

OntoCarBR: Uma Ontologia sobre Carros vendidos no Brasil

Daniel Lucas Albuquerque¹, Vinícius Rosa Buzzi¹, Renato Fileto¹

¹Dep. de Informática e Estatística (INE), Univ. Federal de Santa Catarina (UFSC)
Campus Reitor João David Ferreira Lima (Trindade), Florianópolis-SC – Brasil

daniel.lucas.f@posgrad.ufsc.br, vinicios.buzzi@gmail.com,
r.fileto@ufsc.br

Abstract. *Obtaining information about features of cars marketed in Brazil for reuse in applications is currently challenging due to the unavailability of open, complete, structured, and up-to-date data. This work presents OntoCarBR, a domain ontology aimed at facilitating the reuse, integration, and retrieval of car information for future applications such as text-based information retrieval and fraud detection. OntoCarBR was evaluated on the integration of two datasets, enabling easy retrieval of car records from the FIPE Table based on attributes that were not originally present in this dataset.*

Resumo. *Obter informações sobre características de carros comercializados no Brasil para reuso em aplicações é atualmente difícil devido à indisponibilidade de dados abertos, completos, estruturados e atualizados. Este trabalho apresenta a OntoCarBR, uma ontologia de domínio cujo objetivo é facilitar o reuso, integração e recuperação de informação, visando aplicações futuras como recuperação de informação de textos e detecção de fraudes. A OntoCarBR foi avaliada na integração de dois conjuntos de dados, permitindo recuperar registros de carros da Tabela FIPE com facilidade a partir de atributos que não estavam originalmente presentes em tal tabela.*

1. Introdução

Obter informações sobre características de carros comercializados no Brasil, a partir da Tabela FIPE, representa um desafio. Dois problemas são facilmente observados no sistema original: (1) a não disponibilização de uma API oficial documentada¹ dificultando a obtenção dos dados sobre cada modelo de carro; e (2) informações como o próprio modelo, pacotes de opcionais, cilindrada, potência, quantidade de portas, transmissão, carroceria, entre outras, são agrupadas em um campo semi-estruturado, tornando difícil a identificação automatizada dos carros.

Ontologias têm como um dos principais objetivos o compartilhamento de um entendimento comum da estrutura da informação entre pessoas e agentes [Gruber 1993]. Este trabalho apresenta a OntoCarBR, uma ontologia para representação de carros no cenário brasileiro. A OntoCarBR tem como objetivo estruturar informações de carros provenientes da Tabela FIPE, garantindo interoperabilidade entre diferentes fontes de dados e proporcionando o enriquecimento semântico da informação.

¹As APIs <https://fipeapi.com.br/>, <https://apifipe.com.br> e <https://brasilapi.com.br>, embora documentadas, não são oficiais.

A OntoCarBR foi avaliada de forma qualitativa quanto à facilidade de acesso à informação e capacidades de integração e recuperação de registros de carros. Para isso, o trabalho instanciou um Grafo de Conhecimento (KG) a partir da OntoCarBR, e realizou Consultas de avaliação no *endpoint* SPARQL gerado. Os dados utilizados para gerar o KG foram obtidos de uma API não documentada da Tabela FIPE e de um conjunto de dados adicional. A seção 2 discute trabalhos relacionados, a seção 3 detalha o método utilizado, a seção 4 descreve a avaliação experimental e a seção 5 apresenta a conclusão.

2. Trabalhos Relacionados

Ontologias são descrições explícitas e formais em um domínio de discurso [Noy et al. 2001]. Elas provêm mecanismos para capturar informações sobre classes ou objetos e relações entre eles [Obiniyi et al. 2014]. Podem definir esquemas para Grafos de Conhecimento, a abordagem padrão para adicionar conhecimento em sistemas de Inteligência Artificial e resolver tarefas complexas [Peng et al. 2023].

Diversas ontologias foram propostas para representar características de veículos. A *iCity* [Katsumi and Fox 2020] propõe uma estrutura para veículos, posições geográficas e outros conceitos no domínio urbano, com foco na integração com outras ontologias para aplicações inteligentes. A *Vehicle Sales Ontology* (VSO)² representa diversos tipos de veículos no e-commerce, estruturando preços e períodos de validade. Ela opera sobre a GoodRelations (GR) [Hepp 2008], um vocabulário padrão para modelagem de produtos e serviços na Web. A Generic Automotive Ontology (GAO)³, coordenada pelo W3C, busca melhorar a interoperabilidade de dados na indústria automotiva, promovendo extensões para o Schema.org. A GAO origina ontologias específicas, como a VSO, a *Car Options Ontology* (COO) e a *Used Cars Ontology* (UCO).

