Visualisation of Corpus Query Results

Henrik Dittmann

European Academy Bozen/Bolzano, Italy

CLARIN EDC and Federated Search Workshop,
MPI for Psycholinguistics, Nijmegen,
9 May 2011

Language corpora (text)

- grow in size
- grow in complexity:
  annotations, metadata, federated search

Visualisations

- present information
- aid discovery and decision making
- provide a way to interact with language data
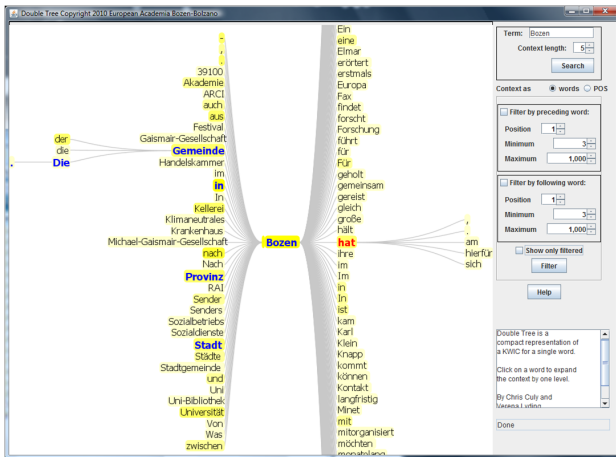
to structure the information and aid exploration

Structured Parallel Coordinates
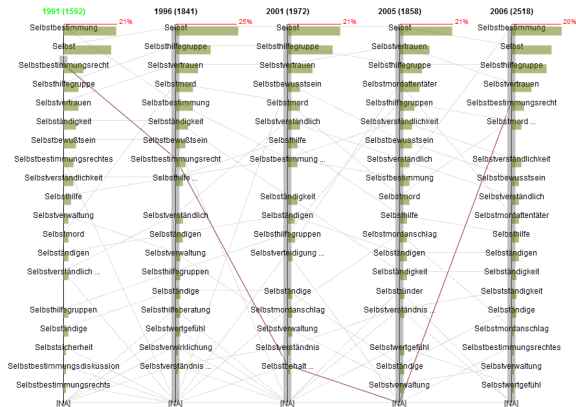
to provide means for interaction and aid in specifying finer-grained queries

Aggregation in this talk refers not to the aggregation of search results from different resources, but to the aggregation of individual hits from one or several queries.

Faceted search

- For the time series application we retrieve a list of tuples (word, year) from our CQP web service that we aggregate in the client application:
  ```
  "Selbst.*"; tabulate Last match
  .. matchend word, match year;
  ```
- Other examples for metadata to group the results by: author, text genre, corpus (in federated search), etc.
- in SRU 2.0 draft: *faceted search* facility

- aggregation over more than one metadata property
- fine-grained control over the text elements over which the aggregation is done
- aggregation over queries (if the query language is not powerful enough to adequately describe the phenomenon under investigation)

Interactivity

An interactive visualisation should be able to get back to the individual results from the aggregation.

Architectural options

- to have a client talk to a service that exposes high-level aggregation as well as capable linguistic corpus search
- to have a separate service that is based on a database with high-level aggregation and possibly statistical functions which receives its data from query results to a separate corpus query service
- to do all aggregation in the client

Thank you!

henrik.dittmann@eurac.edu

http://www.eurac.edu/linfovis