



Universidade Estadual de Maringá

Departamento de Informática



Trabalho de IIA – Processamento de Linguagem Natural

Modalidade: individual ou em grupo (máximo 2 integrantes)

Objetivo: utilizar técnicas de PLN para analisar textos científicos.

Materiais:

Baixar 10 artigos científicos do portal <https://ieeexplore.ieee.org/Xplore/home.jsp>
(escolha um termo de busca relevante na computação)
PyPDF2, python-pdfbox ou qualquer outra à escolha da equipe

Descrição do trabalho: modelar uma ontologia sobre a estrutura de um artigo científico, e a partir dela implementar um software que execute as seguintes tarefas:

- 30% da nota,
 - ler arquivos pdf de um diretório
 - utilizar uma biblioteca para leitura de pdf (ex.: PyPDF2, python-pdfbox)
 - realizar pré-processamento de texto (remoção de *stop-words*, lematização, *steaming*)
 - utilização de modelos de linguagem como *bag-of-words*, n-grama, etc
 - identifique os 10 termos mais citados nos artigos (não considerando as referências bibliográficas – *references*)
 - extraia as referências bibliográficas
- 30% da nota, extrair informações sobre:
 - Objetivo (*objective*) do artigo (normalmente na introdução), ex.: "The objective of this article is to contribute to the debate on whether or not marketing boards have the capacity to de-crease transaction costs faced by producers with an evaluation of the magnitude of transaction costs incurred by producerswhen selling their milk through a marketing board"
 - Problema (*problem*) do artigo (normalmente na introdução), ex.: "That way, our article overcomes a transaction costs economics' basic selection problem"
 - método ou metodologia (method, methodology, interviews, survey, content analysis, etc), ex.: "Interviews and surveys were conducted to estimate transaction costs faced by producers marketing through the Qu´ebec milk marketing board in Canada andbilateral contracts in England and Wales in the United Kingdom using the measurement methodology of the cost of exchange"
 - Contribuição (contributes to) do artigo (não pode ocorrer junto com a palavra objective), ex.: "Our study contributes to this neglected literature whilefocusing on the comparison of two distinct hybrid coordination mechanisms used in developed countries"
- (5% da nota) Salvar os dados extraídos em arquivo, onde cada informação será separada por ";;"



Universidade Estadual de Maringá

Departamento de Informática



- (10% da nota) Criar uma interface de busca simples que liste quais arquivos contém um termo de busca, ordene por exemplo utilizando o cálculo de "page rank".

(25% da nota) Criar uma apresentação em slides contendo:

1. apresentação: universidade, curso, disciplina, título do trabalho, equipe, professor
2. introdução (descrição/modelagem do problema a ser resolvido)
3. fundamentação (breve)
4. materiais e métodos
5. dados sobre os experimentos
6. conclusões
7. bibliografias utilizadas: sites, livros, etc.

O código fonte, executável, dados de teste do problema e slides deverão ser armazenados na nuvem e enviado um link de download para o e-mail wigarashi@uem.br até a data limite de entrega do trabalho, contendo: título do email – Trabalho 2 de IIA; no corpo do email: o nome completo dos integrantes da equipe e RA. Trabalhos entregues em atraso terão desconto na nota final.

Obs: A avaliação do trabalho será realizada utilizando os seguintes critérios: originalidade, complexidade do problema abordado, apresentação dos resultados, completude dos itens especificados, português correto, código compilável e executável. Será utilizado software de verificação de similaridade, assim se houver indícios de cópia de trabalho, os trabalhos envolvidos terão nota 0.