Concept Mapping: Connecting Educators
Proc. of the Third Int. Conference on Concept Mapping
Tallinn, Estonia & Helsinki, Finland 2008

A CONCEPT SENSE DISAMBIGUATION ALGORITHM FOR CONCEPT MAPS

Alfredo Simón-Cuevas[1], Luigi Ceccaroni[2], Alejandro Rosete-Suárez[1], Amhed Suárez-Rodríguez[1], Manuel de la Iglesia-Campos[1]
[1]Centro de Estudios de Ingeniería de Sistemas (CEIS), Facultad de Ingeniería Informática, Instituto Superior Politécnico "José Antonio Echeverría", C. Habana, Cuba
[2]Departament de Llenguatges i Sistemes Informàtics (LSI), Universitat Politècnica de Catalunya (UPC), Campus Nord, Edif. Omega, C. Jordi Girona, 1-3, 08034 Barcelona, Spain
Email: {asimon, rosete, asuarez, miglesia}@ceis.cujae.edu.cu, luigi@lsi.upc.edu

Abstract. Concept maps are a graphically rich tool for representing knowledge in natural language. An important aspect for their automatic or semi-automatic processing, including concept mapping, formalization and evaluation, is the identification of the most rational sense of the concepts. In this paper, we present an algorithm for concept sense disambiguation based on contextual analysis, domain information and gloss. The algorithm takes advantage of the whole map's topology and use WordNet as sense repository. Results of preliminary experimental evaluations of the concept disambiguation algorithm applied to several concept maps in the Spanish language are presented and compared with the state of the art.

1 Introduction

Concept maps (CMs), defined by Novak & Gowin (1984), are a graphically-rich tool for organizing and representing knowledge in natural language. In this paper, we consider the process in which knowledge represented in a CM is automatically recognized, in such a way that it can be semantically analyzed and processed by machines. This process is related to CM formalization (e.g., Brilhante, Macedo, & Macedo, 2006; Simón, Ceccaroni, & Rosete, 2007), the automatic or semi-automatic process of CM construction (Reichherzer, Cañas, Ford, & Hayes, 1998; Cañas & Carvalho, 2004; Richardson, Goertzel, & Fox, 2006), and other processes in which WordNet (Miller, Beckwidth, Fellbaum, Gross, & Miller, 1990) is used as knowledge base in automatic CM analysis, such as the evaluation of interactive CM construction (Kornilakis, Grigoriadou, Papanikolaou, & Gouli, 2004).

CMs can be considered a structural and unrestricted knowledge representation in natural language; therefore, the identification of the more rational sense of concepts can be an interesting aspect for the automatic or semi-automatically processing of knowledge represented in CMs; for example, in the case of concepts with different meanings (ambiguous concepts), it is possible that the system suggests the automatic construction of meaningless propositions. Word sense disambiguation (WSD) (Agirre & Edmonds, 2006) has been broadly studied in the cases where documents or texts are used as context. Nonetheless, few works to solve this problem in the CM context exist; the main contribution being the one reported by Cañas et al. (2003), which shows some limitations when applied to CMs in Spanish language. We report in this work a novel algorithm for *concept sense disambiguation* (CSD), which tries to assign the more rational sense of a given concept in the CM, using WordNet (Miller, Beckwidth, Fellbaum, Gross, & Miller, 1990) as sense repository, *domain information, contextual analysis*, and *the gloss*.

Along the paper, to represent the English translation of the Spanish terms used, the following notation will be used:

español ("Spanish")

This paper begins (section 2) with an overview of the WordNet knowledge-base. Section 3 describes the main aspects considered to define the CSD algorithm, which is presented in section 4, together with an example. Results of preliminary experiments on several CMs in the Spanish language and comparison with the state of the art are reported in section 5.

2 Overview of WordNet

WordNet is a lexical knowledge-base (Miller, Beckwidth, Fellbaum, Gross, & Miller, 1990), whose basic structure is the *synset* (equivalent to sense). Synsets are distributed in form of a semantic network and interconnected among themselves by several types of lexical and semantic relations; the algorithm proposed uses WordNet's *hypernymy*, *hyponymy*, *meronymy*, *holonymy*, *gloss* and *rgloss* relations. The *synset* defines the meaning of a word, which in the case of *polysemy* can be found in various *synsets*; a meaning description (*gloss)* is included in each *synset's* structure. In addition to the *synset's* structure, general domain taxonomy (e.g.

Chemistry, Geography and Philosophy) is associated to it. The domains are associated to *synsets* in such a way that a *synset* can belong to one or several of these domains.

