

Proposta de Disciplina Optativa

Disciplina: Tópicos Especiais em Gerência de Dados da Web (GDW)

Professor Responsável: Thiago Magela Rodrigues Dias

Nº de créditos: 04

Carga horária: 60 h (30 h teórica / 30 h prática)

Ementa:

Características dos dados da Web, Modelagem de dados semi-estruturados, Linguagens de consulta, Acesso a dados da Web, Integração de dados, Web Semântica, Ambientes e aplicações da 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-estururados
- 2.2 Dados estruturados
- 2.3 Dados semi-estruturados

3. Extração de dados da Web

- 3.1 Visão geral
- 3.2 Métodos tradicionais
- 3.3 Métodos baseados em árvores DOM
- 3.4 Métodos probabilísticos

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. Web Semântica

- 5.1 Conceitos básicos
- 5.2 RDF e RDF Schema
- 5.3 Linked Data
- 5.4 Consulta a dados RDF
- 5.5 Ontologias, QWL e lógica descritiva

Bibliografia Básica

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.

Bibliografia Complementar

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.