

Língua Natural

Instituto Superior Técnico - Campus TagusPark

Mini Projecto 1 (2015/2016)

Número do grupo: G03

Nome: Bruno Alexandre Pires Henriques

Número: 72913

Nome: Tiago Manuel Ferrão dos Santos

Número: 72960

1. Introdução

O projecto explicado neste relatório tem como objectivo criar uma baseline para uma aplicação de Língua Natural cujo propósito é identificar uma resposta para uma determinada frase através de um *corpus* constituído por frases e pares de pergunta resposta associados a essa frase. Através de um corpus com as respostas anotadas avaliou-se o seu desempenho tendo em conta a resposta escolhida.

2. Arquitectura

Na aplicação, o utilizador apenas interage com um módulo, o módulo 'Interface'. Este módulo tem ambas as funções pedidas no enunciado: 'sss' e 'myAvalia'. A integração deste módulo com os restantes encontra-se ilustrada na Figura 1.

A função 'sss' cria uma instância do módulo 'AnswerPicker' com o caminho para o ficheiro de perguntas para seguidamente usá-lo, obtendo uma resposta para a pergunta fornecida pelo utilizador a partir do ficheiro. A instância de 'AnswerPicker', na sua génese, cria um mapa entre perguntas e respostas para o ficheiro em questão através de comparações e expressões regulares para quando receber um pedido poder dar uma resposta de forma eficiente evitando, portanto, fazer procuras sucessivas no ficheiro. A resposta escolhida é a mais frequente dentro das respostas que correspondem a frases idênticas à pergunta após uma normalização de ambas, que retira pontuação, acentos e maiúsculas. O 'AnswerPicker' utiliza o módulo 'RegexUtil' para normalizações.

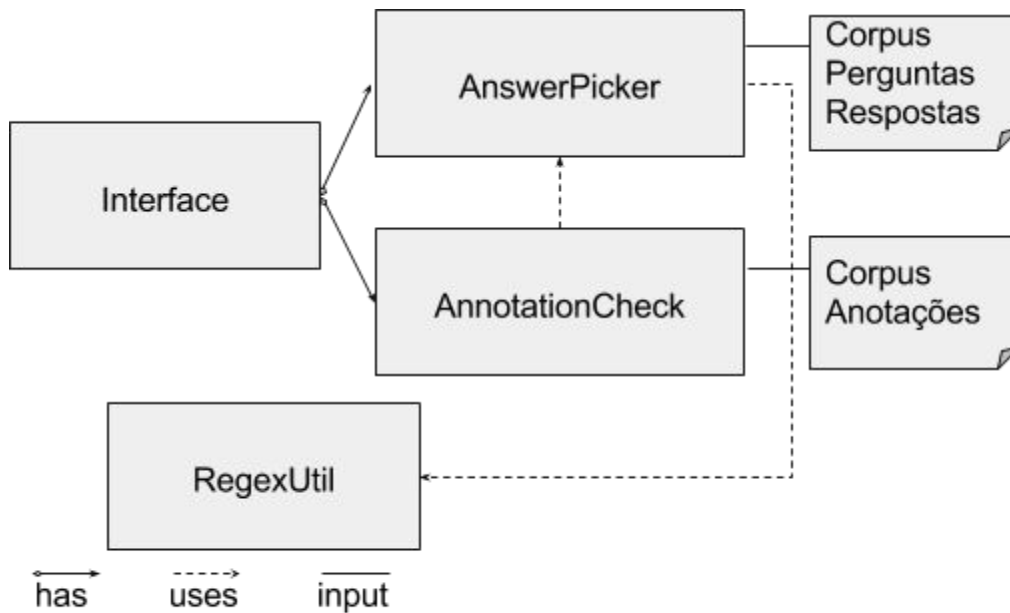


Figura 1 - Arquitectura da aplicação

A função ‘myAvalia’ inicialmente tem um comportamento semelhante ao da função ‘sss’, criando uma instância de ‘AnswerPicker’, no entanto, em vez de a usar para obter respostas directamente, fornece-a a uma instância do módulo ‘AnnotationCheck’. Este módulo vai obter respostas para todas as perguntas no ficheiro fornecido pelo utilizador e procurar a anotação correspondente a essas respostas (através de expressões regulares também). Seguidamente calcula a *accuracy* para esse conjunto de perguntas respostas usando a seguinte fórmula:

$$accuracy = \frac{count(plausible\ answers)}{count(answers)}$$

3. Setup Experimental

Para desenvolvimento desta aplicação, usámos 2 *corpora* para fins experimentais, um com perguntas, frases e respostas agrupadas da seguinte forma:

```

User Input: Estou a fazer uma pergunta?
...
T - Que perguntas?
A - Todas as perguntas.
...
  
```

E o outro com todas as respostas associadas à pergunta anotadas de acordo com a sua

plausibilidade em relação à pergunta em questão. Sendo que respostas anotadas com ‘y’ são consideradas plausíveis, as com ‘n’ não são plausíveis de todo, e as anotadas com ‘m’ poderão ser plausíveis dependendo do contexto. Segue-se um exemplo:

```
...
User Input - Estou a fazer uma pergunta?
A - Olha este, com a mania que é Platão. : y
A - Tiago, já alimentaste o Bruno hoje? : m
A - BANANAS têm potássio, e potássio dá-me gases! : n
...
```

4. Resultados

Através da função ‘myAvalia’ avaliou-se a escolha das perguntas, sendo que para todos os inputs possíveis a accuracy foi a seguinte:

- Considerando apenas as respostas anotadas com ‘y’ como respostas plausíveis):

$$accuracy = \frac{count('y')}{count(Perguntas)} = \frac{43}{232} \approx 18.53\% (2 c.d.)$$

- Considerando as respostas anotadas com ‘y’ ou com ‘m’:

$$accuracy = \frac{count('y') + count('m')}{count(Perguntas)} = \frac{43 + 11}{232} \approx 23.28\% (2 c.d.)$$

5. Discussão dos resultados e conclusão

O resultado esperado está de encontro com o esperado na medida que não foram utilizadas técnicas sofisticadas de normalização ou de comparação de palavras como, por exemplo:

- *Stemmers* para reduzir as frases à sua essência: ‘Não’ vs ‘Não, de certeza’
- Remoção de ‘stop-words’ frequentes, como por exemplo, “é que” (441 matches no ficheiro ‘PerguntasPosSistema.txt’);
- Medidas de comparação de palavras: *Jaccard*, *dice* ou *f-measure*.

Adicionalmente, as anotações marcadas como ambíguas dependente do contexto, i.e. ‘m’, foram consideradas como ‘*true negative*’ pelo que contribuiu para a redução da accuracy em cerca de 5% pelo que, mesmo assim não deu grande impacto. Na segunda parte do projecto estas ambiguidades estes casos serão tratados com maior detalhe e, juntando os procedimentos acima referidos, acreditamos que é possível aumentar bastante a accuracy apresentada reduzindo a quantidade de *false negative* e aumentando o número de *true positives*.