho da edificação está intimamente associado a todos os projetos de implantação e ao desempenho das fundações, devendo ser atendidas as disposições das Normas aplicáveis, particularmente das ABNT NBR 8044, ABNT NBR 5629, ABNT NBR 11682, ABNT NBR 6122 e ABNT NBR 12722.
6. Do ponto de vista da segurança e estabilidade ao longo da vida útil da estrutura, devem ser consideradas as condições de agressividade do solo, do ar e da água na época do projeto, prevendo-se, quando necessário, as proteções pertinentes à estrutura e suas partes.
7. [Trecho referente ao Capítulo 8 - *trechos usados construção dos padrões*]
8. Os sistemas não podem apresentar: a) rupturas, instabilidades, tombamentos ou quedas que possam colocar em risco a integridade física dos ocupantes ou de transeuntes nas imediações do imóvel; b) partes expostas cortantes ou perfurantes; c) deformações e defeitos acima dos limites especificados nas ABNT NBR 15575-2 a ABNT NBR 15575-6.
9. Devem ser previstas no projeto e na execução formas de minimizar, durante o uso da edificação, o risco de: a) queda de pessoas em altura: telhados, áticos, lajes de cobertura e quaisquer partes elevadas da construção; b) acessos não controlados aos locais com riscos de quedas; c) queda de pessoas em função de rupturas das proteções, as quais devem ser ensaiadas conforme ABNT NBR 14718 ou devem possuir memorial de cálculo assinado por profissional responsável que comprove seu desempenho; d) queda de pessoas em função de irregularidades nos pisos, rampas e escadas, conforme a ABNT NBR 15575-3; e) ferimentos provocados por ruptura de subsistemas ou componentes, resultando em partes cortantes ou perfurantes; f) ferimentos ou contusões em função da operação das partes móveis de

componentes, como janelas, portas, alçapões e outros; g) ferimentos ou contusões em função da dessolidarização ou da projeção de materiais ou componentes a partir das coberturas e das fachadas, tanques de lavar, pias e lavatórios, com ou sem pedestal, e de componentes ou equipamentos normalmente fixáveis em paredes; h) ferimentos ou contusões em função de explosão resultante de vazamento ou de confinamento de gás combustível.

10. Evitar a ocorrência de ferimentos ou danos aos usuários, em condições normais de uso.
11. A edificação habitacional deve atender aos requisitos das Normas específicas.
12. Atendimento aos requisitos especificados nas ABNT NBR 15575-3 a ABNT NBR 15575-5.
13. Devem ser previstos nos projetos a prevenção de infiltração da água de chuva e da umidade do solo nas habitações, por meio dos detalhes indicados a seguir: a) condições de implantação dos conjuntos habitacionais, de forma a drenar adequadamente a água de chuva incidente em ruas internas, lotes vizinhos ou mesmo no entorno próximo ao conjunto; b) sistemas que impossibilitem a penetração de líquidos ou umidades de porões e solos, jardins contíguos às fachadas e quaisquer paredes em contato com o solo, ou pelo direcionamento das águas, sem prejuízo da utilização do ambiente e dos sistemas correlatos e sem comprometer a segurança estrutural. No caso de haver sistemas de impermeabilização, estes devem seguir a ABNT NBR 9575; c) sistemas que impossibilitem a penetração de líquidos ou umidades em fundações e pisos em contato com o solo; d) ligação entre os diversos elementos da construção (como paredes e estrutura, telhado e paredes, corpo principal e pisos ou calçadas laterais).
14. Devem ser previstos no projeto detalhes que assegurem a estanqueidade de partes do edifício que tenham a possibilidade de ficar em contato com a água gerada na ocupação ou manutenção do imóvel, devendo ser verificada a adequação das vinculações entre instalações de água, esgotos ou águas pluviais e estrutura, pisos e paredes, de forma que as tubulações não venham a ser rompidas ou desencaixadas por deformações impostas.
15. O programa de simulação computacional deve ser capaz de estimar as variações da temperatura operativa, das cargas térmicas de refrigeração e de aquecimento e do uso da ventilação natural na UH, definidos separadamente em 8 760 h ao longo do ano, considerando as variações horárias de ocupação, de potência de iluminação e de equipamentos. O programa de simulação computacional deve, ainda: a) estar de acordo com a ASHRAE 140, segundo o procedimento de teste da Classe I; b) modelar efeitos de inércia térmica; c) modelar trocas de calor entre a edificação e o solo; d) calcular cargas térmicas latente e sensível; e) ser capaz de simular o sombreamento proveniente de elementos externos às zonas térmicas, como brises, sacadas e entorno; f) ser capaz de simular os efeitos da ventilação cruzada em um ambiente, ou entre dois ou mais ambientes.
16. O arquivo climático deve possuir informações que sejam representativas do clima da cidade onde a UH está localizada.
17. Caso a cidade de implantação da UH não possua arquivo climático, deve ser utilizado o arquivo climático de uma cidade próxima com clima semelhante. A semelhança entre os climas deve considerar a comparação da distância euclidiana, ponderando latitude, longitude e altitude.

18. O arquivo climático utilizado deve fornecer valores mensais de temperatura média do solo, em graus Celsius ( $^{\circ}\text{C}$ ), para todos os meses do ano, além dos seguintes valores horários representativos das 8760 h do ano meteorológico típico: a) temperatura de bulbo seco, expressa em graus Celsius ( $^{\circ}\text{C}$ ); b) temperatura do ponto de orvalho, expressa em graus Celsius ( $^{\circ}\text{C}$ ); c) umidade relativa, expressa em porcentagem (%); d) pressão atmosférica, expressa em Pascal (Pa); e) intensidade de radiação horizontal de onda longa, expressa em watts-hora por metro quadrado ( $\text{Wh}/\text{m}^2$ ); f) radiação horizontal global, expressa em watts-hora por metro quadrado ( $\text{Wh}/\text{m}^2$ ); g) radiação normal direta, expressa em watts-hora por metro quadrado ( $\text{Wh}/\text{m}^2$ ); h) radiação horizontal difusa, expressa em watts-hora por metro quadrado ( $\text{Wh}/\text{m}^2$ ); i) direção do vento, expressa em graus ( $^{\circ}$ ), considerando o sentido horário a partir da direção Norte; j) velocidade do vento, expressa em metros por segundo ( $\text{m}/\text{s}$ ).
19. Os sistemas de pisos e vedações verticais internas (paredes internas) com ou sem função estrutural que dividem unidades habitacionais autônomas ou entre unidades habitacionais e áreas comuns devem ser projetados, construídos e montados de forma a atender pelo menos aos critérios de desempenho mínimo para cada ambiente estabelecido nas ABNT NBR 15575-3 e ABNT NBR 15575-4.
20. Os sistemas de pisos que compõem os edifícios habitacionais devem atender pelo menos aos critérios mínimos estabelecidos nas ABNT NBR 15575-3 e ABNT NBR 15575-5. Este critério aplica-se também aos sistemas de pisos com função estrutural.
21. Contando unicamente com iluminação natural, os níveis gerais de iluminância nas diferentes dependências das construções habitacionais devem atender ao disposto na Tabela 4.
22. Contando unicamente com iluminação natural, o fator de luz diurna (FLD) nas diferentes dependências das construções habitacionais deve atender ao disposto na Tabela 5 (ver ISO 5034-1)
23. Os requisitos de iluminância natural podem ser atendidos mediante adequada disposição dos cômodos (arquitetura), correta orientação geográfica da edificação, dimensionamento e posição das aberturas, tipos de janelas e de envidraçamentos, rugosidade e cores dos elementos (paredes, tetos, pisos etc.), inserção de poços de ventilação e iluminação, eventual introdução de domo de iluminação etc.
24. A presença de taludes, muros, coberturas de garagens e outros obstáculos do gênero não podem prejudicar os níveis mínimos de iluminância especificados.
25. Nos conjuntos habitacionais integrados por edifícios, a implantação relativa dos prédios, de eventuais caixas de escada ou de outras construções, não podem prejudicar os níveis mínimos de iluminância especificados.
26. Os níveis gerais de iluminação promovidos nas diferentes dependências dos edifícios habitacionais por iluminação artificial devem atender ao disposto na Tabela 6.
27. O projeto deve especificar o valor teórico para a vida útil de projeto (VUP) para cada um dos sistemas que o compõem, não inferiores aos estabelecidos na Tabela 7, e deve ser elaborado para que os sistemas tenham uma durabilidade potencial compatível com a vida útil de projeto (VUP) a serem considerados nos projetos elaborados a partir da exigibilidade desta parte da ABNT NBR 15575.

28. O edifício e seus sistemas devem apresentar durabilidade compatível com a vida útil de projeto (VUP) preestabelecida em 14.2.1.
29. As condições de exposição do edifício devem ser especificadas em projeto, a fim de possibilitar uma análise da vida útil de projeto (VUP) e da durabilidade do edifício e seus sistemas.
30. As especificações relativas à manutenção, uso e operação do edifício e seus sistemas que forem consideradas em projeto para definição da vida útil de projeto (VUP) devem estar também claramente detalhadas na documentação que acompanha o edifício ou subsidia sua construção.
31. Manter a capacidade do edifício e de seus sistemas e permitir ou favorecer as inspeções prediais, bem como as intervenções de manutenção previstas no Manual de Uso, Operação e Manutenção, conforme responsabilidades estabelecidas na Seção 5.
32. Propiciar condições de salubridade no interior da edificação, considerando as condições de umidade e temperatura no interior da unidade habitacional, aliadas ao tipo dos sistemas utilizados na construção. O requisito mencionado deve atender aos critérios fixados na legislação vigente.
33. Os materiais, equipamentos e sistemas empregados na edificação não podem liberar produtos que poluam o ar em ambientes confinados, originando níveis de poluição acima daqueles verificados no entorno. Enquadram-se nesta situação os aerodispersóides, gás carbônico e outros. O requisito mencionado deve atender aos critérios fixados na legislação vigente.
34. Gases de escapamento de veículos e equipamentos não podem invadir áreas internas da habitação. O sistema de exaustão ou ventilação de garagens internas deve permitir a saída dos gases poluentes gerados por veículos e equipamentos. O requisito mencionado deve atender aos critérios fixados na legislação vigente.
35. A altura mínima de pé-direito não pode ser inferior a 2,50 m.
36. Em vestíbulos, halls, corredores, instalações sanitárias e despensas, é permitido que o pé-direito seja reduzido ao mínimo de 2,30 m. Nos tetos com vigas, inclinados, abobadados ou, em geral, contendo superfícies salientes na altura piso a piso e/ou o pé-direito mínimo, devem ser mantidos pelo menos 80 % da superfície do teto, permitindo-se na superfície restante que o pé-direito livre possa descer até o mínimo de 2,30 m.
37. Para os projetos de arquitetura de unidades habitacionais, sugere-se prever no mínimo a disponibilidade de espaço nos cômodos da edificação habitacional para colocação e utilização dos móveis e equipamentos-padrão listados no Anexo F .
38. As áreas privativas devem receber as adaptações necessárias para pessoas com deficiência física ou com mobilidade reduzida nos percentuais previstos na legislação, e as áreas de uso comum sempre devem atender ao estabelecido na ABNT NBR 9050.
39. O projeto deve prever para as áreas comuns e, quando contratado, também para as áreas privativas, as adaptações que normalmente referem-se a: a) acessos e instalações; b) substituição de escadas por rampas; c) limitação de declividades e de espaços a percorrer; d) largura de corredores e portas; e) alturas de peças sanitárias; f) disponibilidade de alças e barras de apoio.

40. No projeto e na execução das edificações térreas e assobradadas, de caráter evolutivo, deve ser prevista pelo incorporador ou construtor a possibilidade de ampliação, especificando-se os detalhes construtivos necessários para ligação ou a continuidade de paredes, pisos, coberturas e instalações.
41. O incorporador ou construtor deve anexar ao manual de uso, operação e manutenção (3.26) as especificações e detalhes construtivos necessários para ampliação do corpo da edificação, do piso, do telhado e das instalações prediais, considerando a coordenação dimensional e as compatibilidades físicas e químicas com os materiais disponíveis regionalmente, sempre que possível.
42. As especificações e detalhes construtivos fornecidos devem permitir no mínimo a manutenção dos níveis de desempenho da construção não ampliada, relativos ao comportamento estrutural, segurança ao fogo, estanqueidade à água, desempenho térmico, desempenho acústico e durabilidade.
43. As propostas de ampliação devem ser devidamente consideradas nos estudos de arquitetura, devendo atender aos níveis de funcionalidade previstos nesta Norma.
44. Os elementos e componentes da habitação (trincos, puxadores, cremonas, guilhotinas etc.) devem ser projetados, construídos e montados de forma a não provocar ferimentos nos usuários.
45. Relativamente às instalações hidrossanitárias, devem ser atendidas as disposições da ABNT NBR 15575-6. Os elementos e componentes que contam com normalização específica (portas, janelas, torneiras e outros) devem ainda atender aos requisitos das respectivas normas.
46. Os componentes, equipamentos e dispositivos de manobra devem ser projetados, construídos e montados de forma a evitar que a força necessária para o acionamento não exceda 10 N nem o torque ultrapasse 20 N.m.
47. A implantação do empreendimento deve considerar os riscos de desconfinamento do solo, deslizamentos de taludes, enchentes, erosões, assoreamento de vales ou cursos d'água, lançamentos de esgoto a céu aberto, contaminação do solo ou da água por efluentes ou outras substâncias, além de outros riscos similares.
48. Além do descrito anteriormente, devem ser atendidos os requisitos das ABNT NBR 8044 e ABNT NBR 11682, bem como da legislação vigente.
49. Recomenda-se que os empreendimentos sejam construídos mediante exploração e consumo racionalizado de recursos naturais, objetivando a menor degradação ambiental, menor consumo de água, de energia e de matérias-primas. Na medida do possível, devem ser privilegiados os materiais que causem menor impacto ambiental, desde as fases de exploração dos recursos naturais até a sua utilização final.
50. Recomenda-se a utilização de madeiras cuja origem possa ser comprovada mediante apresentação de certificação legal ou provenientes de plano de manejo aprovado pelos órgãos ambientais.
51. Durante a construção, deve-se implementar um sistema de gestão de resíduos no canteiro de obras, de forma a minimizar sua geração e possibilitar a segregação de maneira adequada para facilitar o reuso, a reciclagem ou a disposição final em locais específicos.

52. No caso de reuso de água para destinação não potável, esta deve atender aos parâmetros estabelecidos na Tabela 8.
53. As instalações elétricas devem privilegiar a adoção de soluções, caso a caso, que minimizem o consumo de energia, entre elas a utilização de iluminação e ventilação natural e de sistemas de aquecimento baseados em energia alternativa.
54. Convém a adoção de soluções que minimizem o consumo de energia, entre elas a utilização de iluminação e ventilação natural e de sistemas de aquecimento baseados em energia alternativa.

## APÊNDICE F - Lista de ontologias avaliadas

Ontology	Acronym	Exist?	Update year	Use	URL
Digital Construction	DiCON	YES			
<i>Digital Construction - Agents</i>	DICA	YES			
<i>Digital Construction - Entities</i>	DICE	YES			
<i>Digital Construction - Information</i>	DICI	YES			
<i>Digital Construction - Processes</i>	DICP	YES			
<i>Digital Construction - Contexts</i>	DCC	YES			
<i>Digital Construction - Variables</i>	DICV	YES	2022	Ontologies Suite	<a href="https://digitalconstruction.github.io/v/o.5/index.html">https://digitalconstruction.github.io/v/o.5/index.html</a>
<i>Digital Construction - Occupant Behavior</i>	DICOB	YES			
<i>Digital Construction - Indoor Air Quality</i>	-	YES			
<i>Digital Construction - Building Acoustics</i>	-	YES			
<i>Digital Construction - Energy Systems</i>	DICES	YES			
<i>Digital Construction - Building Materials</i>	DICBM	YES			
<i>Digital Construction - Lifecycle</i>	DICL	YES			
SEAS - Building Ontology	SEAS	YES	2016	Building Components	<a href="https://ci.mines-stetienne.fr/seas/BuildingOntology-1.0">https://ci.mines-stetienne.fr/seas/BuildingOntology-1.0</a>
Smart Appliances REFerence Ontology	SAREF core	YES	2020	Ontologies Suite	<a href="https://labs.etsi.org/rep/saref/saref-core">https://labs.etsi.org/rep/saref/saref-core</a>
<i>Smart Appliances REFerence Ontology for Buildings</i>	SAREF4BLDG	YES	2021	Asset Management	<a href="https://labs.etsi.org/rep/saref/saref4bldg">https://labs.etsi.org/rep/saref/saref4bldg</a>
<i>Smart Appliances REFerence Ontology for Energy</i>	SAREF4ENER	YES	2020	Asset Management	<a href="https://labs.etsi.org/rep/saref/saref4ener/-/tree/develop-v1.1.2/ontology">https://labs.etsi.org/rep/saref/saref4ener/-/tree/develop-v1.1.2/ontology</a>
Building Topology Ontology	BOT	YES	2021	Building Components	<a href="https://w3c-lbd-cg.github.io/bot/">https://w3c-lbd-cg.github.io/bot/</a>
Building Product Ontology	BOT	YES	2019	Building Components	
IFC Ontology	ifcOWL	YES	2019	Building Components	<a href="https://github.com/buildingSMART/ifcOWL">https://github.com/buildingSMART/ifcOWL</a>
Ontology Modelling for Intelligent Domotic Environments	DogOnt	YES	2019	IoT Communication	<a href="http://iot-ontologies.github.io/dogont/documentation/index-en.html">http://iot-ontologies.github.io/dogont/documentation/index-en.html</a>

BRICK	BRICK	YES	2022	Asset Management	<a href="https://github.com/BrickSchema/Brick">https://github.com/BrickSchema/Brick</a>
Haystack Tagging Ontology	HTO	YES	2023	IoT Communication	<a href="https://project-haystack.org/download">https://project-haystack.org/download</a>
Ontology for property management	OPM	YES	2018	Asset Management	<a href="https://w3c-lbd-cg.github.io/opm/">https://w3c-lbd-cg.github.io/opm/</a>
Real Estate Core Ontology	RealEstateCore ou RECore	YES	2023	Digital Twins	<a href="https://github.com/RealEstateCore/rec">https://github.com/RealEstateCore/rec</a>
e-COGNOS	e-COGNOS	NO	-		
Construction Quality Inspection and Evaluation Ontology	CQIEOntology	NO	-		
urban product ontology	UPO	NO	-		
Ontology for Processes in Infrastructure and Construction	IC-PRO-Onto	NO	2010		
LexiCon	LexiCon	NO	-		
Concepts Library Netherlands	CB-NL	NO	2023		<a href="https://github.com/bimloket/CB-NL">https://github.com/bimloket/CB-NL</a>
BIMERR	Building Ontology	YES	2020	Retrofit and Asset Management	<a href="https://bimerr.iot.linkeddata.es/def/building/">https://bimerr.iot.linkeddata.es/def/building/</a>
Bosh Building Technologies		YES	2021	Digital Twins	<a href="https://github.com/boschglobal/building-technologies-ontology-central">https://github.com/boschglobal/building-technologies-ontology-central</a>
Domain Ontolgy for Construction Knowledge	DOCK 1.0	NO	2012		
Feature Ontology	-	NO	2011		
Ontology of Technical Document	OntoDT	NO	2013	Compliance Checking	
design heuristics ontology	-	NO	2010	Shape Grammar	
e-COGNOS Knowledge Management Infrastructure	e-CKMI	NO	2003	Business Construction Processes	
Construction On-Site Ontology	IoC-OSO	YES	2021	Operations	<a href="https://github.com/Internet-of-Construction/IoC-OSO">https://github.com/Internet-of-Construction/IoC-OSO</a>
OntoArq	OntoArq	NO	2016	Architecture Documentation	
Digital Buildings Ontology	DBO	YES	2023	Asset Management	<a href="https://github.com/google/digitalbuildings">https://github.com/google/digitalbuildings</a>

## APENDICE G - Comparação dos modelos de linguagem da biblioteca SpaCy

Figura 81 - Comparação entre atributos dos modelos da biblioteca Spacy

### Modelo Pequeno

LANGUAGE	<b>PT</b> Portuguese
TYPE	<b>CORE</b> Vocabulary, syntax, entities
GENRE	<b>NEWS</b> written text (news, media)
SIZE	<b>SM</b> 12 MB
COMPONENTS <sup>?</sup>	<a href="#">tok2vec</a> , <a href="#">morphologizer</a> , <a href="#">parser</a> , <a href="#">lemmatizer</a> , <a href="#">senter</a> , <a href="#">attribute_ruler</a> , <a href="#">ner</a>
PIPELINE <sup>?</sup>	<a href="#">tok2vec</a> , <a href="#">morphologizer</a> , <a href="#">parser</a> , <a href="#">lemmatizer</a> , <a href="#">attribute_ruler</a> , <a href="#">ner</a>
VECTORS <sup>?</sup>	0 keys, 0 unique vectors (0 dimensions)

### Modelo Médio

LANGUAGE	<b>PT</b> Portuguese
TYPE	<b>CORE</b> Vocabulary, syntax, entities, vectors
GENRE	<b>NEWS</b> written text (news, media)
SIZE	<b>MD</b> 40 MB
COMPONENTS <sup>?</sup>	<a href="#">tok2vec</a> , <a href="#">morphologizer</a> , <a href="#">parser</a> , <a href="#">lemmatizer</a> , <a href="#">senter</a> , <a href="#">attribute_ruler</a> , <a href="#">ner</a>
PIPELINE <sup>?</sup>	<a href="#">tok2vec</a> , <a href="#">morphologizer</a> , <a href="#">parser</a> , <a href="#">lemmatizer</a> , <a href="#">attribute_ruler</a> , <a href="#">ner</a>
VECTORS <sup>?</sup>	500k keys, 20k unique vectors (300 dimensions)

### Modelo Grande

LANGUAGE	<b>PT</b> Portuguese
TYPE	<b>CORE</b> Vocabulary, syntax, entities, vectors
GENRE	<b>NEWS</b> written text (news, media)
SIZE	<b>LG</b> 541 MB
COMPONENTS <sup>?</sup>	<a href="#">tok2vec</a> , <a href="#">morphologizer</a> , <a href="#">parser</a> , <a href="#">lemmatizer</a> , <a href="#">senter</a> , <a href="#">attribute_ruler</a> , <a href="#">ner</a>
PIPELINE <sup>?</sup>	<a href="#">tok2vec</a> , <a href="#">morphologizer</a> , <a href="#">parser</a> , <a href="#">lemmatizer</a> , <a href="#">attribute_ruler</a> , <a href="#">ner</a>
VECTORS <sup>?</sup>	500k keys, 500k unique vectors (300 dimensions)

Fonte: adaptado de <https://spacy.io/models/pt>

O modelo com grande quantidade de atributos possui todas as palavras vetorizadas, mas é quarenta e cinco vezes maior do que o modelo pequeno, que também não possui vetores predefinidos. A versão grande do modelo em inglês possui atributos similares: 514k palavras (Keys), 514k vectors (300 dimensions) e é treinado em *corpora* de linguagem diferentes como o WordNet 3.0, ClearNLP e OntoNotes5 e Explosion Vectors.

