Towards Automatic Evaluation of Wordnet Synsets

J. Ramanand and Pushpak Bhattacharyya

Department of Computer Science and Engineering,
Indian Institute of Technology, Bombay
{ramanand@it, pb@cse}.iitb.ac.in

Increasing and varied applications of wordnets call for the creation of methods to evaluate their quality. However, no such comprehensive methods to rate and compare wordnets exist. We begin our search for wordnet evaluation strategies by attempting to validate synsets. As synonymy forms the basis of synsets, we present an algorithm based on dictionary definitions to verify that the words present in a synset are indeed synonymous. Of specific interest are synsets in which some members "do not belong". Our work, thus, is an attempt to flag human lexicographers' errors by accumulating evidences from myriad lexical sources.

1 Introduction

Lexico-semantic networks such as the Princeton WordNet ([1]) are now considered vital resources in several applications in Natural Language Processing and Text mining. Wordnets are being constructed in different languages as seen in the EuroWordNet ([2]) project and the Hindi WordNet ([3]). Competing lexical networks, such as ConceptNet ([4]), HowNet ([5]), MindNet ([6]), VerbNet [7], and FrameNet ([8]), are also emerging as alternatives to wordnets. Naturally, users would be interested in knowing not only the relative merits from among a selection of choices, but also the intrinsic value of such resources. Currently, there are no measures of quality to evaluate or differentiate these resources.

A study of lexical networks could involve understanding the size and coverage, domain applicability, and content veracity of the resource. This is especially critical in cases where wordnets will be created by automated means, especially to leverage existing content in related languages, in contrast to the slower manual process of wordnet creation which has been the traditional method.

The motivation for evaluating wordnets is to help answer questions such as the following:

1. How to select one lexico-semantic network over another?
2. Is a given wordnet sound and complete?
3. Is this resource usable, scalable, and deployable?
4. Is this wordnet suitable for a particular domain or application?

A theory of evaluation must address the following issues:

1. Establishing criteria to measure intrinsic quality of the content held in these lexical networks.

2. Establishing criteria to make useful comparisons between different lexico-semantic networks.
3. Methods to check if a network's quality has improved or declined after content updates.
4. Quality of content in the synsets and relationships between synsets

This paper is organized as follows: In Section 2, we briefly survey work related to the area of ontology evaluation. This is followed in Section 3 by an introduction to the novel problem of validating synonyms in a synset. In Section 4, we describe our dictionary-based algorithm in detail. We discuss the experimental setup and results in Section 5. Finally, in Section 6, we present the key conclusions from our work.

2 Related Work

2.1 Evaluations of Lexico-Semantic Networks

Our literature survey revealed that, to the best of our knowledge, there have been no comprehensive efforts to evaluate wordnets or other lexico-semantic networks on general principles. [9] describes a statistical survey of WordNet v1.1.7 to study types of nodes, dimensional distribution, branching factor, depth and height. A syntactic check and usability study of the BalkaNet resource (wordnets in Eastern European languages) has been described in [10]. The creators of the common-sense knowledge base ConceptNet carried out an evaluation of their resource based on a statistical survey and human evaluation. Their results are described in [4]. [11] discuss evaluations of knowledge resources in the context of a Word Sense Disambiguation task. [12] apply this in a multi-lingual context. Apart from these, we are not aware of any other major evaluations of any lexico-semantic networks.

2.2 Evaluations of Ontologies

In the related field of ontologies, several evaluation efforts have been described. As lexical networks can be viewed as common-sense ontologies, a study of ontology evaluations may be useful. [13] describes an attempt at creating a formal model of an ontology with respect to specifying a given vocabulary's intended meaning. The paper provides an interesting theoretical basis for evaluations. [14] provides a classification of ontology content evaluation strategies and also provides an additional perspective to evaluation based on the “level" of appraisal (such as at the lexical, syntactic, data, design levels). [15] describes some metrics which have been used in the context of ontology evaluation.

