Criação e exploração de ontologia

Ezequiel Moreira 1 PG 38413

Universidade do Minho, Portugal https://www.uminho.pt/PT

Abstract. Este documento foca-se na definição do domínio e no desenho da ontologia para o projeto final da disciplina de Processamento e Representação de Conhecimento(PRC) de 2018/2019 no primeiro ano de Mestrado de Engenharia Informática(MEI).

Keywords: ontologia elemento químico molécula

1 Introdução

Como parte da disciplina de PRC foi-nos requerido a criação de uma ontologia e sua exploração como parte do projeto final da disciplina, sendo esta exploração feita através de uma aplicação Web.

Ao longo deste relatório será explicado o domínio escolhido, os dados que foram extraídos e finalmente o desenho da ontologia, referindo em seguida e muito rapidamente como a exploração foi feita.

1.1 Objetivos

Os objetivos deste projeto são a definição de uma ontologia sobre o domínio que escolhemos e criação de ferramentas para a explorar, de modo a permitir obter dados sob forma legível e de modo eficiente.

1.2 Nota

Todo o trabalho realizado para este projeto pode ser consultado em [projeto git]

2 Domínio de ontologia e extração de dados

O domínio escolhido para a ontologia que defini foi elementos químicos e moléculas por eles formados.

Este domínio foi escolhido por considerar ser um domínio interresante de se explorar, por conter bastante informação útil para certos sectores de atividade humana e por ser um domínio relativamente imutável em termos de elementos e ter uma estrutura bem definida para os seus objetos.

2 Ezequiel MoreiraPG38413

Originalmente tentei obter os dados através de queries efetuadas a Wikidata e Dbpedia, mas rapidamente descobri que muita da informação disponibilizada nas páginas Wikipedia de elementos e moléculas não estava presente ou estava anémica(por exemplo, os grupos de elementos estavam extremamente incompletos, não tendo associados os elementos para vários deles).

Como tal, decidi obter os dados através das tabelas nas páginas da Wikipedia correspondentes ao domínio, nomeadamente em [elementos] e [moléculas], mas também nas páginas dos grupos de elementos([grupos]), tendo os dados sido extraídos com o auxilio do serviço prestado em [extração de tabelas].

3 Desenho da ontologia

Com os dados obtidos podemos agora pensar no desenho da ontologia.

Este desenho deve ser de tal modo que seja o mais claro possível em termos do seu desenho para tornar as *queries* mais simples e deve conseguir ter o máximo de informação possível dos dados extraidos anteriormente.

Note-se que todas as relações que serão referenciadas nas subsecções seguintes são mútuas, i.e., toda a relação na ontologia que foi definida tem uma relação inversa.

3.1 Elementos

Dos dados dos elementos e dos grupos podemos imediatamente criar 3 classes de objetos na nossa ontologia: *Element* que contém os múltiplos campos de dados associados a cada elemento na tabela extraída, *Element_Group* que corresponde a um grupo de elementos, tendo como dados o nome e estando relacionado com os vários elementos que o constituem um grupo na tabela periódica e *Element_Period*, que é o mesmo que o grupo mas para um período da tabela periódica.

Em seguida, criei uma nova classe para servir de topo a estas 3: *Periodic_table*, que contém relações com todas as outras classes.

Com isto obtemos a seguinte estrutura para os elementos químicos na nossa ontologia:

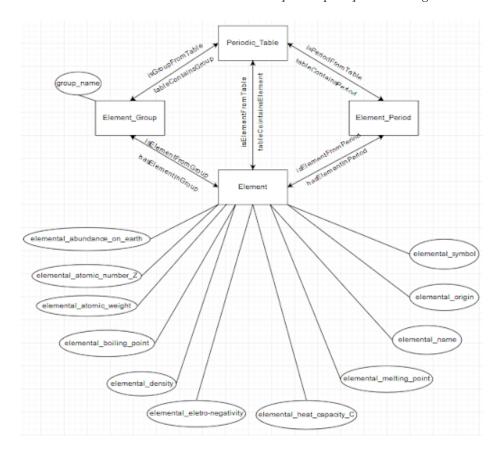


Fig. 1. Desenho de elementos na ontologia final

3.2 Moléculas

Com os elementos definidos podemos agora definir a parte da ontologia correspondente às moléculas.