3 The Disambiguation Process and Concept Maps

Lexical disambiguation in its broadest definition is nothing less than determining the meaning of every word in context, which appears to be a largely unconscious process in people. As a computational problem, its solution presupposes a solution to complete natural-language understanding or common-sense reasoning (Ide & Véronis, 1998). In computational linguistics, one of the kinds of language ambiguity that have received the most attention is that of word senses: its resolution is essential for any practical application, and it seems to require a wide variety of methods and knowledge-sources with no apparent pattern in what any particular instance requires (Agirre & Edmonds, 2006). In this context, the problem is generally called *word sense disambiguation* (WSD), and is defined as the problem of computationally determining which "sense" of a word is activated by the use of the word in a particular context. WSD has been broadly studied in the case of documents or texts contexts; a review of this work is reported by Agirre & Edmonds (2006). Nonetheless, few works exist to solve this problem in CMs.

A CM is an external and simplified representation of part of a person's cognitive structure, and its obtaining is largely non language-based or language-dependent. Rather, it is a derived language from the mental imagery of the person in which ideas can be schematically represented, a feature that generally belongs to natural language. These aspects suggest that both concepts and propositions can be subject to subjectivity, which can derive in ambiguity in some cases. In CMs, WSD has been previously studied by Cañas et al. (2003). They proposed an algorithm to disambiguate the sense of words in CMs, whether they are part of a concept or a linking-phrase, using WordNet (Miller, Beckwidth, Fellbaum, Gross, & Miller, 1990). The algorithm exploits the topology of the CM, by including only the words of key concepts as part of the disambiguation process, and the semantics of the CM, by trying to determine which of the senses in WordNet best matches the context of the CM, using *hypernymy* relations from WordNet (Cañas, Valerio, Lalinde, Carvalho, & Arguedas, 2003). This algorithm was mainly defined to be applied on CMs in the English language and the results obtained with its application on several CMs in Spanish languages were less satisfactory.

In this work, we propose a *knowledge-based method* (Mihalcea, 2006) for concept sense disambiguation (CSD) in CMs, which also use WordNet as sense repository, to be mainly applied to CMs in the Spanish language. In CMs and in WordNet, a concept can be formed by a word (generally a *noun*) or several words (combination of *nouns*, *adjectives* and *verbs*); the method proposed allows disambiguating only the concepts of a CM which are included in WordNet as *synsets*. This kind of disambiguation process on CMs is generally easier than on texts; in a CM, the concepts are explicitly identified and related, while in a text these aspects are not clear and they have to be inferred. To improve the disambiguation process in CMs with respect to the one reported by Cañas et al. (2003), we maintain the contextual analysis in the CM and in WordNet, while increasing the information to take into account in the process; specifically, the use of domain information, considering the experience of Magnini et al. (2002), the gloss, and WordNet's relations such as *hyponymy*, *meronymy*/*holonymy* and *gloss*/*rgloss*, in addition to the *hypernymy* relation used by Cañas et al. (2003), were added to CSD algorithm. In it, the disambiguation process is carried out through heuristic functions, based on *domain, context* and *gloss*.

The **domain** constitutes a fundamental semantic property and a natural way to establish the association between concepts in a CM context (Magnini, Strapparava, Pezzulo, & Gliozzo, 2002). However, a CM integrates many domains; therefore, the most representative domains in the CM should be identified before the disambiguation. These domains are identified analyzing the occurrence frequency of the domains to which the senses of the *most inclusive, most general concepts* belong, and three alternatives have been defined to use them to disambiguate a given concept.

The **context** in which a given concept appears in the CM is explored to determine a corresponding, similar context in WordNet, using the *synset* of the concept at issue and considering the *hypernymy/hyponymy*, *meronymy/holonymy* and *gloss/rgloss* WordNet's relations. The contextual similarity provides a quantitative clue for identifying the most rational sense of a given concept, and a weight factor associated to the context created from each *synset* in WordNet is used to evaluate that similarity.

A similar analysis is carried out with the **gloss**: the algorithm evaluates the overlap between the CM context of a given concept and the context created with all words that form the glosses of the *synsets*, selecting the *synsets* of the gloss with more words in common with the CM context to disambiguate the concept. In the *contextual* and *gloss analyses*, a variable radius is used to select the concepts and words to form the CM context, allowing to take advantage of the whole CM's topology, as a novel way with respect to Cañas et al. (2003), who use two linking-phrases as a fixed distance from the concept to be disambiguated in the selection of the words to conform the CM context.

