Supporting Exploration of Historical Perspectives across Collections

Daan Odijk*[1], Cristina Gârbacea[1], Thomas Schoegje[2], Laura Hollink[2,3], Victor de Boer[2], Kees Ribbens[4,5], and Jacco van Ossenbruggen[2,3]

[1] ISLA, University of Amsterdam, the Netherlands
[2] VU University Amsterdam, the Netherlands
[3] CWI, Amsterdam, the Netherlands
[4] NIOD Institute for War, Holocaust and Genocide Studies, Amsterdam, the Netherlands
[5] Erasmus School of History, Culture and Communication, Rotterdam, the Netherlands

Abstract. The ever growing number of textual historical collections calls for methods that can meaningfully connect and explore these. Different collections offer different perspectives, expressing views at the time of writing or even a subjective view of the author. We propose to connect heterogeneous digital collections through temporal references found in documents as well as their textual content. We evaluate our approach and find that it works very well on digital-native collections. Digitized collections pose interesting challenges and with improved preprocessing our approach performs well. We introduce a novel search interface to explore and analyze the connected collections that highlights different perspectives and requires little domain knowledge. In our approach, perspectives are expressed as complex queries. Our approach supports humanity scholars in exploring collections in a novel way and allows for digital collections to be more accessible by adding new connections and new means to access collections.

1 Introduction

A huge amount of digital material has become available to study our recent history, ranging from digitized newspapers and books to, more recently, web pages about people, places and events. Each collection has a different perspective on what happened and this perspective depends partly on the medium, time and location of publication. For example, Lensen [10] analyses two contemporary novels about the second World War and shows that the writers have a different attitude towards the war than previous generations when it comes to issues of perpetratorship, assignation of blame and guilt. We aim to connect collections and support exploration of these different perspectives.

In this work, we focus on the second World War (WWII), as this is a defining event in our recent history. Different collections tell different stories of events in WWII and there is a wealth of knowledge to be gained in comparing these. Reading a news article on the liberation of the south of the Netherlands in a newspaper collaborating with the occupiers gives the impression that it is just a minor setback for the occupiers. A very different perspective on the same events emerges from the articles in an illegal newspaper of the resistance, leaving the impression that war is ending soon. To complete

* Corresponding author, email: d.odijk@uva.nl.

the picture, these contemporary perspectives can be compared to the view of an historian who—decades later—wrote fourteen books on WWII in the Netherlands and to the voice of thousands of Wikipedians, all empowered with the benefit of hindsight.

We present an interactive search application that supports researchers (such as historians) in connecting perspectives from multiple heterogeneous collections. We provide tools for selecting, linking and visualizing WWII-related material from collections of the NIOD, the National Library of the Netherlands, and Wikipedia. Our application provides insight into the volume, selection and depth of WWII-related topics across different media, times and locations through an exploratory interface. iod We connect digital collections via implicit events, i.e. if two articles are close in time and similar in content, we consider them to be related. Newspaper articles are associated with a clear point in time, the date they were published. However, not all collections have such a clear temporal association. We therefore infer these associations from temporal references (i.e., references in the text to a specific date). Our novel approach to connecting collections can deal with these extracted temporal references.

We provide two insightful visualizations of the connected collections: (1) an exploratory search interface that provides insight into the volume of data on a particular WWII-related topic and (2) an interactive interface in which a user can select two articles/pages for a detailed comparison of the content. Our aim is to provide historical researchers with the means to explore perspectives, not to analyze these for them. These perspectives are expressed in our application through what we call *query contrasts*. These contrasts are, in essence, sets of filters over one or more collections, that are added to search queries and consistently visualized throughout the interface. Example uses of such contrast are comparing newspaper articles versus Wikipedia articles (across collections) or local newspaper versus national newspapers (within a collection).

We focus on events and collections related to WWII. However, as our approach and application can be applied to other digital collections and topics, our work has broader implications for digital libraries. Our main contributions are twofold: (1) we propose and evaluate an approach to connect digital collections through implicit events, and (2) we demonstrate how these connections can be used to explore and analyze perspectives in a working application. Our work is based entirely on open data and open-source technology. We release the extracted temporal references as Linked Open Data and our application as open-source software, for anybody to reuse or re-purpose.

