

PROCESSAMENTO DE CONSULTAS SPARQL EM BASES DE DADOS ABERTAS

Ana Heloísa Bravin Mazur (PIBIC/CNPq/FA/UEM), Profa. Dra. Heloise Manica Paris Teixeira (Orientadora), Profa. Dra. Raqueline Ritter de Moura Penteado (Coorientadora)

Universidade Estadual de Maringá, Centro de Tecnologia, Maringá, PR.

Área e subárea do conhecimento: Ciência da Computação / Sistemas de Informação

Palavras-chave: ontop; ontology based data access; integração de dados.

RESUMO

A tecnologia OBDA permite a criação de uma camada semântica sobre uma base de dados relacional, que possibilita a realização de consultas com a linguagem SPARQL (Protocol and RDF Query Language) para recuperar e manipular informações de um grafo de conhecimento virtual. Existem inúmeras possibilidades de uso e aplicações da linguagem SPARQL. Estudar suas formas de uso e características é essencial para o desenvolvimento de soluções OBDA, como o problema da integração de dados. Este projeto demonstrou o funcionamento de consultas com a linguagem SPARQL através de um estudo de caso prático com dados abertos. Os resultados obtidos podem servir como base para desenvolvedores compreenderem as possibilidades de uso de SPARQL no contexto de OBDA.

INTRODUÇÃO

A cada dia, há um aumento na variedade e quantidade de dados abertos, o que traz oportunidades para explorar e analisar informações em busca de descobertas úteis para tomada de decisões. A integração semântica de dados utiliza uma representação conceitual de dados e seus relacionamentos para eliminar possíveis heterogeneidades. Essa representação conceitual pode ser feita por meio de ontologias, que podem ser definidas como representações formais e explícitas de uma conceitualização compartilhada (Studer *et al.*, 1998).

A tecnologia OBDA (Ontology-Based Data Access) é uma abordagem de integração que permite criar uma camada semântica expressa na forma de uma ontologia sobre uma base de dados relacional, o que possibilita a realização de consultas sobre um grafo de conhecimento mantido virtual. A linguagem SPARQL é uma linguagem de consulta para o formato RDF, capaz de obter e manipular dados em formato de grafos. O Ontop é um sistema OBDA de código aberto que dá suporte a linguagem SPARQL e

aos principais sistemas de banco de dados relacionais. Entender a linguagem SPARQL, suas características, formas de uso e possibilidades de aplicação é essencial para acessar os dados distribuídos em diferentes fontes de dados abertas.

MATERIAIS E MÉTODOS

O presente trabalho dá continuidade ao estudo de caso desenvolvido por Bispo *et al.* (2022). A ontologia OBaS (Ontology for Bariatric Surgery), proposta por Nunes (2021), foi aplicada no contexto do acompanhamento pós cirurgia bariátrica. O acompanhamento pós-operatório de cirurgia bariátrica é composto por várias etapas, e ocorre durante um período de, no mínimo, dois anos (ABESO, 2016). As ferramentas utilizadas foram o editor de ontologias Protégé, e o sistema Ontop. O Ontop é uma ferramenta OBDA de código aberto que possibilita a integração semântica de fontes de dados relacionais. O sistema expõe dados heterogêneos como um grafo de conhecimento unificado em RDF gerado a partir do mapeamento dos termos da ontologia com as fontes de dados (CALVANESE *et al.*, 2016).

As bases de dados utilizadas têm como fonte o portal DATASUS, que disponibiliza dados relacionais de caráter administrativo do SUS. Foram obtidos dados do SIA (Sistema de Informações Ambulatoriais) e do SIH (Sistemas de Informações Hospitalares), que contém informações sobre laudos de APAC (Autorização de Procedimento de Alta Complexidade) de acompanhamento pós cirurgia bariátrica e registros de AIH (Autorização de Internação Hospitalar), respectivamente, abrangendo os anos de 2020 e 2021 no estado do Paraná.

Os dados foram convertidos do formato *dbc* em que são disponibilizados para uma base de dados relacional gerenciada pelo sistema MariaDB, em tabelas separadas para os registros de APAC e AIH. Mapeamentos entre a ontologia e a base de dados foram especificados na linguagem nativa do Ontop, utilizando uma abordagem virtual para modelar entidades individuais presentes em bases de dados diferentes, que consiste em gerar o mesmo URI (Universal Resource Identifier) para entidades equivalentes. O sistema Ontop foi responsável por traduzir as consultas SPARQL feitas sobre a ontologia para consultas SQL sobre o banco de dados, permitindo o acesso aos dados sem o custo da materialização.

RESULTADOS E DISCUSSÃO

Foram desenvolvidos no total 22 mapeamentos, que incluem diferentes subclasses de cirurgias, comorbidades, hábitos e relações. Os mapeamentos, consultas e demais artefatos desenvolvidos estão disponíveis online em <<https://github.com/anabravn/obas-sus>>. De modo a ilustrar as possibilidades da aplicação da abordagem OBDA, são apresentadas duas das consultas SPARQL desenvolvidas utilizando o vocabulário da ontologia OBaS, e seus resultados. As consultas apresentadas fazem uso de agrupamento e funções de agregação, funcionalidades importantes para a tomada de decisão.

```
SELECT (COUNT(?pessoa) AS ?n) ?p WHERE {
  ?pessoa obas:0000174 ?consulta.
  ?consulta obas:1000037 ?p.
} GROUP BY ?p
```

