

DADOS ABERTOS CONECTADOS UTILIZANDO A METODOLOGIA LODMBP

Data de submissão: 09/02/2024

Data de aceite: 01/04/2024

Luara Giovanna Peterlini

Universidade Estadual do Centro-Oeste –
UNICENTRO
Departamento de Ciência da Computação
- DECOMP
Guarapuava - Paraná
<http://lattes.cnpq.br/0335688686747551>

Josiane Michalak Hauagge Dall’Agnol

Universidade Estadual do Centro-Oeste –
UNICENTRO
Departamento de Ciência da Computação
- DECOMP
Guarapuava - Paraná
<http://lattes.cnpq.br/7599246358679552>

Gisane Aparecida Michelon

Universidade Estadual do Centro-Oeste –
UNICENTRO
Departamento de Ciência da Computação
- DECOMP
Guarapuava - Paraná
<http://lattes.cnpq.br/9367680247741161>

Lucélia de Souza

Universidade Estadual do Centro-Oeste –
UNICENTRO
Departamento de Ciência da Computação
- DECOMP
Guarapuava - Paraná
<http://lattes.cnpq.br/8989736627689345>

RESUMO: Nesta pesquisa investigou-se sobre as Melhores Práticas para Dados na Web, segundo o Consórcio W3C (*World Wide Web Consortium*), bem como sobre a interligação de Dados Abertos e as ferramentas mais adequadas para essa tarefa. Em seguida, aplicou-se essa abordagem à interligação de um conjunto de dados das rotas de ônibus que trafegam nas IES públicas da cidade de Guarapuava/PR, de acordo com a Metodologia para LOD baseada nas Melhores Práticas W3C, aplicando conceitos de Dados Abertos Conectados e da Web Semântica.

PALAVRAS-CHAVE: Dados Abertos Conectados, Web Semântica, Metodologia LODMBP, Interligação de dados, Melhores Práticas W3C.

LINKED OPEN DATA USING THE LODMBP METHODOLOGY

ABSTRACT: This research investigated the Data on the Web Best Practices, according to W3C, as well as the interconnection of Open Data and the most appropriate tools for this task. This approach was then applied to the interconnection of a bus routes dataset that serve public Higher Education Institutions in the city of Guarapuava, according with the LOD

Methodology based on W3C Best Practices, addressing concepts of Linked Open Data and of Semantic Web.

KEYWORDS: Linked Open Data, Semantic Web, LODMBP Methodology, linked data, Best Practices W3C.

INTRODUÇÃO

A tecnologia atual está em constante crescimento e evolução, com dados sendo publicados e compartilhados frequentemente. No entanto, surge um desafio em relação a esse tema: é necessário publicar esses dados de forma que sejam facilmente compreensíveis e utilizáveis.

Para enfrentar esse desafio, a Web Semântica surge como uma solução. Proposta por Berners-Lee et al. (2000), é uma abordagem que introduz uma nova forma de publicação e busca de informações no ciberespaço, expandindo as fronteiras da Web ao infundir significado na representação dos dados.

Os dois principais pilares da Web Semântica referem-se i-) ao conceito de Dados Abertos Conectados (*Linked Open Data* – LOD) (Berners-Lee, 2006), que consiste em um conjunto de melhores práticas para a publicação de dados na Web; e, ii-) a interligação de dados na Web, a qual possibilita a interoperabilidade dos conjuntos de dados.

Dessa forma, a Web deixa de ser apenas um espaço global de documentos interligados, para se transformar em um vasto espaço global de dados conectados (Heath; Bizer, 2011).

Nesse sentido, este trabalho tem como objetivo desenvolver uma aplicação em LOD voltada para a interligação de dados visando ampliar o contexto da informação e possibilitar a descoberta de novas informações e o aproveitamento mais eficiente dos dados, buscando aproveitar as vantagens das tecnologias atuais e da Web Semântica.

REVISÃO BIBLIOGRÁFICA

Dados abertos conectados

São dados que estão livremente disponíveis na Web para reuso e redistribuição, estando interligados com outros dados, possibilitando a criação de novos conhecimentos (Berners-Lee, 2006).

