SEVEN publicações acadêmicas

Use of public domain knowledge bases: A case study on the RDF language

Crossref doi
https://doi.org/10.56238/uniknowindevolp-048

Ronnie Carlos Tavares Nunes
Federal University of Santa Catarina
Master in Engineering and Knowledge Management

Márcio Clemes
Federal University of Santa Catarina
Master in Production Engineering

Rogério Cid Bastos
Federal University of Santa Catarina
PhD in Production Engineering

ABSTRACT
The purpose of this paper is to analyze Semantic Web and Linked Data technologies, with the development of a web application that performs a knowledge modeling through the use of ontologies and RDF (Resource Description Framework) language. A public knowledge ontology/vocabulary was used and populated with data from a relational database from Brazil. The aggregated data was converted to RDF format and made available in an RDF database. The Virtuoso© universal server was used for this task, a hybrid database platform that has the functionalities of a traditional relational database management system, object-relational database, virtual database, RDF, XML, free text, web application server and file server functionality in one system. As a result, a web application capable of retrieving information using the basic principles of the Semantic Web was obtained.

Keywords: Semantic Web, RDF Application, Knowledge representation.

1 INTRODUCTION

This article analyzes the technologies of the Semantic Web *and Linked Data* through the development of an application with a graphical interface, makes a brief review of the literature and performs the modeling of knowledge using a relational database, an ontological vocabulary of public domain and a knowledge base of public domain, with the use of the RDF (*Resource Description Framework*) language.

2 *WEB* SEMANTICS AND ONTOLOGY

The Semantic Web is an approach to seek to represent the content of the internet in a way more easily processable by machine and proposes to use intelligent technological tools to take advantage of these representations (Berners-Lee et al., 2001) . It is "a set of standards and best practices for data sharing and the semantics of that data over the *Web* for use by applications." (DuCharme , 2013) It allows you to structure collections of information and inference rules to automate the reasoning process by inserting semantic content into the Internet so that *software* agents can perform tasks for users. It is considered an extension of the *current web*, which aims to enable cooperative work between people and computers (Berners-Lee et al., 2001) .

The Semantic Web depends on ontologies for the structuring of data and its main application is precisely to provide semantics to the Internet. Ontologies serve as metadata schemas that provide a controlled vocabulary of concepts, each with explicit, machine-processable semantic definition. The goal of an ontology is to optimize information retrieval. According to (Freitas, 2005) (Maedche & Staab , 2001) Maedche and Staab (2001), an ontology provides subsidies for data to be processed by machines in the Semantic Web.

The word "ontology" has its origin in the area of philosophy, and was used to denominate the branch of metaphysics that concerns that which exists. One of the most classic definitions is that of , according to which an ontology is an "explicit specification of a conceptualization". Conceptualization, in turn, can be defined as "a simplified and abstract view of the world it is intended to represent" (p. 199). Chart 1 informs some definitions found in the literature. (Blackburn + Marcondes, 1997) (Gruber, 1993)

Table 1: Some definitions of ontology

Concept	Author
An "explicit specification of a conceptualization".	Gruber (1993)
"A formal and explicit specification of a shared conceptualization."	Borst (1997)
"A set of structured terms that describe some topic or domain."	Swartout et al. (1996)
"A set of logical axioms designed to explain the intended meaning of a vocabulary."	Guarino (1997)
"A formal and explicit specification of a shared conceptualization."	Studer, Benjamins and Fensel (1998)
"A representative model of knowledge common to a specific area of expertise or knowledge of domain that provides an explicit description of conceptualizations."	Staab & Studer (2009)
"These are concepts, relationships, attributes, and hierarchies in a domain."	Hassan & Rashid (2021)
"It is the data models that are used to represent the semantics of domain concepts through the ontological term, such as classes (entities) and relationships (properties)."	Yahya, Breslin & Ali (2021)
"They are the result of shared knowledge organized in a way that is machine-readable, capturing a certain conceptualization of the world that is explicitly defined."	Khadir, Aliane & Guessoum (2021)

Source: The authors.

