



DISCIPLINA: Tópicos Especiais em Computação e Algoritmos: Recuperação de Informação	CÓDIGO:
--	---------

VALIDADE: Início: fevereiro de 2017

Eixo:

Carga Horária: Total: **60 horas-aula** Semanal: **4 aulas** Créditos: **4**

Modalidade: **Teórica** Integralização: **Optativa**

Classificação do Conteúdo pelas DCN: **Profissional**

Ementa:

Introdução e história da Recuperação da Informação. Arquitetura e política de coletores automáticos na Web, pré-processamento de dados coletados. Propriedades de documentos e compressão. Índices invertidos. Modelagem e processamento de consultas, *Web Based Models* (*PageRank* e *HITS*). Avaliação de sistemas de RI. Classificação e agrupamento de documentos.

Curso(s)	Período
Engenharia de Computação	5

Departamento/Coordenação: **Departamento de Computação - DECOM**

INTERDISCIPLINARIDADES

Pré-requisitos
Algoritmos e Estruturas de Dados II
Co-requisitos
--
Disciplinas para as quais é pré-requisito / co-requisito
--
Outras inter-relações desejáveis
--

Objetivos: <i>A disciplina deverá possibilitar ao estudante</i>	
1	Conhecer os principais componentes de uma máquina de busca
2	Representar e organizar documentos na Web
3	Realizar coleta e extração de dados na Web
4	Entender os principais critérios que fazem uma página na Web ser considerada de qualidade pelas máquinas de busca
5	Aprender as principais métricas de similaridade de documentos
6	Conhecer métodos automáticos de classificação de documentos

Competências:	
Essenciais	<ul style="list-style-type: none"> • Desenvolver coletores com o objetivo de extrair dados na Web • Desenvolver sistemas de busca na Web • Organizar e representar uma grande quantidade de documentos para extrair informação de maneira rápida • Avaliar a qualidade de métodos de recuperação de informação aprendidos • Desenvolver sistemas para classificação automática de documentos
Relacionadas	<ul style="list-style-type: none"> • Aplicar métodos de aprendizado de máquina em diferentes domínios

Unidades de ensino		Carga-horária (horas-aula)
1	Conceitos Básicos - Introdução e história da Recuperação de informação - Arquitetura de uma máquina de busca	4
2	Coleta: - Tipos de coletores e políticas de coletas - Arquitetura de um coletor	8
3	Preprocessamento - Formato de documentos na Web - Propriedade dos documentos e compressão de texto - Preprocessamento	8
4	Indexação: - Índices invertidos - Índices invertidos completos - Compressão de índice invertidos	8
5	Modelagem e processamento de consultas - Modelagem Booleana - Modelagem Vetorial - Modelagem probabilística: BM1, BM11, BM15 e BM25 - <i>Web based models</i> (<i>Page Rank</i> e HITS)	12
6	Avaliação de sistemas de Recuperação de Informação - Coleções de referência - Precisão e Revocação	6
7	Organização de documentos - Aprendizado de máquina supervisionado e não documentos - Aprendizado de Máquina não supervisionado para agrupamento de documentos	14
Total		60

Bibliografia Básica

1	Baeza-Yates, Ricardo; Ribeiro-Neto, Berthier. Modern information retrieval: the concepts and technology behind search . ACM Press, 2011.
2	Metzler, Donald; Strohman Trevor; Croft , W. Bruce. Search Engines: Information Retrieval in Practice . Pearson. 2010
3	Flach, Peter A. Machine learning : the art and science of algorithms that make sense of data . Cambridge University Press, 2012.

Bibliografia Complementar

1	Witten, I. H.; Moffat, Alistair; Bell, Timothy C. Managing gigabytes: compressing and indexing documents and images . 2nd ed. San Francisco: Morgan Kaufmann, 1999.
2	Liu, Bing. Web Data Mining – Exploring Hyperlinks, Contents and Usage Data . New York: Springer
3	Baeza-Yates, Ricardo e Ribeiro-Neto, Berthier. Recuperação de Informação – Conceitos e Tecnologia das Máquinas de Busca . Porto Alegre: Bookman – Grupo A, 2013.
4	Castelli, Vittorio; Bergman, Lawrence D. Image databases: retrieval of digital imagery . New York: J. Wiley, 2002.
5	Bishop, Christopher M. Pattern recognition and machine learning . New York: Springer, 2006.