



UNIVERSIDADE FEDERAL DE UBERLÂNDIA



FICHA DE COMPONENTE CURRICULAR

CÓDIGO: FACOM32605	COMPONENTE CURRICULAR: Organização e Recuperação da Informação	
UNIDADE ACADÊMICA OFERTANTE: Faculdade de Computação		SIGLA: FACOM
CH TOTAL TEÓRICA: 30 horas	CH TOTAL PRÁTICA: 30 horas	CH TOTAL: 60 horas

1. **OBJETIVOS**

Ao final do curso o aluno estará apto a propor soluções para o problema de recuperar informações (RI) 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 semiestruturados, como por exemplo uma base de dados em XML ou de textos. Além disso, o aluno será apto a avaliar a qualidade de sistemas de recuperação de informação por meio de medidas propostas na literatura.

2. **EMENTA**

Conceitos de documento, palavra e termo. Indexação de documentos: extração de termos, stopwords, stemming, criação de índices. Thesauri. Modelos clássicos de processamento de consultas. Avaliação de Sistemas de RI. RI em Documentos semiestruturados. Classificação de documentos. RI na Web.

3. **PROGRAMA**

1. Introdução à Recuperação da Informação
2. Modelos de RI: caracterização de um modelo, taxonomia de modelos de RI
3. Modelo Booleano
4. Ponderação de termos: TF, IDF, TF-IDF
5. Modelo Vetorial
6. Modelo Probabilístico
7. Modelos alternativos: BM25
8. Indexação de documentos: termos, stopwords, stemming, Thesauri
9. Avaliação de sistemas de recuperação de informação
10. Realimentação de relevantes e expansão de consultas
11. Recuperação em documentos semiestruturados (XML)
12. Classificação de documentos
13. Agrupamento de documentos

14. Web: busca, crawling, indexação, análise de links

15. PageRank e Aprendizado de Ranking

4. BIBLIOGRAFIA BÁSICA

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Recuperação de Informação: Conceitos e Tecnologia das Máquinas de Busca**. Bookman Editora, 2013.

BÜTTCHER, S.; CLARKE, C.; CORMACK, G. **Information retrieval: Implementing and evaluating search engines**. Cambridge: MIT Press, 2010.

MANNING, C.; RAGHAVAN, P.; SCHÜTZE, H. **Introduction to information retrieval**. England: Cambridge University Press, 2008.

5. BIBLIOGRAFIA COMPLEMENTAR

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern information retrieval**. 2nd ed. New York: Addison-Wesley, 2011.

CHOWDHURY, G.; **Introduction to modern information retrieval**. 2nd ed. London: Facet, 2004.

GROSSMAN, D.; FRIEDER, O.; **Information retrieval: Algorithms and heuristics**. Boston: Kluwer, 1998.

KOWALSKI, G.; **Information retrieval architecture and algorithms**. 2nd ed. London: Facet, 2004.

ZIVIANI, N. **Projeto de algoritmos: com implementações em Pascal e C**. Thomson. 3. ed., rev. e ampl. São Paulo: Cengage Learning, 2011.

6. APROVAÇÃO

Prof. Dr. Jefferson Rodrigo de Souza
Coordenador do Curso de Sistemas de Informação

Prof. Dr. Mauricio Cunha Escarpinati
Diretor da Faculdade de Computação



Documento assinado eletronicamente por **Jefferson Rodrigo de Souza, Presidente**, em 21/12/2021, às 12:21, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Mauricio Cunha Escarpinati, Diretor(a)**, em 01/02/2022, às 16:00, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3078450** e o código CRC **5415650F**.