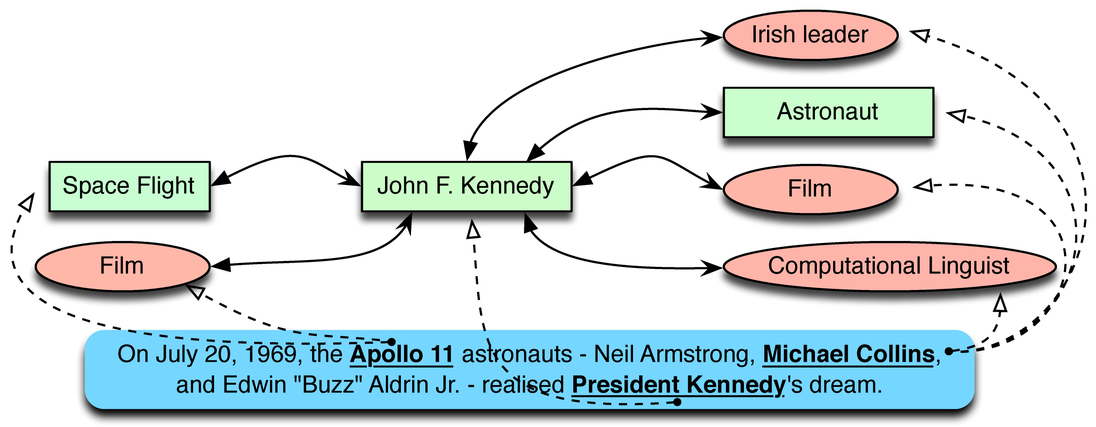
Bringing the Head Closer to the Tail with Entity Linking

Manisha Verma[1], Diego Ceccarelli[2,3]
[1] UCL Department Of Computer Science [2] ISTI-CNR, Pisa, Italy [3] IMT Lucca

Entity Linking Problem



The entity linking task aims at identifying, given a plain document, the small fragments of text (interchangeably called *mentions* or *spots*) referring to any *named entity* that is listed in a given knowledge base, e.g. Wikipedia. The ambiguity of natural language makes it a non trivial task. The same entity can be in fact mentioned with different text fragments, e.g., "President Kennedy" or "John F. Kennedy". On the other hand, the same mention may refer to different entities, e.g., "Michael Collins" may refer to either the well known astronaut, or to the Irish leader and president of the Irish provisional government in 1922.

The annotation is usually organized in three subtasks:

1. **Spotting**: discover the fragments that could refer to an entity. A set of candidate mentions is detected, and for each mention a list of candidate entities is produces;
2. **Disambiguation**: for each spot associated with more than one candidate, a single entity is selected to be linked to the spot;
3. **Ranking**: the list of entites detected is ranked according to some policy, e.g. annotation confidence.

ESAIR '14 – 7th International Workshop on Exploiting Semantic Annotations in Information Retrieval. November 7, 2014. Shanghai, China.