

---

GSI024

# Organização e Recuperação da Informação

## Plano de Curso

Ilmério Reis da Silva

[ilmerio@facom.ufu.br](mailto:ilmerio@facom.ufu.br)

[www.facom.ufu.br/~ilmerio/ori](http://www.facom.ufu.br/~ilmerio/ori)

UFU/FACOM - 2012/1

Arquivo 0 – Plano de curso

---

# Motivação

---

*Sistemas de Informação que incluem o tratamento de dados não estruturados ou semi-estruturados e a interação com usuários que não conhecem a natureza dos dados.*

# Objetivo

---

*Tornar o aluno capaz de propor soluções para o problema de recuperar informações que melhor atendam às necessidades do usuário a partir de uma consulta e de uma coleção de 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.*

# Programa

---

- 1) Introdução à Recuperação de Informação e modelo booleano
  - 2) Dicionário e lista de *postings*: documento, palavra e termo.
  - 3) Indexação: 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
-

# Metodologia

---

- *Aulas expositivas*
- *Debates em sala de aula*
- *Exercícios dissertativos*
- *Exercícios de implementação em laboratório*
- *Seminários*

# Avaliação

---

- ***Itens e pontuação***
  - *Primeira prova, 26/04/2012 – 30 pontos*
  - *Segunda prova, 21/06/2012 – 40 pontos*
  - *Implementações em laboratório – 20 pontos*
  - *Seminário – 10 pontos*

**OBS:**

- 1. Todas as avaliações serão individuais.***
- 2. A avaliação do item seminário inclui preparação de material e participação nas discussões em sala de aula***

# Referências

---

## Bibliografia Básica:

[IIR] Manning, C., Raghavan, P., Schütze, H., *An Introduction to Information Retrieval*, Cambridge University Press, Cambridge, England, 2009, Disponível em: <http://nlp.stanford.edu/IR-book/information-retrieval-book.html>, Acesso em Dezembro de 2009.

### Livro texto

[MIR] Baeza-Yates, R. & Ribeiro-Neto, B., *Modern Information Retrieval*, Addison Wesley, 1999, 513pgs

# Referências

---

## Bibliografia Complementar:

- [IRDS] Frakes, W. B. & Baeza-Yates, R., Eds, *Information Retrieval & Data Structures*, Prentice Hall, NJ, 1992
- [ISR] Korfhage, R. R., *Information Storage and Retrieval*, John Wiley & Sons, NY, 1997
- [DW] Abiteboul, S., Buneman, P., e Suciu, D., *Data on the Web*, Morgan Kaufmann, San Francisco, 2000
- [SCIR] CRESTANI, F.; PASI, G. (Ed.). *Soft Computing in Information Retrieval: Techniques and Applications*, Springer Verlag, 2000
- [IE] MOENS, M.-F. *Information Extraction: Algorithms and Prospects in a Retrieval Context*; Springer Verlag, 2006.
- [PA] ZIVIANI, N. *Projeto de Algoritmos com Implementações em Pascal e C*. 2. ed., Thomson Editora, 004.
- [ART] Artigos, principalmente dos seguintes periódicos e conferências: ACM/TOIS, ACM/SIGIR, ACM/CIKM, JASIS, IEEE TKDE, Esevier/IPM.

---

*FIM*