

UM ESTUDO SOBRE A FERRAMENTA *GRAPHDB* PARA INTEGRAÇÃO DE DADOS NA ÁREA DE SAÚDE

Douglas K. Sakakibara (PIC), Felipe G. C. Scheffel (PIC), Flávio R. Uber (Orientador), Raqueline R. M. Penteadó (Coorientadora)
ra117306@uem.br, ra117741@uem.br, fruber@uem.br, rrmptenteadó@uem.br.

Universidade Estadual de Maringá / Departamento de Informática, PR.

Ciência da computação / Banco de dados

Palavras-chave: GraphDB, integração de dados, ontologias.

Resumo:

Este trabalho apresenta um estudo inicial sobre a ferramenta *GraphDB* no contexto de integração de dados na área de saúde. O estudo envolveu dados reais referentes a cirurgias bariátricas a fim de analisar o desempenho no processamento de consultas e a visualização de resultados no *GraphDB*. Futuramente, o uso da ferramenta pode auxiliar gestores no processo de tomada de decisões.

Introdução

À medida em que o escopo de uma solução que envolva a análise de dados cresce, é natural esperar um aumento proporcional não apenas em volume, mas também na heterogeneidade das fontes. Assim, conforme a indústria atual continua a adotar o paradigma de *Big Data*, uma parcela relevante da atenção de pesquisadores da área é capturada, ainda na fase de preparação, pelo problema da integração de dados.

Em especial na área de saúde, bases de dados de origem e objetivos diferentes podem alimentar consultas e gerar informações que podem auxiliar o gerenciamento e a tomada de decisões tanto no que se refere a empresas particulares quanto na saúde pública de modo geral. Uma base de dados mantida pelo Sistema Único de Saúde (SUS) a respeito da vacinação contra a COVID-19, por exemplo, poderia ser integrada com uma base de dados que armazena dados sobre os atendimentos com cardiologistas. A integração destas bases permitiria deduzir correlações entre vacina e cardiopatias, e ajudar a combater notícias falsas sobre efeitos colaterais.

Muitos conceitos e ferramentas têm papel relevante nas operações que lidam com a integração de dados. A ferramenta *Protégé*, por exemplo, utiliza o *plugin* da ferramenta *Ontop* e os padrões estabelecidos pelo *World Wide Web Consortium* (W3C), para oferecer integração semântica de dados por meio de ontologias, mapeamentos e consultas. Não obstante, diante da necessidade de armazenar grandes quantidades de informação semântica, surgiram os bancos de dados de grafos, entre os quais destacam a Neo4j, opção de código aberto mais popular

atualmente, e o *GraphDB*, ferramenta alvo deste estudo (GÜTING, 1994). Empresas como a emissora britânica BBC, a publicadora holandesa Elsevier e a construtora estatal norueguesa StatNett utilizam o *GraphDB*. A ferramenta também viabiliza a integração de dados por meio do *Ontop*.

Materiais e Métodos

O *GraphDB* é um banco de dados de grafos desenvolvido pelo *Ontotext* com a proposta de facilitar a consulta de conhecimento semântico e agilizar a integração de novas fontes de informação. Este trabalho utilizou a versão 9.9 da edição gratuita da ferramenta, *GraphDB Free*.

Foram realizados experimentos no contexto de cirurgia bariátrica. A ontologia utilizada foi desenvolvida por (NUNES, 2021), a qual é uma extensão da *Ontology for Bariatric Surgery* (OBaS), que tem como base a *Ontology for General Medical Science* (OGMS) e esta tem como base a *Basic Formal Ontology* (BFO). O estudo considerou duas bases de dados com informações sobre procedimentos realizados através do Sistema Único de Saúde Brasileiro (SUS) e foram armazenadas no banco de dados H2.

Analizou-se o desempenho de duas consultas que integram dados das duas bases. A primeira consulta relaciona custo da cirurgia com comorbidades dos pacientes. A segunda relaciona custo da cirurgia com hábitos saudáveis dos pacientes. Para atender as consultas, foram definidos 10 mapeamentos das bases de dados para a ontologia usada no estudo. As consultas foram executadas em dois ambientes. O primeiro integrou o *GraphDB* com o *Ontop*. O segundo utilizou o *Ontop* no *Protégé*.

Resultados e Discussão

A Tabela 1 apresenta os tempos de execução das duas consultas nos dois ambientes. A Figura 1 mostra um *screenshot* do *GraphDB* na execução da segunda consulta. Na tabela, para cada consulta tem-se o tempo da primeira execução e o tempo médio de 10 execuções. Na análise de desempenho foi percebido que as duas consultas tiveram um desempenho similar na sua primeira execução; porém, no segundo ambiente, *Ontop* no *Protégé*, o tempo médio de execução das consultas foi reduzido em até 10 vezes.

As funcionalidades de visualização de dados oferecidas pelo *GraphDB* também foram analisadas nesse estudo inicial. Como o repositório de dados é virtual, ou seja, os dados não foram propriamente inseridos no repositório, as funcionalidades que dependem da inferência semântica ficam desabilitadas, incluindo a visão geral dos grafos e a visualização dos relacionamentos entre as classes da ontologia. Portanto, os grafos visuais foram o foco dessa parte do experimento.

