

APPROVED BY

Director of Institute of Cybernetics
 D.M. Sonkin

Web Data Mining

Field of Study: 09.03.04 Software Engineering

Programme name: Big Data Solutions

Level of Study: Master Degree Programme

Year of admission: 2019

Semester, year: 3,2

ECTS: 6

Total Hours: 216

Contact Hours: 48

- **Lectures:** 24
- **Labs:** 24
- **Practical experience:**


Assessment: exam

Department: Software Engineering

Head of Department


V.S. Sherstnev

Instructor(s)


A.O. Savelev

Web Data Mining

Course Overview

Course Objectives	Course is aimed to formation of student's skills and abilities for professional web data mining.
Learning Outcomes	<p>As a result of mastering the discipline, the student must achieve the following results:</p> <ul style="list-style-type: none"> - Ability to extract information from open access web sources. - Ability to create algorithms for web content classification. - Ability to interpret the web data analysis results. - Understanding the basics of the following technologies: Web Usage Mining, Web Strackture Mining, Web Content Mining.
Course Outline	<ol style="list-style-type: none"> 1. Introduction to Web data mining. 2. Association Rules and Sequential Patterns. 3. Supervised Learning. 4. Unsupervised Learning. 5. Information Retrieval and Web Search. 6. Social Network Analysis. 7. Web Crawling. 8. Structured Data Extraction: Wrapper Generation. 9. Information Integration. 10. Opinion Mining and Sentiment Analysis. 11. Web Usage Mining.
Prerequisites (if available)	<ol style="list-style-type: none"> 1. Introduction to Big Data 2. Data Analysis Methods 3. Big Data Analytics
Course Structure	<p>Introduction to Web data mining. Basic terms and definitions. Web Data Analysis Toolkits description.</p> <p>Association Rules and Sequential Patterns. Objectioves of association rules. Mining of association rules. Main algorithms.</p> <p>Supervised Learning. Overview. Training examples preparation. Problems of supervised learning.</p> <p>Unsupervised Learning. Overview. Main methods. Problems of unsupervised learning.</p> <p>Information Retrieval and Web Search. Information retrieval and search engines. Open-source systems.</p> <p>Social Network Analysis. History. Metrics. Social networking potential. Main SNA methods .</p>

	<p>Web Crawling. Crawling, indexing and data compression</p> <p>Structured Data Extraction: Wrapper Generation. Web Structure Mining. The main purpose for structure mining. Main methods and algorithms.</p> <p>Information Integration. Merging of information from heterogeneous sources.Data consolidation task. Technologies.</p> <p>Opinion Mining and Sentiment Anlaysis. Introduction. Applications. General Challenges. Approaches.</p> <p>Web Usage Mining. User behavior analysis and data analytics.</p>
<i>Facilities and Equipment</i>	<ul style="list-style-type: none"> - Lecture room (computer, projector) – 634034 Tomsk Region, Tomsk, Sovetskaya Street, 84/3,room 313. - Computer class (10 computers) – 634034 Tomsk Region, Tomsk, Sovetskaya Street, 84/3, room. 204.
<i>Grading Policy</i>	<p>In accordance with TPU rating system we use:</p> <ul style="list-style-type: none"> - Current assessment which is performed on a regular basis during the semester by scoring the quality of mastering of theoretical material and the results of practical activities (performance tests, perform tasks, problem solving). Max score for current assessment is 60 points, min – 40 points. - Course final assessment (exam/ credit test) is performed at the end of the semester. Max score for course final assessment is 40 points, min – 22 points. <p>The final rating is determined by summing the points of the current assessment during the semester and exam (credit test) scores at the end of the semester. Maximum overall rating corresponds to 100 points, min pass score is 65.</p>
<i>Course Policy</i>	<p>Class attendance will be taken into consideration when evaluating students’ participation in the course.</p> <p>Students are expected to actively engage in class discussions about the assigned readings.</p>
<i>Teaching Aids and Resources</i>	<p>Compulsory Readings:</p> <ol style="list-style-type: none"> 1. Piwik Web Analytics Essentials [Electronic resource]. - Access mode: http://143.95.72.211/ebooks/Piwik%20Web%20Analytics%20Essentials.pdf, free. 2. Web Data Mining Exploring Hyperlinks, Contents, and Usage Data [Electronic resource]. - Access mode: https://link.springer.com/content/pdf/10.1007%2F978-3-642-19460-3.pdf, free 3. Mining the Social Web [Electronic resource]. - Access mode:

	<p>http://www.webpages.uidaho.edu/~stevel/504/Mining-the-Social-Web-2nd-Edition.pdf, free</p> <p>Additional Readings:</p> <p>4. Advanced Web Metrics withGoogle Analytics [Electronic resource]. - Access mode:https://gentecomgente.files.wordpress.com/2013/10/markiert_brian_clifton_advanced_web_metrics_with_google_analytics__2010.pdf, free</p>
<i>Instructor (-s)</i>	<p>Savelev Aleksey sava@tpu.ru 8 909 540 63 78</p>