

Universidade de Aveiro
Departamento de Eletrónica, Telecomunicações e Informática
MEI

Triplestore Interface

Gil Mesquita 72700

Jorge Loureiro 72982

14/04/2018

Trabalho realizado no âmbito da disciplina

Web Semântica

Índice

Resumo	3
Introdução	4
Descrição do projeto	5
Criação dos triplos	5
Funcionalidades	6
Lista e filtragem de tuplos	6
Adicionar/Remover	6
Visualização em grafo	7
Inferências	7
Análise crítica e conclusão	9
Referências	10

1 Resumo

O propósito deste projeto prende-se com a concepção e adaptação de um modelo de dados que preservasse as suas relações semânticas (triplestore) e o desenvolvimento de uma aplicação web que oferecesse aos seus utilizadores um conjunto de funcionalidades de gestão e manipulação sobre esse mesmo modelo. O objetivo principal foi explorar as tecnologias envolvidas e desta forma consolidar os conceitos teóricos discutidos nas aulas. Para o efeito foi utilizada a framework *django* e uma implementação em python, adaptada, da triplestore fornecida nas aulas.

Neste documento pode encontrar-se uma descrição do dataset original e de que forma este foi adaptado para integrar a triplestore, bem como a descrição das funcionalidades implementadas na interface web. 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.

Acerca da implementação original da triplestore, esta foi alterada a pedido do docente da disciplina, tendo sido utilizadas em vez de listas de dicionários, listas de tuplos.

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, 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 a informação original (CSV) no formato de triplos (sujeito, predicado, objeto). Para este fim foi utilizada a biblioteca *pandas* do *python*.

sex	age	Pstatus	studtime	famsup	activities	romantic	goout	Dalc	Walc
F	18	A		2	no	no	no	4	1
F	17	T		2	yes	no	no	3	1
F	15	T		2	no	no	no	2	2
F	15	T		3	yes	yes	yes	2	1
F	16	T		2	yes	no	no	2	1
M	16	T		2	yes	yes	no	2	1
M	16	T		2	no	no	no	4	1
F	17	A		2	yes	no	no	4	1
M	15	A		2	yes	no	no	2	1
M	15	T		2	yes	yes	no	1	1
F	15	T		2	yes	no	no	3	1
F	15	T		3	yes	yes	no	2	1
M	15	T		1	yes	yes	no	3	1
M	15	T		2	yes	no	no	3	1
M	15	A		3	yes	no	yes	2	1
F	16	T		1	yes	no	no	4	1
F	16	T		3	yes	yes	no	3	1
F	16	T		2	yes	yes	no	2	1
M	17	T		1	yes	yes	no	5	2

Figura 1 - Dataset original

```
# -*- coding: utf-8 -*-
from __future__ import unicode_literals
from grafo import Grafo #Triplestore
from staticvars.staticpaths import pathtocsv #String path to CSV file
import pandas as pd

relations = []
triples = []
pathtocsv = pathtocsv

def maketriples(filepath):
    counter = 1
    csvfile = pd.read_csv(filepath)
    csvfile.drop_duplicates(inplace=True, keep='first')
    for row in csvfile:
        relations.append(row)

    for relation in relations:
        for value in csvfile[relation]:
            triples.append((str(counter), str(relation), str(value)))
            counter += 1
    counter = 1
    return triples

graph = Grafo()
triples=maketriples(pathtocsv)

for triple in triples:
    graph.add(str(triple[0]),str(triple[1]),str(triple[2]))
```

