

DOI: <http://doi.org/10.18352/ts.342> TS •> #38, December 2015, p. 29-41.

Content is licensed under a Creative Commons Attribution 3.0 License. - © Maarten van den Bos & Hermione Giffard

Publisher: www.uopenjournals.org. Website: <http://www.tijdschriftstudies.nl>

The Grapevine: Measuring the Influence of Dutch Newspapers on Delpher

MAARTEN VAN DEN BOS & HERMIONE GIFFARD

m.j.a.vandenbos@uu.nl; h.giffard@uu.nl

ABSTRACT

A measure of which newspapers are the most influential is an important foundation for any large-scale study of newspapers. A reliable measure of significance would allow studies to more strongly weigh articles in more influential papers. In this article, we use digital methods to create a partial citation index of a range of Dutch newspapers with a national circulation published in the period from 1920 to 1939. We find that using the digital archive of the National Library of the Netherlands (available through Delpher) enables us to find citations, but the problems of digitization make it impossible to create any index approaching comprehensiveness. Because source criticism – a crucial part of studying history – is made more difficult by digitization, it merits special notice in the digital age.

KEYWORDS

digital humanities, Delpher, newspaper, citation index, source criticism

INTRODUCTION

Source criticism is a central part of writing history. Digital archives make such analysis even more crucial because search engines imply that all of the documents that they display are equal. Thus far, in a given search, all documents are given equal value: an incidence of a key word in document A has the same weight as an occurrence in document B, regardless of how significant A or B are, even relatively. Searches return all of the documents fulfilling the search query, irrespective of their importance. This sort of equality, which we will call “digital equality”, is an illusion produced by the means of digital access and analysis, and so we aim to provide researchers with a tool to confront it.¹

¹ A. Bingham, ‘The digitization of Newspaper Archives: Opportunities and Challenges for Historians.’ *Twentieth Century British History* 21, 2010, 225-231; Marcel Broersma, ‘Nooit meer bladeren? Digitale krantenarchieven als bron.’ *Tijdschrift voor Mediageschiedenis* 14:2, 2011, 29-55.

In more traditional historical research, a researcher's judgment of significance has always depended on his domain knowledge. But with the wide selection of newspapers available through search engines such as Delpher, which grants access to the vast collection of the National Library of the Netherlands, it becomes less likely that all users will be aware of the status of all the sources made available to them digitally. In order to enable researchers to better evaluate their results from digital searches, we propose to create a citation index for newspapers. A citation index measures the significance of a newspaper by counting references made to it in other newspapers. Our assumption is that references to a given publication imply its influence on other publications, so citations are a proxy for the significance of a newspaper. Citation indices have existed since the mid-twentieth century in both science and the humanities. Theoretical work on such indices suggests that they can be defensibly used as a rough indicator of quality, although they do have bias problems.²

In the introduction we will reflect on the importance of significance in digital historical research. This will be followed by an overview of the Dutch media landscape of the interwar period. Then we will explore the possibilities and difficulties of using Delpher to create our citation index for Dutch newspapers published nationally between 1920 and 1939. We take advantage of the possibility of full text searching in the large digital newspaper collection of the National Library of the Netherlands. However, while the search engine enables us to find citations, we find that the problems of digitization make it impossible to create a comprehensive index. The consequence of the complications of using digital search engines is that the result of our experiment is a partial citation index, giving relative rather than absolute information. We use other tools to try to overcome the problems that we encountered in our first searches as well as to better understand our results. Chief among these is Texcavator, a tool for large-scale text mining developed at the universities of Utrecht and Amsterdam. The use of this tool improves our results, but not to such an extent that they meet the standards for a comprehensive citation index. This raises questions with regard to the use of digital newspaper databases in historical research, which we address in our conclusion.

ON THE IMPORTANCE OF SIGNIFICANCE

There are many measures of influence or significance for evaluating newspapers. Different forms of influence are best measured in different ways. We are not interested, as some other projects are, in un-acknowledged borrowing of content.³ Such work is specifically designed to identify "borrowing" when the source is not credited. We have chosen instead to create a citation index, because citing a source is different from copying (parts of) it. Circulation numbers have also been used as a proxy for influence, but they

² Jonathan Cole and Stephen Cole, 'Measuring the Quality of Sociological Research: Problems in the Use of the Science Citation Index.' *The American Sociologist* 6, 1971: 1, 23-29.

