



IKT407 Web-mining and data analysis

Ole-Christoffer Granmo - Morten Goodwin Olsen



What will you learn?

- **Pattern recognition**
 - How can I use basic statistics to recognise a pattern (such as patterns in the stock market – if a given stock will increase or decrease in value).
- **Crawling / Resource allocation**
 - How is it possible to search the almost infinite world wide web in less than a second with a search engine such as Google.
- **Text Mining / Classification**
 - How can I classify text on a web site into e.g. relevant / not relevant information for me.
 - How can I classify text on a web site into e.g. English / Chinese / German



What will you learn?

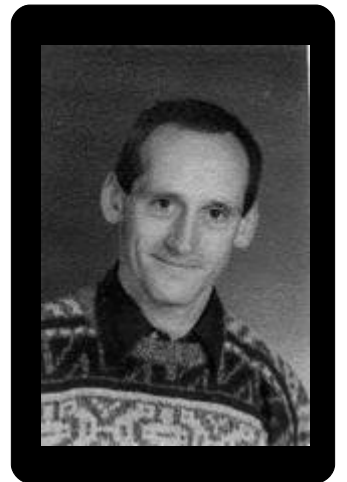
- **Reinforced Learning**
 - How can I use learning techniques to learn which web sites have been updated and focus the crawling on these web sites?
 - How can a I find the optimal action for an AI in a real-time strategy game?
- **Software agents**
 - How can I create software agents such as agents for fixing unstructured (X)HTML to understandable material?
- **Content versus Representation**
 - What is content / representation?
 - How is it presented with CSS / CMS and what Eclipse tools can be used for this?

Who is part of the web-mining course?



- Ole-Christoffer Granmo
 - Pattern Classification and Reinforced learning

- Andreas Prinz
 - Content versus representation



Who is part of the web-mining course?

- Jaran Nilsen (Integrasco)
 - Software agents
- Morten Goodwin Olsen
 - Web and crawling
- Anis Yazidi
 - Student assistant
(not confirmed yet)



Lecture schedule

- 2007-08-23 (Thursday) – Web and Crawling
- 2007-08-30 (Thursday) – Pattern classification
- 2007-09-06 (Thursday) – Reinforced learning
- 2007-09-11 (Tuesday) – Content versus representation
- 2007-09-18 (Tuesday) – Software Agents

Deliverables

- 2007-09-20 – Choose project
- 2007-10-04 – Exam (25% of your grade)
- 2007-10-18 – Motivation and planned experiments
- 2007-11-08 – Experiment results
- 2007-12-30 – Final project report
- 2007-12-07 - Presentation

Why should I choose web-mining?

- The web is, and will be in the future, one of the most important knowledge areas for anyone.
- You will get an introduction to a lot of areas related to web-mining and data analysis.
- You will work in groups and focus on areas interesting for you.
- Related to state of the art research such as the EIAO project (<http://eiao.net>).

Questions?

- Please direct any questions to ether:
 - Ole-Christoffer Granmo - ole.granmo@hia.no
or
 - Morten Goodwin Olsen - morten.g.olsen@hia.no
- For more information see:
 - <http://eiao.net/webmining>
 - Videos of past lectures
 - Previous projects