OPEN MIND WORD EXPERT:
Creating Large Annotated Data Collections with Web Users' Help

Rada Mihalcea
Department of Computer Science
University of North Texas
rada@cs.unt.edu

Timothy Chklovski
Artificial Intelligence Laboratory
Massachusetts Institute of Technology
timc@mit.edu

Abstract

Open Mind Word Expert is an implemented active learning system that aims to create large annotated corpora by tapping into the world's vast pool of knowledge. It does this by relying on the vast number of Web users who contribute their knowledge to data annotation. Open Mind Word Expert focuses on building semantically annotated corpora, by collecting word sense tagging from the general public over the Web. It is available at http://teach-computers.org. During the first nine months of activity, the system yielded 90,000 high quality tagged items at a much lower cost than the traditional method of hiring lexicographers.

1 Introduction

A large range of Natural Language Processing (NLP) applications require large amounts of annotated data in order to ensure good performance and high accuracy. While recent advances in NLP research brought significant improvement in the performance of NLP methods and algorithms, there has been relatively little progress on addressing the problem of obtaining annotated data required by some of the highest-performing algorithms. As a consequence, many of today's NLP applications experience severe training data bottlenecks.

One notoriously difficult problem in understanding text has been Word Sense Disambiguation (WSD). Ambiguity is very common (especially among the most common words — consider the words "table," or the phrase "computer fan"). Humans, however, are so competent at figuring out word senses from context that they usually do not even notice the ambiguities. While a large number of efficient WSD algorithms have been designed and implemented to date within the recent SENSEVAL evaluation frameworks (Kilgarriff and Palmer, 2000), (Preiss and Yarowsky, 2001), and elsewhere, the availability of sense tagged data is still a significant problem.

Most of the efforts in WSD have focused on *supervised* learning algorithms, which usually achieve the best performance at the cost of low recall. The main weakness of these methods is the lack of widely available semantically tagged corpora and the strong dependence of disambiguation accuracy on the size of the training corpus. For instance, one study reports that high precision WSD requires at least 500 examples per ambiguous word (Ng, 1997)[1]. At a throughput of one tagged example per minute (Edmonds, 2000), and with about 20,000 ambiguous words in the common English vocabulary, a simple calculation leads to about 160,000 hours of tagging, which is nothing less than 80 man-years of human annotation work.[2] Since the tagging process is usually

[1]The number of examples required for a word is highly connected to the word entropy. 500 represents an average.

[2]Similar data bottleneck problems are faced by many other NLP applications. High quality part of speech tagging for English requires about 3 million words annotated with their part of speech. The state-of-the-art in syntactic parsing in English is close to 88-89% , performance attainable by training parser models on a corpus of about 600,000 words

done by trained lexicographers, it is very expensive, and limits the size of such corpora to a handful of tagged texts.

In this paper, we present *Open Mind Word Expert*, a Web-based system that aims to create large sense tagged corpora with the help of Web users. The annotation workload is distributed among millions of potential human annotators, which is likely to significantly reduce the cost and the duration of the annotation process. We investigate the amount and quality of the data produced during nine months of deployment of the activity, and present results obtained during preliminary WSD experiments that rely on this sense tagged data.

Open Mind Word Expert is a project that follows the *Open Mind* initiative (Stork, 1999). The basic idea behind broad *Open Mind* initiative is to use the information and knowledge obtainable from the millions of existing Web users, to the end of creating more intelligent software. Other *Open Mind* projects related to natural language and world knowledge include *Open Mind 1001 Questions* (Chklovski, 2003), which acquires knowledge from millions of users, and *Open Mind Common Sense* (Singh, 2002).

2 Sense Tagged Corpora

The availability of large amounts of semantically tagged data is crucial for creating successful WSD systems. Yet, as of today, only few sense tagged corpora are publicly available.[3]

One of the first large scale hand tagging efforts is reported in (Miller et al., 1993), where a subset of the Brown corpus was tagged with WordNet(Miller, 1995) senses. The corpus includes a total of 234,136 tagged word occurrences, out of which 186,575 are polysemous. There are 88,058 noun occurrences of which 70,214 are polysemous.

The next significant hand tagging task was reported in (Bruce and Wiebe, 1994), where 2,476 usages of *interest* were manually assigned with sense tags from the Longman Dictionary of Contemporary English (LDOCE). This corpus was used in various experiments, with classification accuracies ranging from 75% to 90%, depending on the algorithm and features employed.

The high accuracy of the LEXAS system (Ng and Lee, 1996) is due in part to the use of large corpora. For this system, 192,800 word occurrences have been manually tagged with senses from WordNet. The set of tagged words consists of the 191 most frequently occurring nouns and verbs. The authors mention that approximately one man-year of effort was spent in tagging the data set.