Secon, ontologies can be of the following types:

- Knowledge Representation (KR) Ontologies: are those that capture the primitive representation information used to formalize knowledge about a given KR paradigm, and can cite as examples the Frame Ontology *and the OKBC* ontology.
- General or common ontologies: These are used to represent reusable common sense knowledge across the domain.
- Top-level or *upper-level* ontologies: deal with very general concepts and provide general notions about which root terms in the existing ontology should be linked. An example is the IEEE *Standard Upper Ontology* –SUO.
- Domain ontologies: promote reuse in a specific domain and provide vocabularies about concepts within that domain.
- Task ontologies: are the ontologies that describe the vocabulary of a generic task or activity, by specializing terms from the *top-level ontologies.*
- Domain task ontologies (domain-specific): These are domain tasks that are reusable in each domain, but not in other domains, and are therefore independent applications.
- Method ontologies: promote the definition of relevant concepts and applied relationships to specify a reasoning process as well as perform a particular task.
- Application ontologies: These are dependent applications that contain all the definitions necessary to model the knowledge required for a particular application.

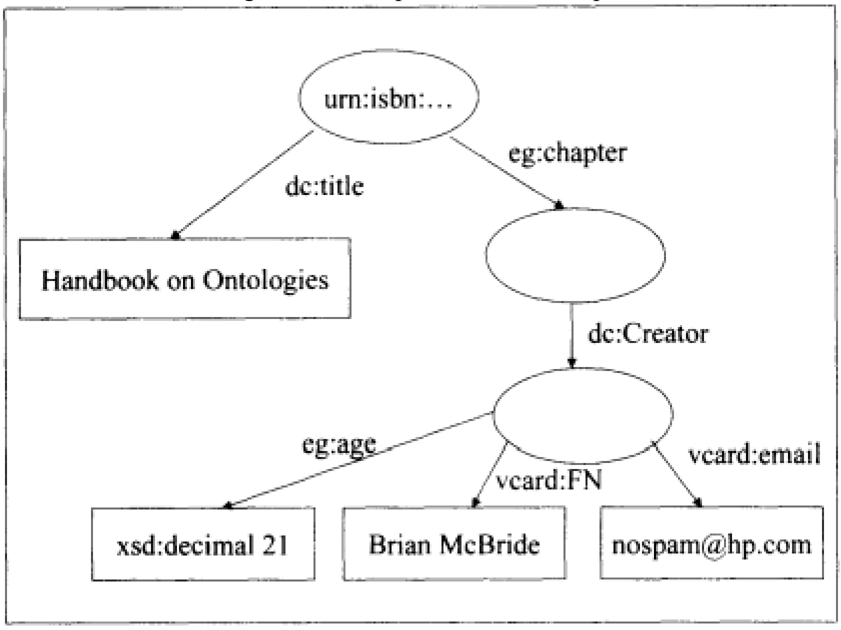
2.1 THE RDF LANGUAGE (*RESOURCE DESCRIPTION FRAMEWORK*)

RDF is a language for information representation that considers that everything being described has properties with values and that resources can be described through declarations. The party that identifies the object of the statement is the subject; the part that identifies a property or characteristic of a resource is called a predicate; and the part that identifies the value of a property is called an object. Because of this, RDF statements are also called RDF triples (de Lima & de Carvalho, 2005) .

RDF is a W3C recommendation, which defines a language for describing resources, forming the basis of the Semantic Web. Just as *the Web is a global infrastructure representing information in documents, the Semantic* Web *is an infrastructure representing information in a* form that can be processed by computer. The fact that the Semantic Web is decentralized imposes a severe restriction on the mechanism it uses to represent information (McBride, 2004) .

