



# Content2Knowledge (C2K)

One of the biggest challenges in computational topic modeling is to identify context-relevant clusters from unstructured textual data. The joint research project *CONTENT2KNOWLEDGE (C2K)* of the University of Turku and the Åbo Akademi University seeks to bridge this gap between by integrating traditional topic-modeling with network analysis in order to develop innovative methods and computational tools for context-aware topic mining, content analysis and knowledge visualisation.

## PROJECT DESCRIPTION:

The **C2K project** started in 2014 as a spin-off from a small-scale project dealing with problems of extracting research data from semi-sized, unstructured text documents.

However, the project rapidly expanded to a larger project seeking to develop computational tools supporting context-aware, exploratory topic modeling, content analysis and knowledge visualisation.

The project seeks to integrate and elaborate traditional text-mining and content analysis methods and network analysis in an innovative way. The main aim is to develop methods helpful for exploring, analysing, structuring and modeling unstructured BigText corpora.

The methodological approach is built around an understanding of **texts as networks**. This approach has several positive effects:

- Identification of thematic paths in a document corpus
- Application of powerful analytical tools (clustering, community detection etc.) for topic modeling
- Creative methods for data and knowledge visualisations

## BASE PROJECT: GERMAN POLITICAL LANGUAGE

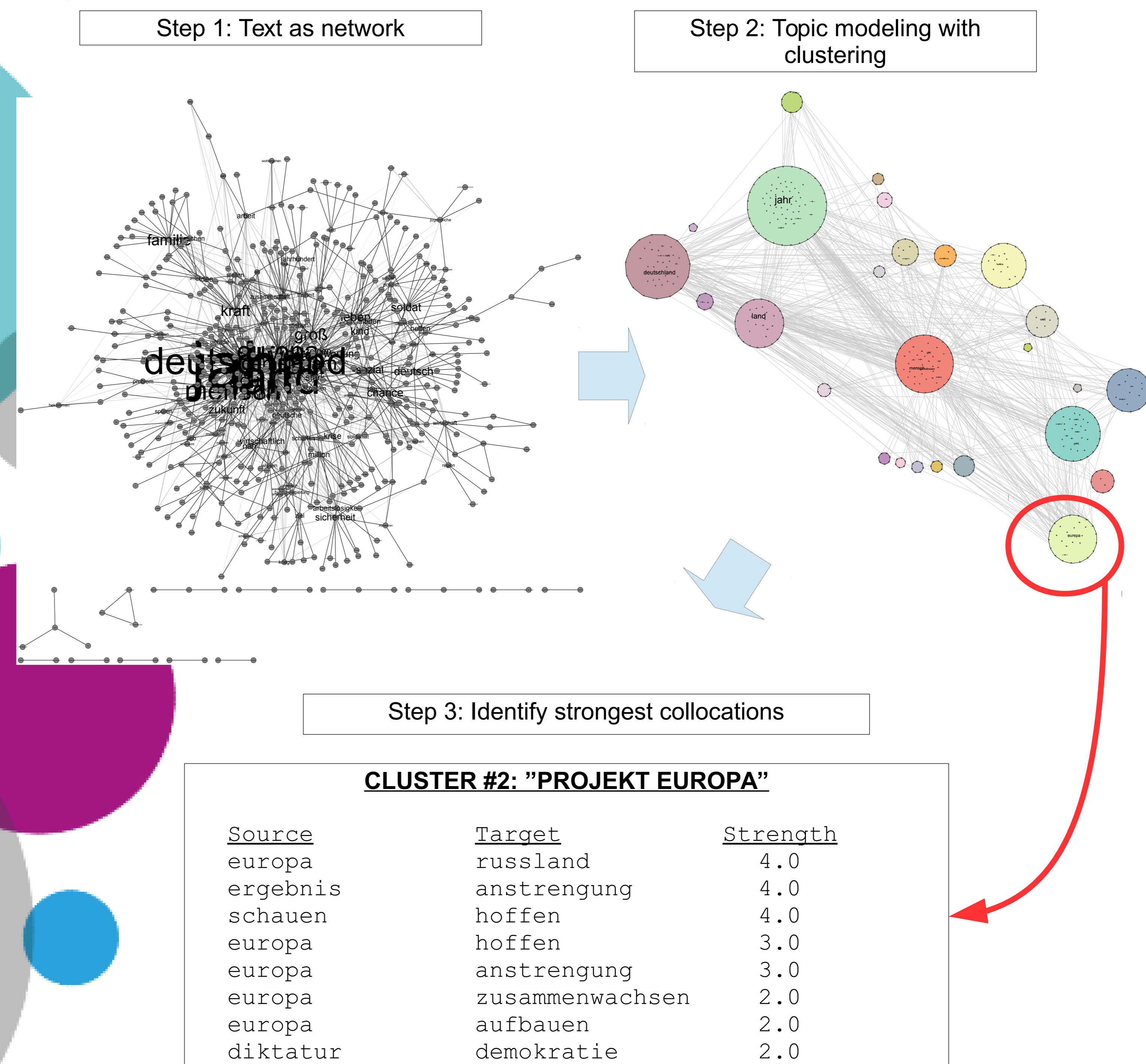
### WHAT WE WANT?

