The cultural-historical value of and problems with digitized advertisements. Historical newspapers and the portable radio, 1950-1969

JESPER VERHOEF

j.verhoef@uu.nl

ABSTRACT

This article demonstrates how a digital newspaper archive such as Delpher offers new possibilities to do justice to the value of newspaper advertisements when conducting historical research. A case study into the way advertisements tried to cater to youngsters in portable radio advertisements (1950-1969) will illuminate how distant, semi-distant, and close reading can further historical enquiry. This case study, at the same time, reveals a major shortcoming of these digitized advertisements, namely the way they are currently indexed – classified advertisements in particular. The article offers two computational approaches that will result in a more fine-grained indexation, and urges the National Library of the Netherlands to experiment with these approaches as well as with crowd sourcing. Only after these measures have been taken will researchers be able to use the full potential of advertisements on Delpher's newspapers.

KEYWORDS

advertisements, classifieds, historical newspapers, digital newspapers, portable radio

Over the last decades, several studies have highlighted the value of historical advertisement for cultural-historical research.[1] Nevertheless, scholars discussing the boons and banes of digital archives have often disregarded the possible value of

---

[1] Renowned examples are T. J. Jackson Lears, *Fables of Abundance. A Cultural History of Advertising in America.* New York: BasicBooks 1994 and Roland Marchand, *Advertising the American dream. Making Way for Modernity, 1920-1940.* Berkeley etc.: University of California Press 1985. For the Netherlands, see Wilbert Schreurs, *Geschiedenis van de reclame in Nederland, 1870-1990.* Utrecht: Het Spectrum 1989. An example of how an advertisement archive may be put to use is offered by Wiebke Schulz, Ineke Maas and Marco H.D. van Leeuwen, 'Employer's Choice. Selection through Job Advertisements in the Nineteenth and Twentieth Centuries.' *Research in Social Stratification and Mobility* 36, 2014, 49–68.

advertisements in digital archives. When Nicholson emphasizes the upsides of using digitized newspapers for historical research, he rightly states that the information in newspaper articles constitutes 'the things that we can learn about the people and the society who produced and read it'.[2] Yet he neglects to mention that advertisements fulfill a similar function; they are 'cultural indicators' as well.[3] Therefore, when using historical newspapers, as Blevins rightfully argues, one should avoid 'biasing certain modes of reading over others', as if editorials were by definition more important to readers than (small) advertisements.[4] Trtovac and Dakic underline that newspaper advertisements should not be overlooked, for 'the information contained in advertisements […] fully reflects the spirit of the past, [as it] indicates development of certain industries, but also covers all aspects of cultural and social life […]'.[5]

This brief introduction does not seek to make the point that advertisements are more valuable than articles; it does indicate, however, that advertisements have merit in their own right. To substantiate this claim, this article explores how the advertisement corpus of a vast newspaper archive like Delpher provides a unique lens through which to analyze cultural-historical questions in a systematic way. It does so by means of a case study on advertisements accompanying the diffusion of the portable radio – sometimes called "portable", "transistor", or "transistor radio" – in the Netherlands in the 1950s and 1960s.[6]

The article consists of two parts. The first section shows how advertisements on Delpher helped to verify and, at the same time, nuance the proposition that youth played a central part in advertisements for portables in the 1950s and 1960s. In doing so, this section highlights how Delpher enables researchers to combine so-called distant, semi-distant, and close reading. The second section examines one major disadvantage of the advertisements on Delpher, namely the fact that separate advertisements are frequently indexed as one. I then suggest approaches to overcome this shortcoming. The National Library of the Netherlands (KB), I maintain in the concluding remarks, should put these into practice.

---

[2] Bob Nicholson, 'The Digital Turn.' *Media History* 19:1, 2013, 59–73 (61). Similarly, Zaagsma mentions newspapers but omits advertisements. Gerben Zaagsma, 'On Digital History.' *BMGN - Low Countries Historical Review* 128:4, 2013, 3–29. Jacobs, lastly, analyzed Dutch digitized newspaper articles, yet refrained from looking into advertisements. Annelies Jacobs, *Het geluid van gisteren. Waarom Amsterdam vroeger ook niet stil was*. Maastricht: Universitaire Pers Maastricht 2014.