All statements are modeled as nodes and arcs in a graph. Figure 1 shows an example of an RDF graph. To record these graphs, the language uses an XML-based syntax, called RDF/XML, which is machine-processable. Through the URIs (*Uniform Resource Identification*), one piece of information can be linked to another on the web, but it can refer to anything, even an object that cannot be retrieved directly from the *Web* (de Lima; de Carvalho, 2005).

Figure 1 - Example of an RDF Graph



Source: McBride (2004).

3 METHODOLOGICAL ASPECTS

This research can be considered as exploratory, with a quantitative approach of technological origin. The Open Refine tools and the Virtuoso® database were used to support data preparation.

4 DISCUSSION AND RESULTS

To explore the potential of *Web* Semantic, the proposed activity was developed using data from a relational database, an ontology and information from the *DBPedia*, a project whose goal is to extract structured information from the *Wikipedia*. The authors chose to use a database made available by the Federal Highway Police where the records of traffic accidents that occurred on federal highways are found. For simplification purposes, the database was left with five fields: the highway or BR on which the accident occurred; the kilometer of the accident (Km); the type of accident; the presumed cause of the accident; the make of the vehicle involved in the accident; and the last field was composed of the URI of the description of the characteristics of that vehicle. Chart 2 shows some records for exemplification purposes.

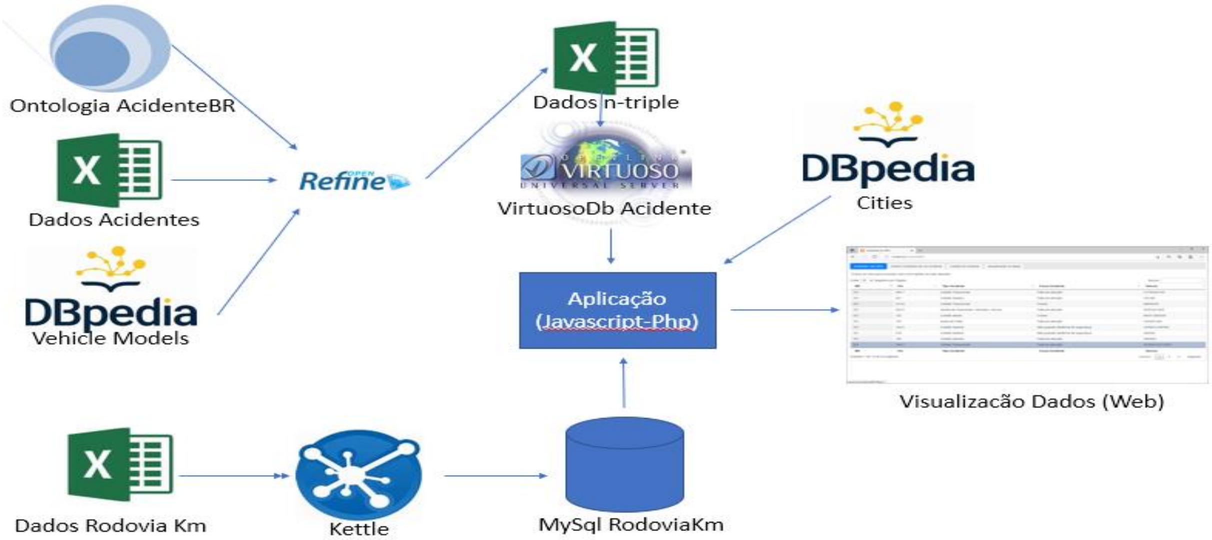
Table 2 – Relational data set

BR	MILES	Type of Accident	Cause of the Accident	Vehicle	URI
282	507,2	Head-on collision	Alcohol intake	GOAL	http://dbpedia.org/resource/Volkswagen_Gol
116	6,5	Rear-end collision	Lack of attention	JETTA	http://dbpedia.org/resource/Volkswagen_Jetta_(A7)
101	214,8	Transverse Collision	Other	AMAROK	http://dbpedia.org/resource/Volkswagen_Amarok
282	207,4	Rear-end collision	Incompatible speed	GOLF	http://dbpedia.org/resource/Volkswagen_Golf

Source: The authors.

