

## Trabalho Prático 2

### Objetivos

Continuação do desenvolvimento do sistema de informação semântico (SI), desenvolvido no TP1, usando a mesma temática, ou outra se for considerado vantajoso, adicionando e complementando com as seguintes funcionalidades:

- Criação de uma ontologia, descrevendo exaustivamente o domínio de conhecimento dos dados utilizados, podendo e devendo ir para além dos conceitos presentes nos dados;
- Utilização da ontologia junto aos dados já existentes, por forma a classificar automaticamente o uso dos dados e otimizar a sua pesquisa – deve ser feito tanto na GraphDB, como no Protégé;
- Criação de um conjunto de regras de inferência que, por um lado, permita o estabelecimento de novas relações entre as diversas entidades e, por outro lado, implemente as classificações automáticas, definidas na ontologia, que os motores de inferência não conseguem processar de forma automática;
- Complementação das funcionalidades do sistema, à custa da informação inferida pela ontologia e pelas regras;
- Complementação do conjunto de dados do SI, através do acesso programático ao *endpoint* sparql da dbpedia e da wikidata;
- Publicação da semântica dos dados do SI, na próprias páginas web do SI.

Deve ser feita utilização das seguintes tecnologias lecionadas:

- Python/Django (programação da aplicação);
- *Triplestore* GraphDB (repositório de dados);
- SPARQL (pesquisa e alteração dos dados);
- RDF (formato dos dados);
- RDFS e OWL (ontologia);
- Protégé (criação e validação da ontologia);
- SPIN (conjunto de inferências);
- SPARQLwrapper (acesso ao *endpoint* da dbpedia e wikidata);
- RDFa e Micro-formatos (publicação da semântica).

São valorizadas as aplicações:

- com maior nível de exploração das tecnologias mencionadas;
- com maior nível de inter-relação entre as tecnologias;

- com interfaces de utilizador mais funcionais e amigáveis.

O trabalho deve ser realizado pelos grupos já estabelecidos para o TP1.

## Entrega do Projeto

- Projeto Web Python/Django para execução em virtual environment python3, com instalação dos requisitos através do ficheiro “requirements.txt” – não são admitidas outras tecnologias ou formatos de requisitos;
- Ficheiro de dados:
  - Ficheiro contendo apenas os factos, sem tipologia – para ser importado para a GraphDB;
  - Ficheiro contendo apenas a ontologia – para ser importado para a GraphDB, para testar a classificação e inferência sobre os dados, previamente importados;
  - Ficheiro com integração da ontologia com os factos, por forma a poder ser testado e validado no Protégé;
  - As regras de inferência em SPIN, devem ser definidas num módulo python independente, por forma a serem explicitamente identificadas;
- Todos estes componentes devem ser entregues por forma a serem facilmente configurados e executados no python3, GraphDB e Protégé, instalados em qualquer máquina;
- Não podem ser usados containers, como Docker, ou outros, para entrega do projeto.

## Relatório

Neste trabalho, o relatório a ser entregue, assume uma importância vital na avaliação do trabalho. Este relatório deve seguir a estrutura de pontos seguinte e na mesma ordem:

1. Introdução ao tema;
2. A definição da ontologia (RDFS e OWL);
3. Conjunto de inferências (SPIN);
4. Novas operações sobre os dados (SPARQL);
5. O uso e integração de dados da Wikidata e/ou DBpedia;
6. A publicação de dados semânticos através de RDFa e microformatos;
7. Funcionalidades da Aplicação (UI), não presentes no TP1 ou complementares;
8. Conclusões;
9. Configuração para executar a aplicação.

## Practical Assignment 2

### Objectives

Continuation of the development of the semantic information system (SI), initiated in TP1, using the same theme or a different one if considered advantageous, by adding and extending it with the following functionalities:

- Creation of an ontology, thoroughly describing the knowledge domain of the data used, going beyond the concepts present in the data whenever possible and appropriate;
- Integration of the ontology with the existing data in order to automatically classify data usage and optimize search operations — this should be implemented both in GraphDB and in Protégé;
- Creation of a set of inference rules that, on one hand, allow the establishment of new relationships between various entities and, on the other hand, implement automatic classifications (as defined in the ontology) that inference engines are unable to process autonomously;
- Enhancement of the system's functionalities by leveraging the information inferred through the ontology and inference rules;
- Expansion of the SI's dataset through programmatic access to the SPARQL endpoints of DBpedia and Wikidata;
- Publishing the semantic annotations of the SI's data directly within its web pages.

The following technologies, covered in class, must be used:

- Python/Django (application programming);
- Triplestore GraphDB (data repository);
- SPARQL (data querying and modification);
- RDF (data format);
- RDFS and OWL (ontology definition);
- Protégé (ontology creation and validation);
- SPIN (set of inference rules);
- SPARQLWrapper (access to DBpedia and Wikidata endpoints);

- RDFa and Microformats (semantic data publication).

Applications will be valued according to:

- The level of exploration and integration of the mentioned technologies;
- The degree of interrelation between the different technologies;
- The quality, functionality, and user-friendliness of their interfaces.

The work must be carried out by the groups already established for TP1.

## Project Submission

- Python/Django Web Project to be executed within a Python 3 virtual environment, with the installation of dependencies managed through the “requirements.txt” file — other technologies or dependency formats are not allowed;
- Data Files:
  - A file containing only the facts, without typology — to be imported into GraphDB;
  - A file containing only the ontology — to be imported into GraphDB, in order to test classification and inference over the previously imported data;
  - A file integrating the ontology with the facts, so it can be tested and validated in Protégé;
  - The SPIN inference rules must be defined in a separate, independent Python module, so they can be explicitly identified;
- All these components must be delivered in a way that allows them to be easily configured and executed on any machine with Python 3, GraphDB, and Protégé installed;
- The use of containers, such as Docker, or any similar technologies, is not allowed for the project submission.

## Report

In this project, the report to be submitted plays a crucial role in the evaluation of the work. This report must follow the structure outlined below, in the same order:

1. Introduction to the topic;
2. Definition of the ontology (RDFS and OWL);
3. Set of inferences (SPIN);
4. New operations on the data (SPARQL);
5. Use and integration of data from Wikidata and/or DBpedia;
6. Publication of semantic data through RDFa and microformats;

7. Application functionalities (UI) not present in TP1 or complementary to it;
8. Conclusions;
9. Configuration to run the application.