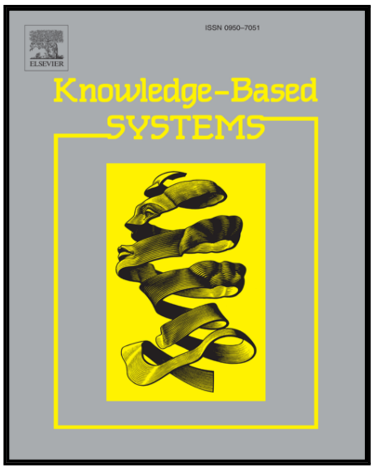
Accepted Manuscript



Spreading semantic information by Word Sense Disambiguation

Yoan Gutiérrez, Sonia Vázquez, Andrés Montoyo

PII: S0950-7051(17)30287-3
DOI: 10.1016/j.knosys.2017.06.013
Reference: KNOSYS 3941

To appear in: *Knowledge-Based Systems*

Received date: 27 October 2016
Revised date: 5 June 2017
Accepted date: 8 June 2017

Highlights

- A Word Sense Disambiguation (WSD) knowledge-based system is presented
- A multidimensional network obtained from different lexical resources is described
- A detailed description of graph creation and algorithms used is provided
- A detailed comparative among relevant WSD approaches is presented
- An extensive evaluation and analysis of the obtained results is provided
- A description about the benefits of this WSD proposal in other Natural Language Processing tasks is provided

Spreading semantic information by Word Sense Disambiguation

Yoan Gutiérrez, Sonia Vázquez, Andrés Montoyo

Department of Software and Computing Systems
University of Alicante
Apartado de correos 99, 03080 Alicante, Spain
Tel.: +34 965 903400 2961
Fax: +34 965 90 93 26

Abstract

This paper presents an unsupervised approach to solve semantic ambiguity based on the integration of the Personalized PageRank algorithm with word-sense frequency information. Natural Language tasks such as Machine Translation or Recommender Systems are likely to be enriched by our approach, which includes semantic information that obtains the appropriate word-sense via support from two sources: a multidimensional network that includes a set of different resources (i.e. WordNet, WordNet Domains, WordNet Affect, SUMO and Semantic Classes); and the information provided by word-sense frequencies and word-sense collocation from the SemCor Corpus. Our series of results were analyzed and compared against the results of several renowned studies using SensEval-2, SensEval-3 and SemEval-2013 datasets. After conducting several experiments, our procedure produced the best results in the unsupervised procedure category taking SensEval campaigns rankings as reference.

Key words: Natural Language Processing, Graph-based, Knowledge-based, Word Sense Disambiguation, PageRank

1. Introduction

The main goal of knowledge technologies is to add meaning to the huge quantity of information that our multilingual societies generate every day. In order to render the knowledge life-cycle progressively more automated, a wide range of advanced techniques are required. The analysis of large data collections therefore implies the need to develop different approaches to automatically represent and manage a high-level of meaningful concepts [1]. Moreover, in order to be able to create efficient Natural Language Processing (NLP) systems it is necessary to transform the information extracted from the words in plain text to meaningful word-senses, that is, to a concept level.

One of the main and frequently studied problems in NLP is how to measure semantic similarity and relatedness. In this case, the problem is to estimate how similar or related two words are in order to establish different semantic relations. Two main approaches are used to tackle this problem: knowledge based approaches [2], [3] and corpus-based approaches [4]. The first one requires lexical resources such as WordNet (WN), Roget's thesaurus, etc. to obtain semantic similarities, while the second one uses co-occurrences to measure the similarity among words.

Email addresses: ygutierrez@dlsi.ua.es, svazquez@dlsi.ua.es, montoyo@dlsi.ua.es (Yoan Gutiérrez, Sonia Vázquez, Andrés Montoyo)

Measuring semantic similarities is related to the problem of Word Sense Disambiguation (WSD) which is a common difficulty in NLP. In WSD the goal is to determine the senses of the words in a text when a word-sense may be different depending on the context in which it appears. For example, let us consider the word *"running"* in different contexts:

- The software is ***running***.
- The athlete is ***running***.
- The coach put great emphasis on ***running***.

We can appreciate that *"running"* has different meanings according to the different contexts in which it may appear [5]. In the first example, the word *"running"* has the meaning of operability, while in the last two examples the meaning is related to the same general concept (Sport).

Detecting the correct senses of words is a difficult task for machines. In order to solve this problem, a wide variety of systems have been developed to determine the right meaning of words in their different contexts. In fact, WSD has been proven to be necessary to improve the results of other tasks such as Machine Translation [6], [7], [8], Opinion Mining [9], [10] or Recommender Systems [11] among others. WSD is therefore considered to be an essential task for all these applications [12], and it is for this reason that many research groups are working on WSD using a wide range of different approaches.

Like other NLP tasks, WSD systems use different kinds of resources, corpora and techniques to obtain the correct senses of words, thus making it difficult to evaluate which system (or procedure) is better since they do not use the same corpus or repositories to obtain the correct word-senses. The need to evaluate different tasks in NLP resulted in the creation of evaluation campaigns like for example SensEval[1]. The main goal of this campaign was initially to measure the strengths and weaknesses of WSD systems with regard to different words, different aspects of language and different languages [13], [14], [15], [16], [17], etc. However, subsequent campaigns added new tasks such as: semantic roles; web people search; affective text; etc. Nevertheless, comparing the best systems of different campaigns SensEval-2 [18], [19], SensEval-3 [20], [21], SemEval-1 [22], SemEval-2 [23], SemEval-2013[2] and SemEval-2015[3], the results are still very low in terms of accuracy.

According to the results of each campaign, the best WSD systems are those based on supervised approaches [24], [21], [22]. However, these systems need to train with large amounts of hand-tagged data [25], [26], which require an effort that cannot easily be repeated for new domains and languages. The most common WSD training corpora include: SEMCOR [27] and DSO corpus [28]. Semcor provides three folders brown1 (around 1000 tagged senses per file), brown2 (around 1000 tagged senses per file) and brownv (around 200 tagged senses per file) each one including 103, 83 and 166 semantically tagged Brown Corpus files respectively. DSO provides a data set of 2,094 examples which are sentences taken from the Brown corpus and the Wall Street Journal corpus that are sense-tagged word occurrences, 121 nouns and 70 verbs. This corpus was first reported in [28], and it contains about 192,800 sense-tagged word occurrences of the 191 most frequently occurring and ambiguous words of English. Owing to the fact that creating these corpora is very time consuming and that only small amounts of training data are available, it can be seen how these supervised systems improve the baseline Most Frequent Sense (MFS) by a small margin (see Table 1).

In order to avoid hand-tagging data task, there is another alternative that does not depend on hand tagged-data: knowledge-based systems which have to use pre-built lexical conceptual resources. Therefore, in order to develop better knowledge-based systems it is necessary to have greater deployment of resources. This fact does not make knowledge-based systems less expensive than supervised systems, but they can cover larger scopes with the same resources and without conducting new training scenarios.

[1]http://www.senseval.org/, last access 10/2016

[2]https://www.cs.york.ac.uk/semeval-2013/task12/index.php

[3]http://alt.qcri.org/semeval2015/task13/index.php?id=results

Knowledge based systems exploit information provided by a Lexical Knowledge Base (LKB) without needing any kind of corpus information. Various procedures for knowledge-based systems have therefore been developed in order to tackle the problem of obtaining hand-tagged data. Among the different knowledge-based procedures for WSD that have been developed in the last few years, this study focuses on graph-based approaches due to their relevant results in SensEval-2, SensEval-3 and SemEval-2013 campaigns: [29], [30], [31], [32], [33] and [34]. The results confirm that graph based approaches are an appropriate means to solve ambiguity.

Thus, graph based approaches have become very popular in recent years. These approaches use different kinds of resources, Lexical Knowledge Bases (LKB) that involve at least a lexical semantic network (i.e. WN), and can be enriched with other domain concepts, as will be detailed in the subsequent sections.

Among the best procedures we can highlight those that obtained relevant results such as the Structural Semantic Interconnection approach [33], which is able to create structural characteristics of word-senses in a context using lexical chains, and those procedures that integrate several semantic resources (Lexical Knowledge Based): [35], [32] and [30].

Most of the works are based on WordNet [5] because this resource has been accepted by the scientific community as a source reference to evaluate different systems, becoming the current word-sense repository in most of the SensEval campaigns. As a consequence other resources based on WordNet have been developed, also known as ISR (Integrated Semantic Resources): WordNet Domains [36]; MultiWordNet (MWN) [37]; EuroWordNet (EWN) [38]; Meaning [39]; UBY [40]; BabelNet [41]; ISR-WN [11]; etc.

All of these resources attempt to provide semantic networks with a common interface. In most cases, ISRs apply lexical integration with a few conceptual resources. However, ISR-WN provides semantic information about additional types of resources such as Semantic Classes (SC) [42] and WordNet Affect (WNA) [43], [44], which are useful in the WSD task. It has been shown to improve WSD results in works of [45], [46], [47].

Thus, the proposal presented in this work uses ISR-WN (Integration of Semantic Resources Based on WordNet) [11] [48] [49] as a knowledge source.

Our LKB also integrates: ISR-WN [48] + XWN (eXtended WordNet) 1.7 [50] + XWN 3.0 [51]+ pSemcor[4] [52] + Word-Sense Frequencies from WN 1.7 [53]. From here on, all of these resources will be integrated in our LKB.

Note that this LKB links WordNet word-senses across the semantic links provided by all aforementioned resources. Each word-sense is weighted according to the frequency information provided by the file cntlist[5] of WN1.7. Notice this frequency statistic has been estimated on SemCor[6].

The main goal of this research is to solve semantic ambiguities by implementing a Multidimensional Semantic Analysis in an unsupervised and knowledge based system with the use of a graph based procedure. The advantages provided by our LKB for dealing with WSD tasks are derived from the diversity of the resources integrated.

Although our proposal is based on previous works regarding the use of PageRank algorithm for WSD, as demonstrated in the evaluation section (see Section 4) our new vision surpasses the results provided by previous approaches [32], [30]. The main contribution of this research consists of applying the PageRank algorithm by using different semantic dimensions (resources) and adding word-sense frequency knowledge in order to improve the results.

This document is structured as follows:

Section 2 provides a brief description of the main knowledge base resource of our research (ISR-WN) followed by an explanation of different graph-based approaches, focusing on the PageRank algorithm.

[4]Word-sense pair relations (collocations) in SemCor providing 310041 semantic relations with WN 1.6 and WN 2.0.

[5]https://wordnet.princeton.edu/man/cntlist.5WN.html, last access 10/2016

[6]http://web.eecs.umich.edu/˜mihalcea/downloads.html #semcor, last access 10/2016

Section 3 describes our proposal for WSD. Section 4 explains the experimental setup and presents the results, and Section 5 discusses the results. Finally, in Section 6 we provide the conclusions and a proposal of further works.

2. Lexical Knowledge Bases and Graph-based Procedures

This section provides appropriate descriptions regarding both the LKB needed to support our WSD procedure (see Section 2.1) as well as detailed descriptions of graph-based procedures (see Section 2.2) that support our algorithms.