**Quadro 45 - Comparação da acurácia dos métodos de cada modelo de língua portuguesa da biblioteca SpaCy**

Métodos	Descrição	Pequeno	Médio	Grande
TOKEN_ACC	Tokenization	1,00	1,00	1,00
TOKEN_P		1,00	1,00	1,00
TOKEN_R		1,00	1,00	1,00
TOKEN_F		1,00	1,00	1,00
POS_ACC	Part-of-speech tags (coarse grained tags, Token. pos)	0,96	0,97	0,97
MORPH_ACC	Morphological analysis	0,95	0,95	0,96
MORPH_MICRO_P		0,97	0,98	0,98
MORPH_MICRO_R		0,97	0,98	0,98
MORPH_MICRO_F		0,97	0,98	0,98
SENTS_P	Sentence segmentation (precision)	0,93	0,93	0,94
SENTS_R	Sentence segmentation (recall)	0,95	0,96	0,96
SENTS_F	Sentence segmentation (F-score)	0,93	0,95	0,95
DEP_UAS	Unlabeled dependencies	0,89	0,90	0,90
DEP_LAS	Labeled dependencies	0,84	0,86	0,86
LEMMA_ACC	Lemmatization	0,97	0,97	0,97
TAG_ACC	Part-of-speech tags (fine grained tags, Tokentag)	0,89	0,90	0,90
ENTS_P	Named entities (precision)	0,88	0,89	0,90
ENTS_R	Named entities (recall)	0,88	0,90	0,90
ENTS_F	Named entities (F-score)	0,88	0,89	0,90

Fonte: Adaptado de <https://spacy.io/models/pt>

Comparativamente, o modelo inglês possui desempenho apenas um pouco melhor nas Dependências (DEP\_UAS 0,92 e DEP\_LAS 0,90), mas precisão pior nas Entidades Nomeadas (ENTS\_P 0,85, ENTS\_R 0,86 e ENTS\_F 0,85) devido à mais rótulos de entidades que o modelo português. Atributos do analisador interno da biblioteca:

Modelos	Atributos da análise de dependência
Pequeno, Medio e Grande	ROOT, acl, acl:relcl, advcl, advmod, amod, appos, aux, aux:pass, case, cccomp, compound, conj, cop, subj, dep, det, discourse, expl, fixed, flat, flat:foreign, flat:name, iobj, mark, nmod, subj, nsubj:pass, nummod, obj, obl, obl:agent, parataxis, punct, xcomp

## APÊNDICE H – Padrões identificados no texto do Capítulo 8 da Norma

Termos CAP8	Padrão POS	Dependência	
<i>Exigências</i>			
dever ser	{VERB}{AUX}{VERB}	{ROOT}{aux:pass}{xcomp}	
dever ser	{VERB}{AUX}{VERB}	{ROOT}{aux:pass}{xcomp}	
dever ser	{VERB}{AUX}{VERB}	{ROOT}{aux:pass}{xcomp}	
deve ser	{VERB}{AUX}{VERB}	{ROOT}{aux:pass}{xcomp}	
dever ser	{VERB}{AUX}{VERB}	{ROOT}{aux:pass}{xcomp}	
dever atender	{VERB}{VERB}{ADP}	{ROOT}{xcomp}{case}	
devem ter	{VERB}{VERB}{NOUN}	{ROOT}{xcomp}{det}{obj}	
deve ser	{VERB}{AUX}{VERB}	{ROOT}{aux:pass}{xcomp}	
dever atender	{VERB}{VERB}{ADP}	{ROOT}{xcomp}{case}	
dever possibilitar	{VERB}{VERB}{SCONJ}	{ROOT}{xcomp}{mark}	
dever atender	{VERB}{VERB}{ADP}{PROPN}	{ROOT}{xcomp}{case}{obj}	
Termos CAP8	Padrão POS	Dependência	Morfologia
<i>Referência à norma</i>			
ABNT NBR xxxx	{PROPN}{PROPN}{NUM}	{obl}{appos}{nummod}	{Gender=Fem Number=Sing}{NumType=Card}
ABNT NBR xxxxx	{PROPN}{ADP}{NUM}	{nmod}{appos}{nummod}	{Gender=Fem Number=Sing}{NumType=Card}
ABNT NBR xxxxx	{PROPN}{ADP}{NUM}	{nmod}{appos}{nummod}	{Gender=Fem Number=Sing}{NumType=Card}
ABNT NBR xxxxx	{PROPN}{ADP}{NUM}	{conj}{appos}{flat:name}	{Gender=Fem Number=Sing}{NumType=Card}
requisito de 8.2	{NOUN}{ADP}{NUM}	{nmod}{case}{nmod}	{Gender=Masc Number=Sing}{NumType=Card}
ABNT NBR xxxxx	{PROPN}{ADP}{NUM}	{nsubj}{case}{obl}	{Gender=Fem Number=Sing}{NumType=Card}
ABNT NBR xxxx	{PROPN}{PROPN}{NUM}	{obl}{appos}{nummod}	{Gender=Fem Number=Sing}{Gender=Masc Number=sing}{NumType=Card}
ABNT NBR xxxxx-x	{PROPN}{ADP}{NUM}	{obl}{case}{obl}	{Gender=Fem Number=Sing}{NumType=Card}
ABNT NBR xxxxx-x	{PROPN}{ADP}{NUM}	{conj}{ROOT}{nummod}	{Gender=Fem Number=Sing}{NumType=Card}
ABNT NBR xxxx	{PROPN}{PROPN}{NUM}	{conj}{appos}{nummod}	{Gender=Fem Number=Sing}{Gender=Masc Number=sing}{NumType=Card}
ABNT NBR xxxxx	{PROPN}{ADP}{NUM}	{obl}{appos}{nummod}	{Gender=Fem Number=Sing}{NumType=Card}

ABNT NBR xxxxx	{PROPN}{ADP}{NUM}	{nmod}{case}{ROOT}	{Gender=Fem Number=Sing}}{NumType=Card}
ABNT NBR xxxxx	{PROPN}{ADP}{NUM}	{nmod}{case}{ROOT}	{Gender=Fem Number=Sing}}{NumType=Card}
ABNT NBR xxxxx	{PROPN}{ADP}{NUM}	{obl}{appos}{nummod}	{Gender=Fem Number=Sing}}{NumType=Card}
ABNT NBR xxxxx	{NOUN}{ADP}{NUM}	{nsubj}{appos}{obj}	{Gender=Fem Number=Sing}}{NumType=Card}
ABNT NBR xxxxx	{NOUN}{ADP}{NUM}	{appos}{apppos}{nummod}	{Gender=Fem Number=Sing}}{NumType=Card}
ABNT NBR xxxxx	{PROPN}{ADP}{NUM}	{obl}{ROOT}{obj}	{Gender=Fem Number=Sing}}{NumType=Card}
ABNT NBR xxxxx	{PROPN}{PROPN}{NUM}	{appos}{flat:name}{nummod}	{Gender=Fem Number=Sing}}{Gender=Masc Number=sing}}{NumType=Card}
ABNT NBR xxxxx	{PROPN}{ADP}{NUM}	{dep}{appos}{nummod}	{Gender=Fem Number=Sing}}{NumType=Card}
ABNT NBR xxxxx	{PROPN}{ADP}{NUM}	{dep}{case}{ROOT}	{Gender=Fem Number=Sing}}{NumType=Card}
ABNT NBR xxxxx	{PROPN}{ADP}{NUM}	{conj}{appos}{nmod}	{Gender=Fem Number=Sing}}{NumType=Card}

<b>Termos CAP8</b>	<b>Padrão POS</b>	<b>Dependência</b>
<i>Componente</i>		
edifício multifamiliar	{NOUN}{ADJ}	{nsubj:pass}{amod}
instalação elétrico	{NOUN}{ADJ}	{nsubj:pass}{amod}
edificação habitacional	{NOUN}{ADJ}	{nmod}{amod}
instalação de gás	{NOUN}{ADP}{NOUN}	{nsubj:pass}{case}{nmod}
ambiente	{NOUN}	{obj}
saída de emergência	{NOUN}{ADP}{NOUN}	{nmod}{case}{nmod}
edifício	{NOUN}	{nmod}
material de revestimento	{NOUN}{ADP}{NOUN}	{nsubj}{case}{nmod}
acabamento	{NOUN}	{conj}
isolamento termoacústico	{NOUN}{ADJ}	{conj}{amod}
sistema	{NOUN}	{nmod}
edificação	{NOUN}	{obj}
sistema construtivo	{NOUN}{ADJ}	{obl}{amod}
porta	{NOUN}	{nmod}
selo corta-fogo	{NOUN}{ADJ}	{conj}{amod}
edifício	{NOUN}	{nsubj:pass}
unidade independente	{NOUN}{ADJ}	{nsubj:pass}{amod}
sistema	{NOUN}	{nsubj}

elemento de compartimentação	{NOUN}{ADP}{ADJ}	{conj}{case}{nmod}
edificação habitacional	{NOUN}{ADJ}	{objt}{amod}
ignição-piloto	{NOUN}	{obj}
cobertura	{NOUN}	{obl}
edificação habitacional	{NOUN}{ADJ}	{nsubj}{amod}
estrutura	{NOUN}	{nmod}
projeto estrutural	{NOUN}{ADJ}	{nmod}{amod}
situação de incêndio	{NOUN}{ADP}{NOUN}	{nmod}{case}{nmod}
projeto estrutural	{NOUN}{ADJ}	{nmod}{amod}
estrutura de aço	{NOUN}{ADP}{NOUN}	{nmod}{case}{nmod}
estrutura de concreto	{NOUN}{ADP}{NOUN}	{nmod}{case}{nmod}
demais estruturas	{DET}{NOUN}	{det}{obl}
edifício habitacional multifamiliar	{NOUN}{ADJ}{ADJ}	{nsubj}{amod}{amod}
sistema de alarme	{NOUN}{ADP}{NOUN}	{obj}{compound}{compound}
extinção	{NOUN}	{conj}
sinalização	{NOUN}	{conj}
iluminação de emergência	{NOUN}{ADP}{NOUN}	{conj}{case}{nmod}
<b>Termos CAP8</b>	<b>Padrão POS</b>	<b>Dependência</b>
<i>Requisito de elemento da construção</i>		
proteção contra descarga atmosférico	{NOUN}{ADP}{NOUN}{ADJ}	{obj}{case}{nmod}{amod}
face interna	{NOUN}{ADJ}	{obl}{amod}
propagação de chamas controladas	{NOUN}{ADP}{NOUN}{VERB}	{nmod}{case}{nmod}{acl}
condição de isolamento	{NOUN}{ADP}{NOUN}	{obj}{case}{nmod}
medida de proteção	{NOUN}{ADP}{NOUN}	{nsubj}{case}{nmod}
estanqueidade		
isolamento		
isolamento de risco		
<b>Termos CAP8</b>	<b>Padrão POS</b>	<b>Dependência</b>

## Método de avaliação

análise de o projeto	{NOUN}{ADP}{NOUN}	{obl:agente}{case}{nmod}
inspeção em protótipo	{NOUN}{ADP}{NOUN}	{conj}{case}{nmod}
análise de o projeto	{NOUN}{ADP}{NOUN}	{ROOT}{case}{nmod}
inspeção em protótipo	{NOUN}{ADP}{NOUN}	{conj}{case}{nmod}
inspeção em protótipo	{NOUN}{ADP}{NOUN}	{obl:agente}{case}{nmod}
ensaio	{NOUN}	{conj}
análise de o projeto	{NOUN}{ADP}{NOUN}	{nsubj}{case}{nmod}
inspeção em protótipo	{NOUN}{ADP}{NOUN}	{conj}{case}{nmod}
análise de o projeto	{NOUN}{ADP}{NOUN}	{ROOT}{case}{nmod}
análise de o projeto	{NOUN}{ADP}{NOUN}	{ROOT}{case}{nmod}
inspeção em protótipo	{NOUN}{ADP}{NOUN}	{obl}{case}{nmod}

## APÊNDICE I - Problema Extração de Padrões

**Frase 1:** Os edifícios multifamiliares devem ser providos de proteção contra descargas atmosféricas, atendendo ao estabelecido na ABNT NBR 5419 e demais Normas Brasileiras aplicáveis, nos casos previstos na legislação vigente.

edifícios multifamiliares ( COMPONENTE )  
 devem ( EXIGENCIA )  
 devem ser providos ( EXIGENCIA )  
 providos ( EXIGENCIA )  
 providos de proteção ( EXIGENCIA )  
 proteção contra descargas ( COMPONENTE )  
 descargas atmosféricas ( COMPONENTE )  
 atendendo ( EXIGENCIA )  
 atendendo ao estabelecido ( EXIGENCIA )  
 ABNT NBR 5419 ( REFERENCIA )  
 Normas Brasileiras ( COMPONENTE )  
 previstos ( EXIGENCIA )  
 previstos na legislação ( EXIGENCIA )  
 legislação vigente ( COMPONENTE )

**Frase 2:** As instalações elétricas das edificações habitacionais devem ser projetadas de acordo com a ABNT NBR 5410 e Normas Brasileiras aplicáveis.

instalações elétricas ( COMPONENTE )  
 edificações habitacionais ( COMPONENTE )  
 devem ( EXIGENCIA )  
 devem ser projetadas ( EXIGENCIA )  
 projetadas ( EXIGENCIA )  
 projetadas de acordo ( EXIGENCIA )  
 ABNT NBR 5410 ( REFERENCIA )

**Frase 3:** Recomenda-se evitar o risco de ignição dos materiais em função de curtos-circuitos e sobretensões.

Recomenda-se ( EXIGENCIA )  
 Recomenda-se evitar ( EXIGENCIA )  
 evitar ( EXIGENCIA )  
 risco de ignição ( COMPONENTE )  
 risco de ignição ( PARAMETRO )  
 ignição dos materiais ( COMPONENTE )  
 materiais em função ( COMPONENTE )  
 função de curtos-circuitos ( COMPONENTE )

**Frase 4:** As instalações de gás devem ser projetadas e executadas de acordo com as ABNT NBR 13523 e ABNT NBR 15526.

instalações de gás ( COMPONENTE )  
 devem ( EXIGENCIA )  
 devem ser projetadas ( EXIGENCIA )  
 projetadas ( EXIGENCIA )  
 executadas ( EXIGENCIA )  
 executadas de acordo ( EXIGENCIA )  
 ABNT NBR 13523 ( REFERENCIA )

ABNT NBR 15526 ( REFERENCIA )

**Frase 5:** A comprovação do atendimento ao requisito de 8.2, pelos critérios estabelecidos em 8.2.1.1 a 8.2.1.3, deve ser feita pela análise do projeto ou por inspeção em protótipo.

comprovação do atendimento ( COMPONENTE )  
 atendimento ao requisito ( COMPONENTE )  
 estabelecidos ( EXIGENCIA )  
 deve ( EXIGENCIA )  
 deve ser feita ( EXIGENCIA )  
 feita ( EXIGENCIA )  
 feita pela análise ( EXIGENCIA )  
 análise do projeto ( COMPONENTE )  
 inspeção em protótipo ( COMPONENTE )

**Frase 6:** Quando houver ambiente enclausurado, devem ser atendidas a ABNT NBR 15526 e outras Normas Brasileiras aplicáveis.

houver ( EXIGENCIA )  
 enclausurado ( EXIGENCIA )  
 devem ( EXIGENCIA )  
 devem ser atendidas ( EXIGENCIA )  
 atendidas ( EXIGENCIA )  
 ABNT NBR 15526 ( REFERENCIA )  
 Normas Brasileiras ( COMPONENTE )

**Frase 7:** As rotas de saída de emergência dos edifícios devem atender ao disposto na ABNT NBR 9077.

rotas ( COMPONENTE )  
 rotas de saída ( COMPONENTE )  
 saída de emergência ( COMPONENTE )  
 emergência dos edifícios ( COMPONENTE )  
 devem ( EXIGENCIA )  
 devem atender ( EXIGENCIA )  
 atender ( EXIGENCIA )  
 devem atender ao ( EXIGENCIA )  
 atender ao disposto ( EXIGENCIA )  
 ABNT NBR 9077 ( REFERENCIA )

**Frase 8:** Análise do projeto ou por inspeção em protótipo.

Análise do projeto ( COMPONENTE )  
 inspeção em protótipo ( COMPONENTE )

## APÊNDICE J- Resultado de Extração de Informações - Capítulo 8 da Norma de Desempenho

1	Os edifícios multifamiliares devem ser providos de proteção contra descargas atmosféricas, atendendo ao estabelecido na ABNT NBR 5419 e demais Normas Brasileiras aplicáveis, nos casos previstos na legislação vigente.	VP	6
		FN	0
		FP	0
		VN	3
2	As instalações elétricas das edificações habitacionais devem ser projetadas de acordo com a ABNT NBR 5410 e Normas Brasileiras aplicáveis.	VP	5
		FN	0
		FP	0
		VN	1
3	Recomenda-se evitar o risco de ignição dos materiais em função de curtos-circuitos e sobretensões.	VP	3
		FN	1
		FP	1
		VN	0
4	As instalações de gás devem ser projetadas e executadas de acordo com as ABNT NBR 13523 e ABNT NBR 15526.	VP	4
		FN	1
		FP	0
		VN	1
5	A comprovação do atendimento ao requisito de 8.2, pelos critérios estabelecidos em 8.2.1.1 a 8.2.1.3, deve ser feita pela análise do projeto ou por inspeção em protótipo.	VP	4
		FN	3
		FP	0
		VN	3
6	Quando houver ambiente enclausurado, devem ser atendidas a ABNT NBR 15526 e outras Normas Brasileiras aplicáveis.	VP	3
		FN	1
		FP	0
		VN	3
7	As rotas de saída de emergência dos edifícios devem atender ao disposto na ABNT NBR 9077.	VP	3
		FN	0
		FP	0
		VN	2
8	Análise do projeto ou por inspeção em protótipo.	VP	2
		FN	0
		FP	0
		VN	1
9	Os materiais de revestimento, acabamento e isolamento termoacústico empregados na face interna dos sistemas ou elementos que compõem a edificação devem ter as características de propagação de chamas controladas, de forma a atender aos requisitos estabelecidos nas ABNT NBR 15575-3 e a ABNT NBR 15575-5 e ABNT NBR 9442.	VP	8
		FN	3
		FP	0
		VN	6
10	A comprovação do atendimento aos requisitos estabelecidos em 8.4.1 deve ser feita por inspeção em protótipo ou ensaios conforme Normas Brasileiras específicas.	VP	4
		FN	2
		FP	0

