

# UCF-WS: Domain Word Sense Disambiguation using Web Selectors

Hansen A. Schwartz and Fernando Gomez

School of Electrical Engineering and Computer Science

University of Central Florida

Orlando, FL 32816

{hschwartz, gomez}@cs.ucf.edu

## Abstract

This paper studies the application of the Web Selectors word sense disambiguation system on a specific domain. The system was primarily applied without any domain tuning, but the incorporation of domain predominant sense information was explored. Results indicated that the system performs relatively the same with domain predominant sense information as without, scoring well above a random baseline, but still 5 percentage points below results of using the first sense.

## 1 Introduction

We explore the use of the Web Selectors word sense disambiguation system for disambiguating nouns and verbs of a domain text. Our method to acquire selectors from the Web for WSD was first described in (Schwartz and Gomez, 2008). The system is extended for the all-words domain task by including part of speech tags from the Stanford Parser (Klein and Manning, 2003). Additionally, a domain adaptation technique of using domain predominant senses (Koeling et al., 2005) is explored, but our primary goal is concerned with evaluating the performance of the existing Web Selectors system on domain text.

In previous studies, the Web Selectors system was applied to text of a general domain. However, the system was not directly tuned for the general domain. The system may perform just as strong for domain WSD since the selectors, which are the core of disambiguation, can come from any domain present on the Web. In this paper, we study the application of the Web Selectors WSD algorithm to an all-words task on a specific domain, the SemEval 2010: Task 17 (Agirre et al., 2010).

## 2 Web Selectors

Selectors are words which take the place of a given target word within its local context (Lin, 1997). In the case of acquiring selectors from the Web, we search with the text of local context (Schwartz and Gomez, 2008). For example, if one was searching for selectors of ‘channel’ in the sentence, “The navigation channel undergoes major shifts from north to south banks”, then a search query would be:

*The navigation \* undergoes major shifts from north to south banks .*

where \* represents a wildcard to match every selector. The query is shortened to produce more results until at least 300 selectors are acquired or the query is less than 6 words. The process of acquiring selectors repeats for every content word of the sentence. Example selectors that might be returned for ‘channel’ include ‘route’, ‘pathway’, and ‘passage’.

Selectors serve for the system to essentially learn the areas or concepts of WordNet that the sense of a word should be similar or related. The target noun or verb is disambiguated by comparing its senses with all selectors for itself (*target selectors*), as well as with *context selectors* for other nouns, verbs, adjective, adverbs, proper nouns, and pronouns in the sentence. Figure 1 shows the overall process undertaken to rank the senses of an ambiguous word. A *similarity* measure is used when comparing with *target selectors* and a *relatedness* measure is used when comparing with *context selectors*. Referring to our previous example, the senses of ‘channel’ are compared to its own (*target*) selectors via *similarity* measures, while relatedness measures are used for the *context selectors*: noun selectors of ‘navigation’, ‘shifts’, ‘north’, ‘south’, and ‘banks’; the verb selectors of