Once the base table that served as the data source was assembled, it was necessary to transform it to RDF. Here, the tool *OpenRefine* was used for the task, through the appropriate plugin. This RDF was then imported into the database named Virtuoso[1]. The application *web* was developed using *Javascript* and PHP. The development process is explained in Figure 2.

Figure 2 - Application Development Process

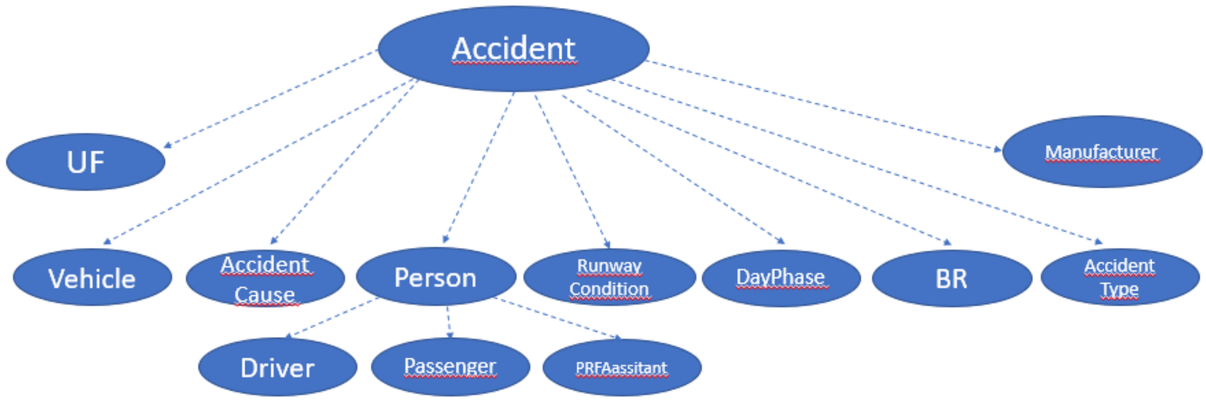


Source: The authors.

The ontology developed can be found in Figure 3. It contains more data than were actually used to populate the database, because the objective was to seek diverse information in the *DBPedia*. For reasons of simplification, additional fields were removed in the development of the application, without affecting the final result, which is to demonstrate the potentiality of the use of the *Web* Semantics. The hierarchy of the classes is shown in Figure 4.

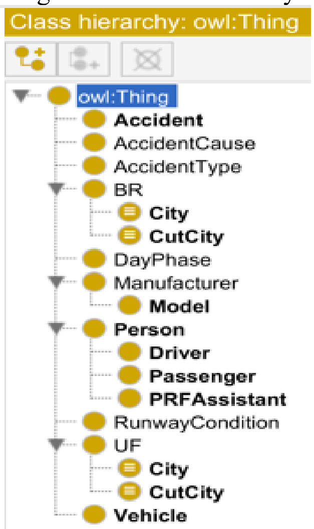
[1] Virtuoso® is a high-performance object-relational SQL database

Figure 3 - Accident Ontology



Source: the authors.

