

Informática e Tecnologias Multimédia

Licenciatura, 1º Ciclo

Plano: Despacho n.º 12419/2016 - 14/10/2016

Ficha da Unidade Curricular: Técnicas Avançadas de Programação

ECTS: 6; Horas - Totais: 162.0, Contacto e Tipologia, TP:30.0; PL:45.0; OT:5.0;

Ano | Semestre: 2 | S2

Tipo: Obrigatória; Interação: Presencial; Código: 814315

Área Científica: Informática

Docente Responsável

Ricardo Nuno Taborda Campos

Professor Adjunto

Docente(s)

Ricardo Nuno Taborda Campos

Professor Adjunto

Objetivos de Aprendizagem

O aluno deverá ser capaz de desenhar a arquitetura de um motor de busca (RI), explorar ferramentas de crawling e scraping, compreender as diferentes fases de processamento da linguagem natural e representação do texto, saber implementar um índice invertido, modelos de RI, e avaliação Cranfield

Objetivos de Aprendizagem (detalhado)

Esta unidade curricular tem por objectivo introduzir os alunos à recuperação de informação e motores de busca. Ao concluir esta unidade o estudante deverá:

- 1) ser capaz de desenhar a arquitetura de um motor de busca (RI);
- 2) saber explorar ferramentas de crawling e scraping;
- 3) compreender e implementar as diferentes fases de processamento da linguagem natural;
- 4) saber interpretar as características e estatísticas de um texto
- 5) saber representar um texto de acordo com as suas características mais relevantes;
- 6) ser capaz de implementar um índice invertido;

- 7) ser capaz de implementar modelos de recuperação de informação;
- 8) compreender a metodologia de avaliação Cranfield

Conteúdos Programáticos

- 1 - Introdução à Recuperação de Informação e Motores de Busca
- 2 - Aquisição de Dados
- 3 - Pre-processamento de Texto
- 4 - Estatísticas do Texto
- 5 - Representação de Texto
- 6 - Índices Invertidos
- 7 - Modelos de Recuperação de Informação
- 8 - Avaliação de Sistema de Recuperação de Informação

Conteúdos Programáticos (detalhado)

- 1 - Introdução à Recuperação de Informação e Motores de Busca
 - 1.1. Definição e objetivos
 - 1.2. Motores de busca
 - 1.3. Aplicações
 - 1.4. Dificuldades e desafios
 - 1.5. Arquitetura de um sistema de pesquisa de informação

- 2. Aquisição de Dados
 - 2.1 Definição e objetivos
 - 2.2 APIs
 - 2.3 Web Scraping
 - 2.4 Web Crawling
 - 2.5 Web Dynamics
 - 2.6 Arquivos da Web

- 3 - Pre-processamento de Texto
 - 3.1 Parsing
 - 3.2 Divisão de frases
 - 3.3 Tokenização
 - 3.4 Stopwords
 - 3.5 Normalização
 - 3.6 Stemming
 - 3.7 Reconhecimento de Entidades
 - 3.8 Part of Speech

- 4 - Estatísticas do Texto
 - 4.1 Lei de Zipf
 - 4.2 Impato na Recuperação de Informação
 - 4.3 Co-ocorrência de palavras

- 5 - Representação do Texto

- 5.1 Características importantes do Texto
- 5.2 Vocabulário controlado vs Vocabulário livre
- 5.3 Bag of Words
- 5.4 Modelo de Espaço Vetorial
- 5.5 Importância dos termos
- 5.6 Matriz Documento - Termo
- 5.7 Matriz Termo - Documento

6 - Índices Invertidos

- 6.1 Definição e objetivos
- 6.2 Estrutura de dados
- 6.3 Desafios
- 6.4 Big Data (Map Reduce)
- 6.5 Ferramentas de Indexação

7 - Modelos de Recuperação de Informação

- 7.1 Algoritmo de RI
- 7.2 Modelo Booleano
- 7.3 Modelo de Espaço Vetorial
- 7.4 Outros Modelos
- 7.5 Processamento de Queries

8 - Avaliação de Sistema de Recuperação de Informação

- 8.1 Noção de Relevância
- 8.1 Métodos de Avaliação
- 8.2 Coleções de Teste
- 8.3 Métricas de Avaliação

Metodologias de avaliação

Avaliação por frequência: Projeto Final (30%) + [Projeto I (20%) + Projeto II (10%) + Projeto III (10%)] + Frequência (30%)

Participação obrigatória em todos os elementos de avaliação (nota mínima de 6 valores em cada um)

Avaliação por exame: Projeto (50%) + Exame (50%)

Software utilizado em aula

Python - Anaconda

Jupyter Notebooks

PyCharm Community

Moodle: plataforma de eLearning do IPT, Centro de eLearning

Estágio

Não Aplicável

Bibliografia recomendada

- Baeza-Yates, R. e Ribeiro-Neto, B. (2010). *Modern Information Retrieval* (pp. 1-944).USA: Addison Wesley Longman Publishing Co. Inc.
- Croft, B. e Metzler, D. e Strohman, T. (0). *Search Engines: Information Retrieval in Practice* Acedido em 12 de fevereiro de 2019 em <http://ciir.cs.umass.edu/irbook/>

Coerência dos conteúdos programáticos com os objetivos

Os conteúdos programáticos estão em coerência com os objetivos da unidade curricular, atendendo a que:

- O ponto 1 dos conteúdos programáticos pretende concretizar o ponto 1 dos objetivos
- O ponto 2 dos conteúdos programáticos pretende concretizar o ponto 2 dos objetivos
- O ponto 3 dos conteúdos programáticos pretende concretizar o ponto 3 dos objetivos
- O ponto 4 dos conteúdos programáticos pretende concretizar o ponto 4 dos objetivos
- O ponto 5 dos conteúdos programáticos pretende concretizar o ponto 5 dos objetivos
- O ponto 6 dos conteúdos programáticos pretende concretizar o ponto 6 dos objetivos
- O ponto 7 dos conteúdos programáticos pretende concretizar o ponto 7 dos objetivos
- O ponto 8 dos conteúdos programáticos pretende concretizar o ponto 8 dos objetivos

Metodologias de ensino

Aulas teórico-práticas expositivas onde se descrevem os conceitos fundamentais.

Aulas práticas de resolução de casos práticos e aplicação dos conceitos a cenários de utilização real.

Coerência das metodologias de ensino com os objetivos

Os objetivos de aprendizagem do curso são atingidos através da realização de um conjunto de exercícios práticos permitindo desta forma que os alunos solidifiquem as competências adquiridas.

Língua de ensino

Português

Pré-requisitos

Preferencialmente o aluno deverá ter tido aproveitamento à UC de Linguagens de Programação

Programas Opcionais recomendados

Não aplicável.

Observações

Docente responsável