Em 2010, Berners-Lee sugeriu um Esquema de Classificação 5 Estrelas para nortear a publicação de Dados Abertos Conectados:

1. Os dados devem estar disponíveis na Internet (em qualquer formato; por exemplo, PDF), desde que com licença aberta.
2. Além da regra 1, devem estar disponíveis de maneira estruturada (por exemplo, um arquivo Excel).

3. Além da regra 2, devem estar disponíveis de maneira estruturada e em formato não proprietário (CSV em vez de Excel).
4. Seguindo todas as regras anteriores, usar URIs¹ para identificar coisas, de forma que os dados possam ser encontrados e consumidos.
5. Todas as regras anteriores, além de interligar seus dados a outros dados, de forma a fornecer um contexto ampliado de informações.

A ideia do autor é que, a cada nova estrela alcançada, os dados tornam-se progressivamente mais utilizáveis, aumentando sua facilidade de uso.

Metodologia LODMBP

A Metodologia para LOD baseada nas Melhores Práticas W3C – LODMBP (Dall’Agnoli; Tacla, 2020) propõem um ciclo de vida para Dados Abertos Conectados. A LODMBP é fundamentada nas 35 Melhores Práticas para Dados na Web (*Data on the Web Best Practices* – DWBPs) (Lóscio; Burle; Calegari, 2017), as quais orientam a criação, publicação e utilização de LOD.

As DWBPs são reconhecidas internacionalmente pela abrangência e pelo detalhamento com que tratam a publicação de dados na Web. Por exemplo a **DWBP1 – fornecer metadados**, apresenta informações relativas ao fornecimento de metadados para auxiliar os humanos a compreenderem os dados e as máquinas a processá-los. As demais DWBPs podem ser consultadas em (Lóscio; Burle; Calegari, 2017).

A LODMBP foi construída mediante a fusão de diversas metodologias de geração e publicação de LOD, sendo que as DWBPs são aplicadas diretamente como elemento essencial em suas etapas. O propósito subjacente a essa metodologia é guiar a incorporação eficaz das DWBPs durante as várias fases do processo LOD, com a finalidade de auxiliar os provedores de dados na definição de uma sequência coerente de aplicação das DWBPs. Isso inclui direcionar por onde iniciar e quais medidas tomar prioritariamente dentro do ciclo de vida do LOD.

A LODMBP é composta por sete etapas:

- **Extração:** Compreende as atividades de obter os dados existentes para, em seguida, extraí-los e armazená-los em repositórios específicos. A DWBP aplicada é a 33.
- **Preparação:** Engloba as ações relacionadas à formatação dos dados. Pode envolver a necessidade de depurar os dados, ou seja, remover informações que não serão divulgadas, assim como converter dados de um formato para outro ou uniformizar os valores nos dados, além de gerar novos dados a partir dos dados já presentes no conjunto. As DWBPs aplicadas são: 9 e 10.
- **Semantificação:** Envolve as ações de conferir significado aos dados, os quais estavam previamente abordados de modo puramente sintático. A semantificação viabiliza a conversão de dados não processados em informações assimiláveis por sistemas computacionais. As DWBPs aplicadas são: 1, 2, 3, 4, 5, 6, 7, 15, 16 e 28.

¹ URI: Universal Resource Identifier

- **Geração:** Realiza a geração dos dados, os quais foram enriquecidos com metadados e informações sobre licença, procedência, versionamento, qualidade entre outras. As DWBPs aplicadas são: 12, 13 e 14.
- **Interligação:** Compreende as ações que permitem que a publicação de dados seja interligada a outros dados, resultando em um contexto ampliado de informações na Web. A DWBP aplicada é a 31.
- **Publicação:** Inclui as atividades ligadas ao armazenamento e à entrega dos dados, permitindo que sejam acessíveis a terceiros. As DWBPs aplicadas são: 17, 18, 19, 20, 21, 23, 24, 25, 29, 30, 32, 34 e 35.
- **Evolução:** Realiza as atividades de manutenção, evolução e atualização do conjunto de dados publicados. As DWBPs aplicadas são: 8, 11, 22, 26 e 27.

