

Peilend.nl: Exploring Dutch-language Online News

Valentin Jijkoun Fons Laan Maarten de Rijke
ISLA, University of Amsterdam, The Netherlands
jijkoun,a.c.laan,derijke@uva.nl

We present Peilend.nl, a system for analysing Dutch-language online news.

Essentially, Peilend.nl consists of three components:

- Data collection: the online news articles and users' comments on the articles (when available) are continuously tracked using Ssscraper, an open source system for collecting dynamic online data¹.
- Data processing: the collected data is indexed using Lucene² and is sent to Fietstas³, a text analysis web service, that performs, in particular, extraction and resolution of named entities; document processing results are available through a REST web service.
- User interface provides functionality such as keyword search and visualization of search results in terms of word and entity clouds.

Peilend.nl provides feedback functionality: logged-in users can correct system's decisions, such as types of entities (person, organization, location), canonical names or URIs for entities. The system uses such feedback to correct the display of the information, and moreover, collects it for future use in retraining entity extractor and resolver.

Peilend.nl is a demonstrator for the technology developed for the online media analysis, where opinions towards entities and topics are studied. Within this user scenario, we will demonstrate the use of simple sentiment analysis techniques, based on hand-crafted and automatically-derived polarity lexicons.

Acknowledgements

Peilend.nl is supported by the DuOMAn project carried out within the STEVIN programme which is funded by the Dutch and Flemish Governments under project nr STE-09-12. The research is also supported by the European Union's ICT Policy Support Programme as part of the Competitiveness and Innovation Framework Programme, CIP ICT-PSP under grant agreement nr 250430, by and by the Netherlands Organisation for Scientific Research (NWO) under project nrs 612.066.512, 612.061.814, 612.061.815, 640.004.802.

¹<http://ilps.science.uva.nl/resources/sssrape>

²<http://lucene.apache.org>

³<http://fietstas.science.uva.nl>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

DIR '11 Amsterdam, The Netherlands
Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...\$10.00.