		VN	2
11	A distância entre edifícios deve atender à condição de isolamento, considerando-se todas as interferências previstas na legislação vigente.	VP	4
		FN	1
		FP	0
		VN	2
12	As medidas de proteção, incluindo no sistema construtivo o uso de portas ou selos corta-fogo, devem possibilitar que o edifício seja considerado uma unidade independente.	VP	6
		FN	1
		FP	1
		VN	2
13	Os sistemas ou elementos de compartimentação que integram as edificações habitacionais devem atender à ABNT NBR 14432 e à parte da ABNT NBR 15575 para minimizar a propagação do incêndio, assegurando estanqueidade e isolamento.	VP	7
		FN	2
		FP	0
		VN	2
14	Para isolamento de risco: análise do projeto e dimensionamento das distâncias seguras, tendo em conta a ignição-piloto por radiação e a convecção através da cobertura.	VP	5
		FN	0
		FP	0
		VN	2
15	Análise do projeto ou inspeção em protótipo atendendo à legislação vigente. Para os sistemas da edificação, consultar as demais partes da ABNT NBR 15575.	VP	4
		FN	1
		FP	0
		VN	4
16	A edificação habitacional deve atender à ABNT NBR 14432 e às normas específicas para o tipo de estrutura conforme mencionado em 8.6.2.	VP	5
		FN	1
		FP	1
		VN	2
17	Análise do projeto estrutural em situação de incêndio.	VP	2
		FN	0
		FP	0
		VN	0
18	Atendimento às seguintes Normas de projeto estrutural: ABNT NBR 14323, para estruturas de aço; ABNT NBR 15200, para estruturas de concreto; para as demais estruturas, aplica-se o Eurocode correspondente, em sua última edição.	VP	6
		FN	0
		FP	2
		VN	3
19	O edifício habitacional multifamiliar deve dispor de sistemas de alarme, extinção, sinalização e iluminação de emergência, conforme proposto nas ABNT NBR 17240, ABNT NBR 13434 (Partes 1, 2 e 3), ABNT NBR 12693, ABNT NBR 13714 e ABNT NBR 10898.	VP	9
		FN	2
		FP	0
		VN	3
20	Análise do projeto e, sendo possível, inspeção em protótipo atendendo à legislação vigente.	VP	3
		FN	0
		FP	0
		VN	2

## APÊNDICE K - Resultado de Extração de Informações - Parte 1 da Norma de Desempenho

1	Para edifícios ou conjuntos habitacionais com local de implantação definido, os projetos de arquitetura, da estrutura, das fundações, contenções e outras eventuais obras geotécnicas devem ser desenvolvidos com base nas características do local da obra (topográficas, geológicas etc.), avaliando-se convenientemente os riscos de deslizamentos, enchentes, erosões, vibrações transmitidas por vias férreas, vibrações transmitidas por trabalhos de terraplenagem e compactação do solo, ocorrência de subsidência do solo, presença de crateras em camadas profundas, presença de solos expansíveis ou colapsíveis, presença de camadas profundas deformáveis e outros.	VP	21
		FN	0
		FP	4
		VN	8
2	Devem ainda ser considerados riscos de explosões oriundas do confinamento de gases resultantes de aterros sanitários, solos contaminados, proximidade de pedreiras e outros, tomando-se as providências necessárias para que não ocorram prejuízos à segurança e à funcionalidade da obra.	VP	7
		FN	3
		FP	2
		VN	3
3	Os projetos devem ainda prever as interações entre construções próximas, considerando-se convenientemente as eventuais sobreposições de bulbos de pressão, efeitos de grupo de estacas, rebaixamento do lençol freático e desconfinamento do solo em função do corte do terreno.	VP	7
		FN	2
		FP	1
		VN	2
4	Tais fenômenos também não podem prejudicar a segurança e a funcionalidade da obra, bem como de edificações vizinhas.	VP	4
		FN	1
		FP	0
		VN	2
5	O desempenho da edificação está intimamente associado a todos os projetos de implantação e ao desempenho das fundações, devendo ser atendidas as disposições das Normas aplicáveis, particularmente das ABNT NBR 8044, ABNT NBR 5629, ABNT NBR 11682, ABNT NBR 6122 e ABNT NBR 12722.	VP	7
		FN	4
		FP	2
		VN	1
6	Do ponto de vista da segurança e estabilidade ao longo da vida útil da estrutura, devem ser consideradas as condições de agressividade do solo, do ar e da água na época do projeto, prevendo-se, quando necessário, as proteções pertinentes à estrutura e suas partes.	VP	7
		FN	4
		FP	2
		VN	1
7	NA	VP	0
		FN	0
		FP	0
		VN	0
8	Os sistemas não podem apresentar: a) rupturas, instabilidades, tombamentos ou quedas que possam colocar em risco a integridade física dos ocupantes ou de transeuntes nas imediações do imóvel; b) partes expostas cortantes ou perfurantes; c) deformações e defeitos acima dos limites especificados nas ABNT NBR 15575-2 a ABNT NBR 15575-6	VP	7
		FN	10
		FP	0
		VN	0
9	Devem ser previstas no projeto e na execução formas de minimizar, durante o uso da edificação, o risco de: a) queda de pessoas em altura: telhados, áticos, lajes de cobertura e quaisquer partes elevadas da construção; b) acessos não controlados aos locais com riscos de quedas; c) queda de pessoas em função de rupturas das proteções, as quais devem ser ensaiadas conforme ABNT NBR 14718 ou devem possuir memorial de cálculo assinado por profissional responsável que comprove seu desempenho; d) queda de pessoas em função de irregularidades nos pisos, rampas e escadas, conforme a ABNT NBR 15575-3; e) ferimentos provocados por ruptura de subsistemas ou componentes, resultando em partes cortantes ou perfurantes; f) ferimentos ou contusões em função da operação das partes móveis de componentes, como janelas, portas, alçapões e outros; g) ferimentos ou contusões em função da dessolidarização ou da projeção de materiais ou componentes a partir das coberturas e das fachadas, tanques de lavar, pias e lavatórios, com ou sem pedestal, e de componentes ou equipamentos normalmente fixáveis em paredes; h) ferimentos ou contusões em função de explosão resultante de vazamento ou de confinamento de gás combustível.	VP	22
		FN	32
		FP	2
		VN	13

10	Evitar a ocorrência de ferimentos ou danos aos usuários, em condições normais de uso.	VP	2
		FN	2
		FP	0
		VN	0
11	A edificação habitacional deve atender aos requisitos das Normas específicas.	VP	3
		FN	1
		VN	1
12	Atendimento aos requisitos especificados nas ABNT NBR 15575-3 a ABNT NBR 15575-5.	VP	4
		FN	0
		FP	0
		VN	1
13	Devem ser previstos nos projetos a prevenção de infiltração da água de chuva e da umidade do solo nas habitações, por meio dos detalhes indicados a seguir: a) condições de implantação dos conjuntos habitacionais, de forma a drenar adequadamente a água de chuva incidente em ruas internas, lotes vizinhos ou mesmo no entorno próximo ao conjunto; b) sistemas que impossibilitem a penetração de líquidos ou umidades de porões e subsolos, jardins contíguos às fachadas e quaisquer paredes em contato com o solo, ou pelo direcionamento das águas, sem prejuízo da utilização do ambiente e dos sistemas correlatos e sem comprometer a segurança estrutural. No caso de haver sistemas de impermeabilização, estes devem seguir a ABNT NBR 9575; c) sistemas que impossibilitem a penetração de líquidos ou umidades em fundações e pisos em contato com o solo; d) ligação entre os diversos elementos da construção (como paredes e estrutura, telhado e paredes, corpo principal e pisos ou calçadas laterais).	VP	25
		FN	14
		FP	1
		VN	7
14	Devem ser previstos no projeto detalhes que assegurem a estanqueidade de partes do edifício que tenham a possibilidade de ficar em contato com a água gerada na ocupação ou manutenção do imóvel, devendo ser verificada a adequação das vinculações entre instalações de água, esgotos ou águas pluviais e estrutura, pisos e paredes, de forma que as tubulações não venham a ser rompidas ou desencaixadas por deformações impostas.	VP	3
		FN	11
		FP	2
		VN	4
15	O programa de simulação computacional deve ser capaz de estimar as variações da temperatura operativa, das cargas térmicas de refrigeração e de aquecimento e do uso da ventilação natural na UH, definidos separadamente em 8 760 h ao longo do ano, considerando as variações horárias de ocupação, de potência de iluminação e de equipamentos. O programa de simulação computacional deve, ainda: a) estar de acordo com a ASHRAE 140, segundo o procedimento de teste da Classe I; b) modelar efeitos de inércia térmica; c) modelar trocas de calor entre a edificação e o solo; d) calcular cargas térmicas latente e sensível; e) ser capaz de simular o sombreamento proveniente de elementos externos às zonas térmicas, como brises, sacadas e entorno; f) ser capaz de simular os efeitos da ventilação cruzada em um ambiente, ou entre dois ou mais ambientes.	VP	12
		FN	18
		FP	2
		VN	5
16	O arquivo climático deve possuir informações que sejam representativas do clima da cidade onde a UH está localizada.	VP	4
		FN	2
		FP	0
		VN	3
17	Caso a cidade de implantação da UH não possua arquivo climático, deve ser utilizado o arquivo climático de uma cidade próxima com clima semelhante. A semelhança entre os climas deve considerar a comparação da distância euclidiana, ponderando latitude, longitude e altitude.	VP	8
		FN	4
		FP	1
		VN	4
18	O arquivo climático utilizado deve fornecer valores mensais de temperatura média do solo, em graus Celsius (°C), para todos os meses do ano, além dos seguintes valores horários representativos das 8760 h do ano meteorológico típico: a) temperatura de bulbo seco, expressa em graus Celsius (°C); b) temperatura do ponto de orvalho, expressa em graus Celsius (°C); c) umidade relativa, expressa em porcentagem (%); d) pressão atmosférica, expressa em Pascal (Pa); e) intensidade de radiação horizontal de onda longa, expressa em watts-hora por metro quadrado (Wh/m <sup>2</sup> ); f) radiação horizontal global, expressa em watts-hora por metro quadrado (Wh/m <sup>2</sup> ); g) radiação normal direta, expressa em watts-hora por metro quadrado (Wh/m <sup>2</sup> ); h) radiação horizontal difusa, expressa em watts-hora por metro quadrado (Wh/m <sup>2</sup> ); i) direção do vento, expressa em graus (°), considerando o sentido horário a partir da direção Norte; j) velocidade do vento, expressa em metros por segundo (m/s).	VP	10
		FN	18
		FP	6
		VN	14
19		VP	6
		FN	8

	Os sistemas de pisos e vedações verticais internas (paredes internas) com ou sem função estrutural que dividem unidades habitacionais autônomas ou entre unidades habitacionais e áreas comuns devem ser projetados, construídos e montados de forma a atender pelo menos aos critérios de desempenho mínimo para cada ambiente estabelecido nas ABNT NBR 15575-3 e ABNT NBR 15575-4.	FP VN	0 6
20	Os sistemas de pisos que compõem os edifícios habitacionais devem atender pelo menos aos critérios mínimos estabelecidos nas ABNT NBR 15575-3 e ABNT NBR 15575-5. Este critério aplica-se também aos sistemas de pisos com função estrutural.	VP FN FP VN	8 2 1 2
21	Contando unicamente com iluminação natural, os níveis gerais de iluminância nas diferentes dependências das construções habitacionais devem atender ao disposto na Tabela 4.	VP FN FP VN	3 2 0 3
22	Contando unicamente com iluminação natural, o fator de luz diurna (FLD) nas diferentes dependências das construções habitacionais deve atender ao disposto na Tabela 5 (ver ISO 5034-1).	VP FN FP VN	4 3 0 3
23	Os requisitos de iluminância natural podem ser atendidos mediante adequada disposição dos cômodos (arquitetura), correta orientação geográfica da edificação, dimensionamento e posição das aberturas, tipos de janelas e de envidraçamentos, rugosidade e cores dos elementos (paredes, tetos, pisos etc.), inserção de poços de ventilação e iluminação, eventual introdução de domo de iluminação etc.	VP FN FP VN	8 6 3 4
24	A presença de taludes, muros, coberturas de garagens e outros obstáculos do gênero não podem prejudicar os níveis mínimos de iluminância especificados.	VP FN FP VN	5 2 2 2
25	Nos conjuntos habitacionais integrados por edifícios, a implantação relativa dos prédios, de eventuais caixas de escada ou de outras construções, não podem prejudicar os níveis mínimos de iluminância especificados.	VP FN FP VN	4 4 0 3
26	Os níveis gerais de iluminação promovidos nas diferentes dependências dos edifícios habitacionais por iluminação artificial devem atender ao disposto na Tabela 6.	VP FN FP VN	4 1 0 3
27	O projeto deve especificar o valor teórico para a vida útil de projeto (VUP) para cada um dos sistemas que o compõem, não inferiores aos estabelecidos na Tabela 7, e deve ser elaborado para que os sistemas tenham uma durabilidade potencial compatível com a vida útil de projeto (VUP) a serem considerados nos projetos elaborados a partir da exigibilidade desta parte da ABNT NBR 15575.	VP FN FP VN	4 8 2 6
28	O edifício e seus sistemas devem apresentar durabilidade compatível com a vida útil de projeto (VUP) preestabelecida em 14.2.1.	VP FN FP VN	2 4 0 1
29	As condições de exposição do edifício devem ser especificadas em projeto, a fim de possibilitar uma análise da vida útil de projeto (VUP) e da durabilidade do edifício e seus sistemas.	VP FN FP VN	3 4 0 2
30	As especificações relativas à manutenção, uso e operação do edifício e seus sistemas que forem consideradas em projeto para definição da vida útil de projeto (VUP) devem estar também claramente detalhadas na documentação que acompanha o edifício ou subsidia sua construção.	VP FN FP VN	3 10 2 3
31	Manter a capacidade do edifício e de seus sistemas e permitir ou favorecer as inspeções prediais, bem como as intervenções de manutenção previstas no Manual de Uso, Operação e Manutenção, conforme responsabilidades estabelecidas na Seção 5.	VP FN FP VN	3 7 0 5
32	Propiciar condições de salubridade no interior da edificação, considerando as condições de umidade e temperatura no interior da unidade habitacional, aliadas ao tipo dos sistemas utilizados na construção. O requisito mencionado deve atender aos critérios fixados na legislação vigente.	VP FN FP VN	8 5 0 5
33	Os materiais, equipamentos e sistemas empregados na edificação não podem liberar produtos que poluam o ar em ambientes confinados, originando níveis de poluição acima daqueles verificados no entorno. Enquadram-se nesta situação os aerodispersóides, gás carbônico e outros. O requisito mencionado deve atender aos critérios fixados na legislação vigente.	VP FN FP VN	10 9 0 4

34	Gases de escapamento de veículos e equipamentos não podem invadir áreas internas da habitação. O sistema de exaustão ou ventilação de garagens internas deve permitir a saída dos gases poluentes gerados por veículos e equipamentos. O requisito mencionado deve atender aos critérios fixados na legislação vigente.	VP	11
		FN	6
		FP	0
		VN	2
35	A altura mínima de pé-direito não pode ser inferior a 2,50 m.	VP	3
		FN	1
		FP	0
		VN	0
36	Em vestíbulos, halls, corredores, instalações sanitárias e despensas, é permitido que o pé-direito seja reduzido ao mínimo de 2,30 m. Nos tetos com vigas, inclinados, abobadados ou, em geral, contendo superfícies salientes na altura piso a piso e/ou o pé-direito mínimo, devem ser mantidos pelo menos 80 % da superfície do teto, permitindo-se na superfície restante que o pé-direito livre possa descer até o mínimo de 2,30 m.	VP	11
		FN	12
		FP	2
		VN	6
37	Para os projetos de arquitetura de unidades habitacionais, sugere-se prever no mínimo a disponibilidade de espaço nos cômodos da edificação habitacional para colocação e utilização dos móveis e equipamentos-padrão listados no Anexo F.	VP	7
		FN	4
		FP	1
		VN	3
38	As áreas privativas devem receber as adaptações necessárias para pessoas com deficiência física ou com mobilidade reduzida nos percentuais previstos na legislação, e as áreas de uso comum sempre devem atender ao estabelecido na ABNT NBR 9050.	VP	7
		FN	4
		FP	0
		VN	3
39	O projeto deve prever para as áreas comuns e, quando contratado, também para as áreas privativas, as adaptações que normalmente referem-se a: a) acessos e instalações; b) substituição de escadas por rampas; c) limitação de declividades e de espaços a percorrer; d) largura de corredores e portas; e) alturas de peças sanitárias; f) disponibilidade de alças e barras de apoio.	VP	10
		FN	4
		FP	0
		VN	4
40	No projeto e na execução das edificações térreas e assobradadas, de caráter evolutivo, deve ser prevista pelo incorporador ou construtor a possibilidade de ampliação, especificando-se os detalhes construtivos necessários para ligação ou a continuidade de paredes, pisos, coberturas e instalações.	VP	6
		FN	7
		FP	0
		VN	4
41	O incorporador ou construtor deve anexar ao manual de uso, operação e manutenção (3.26) as especificações e detalhes construtivos necessários para ampliação do corpo da edificação, do piso, do telhado e das instalações prediais, considerando a coordenação dimensional e as compatibilidades físicas e químicas com os materiais disponíveis regionalmente, sempre que possível.	VP	9
		FN	8
		FP	0
		VN	4
42	As especificações e detalhes construtivos fornecidos devem permitir no mínimo a manutenção dos níveis de desempenho da construção não ampliada, relativos ao comportamento estrutural, segurança ao fogo, estanqueidade à água, desempenho térmico, desempenho acústico e durabilidade.	VP	9
		FN	3
		FP	0
		VN	2
43	As propostas de ampliação devem ser devidamente consideradas nos estudos de arquitetura, devendo atender aos níveis de funcionalidade previstos nesta Norma.	VP	4
		FN	4
		FP	0
		VN	2
44	Os elementos e componentes da habitação (trincos, puxadores, cremonas, guilhotinas etc.) devem ser projetados, construídos e montados de forma a não provocar ferimentos nos usuários.	VP	3
		FN	8
		FP	0
		VN	1
45	Relativamente às instalações hidrossanitárias, devem ser atendidas as disposições da ABNT NBR 15575-6. Os elementos e componentes que contam com normalização específica (portas, janelas, torneiras e outros) devem ainda atender aos requisitos das respectivas normas.	VP	4
		FN	7
		FP	1
		VN	3
46	Os componentes, equipamentos e dispositivos de manobra devem ser projetados, construídos e montados de forma a evitar que a força necessária para o acionamento não exceda 10 N nem o torque ultrapasse 20 N.m.	VP	5
		FN	10
		FP	0
		VN	3
47	A implantação do empreendimento deve considerar os riscos de desconfinamento do solo, deslizamentos de taludes, enchentes, erosões, assoreamento de vales ou cursos d'água, lançamentos de esgoto a céu aberto, contaminação do solo ou da água por efluentes ou outras substâncias, além de outros riscos similares.	VP	9
		FN	5
		FP	1
		VN	1
		VP	4

48	Além do descrito anteriormente, <b>devem ser atendidos</b> os requisitos das <b>ABNT NBR 8044</b> e <b>ABNT NBR 11682</b> , bem como da <b>legislação vigente</b> .	FN	0
		FP	0
		VN	3
49	Recomenda-se que os <b>empreendimentos sejam construídos</b> mediante <b>exploração e consumo racionalizado de recursos naturais</b> , objetivando a <b>menor degradação ambiental</b> , <b>menor consumo de água</b> , de energia e de <b>matérias-primas</b> . Na <b>medida do possível</b> , <b>devem ser privilegiados os materiais que causem menor impacto ambiental</b> , desde as <b>fases de exploração dos recursos naturais até a sua utilização final</b> .	VP	8
		FN	8
		FP	0
		VN	4
50	Recomenda-se a <b>utilização de madeiras</b> cuja <b>origem</b> possa ser <b>comprovada</b> mediante <b>apresentação de certificação legal</b> ou <b>provenientes de plano de manejo</b> aprovado pelos <b>órgãos ambientais</b> .	VP	5
		FN	3
		FP	1
		VN	2
51	Durante a <b>construção</b> , <b>deve-se implementar</b> um <b>sistema de gestão de resíduos</b> no <b>canteiro de obras</b> , de forma a <b>minimizar</b> sua <b>geração</b> e <b>possibilitar</b> a <b>segregação</b> de maneira adequada para <b>facilitar o reuso</b> , a <b>reciclagem</b> ou a <b>disposição final</b> em <b>locais específicos</b> .	VP	6
		FN	5
		FP	1
		VN	3
52	No caso de <b>reuso de água</b> para <b>destinação não potável</b> , esta <b>deve atender</b> aos <b>parâmetros</b> estabelecidos na <b>Tabela 8</b> .	VP	2
		FN	3
		FP	0
		VN	4
53	As <b>instalações elétricas</b> <b>devem privilegiar</b> a adoção de <b>soluções</b> , caso a caso, que <b>minimizem o consumo de energia</b> , entre elas a utilização de <b>iluminação e ventilação natural</b> e de <b>sistemas de aquecimento baseados em energia alternativa</b> .	VP	5
		FN	2
		FP	1
		VN	4
54	Convém a <b>adoção de soluções</b> que <b>minimizem o consumo de energia</b> , entre elas a <b>utilização de iluminação e ventilação natural</b> e de <b>sistemas de aquecimento baseados em energia alternativa</b> .	VP	6
		FN	1
		FP	0
		VN	2