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 e adicioná-los à triplestore. Este código consiste resumidamente na criação de um id incremental por cada linha do CSV (id do sujeito) e na criação de um triplo por cada relação de cada sujeito. Depois de todos os triplos criados, são adicionados um a um à triplestore. O resultado é parcialmente apresentado em seguida na **Figura 3**.

```
('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 3 - Dados no formato de triplos

3.2 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).

3.2.1 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.

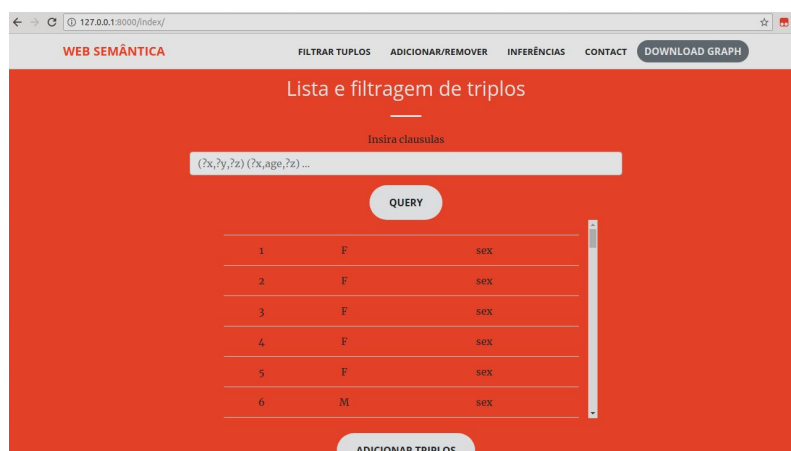


Figura 4 - WebAPP:Lista e filtragem de tuplos

3.2.2 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”).

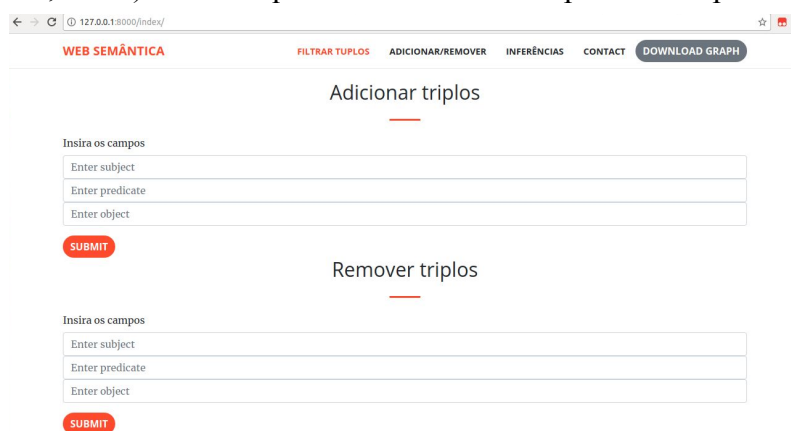


Figura 5 - WebAPP: Adicionar/remover triplos

3.2.3 Visualização em grafo

É dada a possibilidade ao utilizador de obter uma visualização em grafo dos tuplos presentes na triplestore, 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.

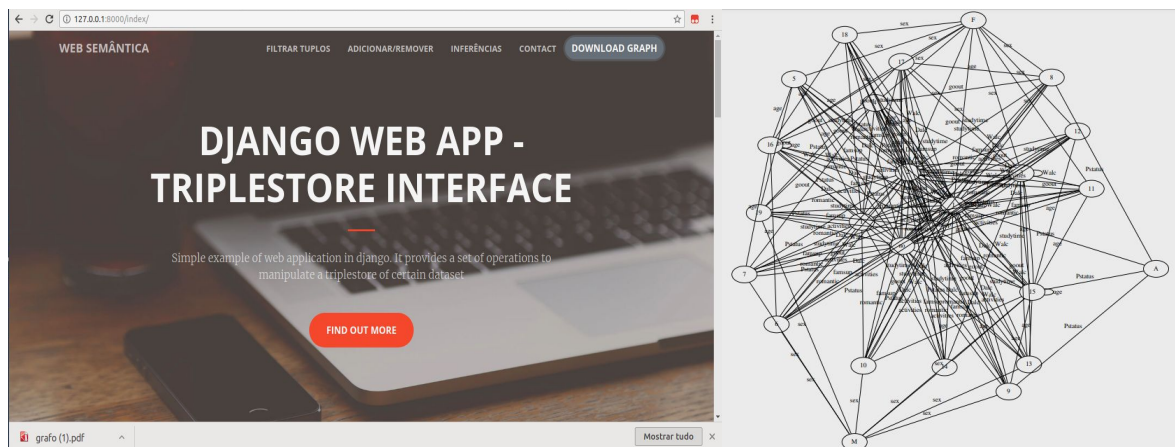


Figura 6 - Download da visualização em grafo

3.2.4 Inferências

Ao utilizador é também dada a possibilidade de inferir novos triplos a partir dos já existentes.

Para esse fim são disponibilizadas três inferências pré-estabelecidas que serão descritas seguidamente.



Figura 7 - Interface para as inferências

Inferência Risco

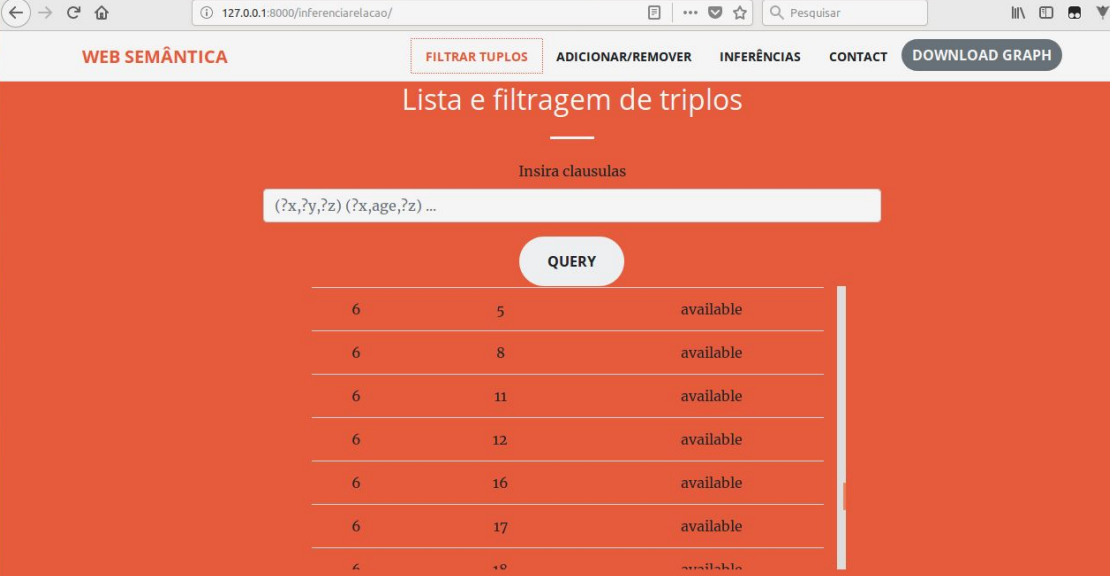
Esta inferência consiste na seguinte dedução: se um dado sujeito (estudante) tem idade inferior a dezoito anos e não tem suporte familiar, é considerado um estudante de risco. Neste sentido é acrescentado um triplo por cada estudante que satisfaça estes requisitos sob a forma: (1,risk,yes), neste exemplo 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 indivíduo "Y". Exemplo: (1,available,2).



Lista e filtragem de triplos		
Insira clausulas		
<input type="text" value="(?x,?y,?z) (?x,age,?z) ..."/>		
QUERY		
6	5	available
6	8	available
6	11	available
6	12	available
6	16	available
6	17	available
6	20	available

Figura 8 - Novos triplos acrescentados (inferência tinder)

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/>