Some ontology evaluation systems have been developed and are in use. One of these is OntoMetric ([16]), a method that helps users pick an ontology for a new system. It presents a set of processes that the user should carry out to obtain the measures of suitability of existing ontologies, regarding the requirements of a particular system. The OntoClean ([17]) methodology is based on philosophical notions for a formal evaluation of taxonomical structures. It focuses on the cleaning of taxonomies. [18]

describes a task-based evaluation scheme to examine ontologies with respect to three basic levels: vocabulary, taxonomy and non-taxonomic semantic relations. A score based on error rates was designed for each level of evaluation. [19] describes an ontology evaluation scheme that makes it easier for domain experts to evaluate the contents of an ontology. This scheme is called OntoLearn.

We felt that none of the above methods seemed to address the core issues particular to wordnets, and hence we approached the problem by looking at synsets.

3 Synset Validation

3.1 Introduction

Synsets are the foundations of a wordnet. A wordnet synset is constructed by putting together a set of synonyms that together define a particular sense uniquely, as given by the principles of minimality and coverage described in the previous section. This sense is explicitly indicated for human readability by a gloss. For instance, the synset {proboscis, *trunk*} represents the sense of *"a long flexible snout as of an elephant"*, as opposed to the synset {luggage compartment, automobile trunk, *trunk*} which is *"a compartment in an automobile that carries luggage or shopping or tools"*. Words with potentially multiple meanings are associated together, out of which a single sense emerges. To evaluate the quality of a synset, we began by looking at the validity of its constituent synonyms.

Before the validation, the following theoretical questions must be addressed:
1. What is the definition of a synonym?
2. What are the necessary and sufficient conditions to determine that synonymy exists among a group of words?

Intuitively, synonymy exists between two words when they share a similar sense. This also implies that one word can be replaced by its synonym in a context without any loss of meaning. In practice, most words are not perfect replacements for their synonyms i.e. they are near synonyms. There could be contextual, collocational and other preferences behind replacing synonyms. [20] describes attempts to mathematically describe synonymy. To the best of our knowledge, *no necessary and sufficient conditions to prove that two words are synonyms of each other* have been explicitly stated.

The foundation of our work is the following: we conjecture that
1. if two words are synonyms, it is necessary that they must share one common meaning out of all the meanings they could possess.
2. A sufficient condition could be showing that the words replace each other in a context without loss of meaning.

The task of synset validation has the following subtasks:
1. Are the words in a synset indeed synonyms of each other?

2. Are there any words which have been omitted from the synset?
3. Does the combination of words indicate the required sense?

In this paper, we attempted to answer the first question above i.e. given a set of words, could we verify if they were synonyms? Our literature survey revealed that though much work had been done in the automated discovery of synonyms (from corpora and dictionaries), no work had been done in automatically verifying whether two words were synonyms. Nevertheless, we began by studying some of the synonym discovery methods available.

3.2 Related Work on Automatic Synset Creation

All these methods are based on web and corpora mining. [21] describes a method to collect synonyms in the medical domain from the Web by first building a taxonomy of words. [22] provides an unsupervised learning method for extracting synonyms from the Web. [23] shows an interesting topic signature method to detect synonyms using document contexts and thus enrich large ontologies. Finally, [24] is a survey of different synonym discovery methods, which also proposes its own dictionary-based solution for the problem. Its dictionary based approach provides some useful hints for our own experiments in synonymy validation.

3.3 Our Approach

We focus only on the problem of checking whether the words in a synset can be shown to be synonyms of each other and thus correctly belong to that synset. As of now, we do not flag omissions in the synsets. It is to be also noted that failure to validate the presence of a word in a synset does not strongly suggest that the word is incorrectly entered in the synset - it merely raises a flag for human validation.