³ Viral texts (Mapping Networks of Reprinting in 19th-Century Newspapers and Magazines), uses the unprecedented ability of computers to match texts to find text reuse among different (American) newspapers. (<http://www.viraltxts.org>).

only give a general picture. A citation index is more useful for a historian who is not so much interested in *how many* people read a newspaper, as in *who* read it: in this case, if other newspapers accredit influence to one of their competitors. Citations are more than just an acknowledgement of authorship. They can serve as a guarantee of validity or reliability and also reflect the assumptions of the newspaper in which the citation appears. Citations are useful because they reveal which newspapers are recognized by other newspapers as authoritative. We are particularly interested in capturing this form of influence.

The danger of using digital sources without any measure of significance becomes apparent when we consider the reasons for including any particular newspaper in a digital archive. Inclusion is not entirely or even primarily, as users might think, due to the influence of a newspaper. In the case of Delpher, the influence of newspapers was considered by experts, but was not decisive in the final choice to include a paper in the archive.⁴ An example of a practical matter influencing inclusion is the quality of the ink of a newspaper. This not only determines whether or not a newspaper survives through time, but also if a mechanized scanner can read it. Of course, archives have always selected sources based on criteria that are not made clear to users (and historians are perhaps too liable to assume that the archives which they use are complete), but digital archives imply the equality of the sources they include in a way that other archives do not.⁵ The problem of digital equality is exacerbated by the fact that newspapers are digitized according to values that are unknown to the researcher.

The influence of these kinds of practicalities on archives is, or should be, something which researchers are naturally aware of. But there are important differences between digital and paper archives. Researchers often assume that digital texts are like traditional texts. Indeed, Delpher confirms this expectation by first presenting researchers with pdf images of a newspaper article rather than the flawed digitally recognized version of the text. There are many problems with the digitization process, which make this assumption a dangerously incorrect one. In fact, digital texts, while more accessible, are inaccurate copies of the original paper version.

Digitization conceals a substantial amount of information from researchers which might seem peripheral, but is of crucial importance to good source criticism. In a library, where at first glance all books may appear equal, books carry traces of use on closer inspection: check-out stamps, worn pages, coffee stains, or underlining. If you get a book

⁴ As Delpher reports: ‘the decision of which newspapers to digitize went first to an academic council. The council’s decision was then tested with respect to practical concerns, which dictated the final decision: did the library hold the newspaper, what was the quality of the newspaper (or record) like, had someone else already digitized it and made it publically available.’

(<https://www.kb.nl/organisatie/onderzoek-expertise/digitaliseringsprojecten-in-de-kb/project-databank-digitale-dagbladen/geselecteerde-titels-en-selectieprocedure/selectieproces>) For recent newspapers, the archive also needs to get permission from the publisher. The archive only represents about 10 per cent of the newspapers ever published in the Netherlands.

(<http://www.delpher.nl/nl/platform/pages/?title=vraag+en+antwoord#collecties>)

⁵ Lisa Jardine, *Temptation in the Archives: Essays in Golden Age Dutch Culture*. London: University College London Press 2015 (1-17).

with uncut pages that are still uncut, then probably the book has not been read since it entered the library. Such traces tell researchers something which might only subliminally influence their work. These traces are, nevertheless, crucial. They add to the researcher's understanding of the significance of a certain work. The digital humanities seem to obviate the need to know what you are looking at, but they actually make it more necessary than ever: it is so easy to mistake obscure, previously inaccessible sources for relevant, influential ones.⁶ The sorts of information about significance that are gathered naturally in libraries and archives, have to be explicitly added to digital archives, which collect and exhibit no such signs of use.

For any researcher trying to answer a specific question, not all articles are equally valuable. An article which might be relevant as far as the subject matter is concerned, may still not be revealing to the specific aims of a historian. For a computer, all articles may be fundamentally equal, but some are deemed to be more relevant to a given query (generally calculated by counting keywords) than others.⁷ For a historian, some articles are more significant than others *and* some articles are more relevant. Unlike the question of relevance, significance is something that we can get a quantitative handle on. It is particularly useful to formalize this information, because such knowledge is needed before one can make truly effective use of a search engine.⁸

THE INTERWAR MEDIA LANDSCAPE IN THE NETHERLANDS

To understand the meaning of citations of Dutch newspapers in other Dutch newspapers, we need to understand the fiercely polarized media landscape of the interwar period between the 1920s and 1930s. This contrasts strongly with the nineteenth century, when liberal papers and magazines, like *De Gids*, *Algemeen Handelsblad*, and the *Nieuwe Rotterdamse Courant*, dominated the Dutch newspaper market.