2.1. Lexical Knowledge Bases

In the vast majority of Natural Language Processing tasks it is necessary to use external resources: dictionaries; thesauri; ontologies; etc. As we have mentioned in the previous section, one of the most frequently used resources is WordNet that serves as a basis to develop other enriched resources.

Due to the fact that we need special information to perform our graph based procedure we have used as part of our Lexical Knowledge Base (LKB) a previously developed resource (ISR-WN) [11], [49]. Figure 1 shows its conceptual model. The description of this resource is available for both WN1.6 [48] and WN2.0 [11], [49].

ISR-WN is a lexical knowledge database that provides an integration of several resources that are aligned to WordNet but are physically isolated. ISR-WN is represented as a complete undirected graph $G = (V, E)$ where their nodes (vertex) can be represented as concepts (synonyms: graph nodes of different resources) $C = \{c_1, c_2, .., c_r\}$ or as word-senses (synsets: graph nodes from WordNet) $S = \{s_1, s_2, .., s_m\}$, that is, $V = C \bigcup S$. The relation between two nodes v_i and v_j is represented with an edge $e_{i,j}$.

In ISR-WN, a WordNet word-sense is connected to WordNet Domains (WND) [36] and WordNet Affect (WNA) [54] concepts, SUMO [55] categories or Semantic Classes (SC) [42] across internal links, as Figure1 shows. In addition ISR-WN includes in its assets a resource named as SentiWordNet (SWN) [43]. SWN provides sentiment polarity information related to word-senses. However, our WSD procedure does not take into account polarities because this information does not contribute to discriminate among word-senses because SWN does not link word-senses.

All the resources contained in our LKB are related through WordNet synsets (word-senses). So, we have a set of word-senses that can be related to different resources.

In order to have a better idea of the distribution of the LKB considered in this WSD work we can show the following statistic:

- ISR-WN:
  - WND (with 172 labels),
  - WNA (with 300 labels),
  - SUMO (with 568 labels), SWN (with 117,659 labels), SC (with 1,231 labels),
  - WN (with 99,643 synsets for WN 1.6[7] and 115,424 synsets for WN2.0 [8])
- XWN (551,551 relations for XWN1.7 and 19,387 relations for XWN3.0).
- pSemcor (31,0041 semantic links among co-occurring pair word-senses) obtained from Semcor (WN 1.6 and WN 2.0 versions)

The connections established among all semantic resources of our LKB can be used to obtain a conceptual multidimensionality from one sentence. Figure 2 shows how to conceptualize a sentence and how words are related throughout a semantic network. This indicates by applying graph connections to a sentence, we are able to use network techniques to solve the semantic ambiguity by favouring those word-senses which have a more accurate representation in the context.

[7]66,025 nouns, 17,915 adjectives, 3,575 adverbs and 12,127 verbs.

[8]79,689 nouns, 18,563 adjectives, 3,664 adverbs and 13,508 verbs.

2.2. Graph based procedures: techniques and resources

As the SensEval and SemEval campaigns have demonstrated, the best unsupervised WSD approaches in the last few years have been those that are graph based. In this section we present different techniques that are used to deal with network issues in order to understand our proposal, which is described in Section 3.

Two of the first proposals were those provided by Dijkstra [56] and [57]. In these works, the authors presented different solutions to deal with the problem of managing a large number of links among graph elements in order to obtain the minimal path (using Breadth First Search).

The usage of network structures to create links among different nodes requires the application of a variety of algorithms and techniques. In order to work with these kinds of structures it is necessary to: obtain minimal paths and groups of elements; detect relevant elements; and apply algorithm optimization, among others.

Various techniques based on the idea of network connections have been developed which have been used in a wide range of scientific contexts. For example, the PageRank technique [58], has been used by the Google Internet search engine [59] in order to rank web sites by popularity while the Cliques technique presented in [60], indegree [61], closeness [62] and betweenness [63], have been used in social networks to establish relationships among users.

In NLP, graph based techniques have been applied to different kinds of tasks. For example, in WSD, various knowledge based systems using graphs have been developed. We could, for example, refer to those that use Structural Semantic Interconnections (SSI) [64]. Another approach is presented in [35] in which WordNet and Framenet are integrated. We will focus on those approaches that use the PageRank technique [32], [65], [66] and [30], in which a Lexical Knowledge Base (LKB) combined with eXtended WordNet is used.

According to the results obtained by graph based approaches in WSD, and in comparison to the best supervised approaches, notably the results are increasingly similar. Table 1 shows the results of the best systems in SemEval according to different categories: Supervised (S); Unsupervised (U); Knowledge Based (KB); and Weak Supervised (WS). Moreover, Table 1 also shows the results obtained by baseline systems using the Most Frequent Sense (MFS).

As observable in the last column in Table 1, the results of unsupervised systems are nearing those of supervised systems. It is important to note that the baseline (MFS) is highly competitive in comparison wit the results of both supervised and unsupervised systems. The last column in Table 1 shows a comparison of the different Recall results between supervised (or weakly supervised) and unsupervised systems.

In our proposal we use the PageRank and Personalized PageRank techniques because of the good results obtained in previous campaigns and their adaptability to different semantic resources. Several studies such as [30], [61] compare various measures and after considering different options we choose the Personalized PageRank as the most appropriate to our procedure.

The PageRank algorithm [58] provides a procedure with which to rank graph vertices according to their relative structural importance. The main idea of PageRank is that whenever a link from v_i to v_j exists in a graph, a vote from node i to node j is obtained, and the rank of node j therefore increases. Moreover, the vote from i to j is more valuable depending on the rank of node i: the more important node i is, the more strength its vote will have. In order to obtain the likelihood of arriving at node i, PageRank can also be considered as a probability distribution that represents the probability of a random walk over the graph ending at node i.

The algorithm representation is as follows:

Let G be a graph with N vertices $v_1 ... v_n$ and d_i be the outdegree (number of connections) of node i. Let M be an $N \times N$ transition probability matrix where $M_{i,j} = \frac{1}{d_i}$, if a link from node i to node j exists, and zero otherwise. The PageRank vector (Pr) is represented in graph G using Equation 1 which has been adapted to the WSD scenario by avoiding nodes without connections [32].

Table 1: SemEval competitions comparison (Precision (P), Recall (R))

System	**SensEval-2 (2001)**				
	P(%)	**R(%)**	**Type**	**Technique**	**R(BS)-R(BU) R(BS)-R(MFS)**
Best system (BS)	69	69	S	Pattern learning [67]	-
Best unsupervised system (BU)	57.5	56.9	U/KB	Mutual Information [68]	12.1
MFS (baseline)	66.9 61.7	64.6 61.7	-	Most Frequent Sense [69]	4.4
System	**SensEval-3 (2004)**				
	P(%)	**R(%)**	**Type**	**Technique**	**R(BS)-R(BU) R(BS)-R(MFS)**
Best system (BS)	65.1	65.1	S	Genetic algorithm [70]	-
Best unsupervised system (BU)	58.3	58.2	U/KB	Relevant Domains [71]	6.9
MFS	62.4 60.9	62.4 60.9	-	Most Frequent Sense [21]	2.7
System	**SemEval-1(2007)**				
	P(%)	**R(%)**	**Type**	**Technique**	**R(BS)-R(BU) R(BS)-R(MFS)**
Best system (BS)	59.1	59.1	S	Maximum entropy [72]	-
Best unsupervised system (BU)	52.7	52.7	U	Pseudo-syntactic dependencies [73]	6.4
MFS	51.4	51.4	-	Most Frequent Sense) [22]	7.7
System	**SemEval-2(2010)**				
	P(%)	**R(%)**	**Type**	**Technique**	**R(BWS)-R(BU) R(BWS)-R(MFS)**
Best system (BWS)	57	55.5	WS	Graph based [74]	-
Best unsupervised system (BU)	51.2	49.5	U/KB	Graph based [74]	6
MFS	50.5	50.5	-	Most Frequent Sense [23]	5
System	**SemEval 2013**				
	P(%)	**R(%)**	**Type**	**Technique**	**R(BS)-R(BU) R(BS)-R(MFS)**
Best System (BS)	64.9	64.5	U	Ppr+Freq	-
Best unsupervised system (BU)	64.9	64.5	U	Ppr+Freq	0
MFS	63.0	63.0	-	Most Frequent Sense [75]	1.9

$$Pr = cMPr + (1 - c)v \tag{1}$$

where:

M: is the transition probability matrix.

v: is a $N \times 1$ vector whose elements are $\frac{1}{N}$.

c: is the so called *damping factor*, a scalar value between 0 and 1. These values are usually between 0.85 and 0.95.

Pr: is the probability vector obtained from the random walk to reach each node.

The first term of the sum in Equation 1 represents the voting scheme. The second term represents the probability of a surfer randomly jumping to any node, e.g. without following any paths on the graph. The damping factor, which is usually set in the [0.85..0.95] range, models the way in which these two terms are combined in each step.

The second term in Equation 1 can also be seen as a smoothing factor that makes any graph fulfill the property of being aperiodic and irreducible, and thus guarantees that the PageRank calculation converges to a unique stationary distribution [32].

In order to initialize the Pr vector we assume that each node has an initial value of $\frac{1}{N}$. The vector v has traditionally been initialized to $\frac{1}{N}$. However, according to other scholars [32], [76], v can have different values associated with its elements in order to highlight the most important nodes.

To establish the new approximation of v, several authors have applied the PageRank algorithm to WSD assuming that nodes will be assigned with different values. For example, [30] carried out six experiments using different similarity measures for v in each experiment by employing WordNet similarity[9]. Meanwhile [32] presented the Personalized PageRank in which v is initialized with the word-senses in the input text in order to disambiguate them. In our work we use a variant of the Personalized PageRank empowered with word-sense frequencies (utilizing the normalized values of word-sense frequencies) and the LKB constituted by the semantic connections obtained from ISR-WN, eXtended WordNet and word-sense pair relations of SemCor.

3. Ppr+Freq: Word Sense Disambiguation

As presented in previous sections, and considering the good results obtained by initializing v with different values, we propose a new approach that initializes v with word-sense frequencies. In our case, each element of v will be initialized with values between [0-1].

Word-sense frequencies obtained from each synset (word-sense) of WordNet are used as the first value of the initial importance (v) of each word-sense (see the example of Table 2).

Table 2: Normalized frequencies

Key word-sense	Freq[10]	Normalized frequency (Frequency/Sum)
struggle%1:04:01::	16	0.301886792
struggle%2:41:00::	13	0.245283019
struggle%1:04:02::	11	0.20754717
struggle%2:35:00::	7	0.132075472
struggle%2:38:00::	3	0.056603774
struggle%2:33:00::	2	0.037735849
struggle%1:04:00::	1	0.018867925
Sum	53	

The process represented in Figure 3 used to obtain the correct word-sense of each word comprises 3 steps:

1. Initial graph creation (see Section 3.1)
2. Personalized PageRank application (see Section 3.2)
3. Correct word-sense selection (see Section 3.3)

The first step is to build a subgraph that contains the context words to be disambiguated rather than building a complete graph with all the words in the LKB. Figure 3 shows the overall workflow of how the Ppr+Freq procedure is implemented:

Each step in our Ppr+Freq procedure is described below.

3.1. Initial Graph Creation

The WSD process begins when a sentence or paragraph is selected. All words are extracted and listed $L = [w_1, w_2, ..., w_l]$. The next steps consist of:

I Grammatical discrimination

II Creation of minimal paths (i.e Breadth First Search) among all the word-senses, followed by the creation of the initial graph without repeated elements (G_D).

