

PCI410082 – Gerência de Dados da Web para Ciência da Informação-2
créditos

Prof. **Thiago Magela Rodrigues Dias**

E-mail: **thiagomagela@cefetmg.br**

Carga Horária: **2 créditos**

Oferta: **2/2021**

Dias: **Quartas-feiras (8h as 12h)**

Ementa: Características dos dados da Web, Modelagem de dados semi-estruturados, Linguagens de consulta, Acesso a dados da Web, Integração de dados, Tratamento e análise de dados, Visualização de dados, Ambientes e aplicações da Web.

Objetivo: Compreender a natureza e características dos dados disponíveis na Web, modelar e realizar consultas em grandes repositórios de dados semi-estruturados, realizar coletas, integrações, tratamentos e visualizações de dados Web.

Conteúdo programático:

1. Natureza dos dados da Web

- 1.1 Conceitos básicos e terminologia
- 1.2 Caracterização dos dados da Web
- 1.3 Ambientes e aplicações que utilizam dados da Web
- 1.4 Dados estruturados, semi-estruturados e não-estruturados
- 1.5 Modelos associados

2. Acesso a dados da Web

- 2.1 Dados não-estruturados
- 2.2 Dados estruturados
- 2.3 Dados semi-estruturados
- 2.4 Ferramentas para consulta de dados

3. Extração de dados da Web

- 3.1 Visão geral
- 3.2 Métodos tradicionais
- 3.3 Ferramentas para extração

4. Integração de dados da Web

- 4.1 Abordagens para dados não-estruturados
- 4.2 Abordagens para dados estruturados
- 4.3 Abordagens para dados semi-estruturados

5. Tratamento e Análise de Dados

5.1 Técnicas para tratamento de dados

5.2 Estratégias computacionais para tratamento de dados

5.3 Ferramentas para tratamento de dados

6. Visualização de dados

6.1 Estratégias para visualização de dados

6.2 Data storytelling

6.3 Ferramentas para visualização de dados

Metodologia: Aulas expositivas e de discussão em grupo; seminários sobre temas específicos; palestras técnicas de convidados; trabalhos individuais.

Avaliação: A avaliação deste curso se dará em duas etapas:

Apresentação de seminários 30%.

Atividades práticas (listas) 30%.

Projeto prático 40%.

Cronograma:

Encontros	Data	Conteúdo
01 Aula expositiva	04/08	Natureza dos dados da Web Introdução Dados estruturados, semi-estruturados e não-estruturados
02 Aula expositiva	11/08	Acesso a dados da web Dados não estruturados, estruturados e sem estruturados Ferramentas de acesso a dados
03 Aula expositiva	18/08	Extração de dados da Web Visão geral
04 Aula expositiva	25/08	Extração de dados da Web Métodos e Técnicas

05 Aula expositiva	01/09	Extração de dados da Web Ferramentas para extração de dados
06 Aula expositiva	08/09	Integração de Dados Técnicas e ferramentas Seminário 01
07 Aula expositiva	15/09	Tratamento e análise de dados Técnicas para tratamento de dados Ferramentas para tratamento de dados
08 Aula expositiva	22/09	Visualização de dados Estratégias para visualização de dados Data storytelling Ferramentas para visualização de dados
ProjetoFinal		Entrega de Projeto da Disciplina

BibliografiaBásica

1. Abiteboul, S.; Buneman, P.; Suciu, D. Data on the Web: From Relations to Semistructured Data and XML. Morgan Kaufmann, 2000.
2. Abiteboul, S.; Manolescu, I.; Rigaux, P.; Rousset, M.-C.; Senellart, P. Web Data Management. Cambridge, 2012.
3. Christen, P. Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection. Springer, Berlin, 2012.
4. Liu, B. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data. Second Edition, Springer, Berlin, 2011.

BibliografiaComplementar

5. Chang, C.-H.; Kayed, M.; Girgis, M. R.; Shaalan, K. F. A Survey of Web Information Extraction Systems. IEEE Trans. Knowl. Data Eng., 18(10): 1411-1428, 2006.
6. Chebotko, A. Literature Survey on the Semantic Web and Search. Technical Report TR-DB-032005-C, Department of Computer Science, Wayne State University, 2005.
7. Florescu, D. Managing semi-structured data. ACM Queue, 3(8): 18-24, 2005.
8. Florescu, D.; Levy, A.; Mendelzon, A. Database Techniques for the World-Wide Web: a Survey. SIGMOD Record, 27(3):59-74, 1998.
9. Laender, A.H.F.; Ribeiro-Neto, B.A.; Silva, A.S.; Teixeira, J.S. A Brief Survey of Web Data Extraction Tools. SIGMOD Record, 31(2): 84-93, 2002.
10. Laender, A.H.F.; Silva, A.S. Cooperative Research on Web Data Management at UFMG and UFAM - A Brief Report. In Proc. of LA-WEB 2008, 144-150.
11. McCallum, A. Information Extraction: Distilling Structured Data from Unstructured Text. ACM Queue, 3(9): 48-57, 2005.
12. Sarawagi, S. Information Extraction. Foundations and Trends in Databases, 1(3): 261-377, 2008.
13. Artigos Técnicos de Periódicos e Anais de Congressos.