The remainder of this paper is organized as follows. We discuss related work in Section 2. Next, we describe our approach to connecting digital collections in Section 3. Section 4 describes our exploratory and comparative interfaces. We provide a worked example in Section 5, after which we conclude in Section 6.

2 Related Work

We discuss related work on studying historical perspectives, on connecting digital collections and on exploring and comparing collections.

Historical Perspectives. Historical news has often been used to study public opinion. For example, Van Vree [18] studied the Dutch public opinion on Germany in the period 1930–1939 based on articles from four newspapers, selected to represent distinct population groups. More recently, Lensen [10] showed how two contemporary novels

exhibit new perspectives on WWII. Typically, scholars from the humanities study these perspectives on a small scale. Au Yeung and Jatowt [2] studied how the past is remembered on a large scale. They find references to countries and years in over 2.4 million news articles written in English and study how these are referred to. Using topic modeling they find significant years and topics and compute similarities between countries. Our work provides scholars with the means to explore their familiar research questions on different perspectives on a large scale, without doing this analysis for them.

Connecting Digital Collections. With more collections becoming digitally available, researchers have increasingly attempted to find connections between collections. Common is to link items based on their metadata. When items are annotated with concepts from a thesaurus or ontology, ontology alignment can be used to infer links between items. An example is MultimediaN E-culture, where artworks from museums were connected based on alignments between thesauri used to annotate the collections [16]. An approach that does not rely on the presence of metadata is to infer links based on textual overlap of items. For example, Bron et al. [3] study how to create connections between a newspaper archive and a video archive. Using document enrichment and term selection they link documents that cover same or related events. Similarly, in the PoliMedia project [9] links between political debates and newspapers articles are inferred based on topical overlap. As in our approach, both used publication date to filter documents in the linking process. However, how to score matches using these dates and combine them with temporal references is an open problem. Alonso et al. [1] survey trends in temporal information retrieval and identify open challenges, that include how to measure temporal similarity and how to combine scores for textual and temporal queries.

Exploratory Search. To support scholars in studying historical perspectives, we propose an exploratory search interface. In exploratory search [11], users interactively and iteratively explore interesting parts of a collection. Many exploratory search systems have been proposed; we discuss a few that are closely related to our work. Odijk et al. [14] proposed an exploratory search interface to support historians in selecting documents for qualitative analysis. Their approach addresses questions that might otherwise be raised about the representativeness, reproducibility and rigidness of the document set. de Rooij et al. [8] collected social media content from four distinct groups: politicians, journalist, lobbyist and the public and proposed a search interface to explore this grouped content. Through studying the temporal context and volume of discussion over time, one could answer the questions of who put an issue on the agenda.

Comparing Collections. To support media studies researchers, Bron et al. [4] propose a subjunctive search interface that shows two search queries side-by-side. They study how this fits into the research cycle of media studies researchers. They find that when using the proposed interface, the researcher explore more diverse topics and formulate more specific research questions. ManyPedia [12] allows users to explore different points of view by showing two Wikipedia articles from different languages side-by-side. A similar approach was used to synchronize cross-lingual content [13] on Wikipedia. Similar to the work described above, we provide exploratory search tools that emphasizes different perspectives. Our work differs in that our application provides an end-to-end solution, from connecting multiple collections to exploring and comparing.

3 Connecting Digital Collections

In this section, we describe how we connect multiple heterogeneous digital collections. Simply put, we connect documents from different collections via implicit events using time and content. If we consider two newspaper articles from different newspapers, but published on the same day and with considerable overlap in content, we can infer that it is likely that they cover the same event. However, not all our sources are as neatly dated as the newspapers are. For these, we need to extract the relevant dates based on the document content (detailed in §3.2). We present and validate our approach to connecting collections in §3.4 and §3.5 respectively. First, we describe our collections.

3.1 Collections

We connect three heterogeneous collections, each representing a different kind of data source: 1) a digitized collection of around 100 million Dutch newspapers articles, spanning four centuries, 2) the encyclopedic articles of the Dutch Wikipedia, 3) the digitized book series *Het Koninkrijk der Nederlanden in de Tweede Wereldoorlog*[6] by historian Loe de Jong. Table 1 provides an overview of the size and characteristics of each collection. For conciseness, we will refer to these three collections as the *Newspapers*, *Wikipedia* and *Loe de Jong* collections respectively in the remainder of this paper.