Neste trabalho será explorada a fase de Interligação dos dados, explicada a seguir.

Interligação de dados

A interligação de dados surge como um dos alicerces fundamentais da Web Semântica, uma vez que viabiliza o relacionamento de informações provenientes de diversas fontes. Esse processo possibilita a descoberta de informações adicionais e o aproveitamento mais eficiente dos dados com vistas à expansão do contexto dos dados. Essa união de dados é concretizada mediante o uso de associações semânticas, empregando vocabulários e ontologias que estruturam e relacionam as informações.

A fase de Interligação, presente na abordagem LODMBP, tem o objetivo de ligar conjuntos de informações através da detecção de correspondências entre os dados, por meio dos seus significados, atribuídos com o uso de vocabulários e ontologias padronizadas. Durante esta etapa, a DWBP31 é aplicada, a qual promove o enriquecimento dos dados por intermédio da descoberta de conexões em larga escala entre distintos conjuntos de dados. Essa descoberta é desencadeada por ferramentas automatizadas, as quais se valem de técnicas como aprendizado de máquina, desambiguação, categorização, recuperação de informações, entre outras.

MATERIAIS E MÉTODOS

Ferramentas

OPEN REFINE

O OpenRefine² é uma ferramenta utilizada para aprimorar e remodelar dados. A ferramenta realiza tarefas como detecção e correção de erros de digitação, uniformização de valores, eliminação de duplicatas e até a interligação de dados desorganizados por meio de métodos como reconciliação de entidades.

² URL: <<https://openrefine.org/>>

LIMES

LIMES³ (*Link Discovery Framework for Metric Spaces*) é um *framework* para descoberta de correspondências entre conjuntos de dados, que implementa abordagens eficientes para realizar a descoberta de ligações em larga escala, com base nas características dos espaços métricos. LIMES utiliza um arquivo *script* de configurações, no qual são especificados alguns parâmetros, assim como os dois grafos que serão relacionados (um grafo de origem e outro de destino).

SPARQL

A linguagem de consulta SPARQL (*SPARQL Protocol and RDF Query Language*⁴) desempenha um papel fundamental na Web Semântica, sendo projetada especificamente para recuperar informações de formatos de dados semânticos, como o RDF⁵. Funcionando de maneira semelhante às linguagens de consulta tradicionais, como SQL⁶, SPARQL é otimizada para dados semânticos. Ela proporciona recursos robustos para pesquisar e recuperar informações de grafos, permitindo a combinação de padrões, a filtragem de resultados e a realização de ordenações e agregações.

Openlink Virtuoso

OpenLink Virtuoso é uma plataforma para acesso, integração e gerenciamento de dados nos mais diversos formatos e com licença aberta. Possibilita o uso de APIs e trabalha nas principais linguagens de programação e sistemas operacionais de mercado. Com desempenho superior a muitas outras plataformas de gerenciamento de dados existentes, pode dar suporte até 50 bilhões de triplas (Michelon; Berardi, 2023).

3 URL: <<https://github.com/dice-group/LIMES>>

4 URL: <<https://www.w3.org/TR/rdf-sparql-query/>>

5 URL: <<https://www.w3.org/RDF/>>

6 URL: <https://www.w3schools.com/sql/sql_intro.asp>

INTERLIGAÇÃO DO CONJUNTO UNIVERSIDADES COM A DBPEDIA

Para realizar a interligação do conjunto de dados Universidades com a DBpedia foi utilizada a ferramenta LINES versão 1.7.9 para Linux, a qual opera apenas por meio de linhas de comandos, em um arquivo *script* de configuração.