[3] Daniel Pope, 'Making Sense of Advertisements.' *History Matters. The U.S. Survey Course on the Web* 2003, <http://historymatters.gmu.edu/mse/Ads/> (consulted on July 2, 2015).

[4] Cameron Blevins, 'Space, Nation, and the Triumph of Region. A View of the World from Houston.' *Journal of American History* 101, 2014, 122–147 (128).

[5] Aleksandra Trtovac and Natasa Dakic, 'Bringing Historical Newspaper Advertisement into Research Focus - Europeana Newspapers Project.' *Преглед НЦД* 25, 2014, 2–10 (9).

[6] For reasons of clarity, this article will consistently use the term "portable radio" or, in short, "portable".

In 1966, Philips promised its retailers two photos of famous musicians for every portable radio they ordered. The company stressed that these could be handed out to 'music loving teenagers', which – it was stressed – 'was every teenager!' Furthermore, Philips coined one of its portables 'Popmaster', to highlight that these devices were well suited to listen to new, popular music.[7]

Manufacturers and retailers, these examples show, discovered the youth as potential buyers of the portable radio. They tried to advance the diffusion of the devices by catering to the alleged desires of teenagers and the bourgeoning youth culture of the 1960s.[8] Various scholars have claimed that this youth culture was both reflected in and furthered by the use of portable radios, for the portable reassured 'listeners of their autonomy and identity anytime, anywhere'.[9] Producers had good reason to commence marketing aimed at teenagers, since contemporary market research and sales figures underlined that teenagers were of increasing importance to the radio market.[10] Even the government tried to vie for the favor of young radio listeners by founding the new radio station Hilversum 3 in 1965, geared towards the presumed tastes of youngsters. This was a somewhat belated response to the success of radio Veronica in capturing teenagers' attention after it had started broadcasting in 1960.[11]

---

[7] W.O. de Wit, 'Radio tussen verzuiling en individualisering.' In J.W. Schot and W.O. de Wit (eds.), *Techniek in Nederland in de twintigste eeuw. Deel V: Transport en communicatie.* Zutphen:Walburg Pers 2002, 203–229 (225).

[8] Cf. Heike Weber, *Das Versprechen mobiler Freiheit. Zur Kultur- und Technikgeschichte von Kofferradio, Walkman und Handy.* Bielefeld: Transcript Verlag 2008 (126–30).

[9] Heike Weber, 'Taking Your Favorite Sound Along: Portable Audio Technologies for Mobile Music Listening.' In Karin Bijsterveld and José van Dyck (eds.), *Sound Souvenirs. Audio Technologies, Memory and Cultural Practices.* Amsterdam: Amsterdam University Press 2009, 69–82 (74). See also Andreas Fickers, *Der "Transistor" als technisches und kulturelles Phänomen: die Transistorisierung der Radio- und Fernsehempfänger in der deutschen Rundfunkindustrie 1955 bis 1965.* Bassum: Verlag für Geschichte der Naturwissenschaften und der Technik 1998 (89). Schiffer maintains that the portable 'became a metaphor for freedom and independence' for a young generation. Michael Brian Schiffer, *The Portable Radio in American Life.* Tucson: University of Arizona Press 1991 (181). Cf. Hans Righart, *De eindeloze jaren zestig. Geschiedenis van een generatieconflict.* Amsterdam: De Arbeiderspers 1995 (121).

[10] They were buyers, or were influential in making their parents buy a portable. W.O. de Wit, 'Producenten, consumenten en intermediairen. De introductie en diffusie van de transistorradio in Nederland in de jaren vijftig en zestig.' In Y. Segers et al. (eds.), *Op weg naar een consumptiemaatschappij. Over het gebruik van voeding, kleding en luxegoederen in België en Nederland (19e en 20e eeuw).* Amsterdam: Aksant 2002, 181–201 (193).