The newspaper archive of the National Library of the Netherlands consists of around 100 million digitized newspaper articles, processed using optical character recognition[7]. Each newspaper article has consistent metadata, including publication date. To focus on the relevant articles we filter out articles published before 1933 or after 1949. The books of Loe de Jong are a standard reference work about WWII in the Netherlands, consisting of 14 volumes and published between 1969–1988 in 29 parts, all recently digitized [7]. Early parts focus on chronicling events in specific years whereas some later ones focus on specific themes. Each section was treated as a new document. Note that the documents are still substantially longer than those in the two other collections.

3.2 Extracting Temporal References

We connect the three heterogeneous collections presented above via content and time. For the digitized newspaper collection, we use the publication date of an article. This metadata is clean and checked by the National Library. However, for Wikipedia articles and the books of Loe de Jong no clear dates are present. We extract the dates that these articles refer to through a process of temporal tagging, i.e. we extract references to dates from the article content. For this, we use a custom pipeline for xTAS[8], an extendable toolkit for large-scale text analysis. Concretely, our approach for extracting dates consists of three steps: 1) *preprocessing*, 2) *temporal tagging* and 3) *aggregating*.

In the preprocessing step, we normalize the text and prevent common errors we encountered in the subsequent temporal tagging. For Wikipedia articles, we remove all special syntax, used for formatting and creating tables. We extract the textual content of a Wikipedia article using a MediaWiki syntax parser[9]. For the Loe de Jong collection, we remove XML tags and keep only the textual content. This textual content has been obtained from book scans using optical character recognition (OCR). Therefore, it can contain errors in the recognition of terms. We process this digitized text to remedy

[6] In English: *The Kingdom of the Netherlands during WWII.* [7] http://delpher.nl/kranten
[8] http://xtas.net [9] http://github.com/pediapress/mwlib

Table 1: Statistics for the collections and for temporal reference extraction.

	Newspapers	Wikipedia	Loe de Jong
Number of documents	21,456,471	2,699,044	1,600
Average number of terms per document	105	91	1776
Number of annotated documents		50	20
Total number of unique / individual annotated date references		609 / 834	469 / 713

common OCR errors that we encountered, in particular errors that influence temporal tagging. For example, the numbers 0, 1 and 5 are commonly confused for °, I and S.

For both collections, we also use simple textual replacement rules to prevent common errors we found after an initial evaluation of temporal tagging on our data. A common short-hand way of referring to years is to use an apostrophe followed by only the last to digits: the period '40-'45. As this gives no information on the century being referred to, such a reference is typically ignored by a temporal tagger. However, these references often refer to the 1900s and given that the topic of most of our document collection (WWII), we resolve a reference as above to the period 1940-1945.

After preprocessing, we analyze the content of each Wikipedia article and each document in the Loe de Jong collection using the multilingual cross-domain temporal tagger Heideltime [17]. The aim of a temporal tagger is to find all mentions of time and dates and to pinpoint these as exact as possible to a specific point in time. The output of Heideltime is a set of temporal references normalized according to the TIMEX3 annotation standard [15]. Temporal tagging of historical documents is particularly challenging due to the fact that temporal expressions are often ambiguous and under-specified. For example, *"in the 1930s"* refers to the time interval 1930-1939, while *"august 1945"* refers to the entire month of August, 1945. For most of these challenges, Heideltime is able to extract and normalize temporal references, even if they are under-specified.

The final step to extracting dates is aggregating all temporal references to the document level. We separate each temporal reference based on the annotation granularity (i.e. exact day, specific month or only year). We store and treat them differently both in connecting collections (see §3.4) and in the exploratory visualizations (see §4.1).

3.3 Evaluating Temporal Reference Extraction

We validate our approach to extracting date references with an experiment. We take 50 random documents from the Wikipedia collection and 20 from the Loe de Jong collection. Five judges annotate all date references within the documents. Table 1 details statistics on the annotated documents. The narrative structure of the books of Loe de Jong leads to less frequent temporal references than the encyclopedic style of Wikipedia. We compute inter-annotator agreement over five doubly annotated Wikipedia documents and three documents from the Loe de Jong collection. We observe 97% agreement on the set of unique dates referenced, signaling excellent agreement among the human annotators. We measure the accuracy of the extracted temporal references by comparing the automatically annotated date references to those annotated by the judges.

