
Detecting and Exploiting Semantic Overlap (DAESO)

Emiel Kraahmer

Site visit, Tilburg, 10 June 2009

Voorpaginanieuws

Aanbevolen

Buitenland

Nederland

Economie

Wetenschap/techniek

Entertainment

Sport

Gezondheid

Meest gelezen

[Nieuwsmeldingen](#)

 [RSS](#)



[Socialisten gaan onderuit in Europa](#)

NRC Handelsblad - 25 minuten geleden

De Europese kiezers stemden vooral voor (centrum-) rechtse partijen. De partijen die de afgelopen jaren in Europa hebben gekozen voor méér markt. De socialisten gingen in veel landen onderuit tijdens de Europese verkiezingen. Brussel, 7 juni. ...

[Sociaal-democraten grote verliezers](#)

Volkskrant - 1 uur geleden

BRUSSEL - De sociaal-democraten zijn de grote verliezers geworden van de Europese verkiezingen. In het kielzog van de PVV van Geert Wilders boekten ook in sommige andere EU-landen nationalistische partijen flinke winst. De opkomst was lager dan ooit ...

[Europa schuift naar rechts](#)

Gelderlander - 1 uur geleden

De opkomst bij de verkiezingen voor het Europees Parlement in 27 EU-landen is opnieuw lager. ANP BRUSSEL - Centrumrechtse en rechtse partijen hebben bij de Europese verkiezingen van de afgelopen dagen de beste resultaten geboekt. ...

[Dreun voor socialisten](#)

AD.nl - 8 uren geleden

BRUSSEL - De Europese verkiezingen van 2009 zijn uitgedraaid op een gigantische dreun voor de socialisten. De tweede politieke familie zakt van 217 naar 155 tot 165 zetels. Ze had zelf op zeker 200 zetels gerekend. De verkiezingen legden geen bom onder ...

[EVP blijft grootste fractie in Europees Parlement](#)

Knack - 8 uren geleden

De Europese Volkspartij (EVP) blijft de grootste fractie in het Europees Parlement en versterkt haar positie. Volgens de eerste projecties zou de EVP tussen 263 en 273 zetels in de wacht slepen. Dat is een opvallend resultaat. De EVP zou immers slechts ...

[Nederlaag van links is goed nieuws voor Barroso](#)

Volkskrant - 9 uren geleden

BRUSSEL - De eurosceptische krachten zijn in opmars. Maar de 'oude' fracties blijven de grootste in het Europees Parlement. Premier Balkenende is vrijdag in gesprek met de voorzitter van de Europese Commissie Barroso op de tweede dag van de EU-top in ...

Problem statement

- Similar information can be expressed (“**paraphrased**”) in many different ways.
- Major **stumbling block** for robust NLP applications,
- Automatic detection of semantic overlap between sentences could improve
 - (Multidocument) **Summarization**: less redundancy
 - **QA, IE, IR**: improved recall
 - Intelligent document **merging**
 - Recognizing textual entailments (**RTE**)
 - ...

Detecting and Exploiting Semantic Overlap

- **Resources** exist on the word level (e.g., Wordnet, Cornetto), but are mostly **lacking** for more complex phrases.
- The Stevin **Daeso** project (2006 - 2009) intends to fill this gap.
- Development of:
 - **Resources**: a Dutch parallel corpus
 - **Tools** for annotating the corpus and for detecting semantic overlap
 - **Application** to primarily summarization, but also to QA and IE
- Collaboration between **Tilburg** University (Erwin Marsi, Emiel Krahmer), **Antwerp** University (Iris Hendrickx, Walter Daelemans), **Amsterdam** University (Maarten de Rijke) and **Textkernel** (Jakub Zavrel).

Where are we now?

- **Resources**
 - Manual annotation (500K) + data processing finished
 - Automatic annotation (500K) at the end of the project

- **Tools**
 - Annotation tools finished
 - Sentence alignment + fusion tools in progress

- **Applications**
 - Development of summarization tools and evaluation platform finished
 - First experiments have been performed, more to follow.
 - Plans to look at duplicate detection (instead of QA).

Self-evaluation

- Daeso will **fulfill** its goals: to produce a rich data-collection, useful tools, new insights in applications.
- **IPR** is completely settled.
- **Extra work**: Pycornetto, BOF multidocument summarization corpus
- Some tasks took **more time** than planned (corpus collection, automatic overlap detector), but we foresee no problems in finishing the project.
- We will not be able to exhaust all **research possibilities** (especially wrt applications).
- Follow up research **after DAESO** is finished (Sander Wubben - UvT, Katja Filipova - EML).