



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
COLEGIADO DO CURSO DE SISTEMAS DE INFORMAÇÃO

FICHA DE DISCIPLINA

DISCIPLINA: ORGANIZAÇÃO E RECUPERAÇÃO DA INFORMAÇÃO

CÓDIGO: GSI024		UNIDADE ACADÊMICA: FACULDADE DE COMPUTAÇÃO		
PERÍODO/SÉRIE: 5º		CH TOTAL TEÓRICA:	CH TOTAL PRÁTICA:	CH TOTAL:
OBRIGATÓRIA: (X)	OPTATIVA: ()	30	30	60
OBS:				
PRÉ-REQUISITOS: NÃO HÁ		CÓ-REQUISITOS: NÃO HÁ		

OBJETIVOS

Ao final do curso o aluno estará apto a propor soluções para o problema de recuperar informações nos documentos da coleção que melhor atendam às necessidades do usuário, a partir de uma consulta formulada pelo próprio usuário e de uma coleção documentos não estruturados ou semi-estruturados, como por exemplo uma base de dados em XML, de textos ou de imagens digitais. Além disso, o aluno será apto a avaliar a qualidade de sistemas de recuperação de informação por meio de medidas propostas na literatura.

EMENTA

Conceitos de documento, palavra e termo. Indexação de documentos: extração de termos, stopwords, stemming, criação de índices. *Thesauri*. Modelos de processamento de consultas. Avaliação de Sistemas de RI. RI em Documentos semi-estruturados, multimídia e documentos na Web. Extração da informação. Classificação de documentos. Redução de dimensionalidade.

DESCRIÇÃO DO PROGRAMA

1. Introdução à Recuperação de Informação e modelo booleano
2. Dicionário e lista de *postings* : conceitos de documento, palavra e termo.
3. Indexação de documentos: termos, *stopwords*, *stemming*, *Thesauri*
4. Compressão de índices
5. Peso de termos
6. Modelo Vetorial
7. Avaliação de sistemas de recuperação de informação
8. Realimentação de relevantes e expansão de consultas
9. Recuperação em documentos semi estruturados (XML)
10. Modelo Probabilístico
11. Classificação de documentos
12. Agrupamento de documentos
13. Redução de dimensionalidade
14. Web: busca, *crawling*, indexação, análise de *links*
15. Extração da informação
16. Introdução à Recuperação de Imagens baseada em conteúdo

BIBLIOGRAFIA

Básica:

MANNING, C.; RAGHAVAN, P.; SCHÜTZE, H. An Introduction to Information Retrieval, Cambridge, England: Cambridge University Press , 2009.

MANNING, C.; RAGHAVAN, P.; SCHÜTZE, H. An Introduction to Information Retrieval, Disponível em: <<http://nlp.stanford.edu/IR-book/information-retrieval-book.html>>. Acesso em Dezembro de 2009.

BAEZA-YATES, R.; RIBEIRO-NETO, B. Modern Information Retrieval, Addison-Welsey, 1999.

Complementar:

FRAKES, W. B.; BAEZA-YATES, R. (Ed.). Information Retrieval & Data Structures, Nova Jersey: Prentice Hall, 1992.

CRESTANI, F.; PASI, G. (Ed.). Soft Computing in Information Retrieval: Techniques and Applications, Springer Verlag, 2000

MOENS, M.-F. Information Extraction: Algorithms and Prospects in a Retrieval Context; Springer Verlag, 2006.

ZIVIANI, N. Projeto de Algoritmos com Implementações em Pascal e C. 2. ed., Thomson Editora, 004.

APROVAÇÃO

03, 11, 2011

Carimbo e assinatura do Coordenador do curso
Universidade Federal de Uberlândia
Prof. Dr. Daniel Gomes Mesquita
Coordenador do Curso de Sistemas de Informação
Portaria R N°. 1257/10

08, 11, 2011

Carimbo e assinatura do Diretor da
Unidade Acadêmica

Universidade Federal de Uberlândia
Prof. Ilmério Reis da Silva
Diretor da Faculdade de Computação
Portaria R N°. 757/11