Only in the nineteenth century did newspapers explicitly aim to provide disinterested reporting. In its statement of principles published in 1844, the editors of the *Nieuwe Rotterdamse Courant* specified that even-handedness would be their watchword, even-handedness without blandness. Dominant in the journalistic culture of the era was the idea that political power needed to undergo constant public scrutiny. Journalists were seen as carrying out this scrutiny on behalf of the public.⁹

At the turn of the century, two simultaneous developments significantly changed both the Dutch newspaper landscape and its journalistic culture. Firstly, a number of publications were acquired by large business conglomerates which were interested in profit before content. For example, businessman Hak Holdert bought *De Telegraaf* and changed the editorial policy of the newspaper from impartial and progressive into

⁶ M. Vajcner, 'The Importance of Context for Digitized Archival Collections.' *Journal for the Association of History and Computing* 11 (2008).

⁷ <http://www.delpher.nl/nl/platform/pages/?title=zoekhulp>

⁸ Andreas Fickers, 'Towards a New Digital Historicism? Doing History in the Age of Abundance.' *Journal of European Television History and Culture* 1, 2012, 19-26.

⁹ Frank van Vree, *De Politiek van de openbaarheid*. Groningen: Historische Uitgeverij 2000 (23-30).

“neutral”, a concept consciously kept ambiguous.¹⁰ At the same time, the democratization of Dutch politics and society produced a whole range of new publications of a specific religious or political affiliation.¹¹ After the First World War, newspapers were no longer seen as public watchdogs, but as defenders of interest groups. Reading a newspaper became part of belonging to a certain community. In 1955, more than 75 per cent of Catholic families subscribed to at least one Catholic daily; for Protestant families this was more than 65 per cent. Figures for the interwar period are alleged to have been even higher.¹²

This pattern of subscription did not mean that the news read by a Catholic family was different from the news read by a Protestant family. Yet the tone and point of view of different papers could be diametrically opposed. A good example of the differing interpretations of events by different communities within the Netherlands is shown by the coverage of the fall of the first Colijn-administration in November 1925. From the introduction of universal suffrage in the Netherlands in 1918, confessional parties maintained a small but steady majority in the Dutch parliament. Since the Catholic party was unwilling to form an administration with the social-democrats, they were forced to cooperate with two Protestant parties. In November 1925, this fragile coalition broke up as a result of a member of a small orthodox-protestant party asking for the Dutch diplomatic legation at the Vatican to be revoked. When the member's proposition was accepted by a majority of parliament, including one of the two protestant parties which supported the administration, all of the Catholic ministers resigned in protest.¹³

Every Dutch newspaper covered the political crisis, but their evaluation of the events differed significantly. Catholic newspapers emphasized the importance of the legation to the Vatican in diplomatic affairs and blamed protestant politicians for causing the crisis by refusing to ignore a proposition made by a ‘fringe element’ in their midst because they were afraid of being seen as indulgent towards the Catholic Church. Depending on their party affiliation, Protestants either applauded taking a firm stand against ‘organized Catholicism’s’ attempt to seize power in national and international affairs, or denounced the proposition as dangerous for political stability.¹⁴ This case shows remarkable similarities to classic case studies concerning the coverage of the Russian Revolution of 1917, events in Germany in the 1930s, or the mutiny on a Dutch naval ship

¹⁰ Mariëtte Wolf, *Het geheim van De Telegraaf: geschiedenis van een krant*. Amsterdam: Boom 2009.

¹¹ Van Vree 2000 (31-35).

¹² J.P. Kruyt and W. Goddijn, ‘Verzuiling en ontzuiling als sociologisch proces.’ In A.J.N. den Hollander (ed.), *Drift en Koers. Een halve eeuw sociale verandering in Nederland*. Assen: Van Gorcum 1968, 227-263 (242).

¹³ Herman Langeveld, *Dit leven van krachtig handelen. Hendrikus Colijn 1869-1944, 1, 1869-1933*. Amsterdam: Balans 1998 (247-287).

