



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

PROGRAMA DA DISCIPLINA	
IDENTIFICAÇÃO	
Nome: Mineração de Texto	Código: PGIA7345
Departamento/Unidade Acadêmica: Informática Aplicada	Área:
Carga horária total: 60h	Créditos: 4
Carga horária semanal: 4h	Teóricas: 30 Práticas: 30
Pré-requisitos:	
Có-requisitos:	
Semestre/ano de aplicação:	
EMENTA	
Processamento de Linguagem Natural (subtarefas). Modelos de Linguagens (n-gram, probabilísticos). Representação de documentos (esparsa e densa). Semântica Distribucional (Word Embedding). Classificação de documentos. Agrupamento de documentos. Extração de Informação. Recuperação de Informação. Introdução à Aprendizagem de Máquina Simbólica. Sumarização Automática de Documentos. Análise de Sentimentos. Repositórios Semânticos. Estudo/discussão de Artigos.	
BIBLIOGRAFIA	
BÁSICA 1. Dan Jurafsky e James H. Martin. Speech and Language Processing, 2nd Edition. Prentice Hall, 2008. 2. Christopher D. Manning, Hinrich Schütze. Foundations of Statistical Natural Language Processing (1st Edition). MIT Press, 1999. 3. Ricardo Baeza-Yates e Berthier Ribeiro-Neto. Recuperação de Informação. Conceitos e Tecnologia das Máquinas de Busca (2a. Edição). Editora Bookman, 2013. COMPLEMENTAR 1. Steven Bird, Ewan Klein, e Edward Loper. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media, 2009. 2. Peter Jackson e Isabelle Moulinier. Natural Language Processing for Online Applications: Text retrieval, extraction and categorization (2nd Edition),	

John Benjamins Publishing Company, 2007.

3. Torres-Moreno, Juan-Manuel. Automatic Text Summarization, John Wiley & Sons, 2014.

4. Michael W. Berry, Jacob Kogan. Text Mining: Applications and Theory. Wiley, 2010.

5. Ashok N. Srivastava, Mehran Sahami. Text Mining: Classification, Clustering, and Application. Chapman & Hall. CRC Data Mining and Knowledge Discovery Series, 2009.

Emitido em:

Responsável: