# Sense and Reference Disambiguation in Wikipedia

*Hui SHEN*[1] *Razvan BUNESCU*[1] *Rada MIHALCEA*[2]
(1) School of Electrical Engineering and Computer Science, Ohio University, Athens, OH
(2) Department of Computer Science, University of North Texas, Denton, TX
`hs138609@ohio.edu, bunescu@ohio.edu, rada@cs.unt.edu`

ABSTRACT
Wikipedia articles are annotated by volunteer contributors with numerous links that connect words and phrases to relevant titles in Wikipedia. In this paper, we identify inconsistencies in the user annotation of links and show that they can have a substantial impact on the performance of word sense disambiguation systems that are trained on Wikipedia links. We describe two major types of link annotations – *sense* and *reference* – that are frequently used without being explicitly distinguished in Wikipedia, and present an approach to training sense and reference disambiguation systems in the presence of such annotation inconsistencies. Experimental results demonstrate that accounting for annotation ambiguity in Wikipedia links leads to significant improvements in disambiguation accuracy.

KEYWORDS: word sense disambiguation, Wikipedia.

TITLE AND ABSTRACT IN ROMANIAN

## Dezambiguizare de Sensuri si Referinte in Wikipedia

Articolele din Wikipedia sunt adnotate de editori voluntari cu numeroase link-uri ce conecteaza fraze din articol cu titluri relevante in Wikipedia. In acest articol descriem inconsecvente in adnotarile editorilor si aratam ca ele pot avea un impact substantial asupra performantei sistemelor de dezambiguizare care sunt antrenate cu link-uri din Wikipedia. Descriem doua tipuri majore de adnotari – *sensuri* si *referinte* – care sunt folosite frecvent fara a fi diferentiate explicit in Wikipedia. Prezentam modele de invatare automata pentru dezambiguizare de sensuri si referinte care pot fi antrenate in prezenta acestor ambiguitati de adnotare. Evaluarea experimentala a acestor modele confirma o imbunatatire semnificativa a performantei de dezambiguizare.

KEYWORDS: dezambiguizare de sensuri, Wikipedia.

*Proceedings of COLING 2012: Posters*, pages 1111–1120,
COLING 2012, Mumbai, December 2012.

## 1 Introduction and Motivation

The vast amount of world knowledge available in Wikipedia has been shown to benefit many types of text processing tasks, such as coreference resolution (Ponzetto and Strube, 2006; Haghighi and Klein, 2009; Bryl et al., 2010; Rahman and Ng, 2011), information retrieval (Milne, 2007; Li et al., 2007; Potthast et al., 2008; Cimiano et al., 2009), or question answering (Ahn et al., 2004; Kaisser, 2008; Ferrucci et al., 2010). In particular, the user contributed link structure of Wikipedia has been shown to provide useful supervision for training named entity disambiguation (Bunescu and Pasca, 2006; Cucerzan, 2007) and word sense disambiguation (Mihalcea, 2007; Mihalcea and Csomai, 2007) systems. Articles in Wikipedia often contain mentions of concepts or entities that already have a corresponding article. When contributing authors mention an existing Wikipedia entity inside an article, they are required to link at least its first mention to the corresponding article, by using *links* or *piped links*. Consider, for example, the following Wikipedia annotations from the article about Palermo: *"Palermo is a city in [[Southern Italy]], the [[capital city|capital]] of the [[autonomous area|autonomous region]] of [[Sicily]]"*. The bracketed strings *[[Southern Italy]]* and *[[Sicily]]* identify the title of the Wikipedia articles that describe the corresponding named entities. The same strings are also used in the displayed HTML version of the sentence. If the author wants a different string displayed (e.g., *"autonomous region"* instead of the title string *"autonomous area"*), then the alternative string is included in a piped link, after the title string. Using these rules for expanding simple or piped links, the HTML string that is displayed for the aforementioned example is: *"Palermo is a city in Southern Italy, the capital of the autonomous region of Sicily"*.

