Universidade de Aveiro
Departamento de Eletrónica, Telecomunicações e Informática
MEI

Triplestore Interface - Integração de standards

Gil Mesquita 72700
Jorge Loureiro 72982
17/05/2018
Trabalho realizado no âmbito da disciplina
Web Semântica

Índice

Resumo **3**

Introdução **4**

Descrição do projeto **5**

Criação dos triplos 5

Ontologia 6

Funcionalidades 6

Lista e filtragem de tuplos 6

Adicionar/Remover 7

Visualização em grafo 7

Inferências 8

Publicação de dados 9

Análise crítica e conclusão **10**

Referências **11**

1 Resumo

O propósito deste projeto prende-se com a adaptação do projeto anteriormente desenvolvido de forma a que os dados e toda a lógica aplicacional seguissem os padrões standards atualmente reconhecidos no âmbito da web semântica, de maneira a que os triplos pudessem ser alocados numa base de dados orientada a grafos - graphDB. Outro dos objetivos passou também pela conceção de uma ontologia sobre os dados existentes e publicação dos mesmos para consumo por parte de aplicações externas.

Neste documento pode encontrar-se não só uma descrição das alterações efetuadas face ao primeiro trabalho - adaptação para triplos RDF, querys SPARQL, ontologia e publicação de dados -, mas uma descrição integral da solução obtida . Por fim, é também feita uma análise crítica sobre o trabalho e tecnologias utilizadas.

2 Introdução

Para o desenvolvimento deste projeto foi necessária a escolha de uma fonte de dados com conteúdo potencialmente interessante que permitisse não só a adaptação para o modelo de triplos, mas também tirar o máximo proveito dos dados para inferir novas relações e desta forma gerar mais informação. Nesse sentido foi recolhido um dataset da plataforma *kaggle* no formato CSV que continha informações sociais acerca de estudantes do ensino secundário numa instituição de ensino dos EUA. O dataset continha informação relativa a quinhentos sujeitos e informação sobre: género, idade, estado civil, estado da relação parental, frequência de saídas noturnas, frequência de consumo de álcool, entre outros.

3 Descrição do projeto

Nesta secção será descrito o processo de desenvolvimento do projeto, passando pela fase de adaptação dos dados para o modelo de triplos RDF - Ntriples, criação da web interface e descrição das funcionalidades implementadas.

3.1 Criação dos triplos

Numa fase de tratamento de dados, foi efetuado trabalho no sentido de transformar os triplos para o formato standard RDF - Ntriples. Para o efeito foi utilizada a biblioteca csv do python.

```
('11', 'sex', 'F')
('12', 'sex', 'F')
('13', 'sex', 'M')
('14', 'sex', 'M')
('15', 'sex', 'M')
('16', 'sex', 'F')
('17', 'sex', 'F')
('18', 'sex', 'F')
('19', 'sex', 'M')
('1', 'age', '18')
('2', 'age', '17')
('3', 'age', '15')
('4', 'age', '15')
('5', 'age', '16')
('6', 'age', '16')
('7', 'age', '16')
```

Figura 1 - Triplos originais

```
import csv

Entity = "http://www.student-mat.com/entity/"
Property = "http://www.student-mat.com/pred/"

triples = []

filename = "clean_data/" + input("Filename: ")
file_in = open(filename, 'r', encoding='utf-8')

reader = csv.reader(file_in)
for sub, pred, obj in reader:
    triples.append((sub, pred, obj))
file_in.close()

file_out = open('filent.nt', 'w')
for sub, pred, obj in triples:
    uri_sub = '<' + Entity + str(sub).lower().replace(' ', '_') + '>'
    uri_pred = '<' + Property + str(pred).lower().replace(' ', '_') + '>'
    obj_uri = '"' + obj + '"'
    file_out.write('{} {} {} .\n'.format(uri_sub, uri_pred, obj_uri))
file_out.close()
print("filent.nt created")
```

