

UNIVERSIDADE DE BRASÍLIA
Faculdade do Gama

Sistemas de Banco de Dados 2

Tecnologias de Banco de Dados (TI-BD)

Bases de Conhecimento (KD - Knowledge Database)

Thiago Siqueira Gomes - 190055294

Brasília, DF
2023

Introdução

Com a crescente demanda por ferramentas computacionais para organizar os mais variados sistemas de informações, se fez necessário métodos de armazenamento que dão um passo adiante em relação ao tradicional banco de dados relacional. Mediante a necessidade de ir além do armazenamento dos dados e suas relações, por meio de soluções que possibilitem atribuir significado semântico para os dados armazenados, a base de conhecimento, se apresenta como um poderoso meio para organização de informações.

Uma base de conhecimento, utiliza uma nova abordagem para lidar com informações, sendo uma poderosa ferramenta para organizar uma grande variedade de informações de todos os tipos, e atribuir significados variados para cada uma delas, que se inter relacionam entre si de forma dinâmica. Desse modo, uma Base de Conhecimento é uma importante ferramenta para que empresas, órgãos governamentais e outras instituições organizem seus dados de forma útil para a construção de aplicações diversas e ferramentas para a tomada de decisões.

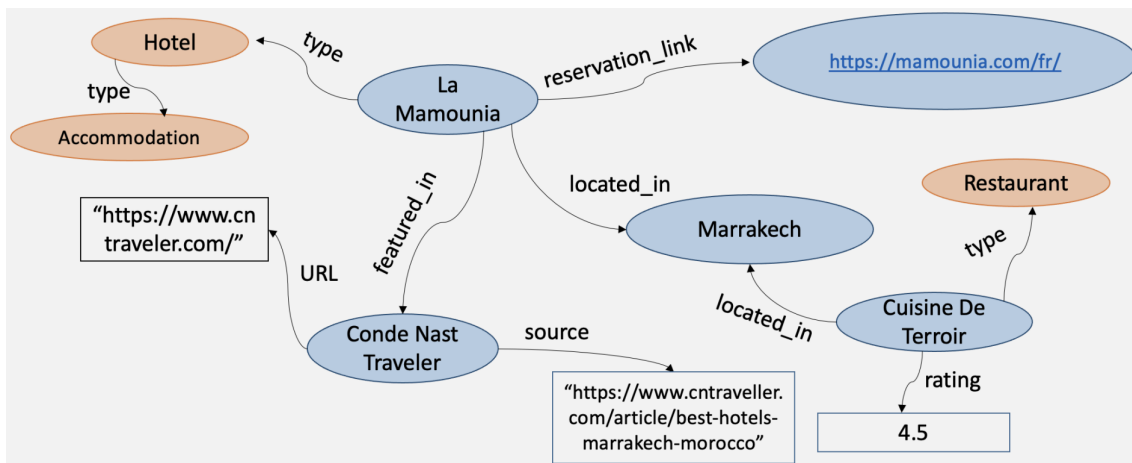
O presente trabalho tem como objetivo definir o que é uma base de conhecimento, abordar o conceito de ontologia, apresentar as principais vantagens e desvantagens associadas a uma base de conhecimento, mostrar as principais ferramentas e falar sobre casos de sucesso que fazem o uso desse método de organizar informações.

Definição Base de Conhecimento

O termo base de conhecimento deriva do inglês, *knowledge database*. Uma base de conhecimento consiste em uma forma de organizar o conhecimento de determinada área, de modo que os significados das informações armazenadas seja preservado e estejam inter relacionados entre si. Uma base de conhecimento vai muito além do simples armazenamento de dados e informações de maneira estruturada, ela abrange diretamente os aspectos semânticos dos seus dados, de modo que é possível extrair através de consultas a essa base, informações difusas sobre determinado assunto

pesquisado, obtendo correlações que podem ir muito além do que se esperava obter inicialmente. A figura 1 mostra de forma visual um exemplo didático da representação de uma base de conhecimento através de uma estrutura de grafos, com associação entre múltiplas informações de diferentes tipos.

Figura 1: Grafo de Conhecimento



Fonte: KEJRIWAL (2022)

As aplicações de bases de conhecimento são as mais diversas, elas podem ser utilizadas como fonte de informações para o treinamento de inteligências artificiais, armazenar informações de preferências do usuário para sistemas de recomendações, como base de dados para assistentes virtuais, em sistemas de suporte a tomada de decisões, e várias outras aplicações em que a semântica das informações seja priorizada.

A estrutura de uma base de conhecimento pode seguir diferentes padrões, mas existem algumas estruturas comuns à maioria delas, as principais são: armazenamento de dados, ontologia, sistema de inferência e interface.