The aim of this process is to create an initial sub-graph (G_D) from the LKB that will contain all the possible word-senses of each word to be disambiguated. It will also contain the vertexes (concepts and other word-senses) that are part of the minimal paths (using Breadth First Search) among these word-senses. Words that do not appear in the sentence or paragraph to be disambiguated are discarded as input words. We have adopted this alternative because recent studies have demonstrated the accuracy of creating specific smaller subgraphs for each target context [32], [30], [31], [61]. A graphical overview of this process can be seen in Figure 4.

The first sub-step (I) of the process comprises the grammatical categorization of each word (w_k). As the SensEval corpora has been used to evaluate our system, we have adopted the SensEval categories to tag each word. However, POS-tagging tools like Freeling [77], NLTK [78], and others can be used to this end. A detailed description of the SensEval data schema and the conversion to the SemCor format is

[9]http://wn-similarity.sourceforge.net/, last access 10/2016

available in Rada Mihalcea[11].

For each word we therefore have a p category in which w_k^p is a POS-tagged k-word. In order to obtain the set of different word-senses for each w_k^p word we define $Senses_D = (w_k, p)$ as $Senses_D = (w_k^p)$. This function obtains the subset of word-senses of each categorized word $sw_k^p = \{sw_k^{p1}, sw_k^{p2}, ..., sw_k^{pn}\}$. In order to obtain all the word-senses of the target context L we apply the function $Senses_D(L^p)$.

These word-senses are represented in the WN dictionary and in our LKB, these are connected to concepts (i.e. labels w_k from WND, WNA, SC and/or SUMO) $V_k = \{v_{k1}, v_{k2}, ..., v_{km}\}$, where $V = C \bigcup S$ (C:Concepts; S:Senses). It is important to stress that the links used by our LKB also consider the relationships of the WN word-senses presented in XWN1.7, XWN3.0 and pSemcor.

Having obtained the grammatical categories and all the possible word-senses of each word we proceeded to sub-step (II). This consists of building a subgraph G_D from the LKB graph (G_{KB}) using the minimal path among all the nodes $v_k^p \in Senses_D(L^p)$ by means of the BFS (Breadth First Search) algorithm. In G_D each word-sense sw_k^{pj} is a concept $v_k^p \in V_k^p$. The set of minimal paths among all nodes that represents a sentence through the G_{KB} is obtained using the function $min_{v_k^p}$. Therefore, $min_{v_k^p}$ collects the minimal paths of all word-senses of each word w_k^p, taking into account different concepts and sematntic links. The nodes and relations included in G_D are represented as $G_D = [\bigcup_{k=1}^{l} = min_{v_k^p} / v_k^p \in V_k^p]$, where G_D is a subgraph of G_{KB}.

It is important to note that the set of concepts V_k^p is represented in the LKB graph (G_{KB}) in which each node $v_h^p \in \bigcup_{h \neq k} V_h^p$, where $h, k \in L$.

Therefore we can say that our initial sub-graph, also known the disambiguation graph (G_D), is considered as a complete graph $G_D = (V, E)$ in which each node represents a concept or a word-sense of the entire LKB. Each relation between v_i and v_j is represented as $e_{i,j}$. It is defined as $G_D \subseteq G_{KB}$.

The information considered to create G_D was described in more detail in Section2.1.

Finally, an initial graph (G_D) is obtained. G_D contains synsets (word-senses) of WN, the semantic concepts (WND, WNA, SC and SUMO) and the new semantic links provided by XWN1.7, XWN3.0 and pSemcor.

The resources included in the LKB are easily configurable to enable the different experiments to be conducted. In the case of selecting WND in the LKB, the Factotum domain is not taken into account because it is a generic domain that does not provide relevant information [79].

3.2. Personalized PageRank Application

Once G_D has been obtained we apply the Ppr+Freq procedure (Step 2) that is described in Table 3 (G_D refers M). In order to set up the initial weightings of G_D nodes, we assign to each element of v the result of $\frac{1}{N}$ to all nodes that do not correspond with the word-senses of the target words. In the case where one node belongs to a word-sense of a target word, it will be initialized using the normalized word-sense frequency values from SemCor (provided by the file cntlist of WN1.7). These values will be between a threshold [0-1] for each target word-sense.

For example, for WN 2.0 the word "struggle" has seven word-senses and its respective frequencies of use. Table 2 shows how the frequencies are normalized.

In our procedure, and following the recommendations of [32] we establish a limitation of 30 iterations because it has been proven that from that point the Pr values have few variations. In addition, we developed some preliminary testing to the algorithm in terms of the number of iterations and we experienced that, over this number of iterations the Pr scores tend to be stable. The value used as a damping factor (c) is 0.85.

[11] http://web.eecs.umich.edu/ mihalcea/downloads.html, last access 02/2017

Table 3: Ppr+Freq Procedure

procedure PageRank()
$c \leftarrow 0.85 \Rightarrow$ damping factor, a scalar value between 0 and 1.
$M \leftarrow NxN$ transition probability matrix where $M_{i,j} = 1/d_i$, an edge from i to j exists; 0 otherwise
$v \leftarrow Nx1$ vector whose elements in i are initialized in sf_i/wsf; otherwise $1/N$
$Pr \leftarrow Pr_i = 1/N$
iteration $\leftarrow$ 30
while iteration > 0 do
$Pr = c * M * Pr + (1 - c) * v \Rightarrow$ iteration for measuring Pr scores
iteration $\leftarrow$ iteration - 1
end while
end procedure

Note: v represents a vector of the G_D. In this case, only those nodes that correspond to word-senses of the target words have a frequency value set, the rest of the nodes have set $1/N$ (i.e. for SUMO, WND, WNA, SC, word-senses from non-target word). sf_i indicates the frequency (See Table 2) regarding a word-sense of the target word; wsf indicates the sum of all frequencies of the word-senses of a specific word.

3.3. Correct word-sense selection

Step 3 comprises the selection of correct word-senses. Each word-sense of the target words has a ranking value associated which can be found in the Pr vector. In order to select the appropriated senses for each word, candidate word-senses that maximize their score in Pr are chosen. Selecting the correct word-senses poses a problem for WSD ranking algorithms because it may be possible find a tie in the ranking score among different word-senses. In our procedure, this is solved by initializing word-sense nodes to their corresponding normalized word-sense frequency values. After conducting several experiments, no ranking score ties occurred.

4. Experimental Setup

In this section different experimental results of our procedure are presented. Firstly, we describe different experimental results conducted with the SensEval2 and SensEval3 corpora and in Section5 we present the results obtained in SemEval-2013. These experiments establish comparisons between our procedure and the state-of-the-art in order to discover improvements to the WSD task. Moreover, we describe notable reviews of WSD campaign systems and highlight their essential features. Next, different evaluations using the aforementioned corpora are carried out.

The evaluation for WSD has been divided into two stages: stage 1 uses the SensEval-2 corpora; and stage 2 uses the SensEval-3 test corpora. In each one of these evaluations, the different behavior of the Ppr+Freq procedure has been analyzed on the basis of different LKB configurations. Moreover, in Section5 we present a comparison with more recent systems.

4.1. SensEval2 Evaluation

In this sub-section, we present eleven approaches with their respective LKB configurations and WSD results. Six experiments (Exp2, ..., Exp7) correspond to the Ppr+Freq procedure, where Exp6 is a voting approach which involves Exp1, ..., Exp5, and five approaches (Exp1, Exp8, ..., Exp11) taken as baselines. In Exp6 the most voted word-sense among all the approaches involved in the voting process is proposed as the correct word-sense (Exp1, ..., Exp5. In this experiment (i.e. Exp6), MFS (i.e. the outputs of the Exp1) can also be used if there is no coincidence among the voters in response to the definitive result.

Note that Exp1 uses MFS and the first word-sense shown in WordNet. This fact is taken into account because some words are not included in the frequency list (obtained from the file cntlist from WN1.7). In order to solve this problem, and always providing Exp1's response, the First Word-sense provided by WordNet is chosen as a second option.

Exp8, ..., 11 are described in more detail in [32] However, to recap the construction of a semantic network that represents the textual input context, results in a disambiguation sub-graph. This sub-graph employs an enriched LKB base on BFS among all word-senses that represent the involved words. Subsequently, three different approaches using the PageRank algorithm are applied. The first one is named Spr (it is not used as baseline) which executes PageRank over G_D without initializing with weightings the involved word-senses. The second, Ppr (Exp8 and Exp10), is similar, but by contrast it assigns weightings to the linked word-senses with words from the input context. The final approach, Ppr_w2w (Exp9 and Exp11), performs a similar procedure to Ppr, but in this case a G_D is built for each target word. In this approach, the initial probability of the word-sense is focused on the adjacent words instead of the target word-senses.

As we can see in Table 4, the first resource corresponds to the use of word-sense frequencies. LKB1.7 and LKB3.0 correspond to new semantic relations that enrich the WN resource, such as KB. ISR-WN (which includes WN, WND, WNA, SUMO and SC) offers new semantic perspectives by conceptualizing sentences from different points of view (see Figure 2). Collocation statistics among word-senses are provided by pSemcor in concordance with the SemCor annotated corpora.

Having stated these considerations, we shall now evaluate the aforementioned approaches.

As shown in Table 4, when MFS is used as a baseline, the procedure is able to obtain relevant results in comparison with the WSD systems from the state-of-the-art mentioned in Table 1. Moreover, the evaluations taken as a reference baseline (Exp 8, 9, 10 and 11) by Agirre and Soroa in 2009 [32] represent the Personalizing PageRank algorithm without the word-sense frequency use. As we can observe, these baselines achieve high ranking results. However, it is also observable that all the Ppr+Freq experiments outperform all the baselines. This leads us to believe that the use of the word-sense frequency in vector v of the Personalizing PageRank provides the WSD process with relevant information. This indicates that our integrated procedure may be better than the baselines: MFS, Ppr and Ppr_w2w.

Depending on the configurations performed, a variable behavior of disambiguation processes is generated in concordance. This is in order to correctly disambiguate some clues from words. We propose to deal with these factors by including the voting approach concerning the responses from Exp1.., Exp5. This voting approach (Exp6) achieves the top results as obtained by our research and it represents a significant improvement on the state-of-the-art. Please note that Experiments 8...11 correspond to Personalized PageRank (Ppr) and Personalized PageRank_w2w (Ppr_w2w) of [32], in which the LKB1.7 and LKB3.0 resources have been used in their LKB configuration.

A similar approach is Exp7, which applies Ppr+Freq by taking as LKB the integration of LKB1.7, LKB3.0, ISR-WN and pSemcor. It would appear that not all the information needed to effectively implement the WSD process was considered during the creation of G_D for Exp7. This was owing to the creation of a huge LKB involving all the semantic networks at the same time. The particularity of this LKB is that coarse-grained semantic resources populate most of the G_D, because its internal links are reduced and the BFS criterion is used to build the G_D. This must therefore be studied as further work in order to determine which are the most representative resources. We would like to point out that between Exp8 and Exp2 the only difference consists of including the word-sense frequency in vector v of PageRank. This detail gives strength to our proposal, which exceeds 6.2 and 2.6 percentage points in Exp8 and Exp1, respectively.