4 Concept Sense Disambiguation Algorithm

In this section, we formally describe the CSD algorithm, which comprises five steps: *preparing the CM*, *selecting a set of CM domains* (D_{cm}), *disambiguating by domain*, *disambiguating by context* and *disambiguating by gloss*. These steps are executed sequentially on a CM and the order was defined to obtain a more efficient processing. The *disambiguation by domain* required fewer queries to WordNet than the *disambiguation by context* and the precision obtained in the process is better; the *gloss* is included in the CSD algorithm as an alternative if some concepts cannot be disambiguated by domain or context. (In the Spanish version of WordNet used in this work, only a few *synsets* with gloss are available.) In the process, concepts, when disambiguated, are added to a set of non-ambiguous concepts with their senses. Before describing the algorithm, let us consider the following basic data:

- C is the set of concepts (c) in the CM;
- $S(c)$ is the set of *synset* (s) corresponding to concept c; e.g., the *synset{ser_vivo#1, ser#1, organismo#1}* corresponding to concept *Organismos* ("organism");
- $S(C)$ is the set of *synsets* corresponding to all concepts in C;
- $D(s)$ is the set of domains (d) associated to s; e.g., the domains *{Chemistry, Physics}* associated to the *synset{nitrógeno#1, número_atómico_7#1}*;
- $D(c)$ is the set of domains associated to the set of *synsets of c*;
- $D(C)$ is the set of domains associated to the set of *synsets* of all concepts in C;
- $CSD(C, d)$ is the subset of concepts in C which have at least one *synset* associated to the domain d: $CSD(C, d) = \{c_i \mid c_i \in C, d \in D(c_i)\}$; e.g., *CSD({Nitrógeno, Atmósfera, Tierra}, Physics) = {Nitrógeno, Atmósfera}*;
- $OF(d, C)$ is the *occurrence frequency* of domain d in the *synsets* of the concepts in C:

$$OF(d,C) = \frac{|CSD(C,d)|}{|C|} \qquad \textbf{(1)}$$

- $D_{ch}(D)$ is the set of child domains of the domains included in D according to the taxonomy of WordNet; e.g., D_{ch}*({Biology, Geography})={Biochemistry, Anatomy, Physiology, Genetics, Topography};*
- $D_p(D)$ is the set of parent domains of the domains included in D according to the taxonomy of WordNet; e.g., D_p*({Biology, Geography})={Pure Science, Earth};*
- $Context_{cm}(c, r)$ is the set of neighbor concepts of a given concept c within a radius r (measured as arcs between two concepts) in the CM and the words (*nouns*, *adjectives* and *verbs*) extracted from the *linking-phrases* used in the proposition in which these concepts are related;
- $Context_{wn}(s, L, C)$ is the set formed by paths between *synset s* and other *synsets s'* in WordNet, with a maximum length of L (measured as arcs between two synsets) from s, such that $s' \in S(C)$ and using *hyperonymy*, *meronymy* and *gloss* relations; e.g. $Context_{wn}$ *({agua#, H2O#1}, 2, {Hidrógeno, Oxígeno})= {({hidrógeno#1, número_atómico_1#1}, 1, 1), ({número_atómico_8#1, O#1, oxígeno#1}, 1, 1),…}*, from the paths: *{agua#, H2O#1} has_mero_madeof {hidrógeno#1, número_atómico_1#1}* and *{agua#, H2O#1} has_mero_madeof {número_atómico_8#1, O#1, oxígeno#1}*;
- $w(Context_{wn}(s, R, C))$ represents the weight of a sense s to disambiguate a concept c:

$$w(Context_{wn}(s,L,C)) = \sum_{k=1}^{|Context_{wn}(s,L,C)|} \frac{\alpha_k}{l_k} \qquad \textbf{(2)}$$

  where l_k is the length of the path (k) and α_k is the number of concepts in C with some *synset* in *k;*
- $gloss(s)$ is the set of words included in the gloss of the synset s in WordNet.

The five steps of the disambiguation process of a CM with only one most general concept are theoretically described below and applied to a practical example in section 4.1.

Step 1. Preparing the CM

Extract all concepts (c_i) and the propositions they belong to from the CM; the *proposition* set *PS* and *concept* set *CS* are created. From *CS*, the following sets are created[1]:

- the *non-ambiguous concept* set NACS = $\{c_i | c_i \in CS, |S(c_i)| = 1\}$;
- the *unknown concept* set UCS = $\{c_i | c_i \in CS, |S(c_i)| = 0\}$;
- the *ambiguous concept* set ACS = $\{c_i | c_i \in CS, |S(c_i)| > 1\}$.