Since many words and names mentioned in Wikipedia articles are inherently ambiguous, their corresponding links can be seen as a useful source of supervision for training named entity and word sense disambiguation systems. For example, Wikipedia contains articles that describe possible senses of the word "capital", such as CAPITAL CITY, CAPITAL (ECONOMICS), FINANCIAL CAPITAL, or HUMAN CAPITAL, to name only a few. When disambiguating a word or a phrase in Wikipedia, a contributor uses the context to determine the appropriate Wikipedia title to include in the link. In the example above, the editor of the article determined that the word "capital" was mentioned in that context with the political center meaning, consequently it was mapped to the article CAPITAL CITY through a piped link.

In order to use Wikipedia links for training a WSD system for a given word, one needs first to define a sense repository that specifies the possible meanings for that word, and then use the Wikipedia links to create training examples for each sense in the repository. Taking the word "atmosphere" as an example, the process might be implemented using the following sequence of steps:

1. Collect all Wikipedia titles that are linked from the anchor word "atmosphere". This results in a wide array of titles, ranging from the general ATMOSPHERE and its instantiations ATMOSPHERE OF EARTH or ATMOSPHERE OF MARS, to titles as diverse as ATMOSPHERE (UNIT), MOOD (PSYCHOLOGY), or ATMOSPHERE (MUSIC GROUP).

2. Create a repository of senses from all titles that have sufficient support in Wikipedia i.e., titles that are referenced at least a predefined minimum number of times using the ambiguous word as anchor. The most frequent titles for the anchor word "atmosphere" are thus assembled into a repository $\mathscr{R}$ = {ATMOSPHERE, ATMOSPHERE OF EARTH, ATMOSPHERE OF MARS, ATMOSPHERE OF VENUS, STELLAR ATMOSPHERE, ATMOSPHERE (UNIT), ATMOSPHERE (MUSIC GROUP)}.

The Beagle 2 lander objectives were to characterize the physical properties of the [[atmosphere]] and surface layers
*Sense* = **Atmosphere**; *Reference* = Atmosphere of Mars; *Label* = A → A(S) → AM

---

The Orbiter has been successfully performing scientific measurements and study of the interaction of the [[Atmosphere of Mars|atmosphere]] with
*Sense* = Atmosphere; *Reference* = **Atmosphere of Mars**; *Label* = A → A(S) → AM

---

In global climate models, the state and properties of the [[atmosphere]] are specified or computed at a number of discrete locations
*Sense* = **Atmosphere**; *Reference* = Atmosphere of Earth; *Label* = A → A(S) → AE

---

The principal natural phenomena that contribute acid-producing gases to the [[Atmosphere of Earth|atmosphere]] are emissions from volcanoes
*Sense* = Atmosphere; *Reference* = **Atmosphere of Earth**; *Label* = A → A(S) → AE

---

An aerogravity assist, or AGA, is a spacecraft maneuver designed to change velocity when arriving at a body with an [[atmosphere]]
*Sense* = **Atmosphere**; *Reference* = Atmosphere ▹ *generic*; *Label* = A → A(O)

---

Assuming the planet's [[atmosphere]] is close to chemical equilibrium, it is predicted that 55 Cancri d is covered in a layer of water clouds
*Sense* = **Atmosphere**; *Reference* = Atmosphere of Cancri ▹ *missing*; A → A(O)

Figure 1: A(S) = Atmosphere (S), A(O) = Atmosphere (O), A = Atmosphere, AE = Atmosphere of Earth, AM = Atmosphere of Mars.

3. Use the links extracted for each sense in the repository as labeled examples for that sense and train a WSD model to distinguish between alternative senses of the ambiguous word "atmosphere" based on features extracted from the word context.

