

# DD2476 Search Engines and Information Retrieval Systems

## Project 4: Wikipedia Information

Contact: Magnus Rosell, FOI ([magnus.rosell@foi.se](mailto:magnus.rosell@foi.se))

*This project is worth 3 ECTS credits. This means that it is expected to require 80 hours of work for each person in the group. The project formulation, method, and results are presented in a report as well as in a poster session. For more details, look at the course homepage, under Project in the menu.*

### Assignment

Build a tool that gives context information about a search string using Wikipedia. The idea is to use this tool within a system with a lot of entities that the user may not know anything about.

The system should receive a search string and return a few informative sentences/lines about that subject. Use Wikipedia as the source and retrieve the  $n$  best matching lines and/or documents. Use some kind of text summarization and present the  $k$  best of these. The first version of the summarization is of course to simply present the  $k$  highest weighted sentences. Feel free to expand upon this main idea.

### About FOI

The Swedish Defence Research Agency (FOI) is one of Europe's leading research institutes in the areas of defence and security. We have 1 000 highly skilled employees with various backgrounds. At FOI, you will find everything from physicists, chemists, engineers, social scientists, mathematicians and philosophers to lawyers, economists and IT technicians. FOI's core activities are research, methodology/technology development, analyses and studies. FOI is an assignment-based authority under the Ministry of Defence.

FOI develops smart solutions ranging from data gathering to innovative presentation systems. The research undertaken at the "Department of decision support systems and information fusion" increases the user's capability to understand and convert available information into effective decisions. FOI's extensive breadth of competence and well-developed contact network, both national and international, gives the agency a unique capability for solving problems in many different areas of expertise.

The research work at our department covers both technology and methods and bridges the divide between purely technical research and, for example, sensor and communications systems and human decision-making. The ultimate aim is to develop improved command and control capabilities, primarily in the defence environment but

not only here. Our particular strength lies in our ability, in concert with others, to take an overall perspective in the development of support tools for command and control and for decision-making processes in complex environments. Our customers include the Swedish Armed Forces and the EU but also Vinnova and the Swedish Environmental Protection Agency.