

MINISTÉRIO DA EDUCAÇÃO
CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA CELSO SUCKOW DA FONSECA
DIRETORIA DE ENSINO (DIREN)
DEPARTAMENTO DE EDUCAÇÃO SUPERIOR (DEPES)
DEPARTAMENTO DE INFORMÁTICA (DEPIN)
CURSO CIENCIA DA COMPUTAÇÃO

DEPARTAMENTO
DEPIN - Departamento Acadêmico de Informática

PLANO DE CURSO DA DISCIPLINA
MINERAÇÃO DE TEXTO

CÓDIGO
GCC 1933

PERÍODO
optativa

ANO
2017

SEMESTRE
2º

PRÉ-REQUISITOS
GCC1311 Programação Orientada a Objetos

CRÉDITOS
4

AULAS/SEMANA		
TEÓRICA	PRÁTICA	ESTÁGIO
4	0	0

TOTAL DE AULAS NO SEMESTRE
72

EMENTA

Visão geral de mineração de textos e aplicações. Processamento de linguagem natural e Representação de documentos. Processo de Descoberta de Conhecimento em Texto (KDT). Análise Exploratória de Texto. Pré-processamento de Texto: Stopwords; Stemming; Dicionário ou Thesaurus. Agrupamento e classificação de textos. Análise de sentimento e mineração de opiniões. Métricas de Avaliação.

BIBLIOGRAFIA

Bibliografia básica

1. KAO, Anne; POTEET, Stephen; Natural language processing and text mining. London: Springer 2007. ISBN 184628175.
2. HAN, Jiawei; KAMBER, Micheline; PEI, Jian. Data mining : concepts and techniques, Waltham, MA : Elsevier, 2012. ISBN 9780123814791.
3. Manning, Christopher; Schütze, Hinrich. Foundations of statistical natural language processing. Cambridge, Mass. : MIT Press, 1999. ISBN 9780262133609.

Bibliografia complementar

1. XUEDONG, Huang,. Spoken language processing: a guide to theory, algorithm, and system development. Upper Saddle River, N.J. : Prentice Hall PTR, 2001. ISBN 0130226165.
2. MCLOUGHLIN, Ian. Applied speech and audio processing : with MATLAB examples. Cambridge, UK : Cambridge University Press, 2009. ISBN 9780521519540.
3. LIU, Huan; MOTODA, Hiroshi. Motoda. Computational methods of feature selection. Boca Raton, FL : Chapman & Hall/CRC Press, c2008. ISBN 9781584888789.
4. Russell, Matthew. Mineração de dados da web social : análise de dados do Facebook, Twitter, LinkedIn e outros sites de mídia social. São Paulo : Novatec, 2011. ISBN 9788575222454.

5. Tan, Pang-Ning. Introdução ao Datamining : mineração de dados. Rio de Janeiro : Ciência Moderna Ltda, 2009. ISBN 9788573937619

OBJETIVO GERAL

Introduzir conceitos, técnicas e ferramentas pertinentes a mineração de texto.

METODOLOGIA

- Aulas expositivas, eventualmente contando com recursos audiovisuais.
- Resolução de exercícios de fixação e propostos.

CRITÉRIO DE AVALIAÇÃO

A avaliação semestral envolve duas provas escritas (P1 e P2). As datas das provas são agendadas conforme o calendário da Instituição. A média parcial (MP) será calculada pelo cômputo da média aritmética simples entre a nota P1 e P2:

$$MP = (P1 + P2) / 2$$

O aluno que faltar a uma das duas provas terá direito a uma avaliação alternativa, denominada segunda chamada, versando sobre todos os tópicos abordados no curso, e cuja data também é agendada entre docente e discentes. A nota obtida nessa 2ª chamada substituirá a da avaliação P1 ou P2 onde o aluno não esteve presente. Caso ele falte às duas avaliações, terá atribuído o grau ZERO em uma delas.

Segundo o regimento do CEFET-RJ, caso o aluno obtenha média parcial inferior a 3,0 (três e zero) estará reprovado diretamente. Graus MP maiores ou iguais a 7,0 (sete e zero) aprovam diretamente o aluno. Em situações onde o aluno tenha grau MP entre 3,0 inclusive e 7,0 exclusive, terá direito a uma prova final (PF), que, juntamente com a média parcial gerará uma nova média, denominada média final (MF). Essa média é calculada da seguinte forma:

$$MF = (MP + PF) / 2$$

Para ser aprovado, o aluno deve alcançar uma MF maior ou igual a 5,0 (cinco e zero). Caso contrário, estará reprovado, devendo repetir a componente curricular.

CHEFE DO DEPARTAMENTO

NOME	ASSINATURA

PROFESSOR RESPONSÁVEL PELA DISCIPLINA

NOME	ASSINATURA
Gustavo Guedes	

PROGRAMA

1. Visão Geral de mineração de textos e aplicações
2. Processo de Descoberta de Conhecimento em Texto (KDT)

3. Comparação de Mineração de Dados e Mineração de Texto
4. Análise Exploratória de Texto
5. Etapas do KDT
6. Pré-Processamento de Texto
 - 6.1. Stopwords
 - 6.2. Stemming
 - 6.3. Dicionário ou Thesaurus
7. Tarefas do Processamento
8. Categorização, Extração da Informação, Clusterização, Sumarização
9. Agrupamento e classificação de textos.
10. Aplicações
 - 10.1. Análise de sentimento
 - 10.2. Mineração de opiniões
 - 10.3. Sistema de recomendação
 - 10.4.
11. Métricas de Avaliação
12. Estudos de Casos