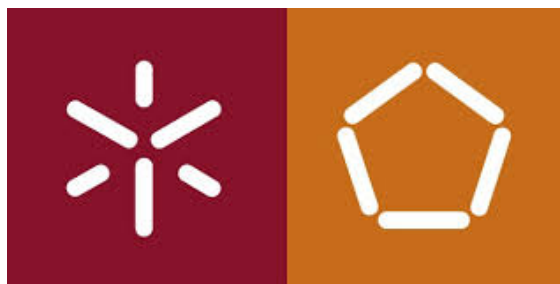


# Universidade do Minho



## *Ligas Europeias*

Mestrado Integrado em Engenharia Informática

### **Processamento e Representação de Conhecimento**

2º Semestre , 2018/2019

A78508 - Fábio Araújo

Gualtar, Braga

14 de Junho de 2019

# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>2</b>
<b>2</b>	<b>Processamento de Conhecimento</b>	<b>3</b>
2.1	Base de Dados . . . . .	3
2.2	API . . . . .	4
<b>3</b>	<b>Representação do Conhecimento - Interface</b>	<b>4</b>
<b>4</b>	<b>Conclusão</b>	<b>5</b>

# 1 Introdução

Este documento trata o desenvolvimento de um projeto no âmbito da cadeira de *Processamento de Linguagens*. O principal objetivo deste projeto passa por implementar o aprendizado, seguindo a dinâmica que o próprio nome da unidade curricular incute, sempre numa primeira fase com o processamento de dados, e numa fase final com a representação de uma forma elegante e em tempo útil da informação mantida.

O processamento de informação passará sempre pela manutenção dos dados em base ontológica, que é o núcleo da unidade curricular, quanto à representação dos dados, distinguem-se duas fases, uma que será a construção da *API* que representará as *querys* em *SPARQL* que responderão às exigências do sistema, por fim, uma outra fase dedicada à amostragem dos dados.

A temática do projeto ficou ao encargo de cada aluno.

## 2 Processamento de Conhecimento

### 2.1 Base de Dados

Como dito anteriormente, o tema do projeto ficou ao encargo de cada um, pelo que, como a tal a primeira abordagem caracterizou-se pela procura da informação tendo o tema desejado em mente. Antes demais, importante indicar o tema desejado, que neste caso se trata da exposição de informação relativa a 5 campeonatos da região europeia. Tendo isso em mente decorreu-se o processo de pesquisa, que veio a tornar-se insuficiente, dado que, os dados recolhidos, de uma forma isolada não continham a informação desejada, mas quando agregada a outra coletânea de dados, tornava-se difícil a leitura destes em conjunto, e, por conseguinte a sua normalização como um só.

Tendo isto em conta, optou-se por outra estratégia, que foi ao encontro da construção em si dos próprios dados, passando pelo *scrapping* de páginas *web*. Embora exista um maior dispêndio de tempo e energia, facilita a estruturação da nossa ontologia dado que não estamos presos a uma estrutura, qualquer que seja o formato da mesma.

No desempenho de *scrapping* foi o **Python** a linguagem escolhida, dado ao seu elevado poder de *scripting* com o auxílio de uma variada panóplias de módulos, com a utilização em concreto das ferramentas **BeautifulSoup** e **Selenium**,

Neste caso, o mecanismo de levantamento e construção da ontologia, são executados em paralelo, isto é, à medida que se retira informação das sucessivas páginas *web*, esta é diretamente injetado num ficheiro à parte que servirá de apêndice ao ficheiro que mantêm a ontologia.

Com esta ontologia pretendia-se basicamente que fosse possível a visualização, de diferentes épocas, de campeonatos distintos, com a amostragem da tabela classificativa e seus intervenientes, bem como os jogos disputados.

As classes que fazem parte do sistema são:

- **Competition** - onde esta tem Season's
- **Season** - época - onde possui classification's e Round's
- **Round** - jornada - onde este tem Game's associados
- **Classification** - onde esta tem as equipas associadas
- **Squad** - plantel que representa os jogadores daquela temporada
- **Player** - jogador com suas características
- etc

A manutenção desta base de dados é efetuada pela plataforma **GraphDB**, que será o servidor de comunicação com a **API**.

## 2.2 API

Esta secção representa o intermediário entre os dados normalizados e a amostragem dos mesmos. Estando nós presentes numa temática ontológica, é com o apoio da linguagem *SPARQL* que são estabelecidas as *queries* à base de dados.

Embora este intervenha na comunicação entre a *Base de Dados* e a *Interface*, este toma um grande papel do lado do processamento de linguagem, pois funciona como filtro à apresentação que os dados irão tomar do lado da *Interface*.

Voltando a atenção novamente para a ontologia criada e para a sua especificação, estabelecem-se algumas necessidades primárias, às quais o sistema a implementar, deve responder, como a representação da tabela classificativa, a representação dos planteis envolvidos, bem como informação relativa ao jogadores.