We also analyzed the behavior of the different Parts of Speech (POS) in the disambiguation process. Figure 5 provides graphic data in which Ppr+Freq outperforms all the results provided by the baselines. The only exceptions are Ppr and Ppr_w2w in the case of noun disambiguation, though with very small margins.

Comparing the results obtained by Ppr+Freq with the SensEval-2 ranking as described in Table 4, these results are superior to several supervised systems. Ppr+Freq would occupy the second place on the ranking provided if the baseline shown in Table 5 was not considered. We should stress that this baseline was not included as a participant system and was subsequently computed [69]. Preiss' work contains frequency information that could be very valuable for our procedure. Our procedure has demonstrated the ability to improve the frequency baseline involved in the Ppr+Freq ranking process. Thus, if frequency information obtains 60.4% of Recall and if the MFS Preiss Recall is 64.6%, then by using the Preiss frequency information our procedure would be better than the Preiss baseline, because Ppr+Freq always outperforms the MFS used.

Table 4: General results of Ppr+Freq evaluated with SensEval-2 (Precision(P), Recall (R))

Experim	**Resources**					**Voting**	**R**	**P**
	MFS +First Sense [12]	**LKB 1.7**	**LKB 3.0**	**ISR - WN**	**pSemcor**			
Exp 1MFS	X						0.604	0.609
Exp 2		**X**					**0.630**	**0.637**
Exp 3			X				0.614	0.619
Exp 4				X			0.615	0.621
Exp 5					**X**		**0.623**	**0.628**
Exp 6						**X**	**0.641**	**0.646**
Exp 7	**X**	**X**	**X**	**X**	**X**		**0.629**	**0.634**
8- Ppr		X					0.568	-
9-Ppr_w2w		X					0.586	-
10- Ppr			X				0.535	-
11-Ppr_w2w			X				0.558	-

For this reason, we plan to study other frequency statistics like Preiss's frequency to measure its impact in our WSD approach.

4.2. SensEval3 Evaluation

The evaluation carried out with the SensEval-3 test corpus follows the same principles applied in the previous section. Table 6 shows some Ppr+Freq approaches that are able to improve on the baselines proposed in this work. The fact that some Ppr+Freq approaches (Exp 3 and Exp 4) do not attain top results is owing to the variability of both test corpora. Moreover, Experiments 5, 6 and 7 in this section obtain the best results with both corpora. Therefore, Ppr+Freq again proves its superiority against all the baselines involved in this research.

Figure 6 describes the disambiguation results obtained after analyzing each POS. In this scenario, it will be observed that the disambiguation of adverbs is 100% effective when Ppr+Freq is used. This result is only comparable with the approach of [30], which is analyzed in the following section. Other relevant results are that several Ppr+Freq experiments outperform the baselines for all POS, as it is shown in Figure 6.

These results (Experiments 5, 6 and 7), when compared with those of the SensEval-3 system (see Table 7), could be positioned in 6th place of the ranking without taking into account the MFS baseline calculated by [21]. The results obtained with our procedure are therefore better than those obtained with the unsupervised WSD systems, thereby making it the leader in this category (unsupervised).

4.3. Overall analysis of both SensEval competitions

Ppr+Freq is one of the unsupervised and knowledge-based procedures with the best efficacy among those reported in the SensEval-2 and SensEval-3 state-of-the-art, as demonstrated by the aforementioned analysis of its behavior.

It outperforms all the approaches in its category and is able to reduce the difference between unsupervised and supervised results systems. Among the most relevant aspects, we can mention that Ppr+Freq allows adverbs to be disambiguated with an effectiveness of 100%, and that the remaining POS obtain scores that are very close to the best reports from the state-of-the-art.

An overall analysis of the behaviour of WSD in sentences of different sizes has also been considered (see Figure 7). The approaches taken into consideration for this test behave similarly while they progress across all word windows. This indicates that the Recalls for these cases are regulated by the difficulty of the corpora, since the MFS and Personalized PageRank approaches have absolutely no dependency between them, and still show a primary correlation. However, we can affirm that Ppr+Freq is linked to MFS and Personalized PageRank, but outperforms both of these baselines.

Table 5: Comparing Ppr+Freq with SensEval-2 ranking (supervised (S), un-supervised (U))

Rank	*Precision*	*Recall*	*System*	*Supervised*
1	0.690	0.690	*SMUaw*	S
-	0.669	0.646	*Baseline-MFS-Preiss*	-
-	**0.646**	**0.641**	**Exp6 (Ppr+Freq)**	**U**
2	0.636	0.636	*CNTS-Antwerp*	S
3	0.618	0.618	*Sinequa-LIA - HMM*	S
-	0.617	0.617	*Baseline-MFS-Chen*	-
-	0.610	0.609	*(Best)RST+Freq*	U
4	0.575	0.569	*UNED - AW-U2*	U
5	0.556	0.550	*UNED - AW-U*	U
6	0.475	0.454	*UCLA - gchao2*	S
7	0.474	0.453	*UCLA - gchao3*	S
8	0.416	0.451	*CL Research - DIMAP*	U
9	0.451	0.451	*CL Research - DIMAP (R)*	U
10	0.500	0.449	*UCLA - gchao*	S
-	0.444	0.433	*(Best)N-Cliques+RV* [80]	U
-	0.436	0.426	*(Best)N-Cliques+RST* [81]	U
-	0.425	0.424	*(Best)RST*	U
11	0.360	0.360	*Universiti Sains Malaysia 2*	U
12	0.748	0.357	*IRST*	U
13	0.345	0.338	*Universiti Sains Malaysia 1*	U
14	0.336	0.336	*Universiti Sains Malaysia 3*	U
15	0.572	0.291	*BCU - ehu-dlist-all*	S
16	0.440	0.200	*Sheffield*	U
17	0.566	0.169	*Sussex - sel-ospd*	U
18	0.545	0.169	*Sussex - sel-ospd-ana*	U
19	0.598	0.140	*Sussex - sel*	U
20	0.328	0.038	*IIT 2*	U
21	0.294	0.034	*IIT 3*	U
22	0.287	0.033	*IIT 1*	U

Obtaining better word-sense frequency information will therefore provide even better results for Ppr+Freq. It is important to remark that several dimensions (resources) have been used separately, but that not all of them constitute part of our integrated LKB i.e., word-sense frequency information and word-sense collocation. To clarify we have two types of resources: statistical and semantic.

- Statistical: word-sense frequency
- Semantic: Word-sense collocation (by exploring the SemCor corpora), XWN1.7, XWN3.0, WN, WND, WNA, SUMO and SC. These elements have enabled us to create links to cover a word disambiguation that has many or few contextual affinities.

Besides the evaluations presented in Table 4 and Table 6 where we studied the impact of building the G_D depending on specific resource configurations, we decided to go deeper in order to understand the impact of $G'_D s$ semantic relationships and concepts when our Ppr+Freq procedure (loading the whole ISR-WN) hits or misses. To this end, we performed some metrics based on the study of the corpus "d001.key" from SensEval-3, which includes 48 sentences to disambiguate. Many features were set to that aim, for instance: Number of Words; Polysemy per word; Mean Polysemy in a sentence; number of POS from words; mean number of POS types per sentence; number of concepts in each G_D; percentage of each concept in each G_D; number of relationships; and percentage of each type of relationship in the G_D. In order to have an overview of the data distribution in G_D we provide the following information: mean concepts is about 3110; mean relations is about 30542; mean words is about 17; and mean words polysemy is about 6 so the initial mean set is 6x17= 102.

Among the aforementioned features we noticed some irregularities regarding the use of SUMO concepts and Stative relationship. As Figures 8 and

Table 6: General results of Ppr+Freq evaluated with SensEval-3 (Precision (P) and Recall (R))

Experim	**Resources**					**Voting**	**R**	**P**
	MFS +First Sense [13]	LKB 1.7	LKB 3.0	ISR - WN	pSemcor			
Exp 1 MFS	X						0.578	0.579
Exp 2		**X**					**0.598**	**0.600**
Exp 3			X				0.576	0.578
Exp 4				X			0.576	0.578
Exp 5					**X**		**0.611**	**0.612**
Exp 6						**X**	**0.611**	**0.613**
Exp 7	**X**	**X**	**X**	**X**	**X**		**0.618**	**0.620**
8- Ppr		X					0.561	-
9-Ppr_w2w		X					0.574	-
10- Ppr			X				0.485	-
11-Ppr_w2w			X				0.516	-

9 show, when the system misses the percentage of SUMO representativeness in each G_D increments. This means that SUMO provides general concepts that affect the grained WSD. In other words, it makes the G_D too generic, which provokes the inclusion of other concepts that do not fit to the sentence's context.

On the other hand, we also noticed the "Stative" relationship, when included in the G_D, appears to affect the WSD procedure (as indicated by Figures 10 and 11). Due to this possibility, we then proceed to rebuild our experiment based on the file "d001.key" avoiding the "Stative" relationship in the G_D and we noticed that the results did not show any variation. With the aim of dealing more in detail with this problem, our future plan is to experiment Ppr+Freq with different ISR-WN configurations (i.e. loading by separate SUMO, WND, SC, WNA and semantic relationships). This is in order to identify the best configuration in the graph based procedure.

To clarify, in Figures 8, 9, 10 and 11: the scale considers a range 0-1.

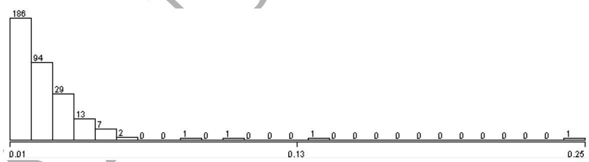


Figure 8: Mean Percentage of SUMO concepts in the G_D for WSD hits

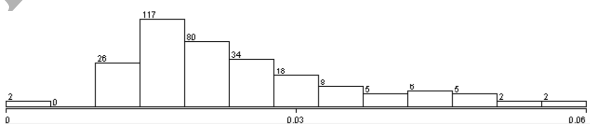


Figure 9: Mean Percentage of SUMO concepts in the G_D for WSD misses

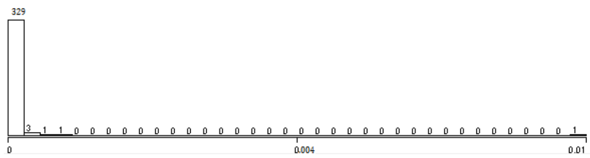


Figure 10: Mean Percentage of Stative relationship in the G_D for WSD hits

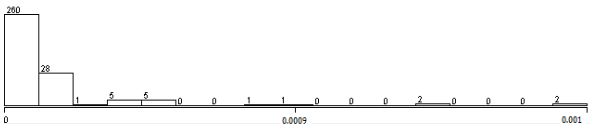


Figure 11: Mean Percentage of Stative relationship in the G_D for WSD misses

4.4. Comparison with other new approaches

In this section, we present a comparative analysis of some of the most relevant WSD approaches.

After studying the state-of-the-art of WSD approaches, we have focused our attention on graph-based procedures involving word-sense frequency.