Figura 2 - Código de adaptação

Na **Figura 2** encontra-se o código com o qual foi possível transformar os dados originais (**Figura 1**) no formato de triplos RDF. Este código consiste resumidamente na criação de prefixos para as entidades e propriedades e adaptação para o formato Ntriples. Depois de todos os triplos criados, os mesmos são adicionados manualmente a uma base de dados orientada a grafos - graphDB. O resultado é parcialmente apresentado em seguida na **Figura 3**.

```
<http://www.student-mat.com/entity/1> <http://www.student-mat.com/pred/sex> "F" .
<http://www.student-mat.com/entity/2> <http://www.student-mat.com/pred/sex> "F" .
<http://www.student-mat.com/entity/3> <http://www.student-mat.com/pred/sex> "F" .
<http://www.student-mat.com/entity/4> <http://www.student-mat.com/pred/sex> "F" .
<http://www.student-mat.com/entity/5> <http://www.student-mat.com/pred/sex> "F" .
<http://www.student-mat.com/entity/6> <http://www.student-mat.com/pred/sex> "M" .
<http://www.student-mat.com/entity/7> <http://www.student-mat.com/pred/sex> "M" .
<http://www.student-mat.com/entity/8> <http://www.student-mat.com/pred/sex> "F" .
<http://www.student-mat.com/entity/9> <http://www.student-mat.com/pred/sex> "M" .
<http://www.student-mat.com/entity/10> <http://www.student-mat.com/pred/sex> "M" .
<http://www.student-mat.com/entity/11> <http://www.student-mat.com/pred/sex> "F" .
<http://www.student-mat.com/entity/12> <http://www.student-mat.com/pred/sex> "F" .
<http://www.student-mat.com/entity/13> <http://www.student-mat.com/pred/sex> "M" .
<http://www.student-mat.com/entity/14> <http://www.student-mat.com/pred/sex> "M" .
<http://www.student-mat.com/entity/15> <http://www.student-mat.com/pred/sex> "M" .
<http://www.student-mat.com/entity/16> <http://www.student-mat.com/pred/sex> "F" .
<http://www.student-mat.com/entity/17> <http://www.student-mat.com/pred/sex> "F" .
<http://www.student-mat.com/entity/18> <http://www.student-mat.com/pred/sex> "F" .
<http://www.student-mat.com/entity/19> <http://www.student-mat.com/pred/sex> "M" .
<http://www.student-mat.com/entity/1> <http://www.student-mat.com/pred/age> "18" .
<http://www.student-mat.com/entity/2> <http://www.student-mat.com/pred/age> "17" .
<http://www.student-mat.com/entity/3> <http://www.student-mat.com/pred/age> "15" .
```

Figura 3 - Dados no formato RDF - Ntriples

3.2 Ontologia

Sobre os triplos RDF foi criada uma ontologia que consiste na definição do domínio e range de cada predicado. Desta forma é acrescentado mais conhecimento sobre os dados, ou seja, são conhecidos os tipos de dados (RDF:Type) de cada sujeito/objeto que tem uma dada relação. Em seguida é parcialmente apresentada a respetiva ontologia.

```
<rdf:RDF xmlns:owl ="http://www.w3.org/2002/07/owl#"
         xmlns:rdf ="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
         xmlns:xsd ="http://www.w3.org/2001/XMLSchema#"
         xmlns:foaf ="http://xmlns.com/foaf/0.1/">

    <owl:DatatypeProperty rdf:about="http://www.student-mat.com/pred/sex">
        <rdfs:domain rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
     <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
    </owl:DatatypeProperty>

    <owl:DatatypeProperty rdf:about="http://www.student-mat.com/pred/age">
         <rdfs:domain rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
     <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
    </owl:DatatypeProperty>

    <owl:DatatypeProperty rdf:about="http://www.student-mat.com/pred/pstatus">
         <rdfs:domain rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
     <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
    </owl:DatatypeProperty>

    <owl:DatatypeProperty rdf:about="http://www.student-mat.com/pred/studytime">
         <rdfs:domain rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
     <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
    </owl:DatatypeProperty>

    <owl:DatatypeProperty rdf:about="http://www.student-mat.com/pred/famsup">
         <rdfs:domain rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
     <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
    </owl:DatatypeProperty>


    <owl:DatatypeProperty rdf:about="http://www.student-mat.com/pred/activities">
         <rdfs:domain rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
     <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
    </owl:DatatypeProperty>

    <owl:DatatypeProperty rdf:about="http://www.student-mat.com/pred/romantic">
         <rdfs:domain rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
     <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
    </owl:DatatypeProperty>
```

