

3.7 Programa de Disciplina (Formulário SUPAC/UFBA)

UNIVERSIDADE FEDERAL DA BAHIA

SUPERINTENDÊNCIA ACADÊMICA PROGRAMA DE DISCIPLINAS SECRETARIA GERAL DOS
CURSOS**OBS: A carga horária básica deve ser de 17h ou múltiplo de 17h, conforme Res. 05/03 do CONSEPE.**

DISCIPLINAS

Código	Nome
	Mineração de texto

Carga Horária				Crédito	Assinatura do Chefe do Departamento	Ano
T	P	E	Total			
	34					

Ementa / Objetivos

Ementa:

Introdução à mineração de texto; Pré-processamento de texto e extração de informação; Categorização de texto; Agrupamento de texto; Modelagem de tópicos; Análise de sentimento; Visualização de texto e Aplicações.

Objetivos:

Neste curso iremos apresentar uma variedade de princípios, técnicas e avanços atuais em mineração de texto.

Metodologia

Conteúdo Programático

1. Introdução à mineração de texto
 - a. Exemplos de aplicação
 - b. Principais desafios
2. Pré-processamento de texto e extração de informação
 - a. Bag of words
 - b. Expressões Regulares
 - c. Modelo N-gram
 - d. Text Encoding Initiative
3. Aplicação dos métodos de aprendizado em texto
 - a. Métodos supervisionados
 - b. Métodos não supervisionados
4. Topic modelling
5. Análise de sentimento
6. Visualização de texto
7. Aplicações
 - a. Redes sociais (Twitter e Facebook)
 - b. Filtro de spam
 - c. Criação de sugestões e recomendações (e.g. Amazon)

Bibliografia

Bibliografia básica:

Charu C. Aggarwal and ChengXiang Zhai (2012). Mining Text Data. Springer.

Pang, B. and Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, 2(1-2): 1–135. Available online at: <http://www.cs.cornell.edu/home/llee/opinion-mining-sentiment-analysis-survey.html>.

Weiss, S. M., Indurkha, N., Zhang, T. (2010). Fundamentals of Predictive Text Mining. Springer: New York. ISBN: 978-1849962254.

Bibliografia complementar:

Manning, C. D., Raghavan, P., and Schütze, H. (2008). Introduction to Information Retrieval, Chapters 6 and 13-18, Cambridge University Press. Available online at: <http://nlp.stanford.edu/IR-book/>