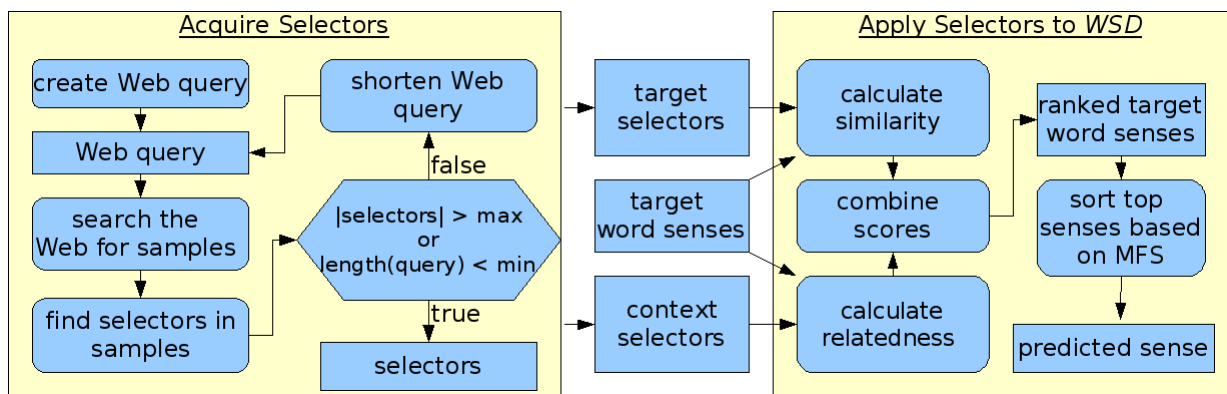


Figure 1: The overall process undertaken to disambiguate a word using Web selectors.

‘undergoes’; plus the adjective selectors of ‘major’. Adverbs, proper nouns, and pronouns are not present in the sentence, and so no selectors from those parts of speech are considered.

For this study, we implemented the Web Selectors system that was presented in (Schwartz and Gomez, 2009). This generalized version of the system may annotate verbs in addition to nouns, and it includes the previously unused context selectors of adverbs. We used the path-based similarity measure of (Jiang and Conrath, 1997) for *target selectors*, and the gloss-based relatedness measure of (Banerjee and Pedersen, 2002) for *context selectors*.

The incorporation of a part of speech tagger was a necessary addition to the existing system. Previous evaluations of Web Selectors relied on the testing corpus to provide part of speech (POS) tags for content words. In the case of SemEval-2010 Task 17, words were only marked as targets, but their POS was not included. We used the POS tags from the Stanford Parser (Klein and Manning, 2003). We chose this system since the dependency relationship output was also useful for our domain adaptation (described in section 2.1). A modification was made to the POS tags given the knowledge that the testing corpus only included nouns and verbs as targets. Any target that was not initially tagged as a noun or verb was reassigned as a noun, if the word existed as a noun in WordNet (Miller et al., 1993), or as a verb if not.

## 2.1 Domain Adaptation

Overall, the Web Selectors system is not explicitly tuned to the general domain. Selectors themselves can be from any domain. However, sense tagged data may be used indirectly within the system.

First, the similarity and relatedness measures used in the system may rely on SemCor data (Miller et al., 1994). Also, the system breaks ties by choosing the most frequent sense according to WordNet frequency data (based on SemCor). These two aspects of the system can be seen as tuned to the general domain, and thus, they are likely aspects of the system for adaptation to a specific domain.

For this work, we focused on domain-adapting the tie breaker aspect of the Web Selectors system. The system defines a tie occurring when multiple sense choices are scored within 5% of the top sense choice. In order to break the tie, the system normally chooses the most frequent sense among the tied senses. However, it would be ideal to break the tie by choosing the most prevalent sense over the testing domain. Because sense tagged domain data is not typically available, Koeling et al. (2005) presented the idea of estimating the most frequent sense of a domain by calculating sense prevalence scores from unannotated domain text.

Several steps are taken to calculate the prevalence scores. First, a dependency database is created, listing the frequencies that each dependency relationship appears. In our case, we used the Stanford Parser (Klein and Manning, 2003) on the background data provided by the task organizers. From the dependency database, a thesaurus is created based on the method of (Lin, 1998). In our approach, we considered the following relationships from the dependency database:

**subject** (*agent, csubj, subjpass, nsubj, nsubjpass, xsubj*)

**direct object** (*dobj*)

**indirect object** (*iobj*)

**adjective modifier** (*amod*)

**noun modifier** (*nn*)

**prepositional modifier** (any preposition, excluding *prep\_of* and *prep\_for*)

(typed dependency names listed in parenthesis)

Finally, a prevalence score is calculated for each sense of a noun or verb by finding the similarity between it and the top 50 most similar words according to the automatically created thesaurus. As Koeling et al. did, we use the similarity measure of (Jiang and Conrath, 1997).

### 3 Results and Discussion

The results of our system are given in Table 1. The first set of results (**WS**) was a standard run of the system without any domain adaptation, while the second set (**WS<sub>dom</sub>**) was from a run including the domain prevalence scores in order to break ties. The results show our domain adaptation technique did not lead to improved results. Overall, **WS** results came in ranked thirteenth among twenty-nine participating system results.

We found that using the prevalence scores alone to pick a sense (i.e. the ‘predominant sense’) resulted in an F score of 0.514 (**PS** in Table 1). Koeling et al. (2005) found the predominant sense to perform significantly better than the first sense baseline (*Isense*: equivalent to most frequent sense for the English WordNet) on specific domains (32% error reduction on a finance domain, and 62% error reduction on a sports domain). Interestingly, there was no significant error reduction over the *Isense* for this task, implying either that the domain was more difficult to adapt to or that our implementation of the predominant sense algorithm was not as strong as that use by Koeling et al. In any case, this lack of significant error reduction over the *Isense* may explain why our **WS<sub>dom</sub>** results were not stronger than the **WS** results. In **WS<sub>dom</sub>**, prevalence scores were used instead of *Isense* to break ties.

We computed a few figures to gain more insights on the system’s handling of domain data. Noun precision was 0.446 while verb precision was 0.449. It was unexpected for verb disambiguation results to be as strong as nouns because a previous study using Web Selectors found noun sense disambiguation clearly stronger than verb sense disambiguation on a coarse-grained corpus

	<b>P</b>	<b>R</b>	<b>F</b>	<b>P<sub>n</sub></b>	<b>P<sub>v</sub></b>
<i>rand</i>	0.23	0.23	0.23		
<i>Isense</i>	0.505	0.505	0.505		
<b>WS</b>	0.447	0.441	0.444	.446	.449
<b>WS<sub>dom</sub></b>	0.440	0.434	0.437	.441	.438
<b>PS</b>	0.514	0.514	0.514	.53	.44

Table 1: (**P**)recision, (**R**)ecall, and (**F**)-score of various runs of the system on the Task 17 data. **P<sub>n</sub>** and **P<sub>v</sub>** correspond to precision results broken down by nouns and verbs.

	<b>P<sub>en1</sub></b>	<b>P<sub>en2</sub></b>	<b>P<sub>en3</sub></b>
<b>WS</b>	0.377	0.420	0.558
<b>WS<sub>dom</sub></b>	0.384	0.415	0.531

Table 2: Precision scores based on the three documents of the English testing corpora (‘en1’, ‘en2’, and ‘en3’).

(Schwartz and Gomez, 2009). Ideally, our results for noun disambiguation would have been stronger than the the *Isense* and *PS* results. In order to determine the effect of the POS tagger (parser in this case) on the error, we determined 1.6% of the error was due to the wrong POS tag at (0.9% of all instances). Lastly, Table 2 shows the precision scores for each of the three documents from which the English testing corpus was created. Without understanding the differences between the testing documents it is difficult to explain why the precision varies, but the figures may be useful for comparisons by others.

Several aspects of the test data were unexpected for our system. Some proper nouns were considered as target words. Our system was not originally intended to annotate proper nouns, but we were able to adjust it to treat them simply as nouns. To be sure this treatment was appropriate, we also submitted results where proper nouns were excluded, and got a precision of 0.437 and recall of 0.392. One would expect the precision to increase at the expense of recall if the proper nouns were more problematic for the system than other instances. This was not the case, and we conclude our handling of proper nouns was appropriate.

Unfortunately, another unexpected aspect of the data was not handled correctly by our system. Our system only considered senses from one form of the target word according to WordNet, while the key included multiple forms of a word. For example, the key indicated *low\_tide-1* was the answer to

an instance where our system had only considered senses of ‘tide’. We determined that for 10.2% of the instances that were incorrect in our WS results we did not even consider the correct sense as a possible prediction due to using an inventory from only one form of the word. Since this issue mostly applied to nouns it may explain the observation that the noun disambiguation performance was not better than the verb disambiguation performance as was expected.