Figure 4 – Class hierarchy

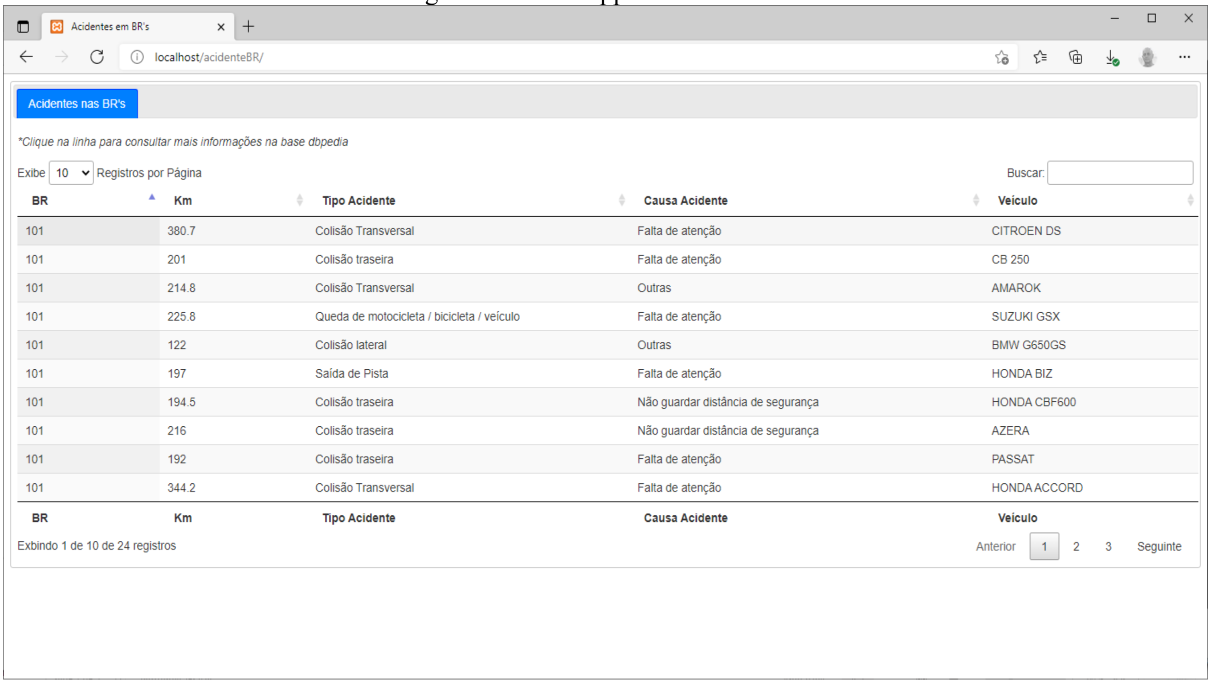


Source: The authors

When running the application, the following characteristics can be observed:

Initially, the screen is presented with the accidents in table format, with the columns BR, Km, Type Accident, Cause Accident and Vehicle, with the option to filter on all the fields of the table. Figure 5 shows this interface.