## APÊNDICE L - Comparação sobre alteração na ordem dos padrões

Versão 1 - com ordem de padrões alterada (REF, COMP, PAD, EXIG e VAL)				Versão 2 - com ordem de padrões alterada (REF, EXIG, COMP, PAD e VAL)			
Frase	Obj. Semânt.	Texto	Rótulo	Frase	Obj. Semânt.	Texto	Rótulo
1	1	<i>base nas características do local</i>	COMPONENTE	1	1	<i>devem ser desenvolvidos</i>	EXIGENCIA
1	2	<i>ocorrência de subsidência do solo</i>	COMPONENTE	1	2	<i>base nas características do local</i>	COMPONENTE
1	3	<i>presença de crateras em camadas</i>	COMPONENTE	1	3	<i>ocorrência de subsidência do solo</i>	COMPONENTE
1	4	<i>presença de solos expansíveis</i>	COMPONENTE	1	4	<i>presença de crateras em camadas</i>	COMPONENTE
1	5	<i>presença de camadas profundas</i>	COMPONENTE	1	5	<i>presença de solos expansíveis</i>	COMPONENTE
1	6	<i>local de implantação</i>	COMPONENTE	1	6	<i>presença de camadas profundas</i>	COMPONENTE
1	7	<i>projetos de arquitetura</i>	COMPONENTE	1	7	<i>local de implantação</i>	COMPONENTE
1	8	<i>riscos de deslizamentos</i>	COMPONENTE	1	8	<i>projetos de arquitetura</i>	COMPONENTE
1	9	<i>trabalhos de terraplenagem</i>	COMPONENTE	1	9	<i>riscos de deslizamentos</i>	COMPONENTE
1	10	<i>compactação do solo</i>	COMPONENTE	1	10	<i>trabalhos de terraplenagem</i>	COMPONENTE
1	11	<i>conjuntos habitacionais</i>	COMPONENTE	1	11	<i>compactação do solo</i>	COMPONENTE
1	12	<i>obras geotécnicas</i>	COMPONENTE	1	12	<i>conjuntos habitacionais</i>	COMPONENTE
1	13	<i>vias férreas</i>	COMPONENTE	1	13	<i>obras geotécnicas</i>	COMPONENTE
1	14	<i>devem ser desenvolvidos</i>	EXIGENCIA	1	14	<i>vias férreas</i>	COMPONENTE
2	1	<i>confinamento de gases resultantes</i>	COMPONENTE	2	1	<i>Devem ainda ser considerados</i>	EXIGENCIA
2	2	<i>riscos de explosões</i>	COMPONENTE	2	2	<i>confinamento de gases resultantes</i>	COMPONENTE
2	3	<i>proximidade de pedreiras</i>	COMPONENTE	2	3	<i>riscos de explosões</i>	COMPONENTE
2	4	<i>prejuízos à segurança</i>	COMPONENTE	2	4	<i>proximidade de pedreiras</i>	COMPONENTE
2	5	<i>funcionalidade da obra</i>	COMPONENTE	2	5	<i>prejuízos à segurança</i>	COMPONENTE
2	6	<i>aterros sanitários</i>	COMPONENTE	2	6	<i>funcionalidade da obra</i>	COMPONENTE
2	7	<i>providências necessárias</i>	COMPONENTE	2	7	<i>aterros sanitários</i>	COMPONENTE
2	8	<i>Devem ainda ser considerados</i>	EXIGENCIA	2	8	<i>providências necessárias</i>	COMPONENTE
3	1	<i>sobreposições de bulbos de pressão</i>	COMPONENTE	3	1	<i>sobreposições de bulbos de pressão</i>	COMPONENTE
3	2	<i>efeitos de grupo de estacas</i>	COMPONENTE	3	2	<i>efeitos de grupo de estacas</i>	COMPONENTE
3	3	<i>desconfinamento do solo em função</i>	COMPONENTE	3	3	<i>desconfinamento do solo em função</i>	COMPONENTE
3	4	<i>interações entre construções próximas</i>	COMPONENTE	3	4	<i>interações entre construções próximas</i>	COMPONENTE
3	5	<i>rebaixamento do lençol freático</i>	COMPONENTE	3	5	<i>rebaixamento do lençol freático</i>	COMPONENTE
3	6	<i>corte do terreno</i>	COMPONENTE	3	6	<i>corte do terreno</i>	COMPONENTE
3	7	<i>projetos</i>	COMPONENTE	3	7	<i>projetos</i>	COMPONENTE
4	1	<i>funcionalidade da obra</i>	COMPONENTE	4	1	<i>não podem prejudicar</i>	EXIGENCIA
4	2	<i>edificações vizinhas</i>	COMPONENTE	4	2	<i>funcionalidade da obra</i>	COMPONENTE
4	3	<i>fenômenos</i>	COMPONENTE	4	3	<i>edificações vizinhas</i>	COMPONENTE
4	4	<i>não podem prejudicar</i>	EXIGENCIA	4	4	<i>fenômenos</i>	COMPONENTE
5	1	<i>ABNT NBR 8044</i>	REFERENCIA	5	1	<i>ABNT NBR 8044</i>	REFERENCIA
5	2	<i>ABNT NBR 5629</i>	REFERENCIA	5	2	<i>ABNT NBR 5629</i>	REFERENCIA
5	3	<i>ABNT NBR 11682</i>	REFERENCIA	5	3	<i>ABNT NBR 11682</i>	REFERENCIA

5	4	ABNT NBR 6122	REFERENCIA	5	4	ABNT NBR 6122	REFERENCIA
5	5	ABNT NBR 12722	REFERENCIA	5	5	ABNT NBR 12722	REFERENCIA
5	6	Normas aplicáveis	REFERENCIA	5	6	Normas aplicáveis	REFERENCIA
5	7	desempenho da edificação	COMPONENTE	5	7	devendo ser atendidas	EXIGENCIA
5	8	projetos de implantação	COMPONENTE	5	8	desempenho da edificação	COMPONENTE
5	9	desempenho das fundações	COMPONENTE	5	9	projetos de implantação	COMPONENTE
5	10	devendo ser atendidas	EXIGENCIA	5	10	desempenho das fundações	COMPONENTE
6	1	ponto de vista da segurança	COMPONENTE	6	1	devem ser consideradas	EXIGENCIA
6	2	estabilidade ao longo da vida	COMPONENTE	6	2	ponto de vista da segurança	COMPONENTE
6	3	condições de agressividade do solo	COMPONENTE	6	3	estabilidade ao longo da vida	COMPONENTE
6	4	água na época do projeto	COMPONENTE	6	4	condições de agressividade do solo	COMPONENTE
6	5	proteções pertinentes	COMPONENTE	6	5	água na época do projeto	COMPONENTE
6	6	devem ser consideradas	EXIGENCIA	6	6	proteções pertinentes	COMPONENTE
7	NA	Capítulo 8 - omitido	NA	7	NA	Capítulo 8 - omitido	NA
8	1	transeuntes nas imediações do imóvel	COMPONENTE	8	1	não podem apresentar	EXIGENCIA
8	2	integridade física	COMPONENTE	8	2	possam colocar em	EXIGENCIA
8	3	sistemas	COMPONENTE	8	3	transeuntes nas imediações do imóvel	COMPONENTE
8	4	defeitos acima dos limites	PARAMETRO	8	4	integridade física	COMPONENTE
8	5	não podem apresentar	EXIGENCIA	8	5	sistemas	COMPONENTE
8	6	possam colocar em	EXIGENCIA	8	6	defeitos acima dos limites	PARAMETRO
9	1	ABNT NBR 14718	REFERENCIA	9	1	ABNT NBR 14718	REFERENCIA
9	2	ABNT NBR 15575-3	REFERENCIA	9	2	ABNT NBR 15575-3	REFERENCIA
9	3	queda de pessoas em altura	COMPONENTE	9	3	Devem ser previstas	EXIGENCIA
9	4	locais com riscos de quedas	COMPONENTE	9	4	devem ser ensaiadas	EXIGENCIA
9	5	queda de pessoas em função	COMPONENTE	9	5	devem possuir	EXIGENCIA
9	6	queda de pessoas em função	COMPONENTE	9	6	queda de pessoas em altura	COMPONENTE
9	7	contusões em função da operação	COMPONENTE	9	7	locais com riscos de quedas	COMPONENTE
9	8	contusões em função da	COMPONENTE	9	8	queda de pessoas em função	COMPONENTE
9	9	dessolidarização	COMPONENTE	9	9	queda de pessoas em função	COMPONENTE
9	9	contusões em função de explosão	COMPONENTE	9	9	queda de pessoas em função	COMPONENTE
9	10	confinamento de gás combustível	COMPONENTE	9	10	contusões em função da operação	COMPONENTE
9	11	uso da edificação	COMPONENTE	9	11	contusões em função da	COMPONENTE
9	12	lajes de cobertura	COMPONENTE	9	12	dessolidarização	COMPONENTE
9	13	rupturas das proteções	COMPONENTE	9	13	contusões em função de explosão	COMPONENTE
9	14	memorial de cálculo	COMPONENTE	9	14	confinamento de gás combustível	COMPONENTE
9	15	irregularidades nos pisos	COMPONENTE	9	15	uso da edificação	COMPONENTE
9	16	ruptura de subsistemas	COMPONENTE	9	16	lajes de cobertura	COMPONENTE
9	17	projeção de materiais	COMPONENTE	9	17	rupturas das proteções	COMPONENTE
9	18	profissional responsável	COMPONENTE	9	18	memorial de cálculo	COMPONENTE
9	19	partes cortantes	COMPONENTE	9	19	irregularidades nos pisos	COMPONENTE
9	20	partes móveis	COMPONENTE	9	20	ruptura de subsistemas	COMPONENTE
9	21	Devem ser previstas	EXIGENCIA	9	21	projeção de materiais	COMPONENTE
				9	21	profissional responsável	COMPONENTE

9	22	<i>devem ser ensaiadas</i>	EXIGENCIA	9	22	<i>partes cortantes</i>	COMPONENTE
9	23	<i>devem possuir</i>	EXIGENCIA	9	23	<i>partes móveis</i>	COMPONENTE
10	1	<i>ocorrência de ferimentos</i>	COMPONENTE	10	1	<i>ocorrência de ferimentos</i>	COMPONENTE
10	2	<i>danos aos usuários</i>	COMPONENTE	10	2	<i>danos aos usuários</i>	COMPONENTE
10	3	<i>condições normais</i>	COMPONENTE	10	3	<i>condições normais</i>	COMPONENTE
11	1	<i>Normas específicas</i>	REFERENCIA	11	1	<i>Normas específicas</i>	REFERENCIA
11	2	<i>edificação habitacional</i>	COMPONENTE	11	2	<i>deve atender aos</i>	EXIGENCIA
11	3	<i>deve atender aos</i>	EXIGENCIA	11	3	<i>edificação habitacional</i>	COMPONENTE
12	1	<i>ABNT NBR 15575-3</i>	REFERENCIA	12	1	<i>ABNT NBR 15575-3</i>	REFERENCIA
12	2	<i>ABNT NBR 15575-5</i>	REFERENCIA	12	2	<i>ABNT NBR 15575-5</i>	REFERENCIA
12	3	<i>Atendimento aos requisitos</i>	COMPONENTE	12	3	<i>Atendimento aos requisitos</i>	COMPONENTE
13	1	<i>ABNT NBR 9575</i>	REFERENCIA	13	1	<i>ABNT NBR 9575</i>	REFERENCIA
13	2	<i>prevenção de infiltração da água</i>	COMPONENTE	13	2	<i>Devem ser previstos</i>	EXIGENCIA
13	3	<i>umidade do solo nas habitações</i>	COMPONENTE	13	3	<i>devem seguir</i>	EXIGENCIA
13	4	<i>condições de implantação dos conjuntos</i>	COMPONENTE	13	4	<i>prevenção de infiltração da água</i>	COMPONENTE
13	5	<i>prejuízo da utilização do ambiente</i>	COMPONENTE	13	5	<i>umidade do solo nas habitações</i>	COMPONENTE
13	6	<i>água de chuva incidente</i>	COMPONENTE	13	6	<i>condições de implantação dos conjuntos</i>	COMPONENTE
13	7	<i>meio dos detalhes</i>	COMPONENTE	13	7	<i>prejuízo da utilização do ambiente</i>	COMPONENTE
13	8	<i>penetração de líquidos</i>	COMPONENTE	13	8	<i>água de chuva incidente</i>	COMPONENTE
13	9	<i>umidades de porões</i>	COMPONENTE	13	9	<i>meio dos detalhes</i>	COMPONENTE
13	10	<i>paredes em contato</i>	COMPONENTE	13	10	<i>penetração de líquidos</i>	COMPONENTE
13	11	<i>direcionamento das águas</i>	COMPONENTE	13	11	<i>umidades de porões</i>	COMPONENTE
13	12	<i>sistemas de impermeabilização</i>	COMPONENTE	13	12	<i>paredes em contato</i>	COMPONENTE
13	13	<i>penetração de líquidos</i>	COMPONENTE	13	13	<i>direcionamento das águas</i>	COMPONENTE
13	14	<i>umidades em fundações</i>	COMPONENTE	13	14	<i>sistemas de impermeabilização</i>	COMPONENTE
13	15	<i>pisos em contato</i>	COMPONENTE	13	15	<i>penetração de líquidos</i>	COMPONENTE
13	16	<i>elementos da construção</i>	COMPONENTE	13	16	<i>umidades em fundações</i>	COMPONENTE
13	17	<i>ruas internas</i>	COMPONENTE	13	17	<i>pisos em contato</i>	COMPONENTE
13	18	<i>lotes vizinhos</i>	COMPONENTE	13	18	<i>elementos da construção</i>	COMPONENTE
13	19	<i>jardins contíguos</i>	COMPONENTE	13	19	<i>ruas internas</i>	COMPONENTE
13	20	<i>sistemas correlatos</i>	COMPONENTE	13	20	<i>lotes vizinhos</i>	COMPONENTE
13	21	<i>segurança estrutural</i>	COMPONENTE	13	21	<i>jardins contíguos</i>	COMPONENTE
13	22	<i>corpo principal</i>	COMPONENTE	13	22	<i>sistemas correlatos</i>	COMPONENTE
13	23	<i>calçadas laterais</i>	COMPONENTE	13	23	<i>segurança estrutural</i>	COMPONENTE
13	24	<i>entorno próximo ao conjunto</i>	PARAMETRO	13	24	<i>corpo principal</i>	COMPONENTE
13	25	<i>Devem ser previstos</i>	EXIGENCIA	13	25	<i>calçadas laterais</i>	COMPONENTE
13	26	<i>devem seguir</i>	EXIGENCIA	13	26	<i>entorno próximo ao conjunto</i>	PARAMETRO
14	1	<i>estanqueidade de partes do edifício</i>	COMPONENTE	14	1	<i>Devem ser previstos</i>	EXIGENCIA
14	2	<i>adequação das vinculações entre instalações</i>	COMPONENTE	14	2	<i>devendo ser verificada</i>	EXIGENCIA
14	3	<i>manutenção do imóvel</i>	COMPONENTE	14	3	<i>estanqueidade de partes do edifício</i>	COMPONENTE
14	4	<i>águas pluviais</i>	COMPONENTE	14	4	<i>adequação das vinculações entre instalações</i>	COMPONENTE

14	5	<i>Devem ser previstos</i>	EXIGENCIA	14	5	<i>manutenção do imóvel</i>	COMPONENTE
14	6	<i>devido ser verificada</i>	EXIGENCIA	14	6	<i>águas pluviais</i>	COMPONENTE
15	1	<i>h ao longo do ano</i>	COMPONENTE	15	1	<i>deve ser capaz</i>	EXIGENCIA
15	2	<i>programa de simulação computacional</i>	COMPONENTE	15	2	<i>h ao longo do ano</i>	COMPONENTE
15	3	<i>variações da temperatura operativa</i>	COMPONENTE	15	3	<i>programa de simulação computacional</i>	COMPONENTE
15	4	<i>uso da ventilação natural</i>	COMPONENTE	15	4	<i>variações da temperatura operativa</i>	COMPONENTE
15	5	<i>programa de simulação computacional</i>	COMPONENTE	15	5	<i>uso da ventilação natural</i>	COMPONENTE
15	6	<i>efeitos de inércia térmica</i>	COMPONENTE	15	6	<i>programa de simulação computacional</i>	COMPONENTE
15	7	<i>potência de iluminação</i>	COMPONENTE	15	7	<i>efeitos de inércia térmica</i>	COMPONENTE
15	8	<i>procedimento de teste</i>	COMPONENTE	15	8	<i>potência de iluminação</i>	COMPONENTE
15	9	<i>trocas de calor</i>	COMPONENTE	15	9	<i>procedimento de teste</i>	COMPONENTE
15	10	<i>efeitos da ventilação</i>	COMPONENTE	15	10	<i>trocas de calor</i>	COMPONENTE
15	11	<i>cargas térmicas</i>	COMPONENTE	15	11	<i>efeitos da ventilação</i>	COMPONENTE
15	12	<i>variações horárias</i>	COMPONENTE	15	12	<i>cargas térmicas</i>	COMPONENTE
15	13	<i>cargas térmicas</i>	COMPONENTE	15	13	<i>variações horárias</i>	COMPONENTE
15	14	<i>sombreamento proveniente</i>	COMPONENTE	15	14	<i>cargas térmicas</i>	COMPONENTE
15	15	<i>elementos externos</i>	COMPONENTE	15	15	<i>sombreamento proveniente</i>	COMPONENTE
15	16	<i>zonas térmicas</i>	COMPONENTE	15	16	<i>elementos externos</i>	COMPONENTE
15	17	<i>deve ser capaz</i>	EXIGENCIA	15	17	<i>zonas térmicas</i>	COMPONENTE
16	1	<i>clima da cidade</i>	COMPONENTE	16	1	<i>deve possuir</i>	EXIGENCIA
16	2	<i>arquivo climático</i>	COMPONENTE	16	2	<i>clima da cidade</i>	COMPONENTE
16	3	<i>UH</i>	COMPONENTE	16	3	<i>arquivo climático</i>	COMPONENTE
16	4	<i>deve possuir</i>	EXIGENCIA	16	4	<i>UH</i>	COMPONENTE
17	1	<i>comparação da distância euclidiana</i>	COMPONENTE	17	1	<i>deve ser utilizado</i>	EXIGENCIA
17	2	<i>cidade de implantação</i>	COMPONENTE	17	2	<i>deve considerar</i>	EXIGENCIA
17	3	<i>arquivo climático</i>	COMPONENTE	17	3	<i>comparação da distância euclidiana</i>	COMPONENTE
17	4	<i>arquivo climático</i>	COMPONENTE	17	4	<i>cidade de implantação</i>	COMPONENTE
17	5	<i>cidade próxima</i>	COMPONENTE	17	5	<i>arquivo climático</i>	COMPONENTE
17	6	<i>clima semelhante</i>	COMPONENTE	17	6	<i>arquivo climático</i>	COMPONENTE
17	7	<i>semelhança</i>	COMPONENTE	17	7	<i>cidade próxima</i>	COMPONENTE
17	8	<i>deve ser utilizado</i>	EXIGENCIA	17	8	<i>clima semelhante</i>	COMPONENTE
17	9	<i>deve considerar</i>	EXIGENCIA	17	9	<i>semelhança</i>	COMPONENTE
18	1	<i>temperatura do ponto de orvalho</i>	COMPONENTE	18	1	<i>utilizado deve</i>	EXIGENCIA
18	2	<i>h do ano meteorológico</i>	COMPONENTE	18	2	<i>temperatura do ponto de orvalho</i>	COMPONENTE
18	3	<i>temperatura de bulbo seco</i>	COMPONENTE	18	3	<i>h do ano meteorológico</i>	COMPONENTE
18	4	<i>intensidade de radiação horizontal</i>	COMPONENTE	18	4	<i>temperatura de bulbo seco</i>	COMPONENTE
18	5	<i>watts-hora por metro quadrado</i>	COMPONENTE	18	5	<i>intensidade de radiação horizontal</i>	COMPONENTE
18	6	<i>watts-hora por metro quadrado</i>	COMPONENTE	18	6	<i>watts-hora por metro quadrado</i>	COMPONENTE
18	7	<i>watts-hora por metro quadrado</i>	COMPONENTE	18	7	<i>watts-hora por metro quadrado</i>	COMPONENTE
18	8	<i>watts-hora por metro quadrado</i>	COMPONENTE	18	8	<i>watts-hora por metro quadrado</i>	COMPONENTE
18	9	<i>meses do ano</i>	COMPONENTE	18	9	<i>watts-hora por metro quadrado</i>	COMPONENTE

