

Automated historical judgement extraction: Analyzing perspectives on big and small heroes through NLP

Miel Groten

VU University Amsterdam
mielgroten@live.nl

Yassine Karimi

VU University Amsterdam
m.y.karimi@student.vu.nl

Serge ter Braake

VU University Amsterdam
s.ter.braake@vu.nl

Antske Fokkens

VU University Amsterdam
antske.fokkens@vu.nl

This year, a movie about Michiel de Ruyter (1607-1676), the most famous admiral from the Netherlands,¹ was released. The fame of De Ruyter has inspired people for centuries. De Ruyter was even voted number seven in the 2004 Dutch tv elections for the 'grandest' Dutch person from history.² A search in the Google Ngram viewer shows that he also became increasingly known in the twentieth century in English language books.³ Not everyone, however, believes he should be venerated as a hero. One group in particular, 'Michiel de Rover', protests against the movie and sees De Ruyter as a sea robber who played a doubtful role in Dutch slavery practices (Jensen 2015, Historiek 2015; Elsevier 2010 for earlier controversy).

De Ruyter is a good example, and many more can easily be found, of how the perspective on people changes over time and is dependent on social, cultural and political factors. It is difficult however, to see patterns in how time tells a different story. Such patterns are extremely interesting, because they could reveal a lot about historiography and identity. What attributes do 'heroes', or villains, get from historians over time? Is there a difference in the way women/men, politicians/painters, protestants/catholics are described over time? What *topoi* do historians use? Are there any 'national' characteristics discernable in historiographical sources? To answer such questions one would have to 'close read' an incredible number of sources from the past. Much more than any person could do in a reasonable amount of time. In this paper we apply NLP (automatic text analysis) tools to have a computer aid us in this task. This paper introduces our methodology and first results in using an NLP pipeline for automated historical judgement extraction.

For our purpose we use two extensive biographical dictionaries from the nineteenth and twentieth centuries to start with (Van der Aa 1852-1878, Blok & Molhuysen 1911-1937). They contain circa 45,000 short biographies on famous Dutch people from all history. There is a lot of overlap in the people described, which makes comparisons possible. Our approach in using NLP

to mine perspectives from these texts is twofold. First, we analyze a relatively small selection of texts manually on positive and negative character traits which were attributed to people's characters and their deeds by the authors of the texts and by third parties mentioned in the texts. Second, we use an NLP pipeline that aims to identify positive and negative traits automatically. It is based on an existing pipeline that was originally built for event extraction, including a Named Entity Recognizer.⁴ In this paper, we will introduce our methodology in more detail, present the results (precision and recall) from our pipeline and will discuss what (preliminary) answers we were able to give to the kind of questions asked in the second paragraph.

References

- Aa, A.J. van der (ed.) (1852-1878) *Het Biographisch Woordenboek der Nederlanden*, Haarlem: J.J. van Brederode.
- Blok, P.J. and Molhuysen. P.C., (ed.) (1911-1937) *Nieuw Nederlandsch Biografisch Woordenboek*, Leiden: A.W. Sijthoff's Uitgeverij.
- Historiek. (29 January 2015) *Michiel de Rover*. [Online] Available from: <http://historiek.net/michiel-de-rover/47743/>.
- Jensen, Lotte (2015) Michiel de Ruyter kan dienen voor elk doel. *De Volkskrant*, 30 January. p.20.
- Elsevier. (19 June 2010) *Stichting Michiel de Ruyter boos op 'zeeheld' Wilders'*. [Online] Available from: <http://www.elsevier.nl/Cultuur--Televisie/nieuws/2010/6/Stichting-Michiel-de-Ruyter-boos-op-zeeheld-Wilders-ELSEVIER268640W/>.

¹ See <http://www.admiralthemovie.com/>.

² See http://en.wikipedia.org/wiki/De_Grootste_Nederlander.

³ Google Books Ngram Viewer:

https://books.google.com/ngrams/graph?content=Michiel+de+Ruyter&year_start=1800&year_end=2000&corpus=15&smoothing=3&share=&direct_url=t1%3B%2CMichiel%20de%20Ruyter%3B%2Cc0.

⁴ For a schematic rendering of the tools used in the pipeline, see: <http://www.biographynet.nl/wp-content/uploads/2014/12/Meertens.pdf>.