Step 2. Selecting a set of CM domains (D_{cm})

Let us consider $r = 1$ and $T = 0.4$[2].

```
While ($|Context_{cm}(most\ general\ concept, r)| < T * |CS|$){$r = r+1$};
$DS = D(Context_{cm}(most\ general\ concept, r))$;
$DS_{max} = \{d_{max} \mid d_{max} \in DS, \forall d_i \in DS\ OF(d_{max}, Context_{cm}(most\ general\ concept, r)) \geq OF(d_i, Context_{cm}(most\ general\ concept, r))\}$;
$D_{cm} = DS_{max} \cup D(most\ general\ concept)$.
```

Step 3. Disambiguating by domain

```
For each $c_i \in$ ACS
    $c_i$ is considered disambiguated by $s_{ij}$ if:
        a. $|\{s_{ij} | s_{ij} \in S(c_i), |D(s_{ij}) \cap D_{cm}| > 0\}| = 1$; or
        b. $|\{s_{ij} | s_{ij} \in S(c_i), |D(s_{ij}) \cap D_{ch}(D_{cm})| > 0\}| = 1$ and $|\{s_{ij} | s_{ij} \in S(c_i), |D(s_{ij}) \cap D_{cm}| > 0\}| = 0$; or
        c. $|\{s_{ij} | s_{ij} \in S(c_i), |D(s_{ij}) \cap D_{p}(D_{cm})| > 0\}| = 1$ and $|\{s_{ij} | s_{ij} \in S(c_i), |D(s_{ij}) \cap D_{cm}| > 0\}| = 0$.
Update concept sets: NACS = NACS $\cup$ $\{c_i\}$, ACS = ACS $-$ $\{c_i\}$.
```

Step 4. Disambiguating by context

```
For each $c_i \in$ ACS
  $r = 1$;
  repeat
    $r = r + 1$; $C_t = Context_{cm}(c_i, r)$; $W_d = 0$; $S_d = \{\}$;
    for each $s_{ij} \in S(c_i)$
      if ($w(Context_{wn}(s_{ij}, L, C_t)) > W_d$), then $S_d = \{s_{ij}\}$;
         $W_d = w(Context_{wn}(s_{ij}, L, C_t))$;
      else
        if ($w(Context_{wn}(s_{ij}, L, C_t)) = w_d$), then $S_d = S_d \cup \{s_{ij}\}$;
  until ($|S_d| = 1$ v $|Context_{cm}(c_i, r)| = |CS|$)
if $|S_d| = 1$, then $c_i$ is disambiguated with $s_{ij}$;
Update concepts sets: NACS = NACS $\cup$ $\{c_i\}$, ACS = ACS $-$ $\{c_i\}$.
```

Step 5. Disambiguating by gloss

```
For each $c_i \in$ ACS
  $r = 1$;
  repeat
    $r = r + 1$; $C_t = Context_{cm}(c_i, r)$; $G_d = \{\}$; $S_d = \{\}$;
    for each $s_{ij} \in S(c_i)$
      if ($|gloss(s_{ij}) \cap (Context_{cm}(c_i, r)| > |G_d|$), then $S_d = \{s_{ij}\}$;
        $G_d = gloss(s_{ij}) \cap Context_{cm}(c_i, r)$;
      else
        if ($|gloss(s_{ij}) \cap (Context_{cm}(c_i, r)| = |G_d|$), then $S_d = S_d \cup \{s_{ij}\}$;
  until ($|S_d| = 1$ v $|Context_{cm}(c_i, r)| = |CS|$)
if $|S_d| = 1$, then $c_i$ is disambiguated with $s_{ij}$;
Update concepts sets: NACS = NACS $\cup$ $\{c_i\}$, ACS = ACS $-$ $\{c_i\}$.
```

[1] The senses of the concepts are found using WordNet, after applying a morphological transformation where needed. The transformation simply consists in obtaining the singular form of the concept if it appears in plural.

[2] Coefficient T defines the percentage of concepts in the CM, to be considered for determining the CM's domains.

4.1 An example

As an example, we apply the CSD algorithm to a CM in Spanish about *Nitrógeno* (“nitrogen”), shown in Figure 1; its English translation is shown in Figure 2. A Spanish version of WordNet[3] was used as sense repository.