On the Wikipedia collection, we observe a mean precision of 98.27% on the set of all automatically extracted dates, with recall at 86.72% of unique annotated dates. These scores are comparable to what is reported on standard datasets [17] and signals that the task of extracting dates from Wikipedia articles is well suited to be done automatically.

On the Loe de Jong collection, we obtain substantially lower precision of 63.86% and recall of 68.04%. The sections of these books pose two distinct challenges for temporal tagging. First, as the books are digitized using OCR, there are errors in detected terms, including in parts of dates. Our preprocessing approach to remedy some of the common errors has doubled both precision and recall (up from 36.86% and 30.48% respectively). The second challenge is more difficult. The books of Loe de Jong are written in a narrative style, where temporal references are often (partially) implicit. For example, a section on famine in the winter of 1944-1945 (referred to as the "hunger winter") often only refers only to days in these winter months, without referring to a year. Given the topic, a reader knows that January 15th refers to January 15th, 1945, but for an automatic approach, this is rather difficult to infer. In fact, half of the fourteen books indicate in the title that they cover only a specific period.

Improving the accuracy of temporal reference extraction on such a collection poses interesting future work for information extraction researchers. Given the length of the documents and thus large number of temporal references, the level of accuracy we obtain after preprocessing is sufficient for ours and similar applications. The extracted temporal references for the Loe de Jong collection are published as Linked Open Data[10]. This enrichment allows for new types of temporal analysis of the collection.

This approach for temporal reference extraction is the first step in connecting these three heterogeneous collections. Using the temporal references for Wikipedia articles and sections of the books of Loe de Jong, combined with the publication dates of newspaper articles, we can find subsets of documents for a specific time period that were either published within that period or refer to a point in time within that period. This provides the researcher with the valuable means to find similar subsets across different collections. However, this does not yet mean that all the documents in the subsets are topically related. For this, we also need to look at the content of the document.

3.4 Combining Temporal and Textual Similarity

To estimate whether two documents refer to the same implicit event, we combine textual similarity with temporal similarity. We measure textual similarity as Manhattan distance over document terms in a TF.IDF weighted vector space. Concretely, we take the subset of maximally 25 terms from a source document, that have the highest TF.IDF score. We then select documents that match at least 30% of these terms and compute similarity as the sum of TF.IDF scores over the terms. More matching terms thus lead to a higher similarity, as does matching a less common term than a more common term.

We measure temporal similarity using a Gaussian decay function. If two documents are from the same date, they are completely temporally similar. The further the two documents are apart in time, the lower the similarity score. In case we are matching two documents based on temporal references, we multiply the scores we obtain for each temporal reference match. The overall similarity between two documents is then computed by multiplying the temporal similarity with the textual similarity. In this way, temporal similarity functions in a similar matter as a temporal document prior would work, giving preference to documents from a specific period.

[10] The exported RDF triples are ingested in the "Verrijkt Koninkrijk" triple store. The updated triple store can be found at http://semanticweb.cs.vu.nl/verrijktkoninkrijk/.

3.5 Evaluating Related Article Finding

We evaluate our approach to measuring similarity using a retrieval experiment to find related documents within the Wikipedia collection. The task is to find documents related to a source Wikipedia article within the Wikipedia collection. We compare two approaches for finding related documents: using only textual similarity and combining temporal and textual similarity.

We sample ten Wikipedia articles out of the 18,361 articles that link to an article with WWII in the title (*"Tweede Wereldoorlog"* in Dutch). We pool the top ten results based on textual similarity and have annotators judge the relatedness of two documents side-by-side on a four-point scale, label from bad to perfect. We obtain the relatedness labels via a crowdsourcing platform. To ensure good quality judgments, we manually create a set of gold standard judgments for twelve document pairs that pilot judges agreed entirely on. Our crowdsourcing judges need to obtain an agreement of over 70% with the gold standard to start judging. During judging, gold standard pairs are intertwined with unjudged pairs. If the judges do not maintain this agreement on the gold standard, they cannot continue and their judgments are not included. We obtain at least three judgments per document pair, more if the three judges do not agree. We obtain 812 judgments (including the judgments for the gold standard) and measure a mean agreement of 69.5% over all document pairs. We compute a final rating for a document pair as the mean rating over all judges for that pair.

