Knowledge-rich WSD based on WordNet++

(Dezambiguizare BOGATA in cunostinte)

The idea:

- **Simple, basic disambiguation algorithms can have higher performance than sophisticated ones when being fed enough knowledge (cunostinte). Here knowledge will be provided by means of semantic relations.**

- **When provided with a vast amount of high-quality semantic relations, simple knowledge-lean disambiguation algorithms compete with state-of-the-art supervised WSD systems.**

(knowledge-lean = lipsit de cunostinte)

The method:

- **Paolo Ponzetto and Navigli (2010) have conceived a methodology for automatically extending WordNet with a great number of new semantic relations, provided by an encyclopedic resource – Wikipedia.**

- **The resulting resource is called WordNet++**

HOW:

Wikipedia pages are automatically associated with WordNet senses and the associative semantic relations from Wikipedia are transferred to WordNet. The result is a much richer lexical resource.

Wikipedia was launched on January 15, 2001 by Jimmy Wales si Larry Sanger.

The Structure of Wikipedia

Wikipedia text is very structured:

- **the pages corresponding to articles are formatted into sections and paragraphs**
- **various relations among pages exist**

The relations among pages include:

1. **Redirect pages (redirecting relations) – used in order to redirect the query towards the actual article page that contains information concerning the entity designated by the query. These pages are used in order to designate alternate expressions for an entity. (EX.: the pages CAR and SICKNESS redirect towards pages AUTOMOBILE and DISEASE respectively). These pages model synonymy.**

2. Disambiguation pages – they collect links for a number of possible entities towards which the initial query might point. They model homonymy.

3. Internal links – articles mentioning other entries in the encyclopedia point towards them through the internal hyperlinks. They model article cross-reference.

EX.: The page about Wikipedia contains hyperlinks pointing towards the two inventors.

In summary:

A Wikipedia page (henceforth, Wikipage) presents the knowledge about a specific concept (e.g. SODA (SOFT DRINK)) or named entity (e.g. FOOD STANDARDS AGENCY). The page typically contains hypertext linked to other relevant Wikipages. For instance, SODA (SOFT DRINK) is linked to COLA, FLAVORED WATER, LEMONADE, and many others. The title of a Wikipage (e.g. SODA (SOFT DRINK)) is composed of the lemma of the concept defined (e.g. soda) plus an optional label in parentheses which specifies its meaning in case the lemma is ambiguous (e.g. SOFT DRINK vs. SODIUM CARBONATE). Finally, some Wikipages are redirections to other pages, e.g. SODA (SODIUM CARBONATE) redirects to SODIUM CARBONATE.

KNOWLEDGE-RICH WSD

- **Introduced in the paper:**

Knowledge-rich Word Sense Disambiguation Rivaling Supervised Systems. Simone Paolo Ponzetto, Roberto Navigli, Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 2010, p. 1522 – 1531.

- **The underlying idea: the ideas of Bunescu and Pasca (2006) and of Mihalcea (2007), according to which the Wikipages can be considered as representing word senses.**

- **The technique that will be used: the knowledge from Wikipedia will be injected into a WSD system by means of a mapping to WordNet.**

Extending WordNet

(Mapping Wikipedia to WordNet)

PHASE I: automatic mapping between Wikipages and WordNet senses

PHASE II: the relations connecting Wikipedia pages are transferred to WordNet; the result is WordNet++

Notations:

- ✓ **The set of all Wikipedia pages –** $Senses_{Wiki}$
- ✓ **The set of all WN senses –** $Senses_{WN}$
- ✓ **A Wikipage –** w

Let

$\mu : Senses_{Wiki} \rightarrow Senses_{WN}$

be the mapping we want to obtain.

Let $w \in Senses_{Wiki}$

$$\mu(w) = \begin{cases} s \in Senses_{WN}(w), \text{ daca se poate stabili o legatura} \\ \varepsilon, \text{ altfel} \end{cases}$$

where $Senses_{WN}(w)$ **is the set of senses of *w*'s *lemma* in WordNet.**

- ✓ **<u>Reminder</u>: The title of a Wikipage (e.g. SODA (SOFT DRINK)) is composed of the lemma of the concept defined (e.g. soda) plus an optional label in parentheses which specifies its meaning in case the lemma is ambiguous (e.g. SOFT DRINK vs. SODIUM CARBONATE).**

<u>Example:</u>

$$\mu(\text{SODA}(\text{SOFT DRINK})) = \text{soda}_n^2,$$

where soda_n^2 **is the corresponding WordNet sense.**

Establishing the mapping

In order to establish a mapping between the two resources, we first perform two operations:

- **we identify various types of disambiguation contexts for Wikipages**
- **we identify various types of disambiguation contexts for WordNet senses**

These two contexts will be intersected in order to obtain the mapping.

1. Disambiguation context for a Wikipage

Given a target Wikipage *w*, that we want to map to a WordNet sense of *w*, we use the following information as disambiguation context:

- ✓ **Sense labels: e.g. given the page SODA (SOFT DRINK), the words soft and drink are added to the disambiguation context.**

- ✓ **Links: the titles' lemmas of the pages linked from the Wikipage *w* (outgoing links). For instance, the links in the Wikipage SODA (SOFT DRINK) include soda, lemonade, sugar, etc.**

- **Categories: Wikipages are classified according to one or more categories, which represent meta-information used to categorize them. For instance, theWikipage SODA (SOFT DRINK) is categorized as SOFT DRINKS. Since many categories are very specific and do not appear in WordNet (e.g., SWEDISH WRITERS or SCIENTISTS WHO COMMITTED SUICIDE), we use the lemmas of their syntactic heads as disambiguation context (i.e. writer and scientist). To this end, we use the category heads provided by Ponzetto and Navigli (2009) in:**

Simone Paolo Ponzetto and Roberto Navigli. 2009. Large-scale taxonomy mapping for restructuring and integrating Wikipedia. In: Proceedings of IJCAI-09, pages 2083–2088.