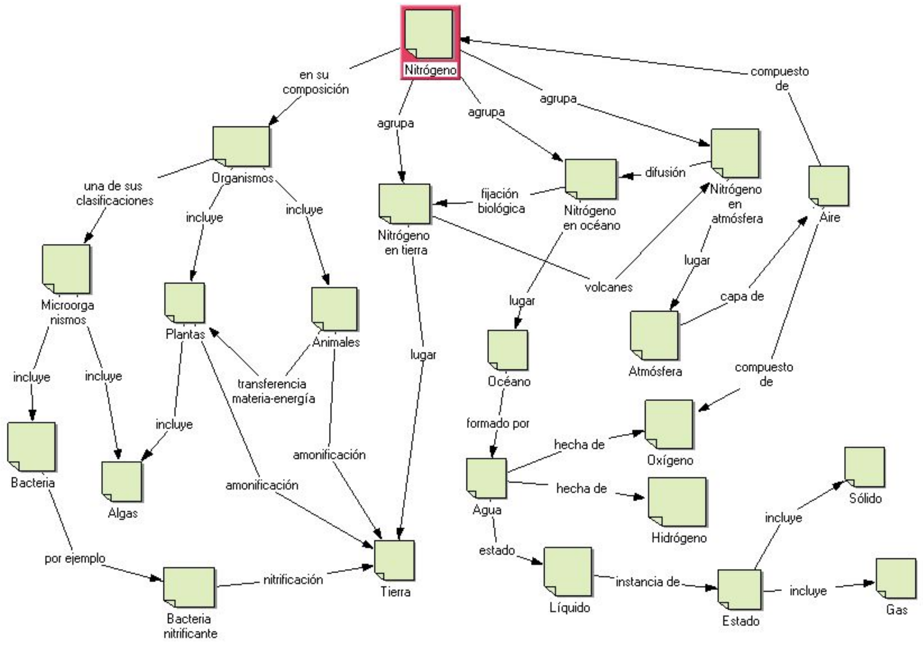

Figure 1. Concept map of *Nitrógeno* (“nitrogen”), in Spanish

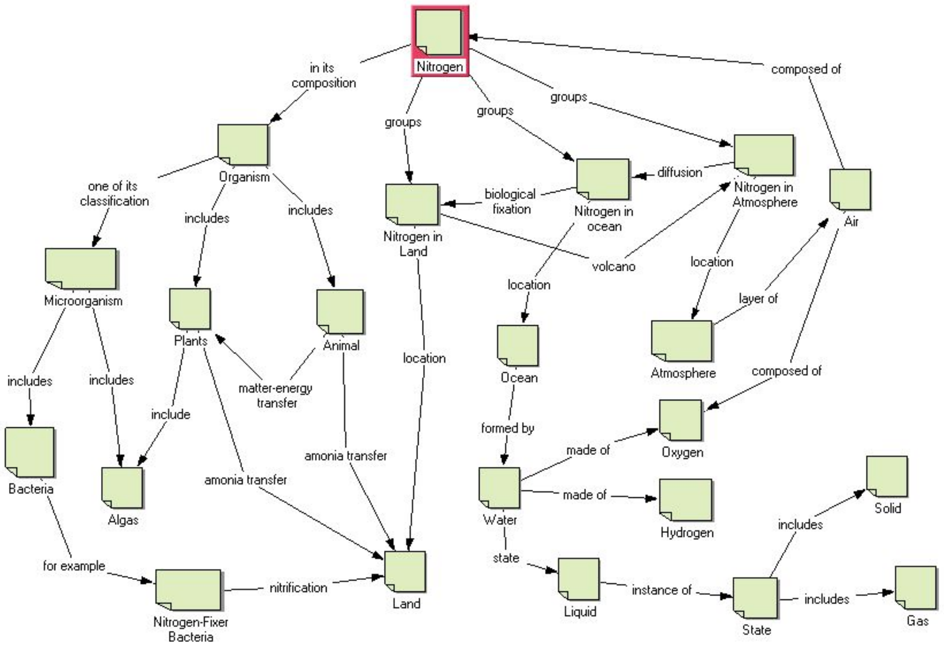

Figure 2. Concept map of the *Nitrógeno* (“nitrogen”), in English

In **Step 1**, all *concepts* and *propositions* are extracted and 11 ambiguous concepts (50%), eight non-ambiguous concepts (30%) and four unknown concepts (20%) are identified.

In **Step 2**, 21 domains are identified in WordNet from the synsets corresponding to the 19 ambiguous and non-ambiguous concepts; from these WordNet domains, *Chemistry*, *Physics*, *Biology*, *Geography*, and *Astronomy* are identified as CM domains.

[3] Concretely, the version developed by the Natural Language Processing Group (TALP) of the Software Department (LSI) of the Technical University of Catalonia (UPC) (Farreres, Rigau, & Rodríguez, 1998).

