Entity Linking with YAGO

Thomas Rebele
Télécom ParisTech

Fabian Suchanek
Télécom ParisTech

January 7, 2016

YAGO is a knowledge base with more than 10 million entities (such as people, companies, and cities) and more than 120 million facts about these entities. It knows, e.g., birth dates of people, creation dates of companies, and names and sizes of cities. YAGO is extracted from Wikipedia in different languages. Human evaluation showed an accuracy of over 95 %, which distinguishes it from its competitors.

Knowledge bases like YAGO enhance the possibilities of text analysis. Entities that are mentioned in a text can be linked to the corresponding entities of the knowledge base. This is called entity linking, and enables applications such as automatic link generation for websites, trend analysis, or fact checking. An analysis by Huet et al. on the French newspaper 'Le Monde' illustrates this by analyzing mentions of men and women and their age over time.

Traditional approaches restrain themselves to cases where the entity name is mentioned explicitly (such as "Barack Obama"). Our goal is to extend entity linking also to nominal anaphoras such as "the president", where it is clear from the context which entity is meant. By covering also the nominal anaphoras, we hope to drastically increase the number of annotations, and thus ultimately their usefulness.