Course Motivation

How does Google work?
Course Motivation

How does Google work?

How do search engines work?
Course Motivation

How does Google work?

How do search engines work?

Algorithms for web indexing and searching
Course Outline

1. Introduction to Course
2. General Anatomy of Web Search Engines
3. Building blocks of Search Engines
   (a) Web Crawlers
       - Anatomy of crawlers
       - Crawling strategy
   (b) Index
       - Inverted files
       - Suffix trees
       - Signature files
       - Compression
       - Issues of efficient construction
       - Duplicate removal
(c) Types of Queries

(d) Ranking

- Text based methods
  - Vector based methods
  - Latent semantic indexing
- Link based methods
  - PageRank
  - HITS
  - SALSA
  - Others
Course Outline

4. Further topics
   (a) Clustering
   (b) Automatic Categorization/Hierarchy Building
   (c) Evaluation of search engines
   (d) Structure of and Models for the Web Graph
   (e) Data Mining
<table>
<thead>
<tr>
<th><strong>Prerequisites:</strong></th>
<th>dADS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Literature:</strong></td>
<td>Handouts</td>
</tr>
<tr>
<td><strong>Course language:</strong></td>
<td>Danish or English</td>
</tr>
<tr>
<td><strong>Credits:</strong></td>
<td>2 points/10 ECTS</td>
</tr>
<tr>
<td><strong>Evaluation:</strong></td>
<td>Programming project</td>
</tr>
</tbody>
</table>

**Course page:**

http://www.daimi.au.dk/~gerth/webalg02/index.html
Programming Project

Implement a Web Search Engine
Programming Project

Implement a Web Search Engine

Distributed project
Groups (2–4 persons) doing:

- Web crawling
- Index building
- Ranking
- Query interface

Start: index Aarhus University website
Goal: index domain .dk