

UTILIZAÇÃO DE BASES DE CONHECIMENTO DE DOMÍNIO PÚBLICO: UM ESTUDO DE CASO SOBRE A LINGUAGEM RDF

Ronnie Carlos Tavares Nunes¹

Márcio Clemes²

Alexandre Leopoldo Gonçalves³

Rogério Cid Bastos⁴

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.

Resumo: O objetivo deste artigo é analisar as tecnologias de *Web Semântica* e *Linked Data*, com o desenvolvimento de uma aplicação web que realiza uma modelagem de conhecimento, através do uso de ontologias e da linguagem RDF (*Resource Description Framework*). Foi utilizada uma ontologia/vocabulário de conhecimento público populada com dados oriundos de uma base relacional do Brasil. Os dados agregados foram convertidos para o formato RDF e disponibilizados em um banco de dados neste formato. Foi utilizado o servidor universal Virtuoso® para esta tarefa, uma plataforma híbrida de banco de dados que possui as funcionalidades de um sistema tradicional de gerenciamento de banco de dados relacional, banco de dados objeto-relacional, banco de dados virtual, RDF, XML, texto livre, servidor de aplicações web e funcionalidade de servidor de arquivos em um único sistema. Como resultado, obteve-se uma aplicação web capaz de recuperar informações utilizando os princípios básicos da *Web Semântica*.

Palavras-chave: Web Semântica; Aplicação RDF; Representação do conhecimento.

¹Mestre em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina (UFSC), Florianópolis – Brasil. ORCID: <https://orcid.org/0000-0002-2373-511X>. E-mail: rocatan@ig.com.br

² Mestre em Engenharia de Produção, Universidade Federal de Santa Catarina - UFSC, Florianópolis – Brasil. ORCID: <https://orcid.org/0000-0003-1260-3901>. E-mail: marcio.clemes@ufsc.br

³ Professor Dr. do Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento – Universidade Federal de Santa Catarina (UFSC), Florianópolis – Brasil. ORCID: <https://orcid.org/0000-0002-6583-2807>. E-mail: a.l.goncalves@ufsc.br

⁴ Professor Dr. do Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento – Universidade Federal de Santa Catarina (UFSC), Florianópolis – Brasil. ORCID: <https://orcid.org/0000-0002-3775-1516>. E-mail: rogerio.bastos@ufsc.br

1. INTRODUÇÃO

Este artigo analisa as tecnologias da Web Semântica e *Linked Data* através do desenvolvimento de uma aplicação com interface gráfica, faz uma breve revisão da literatura e, realiza a modelagem do conhecimento utilizando um banco de dados relacional, um vocabulário ontológico de domínio público e uma base de conhecimento de domínio público, com o uso da linguagem RDF (*Resource Description Framework*).

2. WEB SEMÂNTICA E ONTOLOGIA

A *Web Semântica* é uma abordagem para busca representar o conteúdo da internet de uma maneira mais facilmente processável por máquina e propõe utilizar ferramentas tecnológicas inteligentes para tirar proveito dessas representações (Berners-Lee et al., 2001). É “um conjunto de padrões e melhores práticas para o compartilhamento de dados e a semântica desses dados pela *Web* para uso por aplicações” (DuCharme, 2013). Ela permite estruturar coleções de informações e regras de inferência para automatizar o processo de raciocínio, inserindo conteúdo semântico na Internet, de modo que os agentes de *software* possam realizar tarefas para os usuários. É considerada uma extensão da *web* atual, que visa possibilitar o trabalho cooperativo entre pessoas e computadores (Berners-Lee et al., 2001).

A *Web Semântica* depende de ontologias para a estruturação dos dados e a sua principal aplicação é justamente prover semântica à Internet (Freitas, 2005). Ontologias servem como esquemas de metadados que fornecem um vocabulário controlado de conceitos, cada um com definição semântica explícita e processável por máquina (Maedche & Staab, 2001). A meta de uma ontologia é otimizar a recuperação de informação. Segundo Maedche e Staab (2001), uma ontologia fornece subsídios para que os dados sejam processados por máquinas na *Web Semântica*.

