



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PÓS GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA APLICADA
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
Programa de Pós-Graduação: Informática Aplicada	Área: Computação
Carga horária total: 60h	Créditos: 4
Carga horária semanal: 4h	Teóricas: 30h Práticas: 30h
Pré-requisitos: Não há.	
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	
BIBLIOGRAFIA BÁSICA • Dan Jurafsky e James H. Martin. Speech and Language Processing, 2nd Edition. Prentice Hall, 2008. • Christopher D. Manning, Hinrich Schütze. Foundations of Statistical Natural Language Processing (1st Edition). MIT Press, 1999. • Ricardo Baeza-Yates e Berthier Ribeiro-Neto. Recuperação de Informação: Conceitos e Tecnologia das Máquinas de Busca (2a. Edição). Editora Bookman, 2013. BIBLIOGRAFIA COMPLEMENTAR • Steven Bird, Ewan Klein, e Edward Loper. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media, 2009. • Peter Jackson e Isabelle Moulinier. Natural Language Processing for Online Applications: Text retrieval, extraction and categorization (2nd Edition), John Benjamins Publishing Company, 2007. • Torres-Moreno, Juan-Manuel. Automatic Text Summarization, John Wiley & Sons, 2014. • Michael W. Berry, Jacob Kogan. Text Mining: Applications and Theory. Wiley, 2010.	

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