A manutenção desta **API**, é baseada no plano das aulas da unidade curricular em questão, e é caracterizada, pela construção de um servido em **NodeJS**, que somente fará a comunicação entre a *Base de Dados* e a *Interface*.

Em baixo encontram-se algumas dessas *queries*.

```
PREFIX : <http://www.prc.di.uminho.pt/football#>
select ?pos ?squad ?pg ?win ?draw ?loss ?sc ?p where{
  :seasonprimeira-liga2018 :hasClassification ?class .
  ?class :classification ?pos;
         :squad_name ?squad;
         :played_games ?pg;
         :wins ?win;
         :draws ?draw;
         :losses ?loss;
         :scored_goals ?sc;
         :points ?p.
}
```

```
PREFIX : <http://www.prc.di.uminho.pt/football#>
select ?jnr ?pname ?pos where{
  :squadsc-braga2018 :hasPlayer ?p .
  ?p :jersey_number ?jnr;
     :name ?pname;
     :position ?pos.
}
```

## 3 Representação do Conhecimento - Interface

A interface é o processo final que tem como tarefa a apresentação, de uma forma elegante, da informação contida na base de dados. Neste caso a ferramenta utilizada, mais uma vez lecionada nas aulas, foi o **Vue**, com o apoio estilístico do *addon* **Vuetify**.

Dentro do que se pretendia, esta *Interface* encontra-se capaz de representar a necessidades anteriormente faladas, como a visualização da tabela classificativa, dos plantéis e de alguma informação dos jogadores.

Isto é facilmente comprovado, pelas imagens em baixo.

Pos	Equipa	JJ	V	E	D	GM	GS	P
1	SL Benfica	34	28	3	3		31	87
2	FC Porto	34	27	4	3		20	85
3	Sporting CP	34	23	5	6		33	74
4	SC Braga	34	21	4	9		37	67
5	Vitória SC	34	15	7	12		34	52
6	Moreirense FC	34	16	4	14		44	52
7	Rio Ave FC	34	12	9	13		52	45
8	Boavista FC	34	13	5	16		40	44
9	Belenenses SAD	34	10	13	11		51	43
10	CD Santa Clara	34	11	9	14		45	42
11	CS Marítimo	34	12	3	19		44	39
12	Portimonense	34	11	6	17		59	39
13	Vitória FC	34	8	12	14		39	36
14	CD Aves	34	10	6	18		49	36
15	CD Tondela	34	9	8	17		54	35
16	GD Chaves	34	8	8	18		57	32
17	CD Nacional	34	7	7	20		73	28
18	CD Feirense	34	3	11	20		64	20

figureTabela classificativa *Liga Zon Sagres* 2018/19

Camisola	Nome	Posição
1	Mile Štilar	Guarda-Redes
2	Germain Conti	DefesaCentral
3	Alejandro Grimaldo	LateralEsquerdo
5	Ljubomir Fejsa	MedioDefensivo
6	Ruben Dias	DefesaCentral
7	Sebastien Corchia	LateralDireito
8	Gabriel	MedioCentro
10	Jonas	SegundoAvançado
11	Franco Cervi	MedioEsquerdo
14	Haris Seferovic	Pontadelanica
15	Yuri Ribeiro	LateralEsquerdo
17	Andrija Zivkovic	MedioDireito
18	Eduardo Salvo	ExtremoDireito
20	Filip Krivonovic	MedioOfensivo
21	Pizzi	MedioDireito
22	Andreas Samaris	MedioDefensivo
23	Tyronne Ebuehi	LateralDireito
27	Rafa Silva	MedioEsquerdo

figureTabela com os jogadores do *SL Ben-*  
*fica*

## 4 Conclusão

Primeiramente é importante realçar que um senso de liberdade foi incutido na realização deste projeto, com a escolha do tema a recair sobre os alunos, o que proporciona uma maior liberdade, no que toca a lidar com estruturas de *datasets* complicados, isto permite sem dúvida lidar com esse aspeto de uma forma mais flexível, permitindo aos alunos ir mais longe.

Agora em forma de reflexão o facto, de neste projeto se ter optado pela opção de busca direta pela informação e não pela via da transformação/conversão de dados, veio enriquecer o produto final, dado que, se possuímos uma visão mais distante, não tão focada na chegada, mas no caminho, a realização deste projeto deu uma perspetiva interessante das dificuldades que se encontram por forma a possuir a melhor representação dos dados, desde do momento em que são recolhidos até ao momento em que são observados. Tendo pelo meio a noção do trabalho que requer o seu processamento. Neste aspeto, este trabalho completa bem esse requisito, de trabalho de início a fim.

No entanto, é perceptível que o trabalho nuclear apresentado neste documento, assenta principalmente no *Processamento* dos dados e menos na *Representação* dos mesmos. Nesse aspeto existem melhorias consideráveis que devem ser tomadas em conta para possuir uma visualização mais prazerosa de todo o trabalho efetuado.

Este forma de conclusão, este projeto, é um projeto com conta peso e medida que incute aos alunos diferentes dinâmicas, com a implementação de três componentes distintos, mas comunicantes que formarão um sistema global.