In our application, related documents are presented to find interesting alternative perspectives from different collections. The related documents are presented as a ranked list, very similar to a standard information retrieval setting. Given this setting and the annotations on a four-point scale, we choose nDCG@10 as our evaluation metric. The nDCG metric can incorporate graded relevance and gives more importance to results higher in the ranked list. We compute nDCG only on the top ten results, as we expect that lower documents are unlikely to be inspected by a user. An nDCG score of 1 indicates that documents are ranked perfectly in order of their relevance score and a score of 0 would mean that all retrieved documents have the lowest relevance score.

Using only textual similarity, we measure an nDCG value of 0.861 and an average rating in the top ten of 2.6 on a scale from 1 to 4. This suggests that the retrieved documents are already of good quality and ranked in a reasonable order. By combining textual and temporal similarity we improve the nDCG score with 3.8% to 0.894. A detailed look at each of the ten source documents shows improvements in nDCG scores up to 25%, but also decreased scores up to 5%. The results suggest that we effectively retrieve related documents and that combining textual and temporal similarity improves effectiveness over only using textual similarity. There are interesting challenges for future research in new approaches for incorporating temporal similarity. For example, intuitively, a match on year references is less strong a match than one on day references. One could therefore weigh matches based on the least fine-grained granularity.

4 Search Interface to Explore Perspectives

We described how we connect collections and their content in Section 3 as our first main contribution. Our second contribution is a novel search and analysis application[11].

[11] The fully functional application can be accessed at http://qhp.science.uva.nl.

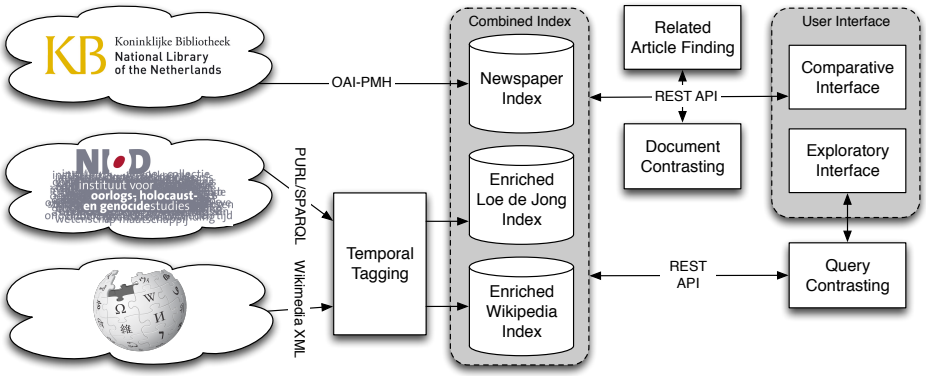


Fig. 1: Architecture to support exploration of perspectives in collections. The interface (right) interacts with a combined index (center) collected from three sources (left).

We provide new means of exploring and analyzing perspectives in these connected collections. Our architecture is modeled in Fig. 1. At the core of the application we use proven open-source technology such as xTAS and Elasticsearch. For each collection, we build a separate index that is exposed to the user interfaces as a combined index.

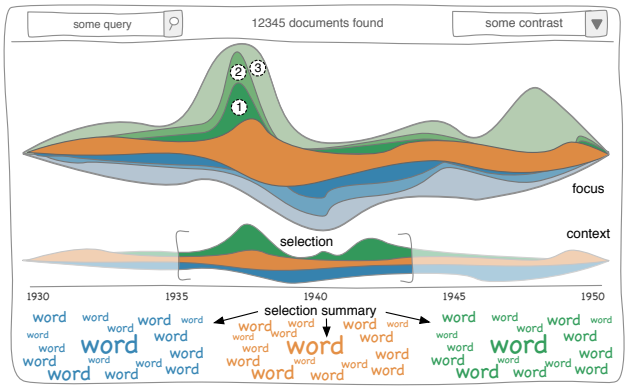
A researcher interacts with our application through two separate, but connected interfaces: 1) an exploratory search interface for a broad overview, and 2) a comparative interface for detailed analysis. We will discuss the flow between these interfaces with a worked example of a researcher interacting with the application in Section 5. First, we will describe each of the two interfaces in more detail below. The comparative interface communicates directly with the combined index, supported by related article finding and document contrasting services. For the exploratory interface, all requests to the index are processed through the query contrasting system.

