

LILIANE SOARES DA COSTA

**ENRIQUECIMENTO SEMÂNTICO DE INFORMAÇÃO GEOGRÁFICA  
VOLUNTÁRIA UTILIZANDO *LINKED DATA* E TESAURO**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, para obtenção do título de *Magister Scientiae*.

VIÇOSA  
MINAS GERAIS – BRASIL  
2018

**Ficha catalográfica preparada pela Biblioteca Central da Universidade  
Federal de Viçosa - Câmpus Viçosa**

T

C837e Costa, Liliane Soares da, 1990-  
2018 Enriquecimento semântico de informação geográfica  
voluntária utilizando Linked Data e Tesauro / Liliane Soares da  
Costa. – Viçosa, MG, 2018.

ix,91f. : il. (algumas color.) ; 29 cm.

Inclui apêndices.

Orientador: Jugurta Lisboa Filho.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f.55-59.

1. Sistemas de Informação geográfica. 2. Tesauro.  
3. Metadados. I. Universidade Federal de Viçosa. Departamento  
de Informática. Programa de Pós-graduação em Ciência da  
Computação. II. Título.

CDD 22. ed 003

LILIANE SOARES DA COSTA

**ENRIQUECIMENTO SEMÂNTICO DE INFORMAÇÃO  
GEOGRÁFICA VOLUNTÁRIA UTILIZANDO LINKED DATA E  
TESAURO**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, para obtenção do título de *Magister Scientiae*.

APROVADA: 07 de fevereiro de 2018.

  
Alcione de Paiva Oliveira

  
Mauro Nacif Rocha

  
Júrgurta Lisboa Filho  
(Orientador)

*À minha família, amigos e professores, por  
lapidarem diariamente o meu ser.*

## AGRADECIMENTOS

Às agências de pesquisa e financiamento Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e Fundação de Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG) por todo suporte financeiro para a realização do mestrado com dedicação exclusiva.

Aos mestres responsáveis por me passar conhecimentos e desafios, especialmente ao orientador Jugurta Lisboa e coorientadora Alexandra por todo o apoio, atenção e conhecimento compartilhado e por estarem sempre disponíveis quando precisei.

À minha mãe (in memoriam) pela educação e amor a mim dedicados em vida, e ainda hoje por sentir sua proteção. Ao meu pai por me ensinar a ser uma pessoa batalhadora assim como ele foi e ainda é.

Aos meus oito irmãos, pela força da união, sempre batalhar e torcer pela vitória uns dos outros. Cada um tem um papel fundamental na minha vida. Agradeço especialmente a minha irmã Lenice, por ter se tornado uma mãe para mim, não medindo esforços para que meus sonhos sejam alcançados; e a minha irmã e melhor amiga Eliane, pelo seu amor, compreensão e companheirismo desde sempre.

Aos meus amigos e colegas de departamento por formamos uma família aqui em Viçosa e termos compartilhado tantos momentos e coisas boas.

Ao Italo, que foi um grande amigo e também orientador em todo o tempo do mestrado, mesmo à distância sempre se pôs disponível a auxiliar quando necessário.

A todos que acreditaram nas minhas conquistas, meu sincero muito obrigada.

Por fim agradeço a Deus, minha força maior, sempre tão generoso comigo. Obrigada por sempre me acolher e propiciar momentos únicos e conquistas tão importantes.

## SUMÁRIO

<b>LISTA DE FIGURAS .....</b>	<b>v</b>
<b>LISTA DE TABELAS .....</b>	<b>vi</b>
<b>RESUMO .....</b>	<b>vii</b>
<b>ABSTRACT .....</b>	<b>ix</b>
<b>1 INTRODUÇÃO .....</b>	<b>1</b>
1.1. Objetivos.....	2
1.2. Organização da dissertação .....	3
<b>2 ARTIGOS.....</b>	<b>4</b>
2.1. Artigo I: An Approach for Semantically Enriching Volunteered Geographic Data with Linked Data.....	5
2.2. Artigo II: Enriquecimento Semântico de Informação Geográfica Voluntária obtida por um Sistema de Combate ao Desperdício de Água .....	18
<b>3 CONCLUSÕES GERAIS E TRABALHOS FUTUROS .....</b>	<b>54</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>56</b>
<b>APÊNDICE A .....</b>	<b>61</b>
Levantamento das palavras da base de dados do sistema Gota D'Água .....	61
<b>APÊNDICE B .....</b>	<b>75</b>
Tesouro de recursos hídricos baseado em termos do Sistema Gota D'Água.....	75

## LISTA DE FIGURAS

### Artigo I

FIGURA 1 - Steps of the method for VGI enrichment with Linked Data.....	14
FIGURA 2 - Examples of buffer shapes .....	15
FIGURA 3 - Identification of contribution location .....	16

### Artigo II

FIGURA 1 - Tripla RDF que descreve a informação "o nome da empresa Petrobras é Petróleo Brasileiro S.A." .....	27
FIGURA 2 - Fluxo do processo proposto de enriquecimento semântico de VGI .....	34
FIGURA 3 - Exemplo de contribuição enriquecida semanticamente pelo algoritmo proposto utilizando como fundo mapa oriundo do OpenStreetMap .....	38
FIGURA 4 - Exemplo de contribuição enriquecida semanticamente pelo algoritmo proposto utilizando como fundo mapa oriundo do Google Maps na interface do sistema Gota D'Água.....	39
FIGURA 5 - Página inicial do Sistema Gota D'Água.....	40
FIGURA 6 - Processo pra retornar contribuições VGI semanticamente semelhantes com base em uma consulta fornecida por um usuário.....	50

## LISTA DE TABELAS

### Artigo II

TABELA 1 - Lugares encontrados em um raio de 100 metros a partir da coordenada geográfica de uma contribuição VGI.....	36
TABELA 2 - Amostra das tuplas retornadas pela consulta Q1.....	41
TABELA 3 - Amostra das tuplas retornadas pela consulta Q2.....	42
TABELA 4 - Amostra das tuplas retornadas pela consulta Q3.....	43
TABELA 5 - Amostra das tuplas retornadas pela consulta Q4.....	44
TABELA 6 - Categorias de matéria.....	46

## RESUMO

COSTA, Liliane Soares da, M.Sc., Universidade Federal de Viçosa, fevereiro de 2018. **Enriquecimento Semântico de Informação Geográfica Voluntária utilizando Linked Data e Tesouros.** Orientador: Jugurta Lisboa Filho. Coorientadora: Alexandra Moreira.

Informação Geográfica Voluntária (VGI) refere-se a um tipo de informação com posicionamento geográfico e que é coletada e/ou compartilhada voluntariamente pelos usuários por meio da Internet. VGI representa uma mudança de paradigma na forma como a informação geográfica é produzida e compartilhada. Porém, ainda há diversas limitações para se utilizar essas informações, como questionamentos sobre a qualidade dos dados, dificuldade de utilização e recuperação, dentre outros. Uma alternativa para melhorar o uso dessas informações é fazer uso do enriquecimento semântico, que pode ser visto como o processo de atribuir maior significado aos dados através da aplicação de recursos auxiliares, com o objetivo de facilitar a compreensão, a integração e o processamento dos dados por pessoas e máquinas. O enriquecimento semântico é uma forma potencial para amenizar alguns problemas relacionados à VGI. Esta pesquisa propõe um método para enriquecimento semântico de VGI utilizando os conceitos de Linked Data e tesouro. O método é composto de duas etapas, uma automática e outra manual. A primeira etapa inclui um algoritmo para automatizar parte do processo, ligando as contribuições VGI à pontos de interesse (PoIs), ou seja, lugares descritos em bases de conhecimento que possam ser de interesse do usuário. Nesta etapa, diversas ferramentas e artefatos são utilizados, como acesso a uma base de Linked Data e um sistema de banco de dados espaciais PostgreSQL+PostGIS com informações sobre PoIs de uma cidade, obtidos diretamente do sistema OpenStreetMap. Além disso, foi utilizado o sistema de banco de dados Virtuoso para armazenar os arquivos RDF gerados pelo algoritmo apresentado. Para verificar a viabilidade do método proposto, foram utilizadas contribuições provenientes de um sistema VGI relacionado com o combate ao desperdício de água, chamado Gota D'Água. Desse modo, na etapa manual do enriquecimento semântico, foi construído um tesouro no domínio de recursos hídricos, com base nos termos encontrados nas contribuições voluntárias. Com base no tesouro, é proposto um processo que retorna contribuições VGI semanticamente semelhantes a partir de uma consulta formulada pelo usuário. Por fim, a viabilidade do método foi verificada por meio

da realização de consultas SPARQL que demonstram análises que somente são possíveis após o processo de enriquecimento semântico.

## ABSTRACT

COSTA, Liliane Soares da, M.Sc., Universidade Federal de Viçosa, February, 2018. **Semantic Enrichment of Voluntary Geographic Information using Linked Data and Thesaurus**. Adviser: Jugurta Lisboa Filho. Co-Adviser: Alexandra Moreira.

Volunteer Geographic Information (VGI) refers to a type of information with geographic positioning that is collected and / or shared voluntarily by users via the Internet. VGI represents a paradigm shift in the way geographic information is produced and shared. However, there are still several limitations using this information, such as questions about data quality, difficulty of use and recovery, among others. An alternative to improving the use of this information is to make use of semantic enrichment, which can be seen as the process of assigning greater meaning to the data through the application of ancillary resources, in order to facilitate the understanding, integration and processing of the data by people and machines. Semantic enrichment is a potential way to alleviate some problems related to VGI. This research proposes a method for semantic enrichment of VGI using the concepts of Linked Data and Thesaurus. The method is composed of two steps, one automatic and one manual. The first step includes an algorithm to automate part of the process, linking the VGI contributions to points of interest (PoIs), that is, places described in knowledge bases that may be of interest to the user. In this step, various tools and artifacts are used, such as access to a Linked Data database and a PostgreSQL + PostGIS spatial database system with information about a city's PoI, obtained directly from the OpenStreetMap system. In addition, the Virtuoso database system was used to store the RDF files generated by the presented algorithm. In order to verify the feasibility of the proposed method, contributions were made from a VGI system related water waste prevention, called Gota D'Água. Thus, in the manual stage of semantic enrichment, a thesaurus was constructed in the domain of water resources, based on the terms found in the voluntary contributions. Based on the thesaurus, a process is proposed that returns semantically similar VGI contributions from a query formulated by the user. Finally, the feasibility of the method was verified through SPARQL queries that demonstrate analyzes that are only possible after the process of semantic enrichment.

# 1 INTRODUÇÃO

Devido ao enorme volume de dados e informações disponibilizados e consumidos diariamente, a World Wide Web vem evoluindo para atender as mais diversas demandas dos usuários. Essa evolução permite que a participação do usuário seja mais simples e efetiva, tornando mais fácil a disseminação e visibilidade dos conteúdos (PALETTA, 2017).

Considerando a maior interação do usuário na Web, surge o conceito de Informação Geográfica Voluntária (*Volunteered Geographic Information – VGI*) que, segundo Goodchild (2007a), diz respeito a dados que possuem alguma característica espacial e que são gerados a partir de algum tipo de contribuição voluntária. Informação produzida voluntariamente possui vantagens como: ser gratuita; poder ser obtida rapidamente; e gerar dados que ainda não foram disponibilizados por agências oficiais (GOODCHILD; LI, 2012).

No entanto, ainda existe a necessidade de se descobrir, acessar e utilizar mais facilmente as contribuições VGI existentes na Web. Uma das formas de amenizar tais necessidades é adicionando semântica às contribuições VGI. Mas, se tratando de estruturação semântica, contribuições VGI geralmente não apresentam semântica suficientemente formal para ser utilizada computacionalmente e detalhada o suficiente para suprir as necessidades das aplicações.

Paralelamente, nos últimos anos, um número crescente de provedores de dados vem adotando um conjunto de tecnologias e boas práticas da Web Semântica para publicar e interligar dados estruturados na Web, formando assim a Web de Dados. Os princípios determinados pela abordagem Linked Data e adotados pela Web de Dados oferecem simplicidade e flexibilidade para que dados sejam representados, interligados e possam interoperar na Web. Estes dados são estruturados por meio de fatos expressos como triplas (sujeito-predicado-objeto), seguindo o padrão *Resource Description Framework* (RDF), que contém identificadores no formato *Uniform Resource Identifier* (URI) e descritores definidos por vocabulários (BIZER, HEATH, BERNERS-LEE, 2009).

Os relacionamentos semânticos expressos em Linked Data permitem que os dados se relacionem entre si, facilitando sua descoberta entre diferentes repositórios de dados, além de permitir que pesquisas semânticas sejam realizadas a partir desses dados (SCHADE; GRANELL; DÍAZ, 2010).

Devido à utilização de relacionamentos semânticos para relacionar os dados e assim melhorar seu compartilhamento, o conceito de Linked Data tem sido sugerido como uma abordagem no enriquecimento semântico de VGI (BALLATORE,2011; STADLER,2012; RONZHIN, 2015).

Uma maneira de se inserir semântica nos dados é adicioná-la de forma manual. Entretanto, esse método não é aconselhável, por consumir um tempo considerável e permitir que o publicador ou analista cometa erros durante o processo. Por esses motivos, existe a necessidade de métodos automáticos para auxiliar o processo de anotação. Adicionalmente, é interessante promover um nível de transparência para o usuário abstraindo toda a complexidade das anotações na geração de metadados enriquecidos.

Portanto, este trabalho propõe a utilização de abordagens automáticas e manuais de enriquecimento semântico de contribuições VGI, visando a realização de análises semânticas sobre bases de VGI.

Como caso de estudo, o processo de enriquecimento semântico automático de dados VGI, as análises semânticas e a verificação da confiabilidade serão implementados e testados no sistema Gota D'Água. Este sistema é fruto de um projeto de pesquisa do Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Viçosa (UFV), baseado em VGI, e visa coletar dados do usuário a respeito do desperdício e da falta de água.

## **1.1. Objetivos**

O objetivo geral desta pesquisa é enriquecer semanticamente contribuições VGI utilizando *Linked Data* e tesauro. Especificamente, os objetivos desta pesquisa são:

- a) Relacionar problemas identificados por voluntários aos locais afetados, sendo que locais e contribuições serão expressos em *Linked Data*;
- b) Propor e realizar análises semânticas de modo a demonstrar a aplicabilidade do enriquecimento semântico de dados VGI no processo de tomada de decisões;
- c) Utilizar um tesauro com base no conjunto de palavras contidas nas contribuições voluntárias para trazer mais conteúdo informacional;
- d) Aplicar o método propostos (*Linked Data* e tesauro) no sistema Gota D'Água, o qual foi selecionado como caso de estudo.

## **1.2. Organização da dissertação**

Esta dissertação foi elaborada de acordo com um dos formatos recomendados pela Comissão do Programa de Pós-Graduação em Ciência da Computação da UFV. A dissertação está organizada como uma coletânea de artigos produzidos resultantes da pesquisa. Dentre os artigos, o primeiro foi publicado em uma conferência e o segundo será submetido à avaliação para publicação em uma revista científica.

A dissertação está organizada da seguinte forma:

O Capítulo 2 é composto de dois artigos resultantes da pesquisa. O Artigo I (Seção 2.1) propõe um algoritmo para o enriquecimento semântico de VGI com base em sua localização espacial. O Artigo II (Seção 2.2) apresenta o método completo, envolvendo uma etapa automática e outra manual, que possibilita o enriquecimento semântico de VGI também com base em termos descritos em um tesauro e os resultados alcançados.

No Capítulo 3 são apresentadas as conclusões gerais, discutindo os resultados obtidos e avanços alcançados no Artigo II. Ainda nesse capítulo, algumas oportunidades de pesquisa em aberto são apresentadas.

O Apêndice A inclui o levantamento de termos feito na base de dados do Sistema Gota D'Água. Estes termos são apresentados em listas com diferentes categorias. Por último, o Apêndice B, apresenta o recorte do tesauro sobre recursos hídricos.

## 2 ARTIGOS

Este capítulo apresenta dois artigos resultantes da pesquisa desta dissertação. O primeiro artigo (Seção 2.1), intitulado “*An Approach for Semantically Enriching Volunteered Geographic Data with Linked Data*” apresenta a proposta de enriquecimento semântico de dados voluntários geográficos. Este artigo foi publicado na *International Conference on Computational Science and its Applications (ICCSA 2017)*, ocorrida em Trieste, Itália.

No segundo artigo (Seção 2.2), intitulado “*Enriquecimento Semântico de Informação Geográfica Voluntária obtida por um Sistema de Combate ao Desperdício de Água*” é apresentado o método de enriquecimento semântico de dados voluntários geográficos, os resultados obtidos e a discussão dos mesmos. Este artigo será submetido a uma revista científica da área, ainda não definida, após sua apresentação à banca de defesa desta dissertação.

## 2.1. Artigo I: An Approach for Semantically Enriching Volunteered Geographic Data with Linked Data

Liliane Soares da Costa, Italo Lopes Oliveira, Alexandra Moreira, Jugurta Lisboa-Filho

In: INT. CONFERENCE ON COMPUTACIONAL SCIENCE AND ITS APPLICATIONS, 2, 2017, Trieste, Italy. **Proceedings...** Cham: Springer LNCS vol 10406, Part 3 – Lecture Notes in Computer Science book series, 2017. pp. 296-306

### RESUMO

A informação geográfica voluntária, que se refere às informações geográficas coletadas e compartilhadas voluntariamente, representa uma mudança de paradigma na forma como a informação geográfica é criada e compartilhada. Porém, ainda há dificuldades para utilizar essas informações de forma adequada. Uma alternativa para melhorar o uso dessas informações é fazer uso do enriquecimento semântico. O enriquecimento semântico é uma forma potencial para amenizar diversos problemas que assolam dados VGI, como baixa qualidade dos dados, falta de confiabilidade, dificuldade de utilização e recuperação, entre outros. Este trabalho discute a possibilidade do enriquecimento semântico de dados geográficos voluntários utilizando Linked Data e apresenta um algoritmo simplificado para automatizar tal processo.

**Palavras-chave:** Informação Geográfica Voluntária, Enriquecimento semântico, Linked Data.

### ABSTRACT

Volunteered geographic information (VGI), which pertains to geographic information voluntarily collected and shared, represents a paradigm shift in the way geographic information is created and shared. However, there are still hurdles in properly using such information. One alternative to improve the use of VGI is to use semantic enrichment. Semantic enrichment is a potential way of mitigating several issues that plague VGI such as low data quality, unreliability, and difficulty in use and recovery, among others. The present study discusses the possibility of semantically enriching volunteered geographic data using Linked Data and presents a simplified algorithm to automate this process.

**Keywords:** Volunteered geographic information, Semantic enrichment, Linked Data.

## 1 INTRODUCTION

A huge amount of data and information is made available and consumed daily on the Internet. The evolution of web search engines has been revolutionizing the way these data are discovered, accessed, integrated, and used (HEATH; BIZER, 2011).

The convergence of interactive technologies via web with the growing amount of user-generated content on the internet has spawned a new way of producing geographic information. Citizens are using mobile devices to collect geographic information using web-based mapping interfaces to tag and annotate geographic characteristics or, for example, add geotag photographs. Such actions originated the term Volunteered Geographic Information (VGI), representing a paradigm shift in the way geographic information is created and shared (ELWOOD; GOODCHILD; SUI, 2012). However, just like conventionally generated geographic data (e.g., by companies or governmental agencies), properly discovering and using VGI still poses some challenges such as, for example, ambiguity in terminology. One of the ways to mitigate these issues is by adding semantics to VGI using Linked Data.

Semantic enrichment can be seen as the process of attributing greater meaning to metadata and data through the application of ancillary resources, in order to facilitate the understanding, integration and processing of data by people and machines. In other words, the semantic enrichment makes the data and metadata more qualified through the use of semantic attributed by pre-existing vocabularies, synonyms and provenance information (CLARKE, 2009).

The Linked Data concept arose to aid in the discovery, access, and use of online data. Linked Data are semi-structured data that allow specifying semantic relationships among themselves. Using such concept, data on the internet become semi-structured nodes of a semantic network, thus facilitating their use by systems and helping them be discovered and accessed (HEATH; BIZER, 2011).

The semantic relationships expressed by Linked Data allow data to be related to each other, which makes it easier for them to be discovered across different data repositories, besides allowing semantic searches to be performed from these data (SCHADE; GRANELL; DÍAZ, 2010).

Given the use of semantic relationships to relate data and, therefore, improve sharing, the Linked Data concept has been suggested as an approach in semantic

enrichment of VGI (BALLATORE; BERTOLOTTI, (2011); STADLER, (2012); RONZHIN, (2015)).

Nonetheless, manually adding semantics to VGI is a costly and tedious task. Automated semantic enrichment approaches allow VGI to be semantically enriched as needed by the application domain, besides allowing for a more thorough description of user-generated data. Requiring users to describe the semantics and semantic relationships of previously produced volunteered geographic data is impractical either due to the lack of user knowledge on how to properly specify data semantics or due to the burden of this task, which would prevent users from producing new geographic data.

This paper presents an approach that uses spatial operations to automate VGI semantic enrichment, which enables semantic analyses and, consequently, improves data discovery. Moreover, adding semantics to VGI allows finding or solving inconsistencies and ambiguities, thus improving the quality of user-generated data.

The remaining of the paper is structured as follows. Section 2 describes the basic concepts involved in the research. Section 3 presents some related works. Section 4 presents the work proposal, while Section 5 makes some final considerations.

## **2 THEORETICAL FRAMEWORK**

### **2.1. Volunteered Geographic Information**

With the use of different types of technologies and tools, people are able to widely collect, publish, and make available data on different subjects and in different ways. This way, the amount of geospatial data generated is on the rise.

The advances associated with Web 2.0, particularly in the context of web mapping technologies, have driven innovation in online geographic information collection, sharing, and interaction. That makes way for VGI, which concerns geographic information voluntarily collected and shared by the general public (GOODCHILD, 2007).

The marked increase in VGI contributions leads to several new platforms and projects that use data and technologies in spatial decision-making, participatory planning, and citizen science. The potential use of VGI has been proven for urban management, damage from floods, fires, and earthquakes, and other important cases of risk, crisis, and natural disaster management (NEIS; ZIELSTRA, 2014).

Volunteered information may also be provided anonymously. This way, not all VGI contributions are altruistic or unbiased. Thus, according to Cooper et al. (2011), the

risks of using low-quality VGI are the same when the data come from an official provider since the source of the data does not impact the results of their use. The main difference may be that an official agency or commercial vendor could be considered legally liable for the data, although, in practice, that hardly ever happens because of exemption of liability.

To be able to contribute to any VGI platforms, the information has to correspond to a geographical position, being often necessary connection with Internet, smartphone or personal computer. VGI is sometimes the cheapest and often the only source of geographic data, particularly in areas where access to geographic information is considered a national security issue (NEIS; ZIELSTRA, 2014).

VGI may be found in simple ways, such as information on Wikipedia, which also provides spatial information in some cases. Other examples of projects involving volunteered geographic contributions are OpenStreetMap<sup>1</sup> (OSM) and Wikimapia<sup>2</sup>.

OSM is known as one of the most complex and promising VGI projects. It allows users to download geospatial data free of cost and use them in personal projects (ZIELSTRA; ZIPF, 2010). Wikimapia allows users to use polygons to indicate points of interest (POI) in maps. It focuses mainly on describing places (MUMMIDI; KRUMM, 2008). Projects such as OSM feature tools able to capture, display, produce, and spread information on a larger scale, besides enabling experiences for several purposes.

### **5.1. Linked Data**

Linked Data refers to a style of publishing and interlinking semantic data on the web. The data are made available online in a way that they are easily processed by machines and their meaning/semantics are defined explicitly. Moreover, these data are linked bidirectionally to other external datasets (BIZER; HEATH; BERNERS-LEE, 2009).

The goal is to allow people to share data with well-defined semantic on the web as easily as documents are currently shared. Therefore, it targets massively sharing and reusing information in a global data space, besides allowing new data to be discovered

---

<sup>1</sup> <http://www.openstreetmap.org>

<sup>2</sup> <http://www.wikimapia.org>

(AZEVEDO, 2014). The more a piece of data is interlinked to other data, the greater its value and usefulness.

One of the most commonly employed formats for linked data structuring is Resource Description Framework (RDF<sup>3</sup>). RDF is a generic data model based on graph structure that describes things in the world. The RDF model codes data as triplets: **subject, predicate, object**. The **subject** and **object** of a triplet are both Uniform Resource Identifiers (URI) that identify each resource, while the object may take on the value of a string. The **predicate** specifies how the **subject** and **object** are related and is also represented by a URI of a resource that describes such relationship (BIZER; HEATH; BERNERS-LEE, 2009).

Some Linked Data principles were introduced by Berners-Lee et al. (2006), namely: Using URIs as resource names, use URIs for people, fruits, cars, but also for abstract concepts as love, war, statistics; using HTTP URIs so that people are able to find those names, whenever there is an HTTP GET request, something must be returned, in this case RDF; when someone searches for an URI, ensure that useful information can be obtained through those URIs, which must be represented in RDF format; and including links to other URIs so that other resources can be discovered, these links are RDF properties.

According to Berners-Lee et al. (2009), similarly to geographic data, Linked Data must be published with several metadata in order to aid in their discovery and use. It is suggested by Hartig (2009) that metadata related to Linked Data must contain, at least, the date of creation and method of creation.

The most well-known Linked Data example is project DBpedia<sup>4</sup>, which is a semantic wiki. Wikipedia provides its information in a searchable format, usually RDF. Such queries may be performed using the SPARQL<sup>5</sup> language.

### 3 RELATED WORKS

According to Ronzhin (2015), integrating a VGI set to the Linked Open Data (LOD) cloud provides advantages beyond only resources that are directly interlinked with the VGI set. That is because of the great interconnectivity between the set of data published as linked data.

---

<sup>3</sup> <https://www.w3.org/RDF/>

<sup>4</sup> <http://wiki.dbpedia.org/>

<sup>5</sup> <http://www.sparql.org/>

The study Ronzhin (2015) seeks to investigate to what extent the LOD cloud could help semantically enrich VGI in order to obtain better search results in the context of operations and crisis. Based on that, it is said that the use of URIs of an open knowledge database such as LinkedGeoData<sup>6</sup> (LGD) eliminates the possibility of ambiguity when indicating a place. The additional semantics obtained from structured information presented in LGD enables, for example, quick access to basic and useful knowledge on infrastructure objects and public buildings, which could be used to speed up and improve the decision-making process. Therefore, integrating VGI with relevant entities on the LOD cloud makes it possible to semantically enrich unstructured user-generated content with structured information presented in LOD.