18	10	<i>direção do vento</i>	COMPONENTE	18	10	<i>meses do ano</i>	COMPONENTE
18	11	<i>velocidade do vento</i>	COMPONENTE	18	11	<i>direção do vento</i>	COMPONENTE
18	12	<i>metros por segundo</i>	COMPONENTE	18	12	<i>velocidade do vento</i>	COMPONENTE
18	13	<i>arquivo climático</i>	COMPONENTE	18	13	<i>metros por segundo</i>	COMPONENTE
18	14	<i>valores mensais</i>	COMPONENTE	18	14	<i>arquivo climático</i>	COMPONENTE
18	15	<i>temperatura média</i>	COMPONENTE	18	15	<i>valores mensais</i>	COMPONENTE
18	16	<i>valores horários</i>	COMPONENTE	18	16	<i>temperatura média</i>	COMPONENTE
18	17	<i>umidade relativa</i>	COMPONENTE	18	17	<i>valores horários</i>	COMPONENTE
18	18	<i>pressão atmosférica</i>	COMPONENTE	18	18	<i>umidade relativa</i>	COMPONENTE
18	19	<i>onda longa</i>	COMPONENTE	18	19	<i>pressão atmosférica</i>	COMPONENTE
18	20	<i>radiação horizontal</i>	COMPONENTE	18	20	<i>onda longa</i>	COMPONENTE
18	21	<i>radiação normal</i>	COMPONENTE	18	21	<i>radiação horizontal</i>	COMPONENTE
18	22	<i>radiação horizontal</i>	COMPONENTE	18	22	<i>radiação normal</i>	COMPONENTE
18	23	<i>utilizado deve</i>	EXIGENCIA	18	23	<i>radiação horizontal</i>	COMPONENTE
19	1	<i>ABNT NBR 15575-3</i>	REFERENCIA	19	1	<i>ABNT NBR 15575-3</i>	REFERENCIA
19	2	<i>menos aos critérios de desempenho</i>	COMPONENTE	19	2	<i>devem ser projetados menos aos critérios de desempenho</i>	EXIGENCIA
19	3	<i>sistemas de pisos</i>	COMPONENTE	19	3	<i>sistemas de pisos</i>	COMPONENTE
19	4	<i>vedações verticais</i>	COMPONENTE	19	4	<i>sistemas de pisos</i>	COMPONENTE
19	5	<i>paredes internas</i>	COMPONENTE	19	5	<i>vedações verticais</i>	COMPONENTE
19	6	<i>função estrutural</i>	COMPONENTE	19	6	<i>paredes internas</i>	COMPONENTE
19	7	<i>unidades habitacionais</i>	COMPONENTE	19	7	<i>função estrutural</i>	COMPONENTE
19	8	<i>unidades habitacionais</i>	COMPONENTE	19	8	<i>unidades habitacionais</i>	COMPONENTE
19	9	<i>áreas comuns</i>	COMPONENTE	19	9	<i>unidades habitacionais</i>	COMPONENTE
19	10	<i>devem ser projetados</i>	EXIGENCIA	19	10	<i>áreas comuns</i>	COMPONENTE
20	1	<i>ABNT NBR 15575-3</i>	REFERENCIA	20	1	<i>ABNT NBR 15575-3</i>	REFERENCIA
20	2	<i>ABNT NBR 15575-5</i>	REFERENCIA	20	2	<i>ABNT NBR 15575-5</i>	REFERENCIA
20	3	<i>sistemas de pisos com função</i>	COMPONENTE	20	3	<i>devem atender pelo</i>	EXIGENCIA
20	4	<i>menos aos critérios mínimos</i>	COMPONENTE	20	4	<i>sistemas de pisos com função</i>	COMPONENTE
20	5	<i>sistemas de pisos</i>	COMPONENTE	20	5	<i>menos aos critérios mínimos</i>	COMPONENTE
20	6	<i>edifícios habitacionais</i>	COMPONENTE	20	6	<i>sistemas de pisos</i>	COMPONENTE
20	7	<i>critério</i>	COMPONENTE	20	7	<i>edifícios habitacionais</i>	COMPONENTE
20	8	<i>devem atender pelo dependências das construções habitacionais</i>	EXIGENCIA	20	8	<i>critério</i>	COMPONENTE
21	1	<i>iluminação natural</i>	COMPONENTE	21	1	<i>devem atender ao dependências das construções habitacionais</i>	EXIGENCIA
21	2	<i>níveis gerais</i>	COMPONENTE	21	2	<i>iluminação natural</i>	COMPONENTE
21	3	<i>devem atender ao</i>	EXIGENCIA	21	3	<i>níveis gerais</i>	COMPONENTE
22	1	<i>fator de luz diurna dependências das construções habitacionais</i>	COMPONENTE	22	1	<i>deve atender ao</i>	EXIGENCIA
22	2	<i>iluminação natural</i>	COMPONENTE	22	2	<i>fator de luz diurna dependências das construções habitacionais</i>	COMPONENTE
22	3	<i>deve atender ao</i>	EXIGENCIA	22	3	<i>iluminação natural</i>	COMPONENTE
22	4	<i>iluminação natural</i>	COMPONENTE	22	4	<i>deve atender ao</i>	EXIGENCIA

23	1	<i>inserção de poços de ventilação</i>	COMPONENTE	23	1	<i>podem ser atendidos</i>	EXIGENCIA
23	2	<i>introdução de domo de iluminação</i>	COMPONENTE	23	2	<i>inserção de poços de ventilação</i>	COMPONENTE
23	3	<i>requisitos de iluminância natural</i>	COMPONENTE	23	3	<i>introdução de domo de iluminação</i>	COMPONENTE
23	4	<i>disposição dos cômodos</i>	COMPONENTE	23	4	<i>requisitos de iluminância natural</i>	COMPONENTE
23	5	<i>posição das aberturas</i>	COMPONENTE	23	5	<i>disposição dos cômodos</i>	COMPONENTE
23	6	<i>tipos de janelas</i>	COMPONENTE	23	6	<i>posição das aberturas</i>	COMPONENTE
23	7	<i>cores dos elementos</i>	COMPONENTE	23	7	<i>tipos de janelas</i>	COMPONENTE
23	8	<i>orientação geográfica</i>	COMPONENTE	23	8	<i>cores dos elementos</i>	COMPONENTE
23	9	<i>podem ser atendidos</i>	EXIGENCIA	23	9	<i>orientação geográfica</i>	COMPONENTE
24	1	<i>presença de taludes</i>	COMPONENTE	24	1	<i>não podem prejudicar</i>	EXIGENCIA
24	2	<i>coberturas de garagens</i>	COMPONENTE	24	2	<i>presença de taludes</i>	COMPONENTE
24	3	<i>obstáculos do gênero</i>	COMPONENTE	24	3	<i>coberturas de garagens</i>	COMPONENTE
24	4	<i>níveis mínimos</i>	COMPONENTE	24	4	<i>obstáculos do gênero</i>	COMPONENTE
24	5	<i>não podem prejudicar</i>	EXIGENCIA	24	5	<i>níveis mínimos</i>	COMPONENTE
25	1	<i>caixas de escada</i>	COMPONENTE	25	1	<i>não podem prejudicar</i>	EXIGENCIA
25	2	<i>conjuntos habitacionais</i>	COMPONENTE	25	2	<i>caixas de escada</i>	COMPONENTE
25	3	<i>implantação relativa</i>	COMPONENTE	25	3	<i>conjuntos habitacionais</i>	COMPONENTE
25	4	<i>níveis mínimos</i>	COMPONENTE	25	4	<i>implantação relativa</i>	COMPONENTE
25	5	<i>não podem prejudicar dependências dos edifícios habitacionais</i>	EXIGENCIA	25	5	<i>níveis mínimos</i>	COMPONENTE
26	1	<i>níveis gerais</i>	COMPONENTE	26	1	<i>devem atender ao dependências dos edifícios habitacionais</i>	EXIGENCIA
26	2	<i>iluminação artificial</i>	COMPONENTE	26	2	<i>níveis gerais</i>	COMPONENTE
26	3	<i>devem atender ao</i>	EXIGENCIA	26	3	<i>iluminação artificial</i>	COMPONENTE
26	4	<i>devem atender ao</i>	EXIGENCIA	26	4	<i>iluminação artificial</i>	COMPONENTE
27	1	<b>ABNT NBR 15575</b>	REFERENCIA	27	1	<b>ABNT NBR 15575</b>	REFERENCIA
27	2	<i>exigibilidade desta parte</i>	COMPONENTE	27	2	<i>deve ser elaborado</i>	EXIGENCIA
27	3	<i>valor teórico</i>	COMPONENTE	27	3	<i>deve especificar</i>	EXIGENCIA
27	4	<i>vida útil</i>	COMPONENTE	27	4	<i>exigibilidade desta parte</i>	COMPONENTE
27	5	<i>durabilidade potencial</i>	COMPONENTE	27	5	<i>valor teórico</i>	COMPONENTE
27	6	<i>vida útil</i>	COMPONENTE	27	6	<i>vida útil</i>	COMPONENTE
27	7	<i>projeto</i>	COMPONENTE	27	7	<i>durabilidade potencial</i>	COMPONENTE
27	8	<i>sistemas</i>	COMPONENTE	27	8	<i>vida útil</i>	COMPONENTE
27	9	<i>deve ser elaborado</i>	EXIGENCIA	27	9	<i>projeto</i>	COMPONENTE
27	10	<i>deve especificar</i>	EXIGENCIA	27	10	<i>sistemas</i>	COMPONENTE
28	1	<i>durabilidade compatível</i>	COMPONENTE	28	1	<i>devem apresentar</i>	EXIGENCIA
28	2	<i>vida útil</i>	COMPONENTE	28	2	<i>durabilidade compatível</i>	COMPONENTE
28	3	<i>edifício</i>	COMPONENTE	28	3	<i>vida útil</i>	COMPONENTE
28	4	<i>sistemas</i>	COMPONENTE	28	4	<i>edifício</i>	COMPONENTE
28	5	<i>devem apresentar</i>	EXIGENCIA	28	5	<i>sistemas</i>	COMPONENTE
29	1	<i>condições de exposição do edifício</i>	COMPONENTE	29	1	<i>devem ser especificadas</i>	EXIGENCIA
29	2	<i>análise da vida útil</i>	COMPONENTE	29	2	<i>condições de exposição do edifício</i>	COMPONENTE
29	3	<i>durabilidade do edifício</i>	COMPONENTE	29	3	<i>análise da vida útil</i>	COMPONENTE

29	4	<i>devem ser especificadas</i>	EXIGENCIA	29	4	<i>durabilidade do edifício</i>	COMPONENTE
30	1	<i>projeto para definição da vida</i>	COMPONENTE	30	1	<i>projeto para definição da vida</i>	COMPONENTE
30	2	<i>operação do edifício</i>	COMPONENTE	30	2	<i>operação do edifício</i>	COMPONENTE
30	3	<i>especificações relativas</i>	COMPONENTE	30	3	<i>especificações relativas</i>	COMPONENTE
31	1	<i>capacidade do edifício</i>	COMPONENTE	31	1	<i>capacidade do edifício</i>	COMPONENTE
31	2	<i>intervenções de manutenção</i>	COMPONENTE	31	2	<i>intervenções de manutenção</i>	COMPONENTE
31	3	<i>inspeções prediais</i>	COMPONENTE	31	3	<i>inspeções prediais</i>	COMPONENTE
32	1	<i>legislação vigente</i>	REFERENCIA	32	1	<i>legislação vigente</i>	REFERENCIA
32	2	<i>condições de salubridade no interior</i>	COMPONENTE	32	2	<i>deve atender aos condições de salubridade no interior</i>	EXIGENCIA
32	3	<i>temperatura no interior da unidade</i>	COMPONENTE	32	3	<i>temperatura no interior da unidade</i>	COMPONENTE
32	4	<i>condições de umidade</i>	COMPONENTE	32	4	<i>temperatura no interior da unidade</i>	COMPONENTE
32	5	<i>tipo dos sistemas</i>	COMPONENTE	32	5	<i>condições de umidade</i>	COMPONENTE
32	6	<i>requisito</i>	COMPONENTE	32	6	<i>tipo dos sistemas</i>	COMPONENTE
32	7	<i>deve atender aos</i>	EXIGENCIA	32	7	<i>requisito</i>	COMPONENTE
33	1	<i>legislação vigente</i>	REFERENCIA	33	1	<i>legislação vigente</i>	REFERENCIA
33	2	<i>ar em ambientes confinados</i>	COMPONENTE	33	2	<i>não podem liberar</i>	EXIGENCIA
33	3	<i>níveis de poluição</i>	COMPONENTE	33	3	<i>deve atender aos</i>	EXIGENCIA
33	4	<i>sistemas empregados</i>	COMPONENTE	33	4	<i>ar em ambientes confinados</i>	COMPONENTE
33	5	<i>gás carbônico</i>	COMPONENTE	33	5	<i>níveis de poluição</i>	COMPONENTE
33	6	<i>materiais</i>	COMPONENTE	33	6	<i>sistemas empregados</i>	COMPONENTE
33	7	<i>aerodispersóides</i>	COMPONENTE	33	7	<i>gás carbônico</i>	COMPONENTE
33	8	<i>requisito</i>	COMPONENTE	33	8	<i>materiais</i>	COMPONENTE
33	9	<i>não podem liberar</i>	EXIGENCIA	33	9	<i>aerodispersóides</i>	COMPONENTE
33	10	<i>deve atender aos</i>	EXIGENCIA	33	10	<i>requisito</i>	COMPONENTE
34	1	<i>legislação vigente</i>	REFERENCIA	34	1	<i>legislação vigente</i>	REFERENCIA
34	2	<i>Gases de escapamento de veículos</i>	COMPONENTE	34	2	<i>não podem invadir</i>	EXIGENCIA
34	3	<i>ventilação de garagens internas</i>	COMPONENTE	34	3	<i>deve atender aos</i>	EXIGENCIA
34	4	<i>saída dos gases poluentes</i>	COMPONENTE	34	4	<i>deve permitir</i>	EXIGENCIA
34	5	<i>sistema de exaustão</i>	COMPONENTE	34	5	<i>Gases de escapamento de veículos</i>	COMPONENTE
34	6	<i>áreas internas</i>	COMPONENTE	34	6	<i>ventilação de garagens internas</i>	COMPONENTE
34	7	<i>requisito</i>	COMPONENTE	34	7	<i>saída dos gases poluentes</i>	COMPONENTE
34	8	<i>não podem invadir</i>	EXIGENCIA	34	8	<i>sistema de exaustão</i>	COMPONENTE
34	9	<i>deve atender aos</i>	EXIGENCIA	34	9	<i>áreas internas</i>	COMPONENTE
34	10	<i>deve permitir</i>	EXIGENCIA	34	10	<i>requisito</i>	COMPONENTE
35	1	<i>altura mínima</i>	COMPONENTE	35	1	<i>não pode ser inferior</i>	EXIGENCIA
35	2	<i>não pode ser inferior</i>	EXIGENCIA	35	2	<i>altura mínima</i>	COMPONENTE
35	3	<i>2,50 m.</i>	VALOR	35	3	<i>2,50 m.</i>	VALOR
36	1	<i>m. Nos tetos com vigas</i>	COMPONENTE	36	1	<i>devem ser mantidos</i>	EXIGENCIA
36	2	<i>piso a piso</i>	COMPONENTE	36	2	<i>possa descer até</i>	EXIGENCIA
36	3	<i>superfície do teto</i>	COMPONENTE	36	3	<i>m. Nos tetos com vigas</i>	COMPONENTE
36	4	<i>instalações sanitárias</i>	COMPONENTE	36	4	<i>piso a piso</i>	COMPONENTE

36	5	<i>superfícies salientes</i>	COMPONENTE	36	5	<i>superfície do teto</i>	COMPONENTE
36	6	<i>pé-direito mínimo</i>	COMPONENTE	36	6	<i>instalações sanitárias</i>	COMPONENTE
36	7	<i>superfície restante</i>	COMPONENTE	36	7	<i>superfícies salientes</i>	COMPONENTE
36	8	<i>pé-direito livre</i>	COMPONENTE	36	8	<i>pé-direito mínimo</i>	COMPONENTE
36	9	<i>devem ser mantidos</i>	EXIGENCIA	36	9	<i>superfície restante</i>	COMPONENTE
36	10	<i>possa descer até</i>	EXIGENCIA	36	10	<i>pé-direito livre</i>	COMPONENTE
36	11	<i>80 %</i>	VALOR	36	11	<i>80 %</i>	VALOR
36	12	<i>2,30 m.</i>	VALOR	36	12	<i>2,30 m.</i>	VALOR
37	1	<i>projetos de arquitetura de unidades</i>	COMPONENTE	37	1	<i>sugere-se prever no</i>	EXIGENCIA
37	2	<i>disponibilidade de espaço nos cômodos</i>	COMPONENTE	37	2	<i>projetos de arquitetura de unidades</i>	COMPONENTE
37	3	<i>utilização dos móveis</i>	COMPONENTE	37	3	<i>disponibilidade de espaço nos</i>	COMPONENTE
37	4	<i>edificação habitacional</i>	COMPONENTE	37	4	<i>cômodos</i>	COMPONENTE
37	5	<i>sugere-se prever no</i>	EXIGENCIA	37	5	<i>utilização dos móveis</i>	COMPONENTE
38	1	<i>ABNT NBR 9050</i>	REFERENCIA	37	5	<i>edificação habitacional</i>	COMPONENTE
38	2	<i>pessoas com deficiência física</i>	COMPONENTE	38	1	<i>ABNT NBR 9050</i>	REFERENCIA
38	3	<i>áreas de uso comum</i>	COMPONENTE	38	2	<i>devem atender ao</i>	EXIGENCIA
38	4	<i>áreas privativas</i>	COMPONENTE	38	3	<i>devem receber</i>	EXIGENCIA
38	5	<i>adaptações necessárias</i>	COMPONENTE	38	4	<i>pessoas com deficiência física</i>	COMPONENTE
38	6	<i>devem atender ao</i>	EXIGENCIA	38	5	<i>áreas de uso comum</i>	COMPONENTE
38	7	<i>devem receber</i>	EXIGENCIA	38	6	<i>áreas privativas</i>	COMPONENTE
39	1	<i>substituição de escadas por rampas</i>	COMPONENTE	38	7	<i>adaptações necessárias</i>	COMPONENTE
39	2	<i>alturas de peças sanitárias</i>	COMPONENTE	39	1	<i>deve prever para</i>	EXIGENCIA
39	3	<i>limitação de declividades</i>	COMPONENTE	39	2	<i>substituição de escadas por</i>	COMPONENTE
39	4	<i>largura de corredores</i>	COMPONENTE	39	3	<i>rampas</i>	COMPONENTE
39	5	<i>disponibilidade de alças</i>	COMPONENTE	39	4	<i>alturas de peças sanitárias</i>	COMPONENTE
39	6	<i>barras de apoio</i>	COMPONENTE	39	5	<i>limitação de declividades</i>	COMPONENTE
39	7	<i>áreas comuns</i>	COMPONENTE	39	6	<i>largura de corredores</i>	COMPONENTE
39	8	<i>áreas privativas</i>	COMPONENTE	39	7	<i>disponibilidade de alças</i>	COMPONENTE
39	9	<i>projeto</i>	COMPONENTE	39	8	<i>barras de apoio</i>	COMPONENTE
39	10	<i>deve prever para</i>	EXIGENCIA	39	9	<i>áreas comuns</i>	COMPONENTE
40	1	<i>execução das edificações térreas</i>	COMPONENTE	39	10	<i>áreas privativas</i>	COMPONENTE
40	2	<i>possibilidade de ampliação</i>	COMPONENTE	40	1	<i>projeto</i>	COMPONENTE
40	3	<i>continuidade de paredes</i>	COMPONENTE	40	1	<i>deve ser prevista</i>	EXIGENCIA
40	4	<i>caráter evolutivo</i>	COMPONENTE	40	2	<i>execução das edificações térreas</i>	COMPONENTE
40	5	<i>detalhes construtivos</i>	COMPONENTE	40	3	<i>possibilidade de ampliação</i>	COMPONENTE
40	6	<i>deve ser prevista</i>	EXIGENCIA	40	4	<i>continuidade de paredes</i>	COMPONENTE
41	1	<i>ampliação do corpo da edificação</i>	COMPONENTE	40	5	<i>caráter evolutivo</i>	COMPONENTE
41	2	<i>manual de uso</i>	COMPONENTE	40	6	<i>detalhes construtivos</i>	COMPONENTE
41	3	<i>detalhes construtivos</i>	COMPONENTE	41	1	<i>deve anexar ao</i>	EXIGENCIA
41	4	<i>instalações prediais</i>	COMPONENTE	41	2	<i>ampliação do corpo da edificação</i>	COMPONENTE
41	5	<i>coordenação dimensional</i>	COMPONENTE	41	3	<i>manual de uso</i>	COMPONENTE
				41	4	<i>detalhes construtivos</i>	COMPONENTE
				41	5	<i>instalações prediais</i>	COMPONENTE

