



Web-mining and data analysis

10 ECTS – 1 Semester – Autumn

The course will give an overview of current technology for collection, and analysis of web information. The course is closely related to the EIAO project and each student group participating in this course will carry out a project to implement some techniques related to current research issues in the field.

The course is given in English at Agder University College, Grimstad.

Related applications for the techniques taught

Browsers, Search Engines, Assistive Technologies, Crawling, Pattern recognition, Data Mining, Reinforced Learning, Clustering.

The course will use Open Standards like (X)HTML, XML and HTTP. Pattern recognition and statistical techniques for analysis of data will be covered. Such techniques are e.g. used for decision making in spam-filters, intrusion detection systems, as well as in various Internet content filters. Structuring of information, e.g. raw data combined with analysis results, in ontologies will also be dealt with. Also included are Web standards, Quality of Service (accessibility), web retrieval and analysis techniques.

You will learn

The course conveys a practical understanding of technologies for the retrieval of information from the Internet such as browsers and crawlers. Students will become familiar with an application for retrieving information from the web and storing the information in a database and intrusion detection systems.

The course includes techniques for automatic retrieving and processing of information from the Internet. Accessibility of the Internet content is mainly decided in the form of the content (HTML-code, colours, pictures, animations, etc.). In fact, the form of the content of Internet can determine to what extent people with disabilities can utilise the information. The form of the information also decides what kind of terminals can be used, like mobile phones or braille readers. Understanding and assessing accessibility of the Internet content is therefore important, and is the main purpose of the student projects given in this course.



Required skills to attend

The course is designed for the ICT master level, and is open for external candidates. Participants are expected to have a working knowledge of software programming and to be familiar with basic concepts of the Web. Additionally, basic statistical, such as probability theory, knowledge is required.

Organisation of the course

The first part of the course is organised around lectures and some short presentations given by the students to outline their projects.

The second part of the course is based on student projects where the students regularly meet with a mentoring team.

25% of the grade will be based on a written exam, while 75% of the grade will be based on the student projects.

The students will choose from a wide variety of projects related to web-mining and data analysis. Several of the projects will be related to AI work ORTS, a free software RTS game engine, with the possibility of participating in an online competition.

Guest lecturers

Jaran Nilsen from Integrasco – Software agents

More information

For more information about this course, including earlier projects and lecture material, visit:

<http://eiao.net/webmining>.

Related material can be found at:

- European Internet Accessibility Observatory - <http://eiao.net>
- Web Crawling - http://en.wikipedia.org/wiki/Web_crawler
- Pattern recognition - http://en.wikipedia.org/wiki/Pattern_recognition
- Bayesian network - http://en.wikipedia.org/wiki/Bayesian_network.
- Machine Learning - http://en.wikipedia.org/wiki/Machine_learning
- ORTS - <http://www.cs.ualberta.ca/~mburo/orts/>

Lectures:

Ole-Christoffer Granmo – Phone +47 37 25 33 57– E-mail ole.granmo@hia.no

Morten Goodwin Olsen – Phone +47 37 25 33 24– E-mail morten.g.olsen@hia.no

Student assistant:

Anis Yazidi – E-mail anisy06@student.hia.no

