

PLATAFORMAS DE DESENVOLVIMENTO WEB

LICENCIATURA EM INFORMÁTICA WEB

2022/2023

sebastiao@di.ubi.pt

APRESENTAÇÃO DA UNIDADE CURRICULAR

DESCRIÇÃO

- The Web is the most extensive collection of electronically accessible documents, making it the world's richest source of information. The problem with the Web is that this information is not well structured and organised so that it would be easily retrieved. Search engines help access web documents by keywords, but this is still far from what we need to use the knowledge available on the Web effectively. Machine Learning and Data Mining approaches go further and try to extract knowledge from the raw data available on the Web by organising web pages in well-defined structures or by looking into patterns of activities of Web users. These are the challenges in the area of Web Mining. This course focuses on extracting knowledge from the web by applying Machine Learning techniques to classify and cluster hypertext documents. Basic approaches from Information Retrieval and text analysis are also discussed. The students use recent Machine Learning and Data Mining software to implement practical web document retrieval, classification and clustering applications.

CONTEÚDOS PROGRAMÁTICOS

- Web Mining
- *Information Retrieval and Web Search*
 - *Crawling the Web*
 - *Indexing and keyword search*
 - *Document representation*
 - *Relevance Ranking*
 - *Vector space model (TF, IDF, TFIDF), Euclidian distance, cosine similarity*
 - *Relevance feedback*
 - *Advanced text search*
 - *Using the HTML structure in the keyword search*
 - *Evaluating search quality*
 - *Similarity search*

CONTEÚDOS PROGRAMÁTICOS

- *Hyperlink-Based Ranking*
- *Clustering approaches for Web Mining*
- ~~*Evaluating Clustering*~~
- *Classification approaches for Web Mining*
 - *General setting and evaluation techniques*
 - *Nearest Neighbor*
 - *Feature Selection*
 - *Bayesian approaches*
 - *Numeric approaches*
 - *Relational Learning*

OBJECTIVOS DE APRENDIZAGEM

- Introduce students to the basic concepts and techniques of Information Retrieval, Web Search, Data Mining, and Machine Learning for extracting knowledge from the web.
- Develop skills in using recent data mining software for solving practical problems of Web Mining.
- Gain experience in doing independent study and research.

DOCENTES

- Sebastião Pais: sebastiao@di.ubi.pt
 - Aulas teóricas
 - Aulas práticas
- Atendimento:
 - No final das aulas, sempre que se proporcionar
 - Por email a qualquer altura
 - Marcando uma hora para esclarecimentos

AVALIAÇÃO EM PERÍODO ENSINO-APRENDIZAGEM

Elemento de Avaliação	Peso (%) - MAX
Assessment Labs	Submissão Obrigatória
#01 – Web Crawler	30% - 6
#02 – Web Document Classification	30% - 6
#03 – Intelligent Search Engine	40% - 8

■ Assessment Labs:

- Não têm o mesmo peso na avaliação
- Individual ou em grupos de 2 elementos
- **Submissão obrigatória. Aceitam-se Assessment Labs submetidos fora do prazo, com penalização de 2 valores por cada dia de atraso**
- Sujeitos a defesa

AValiação em Época de Exames

Elemento de Avaliação	Peso (%) - MAX
Assessment Labs	Submissão Obrigatória
#01 – Web Crawler	30% - 6
#02 – Web Document Classification	30% - 6
#03 – Intelligent Search Engine	40% - 8

■ Assessment Labs:

- Não têm o mesmo peso na avaliação
- Individual ou em grupos de 2 elementos
- **Apenas serão aceites/avaliados Assessment Labs submetidos em período ensino-aprendizagem**
- Sujeitos a defesa

OUTRAS NOTAS SOBRE A AVALIAÇÃO

■ Assiduidade (Época Ensino-Aprendizagem):

■ **Todos os alunos estão admitidos à época de exames.**

■ Outras notas sobre a avaliação:

- As componentes da avaliação em época Ensino-Aprendizagem têm correspondência direta às componentes da avaliação nas épocas de Exame
- É portanto possível recuperar ou melhorar as componentes da avaliação da época ensino-aprendizagem nas épocas seguintes (Normal, Recurso e Especial)

PLANEAMENTO DAS AULAS

- 2 aulas semanais (blocos de 2 horas)
 - **Teóricas:** 2 horas
 - Apresentação dos conceitos e exercícios práticos
 - **Práticas:** 2 horas
 - Exploração de ferramentas e tecnologias
 - Início e acompanhamento dos Assessment Labs
 - Tutoriais
 - Apoio

INFORMAÇÃO E MATERIAL DE APOIO

- <http://www.di.ubi.pt/~sebastiao/PDW.html>
 - Classificações
 - Slides das aulas teóricas
 - Assessments Labs
 - Outras informações e material de apoio

BIBLIOGRAFIA

- Liu, Bing. Web data mining: exploring hyperlinks, contents, and usage data. Vol. 1. Berlin: Springer, 2011. [[WEB](#)]
- Zdravko Markov and Daniel T. Larose. Data Mining the Web: Uncovering Patterns in Web Content, Structure, and Usage, Wiley, 2007, ISBN: 978-0-471-66655-4.
- Eibe Frank, Mark A. Hall, and Ian H. Witten (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016. [[WEB](#)]
- DuBois, Paul. MySQL. Pearson Education, 2008.
- MySQL - Full-Text Search Functions. [[WEB](#)]
- Alexandre Pereira e Carlos Poupa, “Linguagens Web”, 5.ª ed. Lisboa: Sílabo, 2013. ISBN: 978-972-818-715-8.
- Tim Converse and Joyce Park, “PHP Bible”, 2nd Edition, Wiley, ISBN: 978-0-7645-4955-7.
- Leon Shklar and Rich Rosen, “Web Application Architecture: Principles, Protocols and Practices”, 2nd Edition, Wiley, 2009.
- Tutoriais da W3Schools: <http://www.w3schools.com>
- Alexis Goldstein, Louis Lazaris and Estelle Weyl, “HTML5 and CSS3 for the real world”, 2nd ed., SitePoint, 2015.