For the disambiguation-by-domain (**Step 3**), we consider the ambiguous concept *Organismo* (“organism”) with two *synsets* in WordNet: *1-{organismo#1, ser#1, ser_vivo# }: Biology –cualquier entidad viva* (“any living entity”) and *2-{organismo#2}: Factotum –entidad pública o privada con una función determinada* (“a system considered analogous in structure or function to a living body”). The concept is disambiguated with the *synset 1-{organismo#1, ser#1, ser_vivo#}* using the *Biology* domain.

For the disambiguation-by-context (**Step 4**), we consider the concept *Agua* (“water”) with five *synsets* in WordNet: *1-{H2O#1, agua#1}*: *Chemistry, Geography –líquido incoloro, insípido e inodoro* (“a clear colorless odorless tasteless liquid”), *2-{agua#2, systema_de_aguas#1}*: *Hydraulics - fuente de agua* (“source of water”), *3-{agua#3, masa_de_agua#1}*: *Geography-parte de la superficie de la Tierra cubierta de agua* (“the part of the earth's surface covered with water”), *4-{agua#4}*: *Philosophy - antes, considerada uno de los cuatro elementos que formaban el universo* (“once thought to be one of four elements composing the universe”) and *5-{agua#5, agua_de_lluvia#1, lluvia#2}*: *Factotum- gotas de agua fresco que caen como precipitación desde las nubes* (“drops of fresh water that fall as precipitation from clouds”). The algorithm first selects the concepts and words from the *linking-phrases* to form the CM context: *Nitrógeno en Océano* (“nitrogen in ocean”), *Océano* (“ocean”), *Oxígeno* (“oxygen”), *Hidrógeno* (“hydrogen”), *Líquido* (“liquid”), *Aire* (“air”), *Estado* (“state”), *hecha* (“made”), *formado* (“formed”), *instancia* (“instance”), *lugar* (“place”), *compuesto* (“composed”). Then, the paths between the *synsets* of these concepts/words and each *synset* of *Agua* (“water”) in WordNet (that is, the context in WordNet associated to the CM context) are selected: 142 paths from *synset-1* (*w = 150*), 18 paths from *synset-2* (*w = 22*), 59 paths from *synset-3* (*w = 69*), 20 paths from *synset-4* (*w = 20*) and 11 paths from *synset-5* (*w = 11*). Therefore, the concept *Agua* (“water”) is disambiguated with the sense identified by the *synset1-{H2O#1, agua#1}*, which is the correct sense of the concept in this context.

The rest of ambiguous concepts is disambiguated either by domain or context and the disambiguation by gloss (**Step 5**) is not necessary in this case. The algorithm proposes only one incorrect sense, achieving 90% of precision.

5 Experimental Results

For the experimental process, the same Spanish version of WordNet used in the example in section 4.1 was used as sense repository, and the metrics *precision* [4](PR), *recall*[5] (RE) and *coverage* (CO) (Palmer, Ng, & Dang, 2006) were used to measure the results, which was possible because the correct sense corresponding to the ambiguous concepts was known. We started the tests selecting 20 CMs from the literature, validated by experts and with *at least one ambiguous concept* (according to WordNet). These CMs had, in average, 17 concepts each (81% of the concepts had at least one synset in WordNet), and 16 domains each. All ambiguous concepts in those CMs were included in the evaluation set (a total of 151) and they had an average number of five *synset*, which belonged to more than 30 domains. Two kinds of tests were carried out with the CSD algorithm: (1) the *whole algorithm* was applied to the evaluation set (The results for each CM are shown in Table 1.), and (2) each part (*domain*, *context* and *gloss*) of the algorithm was independently applied to the evaluation set. (The general results are shown in Table 2.) The CSD algorithm guessed some sense tag in 151 cases (100% of coverage) and the correct sense tag in 135 cases, achieving 89,4% of precision and recall.

The *precision* results obtained in the second test confirm the potential usefulness of domain, gloss and the use of other WordNet’s relations, such as *meronymy/holonymy*, in addition to the *hypernymy* relation used by Cañas et al. (2003), in the concept disambiguation process in CMs. Nonetheless, low results were obtained for *recall* and *coverage* when only the domain information or the gloss were used in the disambiguation process. In the first case, this was due to several *synsets* of a same concept being associated to the same domain in WordNet; rendering thus the domain a less disambiguating factor. In the second case, it was due to the few *synsets* with gloss found in WordNet.