Este *script* contém diversas informações, sobre o contexto de origem e destino, **Source** e **Target** respectivamente, tais como:

- **Prefix:** define uma abreviação para um *namespace*, que é uma coleção de identificadores exclusivos para semantificar recursos da Web (vocabulários padrão). Especifica os campos:
 - **Namespace:** URI que representa o vocabulário que está sendo definido em Prefix.
 - **Label:** campo que define o prefixo do vocabulário apresentado em *Namespace*, o qual será usado ao longo do processo de alinhamento.
- **Source:** define os parâmetros da fonte de origem para comparação dos dados, ou seja, de onde os recursos serão selecionados para a comparação. Especifica os campos:
 - **Id:** é o campo que define um identificador para a fonte.
 - **Endpoint:** é o campo que define o endereço web de acesso à fonte.
 - **Var:** é o campo que define o nome da variável que será utilizada para comparação entre os dados de **Source** e **Target**.
 - **Pagesize:** é o campo que especifica o tamanho máximo de resultados.
 - **Restriction:** campo que define uma restrição SPARQL para selecionar apenas recursos que atendem a esse critério específico. Nesta pesquisa foi utilizada a restrição “**?x rdf:type dbo:City**” especificando que a classe a ser comparada é um recurso do tipo “**dbo:City**”, ou seja, os valores que se encontram no campo “**uriCity**”.
 - **Property:** é o campo que define a propriedade da classe especificada em **Restriction** que será utilizada na comparação. Nesta pesquisa foi utilizada a propriedade “**foaf:name**” a qual refere-se ao nome da cidade, ou seja, os valores que se encontram no campo “**cityName**”.
- **Target:** define os parâmetros da fonte de destino para comparação dos dados, ou seja, de onde os recursos serão selecionados para a comparação com os dados de **Source**. Os campos especificados são os mesmos de **Source** porém referem-se às informações da fonte de destino dos dados. Apenas os campos **Restriction** e **Property** são apresentados para melhor compreensão do conteúdo de ligação:
 - **Restriction:** foi utilizada a restrição “**?y rdf:type dbo:City**” especificando que a classe a ser comparada é um recurso do tipo “**dbo:City**”, ou seja, os valores que se encontram no campo “**City**”, sendo o campo de destino.

- **Property:** foi utilizada a propriedade “**rdfs:label**” a qual refere-se ao nome da cidade na DBpedia.
- **Metric:** é o campo que apresenta a medida de similaridade usada para comparar os recursos nas fontes de dados de origem e destino.
- **Acceptance:** define os parâmetros das configurações de aceitação dos resultados do alinhamento. Especifica os seguintes campos:
 - **Threshold:** É o limiar de aceitação utilizado na regra de similaridade. Nesta pesquisa foi utilizada a métrica (0,9) para que fosse possível encontrar a similaridade dos campos “**cityName**” (referente ao nome da cidade onde a IES pública está localizada) e o **nome da cidade** na DBpedia.
 - **File:** arquivo que receberá as relações responsáveis pelas ligações dos dados.
 - **Relation:** regra da relação de interligação. Nesta pesquisa utilizou-se a propriedade “**rdfs:seeAlso**”, que relaciona um recurso a outro para oferecer informações adicionais.
- **Review:** define os parâmetros de revisão dos resultados, gerando um arquivo que contém os dados de ligação com um limiar de aceitação menor que o limiar de **Acceptance**, os quais devem ser validados por um especialista humano para verificar a ligação encontrada. Os campos especificados são os mesmos de **Acceptance**.
- **Execution:** define as configurações de execução. Especifica os campos:
 - **Rewriter:** define o componente responsável pela reescrita de consultas ou processamento de texto. Pode incluir técnicas de normalização, expansão de abreviações, correção ortográfica, entre outros.
 - **Planner:** define o componente responsável pelo planejamento ou organização do fluxo de trabalho do sistema. Pode determinar a sequência de etapas a serem executadas, definir a ordem das operações e otimizar o processo como um todo.
 - **Relation:** refere-se às relações ou conexões entre entidades ou conceitos. Pode envolver a identificação de relações semânticas, hierárquicas, de associação ou qualquer outro tipo de relação entre os dados.