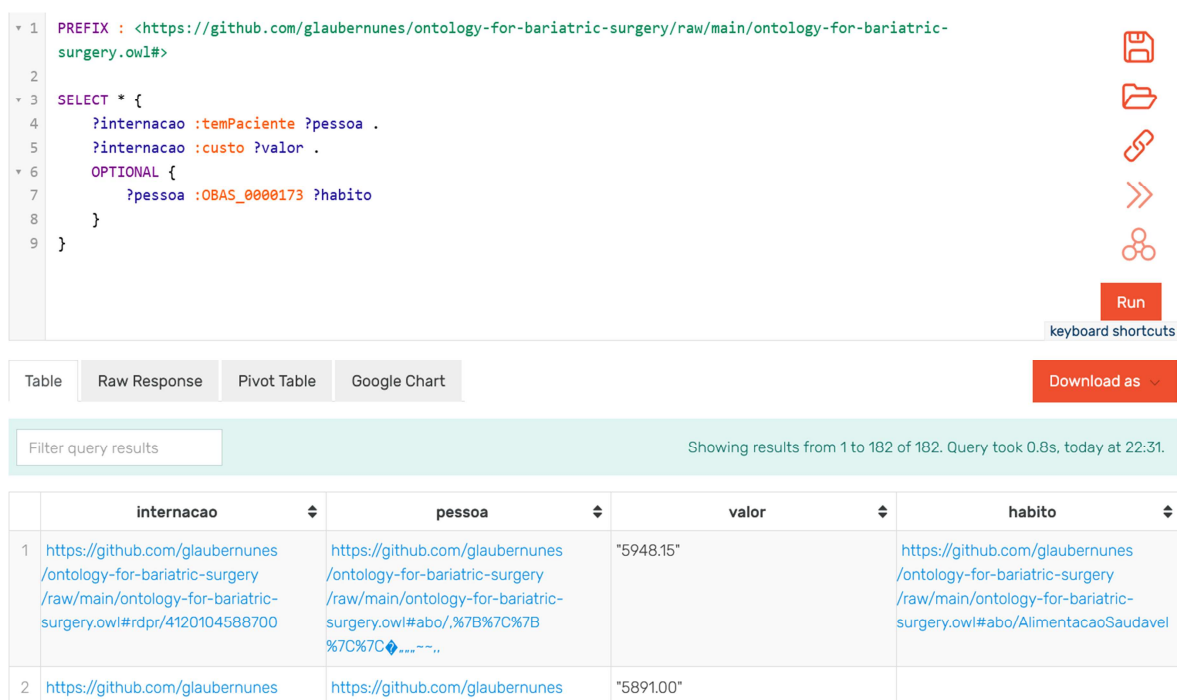


Figura 1 – Execução da segunda consulta no *GraphDB*.

Tabela 1 - Execução da segunda consulta no *GraphDB*.

	Amb. 01	Amb. 01 (10 vezes)	Amb. 02	Amb. 02 (10 vezes)
Consulta 01	1.2s	1.2s	1.05s	93ms
Consulta 02	0.8s	0.8s	0.6s	84ms

A parte inferior da Figura 1 exibe o resultado da segunda consulta de maneira tradicional, ou seja, tabular. Entretanto, o ambiente do *GraphDB* permite a visualização do resultado por meio de nodos e arestas, conforme mostra a Figura 2. A funcionalidade tem grande potencial no quesito de legibilidade humana; porém, para tanto, é necessário uma ontologia abrangente, especialmente se tratando de rótulos, ranqueamento e propriedades de objetos.

Conclusões

O estudo sobre a ferramenta *GraphDB* ainda está em suas fases iniciais. O estudo mostrou que vários pontos podem ser explorados no quesito desempenho de consultas e visualização de dados, a fim de explorar todo o potencial da ferramenta.

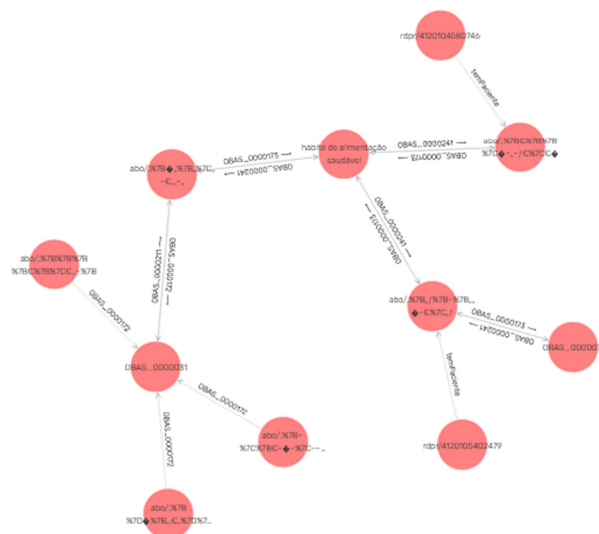


Figura 2 – Exemplo de renderização da consulta para grafos visuais.

Para a análise de desempenho, por exemplo, a execução de outras consultas auxiliaria na detecção de padrões nas execuções, dando subsídios para uma análise detalhada dos resultados. A diversidade de consultas pode ser obtida a partir de um mapeamento mais completo das bases para a ontologia.

A visualização de dados também pode ser explorada a fim de verificar como a interface gráfica do *GraphDB* pode auxiliar gestores na tomada de decisão. Por fim, o estudo também pode levantar os contextos mais vantajosos para a visualização de dados por meio de grafos.

Referências

Güting, R. H. (1994). Graphdb: Modeling and querying graphs in databases. In VLDB, volume 94, pages 12–15. Citeseer.

Nunes, G. M. O. (2021). Multionto: Método de construção de ontologia considerando heterogeneidade de fontes e tipos de conhecimentos - um estudo de caso sobre cirurgia bariátrica. Master's thesis, Universidade Tecnológica Federal do Paraná.