O armazenamento de dados compreende todo o conteúdo conhecido pela base de conhecimento, podendo conter informações e arquivos dos mais variados tipos, como imagens, vídeos, documentos, textos, áudios e todo tipo de dados relevantes para a aplicação.

O termo ontologia vem da filosofia, representando um ramo da metafísica que

estuda a natureza da existência e do ser em si, através da categorização, agrupamento, e da hierarquização de diferentes objetos. Este conceito foi incorporado à ciência da computação, para atribuir sentido e significado a objetos (ANTONIOU; HARMELEN, 2008). Em uma base de conhecimento, a Ontologia é responsável por estabelecer a estrutura semântica, estabelecendo conceitos, classes, hierarquias, organizando e dando significado aos dados que serão armazenados. A ontologia pode ser vista como um conjunto de termos e definições para o conhecimento armazenado. Algumas ferramentas utilizadas para criação de ontologias são: Protégé, TopBraid Composer, PoolParty e OWL API. Existem várias linguagens para a representação de Ontologias, a OWL (Web Ontology Language), a mais utilizada.

O sistema de inferência é a camada de uma base de conhecimento responsável por fazer deduções lógicas através das definições estabelecidas pela ontologia, utilizando lógica proposicional, descritiva, ou outros modelos lógicos, para chegar a conclusões, ou descoberta de novas informações através do que se tem armazenado previamente.

As interfaces de uma base de conhecimento podem ser as mais variadas, dependendo da necessidade da aplicação, ela pode ser uma API para integração a um sistema principal, pode ser uma janela de buscas, pode utilizar uma linguagem própria criada para interagir com a aplicação, entre outras possíveis implementações tanto utilizadas diretamente pelo usuário quanto por outras aplicações.

Vantagens Base de Conhecimento

A utilização de uma base de conhecimento possibilita inúmeras vantagens em relação a formas convencionais de armazenamento de dados, como um banco de dados relacional. Algumas dessas vantagens são: a integração de dados, a reutilização de conhecimento e a flexibilidade.

Através de uma base de conhecimento, é possível agrupar informações de diferentes fontes e formatos que tenham alguma relação semântica entre si,

possibilitando uma conexão e integração de informações muito superior ao que se tem com um banco de dados relacional, em que existe um modelo fixo com relação bem definidas. Através dessa integração de dados existente em uma base de conhecimento, é possível preencher lacunas de dados utilizando inferência, sem comprometer o valor das informações, algo impossível com um modelo de banco de dados relacional.

A reutilização de conhecimento é outro ponto forte de uma base de conhecimento, pois através dela é possível conectar diversas informações de áreas diferentes, aproveitando conceitos já existentes de áreas que não possuem relação direta entre si. Tal versatilidade na reutilização de conhecimento é possível através da ontologia, que define conceitos e classes que organizam o conhecimento de modo em que as relações não precisam ser explicitamente definidas como em um banco de dados relacional.

Em um banco de dados relacional, as tabelas e relacionamentos são pensados e atribuídos com antecedência, modificações envolvendo estruturas do banco já existentes podem implicar em refatorações e reorganização da base de dados. Em uma base de conhecimento não existe uma estrutura rígida pré definida, o que possibilita uma flexibilidade muito maior em termos de evolução da base e consulta das informações. É possível facilmente estabelecer conexões entre dados antigos e novos de forma simples, e fazer consultas complexas envolvendo dados de áreas diferentes e tipos diferentes sem grandes esforços.

Desvantagens Base de Conhecimento

Embora uma base de conhecimento apresenta uma série de vantagens em relação a um banco relacional, também existem desvantagens associadas a esse modelo. Entre elas, pode-se destacar a menor precisão, a complexidade, e os custos.

Por ter uma uma estrutura fixa previamente definida para cada dado e relacionamento estabelecido, um banco de dados relacional apresenta uma

grande precisão para realizar as consultas que ele foi modelado para realizar, trazendo os dados da forma que eles são esperados ao realizar a busca. Devido ao próprio conceito que envolve uma base de conhecimento, ela não é a ferramenta ideal para aplicações que necessitem de uma resposta precisa para uma consulta que é feita de forma recorrente em um padrão fixo, pois podem ocorrer inconsistências que seriam evitadas ao utilizar um modelo de banco de dados relacional, que estabelece uma estrutura rígida de armazenamento de dados e relacionamentos.

Outra desvantagem de uma base de conhecimento, é em relação à sua complexidade de modelagem e gerenciamento. Um banco de dados relacional apresenta uma maior facilidade de uso, sendo útil na resolução de uma grande variedade de problemas, enquanto a complexidade envolvida na estrutura de uma base de conhecimento, torna inviável o uso dessa ferramenta em grande parte dos casos.

