



Universidade do Porto
Faculdade de Engenharia

FEUP

D5: Informações sobre literatura em Linguagem Natural

Relatório Intercalar

Inteligência Artificial

3º ano do Mestrado Integrado em Engenharia Informática e Computação

Grupo D5_4

Bruno Piedade – up201505668@fe.up.pt

Danny Soares – up201505509@fe.up.pt

Sérgio Salgado – up201406136@fe.up.pt

08/04/2018

1. Objetivo

O objetivo do trabalho proposto na unidade curricular Inteligência Artificial, do terceiro ano do Mestrado Integrado em Engenharia Informática e Computação, é desenvolver um sistema que permita obter informações sobre escritores e as suas obras literárias usando Linguagem Natural.

Este sistema será desenvolvido em Prolog, utilizando gramáticas de cláusulas definidas (DCG – *Definite Clause Grammar*), por forma a conseguir interpretar e responder a frases fornecidas pelos utilizadores, em linguagem natural. Para dar resposta a estas frases, o sistema possuirá uma base de conhecimento com dados relevantes de escritores e obras literárias.

2. Descrição

2.1. Especificação

Uma base de conhecimento é o que permite a um sistema dar respostas adequadas a cada pergunta ou declaração que lhe é dirigida. No caso deste trabalho, a base de conhecimento representa o conhecimento do sistema em relação a escritores e às suas obras literárias, tendo portanto que conter informação sobre os escritores (nome, data de nascimento, data de falecimento, sexo, nacionalidade, número de prémios e heterónimos/pseudónimos) e sobre as suas obras literárias (título, autores, ano de publicação, editora e géneros). Além dos escritores e das obras, a base de conhecimento tem ainda duas outras estruturas, uma que representa os géneros dos livros e outra que representa os países. Esta base de conhecimento tem, então, a seguinte estrutura:

- autor(id, primeiro_nome, outros_nomes, data_nasc, data_mort, sexo, id_pais, n_premios, [heretónimos]);
- livro(id, titulo, [autor_ids], ano_pub, editora, [generos_ids]);
- genero(id, nome);
- pais(id, nome, continente, nacionalidade_fem, nacionalidade_masc).

Para além da base de conhecimento, o sistema precisa de uma DCG, que representa a gramática da linguagem natural dos utilizadores. Esta DCG é o que permite ao sistema reconhecer as perguntas ou declarações dos utilizadores, analisar a sua sintaxe e a sua semântica e perceber que tipo de resposta devolver. Os utilizadores podem introduzir frases declarativas, às quais o programa responde afirmativa ou negativamente, dependendo da validade da declaração, ou frases interrogativas, que exigem uma resposta com informação da base de conhecimento. A estrutura inicial da gramática é a seguinte:

frase --> declarativa

frase --> interrogativa

declarativa --> sintagma_nominal, sintagma_verbal, terminal.

interrogativa --> pronome, sintagma_verbal, [?].

sintagma_nominal --> sintagma_nominal_aux.

sintagma_nominal --> sintagma_nominal_aux, [e], sintagma_nominal_aux.

sintagma_nominal_aux --> sintagma_nominal_aux2.

sintagma_nominal_aux --> sintagma_nominal_aux2,
sintagma_preposicional.

sintagma_nominal_aux2 --> sintagma_nominal_aux3.

sintagma_nominal_aux2 --> sintagma_nominal_aux3, sintagma_adjetival.

sintagma_nominal_aux3 --> pronome_indef, nome.

sintagma_nominal_aux3 --> artigo, nome.

```
sintagma_nominal_aux3 --> nome.
```

```
sintagma_preposicional -> preposicao, sintagma_nominal.
```

```
sintagma_adjetival --> sintagma_adjetival_aux.
```

```
sintagma_adjetival --> sintagma_adjetival_aux, [e],  
                        sintagma_adjetival.
```

```
sintagma_adjetival_aux --> adjetivo.
```

```
sintagma_adjetival_aux --> adverbio, adjetivo.
```

```
sintagma_verbal --> verbo, sintagma_dec_nominal.
```

```
sintagma_verbal --> sintagma_nominal, verbo, sintagma_preposicional.
```

```
sintagma_verbal --> sintagma_nominal.
```

```
terminal --> [?].
```

```
terminal --> [.].
```

2.2. Trabalho efetuado

O grupo começou por decidir qual a informação que seria relevante numa base de conhecimento sobre literatura, de modo a conseguir uma representação variada da informação, tornando a experiência do utilizador o mais completa possível, mas contida na área da literatura.

Com a base de conhecimento feita, foi discutida de forma básica a implementação da DCG a utilizar e o tipo de perguntas que seriam aceites por essa DCG. Procedeu-se a um estudo de análise sintática e semântica que seria feita ao input do utilizador e de como representar as regras de produção da gramática em Prolog.

Por fim, foi discutida a criação da interface do utilizador, focada mais na utilização da linguagem Java, mais especificamente, Swing.

2.3. Resultados esperados e forma de avaliação

Para avaliar os resultados do sistema, podem ser utilizadas frases de exemplo, válidas na linguagem natural dos utilizadores, que representam parte das frases que podem ser introduzidas no sistema. Algumas destas frases estão na seguinte lista:

- Quem escreveu Os Maias?
- Quais os livros mais conhecidos de Fernando Pessoa?
- Quem/Quais são os escritores africanos?
- E portugueses?
- Quais (são) os escritores portugueses e espanhóis do século XV?
- Quantos livros de escritores africanos existem após o século XVI?

- E qual é o mais recente?
- Alberto Caeiro é heterónimo de Fernando Pessoa.

As respostas a todas estas frases são conhecidas e facilmente acessíveis na base de conhecimento, portanto os resultados esperados são conhecidos, podendo confirmar-se os resultados dados pelo programa.

3. CONCLUSÕES

Concluindo, acreditamos que o projeto será interessante, que está bem organizado e planeado e que poderemos obter bons resultados com a sua implementação, além de recebermos uma nova perspectiva no que diz respeito a inteligência artificial e bots de resposta a linguagem natural.

4. RECURSOS

→ Apontamentos das aulas sobre linguagem natural, disponibilizados em:

<https://paginas.fe.up.pt/~eol/IA/1718/trabalho.html>

→ http://www.amzi.com/manuals/amzi/pro/ref_dcg.htm