Table 7: Comparing Ppr+Freq with SensEval-3 ranking (supervised (S), un-supervised (U))

Rank	*System*	*Precision*	*Recall*	*Supervised*
1	*GAMBL-AW*	0.651	0.651	S
2	*Sense-Learner*	0.651	0.642	S
3	*Koc University*	0.648	0.639	S
4	*R2D2 English-All-Word*	0.626	0.626	-
-	*MFS Baseline (GAMBL-AW)*	0.624	0.624	
5	*Meaning All-words*	0.625	0.623	S
-	**Exp 7(Ppr+Freq)**	**0.620**	**0.618**	**U**
-	**Exp 6(Ppr+Freq)**	**0.613**	**0.611**	**U**
-	**Exp 5(Ppr+Freq)**	**0.612**	**0.611**	**U**
6	*Meaning simple*	0.611	0.610	S
-	*MFS Baseline (Yuret)*	0.609	0.609	
7	*LCCaw*	0.614	0.606	-
8	*upv-shmm-eaw*	0.616	0.605	-
9	*UJAEN*	0.601	0.588	S
10	*IRST-DDD-00*	0.583	0.582	U
11	*Sussex-Prob5*	0.585	0.568	-
12	*Sussex-Prob4*	0.575	0.55	-
13	*Sussex-Prob3*	0.573	0.547	-
14	*DFA-Unsup-AW*	0.557	0.546	U
15	*KUNLP-Eng-All*	0.51	0.496	U
16	*IRST-DDD-LSI*	0.661	0.496	U
17	*upv-unige-CIAOSENSO-eaw*	0.581	0.48	U
18	*merl.system3*	0.467	0.456	-
19	*upv-unige-CIAOSENSO3-eaw*	0.608	0.451	U
20	*merl-system1*	0.459	0.447	-
21	*IRST-DDD-09*	0.729	0.441	U
22	*autoPS*	0.49	0.433	U
23	*clr-04-aw*	0.506	0.431	-
24	*autoPSNVs*	0.563	0.354	U
25	*merl.system2*	0.48	0.352	-
26	*DLSI-UA-All-Nosu*	0.343	0.275	-

Our procedure includes the use of centrality measures based on the weighting assigned to a relationship between node pairs. As can be observed in Table 8 and Table 9, the lowest results for this type of system (centrality systems) corresponds to N-Cliques+RV [80] and N-Cliques+RST [81], which apply clustering techniques. However, note that neither take into consideration the weighting assigned in the relationship between node pairs.

This fact weakens both proposals in comparison with others that apply ranking algorithms in graph-based procedures (i.e. Ppr_w2w, Sihna07, Mih05, Tsatsa07, Nav05). Among all the systems evaluated in both Table 8 and Table 9, Ppr+Freq procedure obtains the highest scores. Ppr+Freq is able to identify the node's centrality by applying a graph-based ranking algorithm, and it also considers several dimensions such as WN, WND, WNA, SUMO, SC, pSemcor and Freq in its LKB. Our procedure has been the result of a detailed study of some of the WSD procedures that are considered relevant by the scientific community (see Table 9).

Table 9 shows the strengths and weaknesses of the WSD approaches that inspired our work. If we focus on the strengths, the most relevant techniques are graph-based which apply centrality measures with the use of structural weighting assignments. We consider that graph-based approaches have a weakness if they do not include the use of the frequency of word-senses, or the use of different semantic resources to explore different point of views when analyzing the WSD process. With regard to the last point, when studying approaches such as those of Mc.Carthy04 [82] and RST+Freq [46] it is important to remark that their main contribution consists of obtaining the most frequent word-senses in concordance with the context in which they are being used. Both works obtain similar scores and are simultaneously considered to be relevant. The hypothesis defended by both must be taken into consideration by all WSD systems, owing to the fact that the use of the baseline MFS has proven to be effective in all SemEval campaigns (see Section 1). We therefore have decided to empower the initial weightings of the PageRank ranking algorithm with normalized word-sense frequency values, and more specifically to initialize the probability vector v of the Ppr proposal and use a LKB with multidimensional semantic resources and word-sense frequencies. This proposal (Ppr+Freq) overcame the weaknesses previously described to produce one of the best WSD systems in the state-of-the-art.

5. Overall discussion

The Ppr+Freq development was based on two basic pillars: the Multidimensional Semantic Analysis (using the ISR-WN resource, XWN1.7, XWN3.0 and pSemcor) and the insertion of word-sense frequency information. This has been done with the aim of devising a better centrality measure in graph structures. The results obtained by using Ppr+Freq in WSD have outperformed those of other approaches in this area, and this has been made possible by considering multidimensionality and word-sense frequency. The historical margin between the top supervised and the top unsupervised approaches has consequently been reduced from 9.5 to 4.45 perceptual points (see Table 10). This has thus allowed us to empower the efficacy of unsupervised procedures by avoiding the need to collect expensive annotated corpora to train WSD classifiers.

As shown in Table 5 in which SensEval-2 ranking is presented, our best Ppr+Freq procedure attained a perceptual point difference of five. Our proposal reaches top positions in comparison to unsupervised systems according to the rankings of both SensEval-2 (Table 5) and SensEval-3 (Table 7). Ppr+Freq has been created by incorporating the word-sense frequency information into the Ppr proposal, and that both baseline proposals separately obtain poorer results than those of Ppr+Freq (see Table 4 and 7). This indicates that complementing the semantic information provided by Ppr with probabilistic frequency data improves the results.

The graph-based algorithm presented in our work has been previously used by other researchers like [83], [84] and [85]. Agirre et al. [83] also considered shortest paths for creating the subgraphs by using MCR as LKB, but then applying standard PageRank for the final ranking. Based on MRC, this approach made use of the following links: English WordNet 1.6 synsets and semantic relationships; English WordNet 1.7 synsets and semantic relationships; English WordNet 2.0 semantic relationships (to be added to WN1.6); eXtended WordNet (gold, silver and normal); Selectional preference relations; and Coocurrence relations. The best result obtained by [83] was 56.20% of F1 for SensEval-3, which differs from our proposal by a large margin of 6 points. A very similar setting, using DFS (Depth First Search) instead of BFS for the subgraph creation, and applied WSD analysis using many ranking algorithms was presented by [84]. Regarding the subgraph creation each word-sense was associated with a gloss, i.e., a textual definition which explains its meaning. Moreover, the word-senses for each word were ranked according to their frequency of occurrence in the SemCor corpus. This proposal obtained 52.9% of F1 (Degree unsupervised system) and 60.7% of F1 (Degree semi-supervised system) for SensEval-3. Finally, DFS was also used by [85] but Personalized PageRank for the ranking step was applied. The results were 57.9% for SensEval-2 and 59.7% for SensEval-3.

The results presented in this paper for the SensEval-2 and SensEval-3 clearly outperform previous works by a large margin, 6 points and 3...7 points respectively. However, the presented procedure is very similar to those described above, since they are based on widely enriched semantic networks. Arguably, the reason for our good results in SensEval campaign is due to the use of word-sense frequencies for constructing and ranking the contextual sub-graph.

It is thus possible to state that our procedure, which incorporates several semantic and probabilistic analysis points, is able to outperform unsupervised WSD systems.

Table 10 presents the best scores of both supervised and unsupervised systems. In this table, an average

Table 8: Comparisons of relevant WSD approaches when evaluating test corpus from English All Words task of SensEval-2 (Recall (R))

System	Graph Based	R	Noun R	Verb R	Adj R	Adv R	LKB
Ppr+Freq	X	**0.641**	0.690	**0.440**	**0.682**	**0.774**	ISR-WN
RST+Freq [46]	-	0.609	0.610	-	-	-	ISR-WN
Ppr_w2w [32]	X	0.586	**0.704**	0.389	0.583	0.701	WN+XWN1.7 / WN+XWN3.0 / MCR1.6+ XWN1.6
Sihna07 [30]	X	0.564	0.656	0.323	0.614	0.602	WN
Mih05 [29]	X	0.542	0.575	0.365	0.567	0.709	WN
Tsatsa07 [31]	X	0.492	-	-	-	-	WN / WN+XWN
N-Cliques+RV [80]	X	0.433	0.489	0.359	0.239	0.646	ISR-WN
N-Cliques+RST [81]	X	0.426	0.490	0.353	0.231	0.639	ISR-WN
Mc.Carthy04 [82]	-	-	0.630	-	-	-	Corpus

Table 9: Comparisons of relevant WSD approaches when evaluating test corpus from English All Words task of SensEval-3 (Recall (R))

System	Graph Based	R	Noun R	Verb R	Adj R	Adv R	LKB
Ppr+Freq	X	**0.611**	**0.663**	**0.529**	**0.634**	**1.000**	ISR-WN
Nav05 [33]	X	0.604	-	-	-	-	WN
Ppr_w2w [32]	X	0.574	0.641	0.469	0.626	0.929	WN+XWN1.7 / WN+XWN3.0 / MCR1.6+ XWN1.6
Sihna07 [30]	X	0.524	0.605	0.406	0.541	1.000	WN
Mih05 [29]	X	0.522	-	-	-	-	WN
Nav07 [61]	X	-	0.619	0.361	0.628	-	WN

of 9.5 of perceptual point represents the range of difference between both system types. Bearing in mind that our proposal is an unsupervised procedure, although it includes word-sense frequency, it contributes toward reducing the actual margin by 4.45 perceptual points.

We consider Ppr+Freq as a knowledge-based procedure because word-sense frequency is currently viewed as knowledge. Great improvements were made in the semantic studies resulting from Ppr+Freq. New motivations emerged which were channelled towards encouraging unsupervised and multidimensional graph-based procedures to solve WSD problems. We believe that the state-of-the-art in WSD could be improved by considering the aforementioned pillars (word-sense frequency information and semantic resource integration) and searching for more efficient centrality procedures and knowledge.

In addition, we can see that our system for SemEval-2013 was able to outperform both supervised systems and also multilingual campaigns (SemEval-2013). Related to multilingual WSD, our procedure has demonstrated its efficiency in multilingual campaigns, for example, in the SemEval-2013[14] [75]. In this competition our system (considering as LKB: ISR-WN + XWN1.7 + XWN3.0) was evaluated with five languages:

[14]http://www.cs.york.ac.uk/semeval-2013/task12/index.php, last access 10/2016

Table 10: Top scores for supervised and unsupervised systems in SensEval-2 and SensEval-3 (Best Supervised System of the campaign (S), Best Unsupervised System of the campaign (U), Our Best Result (Ap))

Competitions	**Recall (S)**	**Recall (U)**	**S-U =**	**Recall Contribution (Ap)**	**S-Ap =**
SensEval-2 (2001)	69.00%	56.90%	12.10%	64.10%	4.90%
SensEval-3 (2004)	65.10%	58.20%	6.90%	61.10%	4.00%
Average			9.50		**4.45**
Competition	**Recall (S/WS)**	**Recall (U)**	**S(or WS)-U =**	**Recall Contribution (Ap)**	**S(or WS)-Ap =**
SemEval 2013	40.6%	64.5%	- 23.9	64.5%	- 23.9

English, French, Spanish, Italian and German. After evaluating the results, our system addressed by UMCC-DLSI team [47] attained the first position of three other competitors and was consistently able to outperform the MFS baseline (a notoriously hard-to-beat heuristic) in all languages except German.

In Table 11: U means unsupervised; S means supervised; WS: weak supervised.