[11] In 1965, Veronica was at the peak of its popularity, according to Marja Roholl, 'Uncle Sam: An Example for All? The Dutch Orientation Towards America in the Social and Cultural Field, 1945-1965.' In Hans Loeber (ed.), *Dutch-American Relations 1945-1969. A Partnership, Illusions and Facts.* Assen: Van Gorcum 1992, 105–152 (144); Auke Kok, *Dit was Veronica. Geschiedenis van een piraat.* Amsterdam: Rap 2007 (96–115); De Wit 2002, 'Radio tussen verzuiling en individualisering' (226–27); Huub Wijfjes, 'Radio als stemmingsregelaar, 1960-2010.' In Bert Hogenkamp, Sonja de Leeuw and Huub Wijfjes

With the advertisements on Delpher at my disposal, I was able to systemically test the hypothesis that the portable was specifically marketed at teenagers: through the lens of newspaper advertisements I was able to examine to what extent and how advertisements mentioning portable radios explicitly reached out to youngsters.

Because the period between 1950 and 1969 represents the rise and fall of newspaper discourse on the portable radio, this will be the period under scrutiny here.[12] The first step is to gather information through 'distant reading', which has been defined as understanding a corpus 'not by studying particular texts, but by aggregating and analyzing massive amounts of data' – or, even simpler, as reading 'the archive from a distance'.[13] In order to do so, one first has to establish a baseline to contrast the findings to. In this case, this is the sum of advertisements mentioning an equivalent of the portable – which is established through a broad keyword search, resulting in 11,265 advertisements.[14] The distribution is showed in figure 1, which demonstrates that the majority of advertisements were published in the 1960s.
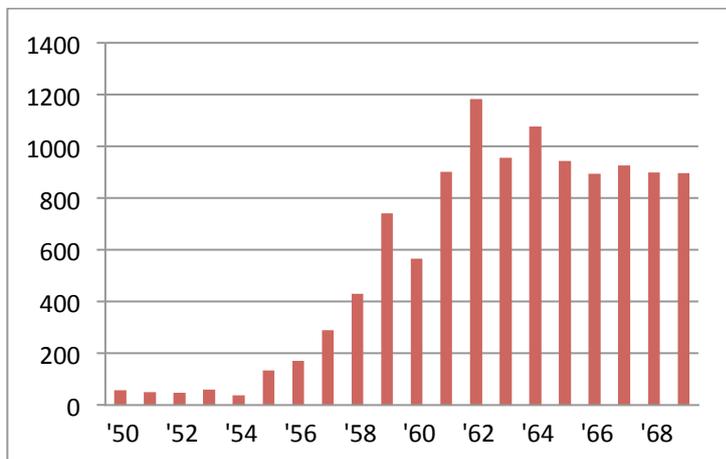

Figure 1: Distribution of digitized newspaper advertisements mentioning "portable radio" or its equivalents, 1950-1969.

To identify whether these advertisements did indeed try to reach out to teenagers, one could, subsequently, narrow this query by adding specific words associated with youth culture. In this case, I narrowed the query by adding Veronica or Hilversum 3. Surprisingly, Delpher revealed – without having read a single advertisement – that only 114 (Veronica: 94; Hilversum 3: 20) portable radio advertisements mentioned these

---

(eds.), *Een eeuw van beeld en geluid. Cultuurgeschiedenis van radio en televisie in Nederland*. Hilversum: Nederlands Instituut voor Beeld en Geluid 2012, 192–225 (200–7).

[12] Jesper Verhoef, 'Het moderne onheil en het lawaaivraagstuk. De draagbare radio en beheerste modernisering, 1955-1969.' *Tijdschrift voor Geschiedenis*, forthcoming.

[13] Kathryn Schultz, 'What is Distant Reading?' *New York Times,* June 24, 2011; Bob Nicholson, 'Counting; or, How to Read Victorian Newspapers from a Distance.' *Journal of Victorian Culture* 17, 2012, 238–246 (246).

