

Proposed 10/22/08

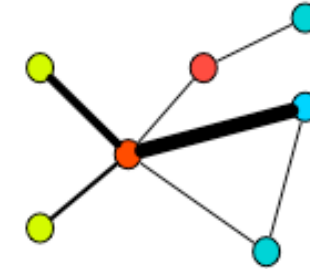
Text Analysis for eParticipation data

Motivation

The manual analysis of eParticipation data is a time consuming process and requires persons that are experts in specific domains. In order to assist these experts, different automated techniques gain more and more importance. Examples are: unsupervised learning/clustering, understanding semantic relations, Word Sense Disambiguation, visualization of high dimensional data..

Project ideas

- Visualization of semantic networks generated on eParticipation documents
- Application of machine learning techniques for text analysis
- Unsupervised clustering of text data
- Lexical analysis of German text
- Automatic generation of ontologies based on eParticipation data
- Evolution of text concepts over time (discussions, blogs)
- Application of text classification tools to security related areas: network traffic, malicious code, event correlation



Some links...

- Visualization: <http://apps.egiz.gv.at/Mitmachen/bin/DocumentVisualizerNew.swf>
- eParticipation: <http://www.mitmachen.at>
- Lexical Parser: <http://nlp.stanford.edu/software/lex-parser.shtml>
- Visualization library: <http://prefuse.org/>
- LingPipe: <http://alias-i.com/lingpipe/>
- Security Related: <http://beatrix.sourceforge.net/>
<http://sourceforge.net/projects/infectclass/>

Master Project

Studies: INF SEW TEL

Prerequisites

- Interested in machine learning
- Java

Advisor / contact

peter.teufl@egiz.gv.at