3. Método de Construção da OntoCarBR

A OntoCarBR foi construída utilizando a metodologia de Noy e McGuiness [Noy et al. 2001] e usou a representação *Web Ontology Language* (OWL), o padrão W3C para representação de ontologias. Para a edição da OntoCarBR foi utilizada a ferramenta Protégé [Musen 2015]. Dada a necessidade de representar os preços dos carros, a OntoCarBR reusa a *Vehicle Sales Ontology* (VSO). A VSO define sua principal classe, *Vehicle*, como subclasse de *ProductOrService* da ontologia GoodRelations, permitindo a herança de importantes propriedades desta, tais como *hasBrand*, *hasManufacturer*, entre outras. Usando a propriedade *owl:equivalentClass*, a OntoCarBR define a classe *Carro* como equivalente à classe *Automobile*, subclasse de *Vehicle*. A figura 1 ilustra as principais classes e propriedades da OntoCarBR.

Os conjuntos de dados utilizados para instanciar classes da OntoCarBR foram a Tabela FIPE e um conjunto de dados adicional sobre carros produzidos no mundo entre 1945 e 2020. Os dados da Tabela FIPE foram obtidos através de um *script* Python que realizou requisições REST sobre uma API⁴ não documentada e obteve 28582 registros de carros (tabela agosto de 2024). Os dados do conjunto de dados adicional foram obtidos diretamente da plataforma Kaggle, em formato *csv*⁵, com 70820 registros.

²URI Vehicle Sales Ontology (VSO): <http://purl.org/vso/ns>

³Generic Automotive Ontology (GAO): <https://www.w3.org/community/gao/>

⁴<https://veiculos.fipe.org.br/api/veiculos//ConsultarMarcas>

⁵<https://www.kaggle.com/datasets/jahaidulislam/car-specification-dataset-1945-2020>

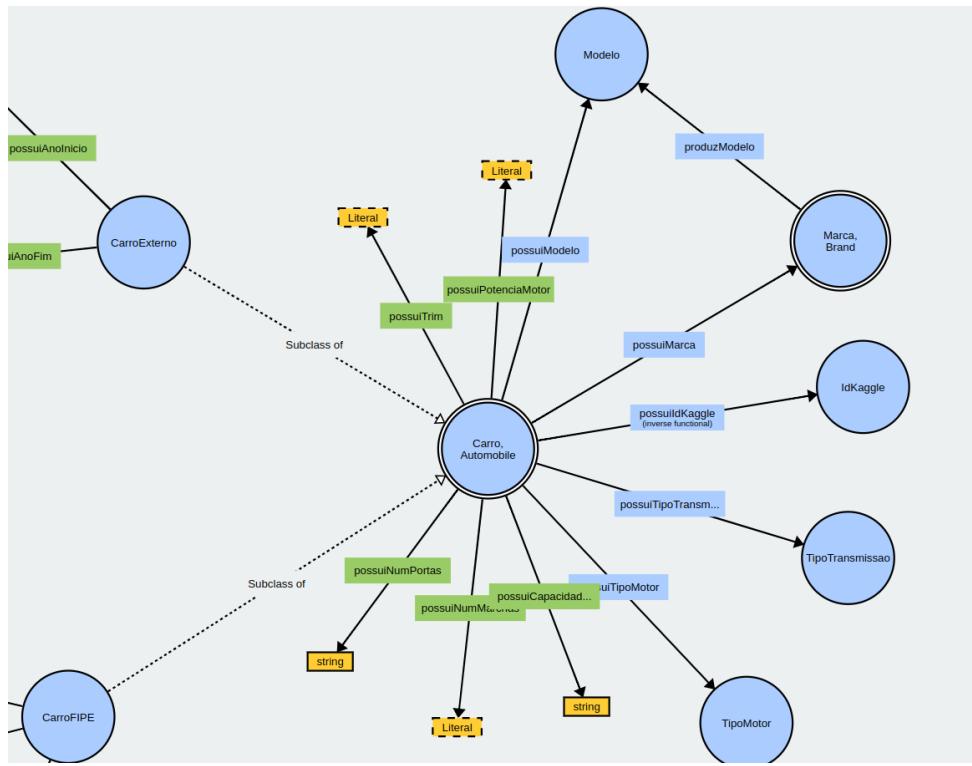


Figura 1. Principais classes e propriedades da OntoCarBR.

Após a obtenção dos dados da Tabela FIPE, o trabalho desmembrou as informações do campo semi-estruturado *modelo*, gerando os novos campos *trim*, *capacidade_motor* (cilindrada), *potencia*, *combustivel* (adicionando os valores *híbrido* e *elétrico*), *portas*, *transmissao*, *carroceria*, *valvulas* e *cilindros*. Os exemplos abaixo ilustram como os dados estavam expressos no campo original *modelo*:

“206 soleil 1.6 16v 110cv 5p” ou “s10 pick-up luxe 4.3 v6 ce”

Em seguida, foi aplicado o conceito de *Data Mashup* para definir equivalências entre os registros dos dois conjuntos de dados e integrá-los. Partindo dos registros da Tabela FIPE, foi concebido um algoritmo recursivo para casamento de registros, o qual aplica filtros baseados em uma sequência predefinida de campos e monitora a quantidade de registros retornados em cada etapa. O processo de integração encontrou 3160 equivalências únicas e 4398 casamentos alternativos que requerem desambiguação.