[14] The query was (((TRANSISTOR OR DRAAGBARE OR PORTABLE OR TRANSPORTABLE) AND RADIO) OR ZAKRADIO). Colonial newspapers are neither included in this query, nor in any other query in this article.

radio stations.[15] Even when taking OCR errors into account, these numbers are not nearly enough to warrant the claim that producers and retailers targeted young people intensely – that is, at least not by mentioning these radio stations. Other possible combinational queries, e.g. with words like "scooter" (*brommer*), need not divert attention from the main point here: distant reading advertisements through keyword searches on Delpher may offer swift, tentative information about the underlying data.

Distant reading alone only aids in answering historical questions or testing hypotheses to a certain extent. Looking at and analyzing the actual advertisements remains crucial for more in-depth insights.[16] Dealing with extensive corpora of digitized advertisements as the one under scrutiny can therefore still be a daunting challenge. Interestingly, Delpher has a built-in function which greatly simplifies this task. When hovering over a result, it usually – it is unclear why not always – zooms in on the part of the advertisement that contains one or more of the search words and gives an excerpt in which the search phrase is highlighted in yellow (see figure 3). This is what I call 'semi-reading'. Especially when dealing with serial advertisements, comprised by the same text and image time and again, this speeds up the process of gauging *distinct* advertisements considerably. In case of small advertisements, such as classifieds, the excerpt sometimes renders clicking (and, consequently, having to wait for the new page to open) redundant.

The last step entails reading all the actual advertisements "manually" – sometimes dubbed 'close reading'. This corroborated my hypothesis: portable radio producers did indeed approach the Dutch youth from time to time. One ad for instance reached out to students who had passed their high school exams. It displayed two teenagers with Philips portables and stated: 'SUCCESSFUL. And how! First we succeeded in passing our finals, afterwards in buying ourselves gifts.'[17] By the same token, a Frisian retailer advertised that he offered a portable to the student with the highest average exam grade.[18] From the late 1950s onwards, portable manufacturers included younger kids in the target group, by offering 'portable radio construction boxes'.[19] The popularity of the portable radio as a consumer item for youngsters was underlined by the fact that companies offered portables as prizes in contests for children, such as coloring contests.[20]

---

[15] The queries were (((TRANSISTOR OR DRAAGBARE OR TRANSPORTABELE OR PORTABLE) AND RADIO) OR ZAKRADIO) AND (VERONICA) and (((TRANSISTOR OR DRAAGBARE OR TRANSPORTABELE OR PORTABLE) AND RADIO) OR ZAKRADIO) AND ("HILVERSUM 3" OR "HILVERSUM III").

[16] This is in line with what most scholars in the field of Digital Humanities claim, namely that distant and close reading should accompany and reinforce one another. E.g. Melvin Wevers and Pim Huijnen, 'Mapping America in Public Discourse.' In Michal Peprník and Matthew Sweney (eds.), *America in Foreign Media. Proceedings of the 19th international colloquium of American studies*. Olomouc: Palcký University Olomouc 2014, 109–125 (121).

[17] In Dutch it said: 'GESLAAGD. En hoe! Eerst voor ons examen, daarna bij het kopen van een cadeau.' *Nieuwsblad van het Noorden*, June 22, 1959.

[18] *Leeuwarder Courant*, June 9, 1961.

[19] For around 15 guilders apiece, these were rather cheap compared with the average portable radio, which cost around 100 guilders. *De Telegraaf*, December 17, 1959; *Het Vrije Volk*, December 20, 1960.

[20] For instance in *Nieuwsblad van het Noorden*, July 2, 1959 and *Nieuwsblad van het Noorden*, April 1, 1965.