¹⁴ This is based on an analysis of 407 articles from national newspapers that appeared on the case in November and December 1925 in national and local newspapers. Retrieved with Delpher, search query: ‘gezantschap Vaticaan’, accessed 27-7-2015. All articles were read manually.

in 1933, which also emphasize the small but significant differences between interwar newspapers of different affiliations.¹⁵

The fact that different newspapers were divided by affiliation complicates any determination of the significance of Dutch newspapers between 1920 and 1939. A Catholic daily would hardly ever quote a Protestant newspaper as an important or reliable source of information; indeed, it might sooner conceal the source of its information. Praise for other Catholic dailies was also rare in Catholic newspapers; frequent positive references to other newspapers could lead readers to change their subscriptions. It is easy, however, to find examples of Catholic dailies dismissing Protestant ones (or the reverse) for misinforming the public. They blamed other newspapers for misinforming their public, but, by doing so, in fact offered a testament to the significance of the other newspaper. A citation index takes advantage of this fact.

Importantly, citations often only acknowledge the significance of particular papers in particular realms. For example, economic news was often drawn from the *Algemeen Handelsblad* and citing this paper served as a guarantee of accuracy. But when a paper with a clear political position was cited, it was mostly cited as a leading representative of a particular position. In this way, newspapers both acknowledged and reinforced judgments about their competitors among their readers. Looking at how and when newspapers are cited by other newspapers gives us further insight into the assumptions that lie behind citations. The way in which newspapers refer to other newspapers gives us a key to their relationship to public discourse and offers part of the answer to the age-old question of whether newspapers shape or only reflect public debate. We find that newspapers were cited because they were the public mouthpieces of different groups in society. Judged representative, they were often cited to enable another paper to engage with the opinions of a particular group (generally in order to criticize them).

CREATING A CITATION INDEX USING DELPHER

We set out to create a citation index for Dutch newspapers using Delpher and the digital newspaper collection of the National Library of the Netherlands. Using the metadata to provide gross filters, we limited our searches to articles in Dutch newspapers with a national circulation.¹⁶ We first looked for references to the titles of each of the national newspapers published in the period between 1920 and 1939 by entering the name of a newspaper on Delpher. A quick search produced an astronomically high number of hits. Searching for “*De Standaard*”, the most prominent Protestant daily in the Netherlands in

¹⁵ Th. Stoelinga, *Russische revolutie en vredesverwachtingen in de Nederlandse pers*. Bussum: Fibula 1975; J.C.H. Blom, *De muiterij op De Zeven Provinciën. Reacties en gevolgen in Nederland*. Bussum: Fibula 1975; F. van Vree, *De Nederlandse pers en Duitsland. Een studie over de vorming van publieke opinie*. Groningen: Historische Uitgeverij 1989.

¹⁶ In limiting our searches in this way, we rely on the accuracy of the metadata of the database which we know is not always correct, but we do not know how often it is incorrect. The ability to judge uncertainty or error rates is a major problem for scholars attempting to use digital archives. Cf. Myriam C. Traub, Jacco van Ossenbruggen, and Lynda Hardman. ‘Impact Analysis of OCR Quality on Research Tasks in Digital Archives.’ Working-paper Centrum Wiskunde & Informatica, Amsterdam.

the 1920s, returned 7,020 articles in national newspapers for the 1920s and 4,757 for the 1930s. A search for “*De Telegraaf*” resulted in 11,181 and 13,613 articles respectively.

Close reading of the highest ranked results made us immediately aware of two problems. The first problem, which inflated the numbers, was that the titles of our selection of newspapers are commonly used Dutch words (see Figure 1). This affects some newspapers more than others. Therefore, it is difficult to take this problem into account in a broad, comparative search.




Figure 1: From *De Telegraaf* 27.01.1979. This example shows that the title of a newspaper, such as *De Standaard*, can be a common Dutch word, ‘standaard’ [stand]. Simply searching for newspaper titles is therefore problematic. Moreover, in its metadata this text was identified as an article. This example shows how even more precise searching (as discussed below using “volgens”) will always produce false positives.

