

PoliMedia

Improving Analyses of Radio, TV & Newspaper Coverage of Political Debates

Max Kemman, Martijn Kleppe

Erasmus University Rotterdam, The Netherlands
kemman@eshcc.eur.nl, kleppe@eshcc.eur.nl

Abstract. Analysing media coverage across several types of media-outlets is a challenging task for academic researchers. The PoliMedia project aimed to showcase the potential of cross-media analysis by linking the digitised transcriptions of the debates at the Dutch Parliament (Dutch Hansard) with three media-outlets: 1) newspapers in their original layout of the historical newspaper archive at the National Library, 2) radio bulletins of the Dutch National Press Agency (ANP) and 3) newscasts and current affairs programs from the Netherlands Institute for Sound and Vision. In this paper we describe generally how these links were created and we introduce the PoliMedia search user interface developed for scholars to navigate the links. In evaluation it was found that the linking algorithm had a recall of 67% and precision of 75%. Moreover, in an eye tracking evaluation we found that the interface enabled scholars to perform known-item and exploratory searches for qualitative analysis.

Keywords. political communication, parliamentary debates, newspapers, radio bulletins, television, cross-media analysis, semantic web, information retrieval

1 Introduction

Analysing media coverage across several types of media-outlets is a challenging task for academic researchers. Up until now, the focus has been on newspaper articles: being generally available in digital, computer-readable format, these can be studied relatively easily. Analyses of visual material like photos or television programs are however rarely undertaken. We expect that both researchers of political communication (e.g. [1]) as well as researchers on television (e.g. [2]) would benefit from a cross-media comparison, providing a better overview of the choices that different media outlets make.

The PoliMedia project¹ aimed to showcase the potential of cross-media analysis by linking the digitised transcriptions of the debates at the Dutch Parliament (Dutch Hansard) with three media-outlets: 1) newspapers in their original layout of the historical newspaper archive at the National Library, 2) radio bulletins of the Dutch National

¹ <http://www.polimedia.nl>

Press Agency (ANP) and 3) newscasts and current affairs programmes from the Netherlands Institute for Sound and Vision.

The PoliMedia search user interface (SUI) allows researchers to browse the debates by date and analyse the related media coverage, as well as searching by name of a politician or any keyword and evaluate the debates in which the politicians appeared and how he or she was covered in the press. The SUI consists of three main levels: 1) the landing page where researchers can enter search terms, 2) the results page (figure 1) with the search results, facets for refinements and a search bar for new queries and 3) the debate page (figure 2) which shows a complete debate and the linked media items. An advantage of PoliMedia is that the coverage in the media is incorporated in its original form, enabling analyses of both the mark-up of news articles as well as the photos in newspapers and the footage of the televised programs. The main research question that can be addressed using the datasets and technology provided by the project is: *What choices do different media make in the coverage of people and topics while reporting on debates in the Dutch parliament since the first televised evening news in 1956 until 1995?*

