

# ACS project proposals

Anna Korhonen, Laura Rimell,  
Tim Van de Cruys, Yufan Guo

# Unsupervised Detection of Adjective Senses

Anna Korhonen & Tim Van de Cruys

- The meaning of many adjectives is highly context dependent. For example, **cool** can mean
  - neither warm or very cold (*a cool autumn day*)
  - unfriendly (*a cool reception*)
  - fashionable a the time (*that's cool*)
- The ability to detect the likely meaning of an adjective in context would be useful many NLP applications
- A lot of research has been done on detection of noun and verb senses - not so much work exist on adjectives

# Unsupervised Detection of Adjective Senses

- The project will involve
  - choosing and extracting features from data (e.g. bags of words, grammatical relations)
  - implementing an unsupervised learning method for adjective sense detection (based on e.g. clustering, topic models, factorization methods, graph-based methods)
  - evaluating results against a gold standard, using human judges and/or in the context of an NLP task
- A student who is interested in machine learning and its application to NLP tasks, programming language of choice

# The Use of Verbs in Social Media

Laura Rimell & Anna Korhonen

- People write quite differently on Twitter than in other genres, possibly closer to the way they talk than the way they normally write
- The project will involve studying the use of verbs in Twitter and addressing the challenges it poses for NLP
- It will involve
  - collecting a corpus from the Twitter Stream
  - applying an existing technique for analysing the meaning and behaviour of verbs to this corpus

# The Use of Verbs in Social Media

- investigating the challenges Twitter poses for automatic analysis of verbs
- finding ways to overcome these either by improving existing methods or by developing a learning method of the student's choosing
- This project will require some manual annotation from the student
- Programming language of choice

# Unsupervised Discovery of Information Structure of Scientific Literature

Anna Korhonen & Yufan Guo

- Scientific writing tends to be fairly conventionalised
- The information (or rhetorical or discourse) structure of a scientific article can be characterised by classifying text into categories such as Background, Objective, Method, Result, and Conclusion
- Such a classification can be useful for human readers (e.g. scientists) as well as for NLP (e.g. information extraction, summarisation)
- Various approaches have been developed for automatic classification (e.g. Teufel and Moens, 2002; Guo et al., 2011)

# Unsupervised Discovery of Information Structure of Scientific Literature

- Existing approaches are (semi)-supervised
- This project will explore using unsupervised learning for the task (e.g. clustering - a range of possible methods to choose from)
- The method will be applied to a corpus of biomedical journals and evaluated in various ways
- A student who is interested in machine learning
- Programming language of choice

# Predicting London's Mayoral Election Results with Twitter

Tim Van de Cruys and Anna Korhonen

- An increasing amount of research into sentiment and mood classification is based on data generated on micro-blogging sites, such as Twitter
- Recent research has looked at relationship between Twitter mood and e.g. stock market fluctuations and the prediction of political election results

# Predicting London's Mayoral Election Results with Twitter

- In May 2012, a new mayor of London will be elected
- This project involves developing a system that predicts the outcome of the election from sentiment data in Twitter
- Possible methods:
  - sentiment classification and opinion mining techniques suitable for analysing micro-blogging data
  - dynamic topic models that try to capture trending topics during the campaign
- A student interested in machine learning, and its application to sentiment classification / topic detection