A palavra “ontologia” tem sua origem na área de filosofia, e era usada para denominar o ramo da metafísica que diz respeito àquilo que existe (Blackburn & Marcondes, 1997). Uma das definições mais clássicas é a de (Gruber, 1993), segundo o qual uma ontologia é uma “especificação explícita de uma conceituação”. Conceituação, por sua vez, pode ser definida como “uma visão simplificada e abstrata do mundo a qual pretende-se representar” (p. 199). O Quadro 1 informa algumas definições encontradas na literatura.

Quadro 1: Algumas definições de ontologia

| Conceito | Autor |
|--|-------------------------------------|
| Uma “especificação explícita de uma conceituação”. | Gruber (1993) |
| “Uma especificação formal e explícita de uma conceitualização compartilhada”. | Borst (1997) |
| “Um conjunto de termos estruturados que descrevem algum tópico ou domínio”. | Swartout et al. (1996) |
| “Um conjunto de axiomas lógicos projetados para explicar o significado pretendido de um vocabulário”. | Guarino (1997) |
| “Uma especificação formal e explícita de uma conceituação compartilhada”. | Studer, Benjamins and Fensel (1998) |
| “Um modelo representativo de conhecimento comum a uma área específica de especialização ou conhecimento de domínio que fornece uma descrição explícita das conceitualizações”. | Staab & Studer (2009) |
| “São conceitos, relacionamentos, atributos e hierarquias em um domínio.” | Hassan & Rashid (2021) |
| “São os modelos de dados que são usados para representar a semântica dos conceitos de domínio através do termo ontológico, como classes (entidades) e relacionamentos (propriedades).” | Yahya, Breslin & Ali (2021) |
| “São o resultado de um conhecimento compartilhado organizado de forma a ser legível por máquina, capturando uma certa conceptualização do mundo que é explicitamente definida.” | Khadir, Aliane & Guessoum (2021) |

Fonte: Os autores.

Segundo (Gómez-Pérez, 2004), as ontologias podem ser dos seguintes tipos:

- Ontologias de Representação do Conhecimento (KR): são aquelas que capturam as informações primitivas de representação usadas para formalizar conhecimento sobre um dado paradigma de KR, podendo citar como exemplos a *Frame Ontology* e a *OKBC ontology*.

- Ontologias geral ou comum: são usadas para representar conhecimento de senso comum reutilizáveis através do domínio.
- Ontologias *top-level* ou *upper-level*: lidam com conceitos bem gerais e provêm noções gerais sobre quais termos da raiz na ontologia existente deve ser ligado. Um exemplo é a IEEE *Standard Upper Ontology* –SUO.
- Ontologias de domínio: promovem o reuso em um domínio específico e proporcionam vocabulários sobre conceitos dentro desse domínio.
- Ontologias de tarefa: são as ontologias que descrevem o vocabulário de uma tarefa genérica ou atividade, pela especialização de termos das ontologias *top-level*.
- Ontologias de tarefa de domínio (domínio específico): são tarefas de domínio reutilizáveis em um dado domínio, mas não em outros domínios, sendo, portanto, aplicações independentes.
- Ontologias de métodos: promovem a definição de conceitos relevantes e relações aplicadas para especificar um processo de raciocínio assim como realizar uma tarefa particular.
- Ontologias de aplicação: são aplicações dependentes que contém todas as definições necessárias para modelar o conhecimento requerido para uma aplicação particular.

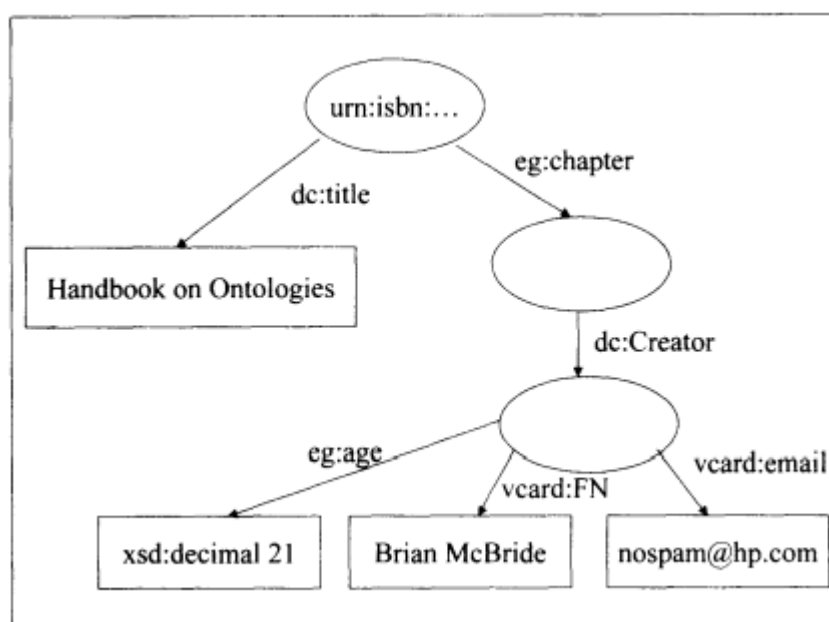
2.1 A LINGUAGEM RDF (*Resource Description Framework*)