This Wikipedia-based approach to creating training data for word sense disambiguation has a major shortcoming. Many of the training examples extracted for the title Atmosphere could very well belong to more specific titles such as Atmosphere of Earth or Atmosphere of Mars. Whenever the word "atmosphere" is used in a context with the sense of "a layer of gases that may surround a material body of sufficient mass, and that is held in place by the gravity of the body," the contributor has the option of adding a link either to the title Atmosphere that describes this sense of the word, or to the title of an article that describes the atmosphere of the actual celestial body that is referred in that particular context, as shown in the first 4 examples in Figure 1. We will call the more general link a *sense* annotation, and the more specific link a *reference* annotation. Correspondingly, Atmosphere will be a sense for the word "atmosphere", whereas Atmosphere of Earth, Atmosphere of Mars, and Atmosphere of Venus will all be references associated with this sense. As shown in bold in Figure 1, different occurrences of the same word may be tagged with a sense or a reference link, an ambiguity that is pervasive in Wikipedia for words like "atmosphere" that have senses with multiple, popular references. There does not seem to be a clear, general rule underlying the decision to tag a word or a phrase with a sense or a reference link in Wikipedia. We hypothesize that, in some cases, editors may be unaware that an article exists in Wikipedia for the actual reference of a word or for a more specific sense of the word, and therefore they end up using a link to an article describing the general sense of the word. There is also the possibility that more specific articles are introduced only in newer versions of Wikipedia, and thus earlier annotations were not aware of these recent articles. Furthermore, since annotating words with the most specific sense or reference available in Wikipedia may require substantial cognitive effort, editors may often choose to link to a general sense of the word, a choice that is still correct, yet less informative than the more specific sense or reference.

| **atmosphere** | Size |
|---|---|
| ATMOSPHERE | 932 |
| *Atmosphere (S)* | *559* |
| *Atmosphere of Earth* | *518* |
| *Atmosphere of Mars* | *19* |
| *Atmosphere of Venus* | *9* |
| *Stellar Atmosphere* | *13* |
| *Atmosphere (O)* | *373* |
| ATMOSPHERE OF EARTH | 345 |
| ATMOSPHERE OF MARS | 37 |
| ATMOSPHERE OF VENUS | 26 |
| STELLAR ATMOSPHERE | 29 |
| ATMOSPHERE (UNIT) | 96 |
| ATMOSPHERE (MUSIC GROUP) | 104 |

| **game** | Size |
|---|---|
| GAME | 819 |
| *Game (S)* | *99* |
| *Video game* | *55* |
| *PC game* | *44* |
| *Game (O)* | *720* |
| VIDEO GAME | 312 |
| PC GAME | 24 |
| GAME (FOOD) | 232 |
| GAME (RAPPER) | 154 |

Table 1: Wikipedia annotations (normal) and manual annotations (*italics*).

To estimate the magnitude of the sense vs. reference annotation ambiguity, we extracted all link annotations for the words “atmosphere” and “game” that were labeled with the sense links ATMOSPHERE and GAME, respectively. We then used the context to manually determine for each sense link annotation the corresponding more specific title, when such a title exists in Wikipedia. The statistics in Table 1 show a significant overlap between the sense and reference categories for words like "atmosphere" that have senses with multiple, popular references. For example, out of the 932 ATMOSPHERE links that were extracted in total, 518 were actually about the ATMOSPHERE OF EARTH, but the user linked them to the more general sense category ATMOSPHERE. On the other hand, there are 345 links to ATMOSPHERE OF EARTH that were explicitly made by the user. The table also shows that sometimes the ambiguous word is linked to a more specific sense, such as STELLAR ATMOSPHERE. We manually assigned *other* links (O) whenever the word is used with a generic sense, or when the reference is not available in the repository of Wikipedia titles collected for that word because either the reference title does not exist in Wikipedia or the reference title exists, but it does not have sufficient support – at least 20 linked anchors – in Wikipedia. We grouped all references and more specific links for any given sense into a special category suffixed with (S), to distinguish them from the other links (generic use, or missing reference) that were grouped into the category suffixed with (O).