Using the Linked Data concept, Azevedo (2014) proposes a solution able to obtain heterogeneous data related to floods in the Rio Doce Basin (water basin located in southeast Brazil) from several public organizations by integrating and providing such data for visualization in a Geographic Information System (GIS). To that end, that author proposes converting data to RDF format, interlinking them, and visualizing them using SPARQL queries. The study employed experimental, proof-of-concept methodology and the Rio Doce Basin was the analysis unit.

In the context of floods, visualizing, interacting with, and publicizing these data are the key for effective disaster management. In that realm, the Linked Data principles are a way of providing information shared on the web by offering a dataset on different subjects and sources and by making connections.

VGI has been successfully utilized in scenarios such as emergency response and is also increasingly integrated into commercial products. Based on an analysis of existing projects and research, Savelyev et al. (2011) propose to extend the idea of VGI by introducing Volunteered Geographic Services (VGS). Instead of contributing information, volunteers can request or offer microservices to their local community. Therefore, it provides a flexible server framework that handles service requests and offers, and also implements a smartphone application developed using Google's Android platform. The server and mobile client are realized following the Linked Data paradigm and using Semantic Web technologies. In this paper, it discusses the idea behind VGS, motivate it using two scenarios, and explain the technical realization. It can be concluded,

---

<sup>6</sup> <http://linkedgeo.org/>

the data produced by VGS users may provide a rich source for spatial and temporal data analysis as well as pattern mining.

Semantic enrichment derived from knowledge on the relationships among geospatial data is a potential source of solutions to traditional issues when recovering geographic information (MOURA, 2013). This way, the study of Moura (2013) discusses the applicability of Linked Data concepts to several issues, such as resolving ambiguity and recognizing the spatial context of documents. In addition, the study presents a list of challenges and opportunities for research on the recovery of geographic information and correlated topics.

#### **4 A METHOD FOR VGI ENRICHMENT WITH LINKED DATA**

The present study is based on the hypothesis that semantic enrichment of VGI enables automatically adding information with no need for user input. The analyses on semantically enriched VGI may improve the decision-making process. Furthermore, the use of semantic relationships generated in such enrichment may help verify the reliability of VGI.

As a study case, the automated VGI semantic enrichment methods, the semantic analyses, and the reliability verification are implemented and tested in a VGI collection system called *Gota D'Água*<sup>7</sup> (Water Drop). This system was developed aiming to collect data through volunteered citizen contributions regarding water waste and shortages, a recurring issue in several regions of Brazil during the dry season.

The main goal of this research is to automate semantic enrichment of volunteered geographic data using Linked Data. The other goals are: Link issues identified by volunteers to the places affected or to user profiles, already converted into Linked Data; propose semantic analyses so as to demonstrate the usefulness of semantic enrichment of VGI; perform the semantic analyses proposed using the SPARQL query language; propose metrics to verify the reliability of semantically enriched VGI; verify whether semantic relationships via Linked Data allow attesting VGI reliability; and applying the approaches proposed in a VGI application.

---

<sup>7</sup> <http://www.gotadaguaufv.com.br/>

The VGI dataset obtained by the *Gota D'Água* system was transformed into Linked Data using tools such as Triplify<sup>8</sup> and -ontop<sup>9</sup>. Those tools transform existing data in a relational database into RDF data.

The POI to be used for the analyses (for example, verifying which places will be impacted by a leak in the water distribution network) are obtained from the LinkedGeoData (LGD) repository. That repository uses information collected by the OpenStreetMap project and makes it available as an RDF knowledge base according to the Linked Data principles. LGD interconnects those data with other knowledge bases of the Linking Open Data (LOD) initiative, such as DBpedia.

Two database management systems (DBMS) are used to store the data: PostgreSQL with the PostGIS extension and the open-source version of Virtuoso<sup>10</sup>. All data (Linked Data), i.e., data in the RDF format, are stored in the Virtuoso database, which allows them to be queried using the semantic query language SPARQL. PostgreSQL, with the PostGIS extension, is used to perform geospatial operations. Although there are DBMS that support spatial queries on RDF data (Strabon and Oracle, for example), geographic operations on RDF data are still less efficient than geographic queries performed in relational databases, which justifies the use of PostgreSQL+PostGIS.

With the data collected by the *Gota D'Água* system and the POI obtained using LinkedGeoData, analyses are proposed to take advantage of the semantic network provided by Linked Data, showing, for instance, how a given type of issue in water distribution may impact places of a certain type (e.g., stores, hospitals). Moreover, the semantic analysis of data is used to verify the levels of reliability of data added to the *Gota D'Água* system. Reliability is measured based on the number of semantic relations an issue or user profile has. The larger the number of semantic relationships (e.g., the same issue has been reported by different users), the higher the likelihood of the data being reliable.

Figure 1 presents the flow of steps (outer boxes) in which the method is applied. As previously described, each step uses the tools needed (inner boxes) for its execution.

---

<sup>8</sup> <http://semanticweb.org/wiki/Triplify.html>

<sup>9</sup> <http://ontop.inf.unibz.it/>

<sup>10</sup> <http://virtuoso.openlinksw.com/>

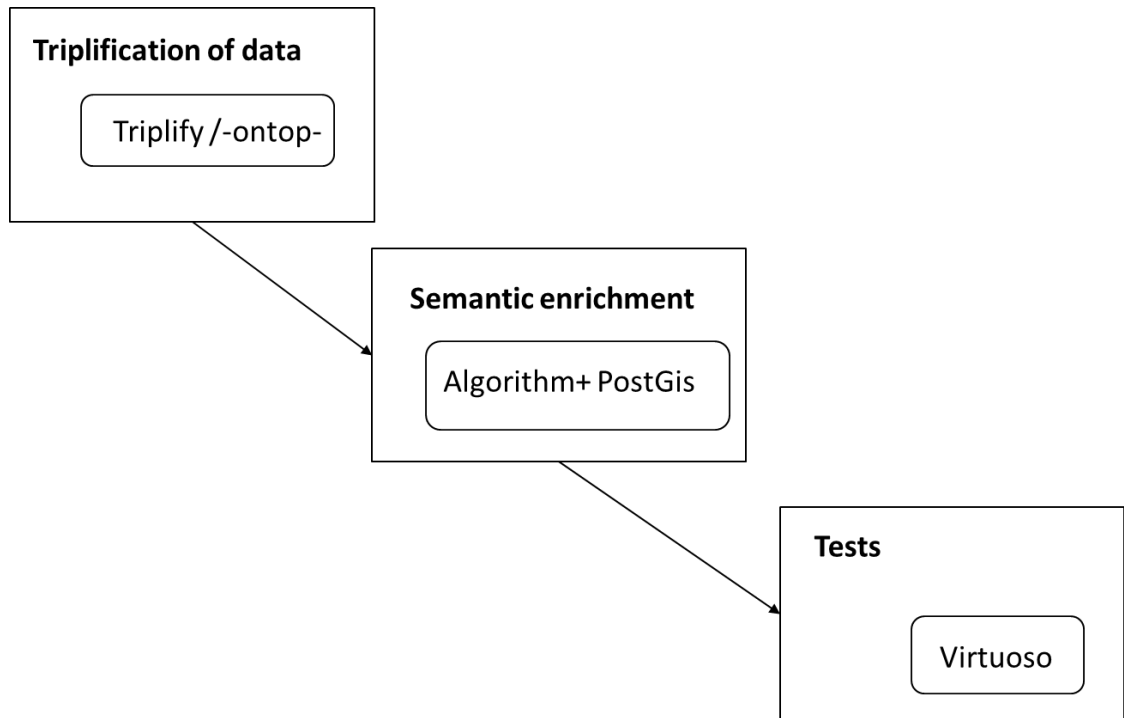


Figure 1. Steps of the method for VGI enrichment with Linked Data.

Algorithm 1 presents the method proposed for semantically enriching VGI.

ALGORITHM 1 Automatic Enrichment of VGI

**Require:**

$V$  // VGI set

$P$  // Knowledge base with semantically well-described places with geographic coordinates

$R$  = empty set of rdf files;

**for each**  $v \in V$  **do**

// Creates buffer around  $v$  in meters

$b = \text{CreateBuffer}(v, v.\text{sizeofBuffer}());$

**for each** ( $p \in P$ ) inside  $b$  **do**

$r.\text{AddAnnotation}(v, p);$

**end for**

$R.\text{add}(r);$

**end for**

**return**  $R;$

As input, the algorithm receives a VGI set  $V$ , a knowledge base containing the places to be used to enrich VGI set  $P$ , and an empty set of RDF files  $R$ , which will receive the annotations generated by the algorithm. For each piece of data  $v \in V$ , not semantically enriched yet, a circular buffer  $b$  is generated around the geographic position of  $v$ , and the

buffer radius is determined by the type of issue specified in  $v$ . For example, in case the issue of VGI  $v$  is a leak, then the buffer radius will be  $x$ , whereas the issue in VGI  $v'$  is water shortage, thus the buffer radius of  $v'$  will be  $x'$ . After the creation of the buffer, it is verified which places  $p \in P$  are geographically within the buffer. For each of those places, a semantic annotation  $r$  is created, whose subject will be the URI for VGI  $v$  and whose object is the URI of place  $p$ . The predicate value will change depending on the type of issue described in  $v$ . After all places  $p$  are annotated to  $v$ , an annotation  $r$  is added to set  $R$ . Finally, set  $R$  is returned.

The algorithm described considers the generation of a buffer around the issue submitted to the system. The buffer may have different shapes and sizes and may be adapted according to the issue. Figure 2 illustrates different buffer shapes.

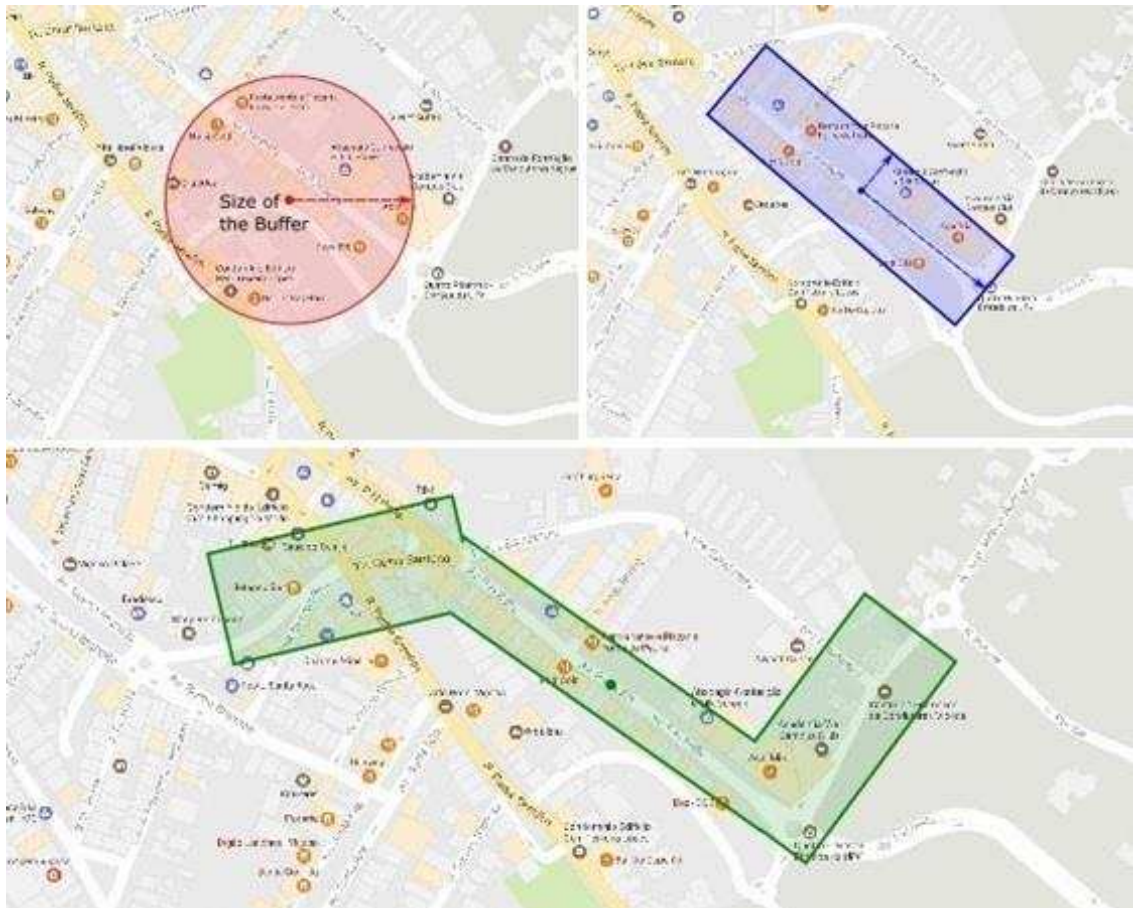


Figure 2. Examples of buffer shapes.

## 5 DISCUSSION

The *Gota D'Água* system, used in this work to obtain VGI, enables users from different parts of Brazil contribute by providing information on different types of

problems related to lack and/or waste of water. When making the contribution, the place indicated by the user is identified by its geographical coordinates, and it is highlighted in the map for visualization, as seen in Figure 3. In addition, users can choose the type of problem (e.g., leak, waste) and provide more detailed information, in form of text, image or video.

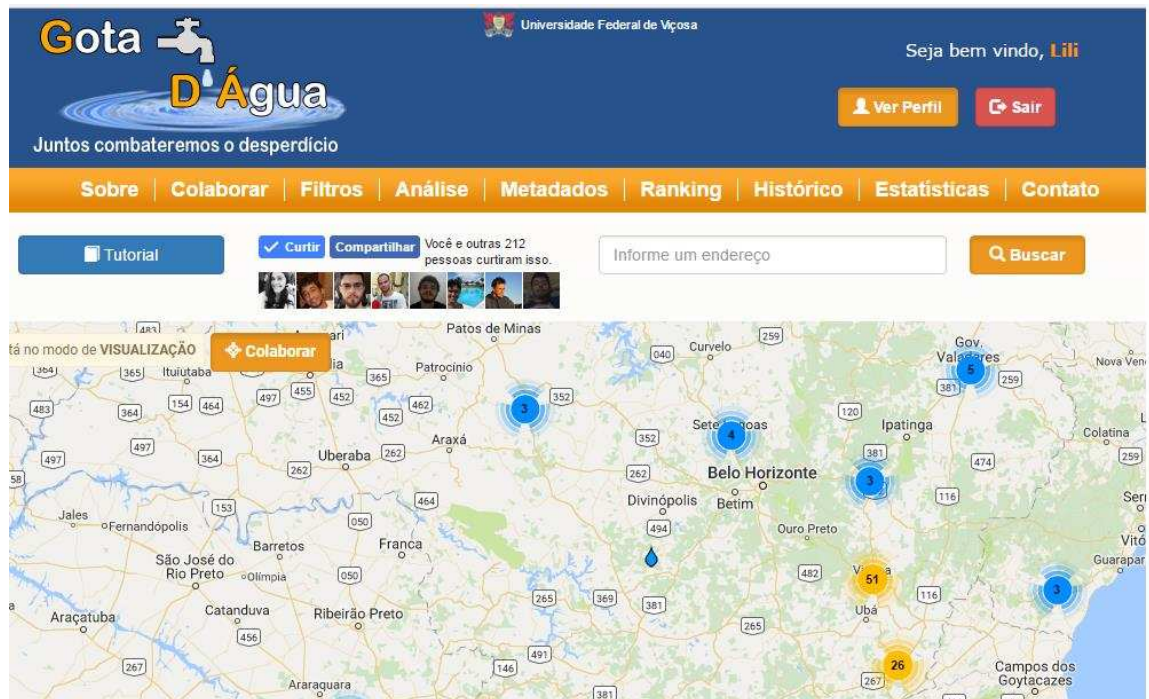


Figure 3. Identification of contribution location

In the context of this work, a VGI has similarities to microblogging data, such as Twitter. Usually, both have low context attached to them. This happens due to the fact that users use small messages, either by limiting the system (e.g., Twitter limits posts to 140 characters), or because the user is very succinct in describing a problem, as occurs in the contributions of the *Gota D'Água* system.

However, although the descriptions of the user contributions in the *Gota D'Água* system are not detailed, in the use of this approach it is expected to demonstrate the semantic enrichment in the VGI starting from the triple coding in RDF.

The generation of RDF files from *Gota D'Água* VGI makes possible to ask questions to be answered by the system through the SPARQL language. Queries in SPARQL become easier than expressing the same query in SQL. In addition to the advantage of using the SPARQL language, a query can be used to formulate complex queries involving several distributed RDF repositories on the Web (W3C, 2013).

The queries to be executed will answer the questions previously proposed after the validation of the semantic enrichment method, and the data will be transcribed in the search results.

A small number of contributions examined will be annotated manually, generating a set of ground truth. The proposed method should generate annotations similar to the set of ground truths, and a certain level of imprecision is expected.

The annotations associated with the described event are the basis for determining the radius of the buffer associated with the VGI. Its limits are delimited at the moment and arbitrarily, but will be used to demonstrate the application of the technique. Its size should be better adjusted to get closer to reality in the final implementation of the system.

The greatest difficulty of the research is to deal with brief description that the VGI contributions have. For example, in the *Gota D'água* system many contributions do not have any type of comments linked to them, based only on the type, geographical location and timestamp of the contribution. This lack of detail makes it impossible to use tools like Named Entity Recognition (NER) / Named Entity Disambiguation (NED) (DERCZYNSKI et al., (2015); HAN; ZHAO, (2009)) or other semantic enrichment methods that could generate new semantic annotations and, consequently, contribute to the results of the queries to be executed.

## **6 FINAL CONSIDERATIONS**

Linked Data is a concept used to integrate data by simplifying the relationship schemas and enriching semantics. Although its use seems advantageous, the technology is new and several challenges must be overcome and areas of application have to be discovered (MOURA, 2013). This paper describes the initial results of a research project whose final goal is to verify the applicability of Linked Data to semantic enrichment of VGI.

The importance of volunteered data is on the rise; however, such data often lack descriptive information. This research aims to add more information – namely, Linked Open Data (LOD) – to VGI in an automated fashion. The research results in a VGI set with annotations that point to LOD, which enables semantic and/or more complex queries to be performed on volunteered data. An example of such query is verifying which types of services will be affected by an issue/disaster.

Although the work proposal is a very simple algorithm, this simplicity can be considered as an advantage. Such simplicity indicates that the algorithm proposed can be easily implemented by any organization or people that want to enrich semantically VGI. Moreover, the functions used in the algorithm (like the buffer creation) are already implemented in several databases, like PostgreSQL (with the extension PostGIS) and OracleDB.

As future works, a more complex algorithm can be proposed to generate more semantic annotation for VGI than the algorithm proposed in this work. To produce more annotations, semantic annotations tasks, like Named Entity Recognition (NER), Named Entity Disambiguation (NED), also called Entity Linking (EL), and Word Sense Disambiguation (WSD) will be integrated in the algorithm. In addition, social media posts can be considered as VGI, like the use of Twitter posts (tweets) in emergency management/communication (WHITE, (2011); GAO; BARBIER; GOOLSBY, (2011)). Although there are several works that enrich semantically social media posts using EL/NED (YAMADA; TAKEDA; TAKEFUJI, (2015), (GUO; CHANG; KICIMAN, (2013)), the use of semantically enriched tweets in VGI systems has not properly studied yet.

## **ACKNOWLEDGEMENTS**

This project was partially funded with resources from the agencies FAPEMIG and CAPES, with the support of the *Companhia Energética de Minas Gerais* (CEMIG).

## 2.2. Artigo II: Enriquecimento Semântico de Informação Geográfica Voluntária obtida por um Sistema de Combate ao Desperdício de Água

Liliane Soares da Costa, Italo Lopes Oliveira, Alexandra Moreira, Jugurta Lisboa-Filho

### RESUMO

Informação Geográfica Voluntária (VGI) refere-se a um tipo de informação com geolocalização, que é coletada e/ou compartilhada voluntariamente por meio da Internet. VGI representa uma mudança de paradigma na forma como a informação geográfica é produzida e compartilhada. Porém, ainda há diversas limitações para se utilizar essas informações de maneira adequada, como questionamentos sobre a qualidade dos dados, dificuldade de utilização e recuperação, dentre outros. Uma alternativa para melhorar o uso dessas informações é fazer uso do enriquecimento semântico, que pode ser visto como o processo de atribuir maior significado aos metadados e dados através da aplicação de recursos auxiliares, com o objetivo de facilitar a compreensão, a integração e o processamento dos dados por pessoas e máquinas. O enriquecimento semântico é uma forma potencial para amenizar alguns problemas relacionados à VGI. Este trabalho propõe um método para enriquecimento semântico de VGI utilizando os conceitos de *Linked Data* e tesouros. O método é composto de duas etapas, uma automática e outra manual. A primeira etapa inclui um algoritmo para automatizar parte do processo, ligando as contribuições VGI à lugares descritos em bases de conhecimento que são de interesse do usuário. Nesta etapa, diversas ferramentas e artefatos são utilizados, como acesso a uma base de *Linked Data*, um banco de dados espaciais com informações sobre pontos de interesse em uma cidade obtidos diretamente do sistema *OpenStreetMap*. Para verificar a viabilidade do método proposto, foram utilizadas contribuições provenientes de um sistema VGI relacionado com o combate ao desperdício de água, chamado Gota D'Água. Desse modo, na etapa manual do enriquecimento semântico, foi construído um tesouro no domínio de recursos hídricos, com base nos termos encontrados nas contribuições voluntárias. Por fim, com base no tesouro, é proposto um processo que retorna contribuições VGI semanticamente semelhantes a partir de uma consulta formulada pelo usuário.

**Palavras-chave:** Informação Geográfica Voluntária, Enriquecimento semântico, *Linked Data*, Tesouro.

### ABSTRACT

Volunteered Geographical Information (VGI) refers to a type of information with geolocation, that is collected and/or shared voluntarily by the internet. VGI represents a shift of paradigm in the way geographic information is produced and shared. Yet, there still are many limitations for using those informations in the adequate way, such as questionings about quality of data, difficulty to use and recovery, among others. One alternative to improve usage of those informations is to use semantic enrichment, that can

be seen as the process of assigning greater meaning to metadata and data through auxiliary resource application, aiming to ease comprehension, integration and data processing by people and machines. Semantic enrichment is a potential way to soften some VGI relative problems. This work proposes a VGI semantic enrichment method using Linked Data and thesaurus concepts. The method has two stages, one automatic and one manual. First stage includes an algorithm to automate part of the process, linking VGI contributions to places described in data bases that are of interest of the user. In this stage, many tools and artifacts are utilized, such as access to a Linked Data base, an spatial data bank with information about points of interest of a city, obtained directly from the OpenStreetMap system. To verify the proposed method viability, contributions from the VGI system Gota D'Água were used, related to prevent water waste. That way, in the manual stage of semantic enrichment, a thesaurus in the hydric resources domain was built, based in terms found in volunteered contributions. Finally, based in the thesaurus, a process is proposed, that returns VGI contributions semantically similar, based on a query made by the user.

**Keywords:** Volunteered geographic information, Semantic enrichment, Linked Data, Thesaurus.

## 1 INTRODUÇÃO

O desenvolvimento da Web ocasionou maior disseminação e visibilidade dos conteúdos. Porém, o rápido crescimento na quantidade de dados disponíveis tornou a Web Semântica uma necessidade e, para viabilizar sua implantação, novas tecnologias foram desenvolvidas (e.g. *Linked Data*) e tecnologias existentes foram aplicadas de modo mais eficiente (e.g., tesauros, ontologias, bases semânticas). Dessa forma, a evolução dos mecanismos de busca e pesquisa na Web vem inovando a maneira com que os dados são descobertos, acessados, integrados e utilizados (HEATH; BIZER, 2011).

Com base no crescimento dos dados disponíveis, as tecnologias da Web Semântica têm como objetivo proporcionar um espaço de informação semântica compartilhada, mudando qualitativamente as experiências na Web (GOODCHILD, 2007a). Com o auxílio de tais tecnologias, a participação dos usuários na Internet se tornou mais ativa, possibilitando a criação de conteúdo e, adaptação e desenvolvimento de aplicações para campos diversos. Um dos campos beneficiados pela maior participação do usuário na Internet é o da informação geográfica.

A convergência de novas tecnologias interativas via Web com a crescente prática de conteúdo gerado por usuário na Internet deu origem a uma nova forma de produção de informação geográfica. Os cidadãos (usuários) estão utilizando dispositivos móveis para coletar informações geográficas, usando interfaces de mapeamento baseado na Web para marcar e anotar características geográficas como, por exemplo, adicionar localização geográfica nas fotografias. Estas ações deram origem ao termo Informação Geográfica Voluntária (*Volunteered Geographic Information - VGI*), representando uma mudança de paradigma na forma como a informação geográfica é criada e compartilhada (ELWOOD; GOODCHILD; SUI, 2012). Entretanto, assim como dados geográficos gerados de maneira convencional (e.g., por empresas ou agências governamentais), descobrir e usar de maneira adequada os dados VGI ainda impõe vários desafios como, por exemplo, ambiguidade nos termos empregados pelo usuário e precisão das coordenadas geográficas. Uma das formas de amenizar tais problemas é adicionar semântica à VGI utilizando o conceito de dados ligados (em inglês, *linked data*).

O conceito de *Linked Data* surgiu para auxiliar a descoberta, acesso e utilização dos dados presentes na Internet. *Linked Data* são dados semi-estruturados que permitem a especificação de relacionamentos semânticos com outros dados que seguem o conceito *Linked Data*. Utilizando-se de tal conceito, os dados presentes na Internet se tornam nodos

semi-estruturados de uma rede semântica, facilitando sua utilização por sistemas e auxiliando sua descoberta e acesso (HEATH; BIZER, 2011). Os relacionamentos expressos na rede semântica formada por *Linked Data* permitem facilitar sua descoberta entre diferentes repositórios de dados, além de permitir que pesquisas semânticas sejam realizadas (CLARKE, 2009).

Devido à utilização de relacionamentos semânticos para associar os dados e assim melhorar seu compartilhamento, o conceito de *Linked Data* tem sido sugerido como uma abordagem no enriquecimento semântico de VGI (SCHADE; GRANELL ;DÍAZ, (2010); STADLER, (2012)).

Entretanto, adicionar manualmente semântica em VGI é uma tarefa custosa, tediosa e suscetível a erros. Além disso, exigir que o usuário descreva a semântica e relacionamentos semânticos dos dados geográficos já produzidos de forma voluntária é impraticável, seja pela falta de conhecimento do usuário em especificar adequadamente a semântica dos dados, seja pelo gasto para especificá-las, o que afastaria o usuário de produzir novos dados geográficos. Portanto, um processo automático se faz necessário para enriquecer semanticamente grandes quantidades de VGI, buscando atender as necessidades de um domínio de aplicação específico e fornecer uma descrição mais completa dos dados gerados pelos usuários.

