

DCCXXX – RECUPERAÇÃO DE INFORMAÇÃO

PRÉ-REQUISITO: DCC205 - Estruturas de Dados

EMENTA: Projeto, implementação, e avaliação de sistemas de recuperação de informação textual, com ênfase em máquinas de busca para a Web.

A – OBJETIVOS

Este curso objetiva investigar técnicas básicas e avançadas para a construção de sistemas de recuperação de informação textual. Ao final do curso, o aluno deverá conhecer em profundidade os principais componentes de uma máquina de busca para documentos Web, a saber: coletor, indexador e processador de consultas.

B – PROGRAMA

1. Coleta: políticas de coleta (seleção, revisitação, civilidade, paralelização); arquitetura de coleta; estruturas de dados
2. Indexação: representação e entendimento de documentos; estruturas de dados; compressão; particionamento; replicação
3. Processamento de consultas: representação e entendimento de consultas; recuperação de resultados candidatos; processamento eficiente; processamento distribuído
4. Modelos de ranqueamento: modelos independentes de consulta (modelos de qualidade, modelos adversariais); modelos dependentes de consulta (modelos léxicos, modelos estruturais, modelos semânticos); modelos contextuais (modelos de personalização, modelos de diversificação)
5. Aprendizado de ranqueamento: representação de instâncias; funções de perda; estratégias de otimização; estratégias de adaptação
6. Avaliação de ranqueamento: modelos de comportamento; projeto experimental; métricas de avaliação; avaliação offline; avaliação online

C – BIBLIOGRAFIA

1. [*Introduction to Information Retrieval*](#), by C. Manning, P. Raghavan, and H. Schütze. Cambridge University Press, 2008.
2. [*Search Engines: Information Retrieval in Practice*](#), by B. Croft, D. Metzler, and T. Strohman. Pearson, 2009.
3. [*Modern Information Retrieval*](#), by R. Baeza-Yates and B. Ribeiro-Neto. Addison-Wesley Professional, 2011.
4. [*Information Retrieval: Implementing and Evaluating Search Engines*](#), by S. Büttcher, C. Clarke, and G. Cormack. MIT Press, 2016.
5. [*Text Data Management: A Practical Introduction to Information Retrieval and Text Mining*](#), by C. Zhai and S. Massung. Morgan & Claypool, 2016.
6. [*Foundations and Trends in Information Retrieval*](#), by several authors
7. [*Synthesis Lectures on Information Concepts, Retrieval, and Services*](#), by several authors