



# Universidade Estadual de Maringá

## Departamento de Informática



### Trabalho 1 – parte 2 - IA

Modalidade: individual ou em grupo (máximo 3 integrantes)

Área: IA - Processamento de linguagem natural (PLN).

Objetivo: utilizar técnicas de PLN para analisar textos científicos.

Materiais:

artigos que serão disponibilizados pelo professor:

[https://drive.google.com/drive/folders/10MfRRyy761bl-SMIoKXJnJfkYzAH-d\\_c?usp=sharing](https://drive.google.com/drive/folders/10MfRRyy761bl-SMIoKXJnJfkYzAH-d_c?usp=sharing)

PDFBox (leitor de pdf em Java)

Maven

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:

- **70% da nota**
  - ler arquivos pdf de um diretório
  - utilizar uma biblioteca para leitura de pdf (ex.: PDFBox)
  - realizar algum tipo de pré-processamento de texto
  - identifique os 10 termos mais citados nos artigos (não considerando as referências bibliográficas – *references*)
  - extraia as referências bibliográficas
  - a identificação de instituições/autor do artigo
- **10% da nota**, extrair parágrafos 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



## Universidade Estadual de Maringá Departamento de Informática



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 “;”
- Analisar precisão e recall **(5% da nota)**

Criar uma apresentação em slides contendo **(10% da nota)**:

apresentação: universidade, curso, disciplina, título do trabalho, equipe, professor

descrição/modelagem do problema a ser resolvido

análise de requisitos: casos de uso, descrição das funcionalidades da implementação

plataforma: dados do computador, SO, linguagem de programação

como se faz para testar os programas

trechos de código

telas de *screenshot* da execução de simulações

bibliografias utilizadas: sites, livros, etc.

Obs.: Bônus para quem implementar a parte de IHC com interface gráfica.

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 [wigarash@gmail.com](mailto:wigarash@gmail.com) até a data limite de entrega do trabalho, contendo 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.**