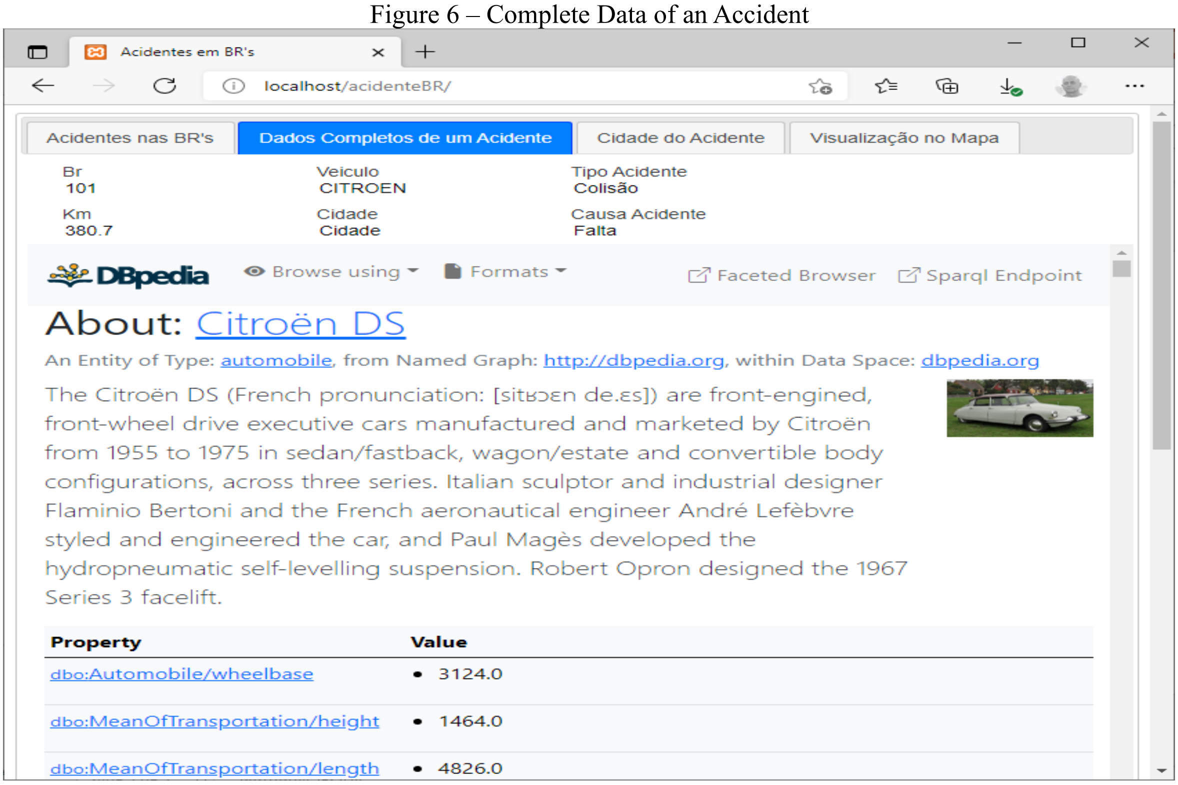
Figure 5 – Main Application Interface



Source: The authors

By clicking on a particular row of the table, the application searches for the data already linked in DBpedia, displayed in the "Complete Data of an Accident" tab, as shown in Figure 6.

Figure 6 – Complete Data of an Accident



Source: The authors

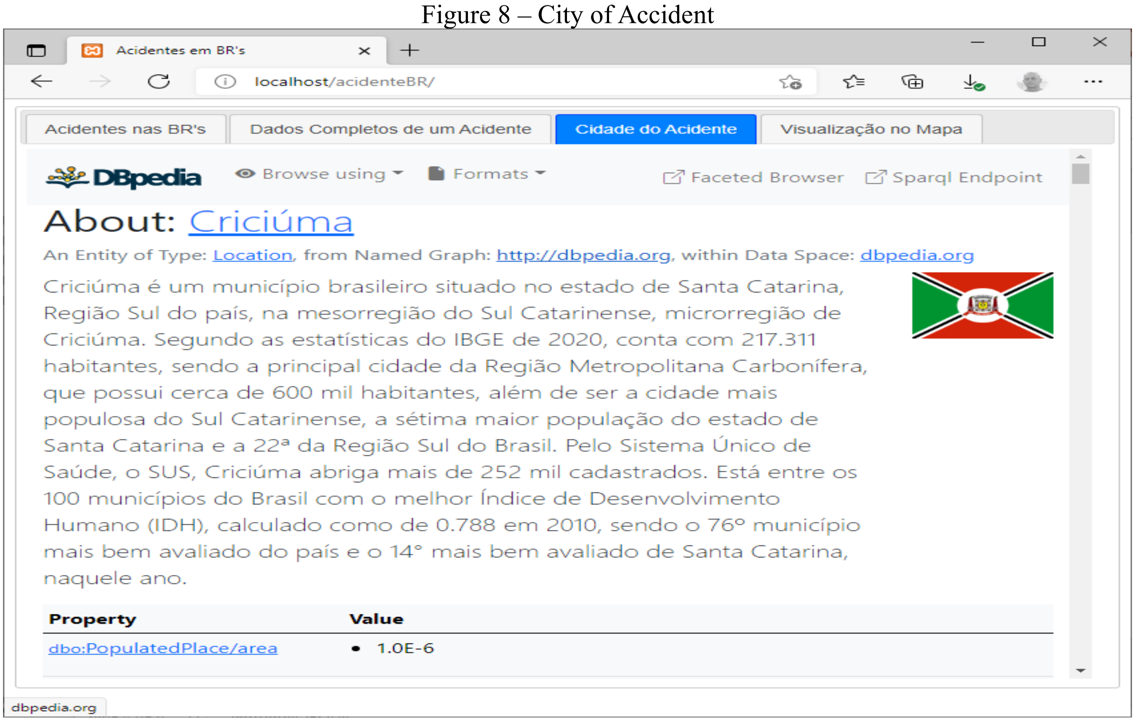
In the "City of Accident" tab, information is presented about the city in which the highway (BR) and the respective kilometer (Km) are located. This information (city) is obtained through DBpedia based on the longitude and latitude obtained from a relational database, through the georeferencing of the stretch of highway. A line of this base is shown in Figure 7 and the city data is presented in Figure 8.