However, the vast majority of advertisements addressed adults. Some advertisements addressed children through their parents: they were spurred to buy their children a portable.[21] The rising average income of Dutch youngsters was apparently not enough reason for producers to speak to them more often through newspaper advertisement.[22] Why they did not do this more often falls beyond the scope of this article. It suffices to conclude that distant reading followed by close reading the advertisements nuances the supposed significance of the youth for portable radio producers.[23]

INDEXATION AND SEGMENTATION

When one combs through the portable radio advertisements, a major disadvantage of the advertisements corpus on Delpher is laid bare: its optical layout recognition (OLR) frequently groups separate advertisements together as one document.[24] The excerpt in figure 2 demonstrates that distinct classified advertisements (*rubrieksadvertenties*) that are placed next to and below each other in the same section are typically perceived as one advertisement.

Advertentie

nden geschikt en een toilet, 3 platen. J. Ph. Aret, Adrién Mildersstraat 78. PHILIPS' nieuwe draagbare radio te koop, 2 weken oud. Van ƒ75 voor ƒ 55. Boekhandel G. P. v. Landegem, Beu ...

Krantentitel	Rotterdamsch nieuwsblad
Datum	03-10-1940

Alle details

Figure 2: Currently, Delpher renders multiple classified advertisements as one segment.

This means that the actual number of advertisements on Delpher is a multitude of the one depicted. It is, however, currently impossible to know exactly how many classified ads were actually published. This complicates research into these specific ads, e.g. inquiring

---

[21] E.g. *Het Vrije Volk*, December 20, 1960 and *Friese Koerier* July 10, 1962.

[22] On average, the income of the working youth rose with 150 percent between 1955 and 1965. Righart 1995 (132).

[23] A point made by Wijfjes as well, see Wijfjes 2012 (197).

[24] The seemingly universal flaw that one can only employ keyword searches, and not 'image searches' – save it for tagged images – will not be touched upon here. See, e.g., Marjo Markkula and Eero Sormunen, 'End-user Searching Challenges Indexing Practices in the Digital Newspaper Photo Archive.' *Information Retrieval* 1, 2000, 259–285; Ka-Ping Yee et al., 'Faceted Metadata for Image Search and Browsing.' In Cockton et al. (eds.), *Proceedings of the ACM CHI 2003 Human Factors in Computing Systems Conference*. New York: ACM 2003, 401–408.

how advertising for romantic partners in newspapers changed over time.[25] Even worse, this OLR imperfection results in so-called false positives: advertisements which are falsely included in a sub-corpus. This is especially problematic when dealing with small classified ads, which were mostly placed by consumers advertising services and products. Since the false positives are so numerous, one has to look into every advertisement to check whether it indeed is about the portable radio. Consequently, the inadequate indexation of advertisements thwarts further analysis by means of digital tools of sub-corpora of advertisements, such as topic modeling, a process that identifies 'clusters of words – topics – that often appear in the same document together'.[26] Since Delpher does not offer these tools on its website, this article will not delve into the specific problem this indexation causes when such tools are used.[27] It suffices to note that all the extra data not related to the original query results in "noise" and as such skews and obscures the outcomes of the tool that is being used.

We need a more fine-grained indexation to use the full potential of advertisements available via Delpher. Two consecutive steps are recommended to improve the current situation. The first step would be for Delpher to distinguish classifieds from "regular" advertisements. Classifieds are advertisement that can be recognized by two features: since advertisers pay per word or even token, they contain a confined number of tokens per advertisement per section (presumably stable over longer stretches of time, depending on the policy of the newspaper), and they are placed together in a section. In the period under scrutiny, the *Leeuwarder Courant* called these sections 'signposts' (*wegwijzers*), whereas *Het Vrije Volk* dubbed them 'sowers' (*zaaiers*). Regular advertisements are usually larger (both in size and number of words used), placed by companies, retailers and the like. Besides, they are often marked by a regular phrase, such as *Ingezonden mededeling*, to inform readers that it is not editorial content.

