

Programa de Disciplina Especialização em Ciência dos Dados

Módulo: II

Disciplina: Técnicas de Recuperação da Informação

Carga Horária: 15 horas (10h teóricas, 5h práticas)

Ofertante: Departamento de Computação e Sistemas – DECSI/UFOP

Objetivo:

Apresentar os principais conceitos e aplicações de recuperação da informação e desenvolver um sistema simples de recuperação de documentos.

Ementa:

Conceitos básicos e aplicações de recuperação da informação. Indexação e meta-dados. Modelos e implementação: modelo booleano, modelo vetor-espço, modelos probabilísticos e modelos de rede de inferência. Avaliação de resultados.

Conteúdo Programático:

1. Conceitos básicos e aplicações.
2. Indexação e meta-dados.
3. Modelos e implementação:
 - a) Modelo booleano.
 - b) Modelo vetor-espço.
 - c) Modelos probabilísticos.
 - d) Modelos de rede de inferência.
4. Avaliação de resultados.

Metodologia:

O conteúdo será introduzido através da metodologia de Ensino-Aprendizagem baseada em Problemas. Uma situação problema de recuperação da informação será introduzida pelo professor, onde o estudante deverá compreender o problema, criar meta-dados sobre os documentos originais e resolver os principais modelos de recuperação para consultas simples. Todas essas etapas serão guiadas pelo professor junto ao conhecimento teórico requerido.

Atividade Prática Proposta:

Os alunos deverão criar um pequeno índice de documentos de uma biblioteca. Tal índice contém as palavras-chave dos documentos dentre outras informações. Posteriormente, deverão executar consultas e avaliar os resultados retornados pelo mecanismo de recuperação de informação. As implementações do mecanismo de recuperação serão dadas pelo professor.

Softwares:

1. A linguagem de programação Python 3.7 (<www.python.org/downloads/release/python-372/>) será utilizada para apresentar a leitura e manipulação dos dados para recuperação de informação.

Bibliografia:

FRAKES, W. B.; BAEZA-YATES, R. *Information Retrieval Data Structures & Algorithms*. 1. ed. EUA: Prentice-Hall, 1992. 512 p. ISBN 978-0134638379.

MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, R. *Introduction to Information Retrieval*. EUA: Cambridge University Press, 2008. 482 p. ISBN 978-052186-571-5.

SINGHAL, A. Modern information retrieval: a brief overview. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, v. 24, n. 4, p. 35–43, 2001.