Visualizing the Dutch Folktale Database

Iwe Everhardus Christiaan Muiser
Twente University (Enschede)
e.c.muiser@utwente.nl

Mariet Theune
Twente University (Enschede)
m.theune@utwente.nl

Humanities researchers are often not well aware of the technical advancements and possibilities offered by computer science. Demand for technical aid is therefore often based on traditional research approaches. Our goal is to bridge the gap between the humanities and computer science by developing tools for the exploration of a folktale collection from a technical perspective. Our intermediate results of these experimental approaches will be presented to folktale researchers to inspire alternative ways of investigation.

In this work we present two visualization approaches based on existing metadata and technologies. Two dimensions of folktales, geographic location and date, are obvious choices for visualization due to their relationship with the historical and geographical environment in which the tales came into existence (Abello et al., 2012). These dimensions can be used to provide a birds eye view of the collection, or a subset of the collection. Our system supports selecting subsets of the data based on search queries and visualizes them on a map in combination with a timeline (figure 1). A second visualization approach that can be added on top of such collections, is to display similarity between documents. Similarity between folktales is of great interest to researchers who are searching for the origins of tales (Tehrani et al., 2013) however, most investigations of narrative similarity have only been carried out on a small scale so far (Nguyen et al., 2014). In our visualization tool, the similarity of folktales can be determined based on (different configurations of) metadata such as folktale type or keywords, or the text of a document. Our system visualizes the similarity between tales as a dynamic network graph (figure 2).

The folktale collection we use is the Dutch Folktale Database, which was established in 1994 at the Meertens Institute in Amsterdam. Since then it went through several phases of development and now contains nearly 44.000 folktales with standardized and spring cleaned metadata (Muiser et al. 2012). Dynamic visualizations like plotting search results on a map, timeline or network graph could lead to new insights into existing data. Visualizations can also be beneficial to detect mistakes in metadata that were made during annotation, like typing errors, divergent formats, or capitalization/ punctuation errors.

The tools are still in development and will soon be tested to see if previous assumptions based on qualitative research can be verified by modern means of search and visualization. Also, it will be interesting to observe whether integration of the tools in the Dutch Folktale Database website

(www.verhalenbank.nl) will make it easier for users to browse through the collection. With our work, we hope to open up new avenues for folktale research that go beyond what was previously possible. The experimental visualization interfaces are available online:

The Dutch Folktale Map Tool: http://www.verhalenbank.nl/verhalenkaart/
The Dutch Folktale Network Tool: http://www.verhalenbank.nl/verhalennetwerk/
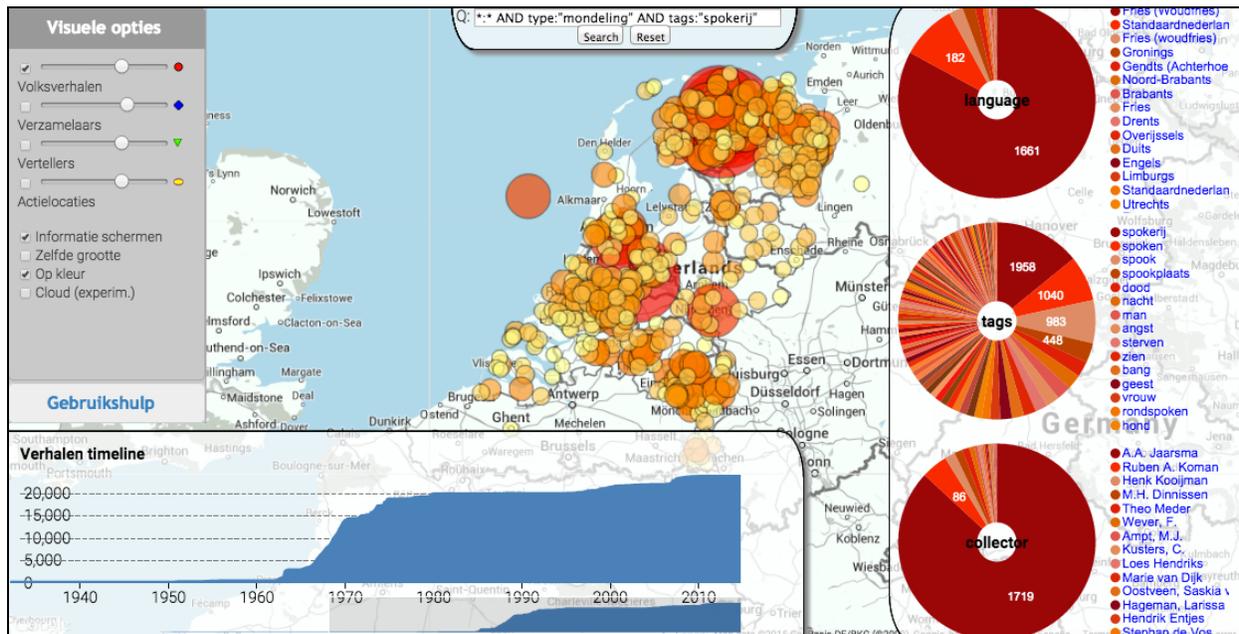

