Kolloquium, 12. Juni 2017

Dayyan Smith

Analysis of NIL Results in an Entity Linking System

Speaker: Mai Linh Pham

Advisor: Yadollah Yaghoobzadeh

To better understand the Mai Linh's work for her thesis, there are some terms that need to be defined. Entity linking is the process of linking mentions from text to a corresponding entity in a knowledge base. NIL results are entities that can *not* be linked to the knowledge base. And fine-grained entity annotation means, that the standard NER tags (LOC, ORG, PER, MISC) have further sub categories. To improve entity linking systems, Mai Linh looks at NIL results and clusters NIL mentions for analysis. The ultimate goal is to examine whether fine-grained types are useful for clustering and analysing NIL mentions.

First an entity annotation tool labels a text, and then an entity linking system links entities to the freebase knowledge base. Then the NIL output is extracted and clustered. For the clustering task fine-grained types are used. The entity annotation system used is called *FIGER*. It is a fine-grained entity annotation system, with 112 tags. It also allows overlapping types, i.e. company, written work, news agency for *New York Times*. The entity linking system used is *WAT*. It has three components: a spotter that scans input text for mentions and returns a list of candidate entities; a disambiguator that ranks candidate entities with different disambiguation algorithms; and a pruner that removes useless annotations, and aims to increase precision. The NIL mentions are those mentions that are annotated by *FIGER* but not linked by *WAT*.

Three clustering approaches are compared: coarse-grained type, fine-grained type, top-level type. In conclusion, fine-grained entity types can be used for clustering semantically related NIL mentions. They take into account lexical and contextual properties of an entity. They are more informative than coarse-grained types.