Our reliance on computer searches thus forced us to limit our search to newspapers with titles which did not consist of frequently used words, such as *De Maasbode*, if we wanted absolute results. Named Entity Recognition (NER) is not yet sophisticated enough to determine if common words are used as proper nouns. Although limiting our search to these newspapers could perhaps be justified, it is problematic in that it moves decision-making about research away from the researcher. We did not want to limit our research in this way, because it would exclude some of the most significant newspapers (*De Standaard* and *De Telegraaf*) from our citation index. Furthermore, because we have no way to measure our errors, we do not know if different newspaper titles are biased in particular ways. We were thus unable to measure, except anecdotally, if “Standaard” is more frequently misrecognized by automated software than “Maasbode”, for example. A second problem we came across was the fact that newspapers often use their own name in their own articles. Even more problematic was the fact that the title banner of some newspapers has been digitized and categorized as an article, so the top hits tended to be those banners (see Figure 2).

De Telegraaf

🔍 Kranten
De Telegraaf

Uitgebreide zoeken



1920-1929

< Resultaten voor **De Telegraaf** | 1920-1929 | Landelijk | Artikel

De Telegraaf

07-11-1929

🔍 Zoomen en navigeren





BUREAU DER REDACTIE:
ST. NICOLAASSTRAAT 37-43
AMSTERDAM

TELEFON 56, 5000 en 51000
TELEGRAM INTERCOMMAAL LETTER T.
POSTBUREAU 3, 2000
GEOGR. AMSTERDAM N. O. 2008

VOORBEREID 7 NOVEMBER 1929

**BURGEMEESTER BOSS
TEGEN ZIJN VERVANGER.**

**SKLAREK-ZAAK LEIDT
TOT CONFLICT.**

**ANDERE EIZING: BOM BEKENDING
„JEWISH“.**

**DE VLEGRAMP IN
ENGELAND.**

De Telegraaf

HET MEEST VERSPREIDE GROOTE DAGBLAD
waars opgenomen bij „Assenrechtelijke Courant“ (Nieuw Jaargang)
Vrijdag 6 N. O. Bureau de l'Amateur pour la Presse. Amsterdam. P. 5, 2, 1000000

● DRIETWEEFDE - DRIET BLADEN

ENGELANDS SEEDERLINGEN.

POOLSEKOP KOFFIEHUIS
BROUWERIJ
Bakker J. Nieuwenhuis
Wolven Schied.

**Hotel SUISSE, A. van
den Bosch.**

**In de RESTAURANTIE v/v
CENTRAAL-STATION**

OGNIBENI

GREENWOOD-HORTUS

WIEVERWACHTING.

**CONFERENTIE ZONDER
MIJNEGENAAR.**

**AMSTERDAM VERANDERT NIET
LEGEN: HUNDE EIZING.**

Our revised search query seemed promising, but the search for “*De Telegraaf*” PROX *in* returned many more results than other citation words, because the Dutch word “in” is a common preposition – implying that it is not solely (or even primarily) used to introduce a reference. So while this refined search is better, it still has numerous problems, making it hard to make sense of results, let alone compare them.

CREATING A CITATION INDEX USING TEXCAVATOR

¹⁸ *Het Volksdagblad* (24-12-1938).

In order to avoid including self-referencing in our index, we wanted to specify that a citation could not occur in the title of a newspaper. In order to limit our search in this way, we turned to Texcavator, a digital tool developed to search and analyze digitized newspapers of the National Library. Texcavator is not only able to search the newspapers with more complex queries than Delpher, but can also be used to analyze and understand search results. Key among its functionalities are timelines and word clouds (as illustrated below), which allow for quick analysis.¹⁹ Because the tool uses a different search syntax than Delpher, our query string was slightly different from the one we used for Delpher, using “~” rather than “PROX”. Elastic search, the syntax used in Texcavator, allows the user to specify the interval for proximity, so we used five words. Thus, “volgens”, for example, has to appear within five words of “Telegraaf”. The program also allows the researcher to limit a search using the metadata provided by the database, so we once again configured our search for articles in national newspapers. After we made the changes described above, our search string looked like this: “*Telegraaf volgens*”~5 - *paper_dc_title: “Telegraaf”*. This query excludes all articles that contain *De Telegraaf* in combination with *volgens* that were published in *De Telegraaf* itself, thus avoiding self-referencing.

