

Search Engines

WS 2009 / 2010

Thursday February 11th, 2010
(Bachelor / Master Projects)

Prof. Dr. Hannah Bast
Chair of Algorithms and Data Structures
Department of Computer Science
University of Freiburg

Bachelor / Master Projects

■ ECTS Points / Working time

- Bachelor project
 - 6 ECTS points = 6 weeks a 30 hours each
- Master's (Team) project
 - 16 ECTS points = 16 weeks a 30 hours each
- Any other kinds of projects here?

Praktikum : 6 ECTS points

■ Phases

- You come up with a cool idea for a project
- We have a joined session with presentations of the ideas
 - where you get feedback from us and the others
- You write a detailed proposal of your project
 - more about that on the next slide
 - includes formulation of *milestones* and *goals*
- YOU START WORKING
- You meet with us at the scheduled milestones
 - and whenever questions / problems arise
- When you have reached your goals you are done

Project Proposal

- A good proposal contains
 - a short, easy to read description of the project
 - a very clear formulation of the **goals** of the project; all projects should have as a goal:
 - software (**well-documented and working!**)
 - a short document describing what you did
 - a detailed work plan at the granularity of days
 - with a number of milestones

**We will spend at least a week on the proposal
it's the basis of everything else to come**

- You should write proper software
 - well modularized, with proper specification of every class
 - every class, method, and member variable documented
 - unit tests and performance tests
 - strict adherence to a stylesheet (your code must pass through a lint-like program with 0 errors / warnings)
 - proper packaging, so that others can easily use it too

**The quality of the software is as important
as the quality of the product you realize**

The Projects

- The most important thing is
 - that you like what you are doing
 - and find the goal worthwhile
 - you are free to define a project completely on your own
 - but we also have some suggestions
 - examples on next slide

Possible Project: Article Search

- Build a system with the following features
 - whenever I read a paper / article, I can easily add it to my "database"
 - everything I have read is instantly searchable
 - articles are automatically associated with their metadata
 - authors
 - year of publication
 - conference / journal where it appeared

(there are databases with all the metadata, the task is to associate an article with the right piece of metadata quickly and reliably)

- anything else that might be useful in this context

Possible Project: Desktop Search

- Build a system with the following features
 - maintains an index of all the files on your disk
 - instantly searchable via all kinds of criteria
 - challenges
 - get the ranking right!
 - incremental index (can't rebuild the whole index every time something changes)
 - getting proper notification when something changes somewhere
 - I would prefer a system that runs on a Windows machine (because most people have that)

Possible Project: Mail Search

- Build a system with the following features
 - instant full-text search in all your personal mail
 - challenges
 - new mails should be instantly searchable
(so, again, incremental indexing needed)
 - should scale up to 10 GB and more
 - consider the specifics of searching in mail
 - different encodings
 - lots of contents repeated by quoting etc.
 - mails grouped into threads