4. Avaliação Experimental

O fluxo de dados da solução até a disponibilização de um *endpoint* SPARQL é representado na figura 2.a. O trabalho utilizou a ferramenta Ontop [Xiao et al. 2020] versão 5.2.0 para gerar um KG Virtual (VKG). Os dados da Tabela FIPE e do conjunto adicional foram instanciados na representação definida pela OntoCarBR. Uma Consulta de avaliação definida como q_1 : “*FIAT 500 2012 1.4 16V 6 MARCHAS*” foi comparada em três cenários distintos: (1) Site da FIPE; (2) OntoCarBR sem enriquecimento dos dados; (3) OntoCarBR com enriquecimento dos dados. O único cenário capaz de satisfazer a necessidade de informação foi o cenário (3) (figura 2.b), em função da equivalência definida entre os registros da Tabela FIPE e do conjunto adicional, utilizando a propriedade *possuiNumMarchas* definida na OntoCarBR.

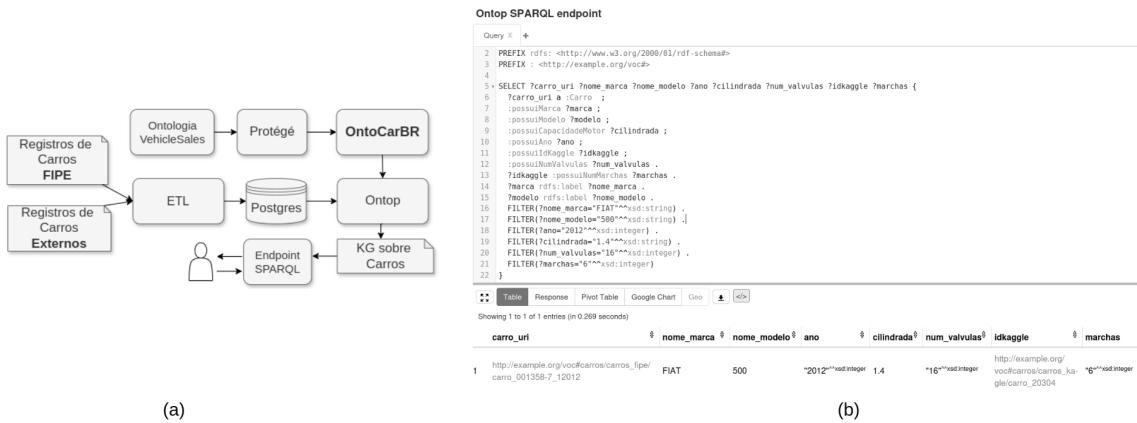


Figura 2. (a) Fluxo de dados; (b) Consulta de avaliação sobre Endpoint SPARQL.

5. Conclusão

O trabalho apresentou a OntoCarBR, uma ontologia sobre carros. Ela se mostrou eficaz para representar, integrar e recuperar informações de carros no cenário brasileiro. Trabalhos futuros incluem: aperfeiçoamento do algoritmo de casamento de registros das diferentes fontes de dados, experimentos mais extensivos para aferir métricas quantitativas de desempenho na recuperação de informação e utilização do grafo de conhecimento produzido para auxiliar na extração de informação de textos descrevendo veículos, tais como descrições de itens em notas fiscais eletrônicas.

Referências

- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2):199–220.
- Hepp, M. (2008). Goodrelations: An ontology for describing products and services offers on the web. In *Knowledge Engineering: Practice and Patterns: 16th Intl. Conf. EKAW 2008*, pages 329–346. Springer.
- Katsumi, M. and Fox, M. (2020). icity transportation planning suite of ontologies. *University of Toronto*.
- Musen, M. A. (2015). The protégé project: a look back and a look forward. *AI matters*, 1(4):4–12.
- Noy, N. F., McGuinness, D. L., et al. (2001). Ontology development 101: A guide to creating your first ontology.
- Obiniyi, A., Oyelade, O., and Auta, G. (2014). Ontology languages, development tools and repositories: Towards a unifying platform. *Intl. Journal of Computer Science and Artificial Intelligence*, 4(3):68.
- Peng, C., Xia, F., Naseriparsa, M., and Osborne, F. (2023). Knowledge graphs: Opportunities and challenges. *Artificial Intelligence Review*, 56(11):13071–13102.
- Xiao, G., Lanti, D., Kontchakov, R., Komla-Ebri, S., Güzel-Kalaycı, E., Ding, L., Corman, J., Cogrel, B., Calvanese, D., and Botoeva, E. (2020). The virtual knowledge graph system ontop. In *Intl. Semantic Web Conference*, pages 259–277. Springer.