RDF é uma linguagem para representação da informação que considera que tudo que está sendo descrito possui propriedades com valores e que recursos podem ser descritos através de declarações. A parte que identifica o objeto da declaração é o sujeito; a parte que identifica uma propriedade ou característica de um recurso é chamada predicado; e a parte que identifica o valor de uma propriedade é chamada de objeto. Por isso, as declarações RDF também são chamadas de triplas RDF (de Lima & de Carvalho, 2005).

A RDF é uma recomendação do W3C, que define uma linguagem para descrever os recursos, constituindo a base da *Web Semântica*. Assim como a *Web* é uma infraestrutura global representando informações em documentos, a *Web Semântica* é uma infraestrutura representando informações em uma forma que pode ser processada por computador. O fato de a *Web Semântica* ser descentralizada impõe uma severa restrição sobre o mecanismo que utiliza para representar a informação (McBride, 2004).

Todas as declarações são modeladas como nós e arcos em um grafo. A Figura 1 mostra um exemplo de grafo RDF. Para registrar esses grafos, a linguagem utiliza uma sintaxe baseada em XML, chamada de RDF/XML, processável por máquina. Através das URIs (*Uniform Resource Identification*), uma informação pode ser ligada a outra na web, mas ela pode se referir a qualquer coisa, mesmo um objeto que não pode ser recuperado diretamente da *Web* (de Lima; de Carvalho, 2005).

Figura 1 - Exemplo de um grafo RDF



Fonte: McBride (2004).

3. ASPECTOS METODOLÓGICOS

Essa pesquisa pode ser considerada como exploratória, com abordagem quantitativa de origem tecnológica. Como suporte para a preparação dos dados foram utilizadas as ferramentas Open Refine e o banco de dados Virtuoso®.

4. DISCUSSÃO E RESULTADOS

Para se explorar o potencial das ferramentas de *Web Semântica*, a atividade proposta foi desenvolvida utilizando dados provenientes de um banco de dados relacional, uma ontologia e informações da *DBPedia*, um projeto cujo objetivo é extrair informações estruturadas da *Wikipedia*. Os autores optaram por utilizar uma base de dados disponibilizada pela Polícia

Rodoviária Federal onde são encontrados os registros de acidentes de trânsito ocorridos em rodovias federais. Para fins de simplificação, a base de dados ficou com cinco campos: a rodovia ou BR na qual aconteceu o acidente; o quilômetro do acidente (Km); o tipo do acidente; a causa presumível do acidente; a marca do veículo envolvido no acidente; e o último campo foi composto pela URI da descrição das características daquele veículo. O Quadro 2 mostra alguns registros para fins de exemplificação.