O Quadro 1 apresenta o resultado da interligação do conjunto de dados Universidades com a DBpedia, informando o *link* da DBpedia que contém informações adicionais sobre a cidade das universidades, neste caso de Guarapuava. A linha do Quadro 1 foi adicionada ao conjunto de dados Universidades, resultando em uma nova versão deste conjunto, a qual foi armazenada no servidor de dados do Grupo LASeD.


```
<http://lod.unicentro.br/SmartGuarapuava/IES-1-2/City_Guarapuava>  
<http://www.w3.org/2000/01/rdf-schema#seeAlso>  
<http://dbpedia.org/resource/Guarapuava> .
```

Quadro 1. Ligação da cidade de Guarapuava com a DBpedia.

A Figura 2 representa a modelagem da interligação dos conjuntos de dados Universidades com a DBpedia. Nota-se que na Figura 2, em azul é representado o conjunto de dados Universidades e em verde o conjunto de dados da DBpedia.

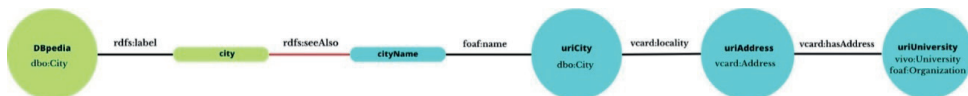


Figura 2. Modelagem da interligação dos conjuntos de dados Universidades com a DBpedia.

Fonte: autores.

Na Figura 2 pode-se observar a interligação dos grafos por meio do recurso “**uriCity**”, o qual possui a propriedade “**foaf:name**” para descrever o objeto “**cityName**”, que contém o nome da cidade. É por meio desta propriedade que é feita a interligação (usando o predicado “**rdfs:seeAlso**”) com a propriedade “**rdfs:label**” do recurso “**dbo:City**” da DBpedia.

RESULTADOS E DISCUSSÕES

Como forma de validar as interligações realizadas nos conjuntos de dados Universidades e DBpedia, foram realizadas consultas no OpenLink Virtuoso, utilizando SPARQL. O Quadro 2 apresenta uma consulta em SPARQL para obter informações sobre os dados das universidades e o *link* da DBpedia referente à cidade de cada universidade, resultado da interligação da DBpedia com as Universidades. No caso desta pesquisa o conjunto utilizado aborda as universidades da cidade de Guarapuava.

```

PREFIX foaf:<http://xmlns.com/foaf/0.1/>
PREFIX vivo: <http://vivoweb.org/ontology/core#>
PREFIX vcard: <http://www.w3.org/2006/vcard/ns#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>

SELECT *
WHERE {?uriUniversidade rdf:type vivo:University ;
      foaf:name ?nomeUNiversidade ;
      foaf:nick ?sigla ;
      dc:identifier ?cnpj .

      ?uriUniversidade vcard:hasAddress ?uriAddress .
      ?uriAddress vcard:locality ?uriCity .
      ?uriCity foaf:name ?cidade .

      ?uriCity rdfs:seeAlso ?linkDB . }

```

Quadro 2. Consulta aos dados das universidades e ao *link* da DBpedia com informações sobre a cidade na qual a universidade está localizada.

Fonte autores.

Primeiramente foi criada uma variável para obter o URI da universidade, sendo ela do tipo **“vivo:University”**, o nome da universidade é obtido na variável **“?nomeUNiversidade”**, descrito pela propriedade **“foaf:name”**. O identificador da universidade é obtido na variável **“?cnpj”**, descrita pela propriedade **“dc:identifier”**, a variável **“?uriAddress”** descrita pelo propriedade **“vcard:hasAddress”** apresenta o URI do endereço da universidade.

O recurso **“?uriAddress”** utiliza o predicado **“vcard:locality”** para relacionar o URI da cidade (informada pela variável **“?uriCity”**), que por sua vez informa o nome da cidade, descrito pela propriedade **“foaf:name”**.