The input to our system is a wordnet synset which provides the following information:
1. The synonymous words in the synset
2. The hypernym(s) of the synset
3. Other linked nodes, gloss, example usages

The output consists of a verdict on each word as to whether it fits in the synset, i.e. whether it qualifies to be the synonym of other words in the synset, and hence, whether it expresses the sense represented by the synset. A block diagram of the system is shown in Fig.1.

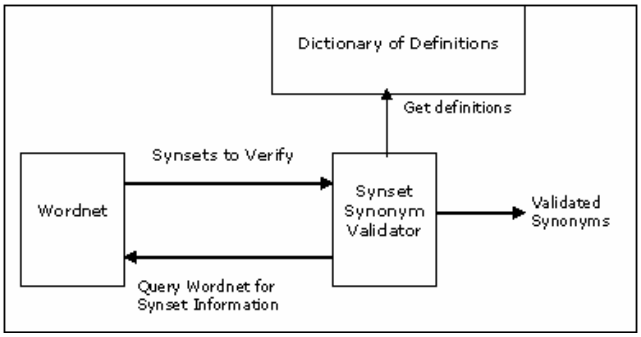

Fig. 1. Block Diagram for Synset Synonym Validation

4 Our Dictionary-based Algorithm

4.1 The Basic Idea

In dictionaries, a word is usually defined in terms of its hypernyms or synonyms. For instance, consider definitions of the word *snake*, whose hypernym is *reptile*, and its synonyms *serpent* and *ophidian* (obtained from the website Dictionary.com [22]):

snake: any of numerous limbless, scaly, elongate *reptiles* of the suborder Serpentes, comprising venomous and non-venomous species inhabiting tropical and temperate areas.
serpent: a *snake*
ophidian: A member of the suborder Ophidia or Serpentes; a *snake*.
This critical observation suggests that dictionary definitions may provide useful clues for verifying synonymy.

We use the following hypothesis:

if a word is present in a synset, there is a dictionary definition for it which refers to its hypernym or to its synonyms from the synset.

Instead of matching synonyms pair-wise, we try to validate the presence of the word in the synset using the hypernyms of the synset and the other synonyms in the synset. A given word belongs to a given synset if there exists a definition for that word, which refers to one of the given hypernym words or one of the synonyms. We use the hypernyms and synonyms to validate other synonyms by mutual reinforcement.

4.2 Algorithm Description

The dictionary-based algorithm consists in applying three groups of rules in order. The first group applies to each word individually, using its dictionary definitions. The second group relies on a set of words collected for the entire synset during the application of the first group. The final group consists of rules that do not use the dictionary definitions. (All definitions in this section are from the website Dictionary.com [25].)

In this section, we describe the steps of the algorithm with examples. The Algorithm has been stated in Section 4.3.

Group 1
Rule 1 - Hypernyms in Definitions
Definitions of words for particular senses often make references to the hypernym of the concept. Finding such a definition means that the word's placement in the synset can be defended.
e.g.
Synset: *{brass, brass instrument}*
Hypernym: *{wind instrument, wind}*
Relevant Definitions:
brass instrument: a musical *wind instrument* of brass or other metal with a cup-shaped mouthpiece, as the trombone, tuba, French horn, trumpet, or cornet.

Rule 2 - Synonyms in Definitions
Definitions of words also make references to fellow synonyms, thus helping to validate them.
e.g.
Synset: *{anchor, ground tackle}*
Hypernym: *{hook, claw}*
Relevant Definitions:
ground tackle: equipment, as *anchors*, chains, or windlasses, for mooring a vessel away from a pier or other fixed moorings.

Rule 3 - Reverse Synonym Definitions
Definitions of synonyms may also make references to the word to be validated.
e.g.
Synset: *{Irish Republican Army, IRA,Provisional Irish Republican Army, Provisional IRA, Provos}*
Hypernym: *{terrorist organization, terrorist group, foreign terrorist organization, FTO}*
Relevant Definitions:
Irish Republican Army: an underground Irish nationalist organization founded to work for Irish independence from Great Britain: declared illegal by the Irish government in 1936, but continues activity aimed at the unification of the Republic of Ireland and Northern Ireland.
Provos: member of the Provisional wing of the *Irish Republican Army*.