41	6	<i>compatibilidades físicas</i>	COMPONENTE	41	6	<i>coordenação dimensional</i>	COMPONENTE
41	7	<i>materiais disponíveis</i>	COMPONENTE	41	7	<i>compatibilidades físicas</i>	COMPONENTE
41	8	<i>incorporador</i>	COMPONENTE	41	8	<i>materiais disponíveis</i>	COMPONENTE
41	9	<i>deve anexar ao</i>	EXIGENCIA	41	9	<i>incorporador</i>	COMPONENTE
42	1	<i>manutenção dos níveis de desempenho</i>	COMPONENTE	42	1	<i>devem permitir no</i>	EXIGENCIA
42	2	<i>segurança ao fogo</i>	COMPONENTE	42	2	<i>manutenção dos níveis de</i>	COMPONENTE
42	3	<i>estanqueidade à água</i>	COMPONENTE	42	3	<i>segurança ao fogo</i>	COMPONENTE
42	4	<i>detalhes construtivos</i>	COMPONENTE	42	4	<i>estanqueidade à água</i>	COMPONENTE
42	5	<i>comportamento estrutural</i>	COMPONENTE	42	5	<i>detalhes construtivos</i>	COMPONENTE
42	6	<i>desempenho térmico</i>	COMPONENTE	42	6	<i>comportamento estrutural</i>	COMPONENTE
42	7	<i>desempenho acústico</i>	COMPONENTE	42	7	<i>desempenho térmico</i>	COMPONENTE
42	8	<i>especificações</i>	COMPONENTE	42	8	<i>desempenho acústico</i>	COMPONENTE
42	9	<i>devem permitir no</i>	EXIGENCIA	42	9	<i>especificações</i>	COMPONENTE
43	1	<i>propostas de ampliação</i>	COMPONENTE	43	1	<i>devendo atender aos</i>	EXIGENCIA
43	2	<i>estudos de arquitetura</i>	COMPONENTE	43	2	<i>propostas de ampliação</i>	COMPONENTE
43	3	<i>níveis de funcionalidade</i>	COMPONENTE	43	3	<i>estudos de arquitetura</i>	COMPONENTE
43	4	<i>devendo atender aos</i>	EXIGENCIA	43	4	<i>níveis de funcionalidade</i>	COMPONENTE
44	1	<i>componentes da habitação</i>	COMPONENTE	44	1	<i>devem ser projetados</i>	EXIGENCIA
44	2	<i>ferimentos nos usuários</i>	COMPONENTE	44	2	<i>componentes da habitação</i>	COMPONENTE
44	3	<i>devem ser projetados</i>	EXIGENCIA	44	3	<i>ferimentos nos usuários</i>	COMPONENTE
45	1	<i>requisitos das respectivas</i>	COMPONENTE	45	1	<i>devem ser atendidas</i>	EXIGENCIA
45	2	<i>instalações hidrossanitárias</i>	COMPONENTE	45	2	<i>requisitos das respectivas</i>	COMPONENTE
45	3	<i>normalização específica</i>	COMPONENTE	45	3	<i>instalações hidrossanitárias</i>	COMPONENTE
45	4	<i>elementos</i>	COMPONENTE	45	4	<i>normalização específica</i>	COMPONENTE
45	5	<i>devem ser atendidas</i>	EXIGENCIA	45	5	<i>elementos</i>	COMPONENTE
46	1	<i>dispositivos de manobra</i>	COMPONENTE	46	1	<i>devem ser projetados</i>	EXIGENCIA
46	2	<i>força necessária</i>	COMPONENTE	46	2	<i>dispositivos de manobra</i>	COMPONENTE
46	3	<i>torque</i>	COMPONENTE	46	3	<i>força necessária</i>	COMPONENTE
46	4	<i>devem ser projetados</i>	EXIGENCIA	46	4	<i>torque</i>	COMPONENTE
46	5	<i>20 N.m</i>	VALOR	46	5	<i>20 N.m</i>	VALOR
47	1	<i>lançamentos de esgoto a céu aberto,</i>	COMPONENTE	47	1	<i>deve considerar</i>	EXIGENCIA
47	2	<i>riscos de desconfinamento do solo</i>	COMPONENTE	47	2	<i>lançamentos de esgoto a céu</i>	COMPONENTE
47	3	<i>implantação do empreendimento</i>	COMPONENTE	47	3	<i>aberto,</i>	COMPONENTE
47	4	<i>deslizamentos de taludes</i>	COMPONENTE	47	4	<i>riscos de desconfinamento do solo</i>	COMPONENTE
47	5	<i>assoreamento de vales</i>	COMPONENTE	47	5	<i>implantação do empreendimento</i>	COMPONENTE
47	6	<i>contaminação do solo</i>	COMPONENTE	47	6	<i>deslizamentos de taludes</i>	COMPONENTE
47	7	<i>água por efluentes</i>	COMPONENTE	47	7	<i>assoreamento de vales</i>	COMPONENTE
47	8	<i>riscos similares</i>	COMPONENTE	47	8	<i>contaminação do solo</i>	COMPONENTE
47	9	<i>deve considerar</i>	EXIGENCIA	47	9	<i>água por efluentes</i>	COMPONENTE
48	1	<i>ABNT NBR 8044</i>	REFERENCIA	48	1	<i>riscos similares</i>	COMPONENTE
48	2	<i>ABNT NBR 11682</i>	REFERENCIA	48	2	<i>deve considerar</i>	EXIGENCIA
				48	2	<i>ABNT NBR 11682</i>	REFERENCIA

48	3	<i>legislação vigente</i>	REFERENCIA	48	3	<i>legislação vigente</i>	REFERENCIA
48	4	<i>devem ser atendidos</i>	EXIGENCIA	48	4	<i>devem ser atendidos</i>	EXIGENCIA
49	1	<i>fases de exploração dos recursos</i>	COMPONENTE	49	1	<i>devem ser privilegiados</i>	EXIGENCIA
49	2	<i>consumo de água</i>	COMPONENTE	49	2	<i>fases de exploração dos recursos</i>	COMPONENTE
49	3	<i>medida do possível</i>	COMPONENTE	49	3	<i>consumo de água</i>	COMPONENTE
49	4	<i>recursos naturais</i>	COMPONENTE	49	4	<i>medida do possível</i>	COMPONENTE
49	5	<i>degradação ambiental</i>	COMPONENTE	49	5	<i>recursos naturais</i>	COMPONENTE
49	6	<i>impacto ambiental</i>	COMPONENTE	49	6	<i>degradação ambiental</i>	COMPONENTE
49	7	<i>utilização final</i>	COMPONENTE	49	7	<i>impacto ambiental</i>	COMPONENTE
49	8	<i>materiais</i>	COMPONENTE	49	8	<i>utilização final</i>	COMPONENTE
49	9	<i>devem ser privilegiados</i>	EXIGENCIA	49	9	<i>materiais</i>	COMPONENTE
50	1	<i>apresentação de certificação legal</i>	COMPONENTE	50	1	<i>possa ser comprovada</i>	EXIGENCIA
50	2	<i>utilização de madeiras</i>	COMPONENTE	50	2	<i>apresentação de certificação legal</i>	COMPONENTE
50	3	<i>plano de manejo</i>	COMPONENTE	50	3	<i>utilização de madeiras</i>	COMPONENTE
50	4	<i>órgãos ambientais</i>	COMPONENTE	50	4	<i>plano de manejo</i>	COMPONENTE
50	5	<i>possa ser comprovada</i>	EXIGENCIA	50	5	<i>órgãos ambientais</i>	COMPONENTE
51	1	<i>sistema de gestão de resíduos</i>	COMPONENTE	51	1	<i>deve-se implementar</i>	EXIGENCIA
51	2	<i>segregação de maneira adequada</i>	COMPONENTE	51	2	<i>sistema de gestão de resíduos</i>	COMPONENTE
51	3	<i>canteiro de obras</i>	COMPONENTE	51	3	<i>segregação de maneira adequada</i>	COMPONENTE
51	4	<i>disposição final</i>	COMPONENTE	51	4	<i>canteiro de obras</i>	COMPONENTE
51	5	<i>locais específicos</i>	COMPONENTE	51	5	<i>disposição final</i>	COMPONENTE
51	6	<i>deve-se implementar</i>	EXIGENCIA	51	6	<i>locais específicos</i>	COMPONENTE
52	1	<i>caso de reuso de água</i>	COMPONENTE	52	1	<i>deve atender aos</i>	EXIGENCIA
52	2	<i>deve atender aos</i>	EXIGENCIA	52	2	<i>caso de reuso de água</i>	COMPONENTE
53	1	<i>adoção de soluções</i>	COMPONENTE	53	1	<i>devem privilegiar</i>	EXIGENCIA
53	2	<i>consumo de energia</i>	COMPONENTE	53	2	<i>adoção de soluções</i>	COMPONENTE
53	3	<i>utilização de iluminação</i>	COMPONENTE	53	3	<i>consumo de energia</i>	COMPONENTE
53	4	<i>sistemas de aquecimento</i>	COMPONENTE	53	4	<i>utilização de iluminação</i>	COMPONENTE
53	5	<i>instalações elétricas</i>	COMPONENTE	53	5	<i>sistemas de aquecimento</i>	COMPONENTE
53	6	<i>ventilação natural</i>	COMPONENTE	53	6	<i>instalações elétricas</i>	COMPONENTE
53	7	<i>energia alternativa</i>	COMPONENTE	53	7	<i>ventilação natural</i>	COMPONENTE
53	8	<i>devem privilegiar</i>	EXIGENCIA	53	8	<i>energia alternativa</i>	COMPONENTE
54	1	<i>adoção de soluções</i>	COMPONENTE	54	1	<i>adoção de soluções</i>	COMPONENTE
54	2	<i>consumo de energia</i>	COMPONENTE	54	2	<i>consumo de energia</i>	COMPONENTE
54	3	<i>utilização de iluminação</i>	COMPONENTE	54	3	<i>utilização de iluminação</i>	COMPONENTE
54	4	<i>sistemas de aquecimento</i>	COMPONENTE	54	4	<i>sistemas de aquecimento</i>	COMPONENTE
54	5	<i>ventilação natural</i>	COMPONENTE	54	5	<i>ventilação natural</i>	COMPONENTE
54	6	<i>energia alternativa</i>	COMPONENTE	54	6	<i>energia alternativa</i>	COMPONENTE

## APÊNDICE M - Trecho do regulamento JGJ 120-2012

**4.5.4** The dynamic consolidation construction shall meet the following requirements:

The tamping point position and its contour line shall be marked out before tamping, the ground elevation shall be measured before and after tamping, and the tamping settlement for tamping on each point shall be calculated;

After each tamping, the tamping pit shall be timely filled and levelled or bulldozed, and the site elevation shall be measured to calculate the site tamping settlement for this tamping;

After full compacting times are completed, the loose soil on site surface layer shall be compacted and rolled with low energy full tamping in the tamping principle of lapping  $1/5 \sim 1/3$  of hammer diameter, and the site elevation after dynamic consolidation shall be measured.

The dynamic consolidation shall be carried out by area, from border area to middle part or from the side of adjacent buildings (structures) to the side far away.

**4.5.5** The construction of dynamic replacement shall meet the following requirements:

The material of dynamic replacement pier should be adopted with coarse particle materials with hard texture and stable performance such as well-graded block stone, gravel and slag, and the content of particle with the size greater than 300mm should not be greater than 30% of the full weight;

The piling principle at tamping point should be interlaced skip piling from inside out;

The site elevation shall be measured after each tamping and uplifting amount of site for each tamping shall be calculated, the part of uplifting amount exceeding design elevation should be timely eliminated.

**4.5.6** For soft soil area and area with lighter groundwater depth, the construction by dewatering in conjunction with low level dynamic consolidation shall meet the following requirements:

The dewatering and drainage system shall be arranged first before the construction of dynamic consolidation; the dewatering system should be adopted with vacuum well point system, the peripheral closed well point should be arranged 3m-4m outside the reinforced area;

The dewatering equipment in tamping area shall be removed at least 2 days after the groundwater level lowers to design water level and is stable;

The low level dynamic consolidation shall be in the principle of piling lightly first and then heavily at little point and for multiple times;

The rest time for each pass of dynamic consolidation should be determined according to that the excess pore water pressure dissipation is not less than 80%;

The construction site in area with lighter groundwater depth should be arranged with longitudinal and horizontal drainage network, drainage network and the maximum spacing of drainage network should not be greater than 15m". (MCPRC, 2012, p.8)

## APÊNDICE N – Teste de extração de informações via LLM

### ChatGPT 3.5 - versão navegador

Inicialmente o modelo foi acessado pela interface pública<sup>53</sup> onde foi criado o prompt de comando para o modelo.

Exemplo de prompt usado na interface pública do ChatGPT

```
Prompt:
Dada a frase "Os edifícios multifamiliares devem ser providos de proteção contra
descargas atmosféricas, atendendo ao estabelecido na ABNT NBR 5419 e demais Normas
Brasileiras aplicáveis, nos casos previstos na legislação vigente.", transforme a
frase para uma sentença em formato XML, mas crie marcações apenas quando encontrar
os seguintes termos:
Edifício multifamiliar = Objeto
possuir = Exigencia
proteção contra descargas atmosféricas = Atributo
ABNT NBR 5419 = ReferenciaNormativa
Resultado:
<Sentenca>
  <Objeto>Os edifícios multifamiliares</Objeto>
  <Exigencia>devem ser providos de</Exigencia>
  <Atributo>proteção contra descargas atmosféricas</Atributo>,
  <Exigencia>atendendo ao estabelecido na</Exigencia>
  <ReferenciaNormativa>ABNT NBR 5419</ReferenciaNormativa>
  <Exigencia>e demais Normas Brasileiras aplicáveis, nos casos previstos na
legislação vigente. </Exigencia>
</Sentenca>
<sentenca>
  <Objeto>Os edifícios multifamiliares</Objeto> <Exigencia>devem ser</Exigencia>
<Atributo>providos de proteção contra descargas atmosféricas</Atributo>,
<Exigencia>atendendo ao estabelecido na</Exigencia> <ReferenciaNormativa>ABNT NBR
5419</ReferenciaNormativa> <Exigencia>e demais Normas Brasileiras aplicáveis, nos
casos previstos na legislação vigente. </Exigencia>
</sentenca>
```

Neste exemplo, não foi usada nenhuma técnica de engenharia de prompts e o resultado se mostrou muito inferior ao obtido pelo processo baseado em regras, inclusive com abandono da ordem estabelecida para o modelo. Desta forma, foi feito o registro de uma chave de acesso de uma API na plataforma OpenAI

<sup>53</sup> <https://chat.openai.com/>

## Teste com ChatGPT3.5 via código

Teste desenvolvido em 04 de janeiro de 2024 usando o modelo GPT3.5. O prompt está reproduzido neste apêndice e pode ser acessado pelo link:

<https://chat.openai.com/share/c48b82f9-6e01-453e-af66-fc63586db6e1>

### Teste 1

```
>>You
Você é um especialista em extração de entidades nomeadas e extração de
informações relacionadas à construção civil.
Dadas as frases abaixo:
"Os edifícios multifamiliares devem ser providos de proteção contra descargas
atmosféricas, atendendo ao estabelecido na ABNT NBR 5419 e demais Normas
Brasileiras aplicáveis, nos casos previstos na legislação vigente.",
"As instalações elétricas das edificações habitacionais devem ser projetadas
de acordo com a ABNT NBR 5410 e Normas Brasileiras aplicáveis.",
"Recomenda-se evitar o risco de ignição dos materiais em função de curtos-
circuitos e sobretensões.",
"As instalações de gás devem ser projetadas e executadas de acordo com as
ABNT NBR 13523 e ABNT NBR 15526.",
"A comprovação do atendimento ao requisito de 8.2, pelos critérios
estabelecidos em 8.2.1.1 a 8.2.1.3, deve ser feita pela análise do projeto ou
por inspeção em protótipo.",
"Quando houver ambiente enclausurado, devem ser atendidas a ABNT NBR 15526 e
outras Normas Brasileiras aplicáveis.",
"As rotas de saída de emergência dos edifícios devem atender ao disposto na
ABNT NBR 9077.", "Análise do projeto ou por inspeção em protótipo."

1 - Encontre todas as referências à normas ou regulamentos. Mostre o
resultado na frase delimitando cada norma ou regulamento com o sinal ||.
2 - Na mesma frase, encontre todos os termos que fazem menção à um componente
da construção civil. Mostre o resultado na frase colocando os termos entre
colchetes.
3 - Ainda na mesma frase, encontre todos os termos que fazem menção à uma
obrigação do regulamento. Mostre o resultado na frase colocando os termos
entre chaves.

>>ChatGPT

Vamos analisar cada frase conforme as instruções fornecidas:
1. "Os edifícios multifamiliares devem ser providos de proteção contra
descargas atmosféricas, atendendo ao estabelecido na ABNT NBR 5419 e demais
Normas Brasileiras aplicáveis, nos casos previstos na legislação vigente."
• Referências a normas ou regulamentos: ||ABNT NBR 5419||, ||Normas
Brasileiras aplicáveis||.
• Termos relacionados à construção civil: [edifícios multifamiliares],
[proteção contra descargas atmosféricas].
```

- Termos relacionados à obrigação do regulamento: {devem ser providos}, {atendendo ao estabelecido}, {nos casos previstos}.
2. "As instalações elétricas das edificações habitacionais devem ser projetadas de acordo com a ABNT NBR 5410 e Normas Brasileiras aplicáveis."
- Referências a normas ou regulamentos: ||ABNT NBR 5410||, ||Normas Brasileiras aplicáveis||.
  - Termos relacionados à construção civil: [instalações elétricas], [edificações habitacionais].
  - Termos relacionados à obrigação do regulamento: {devem ser projetadas de acordo com}.
3. "Recomenda-se evitar o risco de ignição dos materiais em função de curtos-circuitos e sobretensões."
- Não há referências a normas ou regulamentos nesta frase.
  - Termos relacionados à construção civil: [ignição dos materiais], [curtos-circuitos], [sobretensões].
  - Termos relacionados à obrigação do regulamento: {Recomenda-se evitar o risco de}.
4. "As instalações de gás devem ser projetadas e executadas de acordo com as ABNT NBR 13523 e ABNT NBR 15526."
- Referências a normas ou regulamentos: ||ABNT NBR 13523||, ||ABNT NBR 15526||.
  - Termos relacionados à construção civil: [instalações de gás].
  - Termos relacionados à obrigação do regulamento: {devem ser projetadas e executadas de acordo com}.
5. "A comprovação do atendimento ao requisito de 8.2, pelos critérios estabelecidos em 8.2.1.1 a 8.2.1.3, deve ser feita pela análise do projeto ou por inspeção em protótipo."
- Não há referências a normas ou regulamentos nesta frase.
  - Termos relacionados à construção civil: [atendimento], [requisito], [critérios], [análise do projeto], [inspeção em protótipo].
  - Termos relacionados à obrigação do regulamento: {deve ser feita pela análise do projeto ou por inspeção em protótipo}.
6. "Quando houver ambiente enclausurado, devem ser atendidas a ABNT NBR 15526 e outras Normas Brasileiras aplicáveis."
- Referências a normas ou regulamentos: ||ABNT NBR 15526||, ||Normas Brasileiras aplicáveis||.
  - Termos relacionados à construção civil: [ambiente enclausurado].
  - Termos relacionados à obrigação do regulamento: {devem ser atendidas}.
7. "As rotas de saída de emergência dos edifícios devem atender ao disposto na ABNT NBR 9077."
- Referência a normas ou regulamentos: ||ABNT NBR 9077||.
  - Termos relacionados à construção civil: [rotas de saída de emergência], [edifícios].
  - Termos relacionados à obrigação do regulamento: {devem atender ao disposto na}.
8. "Análise do projeto ou por inspeção em protótipo."
- Não há referências a normas ou regulamentos nesta frase.
  - Termos relacionados à construção civil: [análise do projeto], [inspeção em protótipo].
- Termos relacionados à obrigação do regulamento: Não aplicável a esta frase.

Essa análise identificou as referências a normas, os termos relacionados à construção civil e os termos relacionados à obrigação do regulamento em cada uma das frases fornecidas.

### **Análise:**

Apesar de não seguir à risca a marcação nas frases, ele retorna listas de itens com grande precisão. O reconhecimento das entidades é impreciso em apenas uma das frases.

## **Teste 2**

Em seguida foi feito outro teste, mas com 26 novas frases e uma regra extra: “4 - Ainda na mesma frase, encontre todos os termos que fazem menção à valores e suas unidades de medida. Mostre o resultado na frase colocando os termos entre \*\*.”

O resultado está apresentado a seguir:

>>início da resposta

Vamos analisar cada frase conforme as regras fornecidas:

"Devem ser previstos nos projetos a prevenção de infiltração da água de chuva e da umidade do solo nas habitações, por meio dos detalhes indicados a seguir: a) condições de implantação dos conjuntos habitacionais, de forma a drenar adequadamente a água de chuva incidente em ruas internas, lotes vizinhos ou mesmo no entorno próximo ao conjunto; b) sistemas que impossibilitem a penetração de líquidos ou umidades de porões e subsolos, jardins contíguos às fachadas e quaisquer paredes em contato com o solo, ou pelo direcionamento das águas, sem prejuízo da utilização do ambiente e dos sistemas correlatos e sem comprometer a segurança estrutural. No caso de haver sistemas de impermeabilização, estes devem seguir a ABNT NBR 9575; c) sistemas que impossibilitem a penetração de líquidos ou umidades em fundações e pisos em contato com o solo; d) ligação entre os diversos elementos da construção (como paredes e estrutura, telhado e paredes, corpo principal e pisos ou calçadas laterais)."

