



Universidade do Porto
Faculdade de Engenharia

FEUP

D5: Informações sobre literatura em Linguagem Natural

Relatório Final

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

20/05/2018

1. Objetivo

No âmbito da unidade curricular de Inteligência Artificial, do terceiro ano do Mestrado Integrado em Engenharia Informática e Computação, foi desenvolvido um sistema que permite obter informações sobre escritores e as suas obras literárias usando Linguagem Natural.

Este sistema foi desenvolvido em *Prolog*, usando uma gramática de cláusulas definidas (*DCG – Definite Clause Grammar*), e consegue interpretar e responder a frases fornecidas pelos utilizadores, em linguagem natural, neste caso, português. O sistema utiliza uma base de conhecimento com informações sobre escritores e obras, as quais permitem dar respostas corretas e adequadas a cada frase.

2. Especificação

2.1. Análise do Tema e Ilustração de Cenários

O objetivo do nosso sistema é, como já foi descrito, dar respostas a frases que lhe são dirigidas pelo utilizador. Este sistema compreende dois tipos de frases: interrogativas e declarativas. Relativamente às interrogativas, o sistema recebe uma pergunta e responde com a informação devida, enquanto que para as declarativas responde com “Concordo” ou “Discordo” de acordo com a veracidade da frase. As frases apresentadas ao sistema não podem incluir acentos, pelo que palavras acentuadas devem ser escritas sem acentos.

Para ilustrar alguns exemplos de frases suportadas pelo nosso sistema, apresentamos, a baixo, uma lista de frases exemplo, dividida em “Frases declarativas” e “Frases interrogativas”. Estes exemplos representam apenas a estrutura e o conteúdo de frases que o sistema compreende e sabe responder, não sendo estes os únicos exemplos que são aceites.

Frases declarativas

- Alberto Caeiro é heterónimo de Fernando Pessoa.
- A Mensagem é uma poesia.
- Fernando Pessoa é português. (Fernando Pessoa é europeu.)
- A Mensagem é o livro mais conhecido de Fernando Pessoa.
- Fernando Pessoa escreveu A Mensagem.
- Fernando Pessoa está morto.
- Fernando Pessoa nasceu em 1888. (Fernando Pessoa nasceu no século XIX.)
- Fernando Pessoa morreu em 1935. (Fernando Pessoa morreu no século XX.)

Frases interrogativas

- Quem escreveu Os Maias?
- E A Mensagem?
- Quem são os escritores portugueses que nasceram antes do ano 1888?
- Quem nasceu em 1888?
- Quais são os livros mais conhecidos de Camilo Castelo Branco?
- Quais os livros mais conhecidos de escritores ingleses e africanos?
- Quais os escritores portugueses do século XX?
- Quais os livros de escritores portugueses e africanos que existem?
- Quais os livros de escritores portugueses que existem após o século XIX?

2.2. Abordagem

Para tratar do problema em questão, recorremos a uma *DCG*, que nos permitiu desenvolver em *Prolog* um conjunto de predicados que expressam linguagem natural.

A utilização desta *DCG* facilita o reconhecimento de frases com estruturas diferentes, muito ou pouco complexas e até frases iniciadas por “E” que implicam o uso do contexto da frase anterior para responder à nova frase. O sistema analisa a sintaxe e a semântica das frases, de forma a só responder a frases que respeitem a sintaxe e a semântica da linguagem natural utilizada.

Para dar respostas adequadas a um grande leque de perguntas sobre vários escritores e várias obras construímos uma base de conhecimento que o sistema usa. Esta base de conhecimento tem informações sobre escritores, obras, géneros e países. Esta informação permite então localizar geograficamente escritores e obras, permite identificar livros por géneros, etc. Para cada autor é possível saber o seu nome (primeiro e restantes), ano de nascimento e, caso se aplique, de morte, sexo, nacionalidade, prémios e heterónimos, enquanto que para cada livro é possível saber o título, os autores, ano de publicação, géneros e popularidade.