Este artigo propõe um método para o enriquecimento semântico de VGI, utilizando como base relações espaciais de proximidade. Com as contribuições VGI semanticamente enriquecidas pela abordagem proposta, é possível realizar análises semânticas em cima de tais dados, e, por consequência, facilitar a descoberta dos mesmos, permitir a realização de análises semânticas e, conseqüentemente, melhorar a descoberta e uso de tais dados. Além disso, a adição de semântica à VGI permite encontrar ou resolver inconsistências e ambigüidades, melhorando a qualidade dos dados produzidos pelo usuário. O sistema VGI para contribuições relacionadas ao desperdício de água, chamado Gota d'Água<sup>11</sup>, foi utilizado como caso de estudo para demonstrar análises com as contribuições que somente são possíveis após o processo de enriquecimento semântico.

O restante do artigo está estruturado como segue. A Seção 2 faz uma revisão dos principais conceitos envolvidos no trabalho. A Seção 3 cita alguns trabalhos relacionados. A Seção 4 apresenta o método proposto. A Seção 5 descreve o estudo de caso. Na Seção

---

<sup>11</sup> <http://www.gotadaguaufv.com.br/>

6 é feita uma discussão dos resultados obtidos e, por fim, na Seção 7 são feitas algumas considerações finais.

## **2 REFERENCIAL TEÓRICO**

Esta seção apresenta a fundamentação necessária para o entendimento do restante deste trabalho. A subseção 2.1 descreve as características de VGI. As subseções 2.2 e 2.3 apresentam, respectivamente, o conceito *Linked Data* e o processo de enriquecimento semântico baseado neste conceito.

### **2.1. Informação Geográfica Voluntária**

O crescimento de fontes de informação como resultado de computadores em rede e outros dispositivos interconectados ocasionou mudanças significativas na quantidade, disponibilidade e natureza da informação geográfica. Entre as mudanças mais significativas está a quantidade crescente de informações geográficas produzidas de forma voluntária, prontamente disponível (FLANAGIN; METZGER, 2008).

Isso se deve aos avanços associados à Web 2.0, principalmente no contexto das tecnologias de mapeamento na Web, que têm incentivado melhorias no que diz respeito à coleta, compartilhamento e interação com as informações geográficas online, possibilitando a geração de conteúdo pelo usuário (GOODCHILD, 2007a).

O ambiente de mídia atual tem a capacidade de promover, apoiar e sustentar esforços coletivos entre indivíduos. Através de sistemas mais dinâmicos na Web, usuários desempenham funções de consumidor de informações e provedor de informações (GOODCHILD, 2007a). Como consequência, esta estrutura é particularmente adequada à colaboração entre indivíduos.

Outra razão para o número crescente de VGIs é a enorme disseminação de dispositivos equipados com o Sistema de Posicionamento Global (GPS), como *smartphones* e *tablets*. Os usuários deste tipo de equipamento estão produzindo dados espaciais para a Web em uma escala muito maior que anteriormente. A interface gráfica fácil e rápida de usar foi um fator importante para atrair os usuários para o universo de geração de conteúdo. Além disso, estes aparelhos possuem alta mobilidade e ferramentas para compartilhar dados espaciais em redes sociais. Os dados que possuem alguma característica espacial e que são gerados a partir de algum tipo de contribuição voluntária são considerados VGI (GOODCHILD, 2007b).

O aumento acentuado nas contribuições VGI levam ao surgimento de diversas plataformas e projetos que utilizam os dados e tecnologias na tomada de decisão espacial, planejamento participativo e ciência do cidadão.

Um exemplo é o *OpenStreetMap* (OSM), um projeto de mapeamento colaborativo que tem como objetivo criar um conjunto de dados geográficos que sejam livres para os usuários utilizarem e editarem (HAKLAY; WEBER, 2008). O OSM é conhecido como um dos mais complexos e promissores projetos relacionados à VGI. Ele permite aos usuários fazerem downloads de dados geoespaciais sem custo e usá-los em projetos pessoais (ZIELSTRA; ZIPF, 2010). Projetos como o OSM possuem ferramentas capazes de capturar, exibir, produzir e disseminar informações em escalas maiores, além de permitir experiências para diversos fins.

O potencial do uso de VGI tem sido comprovado para fins de gestão urbana, danos causados por catástrofes como inundações, incêndios, terremoto, ou outros casos importantes de risco, crises e gestão de desastres naturais (NEIS; ZIELSTRA, 2014). A mudança do clima global, o crescimento populacional e a propagação de doenças infecciosas, dentre outros fatores, aumentou o número de desastres naturais e antrópicos, em que as respostas a estes desastres envolvem pessoas de diversas regiões por meio da colaboração voluntária (ZOOK et al., 2010). Os sistemas VGI podem ajudar na prevenção e execução de ações durante um fenômeno ou na recuperação de uma região após um desastre natural.

Em comparação com dados de fontes oficiais do Governo ou dados gerados por empresas privadas, VGI requer menos tempo e esforço para sua criação, além de um baixo custo. Portanto, estes sistemas podem ser usados como fonte de informação em casos de desastres naturais, visto que nestas situações necessita-se que os dados sejam gerados em tempo real (GOODCHILD, 2007b).

Informação produzida voluntariamente possui vantagens como: ser gratuita; poder ser obtida rapidamente; e gerar dados que ainda não foram disponibilizados por agências oficiais. Porém, seu conteúdo não possui qualidade garantida (GOODCHILD; LI, 2012). Desta forma, é aconselhável usar métodos para avaliar e garantir a qualidade do conteúdo VGI.

Existe uma notável preocupação com a qualidade VGI devido à grande quantidade de dados que são fornecidos por diferentes indivíduos (FLANAGIN; METZGER, (2008); FOODY et al., (2013)). Estas preocupações também estão relacionadas com a falta de

controle da qualidade durante o processo de criação de dados. No entanto, vários autores (HAKLAY, (2010); ZIELSTRA; ZIPF, (2010); FAN et al., (2016); BROVELLI et al., (2016)) compararam dados voluntários com dados de agências oficiais e concluíram que o nível de qualidade dos dados VGI está se aproximando do nível de qualidade dos dados de agências oficiais, principalmente em áreas urbanas densamente povoadas. Além disso, VGIs podem ser enriquecidos de modo a melhorar a qualidade dos mesmos. Uma forma de enriquecer tais dados é através do uso de *Linked Data* e do processo de enriquecimento semântico (SCHADE; GRANELL; DÍAZ, 2010).

## 2.2. *Linked Data*

*Linked Data* refere-se a um estilo de se publicar e interligar dados semânticos na Web. Os dados são disponibilizados na Web de tal forma que são facilmente processados por máquinas, sendo o seu significado/semântica definido de forma explícita. Além disso, esses dados são ligados bidirecionalmente a outros conjuntos de dados externos (BIZER; HEATH; BERNERS-LEE, 2009).

O objetivo é permitir que as pessoas compartilhem dados com semântica bem definida na Web de forma tão fácil quanto documentos são compartilhados atualmente. Sendo assim, visa à partilha e reutilização de informação de modo massivo, num espaço de dados global, permitindo também a descoberta de novos dados (AZEVEDO, 2014). Quanto mais um dado for interligado com outros dados maior é o seu valor e sua utilidade.

Um dos formatos mais empregados para estruturação de *Linked Data* é o formato *Resource Description Framework* (RDF<sup>12</sup>). RDF é um modelo genérico de dados baseado na estrutura de grafos que descreve as coisas do mundo. O modelo RDF codifica dados na forma de triplas: **sujeito**, **predicado**, **objeto**. O **sujeito** e o **objeto** de uma tripla são ambos *Uniform Resource Identifiers* (URI) que identificam cada recurso, podendo o **objeto** assumir o valor de uma *string*. O **predicado** especifica como o **sujeito** e o **objeto** estão relacionados, e também é representado por um URI de um recurso que descreve tal relacionamento (BIZER; HEATH; BERNERS-LEE, 2009).

Para alcançar e criar dados no estilo *Linked Data*, as tecnologias devem estar disponíveis para um formato comum (RDF), para fazer qualquer conversão ou acesso em tempo real para bancos de dados existentes (relacionais, XML, HTML) (DIMA, 2013).

---

<sup>12</sup> <https://www.w3.org/RDF/>

Alguns princípios de *Linked Data* foram introduzidos por Berners-Lee et al. (2006), sendo eles: usar URI como nome para recursos; usar URIs com HTTP para que as pessoas possam encontrar esses nomes; quando alguém procura por uma URI, garantir que informações úteis possam ser obtidas por meio dessas URIs, as quais devem estar representadas no formato RDF; e incluir links para outras URIs de forma que outros recursos possam ser descobertos.

Segundo Berners-Lee et al. (2009), *Linked Data* devem ser publicados com metadados, auxiliando sua descoberta e sua utilização. É sugerido por Hartig (2009) que os metadados relacionados à *Linked Data* devem conter, no mínimo, o criador, a data de criação e o método de criação.

O exemplo mais conhecido de repositório aberto de *Linked Data* é o projeto DBpedia<sup>13</sup> (LEHMAN et al., 2009), que é uma wiki semântica baseada na Wikipedia. A DBpedia disponibiliza suas informações em um formato RDF, o que possibilita realizar consultas. Estas consultas podem ser realizadas utilizando a linguagem SPARQL<sup>14</sup>.

### 2.2.1. Padrão RDF

O RDF é um padrão da *World Wide Web Consortium* (W3C) projetado para descrever recursos da Web (VIEIRA et al., 2005). O padrão RDF define um modelo e uma linguagem para a representação homogênea de informações associadas a recursos que podem ser identificados por meio da Web. O RDF permite especificar o que são recursos, o que são relacionamentos e quais são os possíveis relacionamentos entre os recursos em um domínio qualquer.

O RDF pode também ser definido como um framework desenvolvido para ser lido por computadores, por apresentar uma forma comum de descrição, atua diretamente sobre problemas de interoperabilidade.

Segundo Vieira et al. (2005), a descrição de recursos do RDF pode ser exemplificada como conteúdo e informação de *copyright*, propriedades de itens de compra, como preço e disponibilidade, conteúdo para busca, etc. Linguagens como RDF permitem a troca e o reuso da definição de conceitos ligados a um domínio ou aplicação. Essas especificações podem ajudar a encontrar informações ou serviços de acordo com

---

<sup>13</sup> <http://wiki.dbpedia.org/>

<sup>14</sup> <http://www.sparql.org/>

as necessidades de um usuário ou cliente, muitas vezes levando em consideração também a sua localização geográfica.

Por meio do RDF pode-se atuar indiretamente sobre os dados, já disponíveis na rede, pela associação de metadados que os descrevem. Tanto quanto desejado, o conteúdo destes metadados pode ser lido por pessoas, ou ser processado por programas de computador (SANTANCH, 2003). No segundo caso, o RDF tem vantagem sobre HTML pela sua capacidade de definir taxonomias a serem utilizadas nas descrições. Programas de computador são capazes de reconhecer automaticamente o papel de cada elemento dentro de uma estrutura, organizada a partir de uma taxonomia, bem como estabelecer sua função e aplicabilidade, conforme o domínio a que está associada tal taxonomia.

A forma que o RDF especifica a descrição de recursos segue a estrutura: recurso-propriedade-valor. Essa estrutura é análoga à tripla sujeito-predicado-objeto. O recurso que está sendo descrito é chamado de sujeito; a propriedade que está sendo declarada é chamada de predicado e o valor desta propriedade é o objeto. Um recurso é qualquer coisa que possa ter um URI, o qual estabelece um meio de identificar qualquer recurso de forma única dentro da Web (e.g. "<http://www.w3c.org/RDF>"). Recursos podem ser, por exemplo, objetos físicos, conceitos abstratos, fatos, etc. Já uma propriedade indica uma informação que relaciona um objeto a um recurso, como "autor" ou "homepage". Por fim, um valor é o valor da propriedade, como "Joao", valor da propriedade "autor". E o valor da propriedade pode ser a URI de outro recurso, possibilitando e facilitando sua adoção para *Linked Data*.

Devido aos dados serem especificados como triplas, os dados em RDF podem ser representados como grafos. O exemplo da Figura 1 apresenta uma tripla, i.e., um grafo com dois nodos e uma aresta. O grafo ilustrado representa a tripla RDF que descreve a informação "o nome da empresa Petrobras é Petróleo Brasileiro S.A." (MENDONÇA, 2013). O recurso Petrobras representa o sujeito, é identificado pela URI <http://dbpedia.org/resource/Petrobras> e é ilustrado pela elipse à esquerda; a propriedade nome representa o predicado, sendo identificada pela URI <http://dbpedia.org/property/name> e é ilustrada pela aresta direcionada do sujeito para o objeto e o valor literal "Petróleo Brasileiro S.A." representa o objeto e é ilustrado pelo retângulo à direita. Outras triplas RDF podem ser incluídas neste grafo, representando outras propriedades da empresa Petrobras e também a interligando a outros recursos.

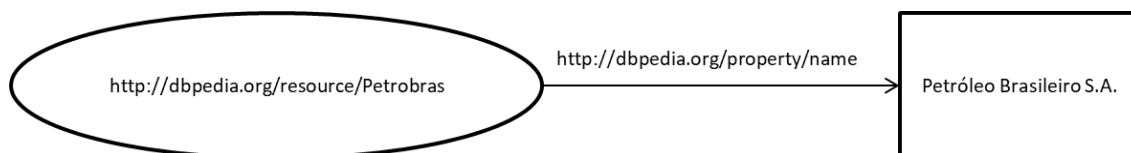


Figura 1. Tripla RDF que descreve a informação "o nome da empresa Petrobras é Petr leo Brasileiro S.A." (MENDONÇA, 2013).

Devido   propriedade de grafos e sem ntica inerentes ao padr o RDF,   necess rio utilizar uma linguagem de consulta que tire proveito de tais caracter sticas para facilitar a recupera o e descoberta de dados. Um exemplo deste tipo de linguagem   a SPARQL, descrita resumidamente na se o seguinte.

### 2.2.2. SPARQL

SPARQL   uma linguagem de consulta desenvolvida principalmente para consultas em grafos RDF (SIRIN e PARSIA, 2007), ou seja, ela   respons vel por extrair informa es de fontes de dados estruturados como triplas. O termo SPARQL   um acr nimo recursivo que significa *SPARQL Protocol and RDF Query Language*,   simultaneamente um protocolo e uma linguagem de consulta a dados representados em RDF.   uma linguagem que permite recuperar valores de dados estruturados e semiestruturados, explorar dados ao consultar rela es desconhecidas, realizar uni es complexas de conjuntos de dados diferentes em uma  nica e simples consulta (PRUD e SEABORNE, 2006).

Pode-se fazer uma analogia entre SPARQL e a *Structured Query Language* (SQL) de consulta a bancos de dados relacionais, considerando que SPARQL tem uma sintaxe adequada a consultas a dados representados como um conjunto de triplas RDF. Dada uma fonte de dados, uma consulta SPARQL consiste em um padr o que   compat vel com ela, e os valores obtidos dessa correspond ncia s o processados para dar a resposta. A fonte de dados a ser consultada pode ser composta por m ltiplas fontes.

Uma consulta SPARQL   composta por: (i) declara es de prefixos, para abreviar URIs; (ii) defini o do conjunto de dados, informando quais grafo(s) RDF est o sendo consultados; (iii) a cl usula de resultado, identificando que informa o deve ser retornada a partir da consulta; (iv) o padr o de consulta, especificando o que consultar dentro do conjunto de dados; e (v) modificadores de consulta, limites, ordena o, e outros que podem modificar o resultado final. Um exemplo de estrutura de uma consulta SPARQL   mostrado no C digo 1.

```

Código 1. Estrutura de consulta SPARQL
1. # declarações de prefixo
2. PREFIX foo: <http://example.com/resources/>
3. ...
4. # definição de conjunto de dados
5. FROM ...
6. # cláusula de resultado
7. SELECT ?x...
8. # padrão de consulta
9. WHERE { ... }
10.# modificadores de consulta
11.ORDER BY ...

```

Segundo Pérez, Arenas e Gutierrez (2006), em uma consulta SPARQL a parte de correspondência de padrões inclui vários recursos interessantes de combinação de padrões de grafos, como partes opcionais, união de padrões, aninhamento, filtragem (ou restrição) de possíveis correspondências e a possibilidade de escolher a fonte de dados para ser acompanhada por um padrão. Os modificadores de solução, uma vez que a saída padrão foi dada (na forma de uma tabela de valores de variáveis), permite modificar esses valores aplicando operadores clássicos como projeção, distinção, ordem, limite e deslocamento. Finalmente, a saída de uma consulta SPARQL pode ser de diferentes tipos: seleções de valores das variáveis que correspondem aos padrões, construção de novas triplas desses valores e descrições de recursos.

As variáveis SPARQL começam com um caractere “?” e podem ser definidas em qualquer uma das três posições de uma tripla (sujeito, predicado, objeto) no conjunto de dados RDF. Uma consulta SPARQL consiste numa estrutura simples de duas cláusulas: SELECT e WHERE. O SELECT identifica as variáveis que aparecerão nos resultados da query e o WHERE mostra o padrão básico do grafo que bate com os dados. A cláusula SELECT retorna uma tabela de variáveis com os valores que satisfazem a consulta.

A linguagem SPARQL especifica quatro formas de consultas com diferentes tipos de propósitos para recuperar os dados (MENDONÇA, 2013). A consulta SELECT extrai dados RDF e retorna os resultados estruturados em formato de tabela; a consulta CONSTRUCT possibilita a construção de novos grafos RDF a partir dos dados extraídos; a consulta ASK provê uma resposta booleana (verdadeiro ou falso) para perguntas que verificam se um padrão de consulta possui resultados e a consulta DESCRIBE retorna um grafo RDF contendo informações sobre um determinado recurso identificado por uma URI. Em cada forma de consulta, um bloco WHERE deve ser usado para restrição do

conjunto de resultados, sendo a restrição opcional para consultas do tipo ASK e DESCRIBE.

A linguagem SPARQL está em processo de padronização, e nesse processo a semântica do idioma desempenha um papel fundamental. A formalização de uma semântica será benéfica por vários motivos: ajuda a identificar relacionamentos entre os construtores que ficam escondidos nos casos de uso, identifica noções redundantes e contraditórias, estuda a expressividade e a complexidade do idioma, ajuda na otimização, etc.

### **2.3. Enriquecimento Semântico de dados**

A Web Semântica é responsável por adicionar significado às informações disponíveis na Web. Esse significado deve ser entendido tanto por humanos como por programas computacionais (VIEIRA et al., 2005).

Com avanço do uso da Web Semântica, sistemas que permitem que os usuários gerem conteúdo para a Web através de experiências do mundo real (e.g., mídias sociais, sistemas VGI) são vistos como sistemas que terão forte impacto na aprendizagem informal. No entanto, esses conteúdos não possuem um domínio de aprendizagem bem definido e normalmente são escritos em linguagem natural. Sendo assim, para que a informação seja compreensível e tratável por um agente computacional, e, portanto, seja parte da Web Semântica, é necessário o uso de técnicas e ferramentas que proporcionem um entendimento mínimo sobre o conteúdo (MOREIRA et al., 2014).

Para que isso seja possível, é importante fazer uso do que diz respeito ao enriquecimento semântico, que pode ser visto como o processo de atribuir maior significado aos metadados e dados através da aplicação de recursos auxiliares, com o objetivo de facilitar a compreensão, a integração e o processamento dos dados por pessoas e máquinas. Ou seja, o enriquecimento semântico torna os dados e metadados mais qualificados e descritivos, através do uso da semântica atribuída por recursos semanticamente bem-descritos, como recursos de bases de conhecimento, ontologias, vocabulários existentes, entre outros (LIRA, 2014).

Segundo IBM (2016), enriquecimento semântico diz respeito ao processo de enriquecer conteúdo com contexto de dados por marcação, além de categorizar e/ou classificar os dados uns em relação aos outros e a dicionários ou outras fontes base de referência. Isso significa acrescentar informações contextuais a algum conjunto de dados

existente. Por exemplo, acrescentar dados de tráfego a mapas de estrada em que os dados fornecem o contexto de condições de estrada, a probabilidade de atraso, tempo de obstruções projetadas, condição da estrada, etc.

Sistemas de pesquisas que retornam resultados mais relevantes tornam os usuários mais produtivos, além de tornar o conteúdo encontrado mais útil. Sendo assim, enriquecimento semântico é visto como meio de melhorar o resultado de uma busca. Apesar dos benefícios de se adicionar semântica em dados, questões como custo, precisão e escalabilidade necessitam ser levadas em consideração, antes de se adotar uma abordagem ou processo de enriquecimento semântico. Uma estratégia semântica bem pensada irá minimizar os custos envolvidos no processo (CLARKE; HARLEY, 2014).

Para realizar o enriquecimento semântico e obter conceitos adicionais, alguns recursos e técnicas são usados, dentre eles a anotação semântica, que consiste na atribuição de significado (semântica) aos elementos de um esquema de origem, que segundo Clarke e Harley (2014) pode ser de forma manual ou automatizada.

Na forma manual, uma pessoa que possui os conhecimentos adequados, lê o conteúdo e aplica a marcação. Esse processo às vezes é chamado de indexação semântica. A marcação manual é ideal quando o uso pretendido de marcação requer um alto grau de precisão. Mas pode ter um custo proibitivo para grandes volumes de conteúdo, porque é intensiva a mão-de-obra e difícil de dimensionar para grandes volumes de trabalho. Alguns tipos de conteúdo, como multimídia, não são acessíveis a sistemas automatizados e a marcação manual pode ser a melhor opção.

Na marcação automatizada, o software analisa o conteúdo, adicionando marcação com base em correspondência de conceitos, padrões estatísticos e análise linguística. A maioria dos sistemas automatizados conta com ajustes de algoritmos utilizados para a marcação para se adequarem a um conjunto de dados e a um campo de conhecimento específico, aumentando assim o nível de precisão que pode ser alcançado através da automação. A forma automatizada é altamente escalável e às vezes, é a única opção para conjuntos de conteúdo muito grandes (e.g. postagens de mídias sociais). No entanto, abordagens automáticas podem levar a falsos positivos (e.g. tags aplicadas em partes dos dados não relevantes), conceitos perdidos e outras imprecisões.

Além disso, destacam-se os processos de enriquecimento semântico: Vinculação e Mapeamento de Recursos, o qual pode ser aplicado juntamente com anotações semânticas; e Conversão para Modelos de Dados Semânticos. A Vinculação e

Mapeamento de Recursos é um processo que consiste em descobrir links entre as combinações semânticas dos dados e metadados com outros recursos na Web de dados. É muito utilizado para interligar recursos na nuvem *Linked Open Data*. A Conversão para Modelos de Dados Semânticos consiste em modelar os dados num formato semântico e estruturado, como RDF/XML, facilitando sua manipulação por aplicações que consomem esses modelos de dados (LIRA, 2014).

Visto que a Web Semântica é uma Web de dados, as tecnologias da Web Semântica (RDF, OWL, SKOS, SPARQL) fornecem um ambiente onde uma aplicação pode consultar esses dados e ou fazer inferências usando vocabulários controlados.

Estas tecnologias integradas na Web Semântica começam a produzir resultados, desde aplicações básicas, usando a descrição RDF (no âmbito da descrição de recursos), vincularem dados de vários sites da Web usando uma linguagem padrão SQL, a SPARQL que consulta RDF até ligações já prontas em XML ou ontologias em OWL (CYGANIAK, 2014).

### **3 TRABALHOS RELACIONADOS**

Segundo Ronzhin (2015), quando um conjunto de VGI é integrado à repositórios abertos de *Linked Open Data* (LOD), pode-se obter vantagens que vão além de apenas recursos que estão diretamente interligados com o conjunto de VGI. Isso se deve à grande interconectividade entre os conjuntos de dados publicados como dados vinculados.

No seu trabalho, Ronzhin (2015) procura investigar até que ponto repositórios LOD podem ajudar a enriquecer semanticamente VGI, a fim de obter melhores respostas em consultas no contexto de operações e crise. A partir disso, afirma-se que o uso de URIs de uma base de conhecimento aberta como, por exemplo, o *LinkedGeoData* (LGD), elimina a possibilidade de ambiguidade ao apontar um local. A semântica adicional obtida a partir de informações estruturadas apresentadas no LGD permite, por exemplo, o acesso rápido a conhecimentos básicos e úteis sobre objetos de infraestrutura e edifícios públicos, que podem ser utilizados para agilizar e melhorar o processo de tomada de decisão. Portanto, a integração de VGI com entidades relevantes na nuvem LOD torna possível o enriquecimento semântico de conteúdos não estruturados gerados pelo usuário com informações estruturadas apresentadas em LOD.

Envolvendo o conceito de Linked Data, Azevedo (2014) propõe uma solução capaz de obter dados heterogêneos relacionados a inundações na Bacia do Rio Doce (bacia

hidrográfica localizada no sudeste do Brasil). A partir de diferentes fontes de dados providas de diversas organizações públicas, o autor integra e disponibiliza os dados para visualização em um Sistema de Informação Geográfica (GIS). Para tal, propõe que os dados sejam convertidos para o formato RDF, interligando e visualizando-os com auxílio de consultas SPARQL. O estudo teve a metodologia experimental, com base em prova de conceito, sendo a Bacia Hidrográfica do Rio Doce, a unidade de análise.

Sorrentino et al. (2013) propõem interligar dados semanticamente enriquecidos com repositórios Linked Data através de uma metodologia experimental semiautomática para facilitar os provedores de recursos na publicação de dados públicos em repositórios LOD e para ajudar os consumidores (empresas e cidadãos) a acessá-los e consultá-los com eficiência. O autor apresenta um método para publicação, ligação e enriquecimento semântico dos dados abertos, realizando a anotação semântica automática de elementos de esquema. O método foi aplicado em um conjunto de dados fornecidos pelo Projeto de Pesquisa sobre Prevenção de Jovens, na Itália, que realiza realizando uma investigação sobre a situação precária de jovens que vivem no distrito de Modena.

Mendonça (2013) descreve uma abordagem de coleta e publicação de dados para o processo de publicação de *Linked Data*. Esta é uma abordagem de coleta e publicação de dados de proveniência para o processo de publicação de *Linked Data* realizado dentro dos limites de uma organização. Proveniência de um recurso no contexto de *Linked Data* é considerada como o conjunto de informações relacionadas às entidades, aos processos e aos agentes envolvidos na produção e disponibilização do recurso. Assim, um agente, denominado Agente Coletor de Proveniência, coleta, interliga e armazena temporariamente os dados de proveniência prospectiva e retrospectiva, durante a execução do processo de publicação de *Linked Data*. Posteriormente, a proveniência coletada é também publicada no contexto de *Linked Data*, a fim de que os dados de domínio e seus respectivos dados de proveniência possam ser explorados conjuntamente, por meio de consultas SPARQL. A abordagem oferece apoio para a avaliação da qualidade e da confiabilidade dos dados publicados no contexto de *Linked Data*. Sendo que, uma das limitações apresentadas é a necessidade de uma estratégia para gerenciar o grande volume de dados gerado pela publicação da proveniência.