Recently, the SENSEVAL competitions have been providing a good environment for the development of supervised WSD systems, making freely available large amounts of sense tagged data for about 100 words. During SENSEVAL-1 (Kilgarriff and Palmer, 2000), data for 35 words was made available adding up to about 20,000 examples tagged with respect to the Hector dictionary. The size of the tagged corpus increased with SENSEVAL-2 (Preiss and Yarowsky, 2001), when 13,000 additional examples were released for 73 polysemous words. This time, the semantic annotations were performed with respect to WordNet.

Additionally, (Kilgarriff, 1998) mentions the Hector corpus, which comprises about 300 word types with 300-1000 tagged instances for each word, selected from a 17 million word corpus.

With *Open Mind Word Expert* we aim to create a very large sense tagged corpus by making use of the incredible resource of knowledge constituted by the millions of Web users. We use techniques for active learning to utilize this resource efficiently.

3 Open Mind Word Expert

Open Mind Word Expert is a Web-based application that allows contributors to annotate words with their WordNet senses. Tagging is organized by word: for each ambiguous word for which we want to build a sense tagged corpus, users are presented with a set of natural language (English) sentences that include an instance of the ambiguous word.

The overall process proceeds as follows. Ini-

manually parsed within the Penn Treebank project, an annotation effort that required approximately 2 man-years of work (Marcus et al., 1993). Information extraction, automatic summarization, anaphora resolution, and other tasks also strongly require large annotated corpora.

[3] See http://www.senseval.org for a complete list of resources.

tially, example sentences are extracted from a large textual corpus. If other training data is not available, a number of these sentences are presented to the users for tagging in *Stage 1*. Next, this tagged collection is used as training data, and active learning is used to identify in the remaining corpus the examples that are "hard to tag". These are the examples that are presented to the contributors for tagging in *Stage 2*. For all tagging, users are asked to select the sense they find to be the most appropriate in the given sentence. The selection is made from a drop-down list containing all WordNet senses of the current word, plus two additional choices, "unclear" and "none of the above." The results of any automatic classification or the classification submitted by other users are not presented so as to not bias the contributor's decisions. Based on early feedback from both researchers and contributors, a future version of Open Mind Word Expert may allow contributors to specify more than one sense for a given instance. As will be elaborated below, the current approach of collecting redundant tagging already addresses this to some degree.

A prototype of the system has been implemented and is available at http://www.teach-computers.org. Figure 1 shows a screen shot from the system interface, illustrating the screen presented to users when tagging the noun "plane".

3.1 Data

The starting corpus we use is formed by a mix of three different sources of data, namely the *Penn Treebank* corpus (Marcus et al., 1993), the *Los Angeles Times* collection, as provided during TREC conferences,[4] and *Open Mind Common Sense*[5], a collection of about 400,000 commonsense assertions in English as contributed by volunteers over the Web (Singh, 2002)[6]. A mix of several sources, each covering a different spectrum of usage, is used to increase the coverage of word senses and writing styles.

Future versions of *Open Mind Word Expert* will include example sentences extracted from the British National Corpus,[7] and the American National Corpus[8] (the latter as soon as it will become available).

3.2 Active Learning

To minimize the amount of human annotation effort needed to build a tagged corpus for a given ambiguous word, *Open Mind Word Expert* includes an active learning component that has the role of selecting for annotation only those examples that are the most informative.

According to (Dagan et al., 1995), there are two main types of active learning. The first one uses memberships queries, in which the learner constructs examples and asks a user to label them. In natural language processing tasks, this approach is not always applicable, since it is hard and not always possible to construct meaningful unlabeled examples for training. Instead, a second type of active learning can be applied to these tasks, which is *selective sampling*. In this case, several classifiers examine the unlabeled data and identify only those examples that are the most informative, that is the examples where a certain level of disagreement is measured among the classifiers.

We use a simplified form of active learning with selective sampling, where the instances to be tagged are selected as those instances where there is a disagreement between the labels assigned by two different classifiers. The two classifiers are trained on a relatively small corpus of tagged data, which is formed either with (1) Senseval training examples, in the case of Senseval words, or (2) examples obtained with the *Open Mind Word Expert* system itself, when no other training data is available.

