Assaf Toledo

Semantic Modeling of Textual Entailment:

Proof-Based Annotation in a Compositional Framework

This thesis introduces a new computational framework and annotation methodology for investigating textual entailment in a theory-based paradigm. This paradigm is premised on the assumption that entailment recognizers could be made more accurate if an explicit linguistic theory explains at least some of the main data that they are designed to cover.

The proposed framework is an annotation platform which allows human annotators to create entailment data that are accounted for by a standard semantic model. The platform integrates a typed-lexicon, a stochastic parser, a theorem prover and a user interface. It is a sound proof system; hence, when the annotations are used successfully for deduction, this indicates that the underlying semantic theory accounts for the entailment. The platform is used within a methodology of Annotating-By-Proving: A premise and a conclusion of a positive pair are considered well-annotated only if the annotations support an inferential chain. An extension of this methodology also covers negative pairs. The platform provides annotators with an immediate feedback on their annotations.

This general approach is used for developing a semantic model incorporating some of the most common inferential phenomena in the Recognizing Textual Entailment corpora: appositive, intersective and restrictive modification, as well as simple existential and universal quantification. Human Annotators used the platform to generate a dataset of 600 annotated positive and negative entailments explained by this semantic model. The corpus, SemAnTE (Semantic Annotation of Textual Entailment), is available online. The explicitness of the semantic theory, the simplicity of its representations, and the standard conventions of tagged parse trees all suggest that the model is learnable and holds promise for developing better performing entailment recognizers.

Netherlands Graduate School of Linguistics

: LOT

Utrecht Institute of Linguistics OTS

UiL OTS

Netherlands Graduate School of Linguistics

LOT

ndelijke Onderzoekschool Taalwetenschap