Here *Irish Republican Army* can be validated using the definition of *Provos*.

Rules 4 and 5 - Partial Hypernyms and Synonyms in Definitions
Many words in the wordnet are multi-words, i.e., they are made up of more than one word. In quite a few cases, such multi-word hypernyms are not entirely present in the definitions of words, but parts of them can be found in the definition.
e.g.
Synset: *{fibrinogen, factor I}*
Hypernym: *{coagulation factor, clotting factor}*
Relevant Definitions:
fibrinogen: a globulin occurring in blood and yielding fibrin in blood *coagulation*.

Group 2
Rule 6 – Bag of Words from Definitions
In some cases, definitions of a word do not refer to synonyms or hypernym words. However, the definitions of two synonyms may share common words, relevant to the context of the sense. This rule captures this case.

When a word is validated using Group 1 rules, the words of the validating definition are added to a collection. After applying Group 1 rules to all words in the synset, a bag of these words (from all validating definitions seen so far) is now available. For each remaining synonym yet to be validated, we look for any definition for it which contains one of the words in this bag.
e.g.
Synset: *{serfdom, serfhood, vassalage}*
Hypernym: *{bondage, slavery, thrall, thralldom, thraldom}*
Relevant Definitions
serfdom: person (held in) *bondage*; servitude
vassalage: dependence, subjection, *servitude*

serfdom is matched on account of its hypernym *bondage* being present in its definition. So the Bag of Words now contains *"person, bondage, servitude"*.
No definition of *vassalage* could be matched with any of the rules from 1 to 5. But Rule 6 matches the word *servitude* and so helps validate the word.

Group 3
Rules 7 and 8 - Partial Matches of Hypernyms and Synonyms
Quite a few words to be validated are multi-words. Many of these do not have definitions present in conventional dictionaries, which make the above rules inapplicable to them. Therefore, we use the observation that, in many cases, these multi-words are variations of their synonyms of hypernyms i.e. the multi-words share common words with them. Examples of these are synsets such as:

1. *{dinner theater, dinner theatre}*: No definition was available for *dinner theatre*, possibly because of the British spelling.
2. *{laurel, laurel wreath, bay wreath}*: No definitions for the two multi-words.

3. *{Taylor, Zachary Taylor, President Taylor}*: No definition for the last multi-word.

As can be seen above, the multi-word synonyms do share partial words. To validate such multi-words without dictionary entries, we check for the presence of partial words in their synonyms.
e.g.
Synset: *{Taylor, Zachary Taylor, President Taylor}*
Hypernym: *{President of the United States, United States President, President, Chief Executive}*
Relevant Definitions:
Taylor, Zachary Taylor: (1784-1850) the 12th President of the United States from 1849-1950.
President Taylor: - no definition found -

The first two words have definitions which are used to easily validate them. The third word has no definition, and so rules from Group 1 and 2 do not apply to it. Applying the Group 3 rules, we look for the component words in the other two synonyms. Doing this, we find *"Taylor"* in the first synonym, and hence validate the third word.

A similar rule can be defined for a multi-word hypernym, wherein we look for the component word in the hypernym words. In this case, we would match the word *"President"* in the first hypernym word.

We must note that, in comparison to the other rules, these rules are likely to be susceptible to erroneous decisions, and hence a match using these rules should be treated as a weak match. The reason for creating these two rules is to overcome the scarcity of definitions for such multi-words.

