The Digital Archiving of Historical Political Cartoons: An Introduction

Junte Zhang
Meertens Institute & NIOD
Institute for War, Holocaust
and Genocide Studies

Kees Ribbens
Erasmus University Rotterdam
& NIOD Institute for War,
Holocaust and Genocide
Studies

Rob Zeeman
Meertens Institute

Royal Netherlands Academy of Arts and Sciences

ABSTRACT

Political (editorial) cartoons often capture the *Zeitgeist* of society and convey a message. Increasingly, historians study them to understand commentaries of past events or personalities. Visual culture as an academic subject could be greatly enhanced if this information can be digitally archived. We employ crowdsourcing to obtain valuable metadata by guiding volunteers' feedback using an online survey with 31 targeted questions. We provide intellectual access to a set of about 300 cartoons of a single creator spanning over multiple years in a highly interactive search engine.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Search process; H.3.7 [**Digital Libraries**]: Systems issues, user issues; H.5.2 [**Information interfaces and presentation**]: Graphical user interfaces (GUI)

General Terms

Design, Human Factors

Keywords

metadata, crowdsourcing, e-Humanities, cartoons

1. INTRODUCTION

Newspapers often have political (editorial) cartoons that contain a commentary about events or personalities [4] which is being disseminated. For historians, these capture the *Zeitgeist* of the period of time of their study, and become an invaluable source of information. These print newspapers are stored in libraries and get digitally archived – for example by the National Library of the Netherlands – for long-term preservation to continued access. Digital archiving is the management of the life cycle of digital assets (records) [2], from preservation to continued use.

In the Radical Political Representation project, we aim to digitally archive historical political cartoons created by a single cartoonist and published before and during the Second World War, so we gain insight into different points of view and support the study of visual culture using a computational approach. This is made possible because newspaper pages have been digitized as images, which contain cartoons. These cartoons are not yet machine-readable, therefore providing intellectual access is the best option. It has been proposed in [5] to detect the text lines in cartoons using OCR, but this is difficult because it involves handwritten texts. In [3] it is pointed out that "more descriptive areas by which images might be accessed are largely neglected," and argued that subject indexing as a field of academic work is *aboutness* – and VRA Core 4.0 is referred to as a metadata schema to record bibliographic information.

Our aim is to transcribe a cartoon, and move beyond standard bibliographic information by comprehensively capturing its meaning(s) for historical research by eliciting user feedback using crowdsourcing. So we address the following question: How can we provide intellectual access to, and allow for, advanced use of these cartoons?

2. CROWDSOURCING OF CARTOONS

The objects of our study are so far 286 cartoons published by Maarten Meuldijk in the weekly *Volk en Vaderland* (VoVa) of the National Socialist Movement in the Netherlands from 1937 to 1942. Pages on which they occur have been digitized by the National Library. To obtain descriptions about the cartoons, we experiment with crowdsourcing to see whether crowdsourcing is applicable in our context.

The search tasks that we have in mind are more complex, therefore we created a comprehensive survey that captures the questions historians typically would ask about a cartoon. This also requires more contextual knowledge. Fig.1(a) shows the VoVa Annotation Editor developed in Adobe Flex, where we guide users through a set of 31 targeted questions in 8 stages, and aid them by offering answers of these questions with pre-defined multiple choice answers in combination with open answers. Users can zoom in/out on a cartoon, but also read contextual information related to the cartoon in the articles on the page – a strategy used by a number of users. There are no time limits and a cartoon is randomly assigned and stays assigned to a user until completion. To control for the completion of a cartoon description, we validate all questions for at least 1 given answer.

We invited interested volunteers online and in printed national media. In total 189 users registered, where eventually 83 volunteers participated with at least 1 completed description of a cartoon and with 5 users completing more than 10.

(a) The VoVa Annotation Editor, where volunteers can provide valuable metadata about the cartoon, ranging from plain descriptions to their opinion of a cartoon.