#### 4 Conclusion

In this paper we examined the application of the Web Selectors WSD system to the SemEval-2010 Task 17: All-words WSD on a Specific Domain. A primary goal was to apply the pre-existing system with minimal changes. To do this we incorporated automatic part of speech tags, which we found only had a small impact on the error (incorrectly tagged 0.9% of all target instances). Overall, the results showed the system to perform below the *Isense* baseline for both nouns and verbs. This is a lower relative performance than past studies which found the disambiguation performance above the *Isense* for nouns. One reason for the lower noun performance is that for 10.2 % of our errors, the system did not consider the correct sense choice as a possibility. Future versions of the system will need to expand the sense inventory to include other forms of a word (example: ‘low\_tide’ when disambiguating ‘tide’).

Toward domain adaptation, we ran an experiment in which one aspect of our system was tuned to the domain by using domain prevalence scores (or ‘predominant senses’). We found no improvement from using this adaptation technique, but we also discovered that results entirely based on predictions of the domain predominant senses were only minimally superior to *Isense* (F-score of 0.514 versus 0.505 for *Isense*). Thus, future studies will examine better implementation of the predominant sense algorithm, as well as explore other complimentary techniques for domain adaptation: customizing similarity measures for the domain, or restricting areas of WordNet as sense choices based on the domain.

#### Acknowledgement

This research was supported by the NASA Engineering and Safety Center under Grant/Cooperative Agreement NNX08AJ98A.

#### References

- Eneko Agirre, Oier Lopez de Lacalle, Christiane Fellbaum, Shu kai Hsieh, Maurizio Tesconi, Monica Monachini, Piek Vossen, and Roxanne Segers. 2010. Semeval-2010 task 17: All-words word sense disambiguation on a specific domain. In *Proceedings of SemEval-2010*. Association for Computational Linguistics.
- Satanjeev Banerjee and Ted Pedersen. 2002. An adapted lesk algorithm for word sense disambiguation using wordnet. In *Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics*, Mexico City, Mexico.
- Jay J. Jiang and David W. Conrath. 1997. Semantic similarity on corpus statistics and lexical taxonomy. In *Proceedings of ROCLING X*, Taiwan.
- Dan Klein and Christopher D. Manning. 2003. Fast exact inference with a factored model for natural language parsing. In *In Advances in Neural Information Processing Systems 15*, pages 3–10.
- Rob Koeling, Diana McCarthy, and John Carroll. 2005. Domain-specific sense distributions and predominant sense acquisition. In *Proceedings of the conference on Human Language Technology and Experimental Methods in NLP*, pages 419–426, Morristown, NJ, USA.
- Dekang Lin. 1997. Using syntactic dependency as local context to resolve word sense ambiguity. In *Proceedings of the 35th annual meeting on Association for Computational Linguistics*, pages 64–71.
- Dekang Lin. 1998. Automatic retrieval and clustering of similar words. In *Proceedings of COLING-ACL 98*, pages 768–774, Montreal, Canada. Morgan Kaufmann.
- George Miller, R. Beckwith, Christiane Fellbaum, D. Gross, and K. Miller. 1993. Five papers on wordnet. Technical report, Princeton University.
- George A. Miller, Martin Chodorow, Shari L, Claudia Leacock, and Robert G. Thomas. 1994. Using a semantic concordance for sense identification. In *In Proc. of ARPA Human Language Technology Workshop*.
- Hansen A. Schwartz and Fernando Gomez. 2008. Acquiring knowledge from the web to be used as selectors for noun sense disambiguation. In *CoNLL 2008: Proceedings of the Twelfth Conference on Computational Natural Language Learning*, pages 105–112, Manchester, England, August.
- Hansen A. Schwartz and Fernando Gomez. 2009. Using web selectors for the disambiguation of all words. In *Proceedings of the NAACL-2009 Workshop on Semantic Evaluations: Recent Achievements and Future Directions*, pages 28–36, Boulder, Colorado, June.