Por fim, com a variável **“?uriCity”** utilizando a propriedade **“rdfs:seeAlso”** obtemos o *link* da DBpedia com informações adicionais da cidade na qual a universidade está localizada, enriquecendo a compreensão acerca do conjunto de dados Universidades.

A Figura 3, complementada pela Figura 4, apresenta os resultados parciais da consulta do Quadro 2.

uriUniversidade	nomeUNiversidade	sigla	cnpj	uriAddress
http://lod.unicentro.br/SmartGuarapuava/IES-1-2/University_Universidade_Estadual_do_Centro_Oeste_do_Parana_Campus_Cedeteg	"Universidade Estadual do Centro Oeste do Parana"	"CEDETEG"	"77.902.914/0001-72"	http://lod.unicentro.br/SmartGuarapuava/IES-1-2/University_Universidade_Estadual_do_Centro_Oeste_do_Parana_Campus_Cedeteg/vcard:hasAddress
http://lod.unicentro.br/SmartGuarapuava/IES-1-2/University_Universidade_Tecnologica_Federal_do_Parana_Campus_Guarapuava	"Universidade Tecnologica Federal do Parana"	"UTFPR"	"75.101.073/0013-23"	http://lod.unicentro.br/SmartGuarapuava/IES-1-2/University_Universidade_Tecnologica_Federal_do_Parana_Campus_Guarapuava/vcard:hasAddress
http://lod.unicentro.br/SmartGuarapuava/IES-1-2/University_Centro_Universitario_Campo_Real	"Centro universitario Campo Real"	"Campo_Real"	"03.291.761/0001-38"	http://lod.unicentro.br/SmartGuarapuava/IES-1-2/University_Centro_Universitario_Campo_Real/vcard:hasAddress
http://lod.unicentro.br/SmartGuarapuava/IES-1-2/University_Faculdade_Guarapuava	"Faculdade Guarapuava"	"Guarapuava"	"09.150.706/0001-04"	http://lod.unicentro.br/SmartGuarapuava/IES-1-2/University_Faculdade_Guarapuava/vcard:hasAddress
http://lod.unicentro.br/SmartGuarapuava/IES-1-2/University_Universidade_Estadual_do_Centro_Oeste_do_Parana_Campus_Santa_Cruz	"Universidade Estadual do Centro Oeste do Parana"	"UNICENTRO"	"77.902.914/0001-72"	http://lod.unicentro.br/SmartGuarapuava/IES-1-2/University_Universidade_Estadual_do_Centro_Oeste_do_Parana_Campus_Santa_Cruz/vcard:hasAddress
http://lod.unicentro.br/SmartGuarapuava/IES-1-2/University_Faculdade_Guairaca	"Faculdade Guairaca"	"Guairaca"	"06.000.722/0001-18"	http://lod.unicentro.br/SmartGuarapuava/IES-1-2/University_Faculdade_Guairaca/vcard:hasAddress

Figura 3. Resultado parcial da consulta Universidades e DBpedia.

Fonte: Sparql *Conductor Query* local dos autores.

uriAddress	uriCity	cidade	linkDB
http://lod.unicentro.br/SmartGuarapuava/IES-1-2/Address_Alameda_Elio_Antonio_Dalla_Vecchia_838	http://lod.unicentro.br/SmartGuarapuava/IES-1-2/City_Guarapuava	"Guarapuava"	http://dbpedia.org/resource/Guarapuava
http://lod.unicentro.br/SmartGuarapuava/IES-1-2/Address_Av_Profa_Laura_Pacheco_Bastos_800	http://lod.unicentro.br/SmartGuarapuava/IES-1-2/City_Guarapuava	"Guarapuava"	http://dbpedia.org/resource/Guarapuava
http://lod.unicentro.br/SmartGuarapuava/IES-1-2/Address_R_Comendador_Norberto_1299	http://lod.unicentro.br/SmartGuarapuava/IES-1-2/City_Guarapuava	"Guarapuava"	http://dbpedia.org/resource/Guarapuava
http://lod.unicentro.br/SmartGuarapuava/IES-1-2/Address_R_Novo_Ateneu_1015	http://lod.unicentro.br/SmartGuarapuava/IES-1-2/City_Guarapuava	"Guarapuava"	http://dbpedia.org/resource/Guarapuava
http://lod.unicentro.br/SmartGuarapuava/IES-1-2/Address_R_Salvatore_Renna_875	http://lod.unicentro.br/SmartGuarapuava/IES-1-2/City_Guarapuava	"Guarapuava"	http://dbpedia.org/resource/Guarapuava
http://lod.unicentro.br/SmartGuarapuava/IES-1-2/Address_R_XV_de_Novembro_7050	http://lod.unicentro.br/SmartGuarapuava/IES-1-2/City_Guarapuava	"Guarapuava"	http://dbpedia.org/resource/Guarapuava