Quadro 2 – Conjunto de dados relacionais

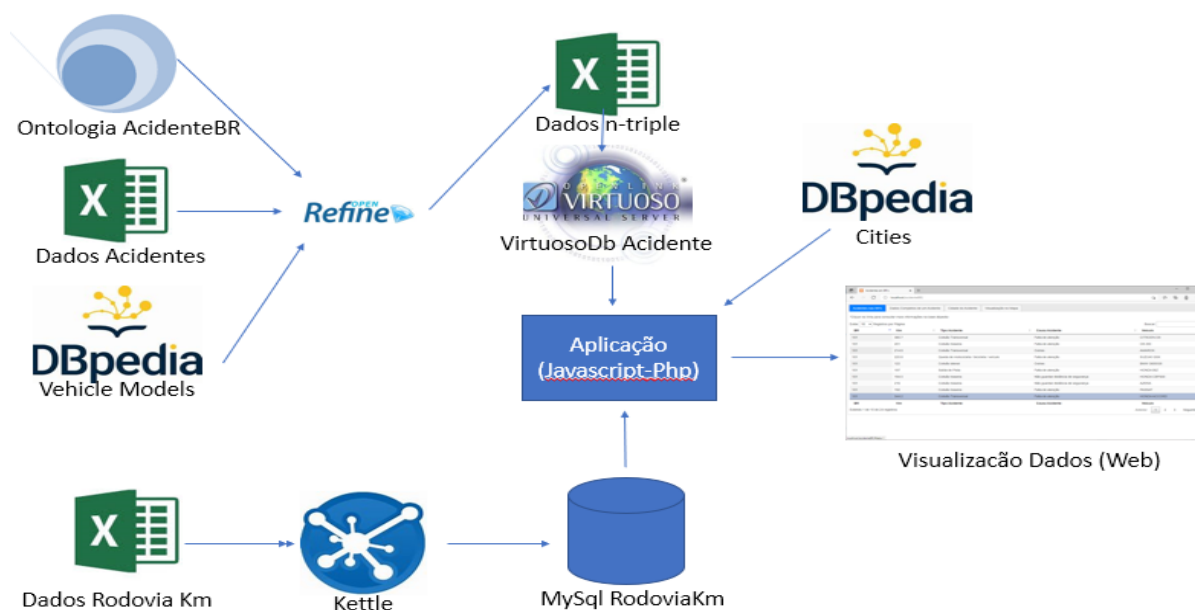
| BR | KM | Tipo de Acidente | Causa do Acidente | Veículo | URI |
|-----|-------|---------------------|-------------------------|---------|---|
| 282 | 507,2 | Colisão frontal | Ingestão de álcool | GOL | http://dbpedia.org/resource/Volkswagen_Gol |
| 116 | 6,5 | Colisão traseira | Falta de atenção | JETTA | http://dbpedia.org/resource/Volkswagen_Jetta_(A7) |
| 101 | 214,8 | Colisão Transversal | Outras | AMAROK | http://dbpedia.org/resource/Volkswagen_Amarok |
| 282 | 207,4 | Colisão traseira | Velocidade incompatível | GOLF | http://dbpedia.org/resource/Volkswagen_Golf |

Fonte: Os autores.

Uma vez montada a tabela base que serviu como origem dos dados, foi necessário transformá-la para RDF. Aqui, a ferramenta *OpenRefine* foi utilizada para a tarefa, através do plugin apropriado. Esse RDF então foi importado para o banco de dados chamado Virtuoso⁵. A aplicação *web* foi desenvolvida utilizando *Javascript* e PHP. O processo de desenvolvimento é explicitado na Figura 2.

⁵ Virtuoso® é um banco de dados SQL objeto-relacional de alto desempenho

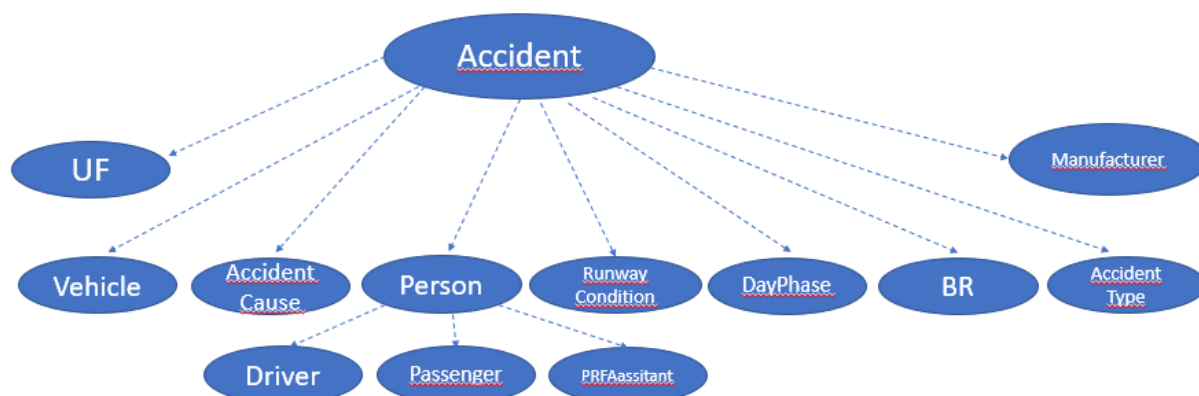
Figura 2 - Processo de desenvolvimento da aplicação



Fonte: Os autores.