O enriquecimento semântico derivado do conhecimento sobre relacionamentos entre dados geoespaciais é uma potencial fonte de soluções para problemas tradicionais em recuperação de informação geográfica (MOURA, 2013). Sendo assim, Moura (2013)

discute a aplicabilidade dos conceitos de *Linked Data* a diversos problemas, como resolução de ambiguidade e reconhecimento do contexto espacial de documentos. Além disso, o trabalho apresenta uma lista de desafios e oportunidades para a pesquisa sobre recuperação de informação geográfica e tópicos correlatos.

Diferentemente dos trabalhos relatados, este trabalho propõe um método automatizado de anotação semântica de contribuições VGI, além de propor um processo para encontrar contribuições VGI semelhantes tirando proveito das contribuições semanticamente enriquecidas e de tesouros construídos manualmente a partir das contribuições.

#### **4 MÉTODO PARA ENRIQUECIMENTO DE VGI COM *LINKED DATA* E CRIAÇÃO DE TESAURO**

Esta seção apresenta o processo proposto para o enriquecimento semântico de VGI com *Linked Data*. O processo proposto visa enriquecer semanticamente as contribuições VGI a partir (i) das suas coordenadas geográficas, anotando as contribuições com possíveis locais de interesse (*Place of Interest* - PoI) para o usuário no contexto em que se encontra a contribuição e (ii) dos textos das contribuições, buscando encontrar palavras relevantes e atribuir o significado correto às mesmas para a geração de um tesouro.

A Figura 2 apresenta um fluxo do processo proposto. A entrada do processo são as contribuições VGI que se deseja enriquecer semanticamente e uma base de dados (e.g., base de conhecimento, repositório LOD) contendo os PoIs da região em que as coordenadas geográficas das contribuições apontam. Tanto as contribuições VGI como a base de dados contendo os PoIs passam por uma etapa de Pré-Processamento, onde ruídos (e.g., erros gramaticais, inconsistências temporais e espaciais), duplicatas e entre outros problemas nos dados são corrigidos. Após o pré-processamento, as contribuições VGI são enriquecidas semanticamente por duas tarefas distintas e que podem ocorrer em paralelo, sendo elas: Enriquecimento Semântico Automático e Enriquecimento Semântico Manual.

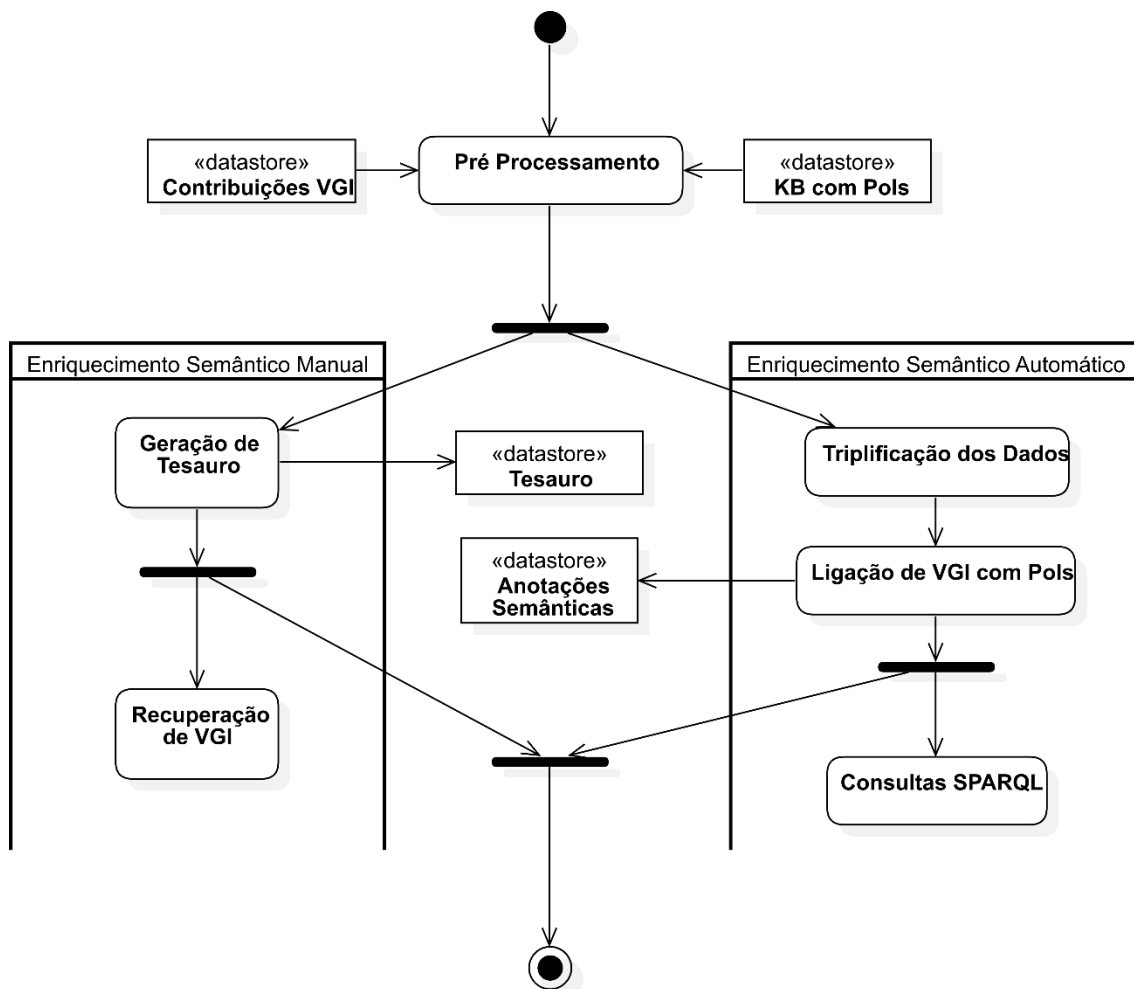


Figura 2: Fluxo do processo proposto de enriquecimento semântico de VGI

O Enriquecimento Semântico Manual consiste na geração de um tesouro que descreva as contribuições VGI que se deseja enriquecer semanticamente, exigindo apenas que as mesmas possuam elementos textuais. Para a geração do tesouro, foi adotado processo semelhante ao utilizado em CEPIS (2005).

A geração de um tesouro pode ser semi-automatizada com ferramentas que identificam o sentido correto das palavras relevantes (ferramentas de *Word Sense Disambiguation* - WSD) e ferramentas que identificam menções a entidades no mundo real e ligam tais menções a recursos semanticamente bem-descritos que as descrevem corretamente no contexto que elas se encontram (ferramentas de *Entity Linking* – EL) (GAO; BARBIER; GOOLSBY, 2011). No entanto, tais ferramentas ainda não apresentam resultados satisfatórios quando o texto a ser anotado possui pouco contexto como, por exemplo, postagens de mídias sociais e contribuições VGI. Portanto, a utilização de tais ferramentas não foi considerada no processo durante a geração do tesouro.

Já o Enriquecimento Semântico Automático consiste em anotar as contribuições VGI com os PoIs do usuário para o contexto da contribuição. Para extrair todo o potencial do enriquecimento semântico, i.e., facilitar sua descoberta e uso por aplicações, as contribuições VGI foram triplificadas adotando o padrão RDF. Para realizar a triplificação, um algoritmo foi desenvolvido. Apesar de existirem ferramentas para triplificação de dados, como a *Triplify*<sup>15</sup>, devido à estrutura simplificada dos dados utilizados no caso de estudo (Seção 5), um simples algoritmo foi o suficiente para a realização da tarefa. Além disso, devido à sua simplicidade e por não ser o foco deste trabalho, o algoritmo de triplificação não é apresentado.

Após a triplificação das contribuições VGI, a próxima etapa do processo é o enriquecimento semântico automático das contribuições, mais especificamente, a ligação das contribuições com PoIs.

#### **4.1. Algoritmo para o enriquecimento semântico automático de VGI**

O Algoritmo 1 consiste no processo proposto para o enriquecimento semântico automático de contribuições VGI para a ligação das contribuições com PoIs que interessam o usuário que postou a contribuição. Nota-se que o algoritmo utiliza operações geoespaciais para encontrar os possíveis PoIs de interesse do usuário. Portanto, o Algoritmo 1 funciona com qualquer biblioteca ou Sistema de Gerenciamento de Banco de Dados (SGBD) com suporte a operações geoespaciais entre diferentes geometrias.

---

<sup>15</sup> <http://aksw.org/Projects/Triplify.html>

ALGORITMO 1 Enriquecimento Automático de VGI com locais de interesse

**Require:**

$V = \{v_0, \dots, v_n\}$  // Conjunto de VGI  
 $P = \{p_0, \dots, p_m\}$  // Base de conhecimento que contém lugares com semântica bem definida e coordenadas geográficas  
 $T_r \in \mathbb{R}$  // Threshold em metros para distância espacial

**output:**

SA // conjunto de anotações semânticas inicialmente vazio

**begin**

$SVP \leftarrow (\Pi v \leftarrow V.cod, p \leftarrow P.cod, names, geoDist(V \bowtie (geoDist \leftarrow dist(v.geom, p.geom)) \leq T_r));$

**for each**  $v \in V$  **do**

**for each**  $(v, p, geoDist) \in SVP$  **do**

$R.AddAnnotation(v, p);$

**end for**

$SA.add(R);$

**end for**

**return** SA;

**end**

A entrada do algoritmo consiste de: um conjunto de contribuições VGI  $V$ ; uma base de conhecimento  $P$  contendo os PoIs a serem usados para enriquecer o conjunto VGI  $V$ ; e um valor em metros  $Tr$  para o tamanho do *buffer* a ser utilizado nas operações geoespaciais. A saída do algoritmo é um conjunto  $SA$  de anotações semânticas como arquivos RDF, sendo inicialmente um conjunto vazio. Para cada contribuição  $v \in V$ , é gerado um *buffer* circular com raio  $Tr$  em volta da posição geográfica de  $v$ . Nota-se que, apesar de não estar presente no algoritmo para se aproximar com o que realmente foi implementado, o *buffer* a ser gerado pode assumir diversas outras geometrias (e.g., retangular, geometrias complexas) de modo a melhor se adequar as necessidades de diversas aplicações. Após a criação do *buffer* é verificado quais os lugares  $p \in P$  se encontram geograficamente dentro do *buffer*. Para cada um desses lugares é criada uma anotação semântica  $r \in R$ , a qual irá conter como sujeito a URI para a VGI  $v$  e como objeto a URI para um PoI  $p$ . O valor do predicado será modificado dependendo do tipo de problema que  $v$  descreve. Por exemplo, se o tipo de problema de uma contribuição  $v$  for “vazamento”, o valor do predicado é “influencia”, enquanto que se o tipo de problema for “desperdício”, o valor do predicado é “notificar”. Após todos os PoIs  $p$  serem anotados

à  $v$ , a anotações  $r_i$  contidas em  $R$  são adicionadas ao conjunto  $SA$ . Por fim, o conjunto  $SA$  é retornado.

Para exemplificar o funcionamento do algoritmo, foi selecionada aleatoriamente uma contribuição VGI da base do sistema Gota D'Água (utilizado como caso de estudo e descrito na Seção 5). Ao ser aplicado o algoritmo, um *buffer* de raio de 100 metros é criado ao redor da coordenada geográfica da contribuição, verificando quais dos PoIs registrados no OSM estão contidos nesse *buffer*. A Tabela 1 mostra quais lugares foram encontrados ao redor da contribuição. As Figuras 3 e 4 mostram o *buffer* envolvendo os estabelecimentos próximos da contribuição, em diferentes representações. Enquanto a Figura 3 é representada no OSM, a Figura 4 é representada no *Google Maps*. Destaque-se que a escolha da base de conhecimento contendo os PoIs afeta o número de anotações semânticas geradas. Como é possível perceber nas Figuras 3 e 4, devido à diferença entre o número de PoIs registrados entre o OSM e o Google Maps, o resultado retornado pelo Algoritmo 1 varia significativamente de acordo com a base de PoIs.

**Tabela 1: Lugares encontrados em um raio de 100 metros a partir da coordenada geográfica de uma contribuição VGI**

Código do local	Características do local
2959454656	"name"=>"Lanchonete do Dênis", "amenity"=>"fast_food"
2959454657	"name"=>"Bar Norte Mineiro", "amenity"=>"bar"
2959454658	"name"=>"Droga Farma", "amenity"=>"pharmacy", "opening_hours"=>"24/7"
2959454659	"name"=>"Alfa Hotel", "tourism"=>"hotel"
2959454660	"name"=>"Orquídea Hotel", "tourism"=>"hotel"
2966424164	"name"=>"Clic Clic", "shop"=>"art"
3412042504	"name"=>"Subway", "amenity"=>"fast_food", "cuisine"=>"sandwich"

A Figura 3 mostra, no sistema OSM, o *buffer* de 100 metros criado a partir do local de uma VGI. Pode-se perceber que todos os locais indicados na Tabela 1 (representados por um círculo amarelo) estão contidos na região do *buffer*. Como o OSM é constantemente atualizado, é possível que alguns lugares que estejam também dentro da região do *buffer* não apareçam.



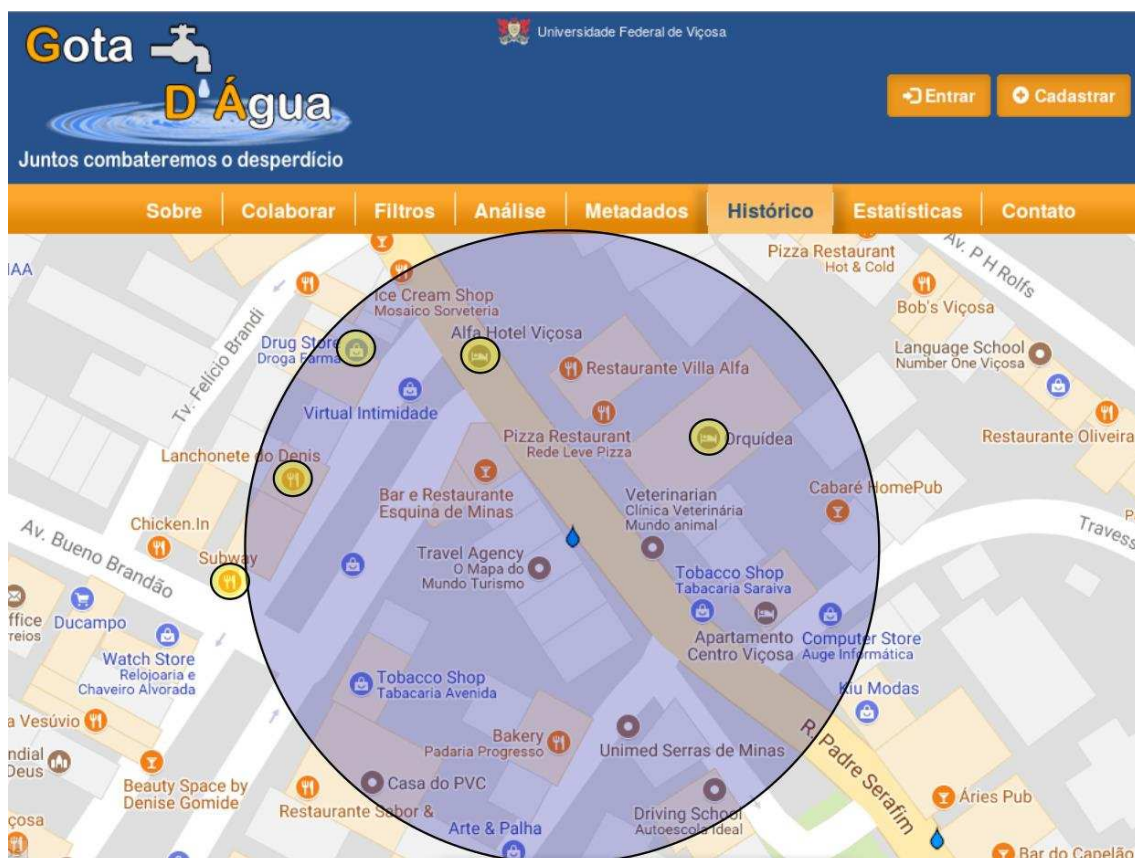


Figura 4: Exemplo de contribuição enriquecida semanticamente pelo algoritmo proposto utilizando como fundo mapa oriundo do Google Maps na interface do sistema Gota D'Água.

## 5 ESTUDO DE CASO: SISTEMA GOTA D'ÁGUA

Como caso de estudo, neste trabalho os métodos de enriquecimento semântico automático de VGI e as análises semânticas são implementados e testados em um sistema de coleta de VGI chamado Gota D'Água<sup>16</sup>. Este sistema foi desenvolvido visando coletar dados por meio de contribuições voluntárias do cidadão, referentes ao desperdício e falta de água, problema que tem sido recorrente em diversas regiões brasileiras no período de estiagem.

O Gota D'Água possibilita que usuários de diferentes partes do Brasil possam contribuir com informações sobre diferentes tipos de problemas relacionados à falta e/ou desperdício de água. Ao fazer uma contribuição, o lugar indicado pelo usuário é identificado pelas suas coordenadas geográficas, e este fica em destaque no mapa para visualização, como visto na Figura 5. Além disso, o usuário pode escolher o tipo de

<sup>16</sup> <http://www.gotadaguaufv.com.br/>

problema identificado (e.g., vazamento, desperdício) e fornecer outras informações mais detalhadas, na forma de texto livre, imagem ou vídeo.

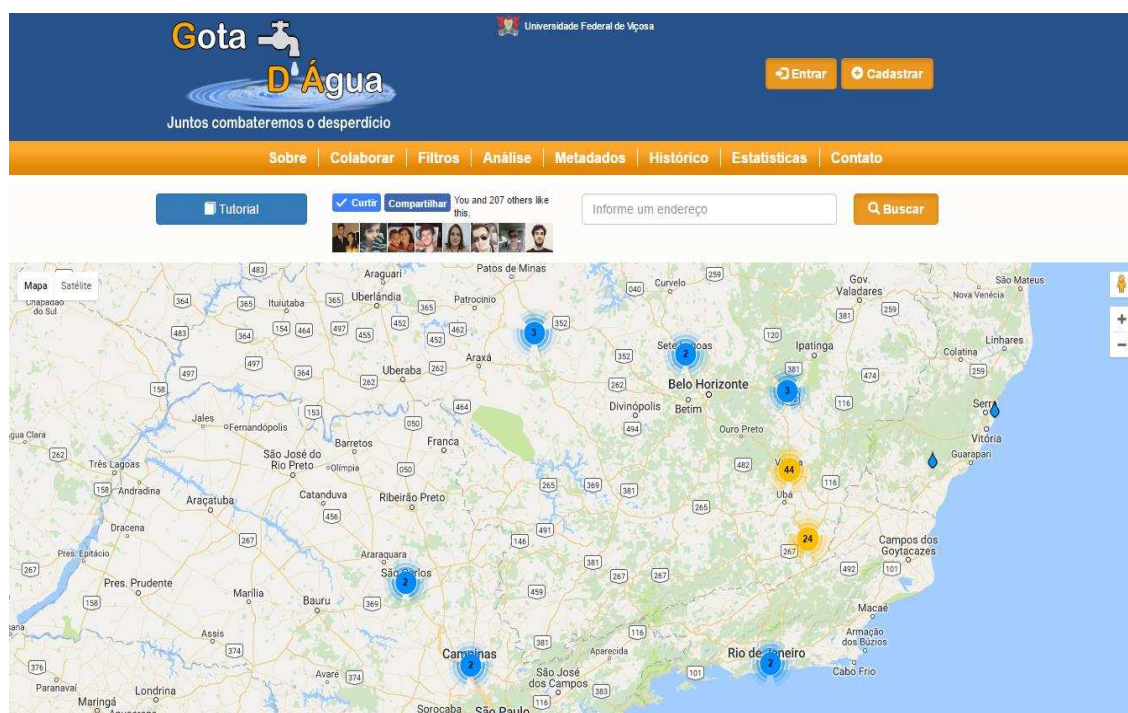


Figura 5: Página inicial do Sistema Gota D'Água

## 5.1. Ferramentas utilizadas

Para a realização dos passos mostrados na Figura 2, diversas ferramentas e artefatos foram utilizados. Em relação aos dados, as contribuições são provenientes do sistema Gota D'Água, sendo realizado um pré-processamento nas contribuições para tratar questões como privacidade (substituindo o nome dos colaboradores por alias) e possíveis ruídos gerados pelo sistema (e.g. contribuições duplicadas). Os PoIs utilizados como recursos para enriquecimento semântico são obtidos a partir do repositório *LinkedGeoData* (LGD). Tal repositório utiliza as informações coletadas pelo projeto *OpenStreetMap* e as tornam disponíveis como uma base de conhecimento RDF de acordo com os princípios de *Linked Data*. O LGD interliga esses dados com outras bases de conhecimento da iniciativa *Linking Open Data* (LOD), como o *DBpedia*. Para a realização das operações geoespaciais propostas no Algoritmo 1, os PoIs foram inseridos em um banco PostgreSQL+PostGIS através da ferramenta *Osmosis*<sup>17</sup>. Apesar de existirem SGBDs que suportem consultas espaciais em dados RDF (Strabon e Oracle, por

<sup>17</sup> <http://wiki.openstreetmap.org/wiki/Osmosis>

exemplo), operações geográficas em dados RDF ainda são menos eficientes que consultas geográficas realizadas em SGBD relacionais. Além disso, o PostgreSQL é um banco escalável e de uso livre. Por fim, o sistema Gota D'Água também utiliza o PostgreSQL, o que reforçou a sua escolha.

As contribuições VGI foram semanticamente anotadas com o Algoritmo 1, implementado em Java, além da utilização da biblioteca Apache Jena<sup>18</sup> para a geração das anotações semânticas utilizando o padrão RDF. Os dados triplicados e as anotações semânticas geradas pelo algoritmo (ambos seguem o princípio *Linked Data*) foram armazenados no banco de dados Virtuoso, permitindo sua consulta em SPARQL. Por fim, o tesouro desenvolvido neste trabalho (detalhado na Seção 5.4) foi baseado no Tesouro de Engenharia Sanitária e Ambiental (CEPIS, 2005).

## 5.2. Exemplos de Consultas em SPARQL

Para demonstrar algumas das diversas possibilidades e vantagens que o enriquecimento semântico providência a VGI, algumas consultas em SPARQL foram propostas para análise dos dados existentes no sistema Gota D'Água. Tais consultas visam demonstrar possíveis análises que somente são possíveis após o processo de enriquecimento semântico. Todas as consultas foram testadas no banco de dados Virtuoso, cuja instância contém as contribuições do sistema Gota D'Água triplicadas e as anotações semânticas das contribuições.

### **Q1. Recuperar os Ids das contribuições, palavras-chave descritas nas contribuições e os nomes dos PoIs ligados as mesmas através das anotações semânticas.**

Esta consulta tem como objetivo identificar os pares <palavra-chave, PoI> existentes para cada contribuição VGI. Desse modo, é possível verificar quais palavras-chave aparecem para determinados PoIs, verificando-se, por exemplo, se os PoIs retornados estão relacionados a uma grande quantidade de palavras-chave (indicando que diferentes problemas ocorrem perto dele) ou se uma palavra-chave é recorrente ao PoI (indicando que o problema se repete  $n$  vezes ou que diferentes usuários relataram o mesmo problema). O Código 2 apresenta a Q1 na linguagem SPARQL executada neste trabalho.

---

<sup>18</sup> <https://jena.apache.org/>

Código 2. Q1 na linguagem SPARQL

```
1. SELECT ?contribuicao ?chave ?lugar
2. WHERE{?contribuicao <gd:keyWords> ?chave.
3. ?a <oa:hasTarget> ?contribuicao; <oa:hasBody> ?l.
4. ?l rdfs:label ?lugar}
5. ORDER BY ?lugar
```

Na linha 1 do Código 2 são recuperados todos os pares <palavra-chave, PoI> de todas as contribuições. A propriedade <gd:keyWords> foi definida pelos autores deste trabalho, sendo *gd* o prefixo para <<http://www.gotadaguaufv.com.br/>>. Na linha 3 é especificado que, para cada anotação semântica (associada à variável *?a*) se deseja o PoI apontado pela anotação semântica (variável *?l*). Por fim, o nome do PoI é associado à variável *?lugar* (linha 4), sendo que as tuplas retornadas são ordenadas de forma crescente pelo nome dos PoIs retornados (linha 5).

A Tabela 2 apresenta uma amostra das tuplas retornadas pela consulta Q1 executada no banco de dados Virtuoso. Na Tabela 2 é possível perceber que, para o PoI Alfa Hotel, duas palavras-chave distintas, cada uma em uma contribuição diferente, foram associadas ao PoI. Ambas palavras-chaves ("vazamento" e "cano quebrado") remetem à um problema de vazamento de água, possivelmente da rede de distribuição ou de esgoto. Através dessas tuplas é possível traçar dois possíveis cenários, que podem ser verificados através do *timestamp* de criação da contribuição: (i) caso ambas contribuições foram reportadas no sistema em um espaço de tempo curto, a confiabilidade de que tal problema realmente ocorreu aumenta, i.e., a confiabilidade de ambas contribuições aumenta e; (ii) caso ambas contribuições foram reportadas com uma diferença no espaço de tempo grande, significa que o problema é persistente naquele PoI devido à algum motivo, o qual deveria ser investigado para diminuir prejuízos à companhia de distribuição de água e esgoto.

**Tabela 2. Amostra das tuplas retornadas pela consulta Q1**

<b>Contribuição</b>	<b>Chave</b>	<b>Lugar</b>
http://gotadagua.com.br/96	Agua	Academia Body Move
http://gotadagua.com.br/111	Desperdício	Academia Vila Fit
http://gotadagua.com.br/111	mangueira.	Academia Vila Fit
http://gotadagua.com.br/111	Calçada	Academia Vila Fit
http://gotadagua.com.br/2	água, registro quebrado	Aconchego Bar e Restaurante
http://gotadagua.com.br/25	Vazamento	Alfa Hotel
http://gotadagua.com.br/18	cano quebrado	Alfa Hotel
http://gotadagua.com.br/3	Vazamento	Armazem
http://gotadagua.com.br/3	Desperdício	Armazem
http://gotadagua.com.br/86	Vazamento	Armazem
http://gotadagua.com.br/1	Falta d'água, linked data, exemplo\r\n	Artes Gerais
http://gotadagua.com.br/18	Cano quebrado	Bar Norte Mineiro
http://gotadagua.com.br/96	Agua	Bar do Gomes

**Q2. Recuperar o número de vezes que um PoI foi apontado por uma anotação semântica no processo de enriquecimento semântico.**

Esta consulta tem como objetivo retornar o número de contribuições que cada PoI foi ligado. Com os resultados da consulta Q2 é possível identificar os PoIs mais recorrentes quando ocorre algum problema relacionado à falta/desperdício de água. O Código 3 apresenta a Q2 na linguagem SPARQL.