The first classifier is a Semantic Tagger with Active Feature Selection (STAFS). This system is one of the top ranked systems in the *English lexical sample* task at SENSEVAL-2. The system consists of an instance based learning algorithm improved with a scheme for automatic feature selection. It relies on the fact that different sets of features have different effects depending on the ambiguous word considered. Rather than creating a general learning model for all polysemous

[4] http://trec.nist.gov

[5] http://commonsense.media.mit.edu

[6] See also (Singh et al., 2002) for additional details regarding the quality of free-form entered information, evaluation, bias, and the level of difficulty of the collected knowledge.

[7] http://www.hcu.ox.ac.uk/BNC/

[8] http://americannationalcorpus.org/

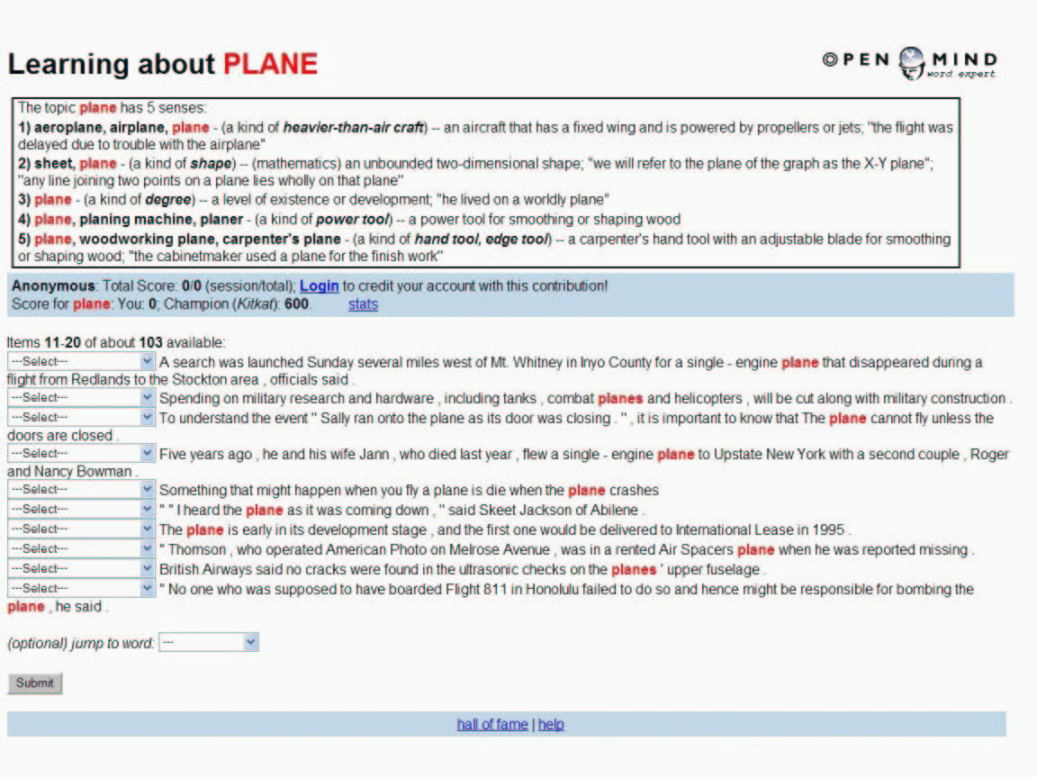

Figure 1: Screen shot from *Open Mind Word Expert*

words, STAFS builds a separate feature space for each individual word. The features are selected from a pool of eighteen different features that have been previously acknowledged as good indicators of word sense, including: part of speech of the ambiguous word itself, surrounding words and their parts of speech, keywords in context, noun before and after, verb before and after, and others. An iterative forward search algorithm identifies at each step the feature that leads to the highest cross-validation precision computed on the training data. More details on this system can be found in (Mihalcea, 2002).

The second classifier is a COnstraint-BAsed Language Tagger (COBALT). The system treats every training example as a set of soft constraints on the sense of the word of interest. WordNet glosses, hyponyms, hyponym glosses and other WordNet data is also used to create soft constraints. Currently, only "keywords in context" type of constraint is implemented, with weights accounting for the distance from the target word. The tagging is performed by finding the sense that minimizes the violation of constraints in the instance being tagged. COBALT generates confidences in its tagging of a given instance based on how much the constraints were satisfied and violated for that instance.

Both taggers use WordNet 1.7 dictionary glosses and relations. The performance of the two systems and their level of agreement were evaluated on the Senseval noun data set. The two systems agreed in their classification decision in 54.96% of the cases. This low agreement level is a good indication that the two approaches are fairly orthogonal, and therefore we may hope for high disambiguation precision on the agreement set. Indeed, the tagging accuracy measured on the set where both COBALT and STAFS assign the same label is 82.5%, a fairly high figure.