A supervised learning algorithm that uses the extracted links for training a WSD classification model to distinguish between categories in the sense repository assumes implicitly that the categories, and hence their training examples, are mutually disjoint. This assumption is clearly violated for words like “atmosphere,” consequently the learned model will have a poor performance on distinguishing between the overlapping categories. Alternatively, we can say that sense categories like ATMOSPHERE are ill defined, since their supporting dataset contains examples that could also belong to more specific, reference categories such as ATMOSPHERE OF EARTH or ATMOSPHERE OF MARS. We see two possible solutions to the problem of inconsistent link annotations:

1. Group related senses and references into one general sense, such that all categories in the resulting repository become disjoint. For the word "atmosphere", we could augment the general category ATMOSPHERE to contain all the links previously annotated as ATMOSPHERE, ATMOSPHERE OF EARTH, ATMOSPHERE OF MARS, ATMOSPHERE OF VENUS, or STELLAR ATMOSPHERE. Correspondingly, the new sense repository would be reduced to $\mathscr{R}$ = {ATMOSPHERE, ATMOSPHERE (UNIT), ATMOSPHERE (MUSIC GROUP)}.

2. Keep the original sense and reference repository, but change the definition of some sense categories such that all categories in the repository become mutually disjoint. Correspondingly, the WSD model will be trained to categorize as ATMOSPHERE (O) all contexts of the word "atmosphere" in which either the word is used with a generic sense, or the corresponding reference does not belong to the Wikipedia title repository. The sense repository then becomes $\mathscr{R}$ = {ATMOSPHERE (O), ATMOSPHERE OF EARTH, ATMOSPHERE OF MARS, ATMOSPHERE OF VENUS, STELLAR ATMOSPHERE, ATMOSPHERE (UNIT), ATMOSPHERE (MUSIC GROUP)}.

The first solution is straightforward, however it has the disadvantage that the resulting WSD model will never link words to specific reference titles in Wikipedia like ATMOSPHERE OF EARTH or ATMOSPHERE OF MARS. The rest of this paper describes a feasible implementation for the second solution, which has the advantage that it results in a WSD system that can make more fine grained annotations, down to the reference level. While leading to a more useful system, this second approach is however complicated by the link annotation ambiguity. A WSD system that is trained on sense and reference links extracted automatically from Wikipedia needs to account for the fact that links annotated as ATMOSPHERE may belong either to the general ATMOSPHERE (O) sense category, to the more specific sense STELLAR ATMOSPHERE, or to one of the reference categories ATMOSPHERE OF EARTH, ATMOSPHERE OF MARS, ATMOSPHERE OF VENUS. The distinction between the general sense category ATMOSPHERE and the more specific categories is missing in the Wikipedia link annotations. Since performing an extra step of manual annotation cannot scale to the whole word and phrase vocabulary of Wikipedia, the system needs to be trained with incomplete label information.

## 2 Learning for Sense and Reference Disambiguation

Figure 2 shows our proposed hierarchical classification scheme for disambiguation, using "atmosphere" as the ambiguous word. Shaded leaf nodes show the final categories in the sense repository for each word, whereas the doted frames on the second level in the hierarchy denote artificial categories introduced to enable a finer grained classification into more specific senses or references. Thick arrows illustrate the classification decisions that are made in order to obtain a fine grained disambiguation of the word. Thus, the word "atmosphere" is first classified to have the general sense ATMOSPHERE i.e., "a layer of gases that may surround a material body of sufficient mass, and that is held in place by the gravity of the body". In the first solution, the disambiguation process would stop here and output the general sense ATMOSPHERE. In the second solution, the disambiguation process continues and further classifies the word to

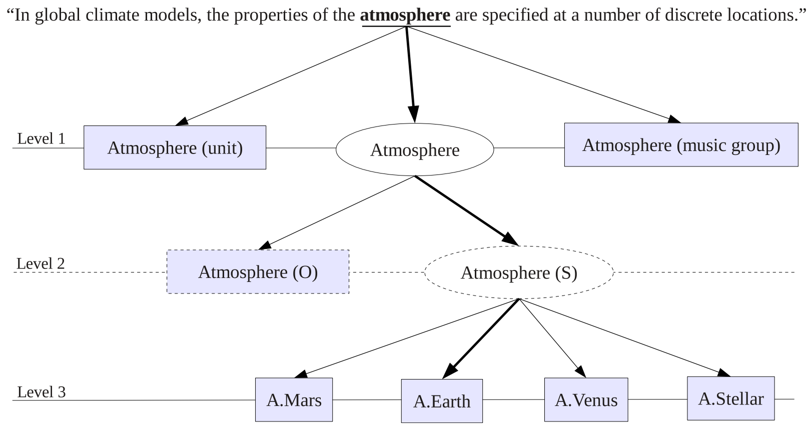


Figure 2: Hierarchical classification for sense and reference disambiguation.