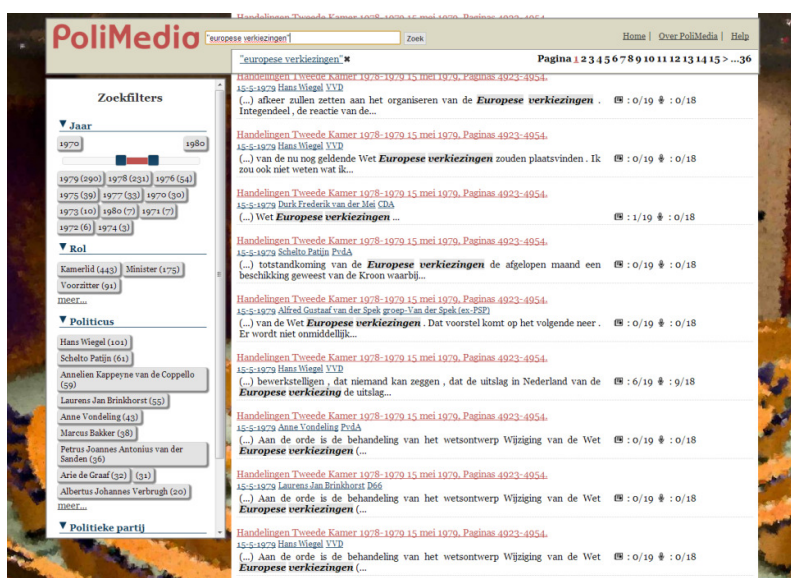


Fig. 1. Screenshot of the PoliMedia search results page

2 Method

The basis of PoliMedia lies in the minutes of Dutch parliament from 1814-1995, containing circa 2.5 million pages of debates with speeches that have been OCR'd and thus allow full-text search. The minutes have been converted to structured data in XML form in previous research [3]. For each speech (i.e. a fragment from a single

speaker in a debate), we extract information to represent this speech; the speaker, the date, important terms (i.e. named entities) from its content and important terms from the description of the debate in which the speech is held. This information is then combined to create a query with which we search the archives of the newspapers, radio bulletins and television programmes. Media items that correspond to this query are retrieved, after which a link is created between the speech and the media item, using semantic web technologies [4].

In order to navigate these links, SUI was developed in which the parliamentary debates are presented with links to the media coverage. The development was based on a requirements study with five scholars in history and political communication, leading to a faceted SUI as depicted in figure 1. Facets allow the user to refine search results, they support the searcher by presenting an overview of the structure of the collection, as well as provide a transition between browsing and search strategies [5]. During development, an initial version of this SUI was evaluated in an eye tracking study with 24 scholars performing known-item and exploratory search tasks [6].



Fig. 2. Screenshot of the PoliMedia debate page

3 Results

From an evaluation of a set of links to newspaper articles, it was found that the recall of the algorithm is approximately 62%, with a precision of 75% [4]. In this context, relevance is indicated as a newspaper article referring to a specific speech or to the entire debate.

From the eye tracking evaluation we found that the faceted SUI enabled users to perform both known-item searches, as well as exploratory searches to analyse a topic over time. However, navigating the debates themselves proved to be rather difficult; as debates can be dozens of pages long, users had difficulty gaining an overview of the debate. To address this issue, the faceted search available on the search results page (figure 1) was also introduced on the debate page (figure 2) in the final version of the interface.

4 Discussion

In the evaluation of the links we found a satisfying balance of recall and precision. However, no links to television programmes could be made. At this point we can make no conclusions about whether this was due to the size of the television dataset, the lack of full-text search or due to lack of suitability of the linking algorithm. The lack of links to television programmes thus remains a question for future research.

We found that the search user interface enabled scholars to navigate the debates and find links to related media coverage. However, the user interface focusses on qualitative research and requires scholars to make an overview themselves by trying several queries and using a combination of facets. The use of these links for quantitative scholarly research has not yet been touched, and is an application we would like to experiment with in future research.

Acknowledgements. We are grateful for the financial support from the CLARIN-NL project PoliMedia – *Interlinking multimedia for the analysis of media coverage of political debates*.

References

1. Van Santen, R.A., Van Zoonen, L.: The personal in political television biographies. *Biography*. 33, 49–67 (2010).
2. Bignell, J., Fickers, A.: *A European Television History*. Wiley-Blackwell, Oxford (2008).
3. Gielissen, T., Marx, M.: Exemplification of parliamentary debates. *Proceedings of the 9th Dutch-Belgian Workshop on Information Retrieval (DIR 2009)*. pp. 19–25 (2009).
4. Juric, D., Hollink, L., Houben, G.: Discovering links between political debates and media. *The 13th International Conference on Web Engineering (ICWE'13)*. , Aalborg, Denmark (2013).
5. Kules, B., Capra, R., Banta, M., Sierra, T.: What do exploratory searchers look at in a faceted search interface? *Proceedings of the 2009 joint international conference on Digital libraries - JCDL '09*. 313 (2009).
6. Kemman, M., Kleppe, M., Maarseveen, J.: Eye Tracking the Use of a Collapsible Facets Panel in a Search Interface. *Theory and Practice of Digital Libraries*. Springer, Malta (2013).