We provide a researcher with the means to explore perspectives in our application via contrasting. In the comparative interface, two documents are contrasted in detail in a side-by-side comparison. In the exploratory interface, a researcher can combine a keyword query with predefined "query contrasts". A query contrast can be seen as a set of filters that each define a collection subset. A single filter functions in a similar way as facet filters. Such a query contrast filter can simply be contrasting different collections (e.g., newspaper versus Wikipedia articles), or different sources (e.g., a collaborating newspaper versus one run by the resistance) or different locations of publishing. Using a set of these filters (what we call a *query contrast*), what is expressed as a simple keyword query turns into a contrasting comparison between different perspectives.

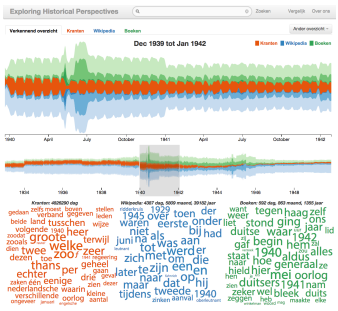
4.1 Exploratory Interface

To allow researchers to explore the three connected collections, we build an exploratory interface as part of our application. This interface is sketched in Fig. 2a.

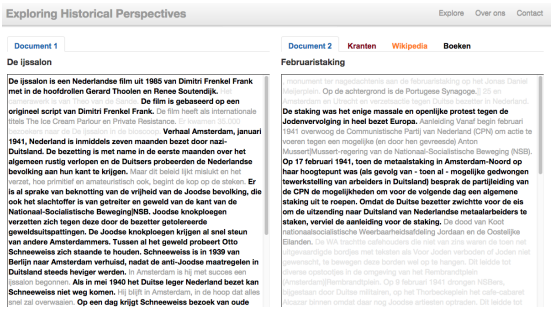
Based on a keyword query and a contrast in the search bar on top, an overview of the search results is presented. Central in the exploratory interface is a visualization that shows the distribution of the volume of documents across time. We visualize this distribution as a streamgraph [5], that can be seen as a streamlined version of a stacked bar chart. A researcher can select a time period of interest while maintaining overview through a Focus+Context interaction design [6]. This allows researchers to focus on a specific period, while at the same time getting an impression of entire time period.



(a) Sketch of the exploratory interface with search on top, a streamgraph to visualize document volume and word clouds to summarize each collection.



(b) Screenshot of (a).



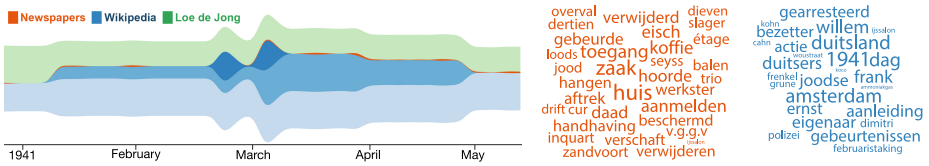
(c) Screenshot of comparative interface.

Fig. 2: Sketch (a) and screenshot of exploratory (b) and comparative (c) interfaces.

The streams in the context visualization are defined by the selected query contrast and consistently color coded based on this. In the simplest case, each represent one of the three connected collections: newspaper articles, encyclopedic articles and sections of the reference books. For each stream, we show a word cloud representing the most significant terms in the documents in each stream for the selected time period. This provides the researcher with a quick overview of what topics these documents cover.

In §3.2, we described how we extract temporal references that can have a granularity of a day, month or even a year. If a stream is based on extracted date references, we distinguish in the focus stream graph between different temporal reference granularity. A stream is then split into three substreams: 1) day references, 2) month references and 3) year references (marked in Fig. 2a). The color coding is kept consistent, but opacity decreases as temporal references become less fine-grained. Similarly, the more fine-grained day references are positioned closer to the center of the stream. If a document refers to a specific day, it refers to that month and year as well. In the visualization, we do not count a reference for a less fine-grained substream if a more fine-grained reference occurs for that day. This way, the combined height of the three substreams at any point of time is equal to the references to that day, month or year.