3. Desenvolvimento

3.1. Ferramentas utilizadas

Para desenvolver este sistema, usámos, como já foi referido neste relatório, a linguagem de programação *Prolog*. Todo o tratamento lógico feito às frases inseridas pelo utilizador é feito em *SWI-Prolog*.

Para proporcionar aos utilizadores uma experiência mais simples e agradável, desenvolvemos uma interface simples e intuitiva para a consola, na qual os utilizadores podem fazer perguntas à sua vontade e têm acesso a uma secção com exemplos de frases que podem usar para testar o sistema. Esta interface foi, também, desenvolvida em *SWI-Prolog*.

O sistema funciona perfeitamente em sistemas operativos *Windows*, *MacOS* e *Linux*.

3.2. Estrutura da aplicação

O código desenvolvido encontra-se organizado em vários ficheiros “.pl” que representam todo o sistema. Os ficheiros são os seguintes:

→ **ui.pl**: interface de utilizador, com tratamento da navegação na aplicação, tratamento das *inputs* dadas ao sistema para serem corretamente passadas à unidade lógica;

→ **lexicon.pl**: léxico do sistema, com informações sintáticas e semânticas, relativas à linguagem natural;

→ **kb.pl**: base de conhecimento do sistema, com informações sobre autores, livros, géneros e países e predicados para retornar estas informações;

→ **query.pl**: predicados úteis à gramática, para verificações de séculos, anos, popularidade de livros, etc;

→ **utils.pl**: predicados úteis a várias secções da aplicação, quer a interface, quer a gramática;

→ **grammar.pl**: DCG implementada, com as regras gramaticais relativas à estrutura das frases.

4. Conclusões

Após terminar o projeto, confirmamos as nossas expectativas iniciais, concluindo que o projeto foi interessante e obtivemos bons resultados. Os objetivos estabelecidos foram todos cumpridos, o que demonstra o esforço e união do grupo.

Os conhecimentos adquiridos com o desenvolvimento do projeto parecem-nos interessantes e que virão a ser úteis no futuro, não só para outras unidades curriculares do curso, mas também para o nosso futuro profissional.

O grupo organizou e repartiu as tarefas, tendo cada elemento cumprido a sua parte. Desta forma, a participação dos três elementos do grupo foi aproximadamente igual, não havendo diferenças a apontar.

5. Melhoramentos para trabalho futuro

Após concluir o trabalho, o grupo acredita que ainda seria possível melhorar o sistema, incluindo alguns tipos de frases que não são reconhecidos ainda.

Futuramente poderá acrescentar-se ao sistema a aceitação de frases negativas (com uso da palavra “não”), do tipo “Fernando Pessoa não escreveu Os Maias.” e a aceitação de frases passivas, do tipo “Os Maias foram escritos por Eça de Queirós.”.

Além de acrescentar novos tipos de frases ao sistema, pode-se aumentar o vocabulário do sistema, incluindo mais sinónimos para permitir que mais frases com o mesmo significado mas estruturas muito diferentes sejam todas aceites.

6. Recursos

6.1. Bibliografia

→ 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

6.2. Divisão de tarefas

Todos os elementos participaram em todos os elementos avaliados neste projeto (desenvolvimento da aplicação e relatório), em proporções equivalentes.

7. Apêndice

7.1. Manual do utilizador

Para utilizar o sistema desenvolvido, é necessário software capaz de interpretar *SWI-Prolog*. Utilizando a consola do *SWI-Prolog* é necessário consultar o ficheiro *grammar.pl* e usar o comando "start." para iniciar a aplicação.

Após iniciar a aplicação, o utilizador pode escolher a opção 1, na qual pode escrever frases à sua vontade, que serão interpretadas pelo sistema, ou a opção 2, na qual lhe é apresentada uma lista de exemplos de frases, das quais pode escolher 1 para ser respondida pelo sistema.

Em ambas as opções o utilizador tem a possibilidade de voltar ao menu inicial, escrevendo na consola "menu" na 1ª opção e "0" na 2ª.