Figura 4. Continuação do resultado da consulta Universidades e DBpedia.

Fonte: Sparql *Conductor Query* local dos autores.

Ao clicar no resultado da consulta do Quadro 2, na variável “linkDB”, obtém-se o direcionamento para a página da cidade de Guarapuava na DBpedia. Na Figura 5 é apresentado o site da DBpedia com as informações adicionais de Guarapuava.

As consultas relativas à interligação foram executadas com sucesso, permitindo a obtenção de novas informações externas aos conjuntos de dados, aumentando o contexto de conhecimentos dos dados.

Ao interligar o conjunto de dados Universidades com a DBpedia, foi possível enriquecer as informações disponíveis sobre as universidades, obtendo dados adicionais provenientes de fontes confiáveis e bem estabelecidas. Por exemplo, consultando o conjunto de dados Universidades foi possível obter informações sobre as universidades públicas de Guarapuava, além de outros dados disponíveis na DBpedia sobre a cidade de Guarapuava, ampliando a compreensão sobre todo o contexto que envolve a universidade e permitindo análises mais abrangentes.

Além disso, pela aplicação da **DWBP-31** - Enriquecer dados por meio da geração de novos dados, ao verificar as diretrizes de Bernes-Lee (2010) o conjunto de dados Universidades-1-2 passou a ser classificado como 5 Estrelas devido a interligação com a DBpedia. Da mesma forma, pela utilização de tecnologias como RDF e URIs na denominação de recursos do conjunto de dados, possibilitou-se a interoperabilidade e a reutilização dos dados, facilitando a integração com outros conjuntos de dados e permitindo que aplicações e agentes inteligentes descubram, entendam, processem e integrem as informações de forma mais eficaz.

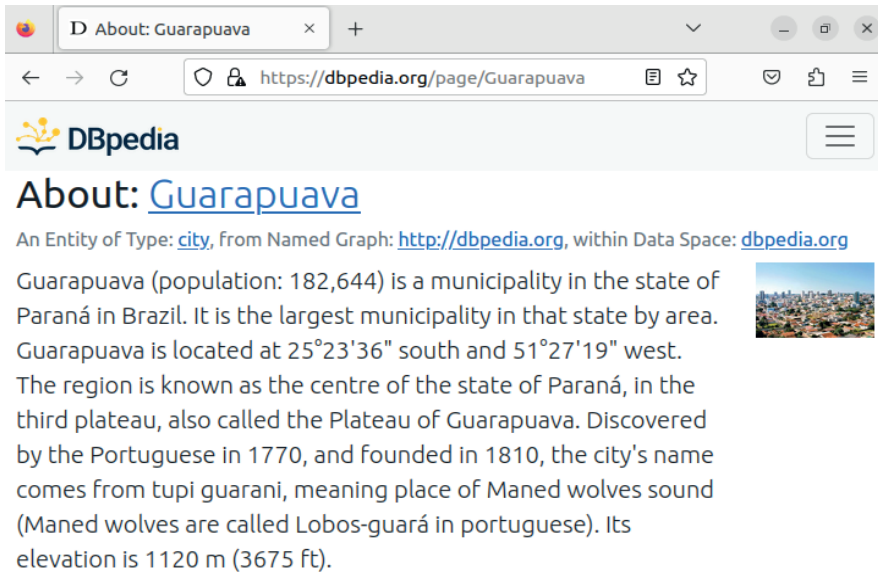


Figura 5. Site da DBpedia com as informações adicionais de Guarapuava.