Fig 4 - Ontologia OWL

3.3 Funcionalidades

Nesta secção serão descritas todas as funcionalidades implementadas no sistema, sendo que, por opção do grupo, todas são parte integrante da interface web desenvolvida em python (django). De notar que todas as funcionalidades foram adaptadas para utilizar queries SPARQL, uma vez que passou a usar-se a base de dados orientada a grafos e os respetivos standards RDF para representação dos triplos.

3.3.1 Lista e filtragem de tuplos

Através desta funcionalidade é dada a possibilidade ao utilizador de fazer uma pesquisa sobre os tuplos pelo padrão que deseja e obter uma vista sobre os mesmos. Este método é bastante flexível e permite efetuar pesquisas complexas, dado que utiliza diretamente a sintaxe do método *query* da triplestore, no entanto cientes de que obriga a que utilizador tenha conhecimento da sintaxe da linguagem de pesquisa.

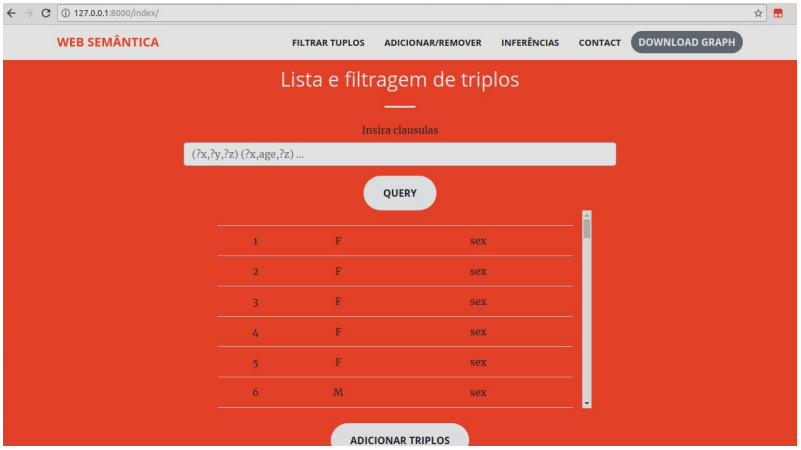

Figura 4 - WebAPP:Lista e filtragem de tuplos

3.3.2 Adicionar/Remover

Esta funcionalidade permite que o utilizador adicione um triplo específico, através da introdução dos capos: sujeito, predicado e objeto. Também permite que seja eliminado um triplo específico com base nos mesmos campos ou eliminar através de um padrão (ex: (None, sex, None) caso em que elimina todos os triplos com o predicado "sex").

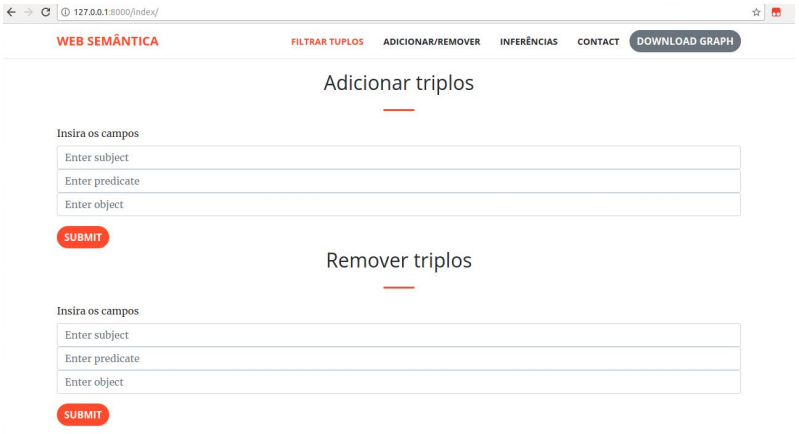

Figura 5 - WebAPP: Adicionar/remover triplos

3.3.3 Visualização em grafo

É dada a possibilidade ao utilizador de obter uma visualização em grafo dos triplos, para além disso, sempre que é aplicada uma filtragem sobre os mesmos, a representação é dinamicamente alterada para construir o grafo de acordo com aquilo que o utilizador quer ver. O procedimento utilizado é, ao clicar no botão "*download graph*" o servidor responde com um pdf que o utilizador pode escolher apenas visualizar ou guardar.