Not depicted in Fig. 2 is the collection search interface, that shows a simple ranked list of documents within any of the three collections. From this search interface, a researcher can select a document to study in more detail in the comparative interface.



(a) Streamgraph (b) Newspapers (c) Wikipedia

Fig. 3: Screenshots for query "ijssalon koco" from October 1941 until March 1942.

4.2 Comparative Interface

The comparative interface shows two documents side-by-side. At first, a selected document is shown on one side, while the other side shows related documents from each of the three collections using the approach described and evaluated in Section 3. When selecting a document from these results, the side-by-side comparison is shown. A researcher can return to the related articles on either side at any time.

When comparing two documents side-by-side, interesting parts of the document are highlighted. Using an approach similar to the textual similarity described in Section 3.4, we compute the similarity of each sentence in a document to the document on the other side. Sentences with a high similarity are shown clearly, whereas sentences with a low similarity are shown partially transparent. This dimming effect draws the attention of the researcher to the interesting parts of a document in the context of another.

5 A Worked Example

We describe a worked example of how our application can be used to study different perspectives on a specific event in WWII. A full-blown evaluation of our application in a user study with historians is planned for future work.

We go back to an event in Amsterdam, early 1941, as described on Wikipedia. During the winter months of '40/'41, oppression of Jewish citizens of Amsterdam is rising. This leads to open street fights between mobs of both sides. The tensions culminated on February 19th in an ice cream parlor called Koco, where a fight broke out between a German patrol and a mob of regular customers set out to defend the shop. Several arrest were made and the Jewish-German owners where arrested and deported. After roundups in other parts of Amsterdam, the tensions finally lead to the "February strike", the only massive public protest against the persecution of Jews in occupied Europe.

Fig. 3 shows screenshots the exploratory search interface, when searching for the name of the shop, contrasting the three distinct collections. From the streamgraph, the historian can clearly see the most documents that mention the shop are from or refer to early 1941. Focusing on this period of interest, the streamgraph depicted in Fig. 3a shows some references to the shop in the buildup towards this event with the bulk in early 1941. A detailed look at the word clouds for newspapers (Fig. 3b), shows that emphasis is given to the perspective of the police. The significant terms include: robbery, thieves, enforcement, gain access, removed and case[12]. On the other hand, the Wikipedia articles referring to this event focus more on the human interest and broader perspectives. The word cloud in Fig. 3c shows the names of the owners and terms as cause, events, February strike, arrested and owner[13].

[12] In Dutch: *overval, dieven, handhaving, toegang, verschaft, verwijderd, zaak*.

[13] In Dutch: *aanleiding, gebeurtenissen, Februaristaking, gearresteerd, eigenaar*.

Fig. 4: Screenshot showing the first sentences of the Dutch Wikipedia article on the movie "The Ice Cream Parlour" compared to Loe de Jong's article on the events around that parlour.

Diving deeper into the different perspectives, the historian searches for articles related to the section 8.2 of Loe de Jong's fourth book, part II, that covers the events around Koco. He finds Wikipedia articles covering the February strike and related events, but decided to have a more detailed look at the article on the movie "The Ice Cream Parlour". Fig. 4 shows a screenshot of the comparison of the content of this article in comparison with the section written by Loe de Jong. The sentences that focus mostly on the movie are faded out, drawing attention to the parts of the article that describe the events in February 1941.

This worked example illustrates how each collection has different perspectives on an important event in WWII, both in comparing subsets of the collection (Fig. 3) and in comparing two documents (Fig. 4). One can easily think of follow-up questions to explore after this, for example: how does the perspective of newspapers from Amsterdam differ from those in the rest of the country?

6 Conclusion

We connected multiple heterogeneous collections through implicit events, via time and content. We show that we can extract temporal references with satisfactory accuracy and that we can use these references for related article finding. For future work, we identified interesting challenges in extracting temporal references from historical narratives, such as the books of Loe de Jong. Furthermore, we consider our proposed approach for using extracted temporal references to improve related article finding as just a first attempt. Similar problems exist outside the historical domain.