Searching for citations using the words “volgens”, “zoals”, and “aldus” showed that *De Telegraaf* was referenced 1,412 times by other newspapers between 1920 and the end of 1939. In comparison, the Catholic daily *De Maasbode* was referenced 388 times and the liberal *Algemeen Handelsblad* 1,077 times. Based on these figures, it can be concluded that *De Telegraaf* was the most frequently cited newspaper of the interwar period. These figures support the argument of previous press historical research that liberal Dutch newspapers such as the *Algemeen Handelsblad* lost their prominence during the interwar period, while “neutral” papers, such as *De Telegraaf*, became more important. In order to further analyze our results, we used Texcavator to make a timeline of the articles referring to *De Telegraaf* as well as a word cloud, which visualizes the most frequently used words in these articles (see Figure 3).

¹⁹ Texcavator is not yet publicly available, but is being made so. On the possibilities offered by Texcavator: J. van Eijnatten, T. Pieters and J. Verheul, ‘TS Tools: Using Texcavator to map public discourse.’ *Tijdschrift voor Tijdschriftstudies* 35, 2014, 59-65.

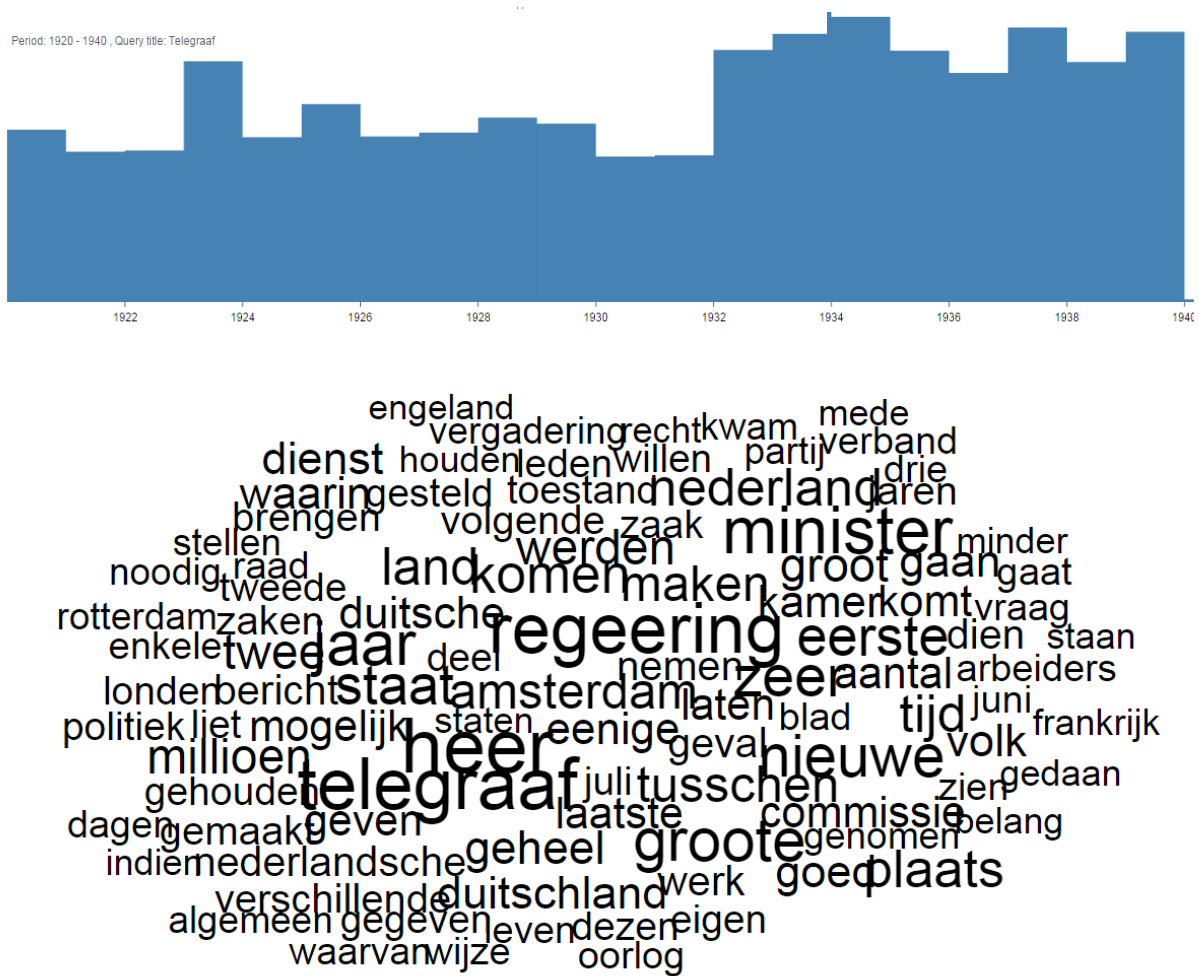


Figure 3: The timeline and word cloud generated by Texcavator using the articles that referenced *De Telegraaf*.