Código 3. Q2 na linguagem SPARQL

```

1. SELECT ?lugar (count (distinct ?contribuicao))
2. WHERE{ ?a <oa:hasTarget> ?contribuicao; <oa:hasBody> ?l.
3. ?l rdfs:label ?lugar}
4. ORDER BY ?lugar

```

Na linha 1 do Código 3 são recuperados todos os PoIs, fazendo a contagem das contribuições relacionadas com cada local recuperado. As linhas 2 e 3 do Código 3 se

comportam exatamente igual às linhas 3 e 4 do Código 2. A linha 4 é responsável por organizar o resultado pelos nomes dos PoIs em ordem crescente.

A Tabela 3 apresenta uma amostra das tuplas retornadas pela consulta Q2 executada no banco de dados Virtuoso. Na Tabela 3 é possível perceber que, para o PoI Academia Vila Fit, foi associada apenas uma contribuição e, por último, os PoIs Cacau Show e Boca Viçosa foram associados à 3 contribuições. Ou seja, há indícios de que nas regiões próximas a estes estabelecimentos ocorrem um número maior de problemas, o qual deve ser investigado pelos órgãos responsáveis.

**Tabela 3. Amostra das tuplas retornadas pela consulta Q2**

Lugar	callret-1
Academia Vila Fit	1
Alfa Hotel	2
Armazem	2
Bar Norte Mineiro	1
Bar do Gomes	1
Biblioteca IME-USP	1
Bob's	2
Boca Viçosa	3
Cacau Show	3

### **Q3. Verificar o número de palavras-chave associadas por PoI e quais são as palavras-chave**

Esta consulta retorna o mesmo resultado da Q1, no entanto, de uma forma mais didática para o usuário. Esta consulta foi proposta para demonstrar que o SPARQL é capaz de realizar operações complexas, como concatenação de *strings*, para retornar resultados que atendam de maneira satisfatória diversas aplicações, como mostra o Código 4. Como mostrado na Tabela 4, é mais fácil verificar as palavras-chave associadas a cada PoI do que no resultado retornado na Tabela 2.

Código 4. Q3 na linguagem SPARQL

```

1. SELECT ?lugar (count(?lugar)) (sql:group_concat(?chave, '
   ') as ?names)
2. WHERE{ ?contribuicao <gd:keyWords> ?chave.
3. ?a <oa:hasTarget> ?contribuicao;<oa:hasBody> ?l.
4. ?l rdfs:label ?lugar}
5. ORDER BY ?lugar

```

**Tabela 4. Amostra das tuplas retornadas pela consulta Q3**

Lugar	Count	Chaves
Academia Body Move	1	Água
Academia Vila Fit	3	desperdício; calçada; mangueira
Aconchego Bar e Restaurante	2	Água; registro quebrado
Armazem	3	vazamento; desperdício; vazamento
Biblioteca IME-USP	1	Vazamento
Bar Norte Mineiro	1	cano quebrado
Bar do Gomes	1	Água
Vitrola Café	3	desperdício; calçada; mangueira
Viçosa Outlet	3	desperdício; calçada; mangueira

**Q4. Verificar o número de vezes que as palavras-chave das contribuições são mencionadas para cada tipo de PoI presente nas anotações semânticas.**

Esta consulta tem como objetivo verificar o número de tipos de PoIs associados para cada palavra-chave. Os tipos de PoIs, ou de qualquer dado que segue o padrão RDF, são definidos pelo atributo “rdf:type”, o qual pode ser substituído pela expressão “a” no SPARQL, como mostra a linha 4 do Código 5. Para facilitar a apresentação dos resultados da consulta Q4, foram recuperados somente tipos expressos na ontologia do *LinkedGeoData*. Para realizar este filtro foi utilizada uma expressão regular (linha 4 do Código 5).

Como mostra a Tabela 5, um grande número de PoIs ligados no processo de enriquecimento semântico são do tipo *Amenity*. Isso ocorre devido ao fato que *Amenities* são lugares que proporcionam lazer e/ou conforto, sendo considerado como um tipo genérico de lugar na ontologia do *LinkedGeoData*. Como o *LinkedGeoData* é baseado no OSM, cujas colaborações muitas vezes são realizadas por usuários sem conhecimento ou treinamento em dados geoespaciais, o número de PoIs com uma classificação genérica é alto.

Código 4. Q4 na linguagem SPARQL

```

1. SELECT ?chave ?TipoLugar (count(?TipoLugar)) AS ?count
   WHERE{
2. ?contribuicao <gd:keyWords> ?chave.
3. ?a <oa:hasTarget> ?contribuicao;<oa:hasBody> ?l.
4. ?l a ?TipoLugar. FILTER regex(?TipoLugar,
   "^http://linkedgeo.org/ontology/", "i") }
5. ORDER BY ?TipoLugar

```

**Tabela 5. Amostra das tuplas retornadas pela consulta Q4**

Chave	TipoLugar	Count
Água	<a href="http://linkedgeodata.org/ontology/AlcoholShop">http://linkedgeodata.org/ontology/AlcoholShop</a>	1
Vazamento	<a href="http://linkedgeodata.org/ontology/Amenity">http://linkedgeodata.org/ontology/Amenity</a>	9
Desperdício	<a href="http://linkedgeodata.org/ontology/Amenity">http://linkedgeodata.org/ontology/Amenity</a>	7
Água	<a href="http://linkedgeodata.org/ontology/Amenity">http://linkedgeodata.org/ontology/Amenity</a>	8

### 5.3. Elaboração do Tesouro

A recuperação de informação se tornou imprescindível entre os diversos membros da comunidade científica e os tesouros vieram como resposta a essa necessidade de recuperação de informação (MOREIRA; MOURA, 2009).

Sintetizando as inúmeras definições de tesouros encontradas na literatura, Sales e Café (2013) descrevem tesouros e definem suas características e objetivos. Tesouro é definido como um vocabulário controlado formado por termos/descriptores semanticamente relacionados, e atua como instrumentos de controle terminológico. Dessa forma, tesouros são utilizados na organização/representação do conhecimento.

Geralmente tesouros são estruturados hierarquicamente, contendo relações genéricas, partitivas, equivalentes, associativas e associativamente (aproximação semântica), que são utilizados para indexar e recuperar informações por meio de seu conteúdo. Eles têm como objetivo organizar informações especializadas, coordenar o vocabulário especializado, auxiliar a consulta do usuário de forma sistemática e potencializar a recuperação da informação, atuando como interfaces entre informações e seus consumidores.

O relacionamento entre os conceitos, que é regido pelas características que possuem, é hierárquico, podendo se manifestar de forma superordenada (do conceito mais específico para o mais geral) e subordinada (do conceito geral para o específico).

Tesouros surgiram como instrumento prático para auxiliar na indexação e busca de documentos. Portanto, ele é um componente importante em um sistema de recuperação de informação, por cumprir o papel de determinar: quais termos podem ser usados no sistema; quais termos podem ser usados na busca para que esta tenha um resultado satisfatório; e permitir a introdução de novos termos em sua estrutura de termos e relações de modo a aproximar a linguagem do usuário à do sistema e realizar alterações de sentidos dos termos existentes (MOREIRA, 2003).

Neste trabalho foi extraído um segmento de elaborado um tesouro referente a recursos hídricos tendo como base as contribuições VGI do caso de estudo. Para esta tarefa, primeiro foi realizado um levantamento das palavras utilizadas nas descrições das contribuições voluntárias armazenadas na base de dados do sistema Gota D'Água. A partir da coleta das palavras foi determinado a entrada na forma de termos gerando duas listas: (i) termos que aparecem no título da contribuição; (ii) termos que se encontram na contribuição, descrevendo o problema relacionado à água. Além disso, realizou-se uma organização dos termos encontrados, classificando-os como termos relacionados à Região/Lugar, termos relacionados à Água e termos que dizem respeito à Qualidade/Característica. Após o levantamento e divisão dos termos em categorias, foi feita a definição de cada um deles com base em referências confiáveis.

A parte do tesouro obtida a partir de recortes feitos no Tesouro de Engenharia Sanitária e Ambiental (CEPIS, 2005), que compila 2.098 descritores e 3.037 não descritores. Este Tesouro foi revisado e atualizado, tendo como referência principal a base de dados bibliográficos da Rede Pan-americana de Informação sobre Saúde Ambiental (REPIDISCA) e as diferentes fontes de informação da Biblioteca Virtual de Saúde Ambiental (BVSA). O Tesouro Ambiental Multilíngüe Geral (GEMET), versão 2004, preparado pelo *Consiglio Nazires desonale delle Ricerche* (CNR) e a *Umweltbundesmt* (UBA), também foram utilizados como referências.

Na seção Semântica do Tesouro Base os descritores estão organizados por temas que formam campos semânticos segundo o esquema de classificação da REPIDISCA. Nesse esquema, o campo da engenharia sanitária e ciências do ambiente está dividido em 10 grandes grupos ou categorias de matéria (Tabela 6).

**Tabela 6: Categorias de matéria**

ID	Categoria
0	Saúde ambiental e saneamento
1	Engenharia ambiental e sanitária
2	Ciências aplicadas e do ambiente
3	Recursos hídricos e contaminação da água
4	Abastecimento de água
5	Águas residuais
6	Resíduos sólidos
7	Contaminação do solo
8	Contaminação do ar
9	Saúde ocupacional

As 10 categorias estão, por sua vez, subdivididas em temas mais específicos e cada uma é representada por um código de quatro dígitos. Isso justifica a numeração que aparece nos termos do tesauro gerado neste trabalho.

A seção semântica permite ao indexador ver os diferentes descritores agrupados sob um determinado tema para facilitar a escolha do descritor pertinente. Na seção alfabética os descritores são apresentados com suas relações preferenciais, hierárquicas e associativas. Os símbolos para representá-los no tesauro são os seguintes:

- NA (NOTA DE ALCANCE) - Define, explica ou limita o significado do descritor para finalidades de indexação;
- UP (USADO PARA) - Indica os sinônimos ou não descritores; eles não são válidos para a indexação;
- TG (TERMO GENÉRICO) - Assinala o termo mais amplo ao qual pertence o descritor;
- TE (TERMO ESPECÍFICO) - Indica os termos específicos (tipos ou classes) associados ao descritor;
- TR (TERMO RELACIONADO) - Indica os termos associados semanticamente com o descritor.

Cada descritor base tem um código para localizá-lo na seção semântica ou classificada. Como pode ser visto no Quadro 1.

**Quadro 1: Exemplo de descritor encontrado no tesauro (CEPIS, 2005).**

Recursos hídricos 3200

**NA** O volume de água superficial ou subterrânea disponível para qualquer uso em uma região específica.

**UP** Afluentes

Massas de água

Fontes da água

Recursos hidráulicos

Recursos hidrológicos

Recursos limnológicos

**TG** Recursos naturais [Saúde Ambiental]

**NT** Águas subterrâneas

Água superficial

Recursos hídricos transfronteiriços

**RT** Abastecimento de água

Bacias

Planejamento de recursos hídricos

Política da água

Recursos marinhos

#### 5.4. Processo para recuperação de VGIs relacionadas com base em tesouro

Como mostrado na Seção 5.3, um tesouro é um recurso que por relacionar e restringir sentidos auxilia num enriquecimento semântico. Portanto, a semântica embutida em um tesouro pode ser explorada por diversas aplicações para obterem resultados mais precisos e completos semanticamente. Para exemplificar como que a semântica presente no tesouro construído a partir de contribuições VGI pode ser utilizado por aplicações, este trabalho propõe um processo que retorna contribuições VGI semanticamente semelhantes com base em uma consulta fornecida por um usuário.

A Figura 6 apresenta o processo proposto. Um usuário fornece uma consulta com termos que ele deseja que as contribuições VGI possuam. Por exemplo, dada a seguinte consulta "Vazamentos próximos a lojas", o usuário deseja todas as contribuições que relatam vazamentos que ocorreram pertos de PoIs do tipo *Shop*. A consulta do usuário é processada no módulo Processamento de Consultas, o qual aplica diversas técnicas de processamento de linguagem natural (e.g., tokenização, *Part-of-Speech*, *Shallow Parsing*) de acordo com a necessidade da aplicação. A consulta processada é utilizada tanto no sub-módulo Comparar c/ Tesouro quanto no sub-módulo Comparar c/ VGI enriquecidas. Ambos os módulos comparam os termos identificados na consulta com os tipos de dados a que o sub-módulo se refere. Para agilizar o processo, ambos utilizam índices. Por fim, as consultas recuperadas por cada sub-módulo são comparadas. As contribuições presentes em ambos os conjuntos são consideradas semelhantes e são retornadas para o usuário.

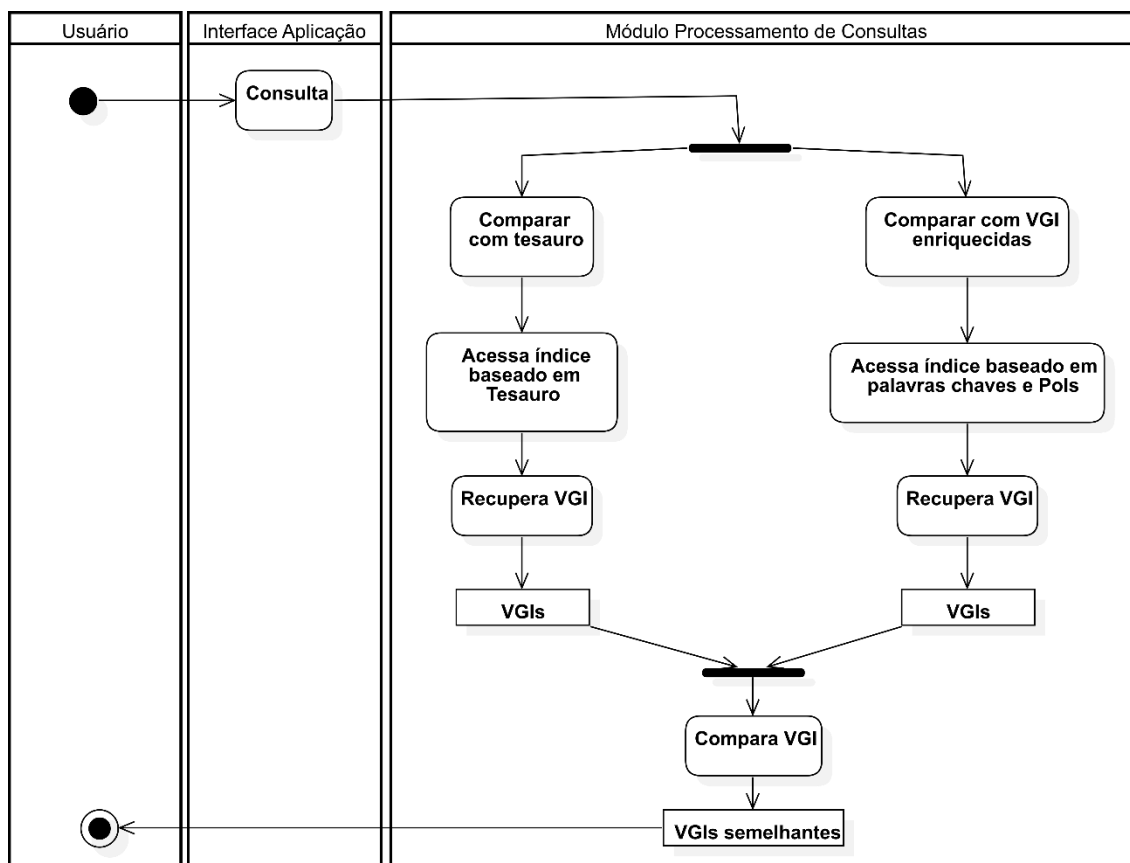


Figura 6: Processo pra retornar contribuições VGI semanticamente semelhantes com base em uma consulta fornecida por um usuário

## 6 DISCUSSÃO

O sistema VGI Gota D'Água, utilizado neste trabalho como caso de estudo, possibilita que usuários de diferentes partes do Brasil possam contribuir fornecendo informação sobre diferentes tipos de problemas relacionados à falta e/ou desperdício de água.

Normalmente uma VGI possui pouco contexto atrelado a ela, devido ao fato dos usuários usarem mensagens pequenas por diversos motivos, como falta de tempo, comodidade, dificuldade em utilizar o sistema, etc.

Segundo Guo et al.(2013), ao trabalhar com enriquecimento semântico com pouco conteúdo textual ou multimídia, como postagens de mídias sociais, os resultados apresentam pouca precisão devido à falta de contexto. Acredita-se que tal fato ocorra também com a VGI que normalmente possui uma contribuição sucinta. Devido às descrições das contribuições do usuário no sistema Gota D'Água serem pouco detalhadas, o objetivo deste trabalho não é ser mais preciso que as abordagens existentes da literatura

e sim demonstrar a viabilidade e aplicabilidade do enriquecimento semântico automático de VGI.

Com a geração de anotações semânticas em VGI, adotando o padrão RDF, é possível realizar perguntas a serem respondidas pelo sistema por meio da linguagem SPARQL, fazendo com que o sistema crie inferências lógicas. A linguagem SPARQL possibilita que uma consulta possa ser usada para formular questões que variam de padrão gráfico simples de correspondência até consultas complexas envolvendo diversos repositórios de RDF distribuídos na Web (W3C, 2013).

Apesar da linguagem SPARQL ser equivalente à SQL, ou seja, é possível traduzir qualquer expressão SPARQL em uma expressão SQL sem perda de semântica (CHEBOTKO; LU; FOTOUHI, (2009); ELLIOTT et al., (2009)), expressar consultas semânticas em SPARQL é mais amigável que expressar a mesma consulta em SQL. Além disso, os SGBDs que possuem suporte à SPARQL possuem módulos de otimização de consulta específicos para consultas semânticas.

A simplicidade do método proposto neste trabalho para enriquecer semanticamente contribuições VGI permite que o mesmo seja implementado em uma variedade de aplicações de diversos domínios, podendo ser adaptado de acordo com as necessidades de cada domínio. Além disso, o método proposto visa utilizar operações geoespaciais já existentes em banco de dados, pois tais bancos implementam operações geoespaciais otimizadas para grandes conjuntos de dados, auxiliando a eficiência do método proposto.

Para avaliar o método descrito na Seção 4, os seguintes passos foram executados: formulação de perguntas relacionadas ao tema distribuição de água, cujas respostas são obtidas via consultas em SPARQL. As consultas SPARQL são executadas com base nos dados do sistema Gota D'Água triplificados e nas anotações semânticas em RDF geradas pelo enriquecimento semântico. Um pequeno número de contribuições é examinado e anotado manualmente. O método proposto gera anotações semelhantes, já sendo esperado certo nível de imprecisão. Após a validação do método de enriquecimento semântico, as consultas são executadas de forma a tentar responder as questões previamente propostas. Os dados obtidos são transcritos nos resultados da pesquisa na Seção 5.2.

A maior dificuldade da pesquisa é lidar com descrição sucinta que as contribuições VGI possuem. Por exemplo, no sistema Gota D'Água, muitas contribuições não possuem nenhum tipo de comentário atrelado a elas, contanto somente com o tipo, a localização geográfica e o *timestamp* da contribuição. Tal falta de detalhes impossibilita o uso de

ferramentas como *Named Entity Recognition* (NER) / *Named Entity Disambiguation* (NED) e outros métodos de enriquecimento semântico que possam vir a gerar novas anotações semânticas e, conseqüentemente, contribuir nos resultados das consultas a serem executadas. Além disso, devido ao número limitado de contribuições, mensagens com pouco contexto e pouco número de atributos por contribuição, as consultas possíveis para demonstrar a aplicabilidade do enriquecimento semântico possuem escopo limitado. No entanto, as consultas apresentadas neste trabalho mostram a utilidade do enriquecimento semântico e permitem responder questionamentos úteis aos órgãos responsáveis pelo abastecimento de água.

## 7 CONCLUSÃO E TRABALHOS FUTUROS

*Linked Data* em conjunto com o padrão RDF é utilizado para integração de dados, simplificando os esquemas de relacionamento e enriquecendo a semântica. Embora sua utilização pareça vantajosa, a tecnologia é relativamente nova e existem diversos desafios a serem explorados e áreas de atuação a serem descobertas (AZEVEDO, 2014). Este artigo descreve os resultados de um projeto de pesquisa cujo objetivo foi verificar a aplicabilidade de *Linked Data* ao enriquecimento semântico de VGI.

A importância dos dados voluntários vem aumentando, no entanto, muitas vezes tais dados possuem poucas informações descritivas atreladas a eles. Esta pesquisa se propôs adicionar mais informações a dados voluntários (VGI) de maneira automática e manual, sendo que os recursos utilizados para o enriquecimento semântico são provenientes de repositórios (LOD). Como resultado da pesquisa, tem-se um conjunto de VGI com anotações que apontam para LOD, o qual permite que consultas semânticas e/ou mais complexas possam ser realizadas sobre as contribuições. Tem-se também um tesouro, que foi utilizado com intuito de obter resultados mais precisos e completos semanticamente, já que a semântica embutida em um tesouro pode ser explorada por diversas aplicações.

Como trabalhos futuros pode-se desenvolver o processo proposto utilizando o tesouro, que retorna contribuições VGI semanticamente semelhantes com base em uma consulta fornecida por um usuário, a fim de mostrar como que a semântica presente no tesouro que foi usado, a partir de contribuições VGI, pode ser utilizada por aplicações. Outra abordagem é a construção de uma ontologia, utilizando como ponto de partida a ontologia presentes nos recursos LOD. Esta por sua vez será útil na descrição e relações

entre VGI, visto que uma ontologia exige registro do conhecimento do domínio em uma linguagem que possa ser processada pelo computador para a realização de inferências.

### **AGRADECIMENTOS**

Este projeto foi parcialmente financiado com recursos das agências FAPEMIG e CAPES, com apoio da empresa Companhia Energética de Minas Gerais (CEMIG).

### 3 CONCLUSÕES GERAIS E TRABALHOS FUTUROS

Um conjunto de tecnologias e boas práticas da Web Semântica tem sido adotado por um número crescente de provedores de dados para publicar e interligar dados de diversas naturezas, utilizando os princípios de Linked Data. Tais princípios podem ser aplicados em contribuições VGI para facilitar sua descoberta, acesso e utilização. Portanto, com base nos estudos e pesquisas realizadas, esse trabalho propôs um processo de enriquecimento semântico automático e manual de contribuições VGI.

A abordagem proposta tem como principais contribuições possibilitar a geração de uma estrutura adequada de dados que melhor descreva as VGIs e auxiliar o usuário na geração de VGI semanticamente enriquecidas. Além disso, a abordagem proposta é simples o suficiente para ser utilizada em uma miríade de aplicações em diversos domínios de conhecimento, além de utilizar método de operações geoespaciais já eficientemente implementadas em bancos de dados. Para avaliação da abordagem proposta, algumas análises semânticas em SPARQL foram propostas, executadas e discutidas.

Ao final do processo de enriquecimento semântico, é obtido um conjunto de VGI enriquecidas descritas por meio de anotações semânticas. Com a nova coleção de dados, é possível responder perguntas de forma mais precisa. Um conjunto de VGI enriquecidas deve apresentar algumas características como: (i) maior quantidade de atributos semânticos, que contenham um significado claro, capaz de descrever melhor os dados; (ii) facilidade de interpretação e processamento do conteúdo dos conjuntos de dados, uma vez que a VGI estará em linguagem estruturada, i.e., definida de forma que possa ser processada por máquina; (iii) termos de vocabulários padrões associados, que permitirá a VGI se integrar a outros dados e recursos na Web.

Como resultado final, considera-se que a abordagem para Enriquecimento Semântico de VGI está de acordo com a proposta sugerida neste trabalho, pois permitem que os mesmos sejam publicados em conjuntos de dados ligados abertos com semânticas bem-descritas, facilitando, portanto, a sua descoberta, acesso e utilização.

Por fim, dentre as sugestões de trabalhos futuros a serem desenvolvidos dentro do tema desta dissertação, destacam-se:

a) desenvolver o processo proposto utilizando o tesouro, que retorna contribuições VGI semanticamente semelhantes com base em uma consulta fornecida por um usuário, a fim de mostrar como que a semântica presente no tesouro pode ser utilizada por aplicações;

b) estruturação de uma ontologia, fundamentada em uma ontologia de alto nível, de forma a melhor integrar as contribuições VGI com diferentes domínios de conhecimento. Dependendo de como a ontologia for estruturada, é possível utilizar termos mais abrangentes as serem utilizados em consultas à conceitos existentes na ontologia, possibilitando, por exemplo, considerar diferentes granularidades espaciais para a categoria do objeto espacial.