Para tal deve-se primeiro observar como uma molécula é estruturada, verificando várias características:

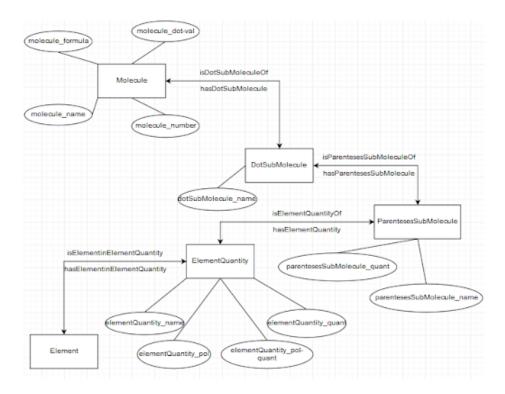
- uma molécula contém uma quantidade inteira positiva de um ou mais elementos anteriormente definidos - ex: H2O
- cada um destes elementos pode ter uma carga positiva ou negativa bem como um número associado a esta carga(correspondendo ao número extra de eletrões que o átomo do elemento tem no caso de ter carga positiva ou em falta para a carga negativa) - ex: D3O+
- uma molécula pode ter submoleculas dentro dela delimitadas por parênteses e com um valor inteiro, representando que esta molécula contém essa quantidade da submolecula dentro de parênteses - ex: CH2(CN)2

4 Ezequiel MoreiraPG38413

– finalmente, uma molécula pode conter um carácter "·" seguido de um número e uma molécula (infelizmente não consegui determinar o significado preciso desta notação) - ex: $Hg(BrO3)2 \cdot 2H2O$

A partir desta lista de características, comecei a definir várias classes para lidar com elas: ElementQuantity, que está relacionada com um elemento e contém campos de dados relacionados com a quantidade do elemento e a carga associada à molécula, ParentesesSubMolecule, que corresponde à divisão por parênteses da molécula e contém dados da formula da submolecula e quantidade associada(estando relacionado com as várias ElementQuantity dos elementos que a constituem), DotSubMolecule que corresponde à divisão pelo carácter "·" da molécula e está associado aos vários elementos obtidos da divisão por parênteses e finalmente temos a molécula completa, que está associado às divisões por "·" e contém dados da sua formula, o seu nome, o seu número CAS e finalmente o valor após o "·" caso tenha essa divisão.

Com tudo isto obtemos a seguinte estrutura na ontologia para a parte das moléculas:



 ${\bf Fig.\,2.}$ Desenho de elementos na ontologia final

Temos assim desenhada a nossa ontologia.

3.3 Processamento de dados obtidos

Com os dados extraídos e ontologia definida falta fazer o processamento e limpeza de dados e inserção na ontologia.

Este processo é complicado e não faz parte do objetivo deste documento, mas consistiu em transformar os dados obtidos das tabelas para um formato que fosse facilmente convertível em xml, sendo que este formato é depois transformado por uso de xsl para a forma ttl que pode ser colocada dentro do motor GraphDB diretamente.

Todas as transformações mencionadas acima podem estão disponibilizadas dentro de [projeto git] na pasta data.

4 Uma referência rápida à exploração de ontologia

A exploração da ontologia foi feita através de 2 servidores (backend e frontend), contactando um(backend) com o serviço GraphDB para executar as queries que permitem obter os dados da ontologia e passando-os para o segundo(frontend) que é responsável por mostrar estes dados de modo simples e legível a um utilizador que queira explorar a ontologia criada.

Ambos estes servidores bem como as instruções para colocar a funcionar o serviço estão disponíveis em [projeto git].

5 Conclusão

Com a ontologia criada e a exploração feita, dou por concluído com sucesso todos os objetivos propostos ao longo deste documento.

Bibliography

```
[projeto git] https://github.com/Ezequiel-Moreira/PRC_projeto_final [elementos] https://en.wikipedia.org/wiki/List_of_chemical_elements [moléculas] https://en.wikipedia.org/wiki/Glossary_of_chemical_formulas [grupos] https://en.wikipedia.org/wiki/Group_(periodic_table) [extração de tabelas] https://www.wikitable2json.com/
```