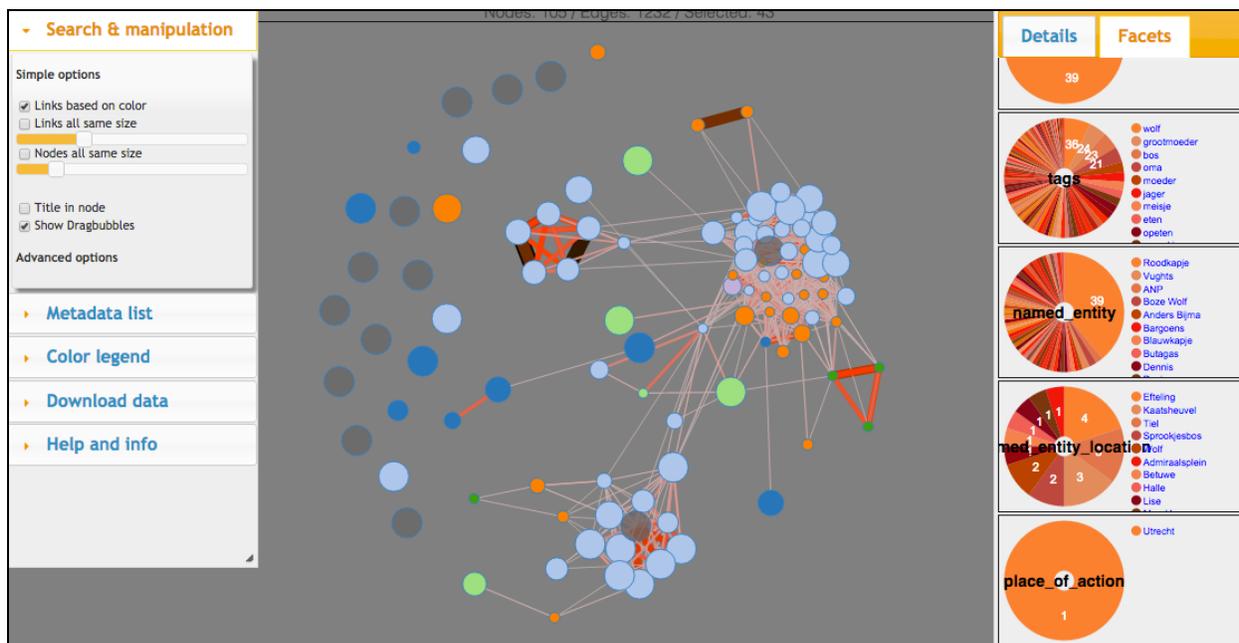Figure 1: Folktale map browser tool

Figure 2: Folktale network browser tool

References

 J. Abello, P. Broadwell, and T. R. Tangherlini (2012) Computational folkloristics. *Communications of the ACM*, vol. 55(7), pp. 60–70.

I. Muiser, M. Theune and T. Meder (2012) Cleaning up and standardizing a folktale corpus for humanities research. *Proceedings of the Second Workshop on Annotation of Corpora for Research in the Humanities (ACRH2)*, pp. 63-74.

D. Nguyen, D. Trieschnigg and M. Theune (2014) Using Crowdsourcing to Investigate Perception of Narrative Similarity. *Proceedings of CIKM'2014*, pp. 321330.

J. J. Tehrani (2013) The phylogeny of Little Red Riding Hood. *PloS one*, vol. 8(11):e78871.