Based on a sample of digitized *Washington Times*, Allen and Hall discuss one way of generating this distinction between classified and regular advertisements computationally:

> Specifically, we developed word lists for each of the five types of sections [a.o. Classified advertisements, Sports, and Society] on which we were focusing. We then found the average frequency for each of the terms across all the pages for that month. Next, we compared the frequencies separately for each page to the frequency for the entire month. If the page frequency exceeded the overall monthly frequency by a large multiplier (e.g., 30

---

[25] Debra L. Merskin and Mara Huberlie, 'Companionship in the Classifieds. The Adoption of Personal Advertisements by Daily Newspapers.' *Journalism & Mass Communication Quarterly* 73, 1996, 219–229.

[26] Robert K. Nelson, 'Mining the dispatch.' *Mining the dispatch*, 2013. <http://dsl.richmond.edu/dispatch/pages/intro> (consulted on August 14, 2015).

[27] Currently, researchers have to request the newspaper data from the National Library of the Netherlands, in order to use tools such as MALLET (topic modeling) to analyze this data. Another example of a tool is Texcavator, specially designed for and applied to turn the digitized (Dutch) newspaper data into word clouds and histograms. This tool is discussed by Joris van Eijnatten, Toine Pieters and Jaap Verheul, 'Using Texcavator to Map Public Discourse.' *Tijdschrift voor Tijdschriftstudies*, 35, 2014, 59–65.

times), that was considered to be a hit. Then, if a minimum number of such matches (e.g., 4) were obtained for a given category we identified it as that type of section.[28]

Their hit ratio turned out to be 1.0 and the false alarm ratio 0.0. In other words: through this procedure the authors were able to pinpoint classified ads perfectly. Another way to discern classifieds by automated indexing, they suggest, is counting words: in their sample 'classified advertisements [in their case complete pages] consistently had the highest word count of any other pages'.[29]

Based on the research on portable radio advertisements, my hypothesis is that the features of Dutch classifieds were similar to their American counterparts. Just like their American counterparts, Dutch advertisements were marked by a specific vocabulary comprised by words such as 'for sale' (*te koop*), 'wanted' (*gevraagd*) etc. They were longer too, as they were habitually published as sections. Therefore, I expect that both approaches will work when using Delpher. Testing them, then, is recommended. When they prove to be effective, the approaches enable users of Delpher to discriminate classifieds from other ads. This, in turn, opens up possibilities for, in my case, specific research into local and regional consumer markets and it raises new research questions. For instance, how big was the market for used portables, measured by classifieds? On what geographical level where such second-hand markets organized?

The second step would be to subdivide the classified sections in smaller, single advertisements – the smallest advertisement units possible. As stated before, discerning them as one-and-the-same when they are in fact separate advertisements makes little sense and hinders further analysis. To my knowledge, no scholarly work has been written on how to do this automatically, but presumably software could be programmed to do this. Particularly in cases in which classifieds are separated by a word in bold (giving an indication of what the ad is about) or by a line, software should be able to distinguish these. As the example in figure 3 shows, these markers were combined as well, which should simplify the automated indexing.

---

[28] Robert B. Allen and Catherine Hall, 'Automated Processing of Digitized Historical Newspapers Beyond the Article Level. Sections and Regular Features.' In Gobinda Chowdhury, Chris Khoo and Jane Hunter (eds.), *The Role of Digital Libraries in a Time of Global Change: 12th International Conference on Asia-Pacific Digital Libraries, ICADL 2010, Gold Coast, Australia, June 21-25, 2010, Proceedings*. Berlin: Springer 2010, 91–101 (97).

[29] *Ibid.*, 100. Naturally, the precision of this measure depends on the OCR quality.

Figure 3: Classified ads in Rotterdamsch Nieuwsblad, October 3, 1940.

Even after improvement of the corpus through these automated techniques, however, results will most likely not be perfect. If only for bad OCR quality, different classifieds might still end up being rendered as one. Libraries should therefore not shy away from the opportunities public participation might offer in improving results. On its website, the KB claims to experiment with crowd sourcing to improve the OCR.[30] Correcting indexation errors could be an additional task that may be crowd sourced. Little knowledge is required to, firstly, distinguish advertisements from classifieds and, secondly, to delineate separate classified ads. Projects that have used crowd sourcing show promising results and deserve to be followed. In the words of Holley, expert in managing large collaborative digital projects: 'Experience shows that the greater the level of freedom and trust you give to volunteers the more they reward you with hard work, loyalty and accuracy.'[31] Additionally, this measure could help the KB to create new virtual communities and user groups – a desirable outcome in itself.[32]

---

[30] http://www.delpher.nl/nl/platform/pages/?title=kwaliteit+(ocr) (Consulted on July 30 2015).