Referências a normas ou regulamentos: ||ABNT NBR 9575||.

Termos relacionados à construção civil: [projetos], [prevenção de infiltração], [conjuntos habitacionais], [água de chuva], [umidade do solo], [sistemas de impermeabilização], [fundações], [pisos], [elementos da construção].

Termos relacionados à obrigação do regulamento: {Devem ser previstos nos projetos}, {de forma a drenar adequadamente}, {sistemas que impossibilitem a penetração}, {ligação entre os diversos elementos}.

Termos relacionados a valores e unidades de medida: Não há nesta frase.

[É importante mencionar que o modelo reconheceu ao todo 9 COMPONENTES, 4 EXIGÊNCIAS e 1 REFERENCIA. Ele também não identificou valores.

"Devem ser previstos no projeto detalhes que assegurem a estanqueidade de partes do edifício que tenham a possibilidade de ficar em contato com a água gerada na ocupação ou manutenção do imóvel, devendo ser verificada a adequação das

vinculações entre instalações de água, esgotos ou águas pluviais e estrutura, pisos e paredes, de forma que as tubulações não venham a ser rompidas ou desencaixadas por deformações impostas."

Referências a normas ou regulamentos: Não há nesta frase.

Termos relacionados à construção civil: [projeto], [estanqueidade], [partes do edifício], [água gerada], [ocupação], [manutenção do imóvel], [vinculações], [instalações de água], [esgotos], [águas pluviais], [estrutura], [pisos], [paredes], [tubulações].

Termos relacionados à obrigação do regulamento: {Devem ser previstos no projeto}, {asseguem a estanqueidade}, {verificada a adequação das vinculações}.

Termos relacionados a valores e unidades de medida: Não há nesta frase.

"O programa de simulação computacional deve ser capaz de estimar as variações da temperatura operativa, das cargas térmicas de refrigeração e de aquecimento e do uso da ventilação natural na UH, definidos separadamente em 8 760 h ao longo do ano, considerando as variações horárias de ocupação, de potência de iluminação e de equipamentos. O programa de simulação computacional deve, ainda: a) estar de acordo com a ASHRAE 140, segundo o procedimento de teste da Classe I; b) modelar efeitos de inércia térmica; c) modelar trocas de calor entre a edificação e o solo; d) calcular cargas térmicas latente e sensível; e) ser capaz de simular o sombreamento proveniente de elementos externos às zonas térmicas, como brises, sacadas e entorno; f) ser capaz de simular os efeitos da ventilação cruzada em um ambiente, ou entre dois ou mais ambientes."

Referências a normas ou regulamentos: ||ASHRAE 140||.

Termos relacionados à construção civil: [programa de simulação computacional], [temperatura operativa], [cargas térmicas], [ventilação natural], [variações horárias], [ocupação], [potência de iluminação], [equipamentos], [efeitos de inércia térmica], [trocas de calor], [edificação], [solo], [cargas térmicas latente e sensível], [sombreamento], [elementos externos], [zonas térmicas], [brises], [sacadas], [entorno], [ventilação cruzada].

Termos relacionados à obrigação do regulamento: {O programa de simulação computacional deve ser capaz de estimar}, {estar de acordo com}, {modelar efeitos}, {modelar trocas de calor}, {calcular cargas térmicas}, {ser capaz de simular}, {ser capaz de simular os efeitos da ventilação cruzada}.

Termos relacionados a valores e unidades de medida: [8 760 h ao longo do ano].

"O arquivo climático deve possuir informações que sejam representativas do clima da cidade onde a UH está localizada.

Caso a cidade de implantação da UH não possua arquivo climático, deve ser utilizado o arquivo climático de uma cidade próxima com clima semelhante. A semelhança entre os climas deve considerar a comparação da distância euclidiana, ponderando latitude, longitude e altitude."

Referências a normas ou regulamentos: Não há nesta frase.

Termos relacionados à construção civil: [arquivo climático], [clima da cidade], [UH - Unidade Habitacional], [cidade de implantação], [clima semelhante], [distância euclidiana], [latitude], [longitude], [altitude].

Termos relacionados à obrigação do regulamento: {O arquivo climático deve possuir informações}, {Caso a cidade de implantação da UH não possua arquivo climático, deve ser utilizado}, {A semelhança entre os climas deve considerar a comparação da distância euclidiana, ponderando latitude, longitude e altitude}.

Termos relacionados a  
>> fim da resposta

O ChatGPT3.5 tem limite de 4.096 caracteres para resposta, já o modelo mais novo, 4.0, limite é de 8.192 tokens<sup>54</sup>.

Desta forma foi utilizada a API dada pela OpenAI para acesso ao ChatGPT 3.5 e com a capacidade de processar mais tokens. Como o modelo não é gratuito como o Lamma e o Boom, a API exige a cobrança por cada token processado na leitura e na resposta. O valor de cada token é \$0.001 para cada mil tokens de input e \$0.002 para output. O modelo GPT 4.0 tem sido apresentado para tarefas complexas e tem custo dez vezes maior, \$0.01/1000 tokens para input e \$0.03/1000 tokens para output.

Tokens usados atuais: 14.772

Uma opção recomendada para melhoria do resultado é o ajuste fino do modelo, mas seria necessário apresentar um conjunto de dados já rotulado de pelo menos 10 exemplos, mas o ideal seria algo entre 50 a 100<sup>55</sup>. Outro parâmetro disponível é a criação de um prompt adequado à tarefa. Para o modelo GPT – 3.5 – turbo – 1106, o exemplo pode ter até 16385 tokens, enquanto o gpt-3.5-turbo – 0613, aceita até 4.096 tokens por exemplo.

Os tokens processados pelos modelos de linguagem não correspondem aos tokens usados em PLN na biblioteca SpaCy e NLTK. Eles são sequencias comuns de caracteres e normalmente são os 4 primeiros caracteres de uma palavra. Com base nestes caracteres os modelos conseguem prever a continuação da palavra no texto. Nas 53 frases da parte 1 da norma, foram considerados 4.520 tokens e 16.424 caracteres<sup>56</sup>.

<sup>54</sup> [https://cookbook.openai.com/examples/how\\_to\\_format\\_inputs\\_to\\_chatgpt\\_models](https://cookbook.openai.com/examples/how_to_format_inputs_to_chatgpt_models)

<sup>55</sup> <https://platform.openai.com/docs/guides/fine-tuning/preparing-your-dataset>

<sup>56</sup> <https://platform.openai.com/tokenizer>

Ele retorna um erro porque excede o limite de 4.096 itens. Então fazendo a análise das 40 primeiras frases:

### Teste 3

```
ChatCompletionMessage(content='
1 - Referências a normas ou regulamentos:
[ABNT NBR 8044], [ABNT NBR 5629], [ABNT NBR 11682], [ABNT NBR 6122], [ABNT NBR
12722], [ABNT NBR 15575-2 a ABNT NBR 15575-6], [ABNT NBR 14718], [ABNT NBR 15575-
3], [ABNT NBR 15575-4], [ABNT NBR 15575-5], [ABNT NBR 9575], [ISO 5034-1], [ASHRAE
140], [ABNT NBR 9050].
2 - Termos que fazem menção a um componente da construção civil:
[topográficas, geológicas], [local da obra], [projetos de arquitetura],
[estrutura], [fundações], [convenções], [outras eventuais obras geotécnicas],
[riscos de deslizamentos], [enchentes], [erosões], [vibrações transmitidas por vias
férreas], [vibrações transmitidas por trabalhos']
role='assistant', function_call=None, tool_calls=None)
```

Após inserir um *loop* para análise por frase, foi obtido o seguinte resultado:

```
ChatCompletionMessage(content='Para edifícios ou conjuntos habitacionais com local
de implantação definido, os projetos de arquitetura, da estrutura, das fundações,
convenções e outras eventuais obras geotécnicas devem ser desenvolvidos com base
nas características do local da obra (topográficas, geológicas etc.), avaliando-se
convenientemente os riscos de deslizamentos, enchentes, erosões, vibrações
transmitidas por vias férreas, vibrações transmitidas por trabalhos de
terraplenagem e compactação do solo, ocorrência de subsidência do solo, presença de
crateras em camadas profundas, presença de solos expansíveis ou colapsíveis,
presença de camadas profundas deformáveis e outros.

Etapa 1: $$topográficas, geológicas etc.$$

Etapa 2: [edifícios], [conjuntos habitacionais], [projetos de arquitetura],
[estrutura], [fundações], [convenções], [outras eventuais obras geotécnicas],
[local da obra], [riscos de deslizamentos], [enchentes], [erosões], [vibrações
transmitidas por vias férreas], [vibrações transmitidas por trabalhos de
terraplenagem e compactação do solo], [ocorrência de subsidência do solo],
[presença de crateras em camadas profundas], [presença de solos expansíveis ou
colapsíveis], [presença de camadas profundas deformáveis]

Etapa 3: {devem ser desenvolvidos}

Etapa 4: **compactação do solo**, role='assistant', function_call=None,
tool_calls=None)

ChatCompletionMessage(content=
Etapa 1: Não há referências a normas ou regulamentos.
Etapa 2: Não há menções a componentes da construção civil.
Etapa 3: Não há menções a obrigações do regulamento.
Etapa 4: Não há menções a valores e unidades de medida.', role='assistant',
function_call=None, tool_calls=None)
```

```

ChatCompletionMessage(content='Os projetos devem ainda prever as interações entre
construções próximas, considerando-se convenientemente as eventuais sobreposições
de bulbos de pressão, efeitos de grupo de estacas, rebaixamento do lençol freático
e desconfinamento do solo em função do corte do terreno.
Etapa 1: [$bulbos de pressão$, $efeitos de grupo de estacas$, $rebaixamento do
lençol freático$, $desconfinamento do solo em função do corte do terreno$]
Etapa 2: [construções, bulbos de pressão, grupo de estacas, lençol freático, solo,
corte, terreno]
Etapa 3: [prever, interações, convenientemente, sobreposições]
Etapa 4: [], role='assistant', function_call=None, tool_calls=None)

ChatCompletionMessage(content='
Etapa 1: Referências à normas ou regulamentos: $$não podem prejudicar$$
Etapa 2: Termos que fazem menção à um componente da construção civil: ("Tais
fenômenos também não podem prejudicar a segurança e a funcionalidade da [obra], bem
como de [edificações] vizinhas.")
Etapa 3: Termos que fazem menção à uma obrigação do regulamento: ("Tais fenômenos
também [não podem prejudicar] a segurança e a funcionalidade [da obra], bem como de
[edificações] vizinhas.")
Etapa 4: Termos que fazem menção à valores e suas unidades de medida: ("Tais
fenômenos também não podem prejudicar a segurança e a funcionalidade da obra, bem
como de edificações vizinhas.")', role='assistant', function_call=None,
tool_calls=None)

E o erro:
RateLimitError: Error code: 429 - {'error': {'message': 'Rate limit reached for
gpt-3.5-turbo in organization org-EyXFR4Td0z1V0J901BtpDMKB on requests per min
(RPM): Limit 3, Used 3, Requested 1. Please try again in 20s. Visit
https://platform.openai.com/account/rate-limits to learn more. You can increase
your rate limit by adding a payment method to your account at
https://platform.openai.com/account/billing.', 'type': 'requests', 'param': None,
'code': 'rate_limit_exceeded'}}

```

Não foi possível, em uma tentativa simples, de configurar o código para que o prompt fosse compreendido como na opção via navegador no site da OpenAi (<https://chat.openai.com/share/c48b82f9-6e01-453e-af66-fc63586db6e1>).

Na experiência codificada, onde as frases seriam informadas por meio de um “loop” ou como uma lista de sentenças, apenas a primeira frase foi interpretada e seus termos classificados.

Ao final de algumas tentativas a interface de custos da plataforma informou que foram contabilizados 27.078 tokens no input e 1.608 tokens para o output das respostas. O custo total para estas tentativas foi de \$0.04.

Foi feito um teste com o ChatGPT4 embutido na plataforma Microsoft Bing a partir dos seguintes prompts:

## Teste 4

### PROMPT

Você é um especialista em extração de entidades nomeadas e extração de informações relacionadas à construção civil.

Dadas as frases abaixo:

"Os edifícios multifamiliares devem ser providos de proteção contra descargas atmosféricas, atendendo ao estabelecido na ABNT NBR 5419 e demais Normas Brasileiras aplicáveis, nos casos previstos na legislação vigente.",

"As instalações elétricas das edificações habitacionais devem ser projetadas de acordo com a ABNT NBR 5410 e Normas Brasileiras aplicáveis.",

"Recomenda-se evitar o risco de ignição dos materiais em função de curtos-circuitos e sobretensões.",

"As instalações de gás devem ser projetadas e executadas de acordo com as ABNT NBR 13523 e ABNT NBR 15526.",

"A comprovação do atendimento ao requisito de 8.2, pelos critérios estabelecidos em 8.2.1.1 a 8.2.1.3, deve ser feita pela análise do projeto ou por inspeção em protótipo.",

"Quando houver ambiente enclausurado, devem ser atendidas a ABNT NBR 15526 e outras Normas Brasileiras aplicáveis.",

"As rotas de saída de emergência dos edifícios devem atender ao disposto na ABNT NBR 9077.",

"Análise do projeto ou por inspeção em protótipo."

1 - Encontre todas as referências à normas ou regulamentos. Mostre o resultado na frase delimitando cada norma ou regulamento com o sinal ||.

2 - Na mesma frase, encontre todos os termos que fazem menção à um componente da construção civil. Mostre o resultado na frase colocando os termos entre colchetes.

3 - Ainda na mesma frase, encontre todos os termos que fazem menção à uma obrigação do regulamento. Mostre o resultado na frase colocando os termos entre chaves.

### RESPOSTA

Claro, aqui estão as frases com as entidades nomeadas destacadas conforme solicitado:

"Os [edifícios multifamiliares] devem ser providos de [proteção contra descargas atmosféricas], atendendo ao estabelecido na ||ABNT NBR 5419|| e demais Normas Brasileiras aplicáveis, nos casos previstos na legislação vigente."

"As [instalações elétricas] das [edificações habitacionais] devem ser projetadas de acordo com a ||ABNT NBR 5410|| e Normas Brasileiras aplicáveis."

"Recomenda-se evitar o risco de ignição dos [materiais] em função de [curtos-circuitos] e [sobretensões]."

"As [instalações de gás] devem ser projetadas e executadas de acordo com as ||ABNT NBR 13523|| e ||ABNT NBR 15526||."

"A comprovação do atendimento ao requisito de 8.2, pelos critérios estabelecidos em 8.2.1.1 a 8.2.1.3, deve ser feita pela análise do projeto ou por inspeção em protótipo."

"Quando houver [ambiente enclausurado], devem ser atendidas a ||ABNT NBR 15526|| e outras Normas Brasileiras aplicáveis."

"As [rotas de saída de emergência] dos [edifícios] devem atender ao disposto na ||ABNT NBR 9077||."

"Análise do projeto ou por inspeção em protótipo."

Por favor, note que as normas e regulamentos estão delimitados por ||, os componentes da construção civil estão entre colchetes [], e as obrigações do regulamento estão entre chaves {}. Espero que isso ajude! Se você tiver mais perguntas ou precisar de mais assistência, fique à vontade para perguntar.

## Teste 5

Novas frases e um incremento processual na regra:

PROMPT

Repita a mesma análise para as frases a seguir obedecendo as regras que são indicadas.

Frases:

Devem ser previstos nos projetos a prevenção de infiltração da água de chuva e da umidade do solo nas habitações, por meio dos detalhes indicados a seguir: a) condições de implantação dos conjuntos habitacionais, de forma a drenar adequadamente a água de chuva incidente em ruas internas, lotes vizinhos ou mesmo no entorno próximo ao conjunto; b) sistemas que impossibilitem a penetração de líquidos ou umidades de porões e subsolos, jardins contíguos às fachadas e quaisquer paredes em contato com o solo, ou pelo direcionamento das águas, sem prejuízo da utilização do ambiente e dos sistemas correlatos e sem comprometer a segurança estrutural.

No caso de haver sistemas de impermeabilização, estes devem seguir a ABNT NBR 9575; c) sistemas que impossibilitem a penetração de líquidos ou umidades em fundações e pisos em contato com o solo; d) ligação entre os diversos elementos da construção (como paredes e estrutura, telhado e paredes, corpo principal e pisos ou calçadas laterais).

Devem ser previstos no projeto detalhes que assegurem a estanqueidade de partes do edifício que tenham a possibilidade de ficar em contato com a água gerada na ocupação ou manutenção do imóvel, devendo ser verificada a adequação das vinculações entre instalações de água, esgotos ou águas pluviais e estrutura, pisos e paredes, de forma que as tubulações não venham a ser rompidas ou desencaixadas por deformações impostas.

O programa de simulação computacional deve ser capaz de estimar as variações da temperatura operativa, das cargas térmicas de refrigeração e de aquecimento e do uso da ventilação natural na UH, definidos separadamente em 8 760 h ao longo do ano, considerando as variações horárias de ocupação, de potência de iluminação e de equipamentos. O programa de simulação computacional deve, ainda: a) estar de acordo com a ASHRAE 140, segundo o procedimento de teste da Classe I; b) modelar efeitos de inércia térmica; c) modelar trocas de calor entre a edificação e o solo; d) calcular cargas térmicas latente e sensível; e) ser capaz de simular o sombreamento proveniente de elementos externos às zonas térmicas, como brises, sacadas e entorno; f) ser capaz de simular os efeitos da ventilação cruzada em um ambiente, ou entre dois ou mais ambientes.

O arquivo climático utilizado deve fornecer valores mensais de temperatura média do solo, em graus Celsius (°C), para todos os meses do ano, além dos seguintes valores horários representativos das 8760 h do ano meteorológico típico: a) temperatura de bulbo seco, expressa em graus Celsius (°C); b) temperatura do ponto de orvalho, expressa em graus Celsius (°C); c) umidade relativa, expressa em porcentagem (%); d) pressão atmosférica, expressa em Pascal (Pa); e) intensidade de radiação

horizontal de onda longa, expressa em watts-hora por metro quadrado (Wh/m<sup>2</sup>); f) radiação horizontal global, expressa em watts-hora por metro quadrado (Wh/m<sup>2</sup>); g) radiação normal direta, expressa em watts-hora por metro quadrado (Wh/m<sup>2</sup>); h) radiação horizontal difusa, expressa em watts-hora por metro quadrado (Wh/m<sup>2</sup>); i) direção do vento, expressa em graus (°), considerando o sentido horário a partir da direção Norte; j) velocidade do vento, expressa em metros por segundo (m/s).

#### Regras:

- 1 - Encontre todas as referências à normas ou regulamentos. Mostre o resultado na frase delimitando cada norma ou regulamento com o sinal ||.
  - 2 - Na mesma frase, encontre todos os termos que fazem menção à um componente da construção civil. Mostre o resultado na frase colocando os termos entre colchetes.
  - 3 - Ainda na mesma frase, encontre todos os termos que fazem menção à uma obrigação do regulamento. Mostre o resultado na frase colocando os termos entre chaves.
  - 4 - Ainda na mesma frase, encontre todos os termos que fazem menção à valores e suas unidades de medida. Mostre o resultado na frase colocando os termos entre \*\*. [neste caso foi convertida para negrito]
- [Nota: Este prompt ficou limitado a 4000 caracteres]

#### RESPOSTA

Claro, aqui estão as frases com as entidades nomeadas destacadas conforme solicitado:

“{Devem ser previstos nos projetos} a prevenção de infiltração da água de chuva e da umidade do solo nas [habitações], por meio dos detalhes indicados a seguir: a) condições de implantação dos [conjuntos habitacionais], de forma a drenar adequadamente a água de chuva incidente em ruas internas, lotes vizinhos ou mesmo no entorno próximo ao conjunto; b) sistemas que impossibilitem a penetração de líquidos ou umidades de [porões] e [subsolo], [jardins] contíguos às [fachadas] e quaisquer [paredes] em contato com o solo, ou pelo direcionamento das águas, sem prejuízo da utilização do ambiente e dos sistemas correlatos e sem comprometer a segurança estrutural.”

“No caso de haver sistemas de impermeabilização, {estes devem seguir} a ||ABNT NBR 9575||; c) sistemas que impossibilitem a penetração de líquidos ou umidades em [fundações] e [pisos] em contato com o solo; d) ligação entre os diversos elementos da construção (como [paredes] e [estrutura], [telhado] e [paredes], corpo principal e [pisos] ou [calçadas] laterais).”

“{Devem ser previstos no projeto detalhes} que assegurem a estanqueidade de partes do [edifício] que tenham a possibilidade de ficar em contato com a água gerada na ocupação ou manutenção do imóvel, devendo ser verificada a adequação das vinculações entre [instalações de água], [esgotos] ou [águas pluviais] e [estrutura], [pisos] e [paredes], de forma que as [tubulações] não venham a ser rompidas ou desencaixadas por deformações impostas.”