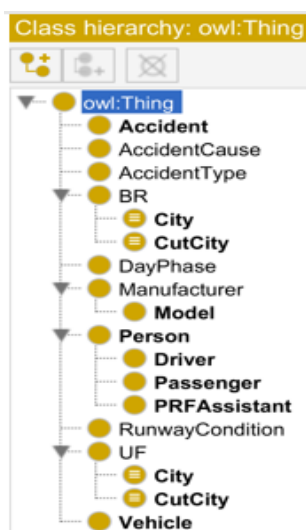
A ontologia desenvolvida pode ser encontrada na Figura 3. Nela constam mais dados do que foram efetivamente utilizados para popular a base de dados, em virtude de que o objetivo era buscar informações diversas na *DBPedia*. Por motivos de simplificação, campos adicionais foram removidos no desenvolvimento da aplicação, sem afetar o resultado final, que é demonstrar a potencialidade do uso da *Web Semântica*. A hierarquia das classes é mostrada na Figura 4.

Figura 3 - Ontologia de Acidentes



Fonte: os autores.

Figura 4 – Hierarquia das classes



Fonte: Os autores

Ao ser executada a aplicação, pode-se observar as seguintes características:

Inicialmente é apresentada a tela com os acidentes em formato de tabela, com as colunas BR, Km, Tipo Acidente, Causa Acidente e Veículo, com a opção de filtrar em todos os campos da tabela. A Figura 5 apresenta esta interface.

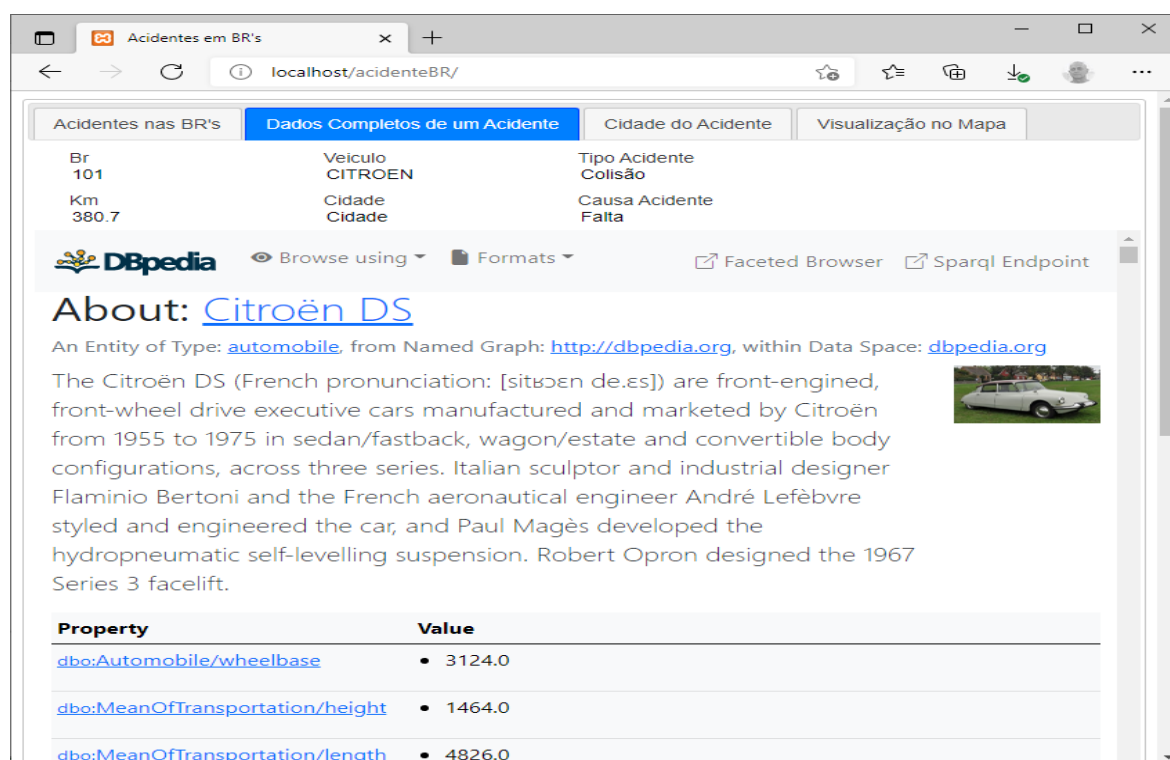
Figura 5 – Interface Principal da Aplicação

