

# DD2476 Search Engines and Information Retrieval Systems

## Project 7: Cross-Cultural Media Monitoring

Contact: Jussi Karlgren, Gavagai ([jussi@kth.se](mailto:jussi@kth.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.*

### Problem

There are topics which are treated differently in different cultural areas, as well as there are topics which are treated similarly. Trying to unearth these differences is interesting for many reasons! Some of these differences may have interesting cultural implications, others may have commercial power.

### Assignment

Build a search-oriented monitor which

1. takes a topical query in Swedish or English (pick either) from the operator,
2. translates the query to some set of languages (determined by the language competence of the project group and the API given below),
3. send the query to a (number of) search engine(s),
4. retrieve resulting texts from the top of the ranked list,
5. process the texts to find the most topical sentences wrt the query,
6. run those sentences through the tonality API at Gavagai (<http://developer.gavagai.se>)
7. display tonality in the various languages wrt the query,
8. in a neat side-by-side interface, and, finally,
9. write a press release on interesting cross-cultural observations you can make from using the test set.

You need to be able to sanity check the retrieved texts, so you'll need to have access to someone in the project group who is able to read the languages you are working with!