T-61.271 Visualization project description
16. Visualization of textual relations

Autumn 2004

Read the project description carefully. Questions and comments can be sent to t61271@mail.cis.hut.fi.

1 General information

The purpose of the project work is to get hands on experience and deeper understanding on one specific area of the course. At the same time you will be forced to think on some of the more general issues that are confronted every time a visualization method is developed or used.

To complete the project you will probably need information that has not been discussed on the lectures or in the exercises. Most of the time relevant information can be easily found on the web, course books or in some cases references to articles will be given. If you have problems finding information on the subject please contact the lecturer or the assistant for hints and guidance.

The project work will be valid for one year after the original deadline.

2 General requirements

- The project should be done by one person. However, discussing it with others is encouraged.
- A short plan on how you will complete the project (programs used, data, methods etc.) should be submitted by 23 October 2004.
- The deadline for the project is 17 January 2005. The reports submitted after the deadline will be rejected. If you have a very good reason that causes you to miss the deadline you can request an extension. The extension must be requested before the deadline.
- To pass the project you have to at least fulfill the tasks given in the specific requirements section.
- To pass with distinction you have to do additional work or the specified tasks should be exceptionally well done.

- If you submit the visualization project in time, but don't pass, you will have a chance to supplement your work after the deadline. The supplemented project works will not be eligible for the higher grade, "passed with distinction".
- If the project requires programming the source code must be included in the project report and it should preferrably be licensed under a license that allows its use and modification freely. Examples of such licenses are GNU General Public License and BSD license, see:
  `http://www.opensource.org/licenses/index.php`
- All program code should be submitted in machine readable format. and if not portable it should have a binary for Windows or Linux included.
- Documentation should be in a format readable with generally available tools. Preferred formats are Postscript and PDF.
- Please note that we need to be able to open and print the documents you send us and run your program code. Please take this into account when planning your project and ask us, if uncertain. You should contact us especially if accessing your project work will require some proprietary software.
- The project report should explain the methods used. It should contain e users guide if the project implements a program and your assessment how well the goals of the project were met.
- Each project report should contain a section that comments on the difficulty of the project and an estimate of the time used for completing it.
- The project reports should contain your email address and full student number. If you submit your work by email, please include the student number also to the subject line.

3 specific requirements

The goal of this project is to create a tool to visualize relations between different parts (chapters, subchapters etc.) of a document. Another option is to visualize the mailbox containing email messages, usenet articles, or similar messages.

The tool should have the following features

1. It should read a structured text file, for example the format could be Latex, HTML, XML, a Unix mbox file etc.
2. It should analyze the content of the text (and possibly the header information) using a suitable statistical method.
3. It should visualize the relations in the document using methods described in Card et al p. 413).

4 Submitting the project work

The project work should be submitted by email to:
`t61271@mail.cis.hut.fi`.