Entity Linking meets Word Sense Disambiguation: A Unified Approach
Andrea Moro, Alessandro Raganato, Tiziano Flati and Roberto Navigli

Abstract
Entity Linking (EL) and Word Sense Disambiguation (WSD) both address the lexical ambiguity of language. But while the two tasks are pretty similar, they differ in a fundamental aspect: in EL the textual mention can be linked to an entity which may or may not contain the exact mention, while in WSD there is a perfect match between the word form (better, its lemma) and a suitable sense.
We present a unified graph-based approach to EL and WSD based on a loose identification of candidate meanings coupled with a densest subgraph heuristic which selects high-coherence semantic interpretations.

The Best of Two Worlds
Our main goal is to bring together the two worlds of WSD and EL:
1. Keep the set of candidate meanings for a given mention as open as possible
2. Provide an effective method for handling this high level of ambiguity.
A key assumption of our work is that the lexicographic knowledge used in WSD is also useful to tackle the EL task, and vice versa: the encyclopedic information utilized in EL helps disambiguate nominal mentions in a WSD setting.

A joint task
Our task is to disambiguate and link all nominal and named entity mentions occurring within a text. The linking task is performed by associating each mention with the most suitable entry of a given knowledge base.
Our definition does not enforce any constraints in terms of what to link, i.e., unlike Wikification and WSD, we can link overlapping fragments of text.

Example. Given the text fragment: “Major League Soccer” we identity and disambiguate several different entity and concept mentions:

| Major League Soccer | major league | league | soccer |

BabelNet
BabelNet is a multilingual knowledge base which consists of roughly nine million concepts and named entities together with their lexicalizations in 50 different languages and 250 million semantic relations. At the core of this resource lies the integration of encyclopedic, i.e., from Wikipedia, and lexicographical knowledge, i.e., from WordNet, Open Multilingual WordNet and OmegaWiki within a unified, multilingual structured network.

The method
1. Given a lexicalized semantic network, we associate with each vertex, i.e., either concept or named entity, a semantic signature, that is, a set of related vertices. This is a preliminary step which needs to be performed only once, independently of the input text.
2. Given a text, we extract all the linkable fragments from the text and, for each of them, list the possible meanings according to the semantic network.
3. We create a graph-based semantic interpretation of the whole text by linking the candidate meanings of the extracted fragments using the previously-computed semantic signatures. We then extract a dense subgraph of this representation and select the best candidate meaning for each fragment.

Semantic Signatures
1. We first weight the edges using triangles: \[
\text{weight}(\varepsilon(i, j, k)) = \frac{1}{|\varepsilon(i, j, k)|} \cdot \left(\text{Sim}(i, j) \cdot \text{Sim}(j, k) \cdot \text{Sim}(i, k)\right)
\]
2. We then run a random walk with restart: \[
P(t+1) = \sum_{i \in |E|} \frac{\text{weight}(\varepsilon(i, t))}{\sum_{j \in |E|} \text{weight}(\varepsilon(j, t))} 
\]
3. The semantic signature of a vertex \(v\) is the set of vertices visited at least \(\mu\) times during the random walk.

Example: SemSign(Hilbert’s program) = {Set theory, ..., Mathematical proof}

Experimental Evaluations
Our experiments on six gold-standard datasets show the state-of-the-art performance of our approach as well as its robustness across languages.

Conclusions
1. We presented a novel, integrated state-of-the-art approach to Entity Linking and Word Sense Disambiguation;
2. Our graph-based approach exploits the semantic network structure to its advantage: two key features of BabelNet, that is, its multilinguality and its integration of lexicographic and encyclopedic knowledge, make it possible to run our general, unified approach on the two tasks of Entity Linking and WSD in any of the languages covered by the unified approach on the two tasks of Entity Linking and WSD in any of the languages covered by the unified approach on the two tasks of Entity Linking and WSD in any of the languages covered by the unified approach on the two tasks of Entity Linking and WSD in any of the languages covered by the unified approach on the two tasks of Entity Linking and WSD in any of the languages covered by the unified approach on the two tasks of Entity Linking and WSD in any of the languages covered by
3. At the core of our approach lies the effective treatment of the high degree of ambiguity of partial textual mentions by means of a 2-approximation algorithm for the densest subgraph problem;

References