be a reference to Atmosphere of Earth. To get to this final classification, the process passes through an intermediate binary classification level where it determines whether the word has a generic sense or a sense that is not covered in the Wikipedia repository, corresponding to the artificial leaf category Atmosphere (O). In such cases, the system stops the disambiguation process and outputs the general sense category Atmosphere. This disambiguation scheme could be used to relabel the Atmosphere links in Wikipedia with more specific, and therefore more informative, references such as Atmosphere of Earth. According to the statistics from Table 1, for ambiguous words like “atmosphere” there is a significant number of instances where a more specific annotation is possible: out of all 933 instances annotated as Atmosphere in Wikipedia, about 60% (559 of them) could have been labeled with more specific titles.

Training word sense classifiers for Levels 1 and 3 is straightforward. For Level 1, Wikipedia links that are annotated by users as Atmosphere, Atmosphere of Earth, Atmosphere of Mars, Atmosphere of Venus, or Stellar Atmosphere are collected as training examples for the general sense category Atmosphere. Similarly, Wikipedia links that are annotated as Atmosphere (unit) and Atmosphere (music group) will be used as training examples for the two categories, respectively. A binary or multiclass classifier is then trained to distinguish between the two or more categories at this level. For Level 3, binary or multiclass classifiers are trained on Wikipedia links collected for each of the specific senses or references.

For the binary classifier at Level 2, we could use as training examples for the category Atmosphere (O) all Wikipedia links that were annotated as Atmosphere, whereas for the category Atmosphere (S) we will use as training examples all Wikipedia links that were annotated specifically as Atmosphere of Earth, Atmosphere of Mars, Atmosphere of Venus, or Stellar Atmosphere. Using this dataset, we could train a traditional binary classification SVM to distinguish between the two categories. We call this approach *Naive SVM*, since it does not account for the fact that a significant number of the links that are annotated by Wikipedia contributors as Atmosphere should actually belong to the Atmosphere (S) category – about 60% of them, according to Table 1. Alternatively, we could treat all Atmosphere examples as unlabeled examples. If we consider the examples in ATMOSPHERE (S) to be positive examples, then the problem becomes one of *learning with positive and unlabeled examples*.

## 2.1 Learning with positive and unlabeled examples

This general type of semi-supervised learning has been studied before in the context of tasks such as text classification and information retrieval (Lee and Liu, 2003; Liu et al., 2003), or bioinformatics (Elkan and Noto, 2008; Noto et al., 2008). In this setting, the training data consists of positive examples $x \in P$ and unlabeled examples $x \in U$. Following the notation of Elkan and Noto (2008), $s(x) = 1$ if the example is positive and $s(x) = -1$ if the example is unlabeled. The true label of an example is $y(x) = 1$ if the example is positive and $y(x) = -1$ if the example is negative. Thus, $x \in P \Rightarrow s(x) = y(x) = 1$ and $x \in U \Rightarrow s(x) = -1$ i.e., the true label $y(x)$ of an unlabeled example is unknown. In the *Biased SVM* formulation of Lee and Liu (2003), a soft-margin SVM is trained on the $s(x)$ values to optimize an estimate of $pr$, the product between the precision and the recall with respect to the partially hidden true labels $y(x)$. Lee and Liu (2003) show that $pr$ can be estimated using only the observed labels $s(x)$. The other approach used in our experiments is based on the *Weighted Samples SVM* formulation of Elkan and Noto (2008), which assumes that labeled examples $\{x|s(x) = 1\}$ are selected at random from the positive examples $\{x|y(x) = 1\}$ i.e., $p(s = 1|x, y = 1) = p(s = 1|y = 1)$. Correspondingly, a first classifier $g(x)$ is trained on the labeling $s$ to approximate the label distribution i.e., $g(x) = p(s = 1|x)$. The probabilistic output of this classifier is used to create a weighted sample of the original training data, and then a second classifier is trained on the weighted sample to approximate the true labels $y(x)$.

