



Anna Aschauer M.A.

Cosmotool – a biographical search engine

Cosmotool

- Based on DARIAH-DE architecture: the data should be published (in Repository or otherwise) and registered in the Collection Registry
- Allows search within biographies
- <https://cosmotool.de.dariah.eu/cosmotool/personsearch/>



Cosmotool

Consists of different components:

- Elasticsearch (schema free search engine):
<https://www.elastic.co/products>
- NLP-toolkit to analyze fulltext
- A toolkit for entities definitions
- A toolkit for geographical visualization

Motivation for the tool development

General presupposition:

- Reconstructing a group of historical actors involved is helpful with any kind of historical research (analyzing a particular historical event, phenomenon, structures etc.)



Dataset

- (Copy of) German Wikipedia and Wikidata (430k) indexed
- only the abstract and the main section with a biographically relevant heading processed
- *Demonstration: Thomas Mann*

Search for new members of a group

Classifier approach in Cosmotool:

- Make a group of biographies and **let Cosmotool suggest new group members** based on the significant terms

*terms that occur with high frequency in the selected set of biographies compared to their occurrences in all other biographies

Search for new members of a group

- *Demonstration*

- A corpus of 32 Wikipedia biographies was selected

Loose criteria :

- 18th century as a time frame
- connection to St. Petersburg,
- Lutheran/Pietist and Germanspeaking

Try it!

- Create a group

Analyze it:

- What are the significant terms?
- What word were you expecting?
- Can you cluster the significant terms into categories?
- How many of the new members meet your criteria?

My results

All significant terms apart from the number „1724“ can be divided into 2 groups:

- proper nouns („blumentrost“, „bering“, „bering“, „steller“, „ostermann“, „gmelin“..) which are the names of German explorers/naturalists, doctors, „politicians“.
- Place names: „sibirien“, „kamtschatka“, „narva“, „poltava“, „petersburg“.
- only 8 out of 14 suggested by Cosmotool members meet the unexplecit criteria (Hannibal, von Campenhausen, Elias von Cyon, Gmelin, Weikard, Billings, de la Croyère, von Hallart)

Support

- For general questions: aschauer@ieg-mainz.de
- For technical questions: tobias.gradl@uni-bamberg.de