Table 1 lists the precision for the agreement and disagreement sets of the two taggers. The low precision on the instances in the disagreement set justifies referring to these as "hard to tag". In *Open Mind Word Expert*, these are the instances that are presented to the users for tagging in the active

System	Precision (fine grained)	Precision (coarse grained)
STAFS	69.5%	76.6%
COBALT	59.2%	66.8%
STAFS∩COBALT	**82.5%**	**86.3%**
STAFS - STAFS∩COBALT	52.4%	63.3%
COBALT - STAFS∩COBALT	30.09%	42.07%

Table 1: Disambiguation precision for the two individual classifiers and their agreement and disagreement sets

learning stage.

3.3 Ensuring Quality

Collecting from the general public holds the promise of providing much data at low cost. It also makes attending to two aspects of data collection more important: (1) ensuring contribution quality, and (2) making the contribution process engaging to the contributors.

To ensure contribution quality, redundant tagging is collected for each item. *Open Mind Word Expert* currently uses the following rules in presenting items to volunteer contributors:

- Two tags per item. Once an item has two tags associated with it, it is not presented for further tagging.
- One tag per item per contributor. We allow contributors to submit tagging either anonymously or having logged in. Anonymous contributors are not shown any items already tagged by contributors (anonymous or not) from the same IP address. Logged in contributors are not shown items they have already tagged.

In all, automatic assessment of the quality of tagging seems possible, and, based on the experience of similar volunteer contribution projects (Singh, 2002), the rate of maliciously misleading or incorrect contributions has been surprisingly low.

Moreover, since we plan to use paid, trained taggers to create a separate test corpus for several of the words tagged with *Open Mind Word Expert*, these same paid taggers could also validate a small percentage of the training data for which no gold standard exists.

4 Results after nine months of activity

During the first nine months of activity, *Open Mind Word Expert* has collected more than 90,000 individual sense taggings from contributors. Of that number, approximately 16,500 tags came from using *Open Mind Word Expert* in the classrooms as a teaching aid (the web site provides special features for this). Future rate of collection depends on the site being listed in various directories and on the contributor repeat visit rate (we are also experimenting with prizes to encourage participation).

There are two main figures that we measured to evaluate the quality of the annotation task. One is *inter tagger agreement*, which represents the agreement between the tags assigned by two different annotators. The other is *replicability*, which measures the degree to which an annotation experiment can be replicated. According to (Kilgarriff, 1999), the capability of recreating a set of annotated data is an indicator for annotation quality that is even more important than the inter-annotator agreement.

4.1 Inter-Tagger Agreement

In terms of inter-annotator agreement, the results obtained so far can be directly compared with the agreement figures previously reported in the literature. (Kilgarriff, 2002) mentions that for the SENSEVAL-2 nouns and adjectives there was a 66.5% agreement between the first two tags collected for each item. About 12% of their tagging consisted of multi-word expressions and proper nouns, which are usually not ambiguous, and which are not considered during our data collection process. So far we measured a 62.8% inter-tagger agreement for single word tagging, plus close-to-100% precision in tagging multi-word expressions and proper nouns (as mentioned earlier, this represents about 12% of the annotated data). This results in an overall agreement of about 66.56% which is reasonable and closely comparable with previous figures.

In addition to raw inter-tagger agreement, the kappa statistic was also determined, which removes from the agreement rate the amount of agreement that is expected by chance (Carletta,

1996). With an average of 5 senses per word, the average value for the chance agreement is 20%[9]. This results in a kappa statistic of 58.2%. Since previous sense annotation experiments have not used this statistic to evaluate the inter-tagger agreement, we have no base for comparison.

4.2 Replicability

To measure the replicability of the tagging process performed through *Open Mind Word Expert*, we had to replicate a tagging experiment where the annotation was performed with "trusted humans." To this end, we used the data set for the noun "interest," created and made available by (Bruce and Wiebe, 1994). In this data set, consisting of 2,369 examples, the annotation was done with respect to LDOCE, and therefore we had first to map the sense entries from this dictionary to WordNet, which is the sense inventory used by *Open Mind Word Expert*. The mapping did not pose any particular problems, and consists of one-to-one mappings for the six LDOCE entries, plus one WordNet entry not defined in LDOCE, for which we discarded all corresponding examples from the Open Mind annotation.

Next, we identified and eliminated all the examples in the corpus that contained collocations (e.g. "interest rate"); these examples accounted for more than 35% of the data. Finally, the remaining 1,438 examples were displayed on the *Open Mind Word Expert* site for tagging.