# 3 Experimental Evaluation

We ran disambiguation experiments on the two ambiguous words *atmosphere* and *game*. Their repository of senses and references have been summarized previously in Table 1. All the WSD classifiers evaluated here use the same set of standard WSD features, such as words and their part-of-speech tags in a window of 3 words around the ambiguous keyword, the unigram and bigram content words that are within 2 sentences of the current sentence, the syntactic governor of the keyword, and its chains of syntactic dependencies of lengths up to two. Furthermore, for each example, a Wikipedia specific feature was computed as the cosine similarity between the context of the ambiguous word and the text of the article for the target sense or reference.

The $\text{Level}_1$ and $\text{Level}_3$ classifiers were trained using the $\text{SVM}^{multi}$ component of the $\text{SVM}^{light}$ package.[1] The WSD classifiers were evaluated in a 4-fold cross validation scenario in which 50% of the data was used for training, 25% for tuning the capacity parameter $C$, and 25% for testing. The final accuracy numbers were computed by averaging the results over the 4 folds. For *atmosphere*, the accuracy was 93.1% at $\text{Level}_1$ and 85.6% at $\text{Level}_3$. For *game*, the accuracy was 82.9% at $\text{Level}_1$ and 92.9% at $\text{Level}_3$.

For the binary classifier at $\text{Level}_2$ we follow the same 4-fold cross validation scheme. We emphasize that our manual labels are used only for testing purposes – the manual labels are ignored during training and tuning, when the data is assumed to contain only positive and unlabeled examples that are automatically collected from Wikipedia without any manual effort. We compare the Naive SVM, Biased SVM, and Weighted SVM, using for all of them the same train/development/test splits of the data and the same features.

[1] http://svmlight.joachims.org

| Accuracy | Naive SVM | Biased SVM | Weighted SVM |
|---|---|---|---|
| atmosphere | 39.9% | **79.6%** | 75.0% |
| game | 83.8% | **87.1%** | 84.6% |
| F-measure | Naive SVM | Biased SVM | Weighted SVM |
| atmosphere | 30.5% | **86.0%** | 83.2% |
| game | 75.1% | **81.8%** | 77.5% |

Table 2: Disambiguation results at $\text{Level}_2$.

Table 2 shows the accuracy and F-measure results of the three methods for $\text{Level}_2$. The Biased SVM and the Weighted Samples SVM outperform the Naive SVM on both accuracy and F-measure. The improvement in performance is particularly substantial for the Biased SVM. Based on these initial results, the Biased SVM could be seen as the method of choice for learning with positive and unlabeled examples in the task of sense and reference disambiguation in Wikipedia.

## Conclusion and Future Work

*Sense* and *reference* annotations of words are frequently used without being explicitly distinguished in Wikipedia. Correspondingly, we showed that inconsistencies in link annotations can have a significant impact on the performance of word sense disambiguation systems that are trained on Wikipedia links. We presented an approach to training sense and reference disambiguation systems that treats annotation inconsistencies under the framework of learning with positive and unlabeled examples. Experimental results on two ambiguous words demonstrate that accounting for annotation ambiguity in Wikipedia links leads to consistent improvements in disambiguation accuracy. An accurate sense and reference disambiguation system has the advantage of enabling finer sense distinctions over a generic word sense disambiguation system. It can be used, for example, to annotate general sense links in Wikipedia with more fine grained annotations, down to the reference level.

Annotation inconsistencies in Wikipedia were circumvented by adapting two existing approaches that use only positive and unlabeled data to train binary classifiers. This binary classification constraint led to the introduction of the artificial specific (S) category on $\text{Level}_2$ in our disambiguation framework. In future work, we plan to investigate a more direct extension of learning with positive and unlabeled data to the case of multiclass classification, which will reduce the number of classification levels from three to two. We also plan to evaluate the new disambiguation method on a larger collection of ambiguous words.