- Development of core political vocabulary for the three German-speaking countries (Germany, Switzerland, Austria)
- Development of innovative methods and tools for (semi-)automated document classification, topic modeling and content analysis through integration of traditional text-mining methods with network analysis
- Innovative tools and methods for knowledge visualisation based on the fundamental understanding of texts as networks

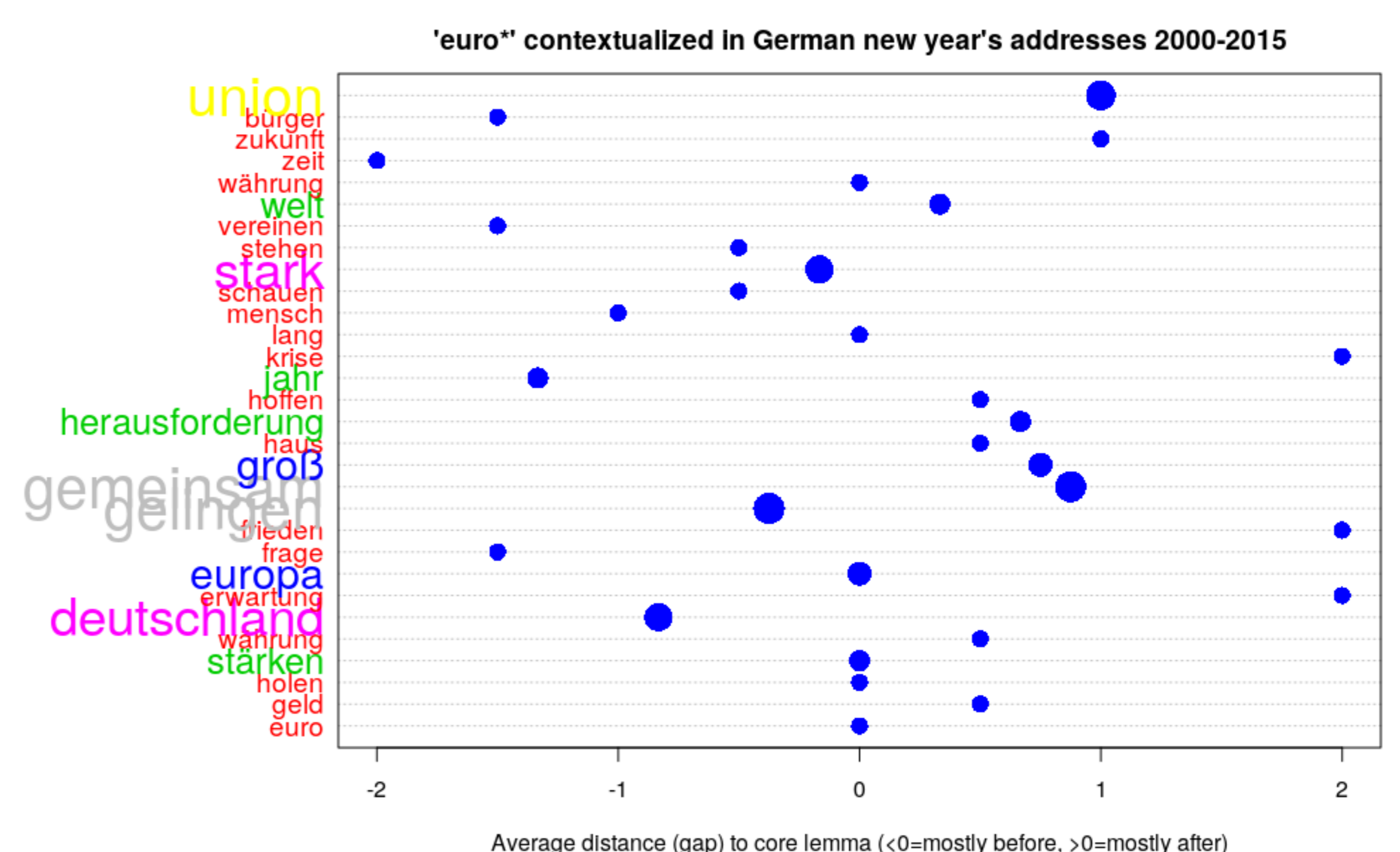
### WHAT COULD YOU DO WITH IT?

- Topic modeling with network analysis (clustering)
- Topic collocation analysis (rhizomes)
- Semantic context analysis
- Thematic pathways analysis, long-term trend analysis

## EXAMPLE #1: CLUSTERS = TOPICS



## EXAMPLE #2: "EUROPE" IN CONTEXT



## CONTACT:

### PROJECT LEADER

Dr Kimmo Elo

Åbo Akademi University, German studies  
University of Turku, European Studies (Social Science)  
Email: [kimmo.elo@abo.fi](mailto:kimmo.elo@abo.fi)  
Mobile: +358-44-342 2122