

# DuOMAn

## Dutch Language Online Media Analysis

STEVIN project: April 2008-March 2011

Valentin Jijkoun, Maarten de Rijke  
ISLA, Universiteit van Amsterdam

September 11, 2008

# DuOMAn: goals

- Development of a set of Dutch language resources and tools for identifying and aggregating sentiments in online data sources
- Sentiment mining:
  - *Who supports/criticizes whom and on what topics?*

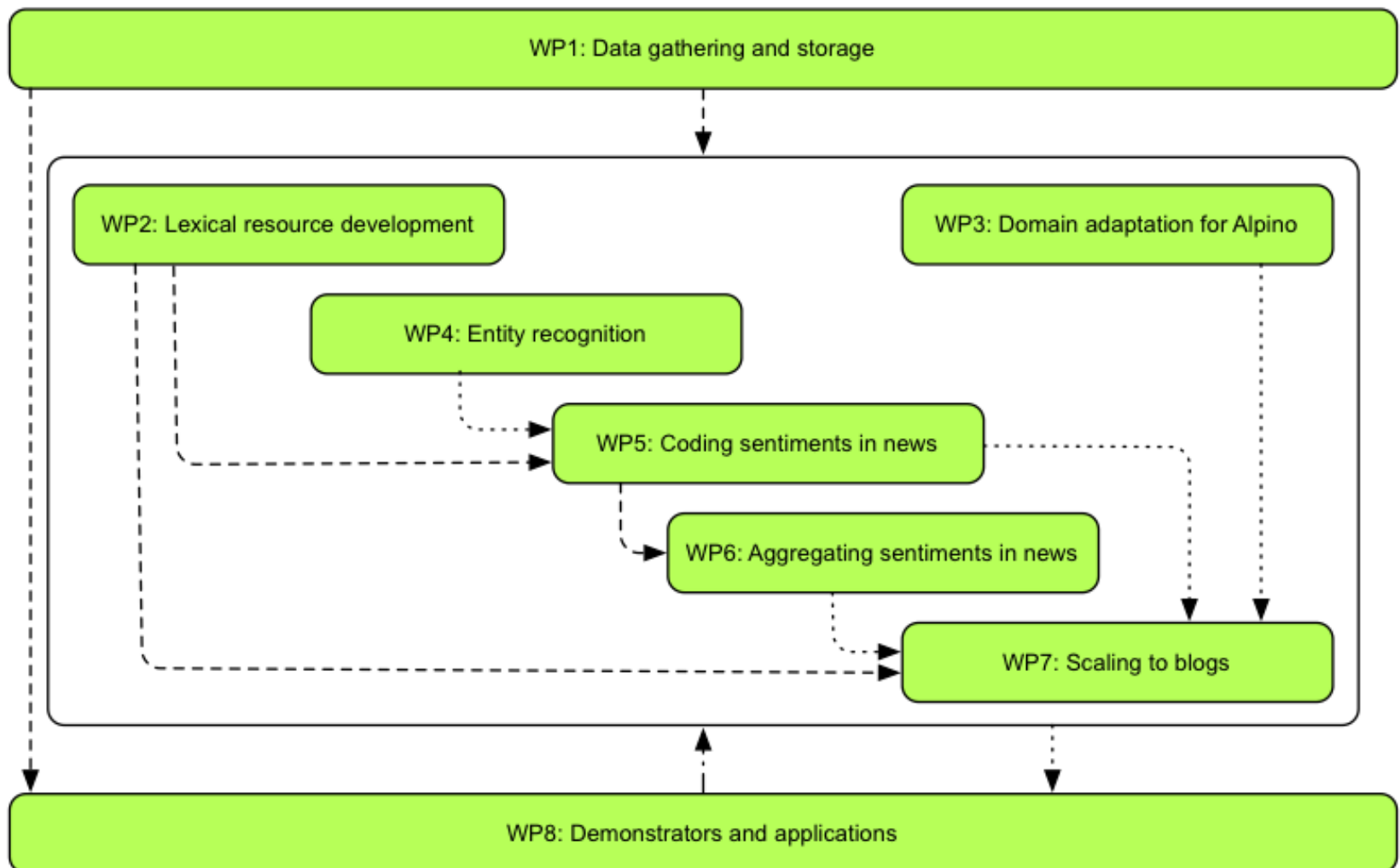
# DuOMAn: goals (2)

- Data:
  - edited (news)
  - user-generated content (news comments, blogs)
- End users:
  - Media-analysts (toolbox)
  - General public (demos)

# DuOMAn consortium

- University of Amsterdam      coordinator
- University College Ghent      coreference resolution
- Rijksuniversiteit Groningen      parsing
- GridLine      lexical resources
- TrendLight      end user, annotations

# Project structure



# DuOMAn: key deliverables

- Methods:
  - Search & discovery for user generated content
  - Aggregation algorithms
- Data:
  - Dutch corpus of blogs and news comments
  - Sentiment lexicons and test sets
  - Annotated corpora of blogs/comments

# DuOMAn: key deliverables (2)

- Tools:
  - Syntactic parser for blog texts (ALPINO)
  - Entity coreference for blog texts (COREA)
  - Sentiment extraction for news and blogs
- Demos:
  - Yearly; first demo: October 2008

# DuOMAn: results so far

- 6 months into the project
- Tracking 11 online newspapers
  - 116,000 news articles
  - 367,000 readers' comments
- First version of sentiment lexicon
  - Cornetto+classification+manual assessment
  - 2836 negative, 1868 positive words

# Thanks!

<http://ilps.science.uva.nl/Projects/DuOMAn>