Out of the 1,438 examples, 1,066 had two tags that agreed, therefore a 74% inter-annotator agreement for single words tagging[10]. Out of these 1,066 items, 967 have a tag that coincides with the tag assigned in the experiments reported in (Bruce and Wiebe, 1994), which leads to an 90.8% replicability for single words tagging (note that the 35% monosemous multi-word expressions are not taken into account by this figure). This is close to the 95% replicability scores mentioned in (Kilgarriff, 1999) for annotation experiments performed by lexicographers.

[9]Note that ideally the chance agreement should take into consideration the entropy of word senses, which implies the availability of sense annotated examples other than those that we are evaluating. Since such examples are not available for all the words in the Open Mind tagged collection, the chance agreement was determined using a simplified assumption of uniform sense distribution

[10]Adding the 35% monosemous multi-word expressions tagged with 100% precision, leads to an overall 83% inter-tagger agreement for this particular word

Number of training examples	Precision		Error rate reduction
	baseline	STAFS	
any	63.32%	66.23%	9%
> 100	75.88%	80.32%	19%
> 200	63.48%	72.18%	24%
> 300	45.51%	69.15%	43%

Table 2: Precision and error rate reduction for various sizes of the training corpus.

4.3 Word Sense Disambiguation using Open Mind Word Expert corpus

For additional evaluations of the quality of the data collected through the *Open Mind Word Expert*, we used these data sets in disambiguation experiments, performed using the STAFS WSD system with a fixed set of features, and consisting in 10-fold cross validation runs. We also computed a simple baseline, consisting of a simple heuristic that assigns the most frequent sense by default (also computed during 10-fold cross validation runs). Table 3 lists all words for which we collected sense tagged data with *Open Mind Word Expert*, the number of items with full inter-annotator agreement, the most frequent sense baseline, and the precision achieved with STAFS[11].

For the 280 words for which data was collected using *Open Mind Word Expert*, the average number of examples per word is 87. The most frequent sense heuristic yields correct results in 63.32% overall. When disambiguation is performed using STAFS, restricting the system to a simple set of features consisting of the word itself, the word's part of speech, and a surrounding context of two (words and their corresponding parts of speech), the overall precision is 66.23%, which represents an error reduction of about 9% with respect to the most frequent sense heuristic, providing additional evidence of usefulness of this corpus.

Moreover, the average for the 72 words which have at least 100 training examples is 75.88% for the most frequent sense heuristic, and 80.32% when using STAFS, resulting in an error reduc-

[11]The Open Mind Word Expert sense tagged corpora used in these experiments is free for download at http://teach-computers.org/download

tion of 19%. When at least 200 examples are available per word, the most frequent sense heuristic is correct 63.48% of the time, and STAFS is correct 72.18% of the time, which represents a 24% reduction in disambiguation error. Table 2 lists the precisions obtained with the most frequent sense heuristic and STAFS, as a function of corpus size. The reduction in error rate grows steadily with the number of training examples. For the words for which more data was collected with *Open Mind Word Expert*, the improvement over the most frequent sense baseline was larger. This agrees with prior work by other researchers (Ng, 1997), (Banko and Brill, 2001), who noted that additional annotated data is likely to bring larger improvements in disambiguation quality.

5 Conclusions and future work

Open Mind Word Expert has the potential of creating a large sense tagged corpus. In this paper we investigated the amount and quality of data collected during the first nine months of deployment of the activity. The experiments performed showed that the inter-tagger agreement, replicability, and disambiguation results obtained on this data are comparable with what can be obtained using data collected with the traditional method of hiring lexicographers, at a much lower cost.

The English sense tagged corpus collected with *Open Mind Word Expert* is continuously growing, and will provide annotated data for the English SENSEVAL-3 lexical sample task. Two new editions, *Romanian Open Mind Word Expert* and *Bilingual Open Mind Word Expert* will soon be deployed. Other languages are also likely to be added in the near future.

Acknowledgments

We want to thank the Open Mind Word Expert contributors who are making all this work possible. We are also grateful to Ted Pedersen and the NLP group at University of Minnesota at Duluth for interesting discussions and important contributions to this data collection process, to Adam Kilgarriff for valuable suggestions, and to all the Open Mind Word Expert users who have emailed us with their feedback and suggestions, helping us improve this activity.

References

M. Banko and E. Brill. 2001. Scaling to very very large corpora for natural language disambiguation. In *Proceedings of the 39th Annual Meeting of the Association for Computational Lingusitics (ACL-2001)*, Toulouse, France, July.

R. Bruce and J. Wiebe. 1994. Word sense disambiguation using decomposable models. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics (ACL-94)*, pages 139–146, LasCruces, NM, June.

