

ShiCo: A Visualization Tool for Shifting Concepts Through Time

Carlos Martinez-Ortiz
Netherlands eScience Center
c.martinez@esciencecenter.nl

Tom Kenter
University of Amsterdam
tom.kenter@uva.nl

Melvin Wevers
Utrecht University
m.j.h.f.wevers@uu.nl

Pim Huijnen
Utrecht University
P.Huijnen@uu.nl

Jaap Verheul
Utrecht University
J.Verheul@uu.nl

Joris van Eijnatten
Utrecht University
J.vanEijnatten@uu.nl

Over time, people use different words to discuss similar concepts. Within a cluster of words used to refer to a concept, words tend to appear and disappear (Kuukkanen 2008). For example, the characterization *anarchists*, used around 1900 to identify people who resort to violence to achieve political goals, has lost currency nowadays. Nevertheless, it is not hard to find a semantic resemblance between the *anarchists* and present-day *terrorists*. In this abstract, we present ShiCo, an open source visualization tool for monitoring these shifts in vocabulary.

ShiCo is a web-based tool that gives its users insight into changes in vocabulary over time, given a small set of user-provided key terms. Such changes in vocabulary (or vocabulary shifts) are calculated from an underlying corpus of time-stamped documents consisting of Dutch newspaper articles¹. It is important to note, however, that any corpus of time-stamped documents can be used.

The ShiCo interface displays a stream graph, which shows the relative importance of terms (with respect to similar terms) over time. Additionally, a set of network graphs display closely related terms over different periods in time. Figure 1 provides screenshots of both types of graphs for the example query *oorlog* (*war*, in Dutch). The ShiCo interface allows the user to adjust various parameters used for generating the vocabularies visualized. ShiCo provides functionality for exporting and importing results, which is important for reasons of reproducibility.

The algorithms used for generating the shifting vocabularies visualized by ShiCo are described in (Kenter et al. 2015). In short, the semantic relations between words in different periods are employed to monitor shifts in vocabulary. In particular, word embeddings – multidimensional distributed

¹ The corpus underlying ShiCo spans 4 decades. It is provided by the National Library of the Netherlands. The full corpus, and more, can be queried at <http://www.delpher.nl>.

representations of words in a semantic vector space (Mikolov et al. 2013) – are employed to capture the meaning of words in a particular time period. Semantic networks of words and their neighbors in the semantic vector space are constructed over time, and an aggregation algorithm is employed to arrive at the final vocabularies being displayed to the user. ShiCo visualizes both the final vocabularies and the underlying semantic networks.

ShiCo consists of two parts: a RESTful API back end and a web-based front end. The API is written in Python. The front end uses AngularJS and NVD3. ShiCo is completely open source, and both the code and the word embedding models needed for calculating the vocabularies are available on Github (<https://github.com/NLeSC/ShiCo>).

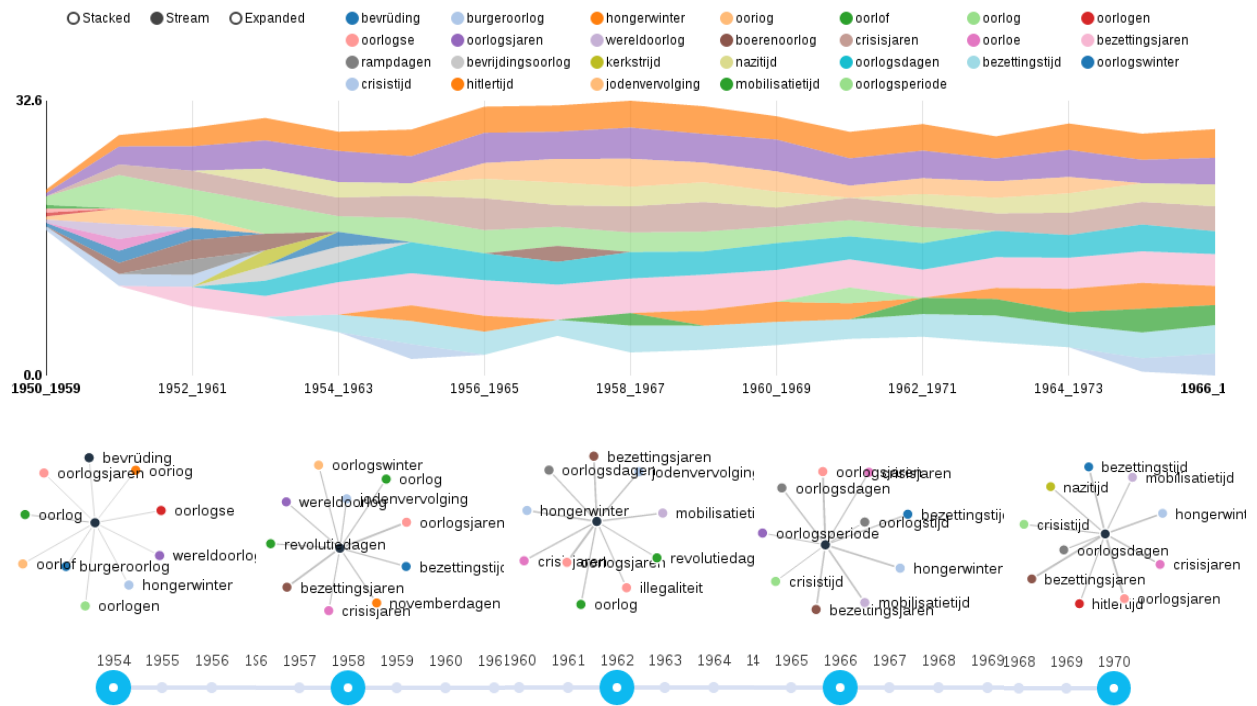


Figure 1: Screenshots of ShiCo interface

Acknowledgements

ShiCo is developed in the context of a Netherlands eScience Center Pathfinder project in cooperation with Utrecht University and the University of Amsterdam.

References

- Kuukkanen, J.-M. (2008) 'Making sense of conceptual change', *History and Theory*, vol. 47, issue 3, October, pp. 351-372.
- Kenter, T.; Wevers, M.; Huijnen, P. and de Rijke, M. (2015) 'Ad Hoc Monitoring of Vocabulary Shifts over Time', *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, ACM
- Mikolov, T.; Chen, K.; Corrado, G. and Dean, J. (2013) 'Efficient estimation of word representations in vector space', *CoRR*.