[31] She elaborates on various crowdsourcing projects and presents a list of useful tips to libraries that consider implementing crowdsourcing, Rose Holley, 'Crowdsourcing. How and Why Should Libraries Do It?' *D-Lib Magazine* 16, 2010. A well-known newspaper archive that has used crowdsourcing to great success is the National Library of Australia, under the Australian Newspapers Digitisation Program, referred to both by Holley and by Allen and Hall 2010 (99). Another example of successful crowdsourcing is touched upon by Mark Patrick Baggett et al., 'Populating the Wilderness. Crowdsourcing Database of the Smokies.' *Library Hi Tech* 32, 2014, 249–259.

[32] Holley 2010, 'Crowdsourcing'.

The KB has made and will continue to make a massive effort digitizing historically invaluable data, of which the vast newspaper corpus stands out. Researchers and the general public should feel blessed, since they have been able to reap the benefits of this work. Multiple research projects already are based on the digital collections of the KB and this number will only increase in the future.[33]

Currently, the KB caters to the needs of scholars using their collections in several ways. It employs fellows as well as researchers-in-residence. The website of the KB states that working with these latter researchers enables the library to 'gather valuable information to improve its services to digital humanities researchers' – in other words: it is conducive to other researchers using the collections as well.[34] Furthermore, the KB Research Lab offers a platform on which 'digital humanities researchers can experiment with their data'.[35]

As much as these initiatives should be applauded, up till now they have proven inadequate to improve the quality of the data, both in terms of OCR and OLR. Therefore, I contend that impetus should be given to this improvement. When it comes to OCR, several projects have already taken up this task.[36] This article has argued that similar initiatives should also be deployed regarding OLR, in particular when it comes to the indexation of advertisements.[37] It has suggested several ways in which the KB, which I believe should coordinate this endeavor, could go about this task. Though there is no holy grail, a *mutual* effort from both the KB and, through crowd sourcing, users of Delpher will greatly improve the quality of the OLR of the digitized corpus.

•> JESPER VERHOEF*, PhD student at Utrecht University, researches how America was constructed in Dutch Public Debates on several mass media (1919-1990). His project is part of the NOW-program 'Transatlantis. Digital Humanities Approaches to Reference Cultures.'*

---

[33] Two of the projects that make extensive use of the newspaper corpus, *Translantis* and *Asymenc*, are elaborated on by Van Eijnatten, Pieters and Verheul 2014.

[34] https://www.kb.nl/en/organisation/research-expertise/researcher-in-residence (consulted on August 21, 2015).

[35] https://www.kb.nl/organisatie/onderzoek-expertise/kb-research-lab (consulted on August 21, 2015).

[36] Text-Induced Corpus Clean-up online processing system (TICCLops), e.g., aims to enhance the text quality fully automatically. It has been developed, amongst other things, 'to provide mainly OCR error post-correction'. Martin Reynaert, 'TICCLops: Text-Induced Corpus Clean-up as Online Processing System.' *COLING 2014*, 2014, 52-56 (52). On the Research Lab website, the National Library presents ALTO-Edit, 'a simple browser based post-correction tool for OCR files in ALTO xml format'. http://lab.kbresearch.nl/enhance/ALTO-Edit (consulted on August 21, 2015).

[37] Cf. Walma in this issue, who champions a more fine-grained indexations of newspaper articles. It remains to be researched whether OLR problems also occur in the categories Delpher distinguishes next to articles and advertisements, namely personal announcement (*Familiebericht*) and captioned illustration (*Illustratie met onderschrift*).