Figure 7 – Data referring to the stretch of a given Km.

IDM_REGIAO	SGRODOVIA	CDTRECHO	DECONDICAO	SGSITUACAO	SGPRE	VLHORIZ	VLVERT	NUKMINICIO	NUKMFIM	NUEXTENSAO	VLLARGPIST	VLLARGACOS
22	BR101	141	Segmento Rural	ASF	101BSC4230	-28.7485596	-49.2738244	378492	388601	10109	7.2	2.5

Source: The authors

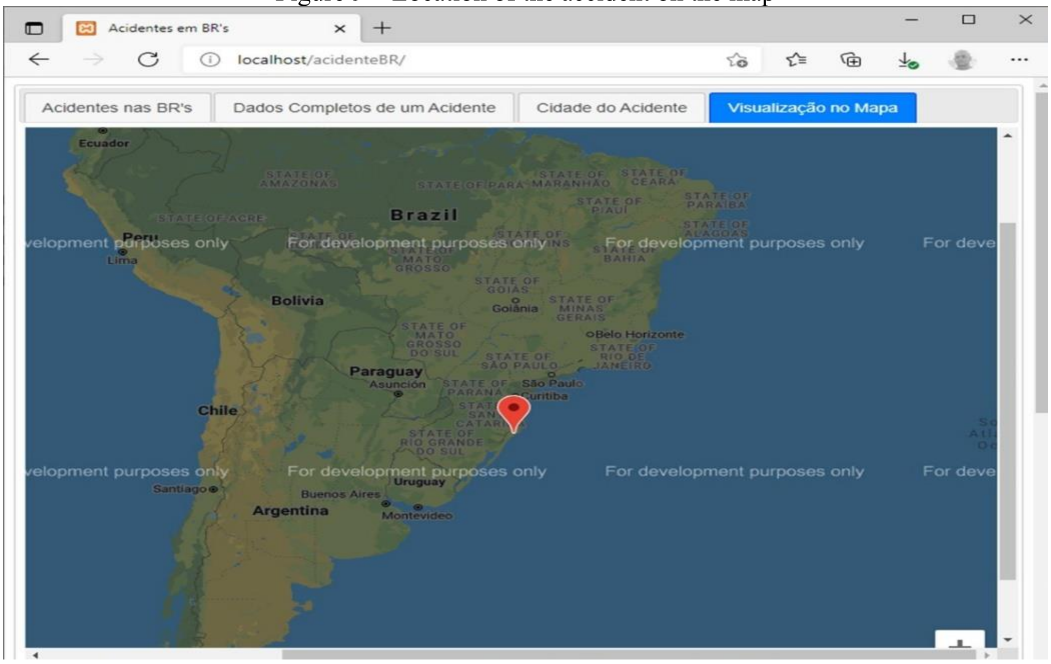
Figure 8 – City of Accident



Source: The authors

Finally, in the 'Map View' tab, (Figure 9) the location of the accident is displayed on the map according to the data reported.