4.3 Algorithm Statement

```
Algorithm 1 – Validating wordnet synsets using a dictionary

1: Input: synset S, words W in synset S, Dictionary of definitions

2: For each word w belonging to W do

3: Apply rules in Group 1:

 - 3.1: (Rule 1) Find a definition for w in the dictionary such that it contains a hypernym word h (repeat with other hypernyms if necessary)

   - 3.2: (Rule 2) Else, find a definition for w containing any synonym of w from the synset

   - 3.3: (Rule 3) Else, find a synonym's definition referring to w
```

```
   - 3.4: (Rule 4) (applicable to multi-words in the
hypernym) Else, find a definition of w referring to a
partial word from a multi-word in the hypernym

   - 3.5: (Rule 5) (applicable to synonyms that are
multi-words) Else, find a definition for w referring to
a partial word from a multi-word synonym

4: Apply the rule 6 in Group 2:

   - 4.1: For every word m from the synset that was
matched by one of the above rules, add the words in the
validating definition for m to a collection of words C.

   - 4.2: For each word w in the synset that has not
been validated, find a definition d of w such that d
has a word appearing in C.

5: Apply rules in Group 3 to each remaining unmatched
word w:

   - 5.1: (Rule 7) See if a partial word from a multi-
word w is found in the synonym to be matched

   - 5.2: (Rule 8) Else, see if a partial word from the
multi-word w is found in a hypernym word h.

6: end for
```

5 Experimental Results

5.1 Setup

The validation was tested on the Princeton WordNet (v2.1) noun synsets. Out of the 81426 noun synsets, 39840 are synsets with more than one word – only these were given as input to the validator. This set comprised of a total of 103620 words.

One of the contributions of our work is the creation of a *super dictionary* which consists of words and their definitions constructed by automatic means from the online dictionary service *Dictionary.com* ([25]) (which aggregates definitions from various sources such as *Random House Unabridged Dictionary*, *American Heritage Dictionary*, etc.) Of these, definitions from *Random House* and *American Heritage* dictionaries were identified and added to the dictionary being created. English stop words were removed from the definitions, and the remaining words were stemmed using Porter's stemmer [26]. The resulting dictionary had 463487 definitions in all for a total of 49979 words (48.23% of the total number of words).

5.2 Results and Discussions

Figs. 2, 3, and 4 summarise the main results obtained by running the dictionary-based validator on 39840 synsets. As shown in Fig. 2, 14844 out of the 18322 unmatched

words did not have definitions in the dictionary. Therefore, there are 88776 words which either have definitions in the dictionary, or are referenced in the dictionary, or are matched by the partial rules. So, considering only these 88776 words, there are 85298 matched words, i.e. a validation value of 96.08%.

In about 9% of all synsets, none of the words in the synset could be verified. Of these 3660 synsets, 2952 (80%) had only 2 words in them. The primary reason for this typically was one member of the synset not being present in the dictionary, and hence reducing the number of rules applicable to the other word.

Failure to validate a word does not mean that the word in question is incorrectly present in the synset. Instead, it flags the need for human intercession to verify whether the word indeed has that synset's sense. The algorithm is not powerful enough to make a firm claim of erroneous placement. In an evaluation system, the validator can serve as a useful first-cut filter to reduce the number of words to be scrutinised by a human expert. In some cases, the non-matches did raise some interesting questions about the validity of a word in a synset. We discuss some examples in the next section.