(b) The VoVA Search Engine, which is used to gain intellectual (advanced) access to the cartoons.

Figure 1: The digital archiving of cartoons with the VoVa Annotation Editor and Search Engine.

3. SERENDIPITY IN CONTEXT

Having obtained the metadata, we want to use it. Since the search engine should serve historians, we design it to support serendipitous search and be highly interactive in order to focus on a high recall (rather than precision). The system has been designed to maximize the user's ability to explore. We have proposed search features to support serendipitous and focused access in [6], and these features have been re-implemented here. The search features primarily deal with query expansion, recommendation, and interactive visualizations of aggregated results. The former is based on using ternary search trees for spellchecking, returning the top term vectors related to the original query, and returning the top terms that have the original query as substring. The latter is based on charts, maps and word clouds.

A user can improve the searching in a session by effectively reducing the information space step by step, i.e. incrementally combining questions. This confirms with the Berryp-

icking model of [1] – queries are not static, but rather evolve, and users "gather information in bits and pieces instead of in one grand best retrieved set." These steps are stored as part of the search trail, so the overview is kept. The user interface of the system is depicted in Fig 1(b). In this example, someone looked for a cartoon about a "Jood" (*Jew*) used as a main keyword to describe a cartoon, with captions under it, and a "ster" (*star*) depicted as a symbol. The search engine treats the questions asked in the survey as facets, and is therefore a straight-forward question-answering system. Facets that always appear are the date of publication of a cartoon, and the education and knowledge levels of the volunteers who provided the descriptions. We show in [7] that making the credibility of the source transparent gives users greater confidence in their selection. We think historians will be aided with this part of the search process.

There are different search strategies possible. Users can search by full-text or focused (within the answers of questions). The query gets highlighted in context given the full-text and the survey question. A dynamic word cloud widget that supports query expansion is not activated, unless the autocompletion is used. Using the Advanced Search option, users can look up a question and then enter a keyword also with the autocompletion feature. Wildcard (empty) queries can be used to obtain the distribution of words of the answers given a question in a word cloud for a quick summary.

4. CONCLUSIONS

We have presented – in a compressed version – the mission statement and some results of the Radical Political Representation project. We completed the first phase of crowd-sourcing, and pending further releases of data by the National Library, we can further digitally archive the complete series of Meuldijk cartoons. The technical infrastructure to digitally archive political cartoons has been set-up.

This means we can expand our scope to other cartoonists in different times – there is no shortage of cartoons. We can refine our survey to allow for more different information needs of historians, or embed our survey as part or extension of a formal metadata schema like VRA Core. We will improve the UI and further implement useful information visualization of results, and evaluate the search engine. It can be used at `www.meertens.knaw.nl/vova/search`.

5. REFERENCES
[1] M. J. Bates. The design of browsing and berrypicking techniques for the online search interface. *Online Review*, 13(5):407–424, 1989.
[2] G. M. Hodge. An information life-cycle approach : Best practices for digital archiving. *Journal of Electronic Publishing*, 5(4):1–14, 2000.
[3] C. Landbeck. Issues in subject analysis and description of political cartoons. *Advances in Classification Research Online*, 19(1), 2008.
[4] C. Sterling. *Encyclopedia of Journalism*. Sage, 2009.
[5] Y. Wu. Searching digital political cartoons. In *Proceedings of the 2010 IEEE International Conference on Granular Computing*, GRC '10, pages 541–545, Washington, DC, USA, 2010. IEEE Computer Society.
[6] J. Zhang. Supporting serendipitous and focused search. In *EuroHCIR*, volume 909 of *CEUR Workshop Proceedings*, pages 79–82, 2012.
[7] J. Zhang, A. Amin, H. S. M. Cramer, V. Evers, and L. Hardman. Improving user confidence in cultural heritage aggregated results. In *SIGIR*, pages 702–703, 2009.