To compare the CSD algorithm with the one reported by Cañas et al. (2003), eight CMs were selected from the 20 used in the previous tests, where all ambiguous concepts were formed by one word. Cañas et al. (2003)

[4] The precision of a system is computed by summing the scores over all test items that the system guesses on, and dividing by the number of guessed-on items.

[5] Recall (or accuracy) is computed by summing the system's score over all items (counting unguessed-on items as zero score), and dividing by the total number of items in the evaluation set.

selected for evaluation one-word concepts that had more than two senses in WordNet. The same version of WordNet from previous tests was used to evaluate both algorithms and the results are shown in Table 3.

Maps	**DC**[a]	**DI**[b]	**ND**[c]	**Step 3 (D)**		**Step 4 (C)**		**Step 5 (G)**		**PR**	**RE**	**CO**
				DC	**DI**	**DC**	**DI**	**DC**	**DI**			
1	12	1	0	3	1	9	0	0	0	0,923	0,923	1,000
2	5	0	0	3	0	2	0	0	0	1,000	1,000	1,000
3	10	0	0	5	0	5	0	0	0	1,000	1,000	1,000
4	10	1	0	6	0	4	1	0	0	0,909	0,909	1,000
5	4	1	0	3	0	1	1	0	0	0,800	0,800	1,000
6	6	2	0	2	0	4	2	0	0	0,750	0,750	1,000
7	9	1	0	2	0	7	1	0	0	0,900	0,900	1,000
8	2	0	0	1	0	1	0	0	0	1,000	1,000	1,000
9	4	1	0	1	0	3	1	0	0	0,800	0,800	1,000
10	3	0	0	1	0	2	0	0	0	1,000	1,000	1,000
11	4	0	0	1	0	3	0	0	0	1,000	1,000	1,000
12	1	1	0	0	0	1	1	0	0	0,500	0,500	1,000
13	9	1	0	4	1	5	0	0	0	0,900	0,900	1,000
14	4	1	0	1	0	3	1	0	0	0,800	0,800	1,000
15	10	1	0	2	0	8	1	0	0	0,909	0,909	1,000
16	2	1	0	1	1	1	0	0	0	0,667	0,667	1,000
17	12	0	0	4	0	8	0	0	0	1,000	1,000	1,000
18	9	1	0	2	0	7	1	0	0	0,900	0,900	1,000
19	15	2	0	10	1	5	1	0	0	0,882	0,882	1,000
20	4	1	0	3	0	1	1	0	0	0,800	0,800	1,000
Total.	**135**	**16**	**0**	55	4	80	12	0	0			
Ave.	6,75	0,8	0	2,75	0,20	4,00	0,60	0	0	**0,872**	**0,872**	**1,000**

[a] **DC**: correctly disambiguated concepts ; [b] **DI**: incorrectly disambiguated concepts ; [c] **ND**: non disambiguated concepts

Table 1: Experimental results obtained with the whole CSD algorithm

Parts of the algorithm	**PR**	**RE**	**CO**
Domain	**0,945**	0,352	0,381
Contextual analysis	**0,841**	**0,841**	**1,000**
Gloss	**0,838**	0,253	0,327

Table 2: Experimental results obtained by each part of the CSD algorithm, applied independently

Maps	**Ambiguous Concepts**	**Cañas et al. (2003)'s algorithm**						**CSD algorithm**					
		DC[a]	**DI**[b]	**ND**[c]	**PR**	**RE**	**CO**	**DC**	**DI**	**ND**	**PR**	**RE**	**CO**
1	5	0	5	0	0,000	0,000	1,000	5	0	0	1,000	1,000	1,000
2	10	8	2	0	0,800	0,800	1,000	10	0	0	1,000	1,000	1,000
3	8	3	5	0	0,375	0,375	1,000	6	2	0	0,750	0,750	1,000
4	2	2	0	0	1,000	1,000	1,000	2	0	0	0,800	0,800	1,000
5	2	1	1	0	0,500	0,500	1,000	1	1	0	0,500	0,500	1,000
6	12	8	4	0	0,667	0,667	1,000	12	0	0	1,000	1,000	1,000
7	17	11	6	0	0,647	0,647	1,000	15	2	0	0,882	0,882	1,000
8	5	4	1	0	0,800	0,800	1,000	4	1	0	0,800	0,800	1,000
Total.	**61**	**37**	**24**	0				**55**	**6**	**0**			
Ave.	7,62	4,62	3,00	0	**0,599**	**0,599**	**1,000**	6,87	0,75	0	**0,842**	**0,842**	**1,000**

[a] **DC**: correctly disambiguated concepts ; [b] **DI**: incorrectly disambiguated concepts ; [c] **ND**: non disambiguated concepts

Table 3: Comparison between the Cañas et al. (2003)'s proposal and the CSD algorithm

The results obtained from the evaluation set selected suggest a significant improvement by the CSD algorithm with respect to the one reported by Cañas et al. (2003) on CMs in Spanish, and confirm the usefulness of increasing the information considered for the disambiguation process.