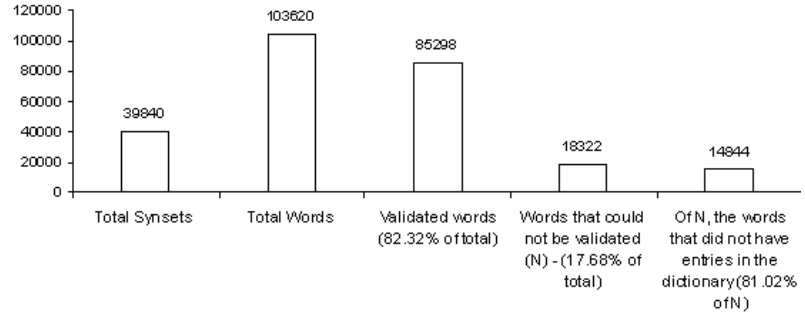

Fig. 2. The Dictionary Approach: Summary of results

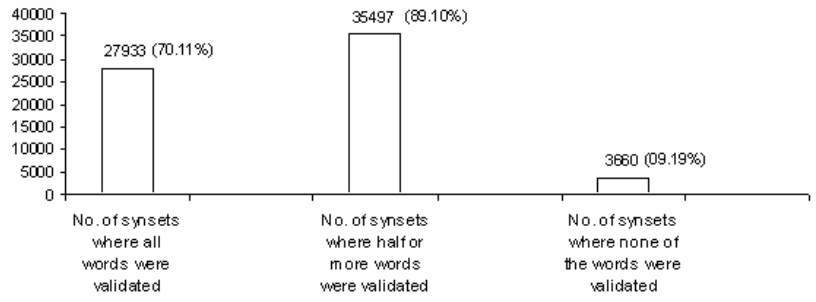

Fig. 3. The Dictionary Approach: A synset perspective

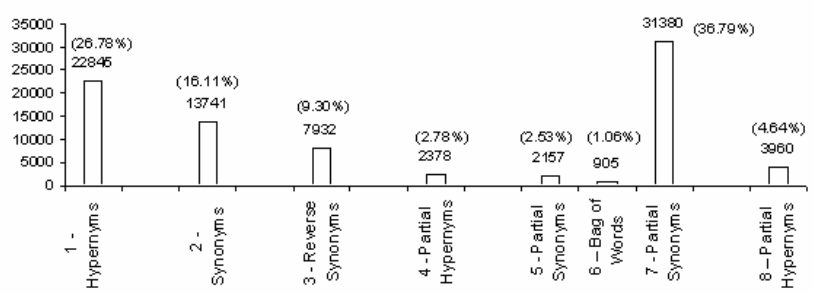

Fig. 4. The Dictionary Approach: Rule-wise summary

5.3 Case Studies

(All sources for definitions in the following examples are from the website Dictionary.com [25])

5.3.1 Possible True Negatives flagged by the validator

The validator could not match about 18% of all words. In most of these cases, the words are indeed correctly placed (as one would expect of a resource manually created by experts) but are flagged incorrectly by the validator as it is as yet not powerful enough to match them. However, consider the following cases of words where non-matches are interesting to study.

Instance 1:
Synset: *{visionary, illusionist, seer}*
Hypernym: *{intellectual, intellect}*
Gloss: *a person with unusual powers of foresight*

The word *"illusionist"* was not matched in this context. This seems to be a highly unusual sense of this word (more commonly seen in the sense of *"conjuror"*). None of the dictionaries consulted provided this meaning for the word.

Instance 2:
Synset: *{bobby pin, hairgrip, grip}*
Hypernym: *{hairpin}*
Gloss: *a flat wire hairpin whose prongs press tightly together; used to hold bobbed hair in place*

It could not be established from any other lexical resource whether *grip*, though a similar sounding word to *hairgrip*, was a valid synonym for this sense. Again, this

could be a usage local to some cultures, but this was not readily supported by other dictionaries.

5.3.2 True Positives correctly flagged by the validator
Here are examples of the validator correctly flagging matches.

Instance 1
Synset: *{smokestack, stack}*
Word to be validated: *smokestack*
Hypernym: *{chimney}*
Relevant Definitions:
smokestack: A large *chimney* or vertical pipe through which combustion vapors, gases, and smoke are discharged.

Instance 2
Synset: *{zombi, zombie, snake god}*
Word to be validated: *snake god*
Hypernym: *{deity, divinity, god, immortal}*
Relevant Definitions:
zombie: a *snake god* worshiped in West Indian and Brazilian religious practices of African origin.