J. Carletta. 1996. Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22(2):249–254.

T. Chklovski. 2003. *Effective Knowledge Acquisition*. Ph.D. thesis, MIT. (proposal).

I. Dagan, , and S.P. Engelson. 1995. Committee-based sampling for training probabilistic classifiers. In *International Conference on Machine Learning*, pages 150–157.

P. Edmonds. 2000. Designing a task for Senseval-2, May. Available online at http://www.itri.bton.ac.uk/events/senseval.

A. Kilgarriff and M. Palmer, editors. 2000. *Computer and the Humanities. Special issue: SENSEVAL. Evaluating Word Sense Disambiguation programs*, volume 34, April.

A. Kilgarriff. 1998. Gold standard datasets for evaluating word sense disambiguation programs. *Computer Speech and Language*, 12(4):453–472.

A. Kilgarriff. 1999. 95% replicability for manual word sense tagging. In *Proceedings of European Association for Computational Linguistics*, pages 277–278, Bergen, Norway, June.

A. Kilgarriff. 2002. English lexical sample task description. In *Proceedings of Senseval-2 Workshop, Association of Computational Linguistics*, pages 17–20, Toulouse, France.

M.P. Marcus, B. Santorini, and M.A. Marcinkiewicz. 1993. Building a large annotated corpus of english: the Penn Treebank. *Computational Linguistics*, 19(2):313–330.

R. Mihalcea. 2002. Instance based learning with automatic feature selection applied to Word Sense Disambiguation. In *Proceedings of the 19th International Conference on Computational Linguistics (COLING-ACL 2002)*, Taipei, Taiwan, August.

G. Miller, C. Leacock, T. Randee, and R. Bunker. 1993. A semantic concordance. In *Proceedings of the 3rd DARPA Workshop on Human Language Technology*, pages 303–308, Plainsboro, New Jersey.

G. Miller. 1995. Wordnet: A lexical database. *Communication of the ACM*, 38(11):39–41.

H.T. Ng and H.B. Lee. 1996. Integrating multiple knowledge sources to disambiguate word sense: An examplar-based approach. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics (ACL-96)*, Santa Cruz.

H.T. Ng. 1997. Getting serious about word sense disambiguation. In *Proceedings of the ACL SIGLEX Workshop on Tagging Text with Lexical Semantics: Why, What, and How?*, pages 1–7, Washington.

J. Preiss and D. Yarowsky, editors. 2001. *Proceedings of SENSEVAL-2, Association for Computational Linguistics Workshop*, Toulouse, France.

P. Singh, T. Lin, E. Mueller, G. Lim, T. Perkins, and W. Li Zhu. 2002. Open mind common sense: Knowledge acquisition from the general public. In *Proceedings of the First International Conference on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems*. Lecture Notes in Computer Science, Heidelberg: Springer-Verlag.

P. Singh. 2002. The public acquisition of commonsense knowledge. In *Proceedings of AAAI Spring Symposium: Acquiring (and Using) Linguistic (and World) Knowledge for Information Access.*, Palo Alto, CA. AAAI.