Word clouds show how digital tools can provide more information about citations than just a number (the relative significance of various newspapers is not taken into account in Figure 3). A word cloud visually represents the content of the articles yielded in response to a search query. The size of the words in the cloud reflects the number of times that a specific word appears in the returned data set. The cloud in Figure 3, generated from articles that referenced *De Telegraaf*, suggests that the references to that newspaper were mostly in articles about politics, since words like “regering” [government/administration] and “minister” [minister] were used more frequently than other words. When we similarly generated a word cloud using the articles referencing the Catholic *De Maasbode*, some of the largest words referred to the Catholic community (see Figure 4). The two clouds are not directly comparable with respect to font size.



Figure 4: The word cloud generated by Texcavator from articles citing De Maasbode.

These kinds of results have to be taken with a grain of salt. Firstly, including citations of *Het Volk* and *De Tijd* in our citation index remains difficult, even in conjunction with citation words, because “volk” [people] and “tijd” [time] are very common words. A search for *Het Volk* in the period that we are interested in yielded 4,960 hits, few of which are references to the newspaper. References to *De Standaard* are subject to a similar problem. This newspaper changed chief editor in 1920. The new chief editor of the newspaper, Henrikus Colijn, was probably the most influential politician of the Dutch interwar period.²⁰ This alone might explain the 1,538 references to the title that we found using Texcavator. However, the timeline of the results for *De Standaard* and “volgens” reveals a further problem (see Figure 5). It shows a burst of articles in 1933, the year when the Conference of London marked the end of the gold standard. Therefore, it is likely that the burst does not reflect an increase in the significance of *De Standaard*, but rather reveals the extensive coverage of the conference and the vigorous political debate in the Netherlands on whether or not to follow its conclusions.²¹ If we allow syntactic confusion between the names of leading newspapers and typical Dutch words of the interwar period to influence our research, then we greatly reduce the number of newspapers for which we can produce a measure of relative significance.

²⁰ Henk te Velde, *Stijlen van leiderschap. Persoon en politiek van Thorbecke tot Den Uyl*. Amsterdam: Wereldbibliotheek 2002 (10-11, 107-152).

²¹ See: Jan Luiten van Zanden, *Een klein land in de twintigste eeuw. Economische geschiedenis van Nederland 1914-1995*. Utrecht: Het Spectrum 1997 (151-161).

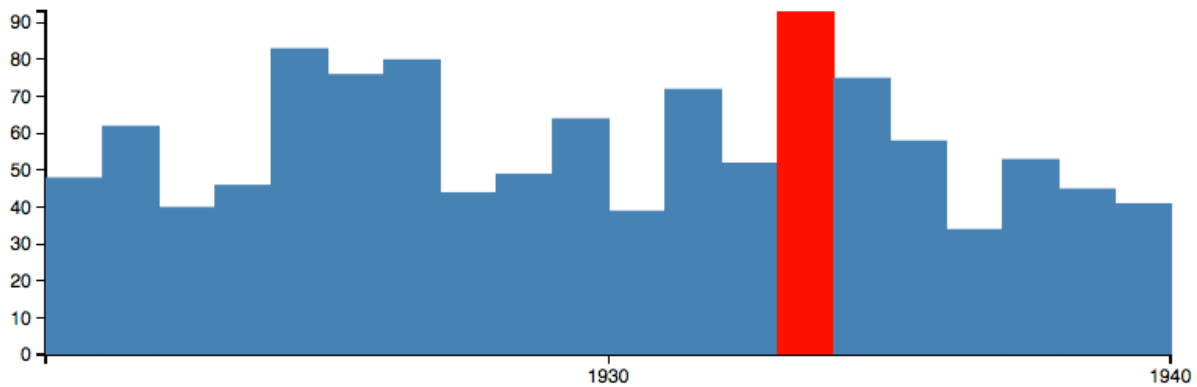


Figure 5: Timeline generated by Texcavator showing the chronological spread of articles using the words “Standaard” and “volgens” but excluding uses in the newspaper *De Standaard*. The red line, known as a burst, represents many articles – the column is not to scale, because if it were, it would render the others invisible. It is likely a reflection of the vigorous Dutch debate about the gold standard, not of references to the newspaper.