“O programa de simulação computacional {deve ser capaz de estimar} as variações da temperatura operativa, das cargas térmicas de refrigeração e de aquecimento e do uso da ventilação natural na UH, definidos separadamente em 8 760 h ao longo do ano, considerando as variações horárias de ocupação, de potência de iluminação e de equipamentos. O programa de simulação computacional deve, ainda: a) estar de acordo com a ||ASHRAE 140||, segundo o procedimento de teste da Classe I; b) modelar efeitos de inércia térmica; c) modelar trocas de calor entre a [edificação] e o solo; d) calcular cargas térmicas latente e sensível; e) ser capaz de simular o sombreamento proveniente de elementos externos às zonas térmicas, como brises, sacadas e entorno; f) ser capaz de simular os efeitos da ventilação cruzada em um ambiente, ou entre dois ou mais ambientes.”

“O arquivo climático utilizado {deve fornecer valores mensais} de temperatura média do solo, em graus Celsius (°C), para todos os meses do ano, além dos seguintes valores horários representativos das 8760 h do ano meteorológico típico: a) temperatura de bulbo seco, expressa em graus Celsius (°C); b) temperatura do ponto de orvalho, expressa em graus Celsius (°C); c) umidade relativa, expressa em porcentagem (%); d) pressão atmosférica, expressa em Pascal (Pa); e) intensidade de radiação horizontal de onda longa, expressa em watts-hora por metro quadrado (Wh/m<sup>2</sup>); f) radiação horizontal global, expressa em watts-hora por metro quadrado (Wh/m<sup>2</sup>); g) radiação normal direta, expressa em watts-hora por metro quadrado (Wh/m<sup>2</sup>); h) radiação horizontal difusa, expressa em watts-hora por metro quadrado (Wh/m<sup>2</sup>); i) direção do vento, expressa em graus (°), considerando o sentido horário a partir da direção Norte; j) velocidade do vento, expressa em metros por segundo (m/s).”

Por favor, note que as normas e regulamentos estão delimitados por ||, os componentes da construção civil estão entre colchetes [], as obrigações do regulamento estão entre chaves {}, e os valores e suas unidades de medida estão entre \*\*. Espero que isso ajude! Se você tiver mais perguntas ou precisar de mais assistência, fique à vontade para perguntar.

## Teste 6 - Microsoft Bing/ ChatGPT 4 - Versão web

Abaixo está o prompt usado para Microsoft Bing utilizando as técnicas de personificação e sequenciamento de ações<sup>57</sup>, O resultado entre os modelos foi reunido e apresentado no Quadro 37. Para comparação foi realizada a marcação humana como padrão-ouro (*gold-standard*) e em seguida comparado com os resultados dos modelos para e mesma sentença.

Prompt:

Você é um especialista em extração de entidades nomeadas e extração de informações relacionadas à construção civil.

Use as seguintes regras para marcar os trechos do texto:

Etapa1 - Encontre todas as referências à normas ou regulamentos. Mostre o resultado na frase delimitando cada norma ou regulamento com o sinal \$\$.

Etapa 2 - Na mesma frase, encontre todos os termos que fazem menção à um componente da construção civil. Mostre o resultado na frase colocando os termos entre colchetes.

Etapa 3 - Ainda na mesma frase, encontre todos os termos que fazem menção à uma obrigação do regulamento. Mostre o resultado na frase colocando os termos entre chaves.

Etapa 4 - Ainda na mesma frase, encontre todos os termos que fazem menção à valores e suas unidades de medida. Mostre o resultado na frase colocando os termos entre \*\*.

Dada a frase: "

<sup>57</sup> <https://platform.openai.com/docs/guides/prompt-engineering>

"Devem ser previstos nos projetos a prevenção de infiltração da água de chuva e da umidade do solo nas habitações, por meio dos detalhes indicados a seguir: a) condições de implantação dos conjuntos habitacionais, de forma a drenar adequadamente a água de chuva incidente em ruas internas, lotes vizinhos ou mesmo no entorno próximo ao conjunto; b) sistemas que impossibilitem a penetração de líquidos ou umidades de porões e subsolos, jardins contíguos às fachadas e quaisquer paredes em contato com o solo, ou pelo direcionamento das águas, sem prejuízo da utilização do ambiente e dos sistemas correlatos e sem comprometer a segurança estrutural. No caso de haver sistemas de impermeabilização, estes devem seguir a ABNT NBR 9575; c) sistemas que impossibilitem a penetração de líquidos ou umidades em fundações e pisos em contato com o solo; d) ligação entre os diversos elementos da construção (como paredes e estrutura, telhado e paredes, corpo principal e pisos ou calçadas laterais)."

Marque os trechos dos textos com as regras informadas

## APÊNDICE O – Teste de implementação da técnica RAG (LangChain)

Os testes estão organizados por tipo de esquema de dados fornecido ao modelo de linguagem. Tanto o JGJ 120-2012 como o ICB 2009 foram usado pela pesquisa de referência (Zhang; El-Gohary, 2016). Aqui os dois regulamentos são usados para comparação dos resultados usando os mesmos rótulos semânticos. O Código completo com os exemplos textuais está na IDE Google Colab<sup>58</sup>. Cada sentença foi extraída três vezes (A, B e C) para avaliar a consistência (repetibilidade) da extração.

### TESTE 1 – JGJ 120-2012

```

compliance_schema = Object(
  id="requirement",
  description="Requirement for construction or evaluation of a building or one of
its components",

  #Atributos
  attributes=[
    Text(
      id= "subject",
      description= "A construction element in the sentence."),
    Text(
      id= "subject_restriction",
      description= "An existential relation between construction elements."),
    Text(
      id= "compliance_checking_attribute",
      description= "An element, object or property that should be evaluated
in this requirement."),
    Text(
      id="deontic_operator_indicator",
      description="An obligation or a possibility."),
    Text(
      id= "comparative_relation",
      description= "A comparative relation between 2 contrction elements."),
    Number(
      id= "quantity_value",
      description= "A integer or float value." ),
    Text(
      id= "quantity_unit_reference",
      description= "The unit of measurement"),
    Text(
      id= "quantity_restriction",

```

<sup>58</sup>

[https://colab.research.google.com/drive/16n9tkZYKso28iwHVmfK\\_GYYYYIt0fSyBS?authuser=2#scrollTo=zAKHt8IBDEFQ](https://colab.research.google.com/drive/16n9tkZYKso28iwHVmfK_GYYYYIt0fSyBS?authuser=2#scrollTo=zAKHt8IBDEFQ)

```

description= "A restriction or condition imposed to the measure.")],
# Exemplos
examples=[
  ("The minimum net area of ventilation openings shall not be less than 1 ft2
for each 150 ft2 of crawl space area",
  [{"subject": "ventilation openings",
    "subject_restriction": "N/A",
    "compliance_checking_attribute": "net area",
    "deontic_operator_indicator": "shall not be",
    "comparative_relation": "less than",
    "quantity_value": "1",
    "quantity_unit_reference": "ft2",
    "quantity_restriction": "for each 150 ft2 of crawl space area"}
  ]),
)

```

## Resultados

<b>Frase 1</b>			
<i>The rest time for each pass of dynamic consolidation should be determined according to that the excess pore water pressure dissipation is not less than 80%;</i>			
<b>Termos semânticos</b>	<b>A</b>	<b>B</b>	<b>C</b>
comparative relation	<i>not less than</i>	<i>not less than</i>	<i>not less than</i>
compliance checking attribute	<i>excess pore water pressure dissipation</i>	<i>excess pore water pressure dissipation</i>	<i>excess pore water pressure dissipation</i>
deontic operator indicator	<i>should be determined according to</i>	<i>should be determined according to</i>	<i>should be determined according to</i>
quantity restriction	-	-	-
quantity unit reference	%	%	%
quantity value	80	80	80
subject	<i>rest time</i>	<i>rest time</i>	<i>rest time</i>
subject restriction	<i>for each pass of dynamic consolidation</i>	<i>for each pass of dynamic consolidation</i>	<i>for each pass of dynamic consolidation</i>

Nesta frase o resultado foi consistente. Em 3 tentativas não houve alteração do resultado e obteve precisão nos termos extraídos.

<b>Frase 2</b>				
<i>The material of dynamic replacement pier should be adopted with with coarse particle materials with hard texture and stable performance such as well-graded block stone, gravel and slag, and the content of particle with the size greater than 300mm should not be greater than 30% of the full weight.</i>				
<b>Termos semânticos</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>C(cont.)</b>
comparative relation	<i>greater than</i>	<i>greater than</i>	N/A	30
compliance checking attribute	N/A	N/A	N/A	<i>content of</i>
deontic operator indicator	<i>should be</i>	<i>greater than</i>	<i>should be adopted with</i>	<i>should not be greater than</i>
quantity restriction	<i>content of particle with the size greater than 300mm should not be greater than</i>	N/A	<i>coarse particle materials with hard texture and stable performance such as</i>	<i>particle with the size greater than 300mm</i>

			<i>well-graded block stone, gravel and slag</i>	
quantity unit reference	%	<i>greater than</i>	N/A	<i>the full weight</i>
quantity value	30	N/A	N/A	%
subject	<i>dynamic replacement pier</i>	<i>greater than</i>	<i>dynamic replacement pier</i>	<i>particle</i>
subject restriction	<i>material of</i>	N/A	<i>material of</i>	<i>size</i>

Alguns termos extraídos de modo correto, mas sem consistência.

## TESTE 2 – Avaliação no ICB 2009

<b>Frase 1</b>			
<i>When the total volume of a given class of concrete is less than 50 cubic yards (38 m<sup>3</sup>), strength tests are not required when evidence of satisfactory strength is submitted to and approved by the building official.</i>			
<b>Termos semânticos</b>	<b>A</b>	<b>B</b>	<b>C</b>
comparative relation	less than	less than	less than
compliance checking attribute	N/A	strength tests	strength tests
deontic operator indicator	are not required	are not required	are not required
quantity restriction	-	38 m <sup>3</sup>	38 m <sup>3</sup>
quantity unit reference	cubic yards	cubic yards	cubic yards
quantity value	50	50	50
subject	total volume	total volume	total volume
subject restriction	of a given class of concrete	of a given class of concrete	of a given class of concrete

Apesar de inconsistente, houve repetição de valores. A alucinação pode ser minimizada ainda mais com adição de prompt, que não foi usado aqui.

<b>Frase 2</b>			
<i>The total area of ventilation openings is permitted to be reduced to 1/1500 of the under-floor area where the ground surface' is covered with a Class I vapor retarder material and the required openings are placed so as to provide cross ventilation of the space</i>			
<b>Termos semânticos</b>	<b>A</b>	<b>B</b>	<b>C</b>
comparative relation	reduced to	reduced to	reduced to
compliance checking attribute	N/A	N/A	N/A
deontic operator indicator	is permitted to be	is permitted to be	is permitted to be
quantity restriction	of the under-floor area where the ground surface' is covered with a Class I vapor retarder material and the required openings are placed so as to provide cross ventilation of the space	of the under-floor area	of the under-floor area
quantity unit reference	N/A	N/A	N/A
quantity value	1/1500	1/1500	1/1500
subject	total area of ventilation openings	total area of ventilation openings	total area of ventilation openings
subject restriction	N/A	N/A	N/A

<b>Frase 3</b>			
<i>Connections that are designed to yield shall be capable of maintaining 80 percent of their design strength at the deformation induced by the design displacement or shall use Type 2 mechanical splice</i>			

Termos semânticos	A	B	C
comparative relation	percent of	80	80
compliance checking attribute	design strength	design strength	design strength
deontic operator indicator	shall be capable of maintaining	shall be capable of	shall be capable of
quantity restriction	at the deformation induced by the design displacement		N/A
quantity unit reference	%	maintaining 80 percent of their design strength at the deformation induced by the design displacement	maintaining
quantity value	80	percent	percent
subject	Connections	Connections	Connections
subject restriction	that are designed to yield	that are designed to yield	that are designed to yield

Frase 4			
<i>Transverse reinforcement shall be extended beyond the pier clear height for at least 12 inches (305 mm).</i>			
Termos semânticos	A	B	C
comparative relator	beyond	beyond	beyond
compliance checking attribute	extended	extended	extended
deontic operator indicator	shall be	shall be	shall be
quantity restriction	at least	at least	at least
quantity unit reference	inches	inches	inches
quantity value	12	12	12
subject	Transverse reinforcement	Transverse reinforcement	Transverse reinforcement
subject restriction	N/A	N/A	N/A

Esta frase mostra que expressões simples, bem definidas são reconhecidas com mais certeza (probabilidade) pelo modelo.

### TESTE 3 – Esquema de dados simplificado, em português e sem incorporar a biblioteca Kor, apenas a LangChain

```

schema = {
  "properties": {
    "componente_edificio": {"type": "string"},
    "menção_à_norma": {"type": "Boolean"}, #exemplo ABNT 00.000
    "exigência": {"type": "string"},
    "valor": {"type": "integer"},
    "unidade_de_medida": {"type": "string"},
  },
  "required": ["componente_edificio", "menção_à_norma", "exigência", "valor", "unidade_de_medida"]}

```

### Resultados

Frase 1
Os componentes, equipamentos e dispositivos de manobra devem ser projetados, construídos e montados de forma a evitar que a força necessária para o acionamento não exceda 10 N nem o torque ultrapasse 20 N.m.

<b>Termos semânticos</b>	<b>A</b>	<b>A(cont.)</b>	<b>B</b>	<b>B(cont.)</b>
Componente edifício	componentes, equipamentos e dispositivos de manobra	componentes, equipamentos e dispositivos de manobra	componentes, equipamentos e dispositivos de manobra	componentes, equipamentos e dispositivos de manobra
Existe menção à norma	False	False	False	False
Expressão deôntica	devem ser projetados, construídos e montados	devem ser projetados, construídos e montados	devem ser	devem ser
Valor exigência	10	20	10	20
Unidade medida	N	N.m	N	N.m
<b>Frase 1 (cont.)</b>				
Os componentes, equipamentos e dispositivos de manobra devem ser projetados, construídos e montados de forma a evitar que a força necessária para o acionamento não exceda 10 N nem o torque ultrapasse 20 N.m.				
<b>Termos semânticos</b>	<b>C</b>	<b>C(cont.)</b>		
Componente edifício	componentes, equipamentos e dispositivos de manobra	componentes, equipamentos e dispositivos de manobra		
Existe menção à norma	False	False		
Expressão deôntica	devem ser projetados, construídos e montados	devem ser projetados, construídos e montados		
Valor exigência	10	20		
Unidade medida	N	N.m		

Aqui o resultado é consistente, sem alteração das informações nas 3 avaliações.

## TESTE 4 – frases em português, com Kor e LangChain.

```

compliance_schema = Object(
  id="requisito",

  description="Requisito para construção ou avaliação do edifício ou de seus componentes",

  attributes=[
    Text(
      id= "elemento construtivo 1"
      description= "O componente construtivo principal da frase."),
    Text(
      id= "elemento construtivo 2",
      description= "Outro componente construtivo da frase."),
    Text(
      id= "elemento construtivo 3",
      description= "Outro componente construtivo da frase."),
    Text(
      id= "elemento construtivo 4",
      description= "Outro componente construtivo da frase."),
    Text(
      id="indicador_de_operador_deontico",
      description="Uma expressão de obrigação ou possibilidade"),
  ]
)

```

```

Text(
  id= "menção à norma",
  description= "citação de uma referência normativa como ABNT 00.000"),

#Exemplos
examples=[
  ("A área líquida mínima das aberturas de ventilação não deve ser inferior a 1 m2 para cada 150 m2 de área de
manutenção, segundo a ABNT NBR 13.450",
  [{"elemento": "aberturas de ventilacao",
    "elemento construtivo 2": "área de manutenção",
    "elemento construtivo 3": "N/A",
    "elemento construtivo 4": "N/A",
    "indicador_de_operador_deontico": "não deve ser",
    "menção à norma": "ABNT NBR 13.450"}]),
)

```

## Resultados

Frase 1			
Em vestibulos, halls, corredores, instalações sanitárias e despensas, é permitido que o pé-direito seja reduzido ao mínimo de 2,30 m.			
Termos semânticos	A	B	C
elemento	Pé-direito	Pé-direito	Pé-direito
relação de restricao	Permitido que	Permitido que	Permitido que
atributo para verificacao	Seja reduzido ao mínimo de	Seja reduzido ao mínimo de	Seja reduzido ao mínimo de
indicador de operador deontico	é	é	é
relacao comparativa	Igual a	Igual a	Igual a
valor de quantidade	N/A	N/A	N/A
unidade de medida	m	m	m
restrição de quantidade	2.30	2.30	2.30

Frase 2				
Os componentes, equipamentos e dispositivos de manobra devem ser projetados, construídos e montados de forma a evitar que a força necessária para o acionamento não exceda 10 N nem o torque ultrapasse 20 N.m.				
Termos semânticos	A	A(cont.)	B	B(cont.)
elemento	componentes, equipamentos e dispositivos de manobra	componentes, equipamentos e dispositivos de manobra	componentes, equipamentos e dispositivos de manobra	componentes, equipamentos e dispositivos de manobra
relação de restricao	N/A	N/A	N/A	N/A
atributo para verificacao	força necessária para o acionamento	torque	força necessária para o acionamento	torque
indicador de operador deontico	Não exceda	Não ultrapasse	Não exceda	Não ultrapasse
relacao comparativa	10	20	10	20
valor de quantidade	N	N.m	N	N.m
unidade de medida	N/A	N/A	N/A	N/A
restrição de quantidade				

Foi feito mais um teste com resultados idênticos

<b>Frase 3</b>			
No caso de reuso de água para destinação não potável, esta deve atender aos parâmetros estabelecidos na Tabela 8.			
<b>Termos semânticos</b>	<b>A</b>	<b>B</b>	<b>C</b>
elemento	água	água	água
relação de restrição	para destinação não potável	para destinação não potável	para destinação não potável
atributo para verificação	reuso de água	reuso de água	reuso de água
indicador de operador deontico	deve atender aos parâmetros estabelecidos	deve atender aos parâmetros estabelecidos na	deve atender aos
relação comparativa	N/A	N/A	parâmetros estabelecidos
valor de quantidade	N/A	N/A	N/A
unidade de medida	N/A	N/A	N/A
restrição de quantidade	na tabela 8	na tabela 8	na tabela 8

## TESTE 5- Esquema de busca com dados simplificados para extrair diversos componentes do texto.

```

compliance_schema = Object(
  id="requisito",
  description="Requisito para construção ou avaliação do edifício ou de seus
componentes",
  attributes=[
    Text(
      id= "elemento construtivo 1",
      description= "O componente construtivo principal da frase."),
    Text(
      id= "elemento construtivo 2",
      description= "Outro componente construtivo da frase."),
    Text(
      id= "elemento construtivo 3",
      description= "Outro componente construtivo da frase."),
    Text(
      id= "elemento construtivo 4",
      description= "Outro componente construtivo da frase."),
    Text(
      id="indicador_de_operador_deontico",
      description="Uma expressão de obrigação ou possibilidade"),
    Text(
      id= "menção à norma",
      description= "citação de uma referência normativa como ABNT 00.000")],
  # Exemplos
  examples=[
    ("A área líquida mínima das aberturas de ventilação não deve ser inferior a
1 m2 para cada 150 m2 de área de manutenção, segundo a ABNT NBR 13.450",
    [{"elemento": "aberturas de ventilacao",
      "elemento construtivo 2": "área de manutenção",

```

```

"elemento construtivo 3": "N/A",
"elemento construtivo 4": "N/A",
"indicador_de_operador_deontico": "não deve ser",
"menção à norma": "ABNT NBR 13.450"}]]],
)

```

## Resultados

<b>Frase 1</b>			
Devem ser previstos nos projetos a prevenção de infiltração da água de chuva e da umidade do solo nas habitações, por meio dos detalhes indicados a seguir: a) condições de implantação dos conjuntos habitacionais, de forma a drenar adequadamente a água de chuva incidente em ruas internas, lotes vizinhos ou mesmo no entorno próximo ao conjunto; b) sistemas que impossibilitem a penetração de líquidos ou umidades de porões e solos, jardins contíguos às fachadas e quaisquer paredes em contato com o solo, ou pelo direcionamento das águas, sem prejuízo da utilização do ambiente e dos sistemas correlatos e sem comprometer a segurança estrutural. No caso de haver sistemas de impermeabilização, estes devem seguir a ABNT NBR 9575; c) sistemas que impossibilitem a penetração de líquidos ou umidades em fundações e pisos em contato com o solo; d) ligação entre os diversos elementos da construção (como paredes e estrutura, telhado e paredes, corpo principal e pisos ou calçadas laterais).			
<b>Termos semânticos</b>	<b>A</b>	<b>B</b>	<b>C</b>
Elemento construtivo 1	prevenção de infiltração da água de chuva e da umidade do solo	prevenção de infiltração da água de chuva e da umidade do solo	prevenção de infiltração da água de chuva e da umidade do solo
Elemento construtivo 2	N/A	N/A	N/A
Elemento construtivo 3	N/A	N/A	N/A
Elemento construtivo 4	N/A	N/A	N/A
Indicador de operador deontico	devem ser previstos	devem ser previstos	devem ser previstos
Menção à norma	ABNT NBR 9575	ABNT NBR 9575	ABNT NBR 9575