D. Stork. 1999. The Open Mind initiative. *IEEE Expert Systems and Their Applications*, 14(3):19–20.

Word	Set size	Baseline	STAFS	Word	Set size	Baseline	STAFS	Word	Set size	Baseline	STAFS
act	119	80.00%	79.50%	activity	103	90.00%	90.00%	afternoon	97	100.00%	100.00%
age	62	78.75%	75.00%	amount	63	58.89%	51.11%	analysis	90	56.67%	46.67%
animal	60	100.00%	100.00%	answer	67	51.54%	44.62%	arc	59	77.14%	82.14%
area	98	60.00%	50.59%	argument	82	25.00%	57.00%	arm	142	52.50%	80.62%
art	107	30.00%	63.53%	aspect	64	54.00%	50.00%	atmosphere	86	28.57%	49.29%
attempt	95	90.71%	90.71%	attention	83	60.00%	55.45%	attitude	107	100.00%	100.00%
audience	84	55.83%	68.33%	author	94	62.31%	73.85%	authority	11	25.00%	30.00%
award	77	61.43%	47.86%	bank	160	91.88%	91.88%	bar	107	61.76%	70.59%
basis	47	98.18%	98.18%	bed	142	98.12%	98.12%	behavior	58	54.62%	45.38%
blood	136	91.05%	91.05%	brother	101	95.45%	95.45%	building	114	87.33%	88.67%
bum	47	40.91%	53.64%	captain	101	47.27%	48.18%	car	144	99.44%	99.44%
cell	126	89.44%	88.33%	chair	38	93.64%	93.64%	chance	115	56.25%	81.88%
channel	103	84.62%	86.15%	chapter	137	68.50%	71.50%	child	105	55.33%	84.67%
church	93	70.00%	75.83%	circuit	197	31.92%	45.77%	circumstance	66	52.50%	50.83%
city	86	89.29%	85.71%	claim	75	51.67%	41.67%	coffee	115	95.00%	95.00%
college	84	93.33%	93.33%	completion	87	56.67%	70.67%	concentration	78	40.00%	56.67%
concern	84	32.50%	60.00%	condition	90	55.56%	52.22%	contrast	50	64.00%	54.00%
cost	36	43.33%	48.89%	country	66	59.17%	55.83%	culture	59	40.71%	44.29%
day	192	34.76%	44.76%	decision	86	55.71%	60.71%	degree	140	71.43%	82.14%
demand	41	10.00%	30.00%	depth	60	43.33%	50.00%	detail	57	36.67%	50.83%
detention	38	65.45%	65.45%	device	106	98.12%	98.12%	difference	76	8.46%	57.69%
difficulty	60	30.00%	56.67%	discussion	44	63.75%	53.75%	distance	54	58.89%	53.33%
distribution	75	60.83%	55.83%	doctor	133	100.00%	100.00%	dog	130	100.00%	100.00%
door	112	54.62%	45.38%	dream	75	50.83%	39.17%	dust	46	63.00%	57.00%
earth	89	80.00%	80.59%	edge	68	47.86%	54.29%	education	56	63.64%	53.64%
effect	72	92.22%	92.22%	effort	66	15.83%	56.67%	election	28	42.00%	64.00%
element	88	75.00%	68.12%	enemy	41	48.00%	52.00%	energy	76	64.62%	56.92%
evening	47	45.45%	66.36%	event	77	32.14%	37.14%	evidence	65	52.73%	54.55%
example	42	26.67%	18.33%	existence	65	85.45%	76.36%	experience	87	44.67%	54.67%
experiment	51	63.33%	68.33%	extent	53	76.25%	97.50%	eye	117	96.11%	96.11%
facility	205	81.60%	74.40%	fact	172	53.16%	47.89%	factor	89	67.65%	64.12%
family	95	61.43%	50.71%	father	160	96.88%	96.88%	fatigue	20	70.00%	70.00%
fear	52	75.71%	75.71%	feature	80	56.25%	55.00%	feeling	48	45.00%	25.00%
fig	72	76.67%	78.89%	film	98	86.47%	79.41%	finger	78	100.00%	100.00%
flower	42	78.33%	54.81%	friend	57	50.83%	50.83%	function	105	24.67%	32.00%
future	73	78.00%	100.00%	gas	52	48.57%	40.00%	girl	68	53.57%	66.43%
glass	93	65.83%	75.83%	god	110	71.82%	81.82%	government	82	79.00%	80.00%
grip	239	45.94%	61.88%	growth	76	42.31%	43.85%	gun	143	94.71%	94.71%
hair	147	96.67%	96.67%	history	57	58.33%	47.50%	holiday	29	100.00%	100.00%
home	46	19.00%	27.00%	hope	67	52.31%	42.31%	horse	138	100.00%	100.00%
hour	79	95.00%	95.00%	idea	41	56.00%	40.00%	image	120	36.67%	71.67%
importance	64	93.00%	93.00%	increase	43	44.29%	32.86%	individual	103	96.15%	96.15%
industry	83	93.64%	93.64%	influence	44	41.25%	40.00%	information	62	56.25%	46.25%
intensity	88	85.00%	76.88%	interest	1066	39.91%	71.08%	item	85	74.62%	74.62%
judgment	85	19.23%	28.46%	kid	106	83.75%	84.38%	knee	29	80.00%	75.45%
labor	59	34.29%	30.00%	lady	9	-	-	language	76	53.08%	51.54%