Figure 9 – Location of the accident on the map



Source: The authors

5 CONCLUSIONS AND FUTURE WORK

The development of this application using the concepts, standards and tools of the Semantic Web allowed a deeper contact with the potential of the proposal to insert semantics in the data structure, to provide greater integration and interoperability of intelligent agents, which can interpret information. Putting into practice the fundamentals of the Semantic Web allowed us to develop notions about the data web and the usability of the *Linked Data architecture,* which explores how the data we deem most important can be restructured from databases to libraries of more comprehensive collections of facts about various entities and concepts.

As a suggestion for future work, the database can be expanded to support an application that uses several *Linked Data records*, in order to automatically obtain descriptions of resources that allow the aggregation of relevant information from various sources.

ACKNOWLEDGMENT

This study was partially funded by the Coordination for the Improvement of Higher Education Personnel - Brazil (Capes) – Financial Code 001.

REFERENCES

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. Scientific American, 284(5), 34–43.

Blackburn, S., & Marcondes, D. (1997). Consultoria da edição brasileira. In D. Dicionário Oxford de Filosofia. Tradução D. Murcho et al. Jorge Zahar.

Borst, W. N. (1997). Construction of engineering ontologies for knowledge sharing and reuse. In Centre for Telematics and Information Technology (CTIT). https://doi.org/10.1006/ijhc.1996.0096.

De Lima, J. C., & de Carvalho, C. L. (2005). Resource description framework (RDF). Technical report, Universidade Federal de Goiás.

DuCharme, B. (2013). Learning SPARQL: querying and updating with SPARQL 1.1. "O'Reilly Media, Inc.".

Freitas, F. L. G. (2005). Ontologias e a Web Semântica. 2005. URL http://www.inf.unisinos.br/renata/cursos/topicosv/ontologias-ws.pdf, 10(08).

Gómez-Pérez, A. (2004). Ontology evaluation. In Handbook on Ontologies (pp. 251–273). Springer.
Gruber, T. R. (1993). A translation approach to portable ontology specifications. Knowledge Acquisition, 5(2), 199–220.

Guarino, N. (1997). Understanding, Building and using Ontologies. International Jurnal of Human-Computer Studies. 46(2-3), 293-310.

Hassan, B. A., & Rashid, T. A. (2021). Artificial intelligence algorithms for natural language processing and the semantic web ontology learning. arXiv preprint arXiv:2108.13772.

Khadir, A. C., Aliane, H., & Guessoum, A. (2021). Ontology learning: Grand tour and challenges. Computer Science Review, 39, 100339.

Maedche, A., & Staab, S. (2001). Ontology Learning for Semantic Web. IEEE Intelligent Systems, 72–79. https://doi.org/10.1109/5254.920602.

McBride, B. (2004). The resource description framework (RDF) and its vocabulary description language RDFS. In Handbook on Ontologies (pp. 51–65). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-24750-0_3.

Staab S, Studer R. Handbook on Ontologies. 2nd ed. Heidelberg: Springer Science & Business Media; 2009.

Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: Principles and methods. Data & Knowledge Engineering, 25(1–2), 161–197. https://doi.org/10.1016/S0169-023X(97)00056-6.

Swartout, B., Patil, R., Knight, K., & Russ, T. (1996). Toward Distributed Use of Large-Scale Ontologies. Proc. of the Tenth Workshop on Knowledge Acquisition for Knowledge-Based Systems, 138–148.

Yahya, M., Breslin, J. G., & Ali, M. I. (2021). Semantic Web and Knowledge Graphs for Industry 4.0. Applied Sciences, 11(11), 5110. MDPI AG. Retrieved from http://dx.doi.org/10.3390/app11115110.