As can be seen in Table 11 our procedure Ppr+Freq as part of the system UMCC-DLSI, reached the top position in SemEval-2013. However, recent unsupervised approaches of state-of-the-art such as Babelfy, and Nasarilexical have outperformed our approach but without achieving statistically significant improvements (x^2, $p < 0.05$). In addition, we would like to remark that our system UMCC-DLSI outperforms once again the MFS baseline frequency (F-1 of 0.578).

The systems of the state-of-the-art included in SemEval-2013 comparison are:

Unsupervised:

- UKB w2w (Ppr_w2w) [32], a state-of-the-art approach for knowledge-based WSD, based on Personalized PageRank [76] over Wordnet
- Babelfly [34] a state-of-the-art approach for knowledge-based WSD, based on Personalized PageRank [76] over Babelnet.

Supervised or weak supervised:

- GETALP [16] which uses an Ant Colony Optimization technique together with the classical measure of [87]
- IMS [25], a state-of-the-art supervised English WSD system which uses an SVM (Super Vector Machine) trained on word-sense-annotated corpora, such as SemCor [27] and DSO [28], among others. This system used the IMS model out-of-the-box with Most Frequent Sense (MFS) as backoff routine since the model obtained using the task trial data performed worse.
- Multi-Objective [15] which views WSD as a multi-objective optimization problem and uses BabelNet as reference knowledge base.
- NASARI+IMS [86], which is based on Nasari WSD framework [86] with the only difference being that in this system they back-off to IMS instead of MFS.

Finally, we considered MUFFIN [17] a WSD approach based on the NASARI vectors that, in contrast, used a WSD framework in which words in context were considered equally important.

6. Conclusions and Future Works

This research presents a graph-based and knowledge based WSD procedure. Discovering the correct word-senses (WSD) is an essential task to improve other NLP systems such as: Recommender Systems; Information Retrieval; Machine Translation; etc, as demonstrated in recent years. Our goal has been to develop a new procedure that is able to improve WSD results and thus assist other NLP tasks. We have used our own previously developed resource called ISR-WN

Table 11: Evaluation on SemEval-2013 using the test dataset (Precision (P), Recall (R))

SemEval-2013 system participation					
System	**Supervision**	**Lang**	**P**	**R**	**F-1**
UMCC-DLSI (our)	U	EN	**0.649**	**0.645**	**0.647**
GETALP-WN ([16])	WS	EN	0.406	0.406	0.406
MFS	-	EN	0.630	0.630	0.630
Systems of recent state-of-the-art					
Babelfy ([34])	U	EN	-	-	**65.9**
UKB w2w ([32])	U	EN	-	-	61.3
Nasarilexical ([86])	U	EN	-	-	**66.7**
IMS ([25])	S	EN	-	-	65.7
GETALP-BN ([16])	WS	EN	-	-	51.4
Nasarilexical + IMS ([86])	mix	EN	-	-	67.0
Muffin ([86])	-	EN	-	-	66.0
Multi-Objective ([15])	S	EN	-	-	72.8

to build a multidimensional network of semantic concepts in order to integrate it into our LKB with other resources: eXtended WordNet (disambiguated glosses of WN 3.0 and WN 1.7); collocations in SemCor and word-sense frequencies of WN 1.7. We have demonstrated that the combination of the frequency of word-sense usage and semantic information from context is useful to obtain better results in WSD.

Our procedure has been compared to others by evaluating the results of several international campaigns: SensEval and SemEval. After conducting several experiments we have demonstrated that our procedure outperforms the best systems in each competition. With regard to the results obtained using the information from SensEval-2, we conclude that our procedure, which uses a voting approach that takes into account word-sense frequencies and a combination of semantic resources to build the LKB, outperforms all the unsupervised systems. However, in the case of supervised systems, our procedure provides better scores with the exception of the SMUaw of [67]. Our procedure obtained a precision of 66.9% while that of SMUaw obtained 69%. SMUaw uses a large corpus of word-word relations using WN1.7, SemCor and a large additional set of word-sense tagged word-word pairs based on heuristics. Our procedure, however, does not need the creation of a corpus because it uses existing resources and is therefore less time consuming. In SensEval-3, our system is ranked in sixth position, being the best of all the unsupervised systems. It obtained a precision of 62% in comparison with the 65.1% of the best supervised system.

According to the results obtained in both competitions our procedure appears to be significantly better than the set of unsupervised systems and has demonstrated that it is closer to the results of supervised systems.

Furthermore, we provide an innovative contribution to current disambiguation systems which are only applied to one language. Our procedure, can also be adapted to perform WSD in different languages, obtaining promising results. In the SemEval-2013 competition our system was evaluated with five languages: English, French, Spanish, Italian and German. After the results had been evaluated, our system attained first position against three other competitors and was consistently able to outperform the MFS baseline (a notoriously hard-to-beat heuristic) in all languages except German.

During the WSD process several factors affected the performance of our procedure, such as: the use of different versions of WN affected the results, and the fact that it is important to initialize the graph during the word-sense frequency inventory. We therefore plan to study these factors in depth in order to improve the results, also considering the impact of using other word-sense frequency inventories.

Moreover, in order to demonstrate the usefulness of WSD in other NLP tasks we integrated our approach as part of a recommender system obtaining promising results [11]. Our procedure has also been integrated in the framework of the EU-funded project SAM (FP7-611312)[15]. Its goal was to build an advanced digital media delivery platform, combining second screen and content to promote communication and social interaction related to broadcasted program content. Finally, our procedure constitutes the central core of the REDES project (TIN2015-65136-C2-2-R)[16]. This project is based on the idea of representing real entities (persons, enterprises, products, etc) in digital word, in order to track and analyze them from different semantic point of views.

As future plan, we would like to experiment Ppr+Freq with different ISR-WN configurations (i.e. loading by separate SUMO, WND, SC and WNA). This is in order to identify the best configuration in the graph based procedure. We also plan to compare other graph based ranking algorithms, study different alternatives of speed so that our Ppr+Freq procedure may be suitable in real time systems.

Acknowledgements

This research work has been partially funded by the University of Alicante, Generalitat Valenciana, Spanish Government, Ministerio de Educación, Cultura y Deporte and ASAP - Ayudas Fundación BBVA a equipos de investigación científica 2016(FUNDACIONBBVA2-16PREMIO) through the projects, TIN2015-65100-R, TIN2015-65136-C2-2-R, PROMETEOII/2014/001, GRE16-01: “Plataforma inteligente para recuperación, análisis y representación de la información generada por usuarios en Internet” and PR16_SOC_0013.

References

[1] A. Montoyo, A. Suárez, G. Rigau, M. Palomar, Combining knowledge- and corpus-based word-sense-disambiguation methods, J. Artif. Int. Res. 23 (1) (2005) 299–330.

[2] Z. Wu, M. Palmer, Verbs semantics and lexical selection, in: Proceedings of the 32nd annual meeting on Association for Computational Linguistics, ACL ’94, Association for Computational Linguistics, Stroudsburg, PA, USA, 1994, pp. 133–138.

[3] C. Leacock, M. Chodorow, Combining local context and wordnet similarity for word sense identification, in: C. Fellfaum (Ed.), MIT Press, Cambridge, Massachusetts, 1998, pp. 265–283.

[4] P. D. Turney, Mining the web for synonyms: Pmi-ir versus lsa on toefl, in: Proceedings of the 12th European Conference on Machine Learning, EMCL ’01, Springer-Verlag, London, UK, UK, 2001, pp. 491–502.

[5] C. Fellbaum, WordNet: An Electronic Lexical Database, The MIT Press, 1998.

[6] Y. S. Chan, H. T. Ng, D. Chiang, Word sense disambiguation improves statistical machine translation, in: Annual Meeting-Association for Computational Linguistics, Vol. 45, Citeseer, 2007, p. 33.

[7] M. Carpuat, D. Wu, Improving statistical machine translation using word sense disambiguation., in: EMNLP-CoNLL, Vol. 7, 2007, pp. 61–72.

[8] L. Sevens, G. Jacobs, V. Vandeghinste, I. Schuurman, F. Van Eynde, Improving text-to-pictograph translation through word sense disambiguation, in: Fifth Joint Conference on Lexical and Computational Semantics (*SEM 2016), 2016, pp. 131–135.

[9] A. Weichselbraun, S. Gindl, A. Scharl, Enriching semantic knowledge bases for opinion mining in big data applications, Knowledge-Based Systems 69 (2014) 78 – 85.

[10] C. Hung, S.-J. Chen, Word sense disambiguation based sentiment lexicons for sentiment classification, Knowledge-Based Systems 110 (2016) 224 – 232.

[15]http://socialisingaroundmedia.com/, last access 10/2016
[16]https://gplsi.dlsi.ua.es/redes/, last access 10/2016

[11] Y. Gutiérrez, S. Vázquez, A. Montoyo, A semantic framework for textual data enrichment, Expert Syst. Appl. 57 (2016) 248–269.

[12] N. Ide, J. Véronis, Introduction to the Special Issue on Word Sense Disambiguation: The State of the Art, Computational Linguistics 24 (1) (1998) 1–40.

[13] J. Yu, W. Hong, C. Qiu, S. Li, D. Mei, A new approach of attribute partial order structure diagram for word sense disambiguation of english prepositions, Knowledge-Based Systems 95 (2016) 142 – 152.

[14] J. Yu, C. Li, W. Hong, S. Li, D. Mei, A new approach of rules extraction for word sense disambiguation by features of attributes, Applied Soft Computing 27 (2015) 411 – 419.

[15] D. Weissenborn, L. Hennig, F. Xu, H. Uszkoreit, Multi-objective optimization for the joint disambiguation of nouns and named entities, in: 53nd Annual Meeting of the Association for Computational Linguistics, July, ACL, 2015, pp. 596–605.

[16] D. Schwab, A. Tchechmedjiev, J. Goulian, M. Nasiruddin, G. Sérasset, H. Blanchon, Getalp system : Propagation of a lesk measure through an ant colony algorithm, in: Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), Association for Computational Linguistics, Atlanta, Georgia, USA, 2013, pp. 232–240.

[17] J. Camacho-Collados, M. T. Pilehvar, R. Navigli, A unified multilingual semantic representation of concepts, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Association for Computational Linguistics, Beijing, China, 2015, pp. 741–751.

[18] ACL (Ed.), Proceedings of the SENSEVAL-2: Second International Workshop on Evaluating Word Sense Disambiguation Systems, 2001.

[19] S. Cotton, P. Edmonds, A. Kilgarriff, M. Palmer, Senseval-2. second international workshop on evaluating word sense disambiguation systems, Association for Computational Linguistics, Toulouse, France, 2001.

[20] ACL (Ed.), Proceedings of the SENSEVAL-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text, 2004.

[21] B. Snyder, M. Palmer, The english all-words task, in: R. Mihalcea, P. Edmonds (Eds.), Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text, Association for Computational Linguistics, Barcelona, Spain, 2004, pp. 41–43.

[22] S. Pradhan, E. Loper, D. Dligach, M. Palmer, Semeval-2007 task-17: English lexical sample, srl and all words, in: Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007), Association for Computational Linguistics, Prague, Czech Republic, 2007, pp. 87–92.

[23] E. Agirre, O. López de Lacalle, C. Fellbaum, S.-K. Hsieh, M. Tesconi, M. Monachini, P. Vossen, R. Segers, Semeval-2010 task 17: All-words word sense disambiguation on a specific domain, in: Proceedings of the 5th International Workshop on Semantic Evaluation, Association for Computational Linguistics, Uppsala, Sweden, 2010, pp. 75–80.

