### DOGUS UNIVERSITY

### CSE 538 Web Search and Mining - Course Instruction Manual

|  |  |
| --- | --- |
| **Institute** | Institute of Science and Technology |
| **Department** | Computer Engineering |
| **Credits / Hour** | 3 (3+0+0) |
| **Course Type** | Elective |
| **Prerequisites** | - |
| **Lecturer** | Prof. Dr. Selim AKYOKUŞ (Room: G 912, Tel: (Ext) 1213)  Electronic mail: sakyokus@dogus.edu.tr |
| **The text book** | * Bing Liu, Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, Springer, 2011, ISBN: 3540378812 |
| **Recommended Texts** | * Mark Levene, An Introduction to Search Engines and Web Navigation, Pearson Education, 2010, ISBN 0321306775 * R. Baeza-Yates, B. Ribeiro-Neto.  [Modern Information Retrieval: the concepts and technology behind search.](http://www.mir2ed.org/) Addison-Wesley, 2011. * Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008. * Soumen Chakrabarti, Mining the Web: Discovering Knowledge from Hypertext Data, Morgan-Kaufmann Publishers, 2003, ISBN 1-55860-754-4 * Pierre Baldi,Paolo Frasconi, Padhraic Smyth, Modeling the Internet and the Web, John Wiley and Sons Ltd, 2003, ISBN 0470849061 |
| **Course Description** | Several aspects of web search engines and web mining. web crawlers, parsing, information retrieval basics such as indexing structures, relevance ranking algorithms, document similarity and clustering, specialized search engines, evaluation, natural language processing, and data mining applied to Web. |
| **Purpose** | The objective of this course is to provide students with an understanding of concepts and techniques associated with web search and mining, including concept, principle, architecture, design, implementation, application of web search and mining. This course also aims to enable students to discuss and critically evaluate the relative strengths and limitations of the different web search and mining methods and approaches, to implement and use some of the important web mining and web search algorithms, apply web mining and web search techniques to real-world web applications. |
| **Learning Outcomes** | The students passing the course will be able to   * understand how web search engines crawl, index, and rank web content, * have in-depth knowledge of the fundamental web mining concepts and techniques, * describe and utilize a range of techniques for web search and mining systems, * understand the functionality of the various web search and web mining methods and components, * appreciate the strengths and limitations of various web mining and web search models, * compare the various approaches to web mining and web search implementations, * apply web mining and web search techniques to real-world web applications, * gain hands-on experience by conducting a term project on designing and developing a data/web mining application, or performing an extensive analysis using data/web mining techniques, * Be able to effectively present and communicate the knowledge they have acquired in the course. |
| **Content of the Course** | This course involves web technologies, web crawling, structured data extraction: wrapper generation, information retrieval models: Boolean model, terms and postings lists, dictionary data structures, and tolerant retrieval, index construction and compression, scoring, term weighting, and vector space model, components of an ir system and performance evaluation of information retrieval systems, association rules and sequential patterns, supervised learning, unsupervised learning, social network analysis, opinion mining and sentiment analysis, web usage mining. |
| **Learning Methods** | A variety of teaching and learning methods are used including formal lectures and paper presentations, the use of several tools and software to gain practical experience of web search and mining systems. Additionally, students will prepare homework and projects on several related problems. |
| **Assessment** | Paper Presentations 05%  HWs 15%  Project 25%  Midterm 15%  Final examination 40% |

### Course Plan

|  |  |
| --- | --- |
| Week | Topics |
| 1 | Introduction to Web Search and Mining  Web Technologies |
| 2 | Web Crawling  Structured Data Extraction: Wrapper Generation |
| 3 | Information Retrieval Models: Boolean Model |
| 4 | The Terms and postings lists, Dictionary Data Structures, and Tolerant Retrieval. |
| 5 | Index Construction and Compression |
| 6 | Scoring, Term Weighting, and Vector Space Model |
| 7 | Components of an IR system and Performance Evaluation of Information Retrieval Systems |
| 8 | Association Rules and Sequential Patterns |
| 9 | Supervised Learning |
| 10 | Unsupervised Learning |
| 11 | Social Network Analysis |
| 12 | Opinion Mining and Sentiment Analysis |
| 13 | Web Usage Mining |
| 14 | Review |