## REFERÊNCIAS BIBLIOGRÁFICAS

- AZEVEDO, P. C. N. **Uma proposta para visualização de linked data sobre enchentes na bacia do Rio Doce**. Dissertação de Mestrado Profissional, Universidade FUMEC, Belo Horizonte, 2014.
- BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked Data - the story so far. **International Journal on Semantic Web and Information Systems**, v. 5, n. 3, p. 01-22, 2009.
- BALLATORE, A; BERTOLOTTO, M. Semantically Enriching VGI in Support of Implicit Feedback Analysis. In: International Symposium on Web and Wireless Geographical Information Systems - **W2GIS**. Springer Berlin Heidelberg, 2011. p. 78-93.
- BERNERS-LEE, T., CHEN, Y., CHILTON, L., CONNOLLY, D., DHANARAJ, R., HOLLENBACH, J., ... & SHEETS, D. Tabulator: Exploring and analyzing linked data on the semantic web. In: **Proceedings of the 3rd international semantic web user interaction workshop**. 2006. p. 159.
- BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked Data - the story so far. **International Journal on Semantic Web and Information Systems**, v. 5, n. 3, p. 01-22, 2009.
- BROVELLI, M. A., MINGHINI, M., MOLINARI, M., & MOONEY, P. Towards an automated comparison of OpenStreetMap with authoritative road datasets. **Transactions in GIS**, v. 21, n. 2, p. 191-206, 2017.
- CEPIS. **Tesouro de engenharia sanitária e ambiental**. Lima; OPS / BVSDE; 296 p. (2005)
- CLARKE, C. A resource list management tool for undergraduate students based on linked open data principles. **The Semantic Web: Research and Applications**, p. 697-707, 2009.
- CLARKE, M.; HARLEY, P. How Smart Is Your Content? Using Semantic Enrichment to Improve Your User Experience and Your Bottom Line. **Science Editor**, v. 37, n. 2, 2014.
- COOPER, A. K., COETZEE, S., KACZMAREK, I., KOURIE, D. G., IWANIAK, A., KUBIK, T.: Challenges for quality in volunteered geographical information. In: **Proceedings of the AfricaGEO 2011 Conference**, Cape Town, South Africa, 2011.
- DERCZYNSKI, L., MAYNARD, D., RIZZO, G., VAN ERP, M., GORRELL, G., TRONCY, R., BONTCHEVA, K.: Analysis of named entity recognition and linking for tweets. **Information Processing & Management**, v. 51, n. 2, p. 32-49, 2015.
- DIMA, C. Intui2: A Prototype System for Question Answering over Linked Data. In **CLEF (Working Notes)**. 2013.
- ELLIOTT, B., CHENG, E., THOMAS-OGBUJI, C., & OZSOYOGLU, Z. M.: A complete translation from SPARQL into efficient SQL. In **Proceedings of the 2009**

- International Database Engineering & Applications Symposium**. ACM, 2009. p. 31-42.
- ELWOOD, S., GOODCHILD, M.F., SUI, DZ.: Researching volunteered geographic information: Spatial data, geographic research, and new social practice. **Annals of the Association of American Geographers**, v. 102, n. 3, p. 571-590, 2012.
- FAN, H., YANG, B., ZIPF, A., & ROUSELL, A. A polygon-based approach for matching OpenStreetMap road networks with regional transit authority data. **International Journal of Geographical Information Science**, v. 30, n. 4, p. 748-764, 2016.
- FLANAGIN, A. J., & METZGER, M. J. The credibility of volunteered geographic information. **GeoJournal**, v. 72, n. 3-4, p. 137-148, 2008.
- FOODY, G. M., SEE, L., FRITZ, S., VAN DER VELDE, M., PERGER, C., SCHILL, C., & BOYD, D. S. (2013). Assessing the accuracy of volunteered geographic information arising from multiple contributors to an internet based collaborative project. **Transactions in GIS**, v. 17, n. 6, p. 847-860, 2013.
- GAO, H., BARBIER, G., GOOLSBY, R.: Harnessing the crowdsourcing power of social media for disaster relief. **IEEE Intelligent Systems**, v. 26, n. 3, p. 10-14, 2011.
- GOODCHILD, M.F.: Citizens as voluntary sensors: spatial data infrastructure in the world of Web 2.0. **International Journal of Spatial Data Infrastructures Research**, v.2, p.4-32, 2007.
- GOODCHILD, M. F. Citizens as sensors: the world of volunteered geography. **GeoJournal**, v. 69, n. 4, p. 211-221, 2007.
- GOODCHILD, M. F., & LI, L. Assuring the quality of volunteered geographic information. **Spatial statistics**, v. 1, p. 110-120, 2012.
- GOMES, H. E. Manual de elaboração de tesouros monolíngues. Programa Nacional de Bibliotecas das Instituições de Ensino Superior. **Ciência da Informação**, v. 19, n. 2, 1990.
- GUO, S., CHANG, M. W., KICIMAN, E.: To Link or Not to Link? A Study on End-to-End Tweet Entity Linking. In **HLT-NAACL**, p. 1020-1030, 2013.
- GUO, W., LI, H., JI, H., & DIAB, M. T.: Linking Tweets to News: A Framework to Enrich Short Text Data in Social Media. In **ACL (1)**, p. 239- 249, 2013.
- HAKLAY, M. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. **Environment and Planning B: Planning and Design**, v. 37, n. 4, p. 682-703, 2010.
- HAKLAY, M., & WEBER, P. Openstreetmap: User-generated street maps. **IEEE Pervasive Computing**, v. 7, n. 4, p. 12-18, 2008.
- HAN, X., ZHAO, J. Named entity disambiguation by leveraging wikipedia semantic knowledge. In: **Proceedings of the 18th ACM conference on Information and knowledge management**. ACM. p. 215-224, 2009.

- HARTIG, O.: Provenance information in the web of data. **LDOW**, v. 538, 2009.
- HEATH, T., BIZER, C.: Linked Data: evolving the Web into a Global Data Space  
**Synthesis lectures on the semantic web: theory and technology**, v. 1, n. 1, p. 1-136, 2011.
- IBM. **Using semantic enrichment to enhance big data solutions**. 2016. Disponível em:  
<<http://www-01.ibm.com/software/ebusiness/jstart/semantic/>>
- LIRA, M. A. B. **Uma Abordagem Para Enriquecimento Semântico de Metadados Para Publicação de Dados Abertos**. Dissertação de mestrado, Universidade Federal de Pernambuco, Recife, 2014.
- MANNENS, E., TRONCY, R., BRAECKMAN, K., VAN DEURSEN, D., VAN LANCKER, W., DE SUTTER, R., & VAN DE WALLE, R. Automatic metadata enrichment in news production. In: **Image Analysis for Multimedia Interactive Services, 2009. WIAMIS'09. 10th Workshop on**. IEEE, p. 61-64, 2009.
- MENDONÇA, R. R., **Uma abordagem para coleta e publicação de dados de proveniência no contexto de Linked Data**. Dissertação de mestrado, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2013.
- MOREIRA, A. **Tesauros e ontologias: estudo de definições presentes na literatura das áreas das ciencias da computação e da informação, utilizando-se método analítico-sintético**. Dissertação de mestrado, Universidade Federal de Minas Gerais, Belo Horizonte, 2003.
- MOREIRA, J. D. C., NETO, F. M. M., DA COSTA, A. A. L., SOMBRA, E. L., DE ALIANÇA NETO, A. S., & DE MEDEIROS VALENTIM, R. A. Um sistema de enriquecimento semântico de perfil de usuário baseado em traços digitais para apoio à aprendizagem informal no contexto da saúde. **RENOTE**, v. 12, n. 1. 2014.
- MOREIRA, M. P., & MOURA, M. A. Construindo tesauros a partir de tesauros existentes: a experiência do TCI-Tesouro em Ciência da Informação. **DataGramZero**, Rio de Janeiro, v. 7, n. 4, p. 0-1001, 2009.
- MOURA, T.H.V.M, DAVIS JR., C. **Linked Geospatial Data: desafios e oportunidades de pesquisa**. André Santanche and Pedro R. Andrade (Eds.), p. 13, 2013.
- MUMMIDI, L.N., KRUMM, J. Discovering points of interest from users' map annotations. **GeoJournal**, v. 72, n. 3-4, p. 215-227, 2008.
- NEIS, P., ZIELSTRA, D. Recent developments and future trends in volunteered geographic information research: The case of OpenStreetMap. **Future Internet**, v. 6, n. 1, p. 76-106, 2014.
- RONZHIN, S.: **Semantic enrichment of Volunteered Geographic Information using Linked Data: a use case scenario for disaster management**. University of Twente, Dissertação de mestrado, 2015.

- PALETTA, F. C; MUCHERONI, M. L. O desenvolvimento da WEB 3.0: Linked Data e DBPEDIA. **Prisma.com**, n. 25, 2017.
- PÉREZ, J., ARENAS, M., & GUTIERREZ, C. Semantics and Complexity of SPARQL. In: **International semantic web conference**. Springer, Berlin, Heidelberg. p. 30-43, 2006
- PRUD, E., & SEABORNE, A. **SPARQL query language for RDF**. (2006)
- SALES, R. D., & CAFÉ, L. M. A. **Os tesouros e as ontologias sob um olhar terminológico**. 2013.
- SANTANCH, A. Aplicações educacionais na Web—o papel de RDF e Metadados. **XIV Simpósio Brasileiro de Informática na Educação—Mini-Cursos**, p. 127-152, 2003.
- SAVELYEV, A., XU, S., JANOWICZ, K., MÜLLIGANN, C., THATCHER, J., & LUO, W. Volunteered geographic services: Developing a linked data driven location-based service. In: **Proceedings of the 1st ACM SIGSPATIAL International Workshop on Spatial Semantics and Ontologies**. ACM, p. 25-31, 2011.
- SCHADE, S., GRANELL, C., DÍAZ, L. Augmenting SDI with Linked Data. In: **Workshop On Linked Spatiotemporal Data, in conjunction with the 6th International Conference on Geographic Information Science (GIScience 2010)**. Zurich, 14th September. 2010.
- SIRIN, E., & PARSIA, B. SPARQL-DL: SPARQL Query for OWL-DL. In: **OWLED v. 258**. 2007.
- SORRENTINO, S., BERGAMASCHI, S., FUSARI, E., & BENEVENTANO, D. Semantic annotation and publication of linked open data. In **International Conference on Computational Science and Its Applications**. Springer, Berlin, Heidelberg, p. 462-474, 2013.
- W3C. **Sparql 1.1 Overview**. W3C, 2013. Disponível em: <https://www.w3.org/TR/sparql11-overview/> Acesso em: Maio, 2017.
- STADLER, C., LEHMANN, J., HÖFFNER, K., AUER, S. Linkedgeodata: A core for a web of spatial open data. **Semantic Web**, v. 3, n. 4, p. 333-354, 2012.
- VIEIRA, R., DOS SANTOS, D. A., DA SILVA, D. M., & SANTANA, M. R. Web semântica: ontologias, lógica de descrição e inferência. **Web e Multimídia: Desafios e Soluções (WebMedia 2005-Minicursos)**, v. 1, p. 127-167, 2005.
- WHITE, C. M. **Social media, crisis communication, and emergency management: Leveraging Web 2.0 technologies**. CRC press, 2011.
- YAMADA, I., TAKEDA, H., TAKEFUJI, Y.: An end-to-end entity linking approach for tweets. In: **5th Workshop on Making Sense of Microposts: Big Things Come in Small Packages, # Microposts 2015, at the 24th International Conference on the World Wide Web, WWW 2015**. CEUR-WS, 2015.

ZIELSTRA, D.; ZIPF, A. A comparative study of proprietary geodata and volunteered geographic information for Germany. **Proceedings of the International Conference On Geographic Information Science (AGILE)**, Guimarães, Portugal, May 11-14, 2010; Springer-Verlag: Berlin, Germany. p. 15. 2010.

ZOOK, M., GRAHAM, M., SHELTON, T., & GORMAN, S. Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. **World Medical & Health Policy**, v. 2, n. 2, p. 7-33, 2010.

## APÊNDICE A

### **Levantamento das palavras da base de dados do sistema Gota D'Água**

O levantamento das palavras utilizadas nas descrições das contribuições VGI foi feito na base de dados do sistema Gota D'Água. A partir da coleta das palavras, foi determinada a entrada na forma de termos gerando duas listas: (i) termos que aparecem no título da contribuição; (ii) termos que se encontram na descrição da contribuição, descrevendo o problema relacionado à água. Além disso, realizou-se uma organização dos termos encontrados, classificando-os como termos relacionados à Região/Lugar, termos relacionados à Água e termos que dizem respeito à Qualidade/Característica. Após o levantamento e divisão dos termos em categorias, foi feita a definição de cada um deles com base em referências confiáveis. As palavras destacadas em negrito são palavras que se encontram no tesauro utilizado na pesquisa.

# 1 LEVANTAMENTO DAS PALAVRAS DA BASE DE DADOS DO SISTEMA GOTA D'AGUA

## 1.1 Termos encontrados no título da contribuição

Lista de termos relacionada à Região/Lugar

	<b>Termo</b>	<b>Notação de Definição – ND</b>
1	Apartamento	<b>ND:</b> é uma unidade habitacional existente em edifícios multifamiliares e em conjuntos habitacionais. (Babelfy - <a href="http://babelfy.org/">http://babelfy.org/</a> )
2	Asa Norte	<b>ND:</b> é um bairro da região administrativa de Brasília, no Distrito Federal, sendo uma área tombada pela UNESCO. Região nobre de Brasília, a Asa Norte possui um arranjo urbano estruturado em conceitos de cidade jardim, modernista e concretista. (Wikipédia - <a href="https://www.wikipedia.org/">https://www.wikipedia.org/</a> )
3	Asa Sul	<b>ND:</b> é um bairro da região administrativa de Brasília, no Distrito Federal, sendo uma área tombada pela UNESCO. (Wikipédia)
4	Calçada	<b>ND:</b> Passeio empedrado, atijolado ou cimentado para trânsito de pedestres e geralmente ao longo das casas. (Michaelis - <a href="http://michaelis.uol.com.br/">http://michaelis.uol.com.br/</a> )
5	Casa	<b>ND:</b> Construção destinada à moradia. (Michaelis)
6	COPASA	<b>ND:</b> Companhia de Saneamento de Minas Gerais – COPASA MG é uma empresa de economia mista. Sua principal atividade é a prestação de serviços em abastecimento de água, esgotamento sanitário e resíduos sólidos. (COPASA - <a href="http://www.copasa.com.br/wps/portal/internet/a-copasa/a-empresa">http://www.copasa.com.br/wps/portal/internet/a-copasa/a-empresa</a> )
7	<b>Edifício</b>	<b>ND:</b> Construção de certa importância, geralmente de alvenaria, que serve de habitação ou para diversas outras atividades. (Michaelis)
8	Estabelecimento	<b>ND:</b> Instituição pública ou privada. (Michaelis)
9	Estacionamento	<b>ND:</b> Lugar onde se estaciona qualquer veículo. Área em que podem ser deixados veículos temporariamente; estacionamento. (Michaelis)
10	Garagem	<b>ND:</b> Lugar para abrigar veículos ou automóveis. (Michaelis)
11	Prédio	<b>ND:</b> Construção que contém uma série de apartamentos individuais. Propriedade imóvel, rural ou urbana. (Michaelis)

12	Quintal	<b>ND:</b> Terreno nos fundos de uma casa, geralmente com jardim, horta ou árvores frutíferas, ou ao lado dela. (Michaelis)
13	<b>Rua</b>	<b>ND:</b> Caminho público em uma cidade, vila etc., ladeado por casas, prédios ou muros. (Michaelis)
14	Vias	<b>ND:</b> Caminho ou estrada por onde se vai de um ponto a outro. Qualquer obra viária (avenida, rua, estrada etc.) que serve para ligar localidades urbanas ou uma cidade a outra, usada para o trânsito de veículos e pedestres. (Michaelis)

Lista de termos relacionada à água

	<b>Termo</b>	<b>Notação de Definição – ND</b>
1	<b>Água</b>	<b>ND:</b> denominação comum do monóxido de hidrogênio, substância composta de um átomo de oxigênio e dois de hidrogênio. (Agência Nacional das Águas - ANA) <b>ND2:</b> fase líquida do monóxido de hidrogênio. (ANA)
2	Banheiro	<b>ND:</b> Cômodo de uma residência ou de uma empresa onde ficam instalados a banheira, o chuveiro, o lavatório, o vaso sanitário e o bidê. (Michaelis) <b>ND2:</b> Qualquer local com vaso sanitário e lavatório; sanitário, toailete. (Michaelis)
3	<b>Bomba</b>	<b>ND:</b> Máquina ou aparelho usado para aspirar ou movimentar fluidos, gasosos ou líquidos, em geral por meio de tubulação. (Michaelis)
4	Bueiro	<b>ND:</b> Abertura, cano ou rego por onde escoam águas. (Michaelis) <b>ND2:</b> Buraco em muro, ou cano subterrâneo em ruas ou estradas, para dar escoamento a águas subterrâneas ou pluviais, ou de rios e riachos. (Michaelis)
5	<b>Caixa d'água</b>	<b>ND:</b> Reservatório de água potável destinado ao abastecimento da população. (Michaelis) <b>ND2:</b> Depósito em que se faz provisão de água. (Michaelis)
6	<b>Cano</b>	<b>ND:</b> Tubo para condução de gases ou de líquidos. (Michaelis) <b>ND2:</b> Construção tubular subterrânea para condução de líquidos, gases ou dejetos. (Michaelis)
7	<b>Chuva</b>	<b>ND:</b> precipitação de água líquida seja sob a forma de gotas de diâmetro superior a 0.5mm, seja sob a forma de gotículas menores, largamente dispersas. (UNESCO, 1983, p. 125)
8	<b>Cloro</b>	<b>ND:</b> Elemento não metálico, univalente e polivalente, de número atômico 17, massa atômica 35,46, que pertence à família dos halogênios, usado principalmente como

		poderoso agente branqueador, oxidante e desinfetante, na purificação de água. (Michaelis)
9	<b>Descarga</b>	<b>ND:</b> Válvula ou bomba que regula a descarga de água em um vaso sanitário. (Michaelis)
10	<b>Desperdício</b>	<b>ND:</b> Ato de gastar em excesso; esbanjamento. (Michaelis) <b>ND2:</b> Falta de aproveitamento; perda. (Michaelis)
11	Desvio	<b>ND:</b> Mudança de direção ou da posição normal. (Michaelis) <b>ND2:</b> Caminho alternativo, fora da rota normal; atalho. (Michaelis)
12	<b>Distribuição</b>	<b>ND:</b> Ato ou efeito de distribuir; organização ou arranjo de algo, de acordo com um critério; classificação, disposição, ordenamento. (Michaelis)
13	<b>Efluentes</b>	<b>ND:</b> Substância líquida com predominância de água produzida pelas atividades humanas (esgotos domésticos, resíduos líquidos e gasosos das indústrias etc.) lançada na rede de esgotos ou nas águas receptoras (cursos d'água, lago ou aquífero), com ou sem tratamento e com a finalidade de utilizar essas águas receptoras no seu transporte e diluição. (ANA) <b>ND2:</b> Esgoto, água descartada, efluentes líquidos de edificações, indústrias, agroindústrias e agropecuária, tratados ou não. (ANA)
14	<b>Encanamento</b>	<b>ND:</b> Sistema de canais ou tubos para encanar águas, gases etc.; tubagem, tubulação. (Michaelis)
15	Escorrer	<b>ND:</b> Fazer correr ou esgotar (um líquido). (Michaelis) <b>ND2:</b> Cair em gotas ou correr em fios; gotejar, fluir. (Michaelis)
16	<b>Falta</b>	<b>ND:</b> Ato ou efeito de faltar; carência, privação, deficiência. (Michaelis) <b>ND2:</b> Inexistência, ausência ou privação de algo; deficiência ou supressão (de um bem, qualidade, faculdade, estado etc.). (Michaelis)
17	<b>Fornecimento</b>	<b>ND:</b> Ação ou efeito de fornecer; abastecimento. (Michaelis)
18	<b>Hidrante</b>	<b>ND:</b> Boca de cano de água, com válvula ou torneira, em via pública ou em prédios, para ligar mangueira usada em combate a incêndios. (Michaelis)
19	<b>Infiltração</b>	<b>ND:</b> Ato ou efeito de infiltrar(-se); ação de líquido ou fluido que se embebe ou penetra nos interstícios dos corpos sólidos. (Michaelis)

		<b>ND2:</b> Passagem da água da superfície do solo para o meio poroso. (ANA)
20	<b>Irrigação</b>	<b>ND:</b> operação agrícola que tem como objetivo suprir artificialmente a necessidade de água da planta, envolvendo a implantação de equipamentos e/ou estruturas e/ou execução de obras, dependendo do método de irrigação. (ANA)
21	Irigador	<b>ND:</b> Diz-se de ou recipiente para regar plantas e jardins; regador. (Michaelis)
22	Jorrar	<b>ND:</b> Sair em jorro, manar com força; brotar, golfar, rebentar. (Michaelis)  <b>ND2:</b> Deitar ou lançar em jorro; fazer sair com ímpeto. (Michaelis)
23	Lavagem	<b>ND:</b> Ato ou efeito de lavar(-se). (Michaelis)
24	Lavar	<b>ND:</b> limpar (algo) por ação de um líquido (esp. água), produto detergente, antisséptico etc.; eliminar as impurezas de (produto, material), submetendo-o a banhos. (Michaelis)
25	<b>Mangueira</b>	<b>ND:</b> Tubo flexível, de lona, borracha ou plástico para a condução de líquidos ou gases; borracha, manga. (Michaelis)
26	<b>Poluição da água</b>	<b>ND:</b> alteração das características naturais da água, tornando-a imprópria para o consumo, e que é provocada por agentes tóxicos que degradam e destroem a fauna e a flora. (Michaelis)
27	Queimada	<b>ND:</b> incêndio de mato, de arvoredo; queima propositada ou acidental de parte de uma floresta ou de um campo. (Michaelis)
28	<b>Racionamento</b>	<b>ND:</b> Ação ou efeito de racionar. (Michaelis)  <b>ND2:</b> Limitação da quantidade de gêneros que cada pessoa ou família pode comprar ou receber. (Michaelis)
29	<b>Rede</b>	<b>ND:</b> Qualquer estrutura ou conjunto que se assemelha a um sistema reticulado. (Michaelis)
30	Torneira	<b>ND:</b> Peça dotada de uma chave, acoplada a um cano, recipiente, tubo etc., usada para deixar sair ou reter o líquido ou o gás nela contido. (Michaelis)
31	<b>Tubulação</b>	<b>ND:</b> Sistema de distribuição de tubos para a passagem de água, gás ou eletricidade ou ainda para o escoamento de água servida e esgoto; canalização. (Michaelis)
32	<b>Vazamento</b>	<b>ND:</b> Ato ou efeito de vazar; vazadura, vazão. (Michaelis)

Lista de termos relacionada à Qualidade/Característica

	<b>Termo</b>	<b>Notação de Definição – ND</b>
1	Abastecido	<b>ND:</b> Que se abasteceu; provido do necessário. (Michaelis)
2	Constante	<b>ND:</b> Que se repete incessantemente; contínuo, incessante, permanente. (Michaelis)
3	Desnecessário	<b>ND:</b> Que não tem utilidade, supérfluo. (Michaelis)
4	Desregulado	<b>ND:</b> Que se desregulou; ao qual se rompeu a regulagem; desajustado. (Michaelis)
5	Doméstico	<b>ND:</b> Concernente à vida da família; familiar, caseiro, íntimo. (Michaelis)
6	Frequente	<b>ND:</b> Repetido muitas vezes; amiudado. (Michaelis)
7	Furado	<b>ND:</b> Que tem buraco ou furo; esburacado, perfurado. (Michaelis)
8	Ilegal	<b>ND:</b> Que contraria os preceitos ou as determinações da lei; sem legitimidade jurídica; ilícito. (Michaelis)
9	Indevido	<b>ND:</b> Que se mostra inconveniente. (Michaelis)
10	Industrial	<b>ND:</b> Que pertence ou se refere à indústria (Michaelis)
11	Lavado	<b>ND:</b> Limpo por banho ou lavagem; aseado. (Michaelis)
12	Movimentado	<b>ND:</b> Posto em movimento; motório. (Michaelis)  <b>ND2:</b> Em que há movimento considerável. (Michaelis)
13	Quebrado	<b>ND:</b> Que se quebrou; separado em dois ou mais pedaços; fragmentado, fraturado, partido. (Michaelis)  <b>ND2:</b> Fendido com ruptura; rachado, trincado. (Michaelis)

## 1.2 Termos encontrados na descrição dos problemas

Lista de termos relacionada à Região/Lugar

	<b>Termo</b>	<b>Notação de Definição – ND</b>
1	Alojamento	<b>ND:</b> Lugar onde se mora por período temporário; aposento, morada, pousada. (Michaelis)
2	Andar	<b>ND:</b> Cada um dos pavimentos superiores nas casas que têm mais de um pavimento; piso. (Michaelis)
3	Apartamento	<b>ND:</b> é uma unidade habitacional existente em edifícios multifamiliares e em conjuntos habitacionais. (Babelfy)
4	Asfalto	<b>ND:</b> Revestimento asfáltico. (Michaelis)
5	Avenida Atlântica	<b>ND:</b> é uma importante via do subúrbio da zona sul de São Paulo, Brasil. Fica situada nas redondezas da represa Guarapiranga. (Wikipédia)
6	Bueiro	<b>ND:</b> Abertura, cano ou rego por onde escoam águas. (Michaelis)

7	Cachoeiro do Itapemirim	<b>ND:</b> é um município brasileiro do estado do Espírito Santo. (Wikipédia)
8	Calçada	<b>ND:</b> Passeio empedrado, atijolado ou cimentado para trânsito de pedestres e geralmente ao longo das casas. (Michaelis)
9	Calçamento	<b>ND:</b> Revestimento de pedras, asfalto, paralelepípedo ou outro material usado para calçar rua ou qualquer outra área; pavimentação. (Michaelis)
10	Canteiro de obra	<b>ND:</b> Local da construção onde se armazenam os materiais (cimento, ferro, madeira etc.) e se realizam os serviços auxiliares durante a obra (preparação da argamassa, dobragem de ferro etc.). (Colégio de arquitetos) <a href="http://www.colegiodearquitetos.com.br/dicionario/2009/02/o-que-e-canteiro-de-obra/">http://www.colegiodearquitetos.com.br/dicionario/2009/02/o-que-e-canteiro-de-obra/</a>
11	Casa	<b>ND:</b> Construção destinada à moradia. (Michaelis)
12	Catedral	<b>ND:</b> Diz-se da igreja principal de uma diocese, sede de um bispado ou arcebispado, onde se encontra a cadeira episcopal. (Michaelis)
13	<b>Cidade</b>	<b>ND:</b> Grande aglomeração de pessoas em um área geográfica circunscrita, com inúmeras edificações, que desenvolve atividades sociais, econômicas, industriais, comerciais, culturais, administrativas etc, (Michaelis)  <b>ND2:</b> Sede municipal, onde se concentram as atividades administrativas. (Michaelis)
14	<b>Cisterna</b>	<b>ND:</b> Reservatório de águas pluviais, abaixo do nível da terra. (Michaelis)
15	COPASA	<b>ND:</b> Companhia de Saneamento de Minas Gerais – COPASA MG é uma empresa de economia mista. Sua principal atividade é a prestação de serviços em abastecimento de água, esgotamento sanitário e resíduos sólidos. (COPASA - <a href="http://www.copasa.com.br/wps/portal/internet/a-copasa/a-empresa">http://www.copasa.com.br/wps/portal/internet/a-copasa/a-empresa</a> )
16	Córrego	<b>ND:</b> Sulco aberto pelas águas correntes; corga, regueiro. (Michaelis)  <b>ND2:</b> Pequeno rio, pouco profundo e de pequeno caudal. (Michaelis)
17	Departamento de Engenharia Civil	<b>ND:</b> Criado em 1978, o Departamento de Engenharia Civil da Universidade Federal de Viçosa oferece hoje os cursos de graduação em Engenharia de Agrimensura, Engenharia Civil e Engenharia Ambiental, além do Programa de Pós-Graduação em Engenharia Civil nas áreas de concentração: Geotecnia, Saneamento e Ambiental, Informações Espaciais e Engenharia da Construção. ( <a href="http://www.dec.ufv.br/">http://www.dec.ufv.br/</a> )