| BR | Km | Tipo Acidente | Causa Acidente | Veículo |
|-----|-------|--|------------------------------------|--------------|
| 101 | 380.7 | Colisão Transversal | Falta de atenção | CITROEN DS |
| 101 | 201 | Colisão traseira | Falta de atenção | CB 250 |
| 101 | 214.8 | Colisão Transversal | Outras | AMAROK |
| 101 | 225.8 | Queda de motocicleta / bicicleta / veículo | Falta de atenção | SUZUKI GSX |
| 101 | 122 | Colisão lateral | Outras | BMW G650GS |
| 101 | 197 | Saída de Pista | Falta de atenção | HONDA BIZ |
| 101 | 194.5 | Colisão traseira | Não guardar distância de segurança | HONDA CBF600 |
| 101 | 216 | Colisão traseira | Não guardar distância de segurança | AZERA |
| 101 | 192 | Colisão traseira | Falta de atenção | PASSAT |
| 101 | 344.2 | Colisão Transversal | Falta de atenção | HONDA ACCORD |

Fonte: Os autores

Ao clicar em uma determinada linha da tabela, a aplicação busca os dados já linkados na DBpedia, exibidos na aba “Dados Completos de um Acidente”, conforme apresentado na Figura 6.

Figura 6 – Dados Completos de um Acidente



Fonte: Os autores

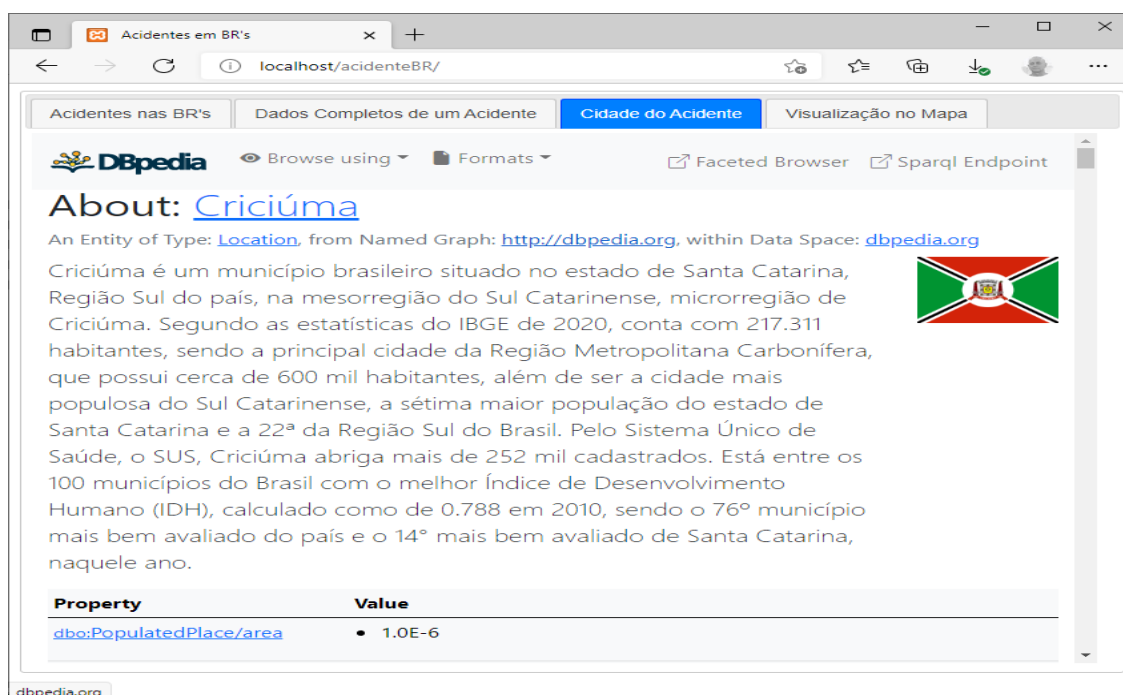
Na aba “Cidade do Acidente”, são apresentadas informações sobre a cidade em que a rodovia (BR) e o respectivo quilometro (Km) estão localizados. Esta informação (cidade) é obtida através da DBpedia com base na longitude e latitude obtidas de uma base de dados relacional, através do georreferenciamento do trecho da rodovia. Uma linha desta base é apresentada na Figura 7 e os dados da cidade são apresentados na Figura 8.