Fonte: <<http://dbpedia.org/page/Guarapuava>>

CONSIDERAÇÕES FINAIS

A Web Semântica desempenha um papel essencial e colaborativo na evolução contínua da Internet. Sua importância se destaca pela habilidade notável de organizar, de maneira eficaz, a vasta e diversificada gama de dados que são publicados on-line, transformando-os em informações interconectadas e repletas de significado. Esse movimento não somente amplia a acessibilidade dos dados, mas também serve como um catalisador para a geração de conhecimento. No entanto, é importante reconhecer que esse processo não é de simplicidade trivial.

Esta pesquisa realizou a interligação de dados, uma abordagem que visa estabelecer uma rede coesa de conjuntos de Dados Abertos Conectados na internet. Essa estratégia proporciona a estruturação e conexão dos dados de modo que sua exploração e enriquecimento se tornem tarefas acessíveis e produtivas.

A materialização prática do conceito de Dados Abertos Conectados foi efetuada ao correlacionar informações relacionadas a universidades, vinculando os dados universitários à DBpedia. A interconexão de conjuntos de dados seguiu os princípios da fase de Interligação conforme definido na metodologia LODMBP desenvolvida pelo grupo LASeD. Não menos importante, os dados se alinham com o padrão de 5 Estrelas estabelecido por Berners-Lee (2010), uma vez que foram interligados com a DBpedia.

Conclui-se que a interligação de conjuntos de dados emerge como uma abordagem

fundamental da Web Semântica, desempenhando um papel crucial, permitindo que informações sejam exploradas de maneira mais eficiente, colaborativa e significativa, tanto para os seres humanos quanto para as máquinas.

Como trabalhos futuros, pretende-se reestruturar o conjunto de dados Universidades para realizar novas interligações com a DBpedia, assim como também é almejado incluir os conjuntos de dados desenvolvidos nesta pesquisa nas APIs desenvolvidas pelo grupo LAsED, as quais fazem a publicação de dados como maneira de disponibilizar dados mais atualizados, enriquecidos e interligados.

REFERÊNCIAS

BERNERS-LEE, T.; CONNOLLY, D.; STEIN, L.A.; SWICK R. **The Semantic Web**. 2000. Disponível em: <<https://www.w3.org/2000/Talks/0906-xmlweb-tbl/text.htm>>. Acesso 5 out. 2022.

BERNERS-LEE, T. **Linked Data, 2006**. Disponível em: <<https://www.w3.org/DesignIssues/LinkedData.html>>. Acesso 10 set. 2022.

BERNERS-LEE, T. **Esquema de Classificação 5 Estrelas. 2010**. Disponível em: <<https://5stardata.info/en/>>. Acesso 08 set. 2022.

DALL'AGNOL, J.; M. H.; TACLA, C.A. **LOD Methodology based on W3C Best Practices – LODMBP**. Relatório Final de Atividades de Pós-doutorado em Engenharia Elétrica e Informática Industrial, UTFPR, Curitiba, 2020.

HEATH, T.; BIZER, C. **Linked Data: Evolving the Web into a Global Data Space**. 2011. Disponível em: <<http://linkeddatabook.com/editions/1.0/>>. Acesso 19 dez. 2022.

LÓSCIO, B. F.; BURLE, C.; CALEGARI, N. **Data on the Web Best Practices**. W3C Recommendation 31 January 2017. Disponível em: <<https://www.w3.org/TR/dwbp/>>. Acesso 20 mar. 2020.

MICHELON, G. A.; BERARDI, R. C. G. **Estudo Comparativo de Ferramentas para Modelagem, Armazenamento e Integração de Dados na Web Semântica para a plataforma da rede ELLAS**, Interfaces. DOI:10.26439/interfaces2023.n018.6613. December 2023.