We presented a novel search interface that supports researchers from the humanities in studying different perspectives on WWII. We showed the value of our application through a worked example of how to study different perspectives on an event in WWII. A full-blown evaluation of our application in a user study with historians is planned for future work. While we focused in this work on events and collections related to WWII, our software and approaches can be applied to any kind of digital collections. We release our work as open data and open-source software[14] to foster future research and applications for digital libraries.

Acknowledgements. This research was supported by Amsterdam Data Science and the Dutch national program COMMIT.

[14] The source code is available on https://bitbucket.org/qhp.

References

[1] O. Alonso, J. Strötgen, R. Baeza-Yates, and M. Gertz. Temporal information retrieval: Challenges and opportunities. In *TWAW Workshop, WWW*, 2011.

[2] C.-M. Au Yeung and A. Jatowt. Studying how the past is remembered: Towards computational history through large scale text mining. In *CIKM '11*, pages 1231–1240, New York, NY, USA, 2011. ACM.

[3] M. Bron, B. Huurnink, and M. de Rijke. Linking archives using document enrichment and term selection. In *TPDL '11*, pages 360–371. Springer, 2011.

[4] M. Bron, J. Van Gorp, F. Nack, M. de Rijke, A. Vishneuski, and S. de Leeuw. A subjunctive exploratory search interface to support media studies researchers. In *SIGIR '11*, pages 425–434. ACM, 2012.

[5] L. Byron and M. Wattenberg. Stacked graphs–geometry & aesthetics. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1245–1252, 2008.

[6] S. K. Card, J. D. Mackinlay, and B. Shneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

[7] V. de Boer, J. van Doornik, L. Buitinck, M. Marx, T. Veken, and K. Ribbens. Linking the kingdom: Enriched access to a historiographical text. In *K-CAP '13*, pages 17–24, New York, NY, USA, 2013. ACM.

[8] O. de Rooij, D. Odijk, and M. de Rijke. Themestreams: Visualizing the stream of themes discussed in politics. In *SIGIR '13*, pages 1077–1078. ACM, 2013.

[9] M. Kleppe, L. Hollink, M. Kemman, D. Juric, H. Beunders, J. Blom, J. Oomen, and G. J. Houben. Polimedia. analysing media coverage of political debates by automatically generated links to radio & newspaper items. In *Proceedings of the LinkedUp Veni Competition*.

[10] J. Lensen. De zoektocht naar het midden: nieuwe perspectieven op de herinnering aan de Tweede Wereldoorlog in Vlaanderen en Duitsland. *Internationale neerlandistiek*, 52(2):113–133, 2014.

[11] G. Marchionini. Exploratory search: From finding to understanding. *Commun. ACM*, 49(4):41–46, Apr. 2006.

[12] P. Massa and F. Scrinzi. Manypedia: Comparing language points of view of wikipedia communities. In *WikiSym '12*, page 21. ACM, 2012.

[13] C. Monz, V. Nastase, M. Negri, A. Fahrni, Y. Mehdad, and M. Strube. Cosyne: a framework for multilingual content synchronization of wikis. In *WikiSym '11*, pages 217–218. ACM, 2011.

[14] D. Odijk, O. de Rooij, M.-H. Peetz, T. Pieters, M. de Rijke, and S. Snelders. Semantic document selection. In *TPDL '12*, pages 215–221. Springer, 2012.

[15] J. Pustejovsky, J. M. Castano, R. Ingria, R. Sauri, R. J. Gaizauskas, A. Setzer, G. Katz, and D. R. Radev. Timeml: Robust specification of event and temporal expressions in text. *New directions in question answering*, 3:28–34, 2003.

[16] G. Schreiber, A. Amin, M. van Assem, V. de Boer, L. Hardman, M. Hildebrand, L. Hollink, Z. Huang, J. van Kersen, M. de Niet, et al. Multimedian e-culture demonstrator. In *Sem. Web Chall. at ISWC '07*, pages 951–958. Springer, 2006.

[17] J. Strötgen and M. Gertz. Multilingual and cross-domain temporal tagging. *Language Resources and Evaluation*, 47(2):269–298, 2013.

[18] F. Van Vree. De Nederlandse Pers en Duitsland, 1930 [-] 1939: Een studie over de vorming van de publieke opinie, 1989.