Figura 7 – Dados referentes ao trecho de um determinado 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 |

Fonte: Os autores

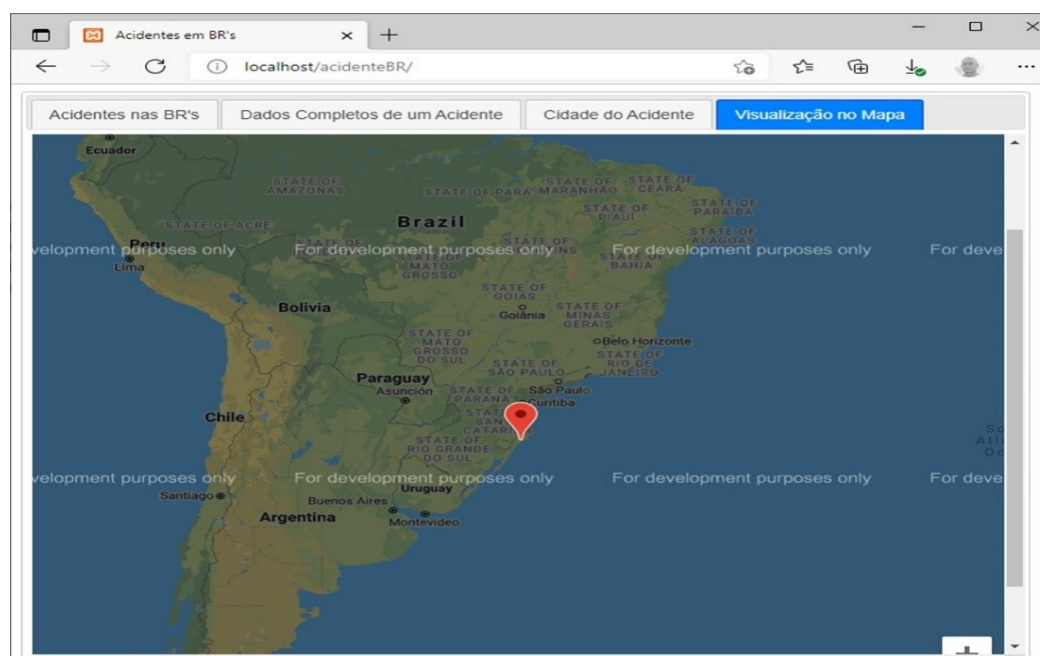
Figura 8 – Cidade do Acidente



Fonte: Os autores

Finalizando, na aba “Visualização no Mapa”, (Figura 9) é exibida a localização do acidente no mapa de acordo com os dados informados.

Figura 9 – Localização do acidente no mapa



Fonte: Os autores

5. CONCLUSÕES E TRABALHOS FUTUROS

O desenvolvimento desta aplicação utilizando os conceitos, padrões e ferramentas da *Web Semântica* possibilitou um contato mais aprofundado com a potencialidade da proposta de inserir semântica na estrutura de dados, para prover maior integração e interoperabilidade dos agentes inteligentes, que podem interpretar a informação. Colocar em prática os fundamentos da *Web semântica* permitiu desenvolver noções sobre a web de dados e a usabilidade da arquitetura de *Linked Data*, que explora como os dados que julgamos mais importantes podem ser reestruturados a partir de bancos de dados para bibliotecas de coleções de fatos mais abrangentes sobre diversas entidades e conceitos.

Como sugestão de trabalhos futuros, pode-se ampliar o banco de dados para suportar uma aplicação que utiliza diversos registros de *Linked Data*, no intuito de obter automaticamente descrições de recursos que permitam agregar informações relevantes de diversas fontes.

AGRADECIMENTOS

Este estudo foi parcialmente financiado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (Capes) – Código financeiro 001.

REFERÊNCIAS

- 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 Journal 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](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>.