

WHAT IS ENTITY LINKING?

Entity linking grounds ambiguous mentions of people, places and things in text to their corresponding node in a knowledge base.

WHY IS IT IMPORTANT?

Linking is a fundamental problem in the domain of knowledge base population, where we wish to automatically extract facts about entities from text and store them in a structured knowledge base.

For example, in recent years, it has become possible to pull out a smartphone and ask questions in natural language like:

“Who is the CEO of **Tesla Motors**?”

To interpret this question, we first must understand that “CEO” refers to a specific type of occupation, and “Tesla Motors” refers to a specific company which produces electric cars – though other interpretations are often possible.

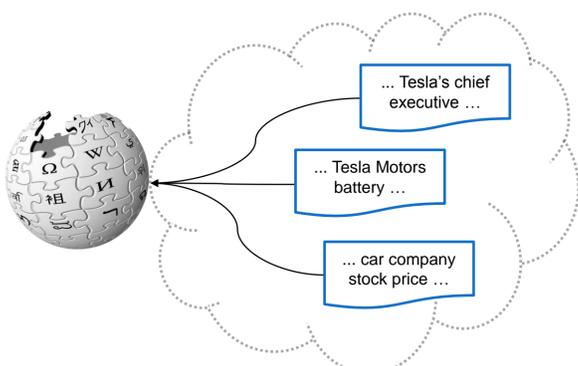
To answer the question, we rely on linking again to identify entities in text about which can then extract facts. For example, we may find the following passage from the web:

“.. **Tesla's** chief executive officer **Elon Musk** announced today...”

from which we can link the query entities and “Elon Musk” to answer the question.

USING LINKS AS ENTITY CONTEXT

- Traditional approaches to entity linking use structured information from Wikipedia to build rich entity models
- However:
 - Wikipedia cannot provide complete coverage over entities in a domain
 - Wikipedia specific features don't generalize to other types with KB
 - Generating Wikipedia like structure is an expensive process, requiring human curation
- We explore the question of whether information derived from mentions of an entity can be used to link to that entity
- To answer this, we use employ the Wikilinks dataset which contains 40 million links from web pages to Wikipedia articles.



ENTITY DISAMBIGUATION

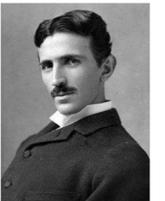
Example

“**Tesla** plans to offer hands-free highway driving in its **Model S** electric sedans in 2015, putting it as much as a year ahead of other luxury brands in offering autopilot functions in automobiles. Other auto makers, including **General Motors** and **Volkswagen**, have said they're aiming to launch hands-free highway driving systems”

Example news text adapted from the Wall Street Journal. “Tesla Aims to Leapfrog Rivals” 10-10-2012

Who or what is “Tesla” referring to?

- Nikola Tesla, the famous electrical engineer and inventor
- Tesla, a standard unit of magnetic flux density
- Tesla Motors, the electric car company headed by Elon Musk



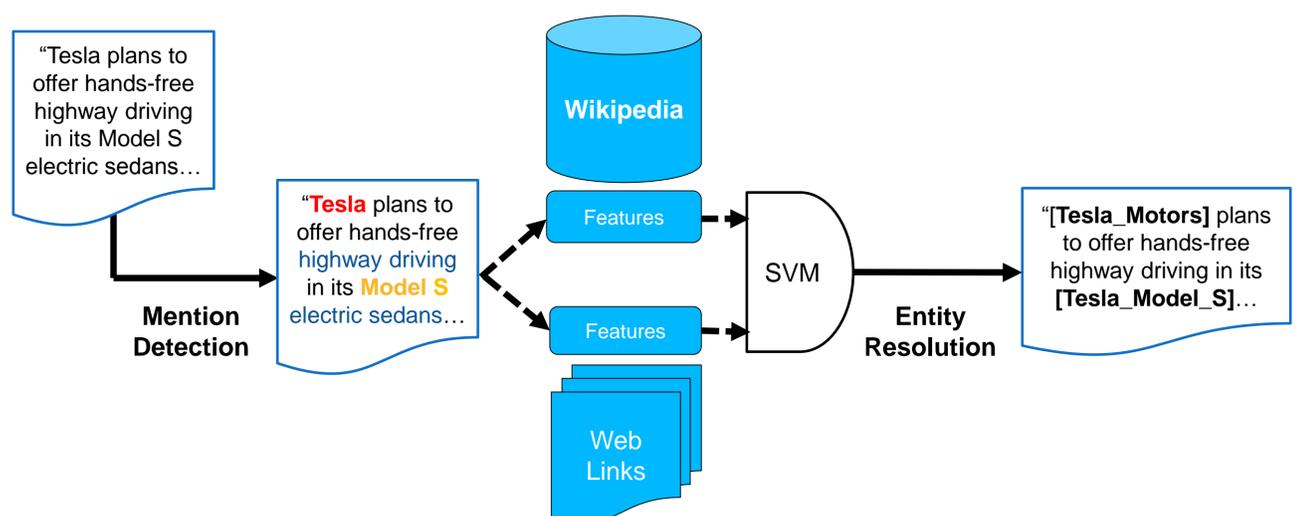
Given this set of candidates, we extract a set of features from the document which help identify the correct candidate.

Standard Disambiguation Features:

- Entity Prior – How prominent is each candidate?
- Name Probability – Probability that a link with “Telsa” used to refer to each candidate.
- Textual Context – Measures similarity between words in the document and words around links to that entity from the web.
- Entity Coherence – How often is one entity observed in the same document as another?

Experiments:

- We build a flexible entity linking system which able to take advantage of feature models derived from Wikipedia, Web links or some combination of both datasets.
- We evaluate our system on the standard AIDA - CoNLL03 dataset of newswire text.
- We perform a detailed analysis and comparison of feature models derived from both Wikipedia and web links.



RESULTS

State-of-the-art Disambiguation Accuracy

We are able to achieve state-of-the-art results by combining both Wikipedia and Web-link derived disambiguation features. 89.2% vs 87.6% (Alhelbawy & Gaizauskas, 2014) on the AIDA / CoNLL dataset.

Web Links can be used in place of a curated Knowledge Base

While much of the current work in entity linking relies of the rich, structured data present in Wikipedia, we demonstrate that web-link derived features perform on average at 97% of their Wikipedia equivalents.

This suggests that a robust entity linking system may be built for any online KB with inbound links.