Os custos também são uma desvantagem associada a uma base de conhecimento, pois em relação aos bancos de dados relacionais, elas possuem maiores custos de hardware e software para o seu funcionamento, além da evolução e manutenção mais caras e demoradas, devido a complexidade desse tipo de aplicação, como descrito anteriormente. Isso torna o uso de uma base de conhecimento financeiramente inviável em muitas aplicações.

Ferramentas e Exemplos de Uso

Algumas das principais ferramentas para implementação de uma base de conhecimento, são o Protégé, TopBraid Composer, PoolParty e o OWL API.

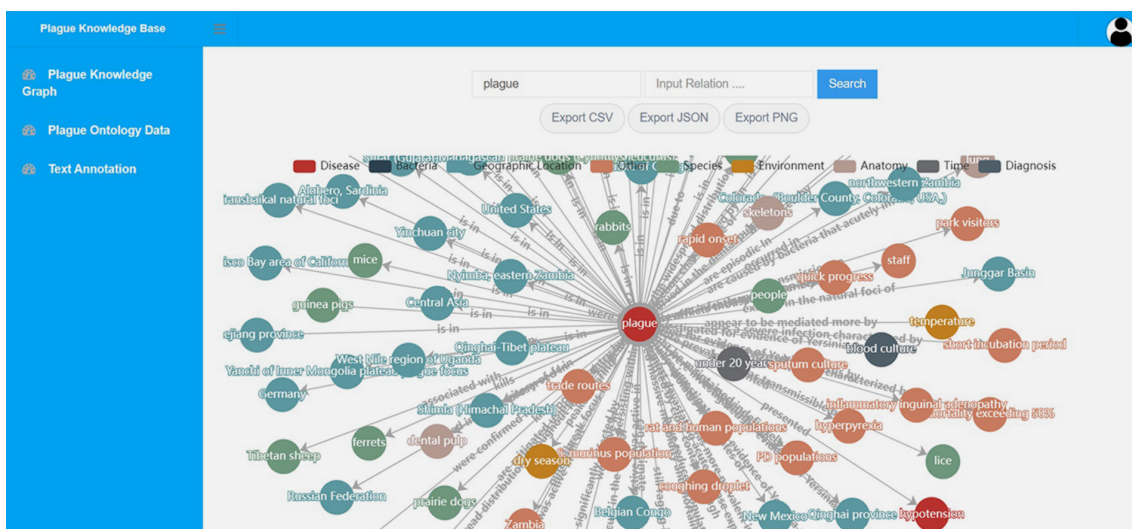
O Protégé e o OWL API são softwares gratuitos de código abertos feitos para modelagem e gerenciamento de ontologias, permitindo o desenvolvimento de ontologias com uma grande variedade de formatos. O TopBraid Composer, PoolParty são ferramentas comerciais também com a finalidade de modelar e gerenciar ontologias.

As bases de conhecimento estão presentes nas mais diversas aplicações, como por exemplo no sistema de recomendação de músicas do Spotify, nos

anúncios e reconhecimento de interesse de postagens utilizados pelo Facebook, no Google que agrupa palavras chaves, informações de sites e imagens no seu motor de busca. Também existem aplicações em que a base de conhecimento representa a parte central de seu funcionamento, como é o caso da aplicação IBM Watson, que utiliza inteligência artificial associada a uma base de conhecimento para armazenar informações das mais diferentes áreas, aprender, e realizar inferências de acordo com o que é armazenado.

Um exemplo bem sucedido de uso de uma base de conhecimento, é a criação da base de conhecimento PlagueKD, que tem o objetivo de mapear as mais diversas informações sobre a Peste Bubônica, doença que assolou a humanidade em diferentes reuniões do globo em diferentes épocas, causada pela bactéria *Yersinia pestis*. O PlagueKD é uma base de conhecimento criada por pesquisadores vinculados a uma universidade chinesa e uma universidade australiana, que através de um mapeamento envolvendo uma ampla literatura sobre Peste Bubônica com 5388 fontes, produziu uma extensa base de conhecimento. Esta base de conhecimento é utilizada por pesquisadores do mundo inteiro, e tem sido muito importante para compreensão da Peste Bubônica, da bactéria *Yersinia*, e para criação de profilaxias. Na figura 2 pode se ter uma visão da interface do PlagueKD (LI et al., 2022).