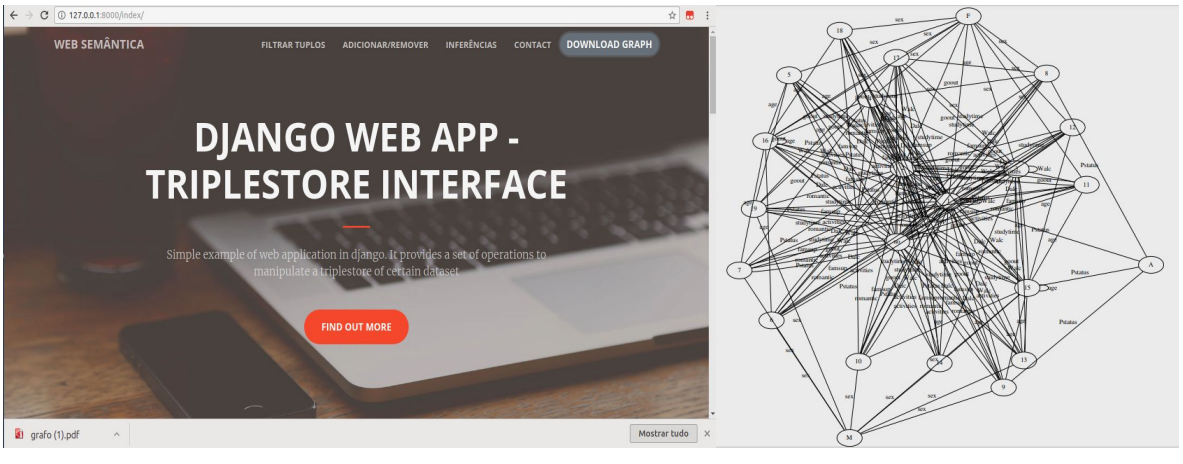

Figura 6 - Download da visualização em grafo

3.3.4 Inferências

Ao utilizador é também dada a possibilidade de inferir novos triplos a partir dos já existentes.
Para esse fim são disponibilizadas quatro inferências pré-estabelecidas que serão descritas seguidamente.

Figura 7 - Interface para as inferências

Inferência Risco
Esta inferência consiste na seguinte dedução: se um dado sujeito (estudante) tem idade inferior a dezoito anos e não tem suporte familiar, é considerado um estudante de risco. Neste sentido é acrescentado um triplo por cada estudante que satisfaça estes requisitos sob a forma: (1,risk,yes), neste exemplo é inferido que o sujeito com id "1" é um estudante de risco.

Inferência Infeliz
Esta inferência é baseada na seguinte regra: se um dado sujeito não tem namorada, sai muito à noite (>2 numa escala de 0-4), e bebe muito álcool (>1 numa escala de 0-4) é considerado um estudante infeliz. Mais uma vez, por cada estudante que satisfaça esta regra é criado um novo triplo, por exemplo: (1,state,unhappy) em que o sujeito com o id '1' para o predicado 'state' tem o valor 'unhappy'.

Inferência Tinder
Uma vez que as inferências anteriores limitam-se a acrescentar um novo predicado e um valor literal por cada sujeito que satisfaça a regra de inferência, decidiu-se criar uma regra que estabelecesse uma relação entre sujeitos (objeto é um apontador para outro sujeito).
Nesse sentido foi considerada a seguinte regra: Dado um sujeito "X" do sexo feminino que não tem namorada e dado outro sujeito "Y" do sexo masculino que também não tem namorada, é inferido que o indivíduo "X" está disponível para o individuo "Y". Exemplo: (1,available,2).