[24] M. Palmer, C. Fellbaum, S. Cotton, L. Delfs, H. T. Dang, English tasks: All-words and verb lexical sample, in: Proceedings of SENSEVAL-2 Second International Workshop on Evaluating Word Sense Disambiguation Systems, Association for Computational Linguistics, Toulouse, France, 2001, pp. 21–24.

[25] Z. Zhong, H. T. Ng, It makes sense: A wide-coverage word sense disambiguation

system for free text, in: Proceedings of the ACL 2010 System Demonstrations, ACLDemos '10, Association for Computational Linguistics, Stroudsburg, PA, USA, 2010, pp. 78–83.

[26] M. T. Pilehvar, R. Navigli, A large-scale pseudoword-based evaluation framework for state-of-the-art word sense disambiguation, Comput. Linguist. 40 (4) (2014) 837–881.

[27] G. A. Miller, C. Leacock, R. Tengi, R. T. Bunker, A semantic concordance, in: Proceedings of the Workshop on Human Language Technology, HLT '93, Association for Computational Linguistics, Stroudsburg, PA, USA, 1993, pp. 303–308.

[28] H. T. Ng, H. B. Lee, Integrating multiple knowledge sources to disambiguate word sense: An exemplar-based approach, in: Proceedings of the 34th Annual Meeting on Association for Computational Linguistics, ACL '96, Association for Computational Linguistics, Stroudsburg, PA, USA, 1996, pp. 40–47.

[29] R. Mihalcea, Unsupervised large-vocabulary word sense disambiguation with graph-based algorithms for sequence data labeling, in: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, HLT '05, Association for Computational Linguistics, Stroudsburg, PA, USA, 2005, pp. 411–418.

[30] R. Sinha, R. Mihalcea, Unsupervised graph-based word sense disambiguation using measures of word semantic similarity, in: Proceedings of the International Conference on Semantic Computing, ICSC '07, IEEE Computer Society, Washington, DC, USA, 2007, pp. 363–369.

[31] G. Tsatsaronis, M. Vazirgiannis, I. Androutsopoulos, Word sense disambiguation with spreading activation networks generated from thesauri, in: Proceedings of the 20th international joint conference on Artifical intelligence, IJCAI'07, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2007, pp. 1725–1730.

[32] E. Agirre, A. Soroa, Personalizing pagerank for word sense disambiguation, in: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, EACL '09, Association for Computational Linguistics, Stroudsburg, PA, USA, 2009, pp. 33–41.

[33] R. Navigli, P. Velardi, Structural semantic interconnections: A knowledge-based approach to word sense disambiguation, IEEE Trans. Pattern Anal. Mach. Intell. 27 (7) (2005) 1075–1086.

[34] A. Moro, R. Navigli, Semeval-2015 task 13: Multilingual all-words sense disambiguation and entity linking, in: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Association for Computational Linguistics, Denver, Colorado, 2015, pp. 288–297.

[35] E. Laparra, G. Rigau, M. Cuadros, Exploring the integration of wordnet and framenet, in: Proceedings of the 5th Global WordNet Conference (GWC'10), Mumbai (India), 2010.

[36] B. Magnini, G. Cavaglia, Integrating subject field codes into wordnet, in: LREC, 2000.

[37] E. Pianta, L. Bentivogli, C. Girardi, MultiWordNet: developing an aligned multilingual database, in: Proceedings of the First International Conference on Global WordNet, 2002, pp. 21–25.

[38] B. J. Dorr, A. Marti, I. Castellon, Spanish eurowordnet and lcs-based interlingual mt, in: Proceedings of the MT Summit Workshop on Interlinguas in MT, San Diego, CA, Citeseer, 1997.

[39] J. Atserias, L. Villarejo, G. Rigau, E. Agirre, J. Carroll, B. Magnini, P. Vossen, The meaning multilingual central repository, in:

In Proceedings of the Second International WordNet Conference, 2004, pp. 80–210.

[40] I. Gurevych, J. Eckle-Kohler, S. Hartmann, M. Matuschek, C. M. Meyer, C. Wirth, Uby: A large-scale unified lexical-semantic resource based on lmf, in: Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, Association for Computational Linguistics, 2012, pp. 580–590.

[41] R. Navigli, S. P. Ponzetto, Babelnet: Building a very large multilingual semantic network, in: Proceedings of the 48th annual meeting of the association for computational linguistics, Association for Computational Linguistics, 2010, pp. 216–225.

[42] R. Izquierdo, A. Suárez, G. Rigau, A proposal of automatic selection of coarse-grained semantic classes for wsd, in: Procesamiento del Lenguaje Natural, 2007, pp. 189–196.

[43] A. Esuli, F. Sebastiani, Sentiwordnet: A publicly available lexical resource for opinion mining, in: In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC06, 2006, pp. 417–422.

[44] C. Strapparava, A. Valitutti, WordNet-Affect: An affective extension of WordNet, in: Proceedings of the 4th International Conference on Language Resources and Evaluation, ELRA, 2004, pp. 1083–1086.

[45] Y. Gutiérrez, A. Fernández, A. Montoyo, S. Vázquez, Umcc-dlsi: Integrative resource for disambiguation task, in: Proceedings of the 5th International Workshop on Semantic Evaluation, SemEval '10, Association for Computational Linguistics, Stroudsburg, PA, USA, 2010, pp. 427–432.

[46] Y. Gutiérrez, S. Vázquez, A. Montoyo, Improving wsd using isr-wn with relevant semantic trees and semcor senses frequency, in: Proceedings of the International Conference Recent Advances in Natural Language Processing 2011, RANLP 2011 Organising Committee, Hissar, Bulgaria, 2011.

[47] Y. Gutiérrez, Y. Castañeda, A. González, R. Estrada, D. D. Piug, J. I. Abreu, R. Pérez, A. Fernández Orquín, A. Montoyo, R. Muñoz, F. Camara, Umcc_dlsi: Reinforcing a ranking algorithm with sense frequencies and multidimensional semantic resources to solve multilingual word sense disambiguation, in: Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), Association for Computational Linguistics, Atlanta, Georgia, USA, 2013, pp. 241–249.

[48] Y. Gutiérrez, S. Vázquez, A. Montoyo, Integration of semantic resources based on wordnet, in: Procesamiento del Lenguaje Natural, 2010.

[49] Y. Gutiérrez, S. Vázquez, A. Montoyo, Enriching the integration of semantic resources based on wordnet, in: Procesamiento del Lenguaje Natural, 2011, pp. 249–257.

[50] D. Moldovan, V. Rus, Explaining Answers with Extended WordNet, in: Meeting of the Association for Computational Linguistics, 2001.

[51] R. Mihalcea, D. I. Moldovan, extended wordnet: progress report, in: In Proceedings of NAACL Workshop on WordNet and Other Lexical Resources, 2001, pp. 95–100.

[52] T. Petrolito, F. Bond, A survey of wordnet annotated corpora, in: Proceedings of the Seventh Global Wordnet Conference, Tartu, Estonia, 2014, pp. 236–245.

[53] G. A. Miller, Wordnet: A lexical database for english., Commun. ACM 38 (11) (1995) 39–41.

[54] A. Valitutti, C. Strapparava, O. Stock, Developing affective lexical resources., PsychNology Journal 2 (1) (2004) 61–83.

[55] A. Zouaq, M. Gagnon, B. Ozell, A sumo-based semantic analysis for knowledge extraction, in: Proceedings of the 4th Language & Technology Conference, 2009.

[56] E. W. Dijkstra, A note on two problems in connexion with graphs, Numerische Mathematik 1 (1959) 269–271.

[57] R. W. Floyd, Algorithm 97: Shortest path, Commun. ACM 5 (6) (1962) 345.

[58] S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine, in: Proceedings of the seventh international conference on World Wide Web 7, WWW7, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, 1998, pp. 107–117.

[59] T. H. Haveliwala, Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search, IEEE Trans. on Knowl. and Data Eng. 15 (4) (2003) 784–796.

[60] R. D. Luce, A. D. Perry, A method of matrix analysis of group structure, Psychometrika 14 (2) (1949) 95–116.

[61] R. Navigli, M. Lapata, Graph connectivity measures for unsupervised word sense disambiguation, in: Proceedings of the 20th international joint conference on Artifical intelligence, IJCAI'07, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2007, pp. 1683–1688.

[62] L. C. Freeman, Centrality in social networks conceptual clarification, Social Networks (1978) 215.

[63] U. Brandes, A faster algorithm for betweenness centrality, Journal of Mathematical Sociology 25 (2001) 163–177.

[64] R. Navigli, A structural approach to the automatic adjudication of word sense disagreements, Nat. Lang. Eng. 14 (4) (2008) 547–573.

[65] S. Reddy, A. Inumella, D. McCarthy, M. Stevenson, Iiith: Domain specific word sense disambiguation, in: Proceedings of the 5th International Workshop on Semantic Evaluation, SemEval '10, Association for Computational Linguistics, Stroudsburg, PA, USA, 2010, pp. 387–391.

[66] A. Soroa, E. Agirre, O. López de Lacalle, W. Bosma, P. Vossen, M. Monachini, J. Lo, S.-K. Hsieh, Kyoto: An integrated system for specific domain wsd, in: Proceedings of the 5th International Workshop on Semantic Evaluation, Association for Computational Linguistics, Uppsala, Sweden, 2010, pp. 417–420.

[67] R. F. Mihalcea, D. I. Moldovan, Pattern learning and active feature selection for word sense disambiguation, in: The Proceedings of the Second International Workshop on Evaluating Word Sense Disambiguation Systems, SENSEVAL '01, Association for Computational Linguistics, Stroudsburg, PA, USA, 2001, pp. 127–130.

[68] D. Fernández-Amorós, J. Gonzalo, F. Verdejo, The uned systems at senseval-2, in: The Proceedings of the Second International Workshop on Evaluating Word Sense Disambiguation Systems, SENSEVAL '01, Association for Computational Linguistics, Stroudsburg, PA, USA, 2001, pp. 75–78.

[69] J. Preiss, A detailed comparison of wsd systems: an analysis of the system answers for the senseval-2 english all words task, Nat. Lang. Eng. 12 (3) (2006) 209–228.

[70] B. Decadt, , B. Decadt, V. Hoste, W. Daelemans, A. V. D. Bosch, Gambl, genetic algorithm optimization of memory-based wsd, in: In Proceedings of ACL/SIGLEX Senseval-3, 2004, pp. 108–112.

[71] C. Strapparava, A. Gliozzo, C. Giuliano, Pattern abstraction and term similarity for word sense disambiguation: Irst at senseval-3, in:

R. Mihalcea, P. Edmonds (Eds.), Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text, Association for Computational Linguistics, Barcelona, Spain, 2004, pp. 229–234.

[72] S. Tratz, A. Sanfilippo, M. Gregory, A. Chappell, C. Posse, P. Whitney, Pnnl: A supervised maximum entropy approach to word sense disambiguation, in: Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007), Association for Computational Linguistics, Prague, Czech Republic, 2007, pp. 264–267.

[73] R. Ion, D. Tufiş, Racai: Meaning affinity models, in: Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007), Association for Computational Linguistics, Prague, Czech Republic, 2007, pp. 282–287.