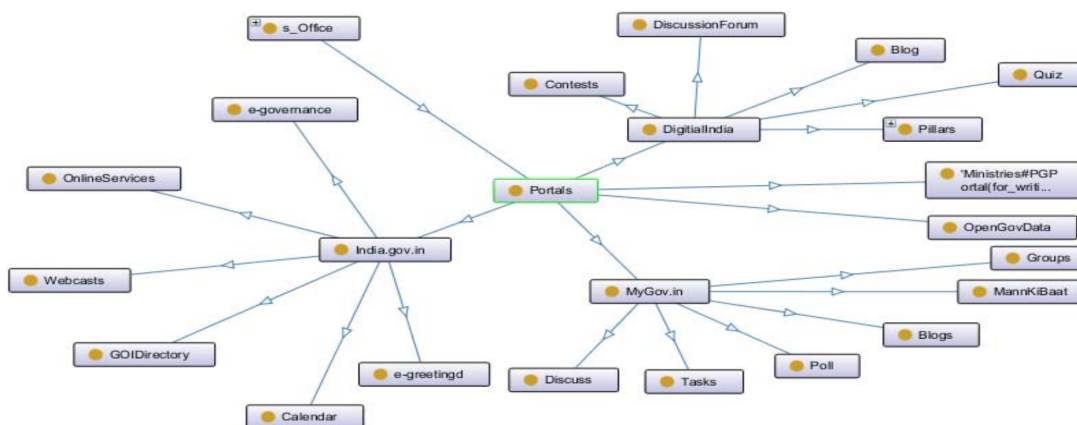
Figura 2: Interface do PlagueKD



Fonte: LI et al. (2022)

Um exemplo de uso bem sucedido de uma base de conhecimento, é a aplicação desenvolvida pelo governo indiano para organizar o conhecimento disponível no portal governamental do país. Utilizando a ferramenta Protégé, foi desenvolvida a ontologia chamada IndiGov-O, que organiza informações sobre ministérios, departamentos, serviços e demais informações governamentais, e disponibiliza o acesso para a população. A figura 3 demonstra a hierarquia da ontologia relacionada aos portais sob o domínio do Primeiro Ministro. O aspecto semântico de uma base de conhecimento permite que o governo descreva seus serviços de forma eficiente, estabelecendo conexões entre diferentes domínios de acordo com o significado de cada um. O sucesso dessa iniciativa adotada pelo governo indiano é um avanço em direção a uma “governança digital” (KUMAR; JOSHI, 2016).

Figura 3: Diferentes portais sob o domínio do Primeiro Ministro



Fonte: KUMAR e JOSHI (2016)

Conclusão

De acordo com o exposto, nota-se que a base de conhecimento é uma ferramenta fundamental para o armazenamento de dados, e possibilita a implementação de soluções computacionais para aplicações que dependem do estabelecimento de relações semânticas entre os dados armazenados.

Apresentando vantagens como maior flexibilidade, integração dos dados, e reutilização de informações previamente armazenadas, e também desvantagens como a sua complexidade, maiores custos e não possuir a

mesma precisão de um banco de dados relacional ao trabalhar com a manipulação de dados que seguem uma estrutura fixa. Desse modo, é essencial que seja feita uma avaliação que leve em conta todos esses pontos positivos e negativos ao implementar uma base de conhecimento como solução.

Aplicações que demandam uma interpretação semântica dos dados armazenados estão cada dia mais presentes no mundo moderno, e a utilização de bases de conhecimento tende a se difundir cada vez mais para resolução desses problemas. Sendo importante solução tanto para empresas privadas, por meio do desenvolvimento de ferramentas como o IBM Watson, tanto soluções estatais que visam o gerenciamento de dados públicos como IndiGov-O implementado pelo governo indiano.

REFERÊNCIAS

KUMAR, Akshi; JOSHI, Arunima. IndiGov-O: An ontology of Indian Government to empower Digital Governance. 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom). **IEEE**, 2016. p. 1233-1238. ISBN: 978-1-4673-6984-8.

ANTONIOU, Grigoris; HARMELEN, Frank Van. Ontologies. **Semantic Web Primer**. MIT Press, 2008. Disponível em: <http://ebookcentral.proquest.com/lib/univbrasilia-ebooks/detail.action?docID=3338781>. Acesso em: 17 abr. 2023.

KEJRIWAL, Mayank. Graphs: A Practical Review of the Research Landscape. **Information**, v. 13, n. 4, p. 161, 2022. Disponível em: <https://doi.org/10.3390/info13040161>. Acesso em: 17 abr. 2023.

Li, J., Gao, J., Feng, B. et al. PlagueKD: a knowledge graph-based plague knowledge database. **Database**, v. 2022, article ID baac100, 2022. DOI: <https://doi.org/10.1093/database/baac100>.