Inferência Orientação
Uma vez que foi criada uma ontologia sobre os dados existentes e, por forma a demonstrar a utilidade da mesma, foi criada uma inferência que tirasse partido dessa mesma ontologia. Desta forma, a inferência consiste no seguinte: Qualquer entidade que seja do tipo

FOAF:Person (informação retirada da ontologia) e tenha a relação “available” - descrita na inferência anterior - é considerada de orientação heterosexual, sendo acrescentado o triplo (entidade, hetero, yes) , no formato Ntriples.

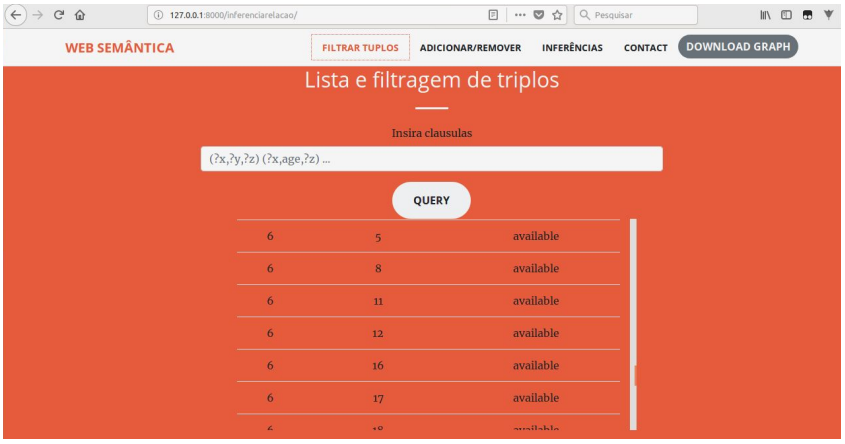

Figura 8 - Novos triplos acrescentados (inferência tinder)

3.4 Publicação de dados

De forma a que fontes externas possam ter acesso à semântica dos dados que são apresentados na interface, os mesmos foram categorizados através de RDFa. No entanto, esta implementação é atualmente apenas para fins demonstrativos, uma vez que não existe qualquer parser para interpretar o modelo dos dados.

```
</thead>
    <tbody>
        {% for sub, pred, obj, obj_type in tuples %}
            <tr>
                <td about="http://www.student-mat.com/entity/{{ sub }}">{{ sub }}</td>
                {% ifequal obj_type 'entity' %}
                        <td rel="http://www.student-mat.com/pred/{{ pred }}">{{ pred }}</td>
                        <td about="http://www.student-mat.com/entity/{{ obj }}">{{ obj }}</td>
                {% else %}
                        <td property="http://www.student-mat.com/pred/{{ pred }}">{{ pred }}</td>
                        <td>{{ obj }}</td>
                {% endifequal %}
            </tr>
        {% endfor %}
        </tbody>
```
Fig 5 - Labels RDFa

4 Análise crítica e conclusão

Sendo a web semântica uma área em crescimento e de grande relevância na atualidade com o surgimento de grandes quantidades de dados, o grupo considera que foi um projeto interessante, na medida em que: permitiu perceber de que forma é que os dados podem estar estruturados na web, garantindo que a informação possa ser interpretada pelas máquinas e desta forma, ao contrário do que acontece na atualidade, estas possam ajudar no acesso e processamento da informação; permitiu perceber de que forma pode ser gerado nova informação com base em dados já existentes; permitiu explorar a framework Django, com a qual o grupo nunca tinha tido contacto no desenvolvimento de aplicações web.

Finalizado este projeto, o grupo considera ainda que foram atingidos todos os objetivos propostos, tendo-se atingido um resultado que foi de encontro com as expectativas.

5 Referências

[1] Dias, T. D., & Santos, N. (2013). Web semântica: conceitos básicos e tecnologias associadas. *Cadernos do IME-Série Informática*, *14*, 80-92.

[2] Bootstrap Themes Built & Curated by the Bootstrap Team. (n.d.). Retrieved from https://themes.getbootstrap.com/

[3] Documentation. (n.d.). Retrieved from https://docs.djangoproject.com/en/2.0/

[4] The Home of Data Science & Machine Learning. (n.d.). Retrieved from https://www.kaggle.com/

[5] Python Data Analysis Library. (n.d.). Retrieved from https://pandas.pydata.org/