18	Detran	<b>ND:</b> Departamento Estadual de Trânsito. Órgão executivo que, entre outras atribuições, administra a documentação dos veículos e condutores, emite, suspende ou cassa a CNH. ( <a href="https://www.detran.mg.gov.br/duvidas-frequentes/glossario">https://www.detran.mg.gov.br/duvidas-frequentes/glossario</a> )
19	DPF	<b>ND:</b> Departamento de Física onde oferece o curso de Física, em tempo integral, na modalidade de bacharelado. ( <a href="http://www.dpf.ufv.br/?page_id=7">http://www.dpf.ufv.br/?page_id=7</a> )
20	<b>Edifício</b>	<b>ND:</b> Construção de certa importância, geralmente de alvenaria, que serve de habitação ou para diversas outras atividades. (Michaelis)
21	Empresa	<b>ND:</b> Sociedade organizada para a exploração de indústria ou comércio; com a finalidade de obter um rendimento monetário através da produção de bens ou de serviços. (Michaelis)
22	Estabelecimento	<b>ND:</b> Instituição pública ou privada. (Michaelis)
23	Estacionamento	<b>ND:</b> Lugar onde se estaciona qualquer veículo. Área em que podem ser deixados veículos temporariamente; estacionamento. (Michaelis)
24	Garagem	<b>ND:</b> Lugar para abrigar veículos ou automóveis. (Michaelis)
25	Igreja Presbiteriana	<b>ND:</b> refere-se às igrejas cristãs protestantes que aderem à tradição teológica reformada (calvinismo) e cuja forma de organização eclesial se caracteriza pelo governo de uma assembleia de presbíteros, ou anciãos. (Wikipédia)  <b>ND2:</b> A Igreja Presbiteriana do Brasil é uma federação de igrejas que têm em comum uma história, uma forma de governo, uma teologia, bem como um padrão de culto e de vida comunitária. Historicamente, a IPB pertence à família das igrejas reformadas ao redor do mundo, tendo surgido no Brasil em 1859, como fruto do trabalho missionário da Igreja Presbiteriana dos Estados Unidos. ( <a href="http://www.ipb.org.br/ipb/historia">http://www.ipb.org.br/ipb/historia</a> )
26	Jardim Bom Clima	<b>ND:</b> Jardim Bom Clima, ou simplesmente Bom Clima, é um bairro situado na zona nordeste do município brasileiro de Juiz de Fora. (Wikipédia)
27	Lavanderia	<b>ND:</b> A parte da casa, hotel, caserna etc. onde a roupa é lavada e passada a ferro. (Michaelis)
28	Motel	<b>ND:</b> Estabelecimento hoteleiro, à beira de estradas movimentadas, que aluga quartos para hóspedes, oferece estacionamento e geralmente possui restaurante. (Michaelis)  <b>ND2:</b> Hotel que aluga quartos com o objetivo de favorecer encontros amorosos, geralmente por poucas horas; hotel de alta rotatividade. (Michaelis)

29	<b>Município</b>	<b>ND:</b> Circunscrição territorial administrada nos seus próprios interesses por um prefeito, à frente da Prefeitura, que executa as leis emanadas do corpo de vereadores eleitos pelo povo, que compõem a Câmara Municipal. (Michaelis)
30	Museu	<b>ND:</b> Instituição onde se expõem obras de arte e objetos de cunho científico ou histórico. (Michaelis)
31	Parque ecológico	<b>ND:</b> um território que se caracteriza pelo cuidado especial que recebem as espécies que habitam no mesmo. É hábito a gestão destes espaços está a cargo do Estado.  <a href="http://conceito.de/parque-ecologico#ixzz4oMn3kaXQ">http://conceito.de/parque-ecologico#ixzz4oMn3kaXQ</a>
32	Parque de exposições	<b>ND:</b> é o local onde se organizam eventos, palestras, feiras, shows ou congressos. Trata-se de grandes edifícios públicos com espaço bastante para acolher as empresas públicas e privadas em eventos sociais do município e áreas metropolitanas. (Wikipédia)
33	Pavimento	<b>ND:</b> Revestimento aplicado à superfície do solo onde se pisa; chão, piso. (Michaelis)
34	PH Rolfs	<b>ND:</b> Rua localizada no bairro de Campus - cidade de Viçosa MG (Michaelis)  <a href="http://www.consultarcep.com.br/mg/vicosa/campus/rua-p-h-rolfs/36570000">http://www.consultarcep.com.br/mg/vicosa/campus/rua-p-h-rolfs/36570000</a>
35	Praça	<b>ND:</b> Lugar aberto e espaçoso, em geral cercado de edifícios; largo (Michaelis)
36	<b>Prédio</b>	<b>ND:</b> Construção que contém uma série de apartamentos individuais. Propriedade imóvel, rural ou urbana. (Michaelis)
37	Quintal	<b>ND:</b> Terreno nos fundos de uma casa, geralmente com jardim, horta ou árvores frutíferas, ou ao lado dela. (Michaelis)
38	<b>Residência</b>	<b>ND:</b> Morada habitual em determinado local. (Michaelis)  <b>ND2:</b> Localidade onde uma pessoa reside. (Michaelis)
39	Ribeirão	<b>ND:</b> Curso de água maior que um regato, porém menor que um rio. (Michaelis)
40	Rio Doce	<b>ND:</b> é um curso de água da Região Sudeste do Brasil, que banha os estados de Minas Gerais e Espírito Santo. Com cerca de 853 km de extensão, seu curso representa a mais importante bacia hidrográfica totalmente incluída na Região Sudeste. (Wikipédia)
41	<b>Rua</b>	<b>ND:</b> Caminho público em uma cidade, vila etc., ladeado por casas, prédios ou muros. (Michaelis)
42	Rua Juquinha Brumado	<b>ND:</b> Rua localizada no bairro Santa Clara, em Viçosa-MG.

43	Rua Professor Carlos Socias	<b>ND:</b> Rua localizada no bairro Clélia Bernardes, em Viçosa-MG.
44	SAAE	<b>ND:</b> Serviço Autônomo de Água e Esgoto de Viçosa, SAAE, é uma autarquia municipal com personalidade jurídica de direito público, autonomia administrativa, econômico-financeira, técnica e patrimônio próprio. Tem como objetivo levar saúde e bem-estar para toda a população, por meio dos serviços prestados na área saneamento básico, fornecendo aos usuários água tratada, realizando o esgotamento sanitário e a gestão dos sistemas de coleta, tratamento, transbordo e dispensa final de resíduos sólidos.  <a href="http://www.saaevicosa.mg.gov.br/">http://www.saaevicosa.mg.gov.br/</a>
45	Sabesp	<b>ND:</b> A Sabesp (Companhia de Saneamento Básico do Estado de São Paulo) é uma empresa brasileira que detém a concessão dos serviços públicos de saneamento básico no Estado de São Paulo. Seu principal acionista é o governo do Estado de São Paulo, que controla a gestão da companhia. (Wikipédia)
46	Samarco	<b>ND:</b> é uma mineradora brasileira fundada em 1977 e atualmente controlada através de uma joint-venture entre a Vale S.A. e a anglo-australiana BHP Billiton, cada uma com 50% das ações da empresa. (Wikipédia)
47	São Joao Nepomuceno	<b>ND:</b> é um município brasileiro do estado de Minas Gerais. Fica localizado na microrregião de Juiz de Fora, por sua vez localizada na mesorregião da Zona da Mata mineira. (Wikipédia)
48	Sete Lagoas	<b>ND:</b> é um município brasileiro do estado de Minas Gerais. Grande polo industrial, localizado a aproximadamente 72 quilômetros de Belo Horizonte, possuía em julho de 2015 uma população estimada de 232.107 habitantes, segundo o IBGE. (Wikipédia)
49	Shopping	<b>ND:</b> é uma edificação que contém um conjunto de estabelecimentos de varejo de diferentes bens de consumo, além de prestação de serviços e lazer (lanchonetes, restaurantes, salas de cinema, teatro, playground etc.), constituindo-se em uma grande área comercial fechada, praticamente independente e isolada do seu entorno imediato, dotada de climatização, escadas rolantes, estacionamento e, eventualmente, atrações musicais e outras. (Wikipédia)
50	Teto	<b>ND:</b> Cobertura de um recinto fechado qualquer. (Michaelis)
51	Tinturaria	<b>ND:</b> Estabelecimento onde se tingem tecidos. (Michaelis)
52	UFV	<b>ND:</b> é uma universidade pública brasileira, com sua sede localizada na cidade de Viçosa, no estado de Minas Gerais. (Wikipédia)

Lista relacionada à Água

	<b>Termo</b>	<b>Notação de Definição – ND</b>
1	<b>Água</b>	<b>ND:</b> denominação comum do monóxido de hidrogênio, substância composta de um átomo de oxigênio e dois de hidrogênio. (ANA)  <b>ND2:</b> fase líquida do monóxido de hidrogênio. (ANA)
2	<b>Banheiro</b>	<b>ND:</b> Cômodo de uma residência ou de uma empresa onde ficam instalados a banheira, o chuveiro, o lavatório, o vaso sanitário e o bidê. (Michaelis)  <b>ND2:</b> Qualquer local com vaso sanitário e lavatório; sanitário, toalete. (Michaelis)
3	<b>Bomba</b>	<b>ND:</b> Máquina ou aparelho usado para aspirar ou movimentar fluidos, gasosos ou líquidos, em geral por meio de tubulação. (Michaelis)
4	<b>Caixa d'água</b>	<b>ND:</b> Reservatório de água potável destinado ao abastecimento da população. (Michaelis)  <b>ND2:</b> Depósito em que se faz provisão de água. (Michaelis)
5	<b>Cano</b>	<b>ND:</b> Tubo para condução de gases ou de líquidos. (Michaelis)  <b>ND2:</b> Construção tubular subterrânea para condução de líquidos, gases ou dejetos. (Michaelis)
6	<b>Chuva</b>	<b>ND:</b> precipitação de água líquida seja sob a forma de gotas de diâmetro superior a 0.5mm, seja sob a forma de gotículas menores, largamente dispersas. (UNESCO)
7	<b>Chuveiro</b>	<b>ND:</b> Aparelho dotado de chapa de metal com numerosos orifícios por onde sai a água canalizada; é usado para tomar banho. (Michaelis)
8	<b>Cloro</b>	<b>ND:</b> Elemento não metálico, univalente e polivalente, de número atômico 17, massa atômica 35,46, que pertence à família dos halogênios, usado principalmente como poderoso agente branqueador, oxidante e desinfetante, na purificação de água. (Michaelis)
9	<b>Curso d'água</b>	<b>ND:</b> é qualquer corpo de água fluente, como rios, córregos, riachos, regatos, ribeiros, entre outros. (Wikipedia)
10	<b>Descarga</b>	<b>ND:</b> Válvula ou bomba que regula a descarga de água em um vaso sanitário. (Michaelis)

11	<b>Desperdício</b>	<b>ND:</b> Ato de gastar em excesso; esbanjamento. (Michaelis) <b>ND2:</b> Falta de aproveitamento; perda. (Michaelis)
12	<b>Distribuição</b>	<b>ND:</b> Ato ou efeito de distribuir; organização ou arranjo de algo, de acordo com um critério; classificação, disposição, ordenamento. (Michaelis)
13	<b>Encanamento</b>	<b>ND:</b> Sistema de canais ou tubos para encanar águas, gases etc.; tubagem, tubulação. (Michaelis)
14	Escoar	Fazer correr ou escorrer (líquido), geralmente de forma lenta. (Michaelis)
15	<b>Escassez de água</b>	<b>ND:</b> falta ou insuficiência de água. (ANA)
16	<b>Esgoto doméstico</b>	<b>ND:</b> efluente líquido referente ao uso doméstico da água. Pode ser resultante das águas cloacais ou dos vasos sanitários e das águas resultantes de outros usos, tais como banho, preparo de alimentos e lavagens. (ANA)
17	<b>Falta</b>	<b>ND:</b> Ato ou efeito de faltar; carência, privação, deficiência. (Michaelis) <b>ND2:</b> Inexistência, ausência ou privação de algo; deficiência ou supressão (de um bem, qualidade, faculdade, estado etc.). (Michaelis)
18	Fluxo	<b>ND:</b> Escoamento ou movimento contínuo que segue um curso, uma corrente; descarga, vazão. (Michaelis)
19	<b>Fornecimento</b>	<b>ND:</b> Ação ou efeito de fornecer; abastecimento. (Michaelis)
20	<b>Hidrante</b>	<b>ND:</b> Boca de cano de água, com válvula ou torneira, em via pública ou em prédios, para ligar mangueira usada em combate a incêndios. (Michaelis)
21	<b>Incêndio</b>	<b>ND:</b> Fogo que se propaga com intensidade e geralmente causa prejuízos materiais. (Michaelis)
22	<b>Infiltração</b>	<b>ND:</b> Ato ou efeito de infiltrar(-se); ação de líquido ou fluido que se embebe ou penetra nos interstícios dos corpos sólidos. (Michaelis) <b>ND2:</b> Passagem da água da superfície do solo para o meio poroso. (ANA)
23	<b>Irrigação</b>	<b>ND:</b> operação agrícola que tem como objetivo suprir artificialmente a necessidade de água da planta, envolvendo a implantação de equipamentos e/ou estruturas e/ou execução de obras, dependendo do método de irrigação. (ANA)
24	Irrigador	<b>ND:</b> Diz-se de ou recipiente para regar plantas e jardins; regador. (Michaelis)
25	Lavagem	<b>ND:</b> Ato ou efeito de lavar(-se). (Michaelis)

26	Lavar	<b>ND:</b> limpar (algo) por ação de um líquido (esp. água), produto detergente, antisséptico etc.; eliminar as impurezas de (produto, material), submetendo-o a banhos. (Michaelis)
27	Litro	<b>ND:</b> Unidade de medida de capacidade correspondente ao volume de 1 dm <sup>3</sup> . (Michaelis)
28	<b>Mangueira</b>	<b>ND:</b> Tubo flexível, de lona, borracha ou plástico para a condução de líquidos ou gases; borracha, manga. (Michaelis)
29	<b>Mictório</b>	<b>ND:</b> Lugar onde se urina. (Michaelis)
30	<b>Potável</b>	<b>ND:</b> Diz-se de água livre de impurezas, sem substâncias tóxicas e cujos níveis de organismos patogênicos são seguros, sendo adequada para o consumo. (Michaelis)
31	<b>Racionamento</b>	<b>ND:</b> Ação ou efeito de racionar. (Michaelis)  <b>ND2:</b> Limitação da quantidade de gêneros que cada pessoa ou família pode comprar ou receber. (Michaelis)
32	Registro	<b>ND:</b> Válvula ou torneira reguladora da passagem de água, vapor ou ar. (Michaelis)
33	<b>Torneira</b>	<b>ND:</b> Peça dotada de uma chave, acoplada a um cano, recipiente, tubo etc., usada para deixar sair ou reter o líquido ou o gás nela contido. (Michaelis)
34	<b>Tratamento</b>	<b>ND:</b> Ato ou efeito de tratar alguém, alguma coisa ou a si mesmo; trato. (Michaelis)
35	<b>Tubulação</b>	<b>ND:</b> Sistema de distribuição de tubos para a passagem de água, gás ou eletricidade ou ainda para o escoamento de água servida e esgoto; canalização. (Michaelis)
36	<b>Vaso sanitário</b>	<b>ND:</b> peça côncava grande, geralmente de louça, onde se urina e defeca; é dotado de canalização que conduz os dejetos produzidos para esgotos ou fossa sanitária; latrina, patente, privada, retreta, retrete, sanitário, vaso. (Michaelis)
37	<b>Vazamento</b>	<b>ND:</b> Ato ou efeito de vazar; vazadura, vazão. (Michaelis)

Lista relacionada à Qualidade/Característica

	<b>Termo</b>	<b>Notação de Definição – ND</b>
1	Consecutivo	<b>ND:</b> Que segue em série; um imediatamente após o outro; seguido, sequente, sucessivo, consequente. (Michaelis)

2	Consertado	<b>ND:</b> Cujo funcionamento ou andamento normal foi corrigido; ajustado, regulado. (Michaelis)
3	Constante	<b>ND:</b> Que se repete incessantemente; contínuo, incessante, permanente. (Michaelis)
4	Desregulado	<b>ND:</b> Que se repete incessantemente; contínuo, incessante, permanente. (Michaelis)
5	Escasso	<b>ND:</b> Que é desprovido de algo; carente, faltar. (Michaelis)
6	Estragado	<b>ND:</b> Que se estragou; danificado, avariado, inutilizado. (Michaelis)
7	Frequente	<b>ND:</b> Repetido muitas vezes; amiudado. (Michaelis)
8	Ferroviário	<b>ND:</b> Relativo à ferrovia. (Michaelis)
9	Interrompido	<b>ND:</b> Que se interrompeu; suspenso. (Michaelis)
10	Leitoso	<b>ND:</b> Que tem a cor ou a consistência do leite; lactescente, (Michaelis)
11	Limpo	<b>ND:</b> Livre de qualquer sujeira. (Michaelis)
12	Movimentado	<b>ND:</b> Posto em movimento; motório. (Michaelis) <b>ND2:</b> Em que há movimento considerável. (Michaelis)
13	Quebrado	<b>ND:</b> Que se quebrou; separado em dois ou mais pedaços; fragmentado, fraturado, partido. (Michaelis) <b>ND2:</b> Fendido com ruptura; rachado, trincado. (Michaelis)
14	Recorrente	<b>ND:</b> Diz-se de algo que torna a surgir após ter desaparecido. (Michaelis)

## **APÊNDICE B**

### **Tesouro de recursos hídricos baseado em termos do Sistema Gota D'Água**

#### **RESUMO**

Esse documento apresenta um tesouro baseado no levantamento das palavras utilizadas nas descrições das contribuições voluntárias armazenadas na base de dados do sistema Gota D'Água. Este tesouro foi obtido a partir de recortes feitos no Tesouro de Engenharia Sanitária e Ambiental (CEPIS, 2005).

**Palavras-chave:** Tesouro, Contribuições voluntárias, Sistema Gota D'Água, Tesouro de Engenharia Sanitária e Ambiental

## 1 TESAURO DE RECURSOS HÍDRICOS BASEADO EM TERMOS DO SISTEMA GOTA D'ÁGUA

Este tesauro foi obtido a partir de recortes feitos no Tesauro de Engenharia Sanitária e Ambiental (CEPIS, 2005), que compila 2.098 descritores e 3.037 não descritores. Este Tesauro foi revisado e atualizado, tendo como referência principal a base de dados bibliográficos da Rede Pan-americana de Informação sobre Saúde Ambiental (REPIDISCA) e as diferentes fontes de informação da Biblioteca Virtual de Saúde Ambiental (BVSA). O Tesauro Ambiental Multilíngüe Geral (GEMET), versão 2004, preparado pelo *Consiglio Nazionale delle Ricerche* (CNR) e a *Umweltbundesamt* (UBA), também foram utilizados como referências.

Na seção Semântica do Tesauro Base os descritores estão organizados por temas que formam campos semânticos segundo o esquema de classificação da REPIDISCA. Nesse esquema, o campo da engenharia sanitária e ciências do ambiente está dividido em 10 grandes grupos ou categorias de matéria (Tabela 6).

As 10 categorias estão, por sua vez, subdivididas em temas mais específicos e cada uma é representada por um código de quatro dígitos. Isso justifica a numeração que aparece nos termos do tesauro gerado.

Na seção alfabética os descritores são apresentados com suas relações preferenciais, hierárquicas e associativas. Os símbolos para representá-los no tesauro são os seguintes:

- NA (NOTA DE ALCANCE) - Define, explica ou limita o significado do descritor para finalidades de indexação;
- UP (USADO PARA) - Indica os sinônimos ou não descritores; eles não são válidos para a indexação;
- TG (TERMO GENÉRICO) - Assinala o termo mais amplo ao qual pertence o descritor;
- TE (TERMO ESPECÍFICO) - Indica os termos específicos (tipos ou classes) associados ao descritor;

TR (TERMO RELACIONADO) - Indica os termos associados semanticamente com o descritor.

Algumas palavras encontradas nas contribuições VGIs não foram incluídas neste recorte do tesauro de Engenharia por questões relacionadas à falta de tempo. Pois, para a devida inclusão seria necessário um tempo maior de dedicação.

**Abastecimento de água 4000**

NA Conjunto funcional de obras, instalações, tubulações, equipamentos e acessórios destinados a produzir água em quantidades, qualidade, regularidade e confiabilidade adequados

*UP Provisão pública de água*

*Provisão de água*

*Redes de abastecimento de água*

*Serviços de fornecimento de água*

*Serviços de água potável*

*Serviços de abastecimento de água potável*

*Sistemas de abastecimento de água*

*Fornecimento de água*

**TR** Água potável

Redes de esgoto

Armazenamento de água

Qualidade da água

Captação de água

Consumo de água [Saúde ambiental]

Demanda de água

Distribuição de água

Empresas de água potável

Escassez de água

Recursos hídricos

Sistemas condominiais da água

Tratamento da água

**Acessórios de redes de água e de esgoto 4406 5200**

*UP Duplo tee*

*Macromedidores*

*Redutor*

*Tees*

**TR** Acessórios sanitários

Aparelhos sanitários

Instalações sanitárias

Limitadores de consumo de água

Encanamento [Saúde ambiental]

Tubulações

**Acessórios sanitários 4406**

*UP Chaves de passo*

*Canos*

*Cano de água*

*Cotovelos*

*Canos de água*

*Grifos*

*Chaves de água*

*Sifões*

*Sumidouros*

*Caixas para graxa*

*Caixas para areia*

*Caixas para graxas*

**TR** Acessórios de redes de água e de esgoto

Aparelhos sanitários

Instalações sanitárias

Limitadores de consumo de água

Encanamento [Saúde ambiental]

Tubulações

**Água 3000**

NA Recomenda-se o uso de descritores específicos

**TE** Água de chuva

Água doce

Esgoto doméstico de lavadouro

Água industrial

Água potável

Água pura

Água para recreação

Águas residuárias

Águas subterrâneas

Águas superficiais

**TR** Meio ambiente

Autodepuração da água

Qualidade da água

Quantidade de água

Poluição da água

Hidrologia

Usos da água

**Água artesiana 3000 3200**

NA Água com pressão suficiente para sair à superfície sem necessidade de bombeamento

**TG** Águas subterrâneas

**TR** Poços artesianos

**Água de chuva 3000**

*UP Água de escoamento*

*Água pluvial*

**TG** Água

**TR** Captação de águas de chuva

Escoamento de água de chuva

Chuvas

**Água doce 3000**

NA Água potável de pouco ou nenhum sabor, por contraposição a águas do mar ou as minerais. Sua mineralização é muito fraca. Este termo designa igualmente as águas continentais, quer sejam superfícies ou subterrâneas

*UP Água fresca*

**TG** Água

**TR** Água potável

Lagos

Rios

**Água industrial 3000 4612**

**TG** Água

**TR** Consumo industrial de água

Usos da água

**Água potável 4800**

NA Água própria para consumo humano

*UP Água de beber*

*Água de uso doméstico*

*Água inócua*

*Água para consumo humano*

*Água segura*

**TG** Água

**TR** Abastecimento de água

Água doce

Consumo doméstico de água

Empresas de água potável

Plantas de tratamento da água

**Água para recreação 3000 4612**

*UP Recreação [Água]*

**TG** Água

**TR** Piscinas

Usos da água

**Água pura 3000**

NA Água bidestilada sobre permanganato potássico e barita, gelada e coletada seguidamente por fusão ao abrigo do ar

**TG** Água

**Águas residuárias 5100**

NA Água que contém resíduos sólidos e líquidos despejados como inúteis logo depois de seu uso doméstico ou municipal. Para o caso de águas residuárias industriais use-se: Efluentes industriais

*UP Água residual*

*Águas cloacais*

*Águas de esgoto*

*Águas negras*

*Águas residuais cruas*

*Águas servidas*

*Águas sépticas*

*Esgotos*

*Esgotos*

*Despejos líquidos*

*Efluentes*

*Efluentes líquidos*

*Líquidos cloacais*

**TG** Água

**TE** Efluentes combinados

Esgotos domésticos

Efluentes industriais

**TR** Redes de esgoto

Poluição da água

**Águas subterrâneas 3000 3200**

*UP Água confinada*  
*Água do subsolo*  
*Água freática*

**TG** Água  
Recursos hídricos

**TE** Aquíferos  
Água artesianas

**TR** Poços

**Águas superficiais 3000 3200**

**NA** As que se acham na superfície de terra formando os oceanos, marés, lagoas, rios, córregos, represas, etc

*UP Água de superfície*  
*Água superficial*  
*Águas de superfície*

**TG** Água  
Recursos hídricos

**TE** Lagos  
Lagoas  
Mananciais  
Rios

**TR** Captação de águas superficiais

**Análise da água 3307 4807**

**NA** Inclui águas residuárias

**TR** Qualidade da água  
Características da água  
Poluição da água

**Aparelhos sanitários 4406**

*UP Acessórios sanitários*  
*Artefatos sanitários*  
*Banheiras*  
*Bebedouros*  
*Bidê*  
*Duchas*  
*Vasos sanitários*  
*Pias*  
*Duchas*

*Mictórios*

**TR** Acessórios sanitários  
Acessórios de redes de água e de esgoto  
Instalações sanitárias  
Limitadores de consumo de água  
Encanamento [Saúde ambiental]  
Tubulações

