Towards improving the precision of a relation extraction system by processing negation and speculation
Vincent Van Asch, Roser Morante, Walter Daelemans
CLiPS - University of Antwerp

GOAL: Evaluating a direct implementation of a negation/speculation module.
Increasing precision of a relation extraction system by adding a negation/speculation filtering module.

BIOGRAPH is a project that aims at putting forward a new methodology for text mining from heterogeneous information sources. The final goal of the project is to use text mining for finding new links between genes and phenotypes, and for gene prioritisation in order to discover non-obvious disease causing genes. BiographTA is the text analysis component of the project and is used to extract relations from biomedical abstracts, with a focus on precision.

Thus, alpha-syntrophin has an important role in synaps formation and in the organization of utrophin.

The binarized version of BioInfer corpus contains 2069 positive relations between 626 different named entities. By retrieving the full abstracts from PubMed, we have a full text corpus of 552 abstracts with annotated relations.

The scope labelling module is optional and filters out negated relations. A relation is negated if both entities are under the scope of a negation and not under the scope of a speculation.

Abstract
Tokenization
Lemmatization
NE - tagging (UMLS)
GDep parsing
SVMLight classification
Scope labelling
Binary relations

SYSTEM DESCRIPTION
The input for the relation extraction system are full abstracts. After some preprocessing steps, the actual relation extraction is done using SVMs. Every pair of named entities (NE) in a sentence is classified and accordingly gets the relation status or not.

The scope labelling module is optional and filters out negated relations. A relation is negated if both entities are under the scope of a negation and not under the scope of a speculation.

EXAMPLE OF FILTERING
[...] talin, but not vinculin or tubulin, appears to co-localize with actin microfilaments [...]