



PROGRAMA DE DISCIPLINA

IDENTIFICAÇÃO

DISCIPLINA: **Mineração de texto** CÓDIGO: **PGIA7345**
DEPARTAMENTO/UNIDADE ACADÊMICA: **DEINFO**
ÁREA: **COMPUTAÇÃO INTELIGENTE E MODELAGEM**
CARGA HORÁRIA TOTAL: **60 h** NÚMERO DE CRÉDITOS: **04**
CARGA HORÁRIA SEMANAL: **4 h** TEÓRICAS: **2 h** PRÁTICAS: **2 h**
PRÉ-REQUISITOS: **NENHUM**
CO-REQUISITOS: **NENHUM**

EMENTA

Processamento de linguagem natural. Associação entre palavras. Recuperação de informação. Classificação e agrupamento de texto. Extração de informação. Sumarização de texto.

CONTEÚDOS

1. Processamento de linguagem natural
 - 1.1. Introdução
 - 1.2. Representação de texto
 - 1.3. Vantagens e desvantagens das representações
 - 1.4. Serviços de pré-processamento
2. Associação de palavras
 - 2.1. Relações paradigmática
 - 2.2. Relações sintagmática
 - 2.3. Métodos baseados em entropia
 - 2.4. Métodos baseados em informações mútuas
 - 2.5. Métodos semânticos e LSA
3. Recuperação de informação
 - 3.1. Modelos probabilísticos
 - 3.2. Indexação e busca
 - 3.3. Avaliação
4. Mineração de texto
 - 4.1. Agrupamento de texto
 - 4.2. Categorização de texto
 - 4.3. Análise de sentimento
 - 4.4. Avaliação
5. Extração de informação
 - 5.1. Arquiteturas
 - 5.2. Algoritmos indutivos
 - 5.3. Métodos probabilísticos
 - 5.4. Avaliação
6. Sumarização de texto
 - 6.1. Métodos extrativos
 - 6.2. Métodos abstrativos
 - 6.3. Avaliação

BIBLIOGRAFIA

BIBLIOGRAFIA BÁSICA:

1. Baeza-Yates, R., & Ribeiro-Neto, B. (2013). *Recuperação de Informação-: Conceitos e Tecnologia das Máquinas de Busca*. Bookman Editora.
2. Feldman, R., & Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge university press.
3. Srivastava, A. N., & Sahami, M. (2009). *Text mining: Classification, clustering, and applications*. Chapman and Hall/CRC.

BIBLIOGRAFIA COMPLEMENTAR:



**UNIVERSIDADE FEDERAL RURAL DE
PERNAMBUCO**

Rua Dom Manoel de Medeiros, s/n, Dois Irmãos - Recife/PE
CEP: 52171-900 | www.ufrpe.br

1. Torres-Moreno, J. M. (2014). *Automatic text summarization*. John Wiley & Sons.
2. Liu, H., & Motoda, H. (Eds.). (1998). *Feature extraction, construction and selection: A data mining perspective* (Vol. 453). Springer Science & Business Media.
3. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2), 1-135.
4. Manning, C. D., Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT press.
5. Weiss, S. M., Indurkha, N., Zhang, T., & Damerou, F. (2010). *Text mining: predictive methods for analyzing unstructured information*. Springer Science & Business Media.

Emitido em: ____/____/____

Responsável: _____