



UNIVERSIDADE FEDERAL DE UBERLÂNDIA



## FICHA DE COMPONENTE CURRICULAR

<b>CÓDIGO:</b> FACOM33602	<b>COMPONENTE CURRICULAR:</b> Recuperação da Informação	
<b>UNIDADE ACADÊMICA OFERTANTE:</b> Faculdade de Computação		<b>SIGLA:</b> FACOM
<b>CH TOTAL TEÓRICA:</b> 30 horas	<b>CH TOTAL PRÁTICA:</b> 30 horas	<b>CH TOTAL:</b> 60 horas

1. **OBJETIVOS**

**Geral:** 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 de documentos não estruturados ou semiestruturados, 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.

**Específicos:**

- Ensinar a extrair de termos
- Entender a construção de índices e sua compressão
- Aprender o modelo Booleano e o modelo Vetorial
- Compreender expansão de consultas
- Analisar o modelo probabilístico
- Estudar como avaliar sistemas de RI
- Aplicar RI em documentos semi-estruturados

2. **EMENTA**

Introdução à Recuperação de Informação e modelo booleano; Indexação de documentos: termos, stopwords, stemização, Thesauri; Compressão de índices; Peso de termos; Modelo Vetorial; Modelo Probabilístico; Avaliação de sistemas de recuperação de informação; Realimentação de relevantes e expansão de consultas; Recuperação em documentos semi estruturados (XML); Tópicos especiais em RI.

3. **PROGRAMA**

1. Introdução à Recuperação de Informação e Modelo Booleano
  1. Conceitos de documento, palavra, stopword e termo

2. Tokenização, lematização e stemização
3. Dicionário e lista de postings
4. Matriz termo-documento
5. Índice invertido
6. Modelo Booleano
2. Indexação de documento:
  1. De blocos baseada em ordenação,
  2. De único passo em memória,
  3. Distribuída
  4. Dinâmica
3. Compressão de índices
  1. Leis de Heaps e de Zipf
  2. Compressão de dicionário: dicionário como uma string, blocking e front coding
  3. Compressão de postings: byte variável, unário e gamma
4. Peso de termos
  1. Recuperação rankeada, pontuação de documentos e frequência dos termos
  2. Esquemas de ponderação
5. Modelo Vetorial
  1. Pontuação de espaço vetorial
  2. Aumentando a velocidade da classificação espaço-vetor
6. Modelo Probabilístico
  1. Princípio de ranqueamento probabilístico
  2. BIM - modelo de independência binária
  3. Algoritmo BM25
7. Avaliação de sistemas de recuperação de informação
  1. Satisfação do usuário e outras medidas de avaliação
  2. Coleções de teste
  3. Apresentação de resultados
8. Realimentação de relevantes e expansão de consultas
  1. Expansão de consultas
  2. Feedback de relevância
  3. Centróide e algoritmo de Rocchio
  4. Tesouro
9. Recuperação em documentos semi estruturados (XML)
  1. Modelos
  2. Aplicações
10. Tópicos especiais em RI
  1. Web (busca, crawling, indexação, análise de links)
  2. Classificação e agrupamento de documentos

### 3. Recuperação da Informação em imagens.

#### 4. BIBLIOGRAFIA BÁSICA

BAEZA-YATES, Ricardo; RIBEIRO NETO, Berthier. **Modern information retrieval: the concepts and technology behind search**. 2nd. ed. New York: Addison Welsey, 2011.

FERNEDA, Edberto. **Introdução aos modelos computacionais de recuperação de informação**. Rio de Janeiro: Ciência Moderna, 2012.

MANNING, Christopher D.; RAGHAVAN, Prabhakar; SCHÜTZE, Hinrich. **Introduction to information retrieval**. New York: Cambridge University Press, 2008.

#### 5. BIBLIOGRAFIA COMPLEMENTAR

CRESTANI, Fabio; PASI, Gabriella (ed.). **Soft computing in information retrieval: techniques and applications**. Heidelberg: Physica, 2000.

ELMASRI, Ramez; NAVATHE, Shamkant B. **Sistemas de banco de dados**. 6. ed. São Paulo: Addison-Wesley, 2011.

FRAKES, William B.; BAEZA-YATES, Ricardo (ed.). **Information retrieval: data structures & algorithms**. Englewood Cliffs: Prentice Hall, 1992.

MAYBURY, M. T. **Multimedia information extraction : advances in video, audio, and imagery analysis for search, data mining, surveillance, and authoring**. [s. l.]: IEEE Computer Society, 2012. ISBN 9781118219546. Disponível em: <https://search.ebscohost.com/login.aspx?direct=true&db=cat08157a&AN=ieee.6381539&lang=pt-br&site=eds-live>. Acesso em: 4 nov. 2022.

MOENS, Marie F. **Information extraction: algorithms and prospects in a retrieval context**. Dordrecht: Springer, 2006.

#### 6. APROVAÇÃO

Prof. Dr. Rafael Dias Araújo  
Coordenador do Curso de Sistemas de Informação  
Campus Monte Carmelo

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



Documento assinado eletronicamente por **Rafael Dias Araújo, Coordenador(a)**, em 25/01/2023, às 22:13, 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 26/01/2023, às 17:18, 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](https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **4043900** e o código CRC **6FD58A92**.