(a)

```
SELECT ?cirurgia (AVG(?valor) AS ?v) WHERE {
  ?pessoa obas:1000014 ?cir.
  ?cir rdf:type ?type.
  ?type rdfs:subClassOf obas:0000140; rdfs:label ?cirurgia
  FILTER(!sameTerm(?type,obas:0000140))
  OPTIONAL {
    ?pessoa obas:1000001 ?internacao.
    ?internacao obas:1000030 ?valor.
  }
} GROUP BY ?cirurgia
```

(b)

Figura 1 – Consultas SPARQL

A **consulta 1** recupera o número de pacientes que apresentaram ganho de peso após a realização da cirurgia bariátrica. A figura 1 (a) apresenta o código em SPARQL. Essa consulta é composta por um padrão de grafo simples, que relaciona uma entidade “pessoa” com uma entidade “consulta” e seleciona o valor da propriedade “reganho de peso”, uma propriedade de dados do tipo booleano. A relação de ganho de peso é associada com uma entidade “consulta”, já que um paciente pode reportar ganho de peso em qualquer momento do tratamento. A figura 2 (a) apresenta os resultados obtidos com a execução da consulta. Dos 33242 pacientes, 33083 não apresentaram ganho de peso, e apenas 159 apresentaram. Dessa forma, é possível perceber que a maioria das cirurgias teve sucesso como tratamento para a obesidade.

A **consulta 2** calcula a média do valor da internação por tipo de cirurgia bariátrica realizada. A consulta seleciona o tipo de cirurgia realizada por um determinado paciente, e obtém o valor da internação, agrupando os dados por tipo de cirurgia e calculando a média do valor de internação. Um padrão de grafos opcional é utilizado pois alguns pacientes não possuem uma internação correspondente. A figura 1 (b) apresenta o código em SPARQL, e a figura 2 (b) apresenta os resultados da consulta. A média do valor da internação é de R\$6398,65 para cirurgias do tipo “derivação gastrointestinal” e R\$5867,42 para cirurgias do tipo “banda gástrica”.

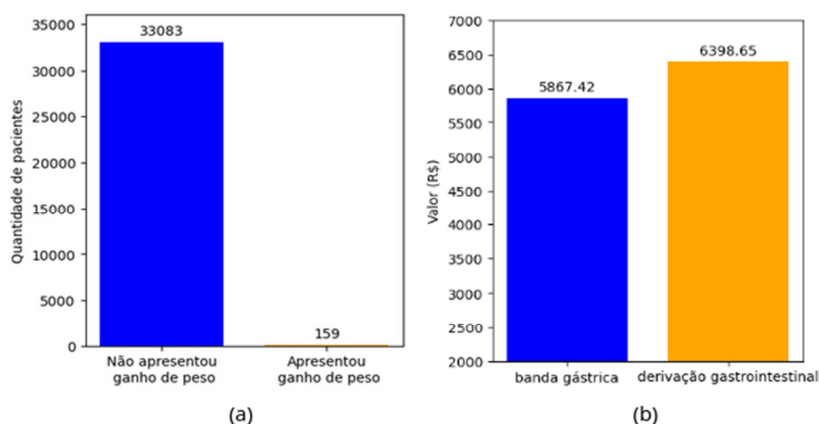


Figura 2 – Resultados das consultas

CONCLUSÕES

Este trabalho mostra uma aplicação do paradigma OBDA utilizando o sistema Ontop no domínio de cirurgia bariátrica. A integração semântica de dados com base em ontologias é uma área promissora, com diversas aplicações. Em contrapartida, o processo de formulação de consultas ainda é relativamente complexo para usuários finais, o que limita os usos desse tipo de sistema. O estudo de caso desenvolvido apresenta um exemplo completo do projeto e execução da integração semântica de dados de duas bases de dados relacionais. Os dados de AIH permitem a consulta de valores de cirurgias e períodos de internação, e juntamente aos dados de APAC permitem o acesso a informações mais específicas sobre o estado dos pacientes após a cirurgia. Em conclusão, esse projeto explorou as possibilidades de uso da SPARQL e sua integração com outras linguagens e ferramentas.

AGRADECIMENTOS

Ao Programa Institucional de Bolsas de Iniciação Científica (PIBIC/CNPq-FA-UEM) pela bolsa concedida à primeira autora.

REFERÊNCIAS

ABESO: Associação Brasileira para o Estudo da Obesidade e da Síndrome Metabólica. **Diretrizes brasileiras de obesidade**. 4. ed. São Paulo: ABESO, 2016.

BISPO, S. L. B.; FUKACE, V. K.; MAZUR, A. H. B.; PENTEADO, R. R. M.; TEIXEIRA, H. M. P.. Integração de Dados Abertos em Saúde com o modelo OBDA: Um Estudo de Caso na Área de Cirurgia Bariátrica. *In: Anais do Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*. [s.l.]: SBC, 2022, p. 401–412.

CALVANESE, D.; COGREL, B.; KOMLA-EBRI, S.; et al. Ontop: Answering SPARQL queries over relational databases. *Semantic Web*, v. 8, n. 3, p. 471–487, 2016.

NUNES, G. M. O.. **Multionto: método de construção de ontologia considerando heterogeneidade de fontes e tipos de conhecimentos – um estudo de caso sobre cirurgia bariátrica**. Dissertação (Mestrado em Computação Aplicada), Universidade Tecnológica Federal do Paraná, 2021.

STUDER, R.; BENJAMINS, V. R.; FENSEL, D.. Knowledge engineering: Principles and methods. **Data & Knowledge Engineering**, v. 25, n. 1, p. 161–197, 1998.