**Aquíferos 3200**

**NA** Formações ou estratos compreendidos dentro da zona de saturação

**TG** Águas subterrâneas

**TR** Poços

**Armazenamento de água 4400**

*UP Água armazenada*

**TR** Abastecimento de água  
Captação de água  
Conservação da água  
Demanda de água  
Tanques  
Reservatórios

**Autodepuração da água 3101**

*UP Autopurificação da água*  
*Capacidade assimilativa*

**TR** Água  
Tratamento da água

**Bacias 3200**

*UP Desenvolvimento de bacias*  
*Fossas*

**TE** Bacias hidrográficas

**TR** Captação de água  
Hidrologia  
Recursos hídricos

**Bacias hidrográficas 3200**

*UP Cuencas de recepção*  
*Cuencas hidrogeológicas*  
*Cuencas hidrogeológicas*

<b>TG</b> Bacias	<i>subterrâneas</i> <i>Exploração de águas subterrâneas</i> <i>Exploração de águas subterrâneas</i>
<b>Bocais 1600</b> <b>TG</b> Equipamento de estações de tratamento	<b>TG</b> Captação de água
<b>Bombas 1600 4204</b> <b>TG</b> Equipamento de estações de Tratamento	<b>TE</b> Poços
<b>TR</b> Cavitação Estações de recalque	<b>TR</b> Águas subterrâneas Fluxo de águas subterrâneas
<b>Caldeiras 1600</b> <b>TG</b> Equipamento de estações de Tratamento	<b>Captação de águas superficiais 4103</b> <b>TG</b> Captação de água
<b>Caminhões cisterna 6501</b> NA Equipamento utilizado para o varrição hidráulica na limpeza de ruas, fazendo a varrição com mangueira de bomba. Também tem caminhões cisterna para a distribuição da água ou para transportar outros líquidos	<b>TE</b> Captação em lagos Captação em rios
<i>UP Cisternas</i> <i>Transporte de água</i>	<b>TR</b> Águas superficiais
<b>TR</b> Distribuição de água	<b>Captação em lagos 4103</b> <b>TG</b> Captação de águas superficiais <b>TR</b> Lagos
<b>Captação de água 4100</b> <i>UP Zonas de captação</i> <i>Áreas de captação</i> <i>Áreas de captação da água</i>	<b>Captação em rios 4103</b> <i>UP Captação em arroios</i> <b>TG</b> Captação de águas superficiais <b>TR</b> Rios
<b>TR</b> Abastecimento de água Armazenamento de água	<b>Caraterísticas da água 3000</b> <i>UP Natureza da água</i> <i>Parâmetros da água</i> <i>Propriedades da água</i>
<b>Captação de águas de chuva 4101</b> <b>TG</b> Captação de água	<b>TR</b> Análise da água Qualidade da água Poluição da água
<b>TR</b> Água de chuva	<b>Cavitação 3103</b> NA Formação e rápido colapso de borbulhas de vapor causadas pela baixa pressão originada pela alta velocidade da água
<b>Captação de águas subterrâneas 4105</b> <i>UP Aproveitamento de águas</i>	<b>TR</b> Bombas
	<b>Chuvas 2108</b> <i>UP Chuva</i>

**TR** Água de chuva  
Escoamento de água de chuva

**Cidades 2102**  
*UP Municipalidades*  
*Municípios*

**TE** Cidades grandes  
Cidades medianas  
Cidades pequenas

**TR** População urbana

**Cidades grandes 2102**  
*UP Grandes cidades*  
**TG** Cidades

**Cidades medianas 2102**  
**TG** Cidades

**Cidades pequenas 2102**  
*UP Pequenas cidades*  
**TG** Cidades

**Cloração 4308 5313**  
*UP Água tratada com cloro*  
*Decloração*

**TG** Desinfecção da água

**TR** Cloradores

**Cloradores 4308**  
*UP Cloradores de alimentação direta*  
*Cloradores de solução à vácuo*  
*Dosificadores de cloro*  
*Hipocloradores*

**TG** Equipamento

**TR** Cloração

**Conflitos pela água 3900**  
*UP Altercados pela água*  
*Contendas pela água*  
*Controvérsias pela água*  
*Desavenças pela água*  
*Disputas pela água*  
*Litígios pela água*

*Lutas pela água*  
*Peléias pela água*  
*Pugnas pela água*  
*Querelas pela água*

**TR** Quantidade de água  
Consumo de água [Saúde ambiental]  
Demanda de água  
Escassez de água  
Legislação ambiental

**Conservação da água 4400**

**NA** Medidas tomadas para economizar a quantidade de água utilizada para qualquer finalidades ou para preservá-la da poluição

*UP Economia de água*  
*Baixo consumo de água*

**TE** Racionamento de água

**TR** Armazenamento de água  
Desperdício de água  
Escassez de água  
Vazamento de água  
Limitadores de consumo de água  
Uso de águas residuárias  
Uso eficiente da água

**Consumo comercial de água 4612**

**TG** Consumo de água [Saúde ambiental]

**TR** Zonas comerciais

**Consumo de água [Saúde ambiental] 4612**

*UP Consumo hídrico*  
*Consumo público de água*  
*Gastos de água*

**TE** Consumo comercial de água  
Consumo doméstico de água  
Consumo industrial de água  
Consumo não contabilizado de água

**TR** Abastecimento de água  
Quantidade de água

Conflitos pela água  
Demanda de água  
Desperdício de água  
Escassez de água  
Limitadores de consumo de água  
Medição de consumo  
Medidores  
Tarifas [Serviços de saneamento]  
Uso eficiente da água  
Usos da água

**Consumo doméstico de água 4612**

**TG** Consumo de água [Saúde ambiental]

**TR** Água potável  
Demanda de água  
Uso eficiente da água

**Consumo industrial de água 4612**

**TG** Consumo de água [Saúde ambiental]

**TR** Água industrial

**Consumo não contabilizado de água 4610**

*UP* Água não contabilizada

**TG** Consumo de água [Saúde ambiental]

**TR** Vazamento de água  
Instalações clandestinas

**Controle da qualidade da água 3300 4800**

**TR** Qualidade da água

**Correntes de água 3103**

*UP* Arroios

*Cursos de água*

*Cursos d'água*

*Derivação de cursos de água*

*Ordenação de cursos de água*

*Ordenação de cursos de água*

*Rede hidrográfica*

**TE** Escoamento em conduto fechado  
Correntes superficiais

**TR** Fluxo de água  
Fluxo de águas subterrâneas  
Rios

**Correntes superficiais 3103**

**TG** Correntes de água

**TR** Águas superficiais

**Crítérios de qualidade da água 3300 4800**

**TR** Qualidade da água  
Normas de qualidade da água  
Usos da água

**Custo da água 3203**

*UP* Cobrança da água

*Preço da água*

**TR** Tarifas [Serviços de saneamento]  
Valor da água

**Demanda de água 4612**

**NA** Quantidade de água requerida para diferentes usos

*UP* Caudal máximo diário

*Caudal máximo horário*

*Caudal médio*

*Necessidade de água*

**TR** Abastecimento de água  
Armazenamento de água  
Conflitos pela água  
Consumo de água [Saúde ambiental]  
Consumo doméstico de água  
Distribuição de água  
Escassez de água  
Uso eficiente da água

**Desinfecção da água 4308 5313**

**NA** Destruição de bactérias patogênicas, geralmente com um antisséptico químico ou desinfetante

**TG** Operações unitárias

**TE** Cloração

**Detergentes 2400 3402**

*UP Espumas*

*Material de limpeza*

**TG** Poluentes conservativos

**TR** Efluentes industriais

**Desperdício de água 4610**

**NA** Aplicado ao mau uso da água

**TR** Conservação da água

Consumo de água [Saúde ambiental]

Escassez de água

Vazamento de água

Usos da água

**Direito das águas 3900**

*UP Código de águas*

*Legislação de água*

*Lei de águas*

*Leis de água*

*Leis sobre água*

**TR** Normas de qualidade da água

**Distribuição de água 4400**

**NA** Parte do sistema de abastecimento de água localizada entre os pontos a partir dos quais a água se torna potável [saída de estações de tratamento, unidades de desinfecção] e os pontos de utilização da água nas instalações prediais

*UP Canalização da água*

**TR** Abastecimento de água

Caminhões cisterna

Vazão de água

Demanda de água

Vazamento de água

Instalações sanitárias

Redes de distribuição de água

Sistemas pressurizados

**Edifícios 2102**

**TG** Habitação

**TR** Urbanização

**Efluentes combinados 5100**

**TG** Águas residuárias

**Efluentes industriais 5400**

*UP Águas residuárias industriais*

*Resíduos líquidos industriais*

**TG** Águas residuárias

**TR** Detergentes

Resíduos industriais

**Empresas de água potável 4600**

**TG** Empresas de saneamento

**TR** Abastecimento de água

Água potável

**Empresas de água potável 4600**

**TG** Empresas de saneamento

**TR** Abastecimento de água

Água potável

**Empresas de limpeza urbana 6200**

**TG** Empresas de saneamento

**Empresas de saneamento 0501**

*UP Instituições de saneamento*

**TE** Empresas de água potável

Empresas de saneamento básico

**TR** Saneamento

**Empresas de saneamento básico 5700**

**TG** Empresas de saneamento

**TR** Redes de esgoto

**Encanamento [Saúde ambiental]4406**

*UP Tubulações hidráulicas*

*Encanamento*

**TR** Acessórios sanitários  
Acessórios de redes de água e de Esgoto

**Equipamento 1600**

NA Recomenda-se seu uso com descritores específicos

*UP Aparelhos*  
*Compra de equipamento*

**TE** Cloradores  
Válvulas de água

**Equipamento de estações de tratamento**

**1600**

*UP Equipamento industrial*  
*Equipamento industrial*

**TG** Equipamento

**TE** Bombas  
Bocais  
Caldeiras  
Filtros

**Escassez de água 4612**

*UP Falta de serviço de água potável*

**TR** Abastecimento de água  
Conflitos pela água  
Conservação da água  
Consumo de água [Saúde ambiental]  
Demanda de água  
Desperdício de água  
Vazamento de água

**Escoamento 3101**

NA Corrente de água que verte ao ultrapassar seu depósito ou leito natural ou artificial

**TE** Escoamento de água de chuva

**TR** Infiltração  
Inundações [Desastres]

**Escoamento de água de chuva 3101 4105 5204**

**TG** Escoamento

**TR** Água de chuva  
Chuvas

**Escoamento em conduto fechado 3103**

*UP Fluxo em canais abertos*

**TG** Correntes de água

**TR** Fluxo de água

**Esgotos domésticos 5100**

*UP Águas residuárias domésticas*

*Águas de esgoto domésticas*

*Águas residuais domésticas*

*Águas residuais domésticas*

*Águas sépticas domésticas*

*Efluentes domésticos*

*Resíduos líquidos domésticos*

*Efluentes domésticos crus*

*Líquidos cloacais domésticos*

**TG** Águas residuárias

**TR** Esgoto doméstico de lavadouro

**Esgoto doméstico de lavadouro 5100**

NA Efluentes domésticos que excluem as

águas que provem dos inodoros

**TG** Água

**TR** Esgotos domésticos

**Estações de recalque 4204 5207**

*UP Centrais de bombeamento*

*Câmaras de bombeamento*

*Estações elevatórias*

*Instalações de bombeamento*

*Plantas elevatórias*

**TR** Bombas

Plantas de tratamento da água

**Filtros 1600 4307 5304**

**TG** Equipamento de estações de Tratamento

**Fluxo de água 3103**

*UP Cálculo de fluxo*  
*Fluxos hidráulicos*

**TR** Correntes de água

**Fluxo de águas subterrâneas 3103**

*UP Correntes subterrâneas*

**TG** Fluxo de água

**TR** Águas subterrâneas  
Captação de águas subterrâneas  
Correntes de água

**Habitação 2102**

**TE** Edifícios  
Habitação social

**TR** Saneamento de residências

**Habitação social 2102**

**TG** Habitação

**Hidrantes 4406**

NA Torneiras contra incêndios. Tomada de água de uma condução a pressão

*UP Bocas de incêndio*  
*Bombas para incêndios*  
*Hidrantes contra incêndios*

**TR** Instalações sanitárias  
Sistemas de combate a incêndio

**Hidrologia 3101**

NA Parte das ciências naturais que estuda as águas

*UP Aspectos hidrológicos*  
*Ciências da água*

**TR** Água  
Bacias  
Inundações [Desastres]

**Incêndios 0800**

**TR** Sistemas de combate a incêndio

**Indústrias 9400**

NA Recomenda-se o uso de descritores específicos

*UP Fabricação*  
*Fábrica*  
*Fábricas*  
*Indústria*  
*Manufaturação*

**TR** Efluentes industriais  
Resíduos industriais

**Infiltração 4105 5503**

**TR** Escoamento

**Instalações clandestinas 4406 4610**

*UP Ligações clandestinas*

**TR** Consumo não contabilizado de água  
Instalações sanitárias

**Instalações domiciliares 4406**

*UP Ligações domiciliares*

**TG** Instalações sanitárias

**TR** Sistemas condominiais da água

**Instalações flexíveis 4406**

*UP Ligações flexíveis*

**TG** Instalações sanitárias

**TR** Tubulações de PVC

**Instalações sanitárias 4406**

*UP Ramal predial*  
*Ligações prediais*

**TE** Instalações domiciliares  
Instalações flexíveis

**TR** Acessórios sanitários  
Acessórios de redes de água e de esgoto

Aparelhos sanitários  
Distribuição de água  
Hidrantes  
Instalações clandestinas  
Medidores  
Sistemas condominiais da água

*Latrinas de fosso seco*  
*Latrinas químicas*  
*Latrinas sépticas*  
*Privada*  
*Toaletes*  
*Vaso*

**Inundações [Desastres]0800**

*UP Enchentes*  
*Cheias*  
*Transbordamentos*

**TR** Escoamento  
Hidrologia

**Irrigação [Agricultura]3709 5512  
7702**

*UP Água para irrigação*  
*Rega*

**TR** Uso de águas residuárias  
Usos da água

**Lagos 3105 3200**

NA Massa continental de águas  
superficiais  
de extensão considerável

**TG** Águas superficiais

**TR** Água doce  
Captação em lagos  
Lagoas

**Lagoas 3105 3200**

NA Depósito natural de água de  
menores  
dimensões que o lago

**TG** Águas superficiais

**TR** Lagos

**Latrinas [Saúde ambiental]5600**

*UP Latrinas*  
*Latrinas bioquímicas*  
*Latrinas de selo hidráulico*  
*Latrinas de fosso anegado*  
*Latrinas de fosso perfurado*

**TR** Saneamento básico

**Legislação ambiental 2900**

*UP Crime ambiental*  
*Delito contra o meio ambiente*  
*Delito ambiental*  
*Direito ambiental*  
*Disposições legais ambientais*  
*Disposições sanitárias*  
*Dispositivos legais ambientais*  
*Dispositivos sanitários*  
*Informação legislativa*  
*Licenças negociáveis de*  
*contaminação*  
*Autorização de poluir negociável*  
*Regulamentação do ambiente*

**TR** Conflitos pela água  
Direito das águas

**Limitadores de consumo de água  
4612**

*UP Apeiros de economia da água*  
*Acessórios de economia da água*  
*Dispositivos de economia da água*  
*Mecanismos aforradores da água*

**TR** Acessórios sanitários

Acessórios de redes de água e de  
esgoto  
Aparelhos sanitários  
Conservação da água  
Consumo de água [Saúde ambiental]  
Vazamento de água  
Uso eficiente da água

**Mananciais 3200**

*UP Olhos de água*  
*Altura de carga de esguicho*  
*Altura de carga de esguicho*

**TG** Águas superficiais

**TE** Mananciais protegidos

**Mananciais protegidos 3200**

*UP Proteção de mananciais*

**TG** Mananciais

**Medição de consumo 4612**

*UP Micromedição*

**TR** Consumo de água [Saúde ambiental]

**Medidores 0000**

**TG** Equipamento de estações de tratamento

**TR** Consumo de água [Saúde ambiental]

Instalações sanitárias

Tarifas [Serviços de saneamento]

**Medição de vazão 3102 4400**

*UP Indicação de vazão*

*Estimativa de vazão*

*Macromedição*

**TR** Vazão de água

**Meio ambiente 2000**

NA Recomenda-se o uso de descritores Específicos

*UP Condição ambiental*

*Condição ambiental*

*Entorno*

*Factor ambiental*

*Ambiente*

*Meio ambiente construído*

*Meio ambiente físico*

*Meio ambiente natural*

**TR** Água

**Normas de qualidade da água 3300 4900**

*UP Standards de qualidade da água*

**TR** Qualidade da água  
Poluição da água

Direito das águas

**Operações unitárias 4300 5300**

NA Cada uma das etapas ou fases que compreende o tratamento da água na qual intervém ou se favorece a ocorrência de um ou mais fenômenos mecânicos, físicos, químicos e/ou biológicos com um fim determinado

*UP Processos unitários*

**TR** Tratamento da água

**Piscinas 0511**

*UP Água de piscinas*

**TR** Água para recreação  
Desinfecção da água

**Plantas de tratamento da água 4324**

**TR** Água potável  
Tratamento da água

**Poços 4105**

**TG** Captação de águas subterrâneas

**TE** Poços artesianos

**TR** Águas subterrâneas  
Mananciais

**Poços artesianos 4105**

**TG** Poços

**TR** Água artesianas

**Poluentes conservativos 2303**

*UP COPs*

*Contaminantes inorgânicos persistentes*

*Poluentes não biodegradáveis*

*Contaminantes orgânicos persistentes*

*Poluentes persistentes*

*Persistência de pesticida*

**TE** Detergentes

**TR** Poluentes da água

**Poluentes da água 3402**

*UP Fontes de contaminação da água*

**TR** Poluição da água

**Poluição da água 3500**

NA Recomenda-se o uso de descritores específicos

*UP Água poluída*

*Contaminação térmica da água*

*Degeneração da água*

*Deterioração da água*

*Contaminação da água*

*Poluição térmica da água*

**TR** Água

Águas residuárias

Qualidade da água

Poluentes da água

Controle da qualidade da água

Normas de qualidade da água

Tratamento da água

**População urbana 2700**

**TR** Cidades

Urbanização

**Qualidade da água 3300 4800**

**TR** Abastecimento de água

Água

Análise da água

Características da água

Poluição da água

Controle da qualidade da água

Critérios de qualidade da água

Normas de qualidade da água

Tratamento da água

**Quantidade de água 4612**

*UP Volume de água*

*Litro*

**TR** Água

Vazão de água

Conflitos pela água

Consumo de água [Saúde ambiental]

**Racionamento de água 4400**

**TG** Conservação da água

**Recursos hídricos 3200**

NA Numa determinada região, a quantidade de águas superficiais ou subterrâneas, disponíveis para qualquer uso

*UP Afluentes*

*Corpos d'água*

*Fontes de água*

*Recursos de água*

*Recursos hidrológicos*

*Recursos limnológicos*

**TE** Águas subterrâneas

Águas superficiais

**TR** Abastecimento de água

Bacias

**Redes de distribuição de água 4402**

NA Conjunto de estruturas hidráulicas [reservatórios, bombas, tubulações, etc.] necessárias para assegurar uma distribuição de água adequada para vários usos

*UP Aquedutos*

**TR** Distribuição de água

Traçado de redes

**Redes de esgoto 5200**

*UP Tubos de esgoto*

*Canalização de desaguentos*

*Captação de esgotos*

*Cloacas*

*Sistemas de esgoto*

*Esgotos*

*Redes cloacais*

*Redes de efluentes sanitários*

*Redes de canalização*

*Redes de desaguentamento*

*Coleta de esgotos*

*Sistemas de esgoto*

**TR** Abastecimento de água  
Águas residuárias  
Empresas de saneamento básico  
Saneamento  
Tubulações

**Reservatórios 4107**

NA Use para abastecimento de água potável

*UP Reservatórios de água*

**TR** Armazenamento de água  
Tanques

**Resíduos 6000**

*UP Refugos*  
*Refugos de fabricação*  
*Desperdícios*  
*Produtos de refugos*  
*Refugos residuais*

**TE** Resíduos sólidos

**Resíduos comerciais 6000**

**TG** Resíduos sólidos

**TR** Resíduos domésticos  
Resíduos industriais  
Resíduos orgânicos  
Zonas comerciais

**Resíduos domésticos 6000**

NA Todo tipo de despejo que normalmente se produz em residências ou conjuntos habitacionais

*UP Lixos domésticos*  
*Restos domésticos*  
*Lixo do município*  
*Despejos urbanos*  
*Resíduos domiciliários*  
*Lixo do município*  
*Resíduos sólidos municipais*  
*Resíduos sólidos urbanos*

**TG** Resíduos sólidos

**TR** Esgotos domésticos

Resíduos comerciais

**Resíduos industriais 6000**

NA Todos os despejos sólidos que se obtêm como resultado de processos industriais e de operações de reparação e limpeza

*UP Lixos industriais*  
*Refugos de fabricação*  
*Despejos industriais*  
*Produtos residuais industriais*

**TG** Resíduos sólidos

**TR** Efluentes industriais  
Indústrias

**Resíduos orgânicos 6000**

*UP Desperdício orgânico*  
*Resíduo verde*  
*Resíduos biodegradáveis*  
*Desperdício orgânico*

**TG** Resíduos sólidos

**TR** Resíduos comerciais

**Resíduos perigosos 6300**

*UP Despejos de materiais perigosos*  
*Despejos de substâncias daninhas*  
*Despejos de substâncias nocivas*  
*Produtos perigosos*  
*Resíduos de produtos perigosos*  
*Substâncias residuais perigosas*  
*Resíduos industriais perigosos*  
*Resíduos químicos perigosos*  
*Locais de resíduos perigosos*

**TG** Resíduos

**TR** Resíduos sólidos

**Resíduos sólidos 6000**

NA Para referir a resíduos líquidos, use:  
Águas residuárias, ou os diferentes tipos de efluentes

*UP Lixo [Saúde ambiental]*  
*Despejos*  
*Restos sólidos*

**TG** Resíduos

**TE** Resíduos comerciais  
Resíduos domésticos  
Resíduos industriais

**Rios 3200**

*UP Regueiras*

**TG** Águas superficiais

**TR** Água doce  
Captação em rios  
Correntes de água

**Saneamento 0501**

**NA** As ações necessárias de controle e obras para restabelecer um balanço ambiental satisfatório

*UP Aspectos sanitários*  
*Fatores sanitários*  
*Higiene ambiental*  
*Higiene do meio*  
*Saneamento ambiental*

**TE** Saneamento básico

**TR** Abastecimento de água  
Redes de esgoto

**Saneamento básico 0501**

**TG** Saneamento

**TR** Latrinas [Saúde ambiental]

**Saneamento de residências 0501**

*UP Higiene de residências*  
*Banheiro*

**TG** Saneamento

**TR** Habitação

**Sistemas de combate a incêndio 0800**

*UP Controle de incêndios*  
*Extinção de incêndios*  
*Instalações contra incêndios*  
*Luta contra incêndios*  
*Sistemas contra incêndios*

**TR** Hidrantes

Incêndios

**Sistemas condominiais da água 4406**

**TR** Abastecimento de água  
Instalações domiciliares  
Instalações sanitárias

**Sistemas pressurizados 4400**

*UP Hidropneumática*  
*Sistemas hidropneumáticos*  
*Sistemas pneumáticos*  
*Tanques hidropneumáticos*

**TR** Distribuição de água

**Substâncias perigosas 6300**

*UP Agentes perigosos*  
*Poluentes perigosos*  
*Poluentes tóxicos*  
*Materiais perigosos*  
*perigosos*  
*Substâncias daninhas*  
*Substâncias nocivas*  
*Substâncias tóxicas*  
*ambientais*

**TR** Resíduos perigosos

**Tanques 4401**

**TR** Armazenamento de água  
Reservatórios

**Tarifas [Serviços de saneamento]4600**  
**5700**

**NA** Use em combinação com outros  
Descritores

*UP Cobros*  
*Preços*

**TR** Abastecimento de água  
Consumo de água [Saúde ambiental]  
Custo da água  
Medidores

**Traçado de redes 4409**  
**5202**

**TR** Redes de esgoto

- Redes de distribuição de água
- Tratamento da água 4300**  
NA Aplica-se a potabilização da água
- UP Acondicionamento da água*  
*Despoluição da água*  
*Potabilidade da água*  
*Tecnologia da água*
- TR** Abastecimento de água  
Autodepuração da água  
Qualidade da água  
Poluição da água  
Operações unitárias  
Plantas de tratamento da água
- Tubulações 4403 5200**  
*UP Tubulação*  
*Tubos*
- TE** Tubulações de PVC
- TR** Acessórios sanitários  
Acessórios de redes de água e de Esgoto  
Redes de esgoto
- Tubulações de PVC 4403 5200**  
*UP Tubulações de plástico*  
*Tubulações de polivinila*
- TG** Tubulações  
**TR** Instalações flexíveis
- Urbanização 2102**  
*UP Expansão urbana*
- TR** Cidades  
Edifícios  
População urbana
- Usos da água 3709**  
*UP Aproveitamento da água*
- TE** Uso eficiente da água
- TR** Água  
Água industrial  
Água para recreação
- Consumo de água [Saúde ambiental]  
Critérios de qualidade da água  
Desperdício de água  
Irrigação [Agricultura]
- Uso de águas residuárias 5512**  
*UP Água de reuso*  
*Aproveitamento de águas residuárias*  
*Reciclagem de esgotos*  
*Recuperação da água*  
*Reuso de águas residuárias*  
*Reuso de efluentes*  
*Reutilização de águas residuárias*  
*Reutilização de efluentes*
- TR** Águas residuárias  
Conservação da água  
Irrigação [Agricultura]
- Uso eficiente da água 4612**  
**TG** Usos da água
- TR** Conservação da água  
Consumo de água [Saúde ambiental]  
Consumo doméstico de água  
Demanda de água  
Limitadores de consumo de água
- Válvulas de água 4406**  
*UP Reguladores de vazão*  
*Válvulas de controle de fluxos*  
*Registro*
- TG** Equipamento
- TE** Válvulas de descarga  
Válvulas de retenção

**Palavras que tiveram mudanças:**

**Bombas:** No levantamento de palavra se encontra Bomba

**Edifícios:** No levantamento de palavra se encontra Edifício

**Esgotos domésticos:** No levantamento de palavra se encontra Esgoto doméstico

**Hidrantes:** No levantamento de palavra se encontra Hidrante

**Válvulas de descarga:** No levantamento de palavra se encontra Descarga. Subentendido como válvula de descarga, devido ao fato do usuário dizer que está frequentemente desregulada

**Redes de distribuição de água:** No levantamento de palavra se encontra apenas Rede, portanto, pelo contexto, se trata de Redes de distribuição de água

**Vazamento de água:** No levantamento de palavra se encontra apenas vazamento, portanto, pelo contexto, se trata de vazamento de água

**Cidades:** No levantamento de palavra se encontra Cidade

**Municípios:** No levantamento de palavra se encontra município

**Produtos químicos:** No levantamento de palavra se encontra Produto químico

**Incêndios:** No levantamento de palavra se encontra incêndio

**Vias:** Entende-se vias públicas pelo contexto

**Água pluvial:** No levantamento de palavra se encontra águas pluviais

**Reservatórios:** No levantamento de palavra se encontra reservatório

**Resíduos líquidos industriais:** No levantamento de palavra se encontra resíduos líquidos e gasosos das indústrias

**Detergentes:** No levantamento de palavra se encontra detergente

**Latrinas:** No levantamento de palavra se encontra latrina