[74] A. Kulkarni, M. Khapra, S. Sohoney, P. Bhattacharyya, Cfilt: Resource conscious approaches for all-words domain specific wsd, in: Proceedings of the 5th International Workshop on Semantic Evaluation, Association for Computational Linguistics, Uppsala, Sweden, 2010, pp. 421–426.

[75] R. Navigli, D. Vannella, Semeval-2013 task 11: Evaluating word sense induction & disambiguation within an end-user application, in: Proceedings of the 7^{th} International Workshop on Semantic Evaluation (SemEval 2013), in conjunction with the Second Joint Conference on Lexical and Computational Semantics (*SEM 2013), Atlanta, USA, 2013, pp. 193–201.

[76] T. H. Haveliwala, Topic-sensitive pagerank, in: Proceedings of the 11th International Conference on World Wide Web, WWW '02, ACM, New York, NY, USA, 2002, pp. 517–526.

[77] X. Carreras, I. Chao, L. Padró, M. Padró, Freeling: An open-source suite of language analyzers., in: LREC, 2004, pp. 239–242.

[78] E. Loper, S. Bird, Nltk: The natural language toolkit, in: Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics - Volume 1, ETMTNLP '02, Association for Computational Linguistics, Stroudsburg, PA, USA, 2002, pp. 63–70.

[79] B. Magnini, C. Strapparava, Experiments in Word Domain Disambiguation for Parallel Texts, in: Proceedings of the ACL Workshop on Word Senses and Multilinguality, Hong Kong, China, 2000.

[80] Y. Gutiérrez, S. Vázquez, A. Montoyo, Word sense disambiguation: A graph-based approach using n-cliques partitioning technique, in: NLDB, 2011, pp. 112–124.

[81] Y. Gutiérrez, S. Vázquez, A. Montoyo, A graph-based approach to wsd using relevant semantic trees and n-cliques model, in: CICLing (1), 2012, pp. 225–237.

[82] D. McCarthy, R. Koeling, J. Weeds, J. A. Carroll, Finding predominant word senses in untagged text., in: ACL, 2004, pp. 279–286.

[83] E. Agirre, A. Soroa, Using the multilingual central repository for graph-based word sense disambiguation., in: LREC, 2008.

[84] R. Navigli, M. Lapata, An experimental study of graph connectivity for unsupervised word sense disambiguation, Pattern Analysis and Machine Intelligence, IEEE Transactions on 32 (4) (2010) 678–692.

[85] E. Agirre, O. L. de Lacalle, A. Soroa, Random walks for knowledge-based word sense disambiguation, Computational Linguistics 40 (1) (2014) 57–84.

[86] J. Camacho-Collados, M. T. Pilehvar, R. Navigli, Nasari: Integrating explicit knowledge and corpus statistics for a multilingual representation of concepts and

entities, Artificial Intelligence 240 (2016) 36 – 64.

[87] M. Lesk, Automated sense disambiguation using machine-readable dictionaries: How to tell a pine cone from an ice cream cone, in: Proceedings of the 1986 SIGDOC Conference, Association for Computing Machinery, Toronto, Canada, 1986, pp. 24–26.

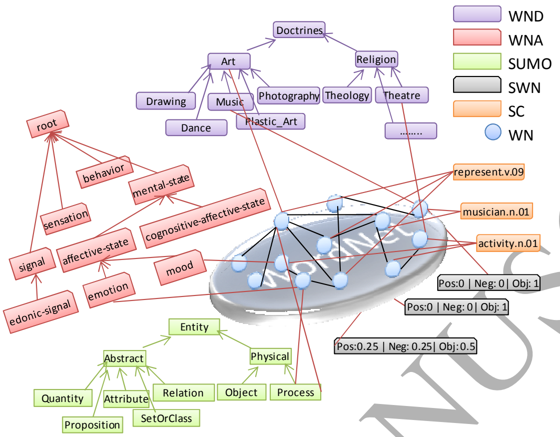


Figure 1: Lexical Knowledge Base of ISR-WN

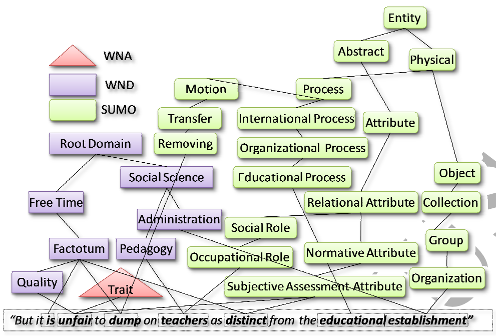


Figure 2: Extraction of Semantic Characteristics

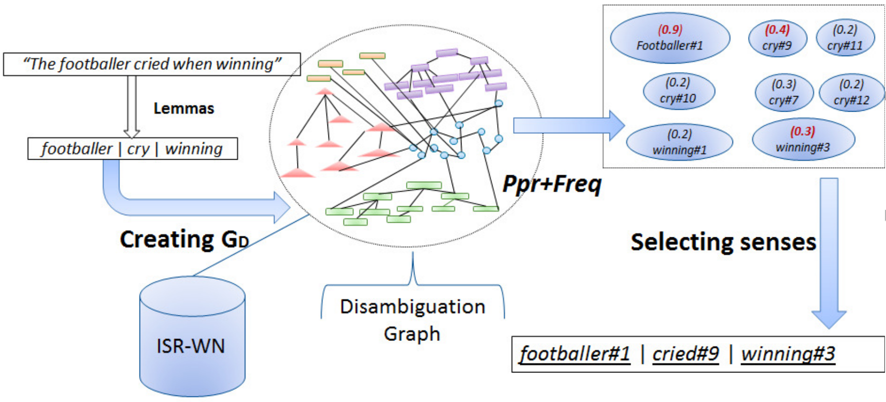


Figure 3: General Workflow of Ppr+Freq

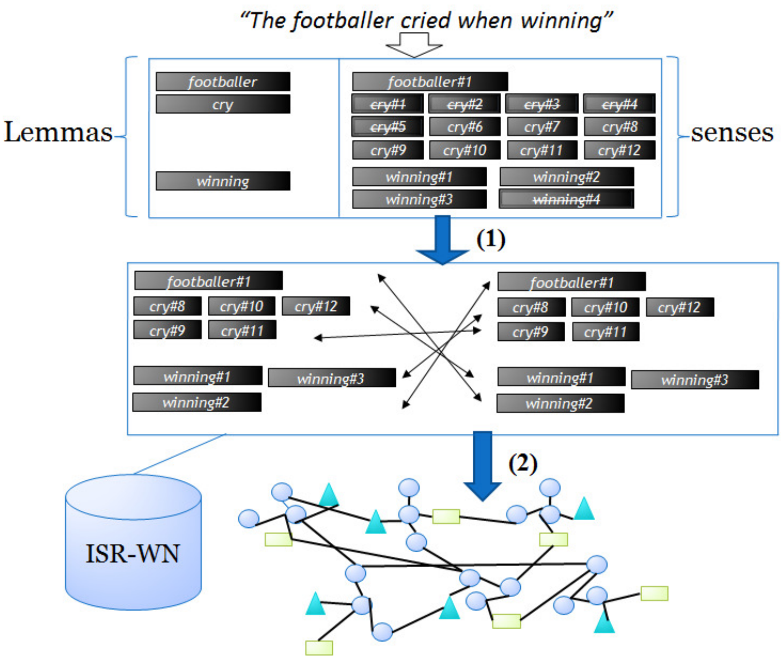
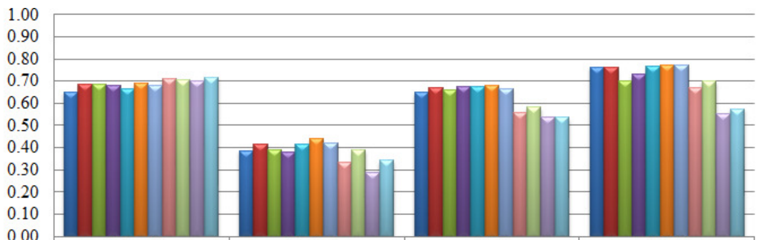
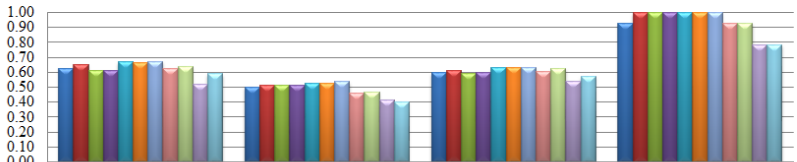


Figure 4: Initial Graph Creation



	Noun R	Verb R	Adj R	Adv R
MFS 1	0.652	0.386	0.649	0.764
Exp 2	0.688	0.413	0.669	0.761
Exp 3	0.686	0.392	0.658	0.701
Exp 4	0.679	0.381	0.675	0.731
Exp 5	0.665	0.417	0.677	0.767
Exp 6	0.690	0.440	0.682	0.774
Exp 7	0.679	0.422	0.667	0.774
8-Ppr	0.711	0.334	0.559	0.671
9-Ppr_w2w	0.704	0.389	0.583	0.701
10-Ppr	0.700	0.286	0.539	0.551
11-Ppr_w2w	0.719	0.344	0.538	0.575

Figure 5: Part of Speech disambiguation with SensEval-2 test corpus



	Noun R	Verb R	Adj R	Adv R
MFS 1	0.627	0.501	0.598	0.929
Exp 2	0.654	0.515	0.609	1.000
Exp 3	0.614	0.515	0.589	1.000
Exp 4	0.614	0.511	0.601	1.000
Exp 5	0.668	0.523	0.631	1.000
Exp 6	0.663	0.529	0.634	1.000
Exp 7	0.671	0.537	0.634	1.000
8-Ppr	0.626	0.46	0.608	0.929
9-Ppr_w2w	0.641	0.469	0.626	0.929
10-Ppr	0.522	0.415	0.542	0.786
11-Ppr_w2w	0.590	0.402	0.572	0.786

Figure 6: Part Of Speech disambiguation with SensEval-3 test corpus

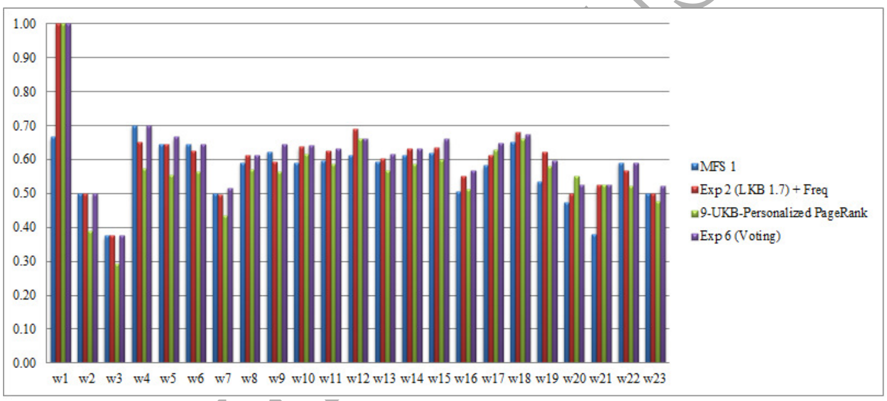


Figure 7: Recalls for word windows when analysing SensEval-2 test corpus