5.3.2 False Negatives flagged by the validator
Here are examples of the validator being unable to match words, despite definitions being present:

Instance 1
Synset: *{segregation, separatism}*
Word to be validated: *segregation*
Hypernym: *{social organization, social organisation, social structure, social system, structure}*
Relevant Definitions:
segregation: The act or practice of segregating
segregation: the state or condition of being segregated

Noun forms of such verbs typically refer to the act, which makes it hard to validate using other words.

Instance 2
Synset: *{hush puppy, hushpuppy}*
Word to be validated: *hush puppy*
Hypernym: *{cornbread}*
Relevant Definitions:
Hush puppy: a small, unsweetened cake or ball of cornmeal dough fried in deep fat.

Establishing the similarity between *cornmeal* and *cornbread* would have been our best chance to validate this word. Currently, we are unable to do this.

6 Conclusions and Future Work

Our observations show that the intuitive idea behind the algorithm holds well. The algorithm is quite simple to implement. No interpretation of numbers is required; the process is just a simple test. The algorithm is heavily dependent on the depth and quality of dictionaries being used. WordNet has several words that were not present in conventional dictionaries available on the Web. Encyclopaedic entries such as *Mandara* (*a Chadic language spoken in the Mandara mountains in Cameroon*), domain-specific words, mainly from agriculture, medicine, and law, such as *ziziphus jujuba* (*spiny tree having dark red edible fruits*) and *pediculosis capitis* (*infestation of the scalp with lice*), phrasal words such as *caffiene intoxication* (sic) were among those not found in the collected dictionary.

Since the Princeton WordNet is manually crafted by a team of experts, we do not expect to find too many errors. However, many of the words present in the dictionary and not validated were those with rare meanings and usages. Our method makes it easier for human validators to focus on such words. This will especially be useful in validating the output of automatic wordnet creations.

The algorithm cannot yet detect omissions from a synset, i.e. the algorithm does not discover potential synonyms and compare them with the existing synset.

Possible future directions could be expanding the synset validation to other parts of a synset such as the gloss and relations to other synsets. The results could be summarized into a single number representing the quality of the synsets in the wordnet. The results could then be correlated with human evaluation, finally converging to a score that captures the human view of the wordnet.

The problem of scarcity of definitions could be further addressed by adding more dictionaries and references to the set of sources.

The presented algorithm is available only for English WordNet. However, the approach should broadly apply to other language wordnets as well. The limiting factors are the availability of dictionaries and tools like stemmers for those languages. Similarly, the algorithm could be used to verify synonym collections such as in Roget's Thesaurus and also other knowledge bases. The algorithm has been executed on noun synsets; they can also be run on synsets from other parts of speech.

We see such evaluation methods becoming increasingly imperative as more and more wordnets are created by automated means.

Acknowledgments. We would like to express our gratitude to Prof. Om Damani (CSE, IIT Bombay) and Sagar Ranadive for their valuable suggestions towards this work.