In order to check the partial results that we were confident in, we compared our measures of citations to citations in different datasets. For example, we looked at the mentions of newspapers in the Dutch parliament, since its minutes have been digitized.²² Here, too, *De Telegraaf* was the title most frequently mentioned, with 688 references between 1920 and 1939. *De Standaard* was mentioned 438 times and *De Maasbode* 365 times. The *Algemeen Handelsblad* was mentioned 640 times, however, which seems to undermine our previous conclusion about the position of liberal newspapers in the interwar period. Still, close reading of the articles shows that “Telegraaf” was often used in parliament to describe changes in Telegraph-laws rather than to cite the national newspaper, which makes the usefulness of this search questionable. Other possible comparisons, such as to the Digital Library of Dutch Literature (Digitale Bibliotheek der Nederlandse Letteren, DBNL), are frustrated by search interfaces that only allow searches for single keywords.²³

CONCLUSION

Digital tools give researchers access to more newspapers than ever before, both major and marginal, so producing a relative measure of significance on a similarly large scale is imperative. A citation index will help to make search engines more effective, and go some way towards fixing the problem of digital equality. Digital researchers no longer pre-select the newspapers that they will study based on criteria deduced from knowledge of the period they study. Instead, a search engine selects sources for the researcher, using its notion of relevance, which may often be misguided. Digital accessibility is clearly changing which titles and articles are being used by researchers, and this is an acute problem for scholars who lack experience to judge whether or not a given quote is representative of community-wide sentiment. This kind of knowledge is of crucial

²² Digitized minutes of Dutch Parliament can be found and searched via: www.statengeneraalidgitaal.nl (accessed 27-7-2015).

²³ Searching the DBNL database is possible via the National Library Labs web-portal: <http://lab.kbresearch.nl/find/DBNLNgrams#> (accessed 27-7-2015).

importance when using a tool such as a computer, which ambivalently presents good and bad quotes together.

Every time designers incorporate a specific definition of a value into digital tools, more interpretation and decision-making is taken away from the user. When researchers are not actively making decisions about significance, they are likely to be less critical of the search results presented to them. So while it is important to think about ways to incorporate new information into search tools, it is also essential to remember that if this kind of information is hidden from them, users can make, or are forced to make, fewer judgments about the specific needs of their research problem.

In this paper, we tried to use digital tools to create a citation index in order to establish an empirically based standard for significance. Because of the many problems that we faced using digital search tools, we were only able to find the relative significance of a few major newspapers. Our inability to answer our initial question brought to our attention many shortcomings of digital archives that remain to be overcome. As long as source criticism remains opaque for digital sources, and admittedly our suggested measure of significance only partly addresses this problem, digital archives will continue to contain too many problems to be used to provide quantitative evidence for historical arguments.

In general, both the fervent exponents and the avid opponents are likely to overestimate the capability of digital tools. Therefore, we need more theoretical reflection on the possibilities and problems of digital archives. Significance is just one measure necessary in order to effectively use the abundance offered by these archives. Our experiment has shown that it is very difficult, if not impossible, to be comprehensive using digital tools. We need to put much more effort into solving the problems, including those of digital equality, which frustrate interpretation.²⁴ An important step towards facilitating the use of digital sources in the same way as any other sources is to really understand both the many sources made available in digital archives and the way in which they are structured and presented. Finding a reliable way to judge significance is part of this indispensable exercise.

•> MAARTEN VAN DEN BOS & HERMIONE GIFFARD *are both postdoc researchers in the Department of History and Art History at Utrecht University for the Asymmetrical Encounters: E-Humanity Approaches to Reference Cultures in Europe, 1815-1992. A project on digital text mining funded by HERA.*

²⁴ Dan Dixon, 'Analysis Tool or Research Methodology: Is There an Epistemology for Patterns?' In David Berry (ed.), *Understanding Digital Humanities*. Basingstoke: Palgrave 2012, 191-209.