6 Conclusions

A concept sense disambiguation algorithm to apply in automatic or semi-automatic processing of concept maps, which uses WordNet as sense repository, has been presented. The algorithm defined explores the context in which the concepts appear in a concept map, and tries to determine which context in WordNet has the best similarity with the context defined in the concept map. This contextual similarity, combined with domain and gloss analysis, allows improving the accuracy of disambiguation of concepts in concept maps in the Spanish language, providing better results with respect to similar research, using the same evaluation set.

7 Acknowledgements

The research described is partly supported by the EC project Laboranova (IST-5-035262-IP). The opinions expressed in this paper are those of the authors and are not necessarily those of the Laboranova partners.

References

Agirre, E., & Edmonds, P. (2006) (Eds.). Word Sense Disambiguation: Algorithms and Applications. Springer. 364 p.

Brilhante, V., Macedo, G., & Macedo, S. (2006). Heuristic Transformation of Well-Constructed Conceptual Maps into OWL Preliminary Domain Ontologies, WONTO'06. Brazil.

Cañas, A., Valerio, A., Lalinde, J., Carvalho, M., & Arguedas, M. (2003). Using WordNet for Word Sense Disambiguation to Support Concept Map Construction. LNCS 2857, Springer-Berlin. 350-359.

Cañas, A. J., & Carvalho, M. (2004) Concept Maps and AI: an Unlikely Marriage?. Paper presented at Simpósio Brasileiro de Informática Educativa (SBIE 2004). Brazil.

Farreres, X., Rigau, G., & Rodríguez, H. (1998). Using WordNet for Building WordNets.Proceedings of COLING-ACL Workshop "Usage of WordNet in Natural Language Processing Systems". Montreal, Canada.

Ide, N., & Véronis, J. (1998). Word Sense Disambiguation: The State of the Art, Computational Linguistics 24(1), 1-40.

Kornilakis, H., Grigoriadou, M., Papanikolaou, K. A., & Gouli, E. (2004). Using WordNet to Support Interactive Concept Map Construction. In Proceeding of the 4th IEEE International Conference on Advanced Learning Technologies (ICALT'04), 600-604.

Magnini, B., Strapparava, C., Pezzulo, G, & Gliozzo, A. (2002). The Role of Domain Information in Word Sense Disambiguation. Natural Language Engineering. Cambridge University Press, 8:359-373.

Miller G., Beckwidth, R., Fellbaum, C., Gross, D., & Miller, K. (1990). Introduction to WordNet: An On-line Lexical Database. International Journal of Lexicography, 3(4), 235-244.

Mihalcea, R. (2006). Knowledge-Based Methods for WSD. In E., Agirre, & P., Edmonds (Eds.), Word Sense Disambiguation: Algorithms and Applications. Springer, 127-132.

Novak, J. D., & Gowin, D. B. (1984). Learning How to Learn. New York: Cambridge University Press.

Palmer, M., Ng, H., & Dang, H. (2006). Evaluation of WSD Systems. In E., Agirre, & P., Edmonds (Eds.), Word Sense Disambiguation: Algorithms and Applications. Springer, 75-106.

Richardson, R., Goertzel, B., Fox, E. A. (2006). Automatic Creation and Translation of Concept Maps for Computer Science-Related Theses and Dissertation. In A. J. Cañas & J. D. Novak (Eds.), Concept Maps: Theory, Methodology, Technology. Proceedings of the Second International Conference on Concept Mapping (Vol. 2, pp. 32-35). San José, Costa Rica: Universidad de Costa Rica.

Reichherzer, T. R., Cañas, A. J., Ford, K. M., & Hayes, P. J. (1998). The Giant: An Agent-based Approach to Knowledge Construction and Sharing. Proceedings of the Eleventh International Florida Artificial Intelligence Research Society Conference. AAAI Press, 136-140.

Simón, A., Ceccaroni, L., & Rosete, A. (2007). Generation of OWL Ontologies from Concept Maps in Shallow Domains. In D., Borrajo, L., Castillo, & J.M., Corchado (Eds.), CAEPIA 2007. LNAI, Vol. 4788. Springer-Verlag, 259-267.