References

1. Miller, G., Beckwith, R., Felbaum C., Gross D., Miller, K.J.: Introduction to Wordnet: an online lexical database. The International Journal of Lexicography, vol. 3 (4), pp. 235-244 (1990)
2. Vossen, P.: EuroWordNet: a multilingual database with lexical semantic networks. Kluwer Academic Publishers (1998)
3. Narayan, D., Chakrabarty, D., Pande, P., and Bhattacharyya, P.: An Experience in building the Indo-Wordnet - A Wordnet for Hindi. In: First International Conference on Global Wordnet (GWC '02) (2002)
4. Liu, H., Singh, P.: Commonsense Reasoning in and over Natural Language. In: The proceedings of the 8th International Conference on Knowledge-Based Intelligent Information and Engineering Systems (2004)
5. Dong, Z., Dong, Q.:An Introduction to HowNet, available from http://www.keenage.com.
6. Richardson, S., Dolan, W., Vanderwende, L.: MindNet: acquiring and structuring semantic information from text. In: 36th Annual meeting of the Association for Computational Linguistics, vol. 2, pp. 1098-1102 (1998)
7. Kipper-Schuler, K.: VerbNet: A broad-coverage, comprehensive verb lexicon. Ph.D. dissertation, University of Pennsylvania (2005)
8. Baker, C. F. , Fillmore, C. J. , Lowe, J. B. : The Berkeley FrameNet project. In: Proceedings of the COLING-ACL (1998)
9. W. Group, "WordNet 2.1 database statistics" http://wordnet.princeton.edu/man/wnstats.7WN.
10. Smrz, P.: Quality Control for Wordnet Development. In: Proceedings of GWC-04, 2nd Global WordNet Conference (2004)
11. Cuadros M., Rigau G.: Quality Assessment of Large-Scale Knowledge Resources. In: Proceedings of Joint SIGDAT Conference on Empirical Methods in Natural Language Processing (EMNLP'06). Sydney, Australia (2006)
12. Cuadros M., Rigau G., Castillo M: Evaluating Large-scale Knowledge Resources across Languages. In: Proceedings of the International Conference on Recent Advances on Natural Language Processing (RANLP'07), Borovetz, Bulgaria (2007)
13. Guarino, N. : Toward a Formal Evaluation of Ontology Quality - (Why Evaluate Ontology Technologies? Because It Works!). IEEE Intelligent Systems, Vol. 19,, No. 4, pp. 74-81, (2004)
14. Brank, J., Grobelnik, M., Mladenic, D.: A survey of ontology evaluation techniques. In: Proceedings of the Conference on Data Mining and Data Warehouses (SiKDD 2005) (2005)
15. Maynard, D. , Peters, W., Li, Y.: Metrics for Evaluation of Ontology based Information Extraction. In: EON2006 at WWW (2006)
16. Hartman, J. , Spyns, P. , Giboin, A., et al.: D1.2.3 Methods for ontology evaluation. Deliverable for Knowledge Web Consortium (2005)
17. Guarino, N., Welty, C.: Evaluating ontological decisions with OntoClean. In: Communications of the ACM, vol. Volume 45, Number 2, pp. 61-65, (2002).
18. Porzel, R., Malaka, R.: A Task-based Approach for Ontology Evaluation. In: ECAI Workshop on Ontology Learning and Population (2004)

19. Velardi, P., et al.: Automatic Ontology Learning: Supporting a Per-Concept Evaluation by Domain Experts. In: Workshop on Ontology Learning and Population (OLP), in the 16th European Conference on Artificial Intelligence (2004)
20. Edmundson, H.P., Epstein, M.: Computer Aided Research on Synonymy and Antonymy. In: Proceedings of the International Conference on Computational Linguistics, (1969)
21. Sanchez , D., Moreno, A.: Automatic discovery of synonyms and lexicalizations from the Web. In: Proceedings of the 8th Catalan Conference on Artificial Intelligence (2005)
22. Turney, P.D.: Mining the Web for synonyms: PMI-IR versus LSA on TOEFL. In: Proceedings of the Twelfth European Conference on Machine Learning, Berlin: Springer Verlag, pp. 491-502 (2001)
23. Agirre, E., Ansa, O., Hovy, E., Martinez, D.: Enriching very large ontologies using the WWW. In: Proceedings of the Workshop on Ontology Construction of the European Conference of AI (ECAI-00) (2000)
24. Senellart, P. P., Blondel, V. D.: Automatic discovery of similar words. In: A Comprehensive Survey of Text Mining, M. W. Berry, Ed. Springer-Verlag (2003)
25. Dictionary.Com, Available at http://dictionary.reference.com/.
26. Porter's Stemmer, Available at http://www.tartarus.org/martin/PorterStemmer/ (as of 1 Jul 2005).