



8ª. LISTA DE EXERCÍCIOS

1. Sumarização Automática Monodocumento

Desenvolva um sistema de sumarização automática (mono-documento) que, após um pré-processamento básico de remoção de stopwords, considere os seguintes pesos dos tokens dos documentos a serem sumarizados:

FEATURE 1. TF-IDF dos termos de uma sentença

FEATURE 2. Atribuir os pesos de acordo com a função gramática da palavra na frase:

- Peso = 3, se o termo é um sujeito
- Peso = 2, se o termo é um objeto
- Peso = 1, todos os outros casos

Use um parser de dependências para encontrar tal informação (Stanford CoreNLP)

Exemplo:

Dado o texto abaixo:

Haemorrhage is a common cause of death in trauma patients.

Although transfusions are extensively used in the care of bleeding trauma patients, there is uncertainty about the balance of risks and benefits and how this balance depends on the baseline risk of death.

Our objective was to evaluate the association of red blood cell (RBC) transfusion with mortality according to the predicted risk of death.

As sentenças acima tem as seguintes dependencias

- **nsubj** (cause-5 , Haemorrhage-1) → Peso = 3
- **nsubjpass** (used-5 , transfusions-2) → peso = 3
- **nsubj** (is-15 , uncertainty-16) → peso = 3
- **nsubj** (depends-28 , balance-27)
- **dobj** (evaluate-5 , association-7) → peso = 2
- **Todos os outros casos** → peso = 1

Onde o termo em amarelo recebe o peso correspondente.

FEATURE 3. os mesmos pesos da *Feature 2* porém normalizados, isto é, a soma de todos os pesos dos termos de uma sentença deve ser dividido pelo comprimento da maior sentença presente neste documento

Em todo os casos, o ranqueamento final de uma sentença vai ser igual a soma dos pesos de todas de suas palavras ou tokens constituintes.

Avaliar usando o dataset de **News_LE_09.zip** que contém notícias (originais) acompanhada de 2 resumos de referência (gold).

Esta avaliação deve ser feita em termos da métrica **R-2** (Recall usando bi-grama) usando a ferramenta de avaliação ROUGE disponível aqui <http://www.rxnlp.com/rouge-2-0/>

PERGUNTAS:

1. Qual combinação de features foi mais efetiva em termos *quantitativos* (ROUGE metrics)?

Para esta avaliação, teste todas as combinações possíveis com até 3 features, seja isoladamente ou de forma incremental da seguinte forma:

- usando Feature 1
- usando Feature 2
- Usando Feature 3
- usando Feature 1+2
- usando feature 2+3
- Usando feature 1+3
- usando feature 1+2+3

2. Qual combinação de features foi mais efetiva em termos *qualitativos* (coesão e coerência)?

3. Houve casos de pronomes soltos (referências quebradas) nos resumos gerados?
O que você sugeriria para evitar este problema?

4. Selecione a melhor combinação do item (1) acima, e adicione a ela a heurística de posição, definida assim:

$\text{sent_pos}(c) = 1 - (\text{índice da sentença no texto, que contém o token T}) / S$

Onde S = é número total de sentenças do documento .

Discuta se esta heurística traz alguma melhoria nos resultados.

Obs.: Ressalta-se que os critérios de pesos acima são apenas usados como um forma de ranquear as frases, por relevância. Depois de selecionadas as frases, se deve gerar um resumo adotando-se a mesma ordem das frases no documento original.