law	106	38.12%	66.88%	length	45	42.22%	62.22%	letter	137	85.00%	81.00%
level	80	37.50%	33.75%	lip	96	90.67%	90.67%	list	102	100.00%	100.00%
literature	57	54.17%	58.33%	manager	91	98.00%	98.00%	manner	53	73.75%	67.50%
marriage	20	60.00%	40.00%	material	196	77.60%	76.40%	matter	46	16.00%	37.00%
meaning	77	55.00%	55.71%	memory	54	42.22%	35.56%	method	60	56.67%	61.67%
mind	57	57.50%	48.33%	minute	93	59.17%	74.17%	mission	14	46.00%	50.00%
moment	63	51.11%	61.11%	money	46	67.00%	62.00%	morning	71	76.25%	76.25%
mother	119	99.00%	99.00%	mouth	151	74.38%	77.50%	music	50	48.00%	70.00%
name	136	98.42%	98.42%	nation	21	73.33%	70.00%	nature	83	80.00%	81.82%
need	73	54.00%	61.00%	neighborhood	67	39.23%	60.00%	newspaper	78	84.00%	78.00%
object	183	96.19%	96.19%	objective	74	100.00%	100.00%	office	209	62.76%	61.03%
officer	103	56.15%	55.38%	onset	97	94.38%	94.38%	organization	46	33.00%	40.00%
pain	83	65.45%	61.82%	paper	81	73.33%	70.00%	particle	95	74.29%	75.71%
party	95	45.71%	49.29%	past	71	55.00%	52.50%	people	120	99.17%	99.17%
performance	61	40.00%	42.86%	phase	71	98.75%	98.75%	plan	243	95.56%	95.56%
plane	46	90.00%	90.00%	plant	126	98.89%	98.89%	policy	97	94.38%	94.38%
portion	5	-	-	possibility	61	44.29%	45.71%	post	49	78.46%	74.62%
presence	5	-	-	pressure	106	72.50%	70.62%	price	84	82.50%	70.83%
problem	60	20.00%	50.00%	procedure	68	50.00%	47.14%	process	96	79.33%	72.00%
product	216	80.74%	81.48%	production	63	48.89%	34.44%	property	99	77.78%	71.67%
purpose	92	75.45%	58.18%	quality	84	75.00%	70.83%	question	53	75.00%	63.75%
radiation	55	72.00%	68.00%	rate	94	59.23%	73.08%	reaction	76	80.77%	68.46%
reality	5	-	-	reason	72	73.33%	67.78%	region	66	65.00%	55.83%
relationship	5	-	-	religion	71	33.75%	61.25%	report	101	66.36%	60.91%
requirement	51	48.33%	36.67%	research	3	-	-	rest	360	51.11%	67.22%
restraint	204	22.92%	46.25%	result	57	37.50%	69.17%	road	93	99.17%	99.17%
role	66	31.67%	54.17%	room	124	100.00%	100.00%	sale	87	74.67%	85.33%
sample	43	45.71%	40.00%	school	82	33.00%	52.00%	sea	205	90.80%	90.80%
season	102	92.50%	92.50%	sense	58	21.54%	75.38%	series	95	44.29%	63.57%
shape	79	52.50%	56.88%	share	93	69.17%	80.00%	shelter	81	85.56%	80.00%
shoulder	92	89.09%	80.00%	site	66	69.17%	66.67%	situation	43	47.14%	44.29%
size	83	74.55%	69.09%	skill	55	66.00%	63.00%	society	94	82.31%	73.85%
soil	61	61.43%	62.86%	soldier	95	98.57%	98.57%	song	116	92.35%	92.35%
sort	98	62.94%	86.47%	source	69	33.33%	60.67%	spade	11	100.00%	100.00%
statement	73	4.00%	64.00%	story	77	50.71%	65.71%	street	78	45.33%	30.67%
stress	24	65.00%	60.00%	structure	112	75.38%	72.31%	student	75	96.67%	96.67%
success	44	38.75%	45.00%	sun	101	63.64%	66.36%	surface	87	50.67%	58.67%
table	89	60.59%	68.24%	team	96	100.00%	100.00%	technique	64	90.00%	90.00%
term	125	71.18%	90.59%	test	9	-	-	text	48	22.50%	49.17%
theory	31	62.50%	55.00%	thought	51	65.00%	53.33%	tissue	95	85.00%	82.86%
town	74	83.64%	76.36%	trade	11	20.00%	-	training	44	100.00%	100.00%
treatment	108	67.78%	66.67%	tree	105	100.00%	100.00%	trial	109	87.37%	86.84%
trouble	73	65.00%	53.00%	type	135	92.78%	92.78%	understanding	29	37.27%	43.64%
unit	108	54.44%	46.67%	use	92	85.45%	82.73%	value	84	22.50%	56.67%
volume	103	63.85%	54.62%	war	86	72.14%	82.14%	water	103	53.85%	72.31%
wave	80	51.25%	51.25%	week	39	70.00%	78.33%	window	60	33.33%	36.67%
woman	65	26.36%	37.27%	work	89	51.18%	61.76%	worker	33	93.33%	93.33%

Table 3: Words with examples sense tagged in *Open Mind Word Expert*: (1) set size, (2) precision attainable with the most frequent sense heuristic, (3) precision attainable with the STAFS WSD system