

Cast Your Net Wide: Finding Historical References in Parliamentary Data

Alex Olieman
alex@olieman.net
University of Amsterdam, the Netherlands

Kaspar Beelen
kasparvonbeelen@gmail.com
University of Amsterdam, the Netherlands

Introduction

Analyzing how the past is molded to fit contemporary needs and priorities represents an intensely researched topic among historians, who have scrutinized (among others) the performance of “commemorative practices,” the transmission of “collective memories,” or, in general, “the language of the past.” However, scholars currently lack adequate digital tools to assist them in this endeavor (Likaka, 2009; Piersma and Ribbens, 2013; Wilson, 2016).

The information needs of historians and other humanities scholars are not always adequately captured by traditional full-text search because query terms often serve as a proxy for more abstract concepts, while both items (the word and the concept) sometimes hardly overlap. This technology demonstration focuses on a specific type of concept, namely historical eras or periods.

Because of their complex and heterogeneous character, historical periods are difficult to capture with simple keywords. For example, the historian looking for references to the Renaissance in diachronic corpora will often encounter documents in which the term appears in its metaphorical sense, e.g. “the renaissance of train travel,” instead of referring to the historical period. Besides yielding many irrelevant hits, keyword search also tends to overlook many relevant documents, because it ignores entities that are related to the query at a *conceptual* level, such as the painter Titian or the Nonsuch Palace.

Traditional full-text retrieval methods will obtain low recall and precision, when the target of a search comprises complex phenomena such as historical events or periods. To overcome these shortcomings,

we introduce “WideNet,” a tool that allows historians and other scholars within the humanities to search for information about specific periods in a multilingual corpus containing the Parliamentary debates from the United Kingdom, Canada, and the Netherlands.

WideNet models historical periods as a container of (hierarchically) related entities. The art historian, interested in how British MPs refer to the Italian Renaissance will, instead of retrieving all occurrences of the term “Renaissance,” be provided with speeches that mention, e.g., composers such as Costanzo Porta, and famous paintings such as The Bacchanal of the Andrians. Whereas earlier work proposed to use such entities as search suggestions (Piersma et al., 2014), we rather prefer a “sculptor’s approach” in which many containers of potentially relevant entities are initially included in the query, and may be deselected based on their empirical relevance.

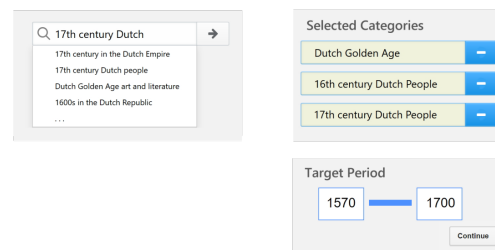


Figure 1: Initial query specification in WideNet

These categories of entities are provided by a knowledge base (KB) that is loaded into the tool in advance. Our demo tool makes use of DBpedia, but any KB that conforms to the SKOS ontology may be loaded. The scholar, using WideNet, starts by selecting one or several root categories from a type-ahead search box (see Figure 1), and can further demarcate the query by selecting a time period, which will be used to prune the underlying entities of the selected categories. WideNet subsequently retrieves the network of narrower categories for each selected root category, and collects the contained entities as potentially relevant query components.

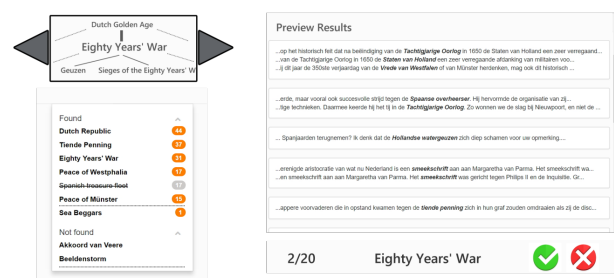


Figure 2: Assessing the relevance of categories and entities

The next step for the WideNet user is to assess which of the retrieved subcategories actually contain entities that lead to relevant results. The interface facilitates this task by showing, per subcategory, which entities are mentioned in the corpus, and how frequently, as well as which entities did not occur (see Figure 2). It also displays a list of preview results, showing limited context, to offer quick clues about the relevance of the category. This preview is also useful to identify individual entities that are not relevant after all, which can be deselected.

After inspecting and selecting relevant categories of entities, the demo interface allows further exploration by providing an environment in which the retrieved documents can be tagged and subjected to close reading. Moreover, the user can examine the results in relation to the parliamentary metadata, i.e. look for saliency by plotting the annotations over time, or study bias by comparing how often different political parties refer to the entities of interest.

WideNet offers a flexible and widely deployable interface that enables scholars to simultaneously search for myriad aspects that have shaped specific historical eras. It provides researchers with a holistic picture of how these periods are discussed in parliament, and thereby helps future scholars to get a more profound understanding of how history ties in with contemporary issues, and how societies deal with their past.

Bibliography

Likaka, O. (2009). *Naming Colonialism: History and Collective Memory in the Congo, 1870–1960*. Madison: University of Wisconsin Press.

Piersma, H., Tames, I., Buitinck, L., Doornik, J., and Marx, M. (2014). "War in Parliament: What a Digital Approach Can Add to the Study of Parliamentary History." *Digital Humanities Quarterly*, 8(1).

Piersma, H. and Ribbens, K. (2013). "Digital Historical Research: Context, Concepts and the Need for Reflection." *BMGN - Low Countries Historical Review*, 128(4): 78–102. DOI: <http://doi.org/10.18352/bmgn-lchr.9352>

Wilson, R. (2016). *The Language of the Past*. London: Bloomsbury Academic.