The disambiguation context of a Wikipage is formed as follows:

Given a Wikipage *w*, we define its disambiguation context, *Ctx(w)*, as the set of words obtained from some or all of the three previously mentioned sources.

2. The disambiguation context of a WordNet sense

Given a WordNet sense *s* and its synset *S*, we use the following information as disambiguation context to provide evidence for a potential link in our mapping *μ*:

- **Synonymy: all synonyms of *s* in synset *S*.**
- **Hypernymy/Hyponymy : all synonyms in the synsets H such that H is either a hypernym (i.e., a generalization) or a hyponym(i.e., a specialization) of S.**
- **Sisterhood: words from the sisters of *S*. A sister synset *S'* is such that *S* and *S'* have a common direct hypernym.**
- **Gloss: the set of lemmas of the content words occurring within the gloss of *s*.**

Given a WordNet sense *s*, we define its disambiguation context, *Ctx(s)*, as the set of words obtained from some or all of these four sources.

The Mapping

Contexts *Ctx(w)* and *Ctx(s)* are intersected in order to obtain the mapping.

Mapping Algorithm

- **We want to link each Wikipage to a WordNet sense. Let *w* be such a Wikipage.**

The steps of the algorithm:

- **Initialization: our mapping μ is empty, i.e. it links each Wikipage *w* to ε (lines 1-2).**

- **For each Wikipage *w* whose lemma is monosemous both in Wikipedia and in WordNet i.e.**

$$|SensesWiki(w)| = |SensesWN(w)| = 1$$

we map *w* to its only WordNet sense w_n^1 (lines 3-5).

- **For each remaining Wikipage *w* for which no mapping was previously found (i.e., $\mu(w) = \varepsilon$, line 7), we do the following:**

  - ✓ **lines 8-10: for each Wikipage *d* which is a redirection to *w*, for which a mapping was previously found (i.e. $\mu(d) \neq \varepsilon$, that is, *d* is monosemous in both Wikipedia and WordNet) and such that it maps to a sense *μ(d)* in a synset *S* that also contains a sense of *w*, we map *w* to the corresponding sense in *S*.**

  - ✓ **lines 11-14: if a Wikipage *w* has not been linked yet, we assign the most likely sense to *w* based on the maximization of the conditional probabilities *p(s|w)* over the senses**

$$s \in Senses_{WN}(w)$$

(no mapping is established if a tie occurs, line 13).

- **As a result of the execution of the algorithm, the mapping μ is returned (line15).**

- **At the heart of the mapping algorithm lies the calculation of the conditional probability $p(s|w)$ of selecting the WordNet sense s given the Wikipage w. The sense s which maximizes this probability can be obtained as follows:**

$$\mu(w) = \underset{s \in Senses_{WN}(w)}{argmax}\ p(s/w) = \underset{s}{argmax}\frac{p(s,w)}{p(w)} =$$

$$= \underset{s}{argmax}\ p(s,w)$$

The latter formula is obtained by observing that *p(w)* does not influence our maximization, as it is a constant independent of *s*. As a result, the most appropriate sense *s* is determined by maximizing the joint probability *p(s,w)* of sense *s* and page *w*.

We estimate *p(s,w)* as:

$$p(s,w) = \frac{scor(s,w)}{\sum_{\substack{s' \in Senses_{WN}(w) \\ w' \in Senses_{Wiki}(w)}} scor(s',w')}$$

where

$$scor(s,w) = \left| Ctx(s) \cap Ctx(w) \right| + 1$$

(we add 1 as a smoothing factor).

Formula interpretation:

We determine the best senses by computing the intersection of the disambiguation contexts of *s* and *w*, and normalizing by the scores summed over all senses of *w* in Wikipedia and WordNet.

The algorithm:

Input: $Senses_{Wiki}$, $Senses_{WN}$

Output: **The mapping** $\mu : Senses_{Wiki} \rightarrow Senses_{WN}$

1: **for each** $w \in Senses_{Wiki}$

2: $\mu(w) := \varepsilon$

3: **for each** $w \in Senses_{Wiki}$

4: **if** $\left|Senses_{Wiki}(w)\right| = \left|Senses_{WN}(w)\right| = 1$ **then**

5: $\mu(w) := w_n^1$

6: **for each** $w \in Senses_{Wiki}$

7: **if** $\mu(w) = \varepsilon$ **then**

8: **for each** $d \in Senses_{Wiki}$ **such that *d* redirects to *w***

9: **if** $\mu(d) \neq \varepsilon$ **and** $\mu(d)$ **is in a synset of *w* then**

10: $\mu(w) :=$ **sense of *w* in synset of μ(d);break**

11: **<u>for</u> each** $w \in Senses_{Wiki}$

12: **if** $\mu(w) = \varepsilon$ **<u>then</u>**

13: **<u>if</u> no tie occurs <u>then</u>**

14: $$\mu(w) := \underset{s \in Senses_{WN}(w)}{argmax}\ p(s/w)$$

15: **<u>return</u>** μ