## Acknowledgments

This material is based in part upon work supported by the National Science Foundation IIS awards #1018613 and #1018590 and CAREER award #0747340. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

## References

Ahn, D., Jijkoun, V., Mishne, G., Muller, K., de Rijke, M., and Schlobach, S. (2004). Using Wikipedia at the TREC QA track. In *Proceedings of the 13th Text Retrieval Conference (TREC 2004)*.

Bryl, V., Giuliano, C., Serafini, L., and Tymoshenko, K. (2010). Using background knowledge to support coreference resolution. In *Proceedings of the 2010 conference on ECAI 2010: 19th European Conference on Artificial Intelligence*, pages 759–764, Amsterdam, The Netherlands.

Bunescu, R. and Pasca, M. (2006). Using encyclopedic knowledge for named entity disambiguation. In *Proceesings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL-06)*, pages 9–16, Trento, Italy.

Cimiano, P., Schultz, A., Sizov, S., Sorg, P., and Staab, S. (2009). Explicit versus latent concept models for cross-language information retrieval. In *International Joint Conference on Artificial Intelligence (IJCAI-09*, pages 1513–1518, Pasadena, CA.

Cucerzan, S. (2007). Large-scale named entity disambiguation based on Wikipedia data. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 708–716.

Elkan, C. and Noto, K. (2008). Learning classifiers from only positive and unlabeled data. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '08, pages 213–220.

Ferrucci, D. A., Brown, E. W., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A., Lally, A., Murdock, J. W., Nyberg, E., Prager, J. M., Schlaefer, N., and Welty, C. A. (2010). Building watson: An overview of the deepqa project. *AI Magazine*, 31(3):59–79.

Haghighi, A. and Klein, D. (2009). Simple coreference resolution with rich syntactic and semantic features. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 1152–1161, Singapore.

Kaisser, M. (2008). The QuALiM question answering demo: Supplementing answers with paragraphs drawn from Wikipedia. In *Proceedings of the ACL-08 Human Language Technology Demo Session*, pages 32–35, Columbus, Ohio.

Lee, W. S. and Liu, B. (2003). Learning with positive and unlabeled examples using weighted logistic regression. In *Proceedings of the Twentieth International Conference on Machine Learning (ICML*, pages 448–455, Washington, DC.

Li, Y., Luk, R., Ho, E., and Chung, K. (2007). Improving weak ad-hoc queries using Wikipedia as external corpus. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 797–798, Amsterdam, Netherlands.

Liu, B., Dai, Y., Li, X., Lee, W. S., and Yu, P. S. (2003). Building text classifiers using positive and unlabeled examples. In *Proceedings of the Third IEEE International Conference on Data Mining*, ICDM '03, pages 179–186, Washington, DC, USA.

Mihalcea, R. (2007). Using Wikipedia for automatic word sense disambiguation. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics*, pages 196–203, Rochester, New York.

Mihalcea, R. and Csomai, A. (2007). Wikify!: linking documents to encyclopedic knowledge. In *Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management*, Lisbon, Portugal.

Milne, D. (2007). Computing semantic relatedness using Wikipedia link structure. In *Proceedings of the New Zealand Computer Science Research Student Conference*, Hamilton, New Zealand.

Noto, K., Saier, Jr., M. H., and Elkan, C. (2008). Learning to find relevant biological articles without negative training examples. In *Proceedings of the 21st Australasian Joint Conference on Artificial Intelligence: Advances in Artificial Intelligence*, AI '08, pages 202–213.

Ponzetto, S. P. and Strube, M. (2006). Exploiting semantic role labeling, wordnet and wikipedia for coreference resolution. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pages 192–199.

Potthast, M., Stein, B., and Anderka, M. A. (2008). Wikipedia-based multilingual retrieval model. In *Proceedings of the 30th European Conference on IR Research*, Glasgow.

Rahman, A. and Ng, V. (2011). Coreference resolution with world knowledge. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, pages 814–824, Stroudsburg, PA, USA. Association for Computational Linguistics.