

Query Expansion with ConceptNet and WordNet: An Intrinsic Comparison

Ming-Hung Hsu, Ming-Feng Tsai, and Hsin-Hsi Chen*

Department of Computer Science and Information Engineering
National Taiwan University
Taipei, Taiwan
{mhhsu, mftsai}@nlg.csie.ntu.edu.tw,
hhchen@csie.ntu.edu.tw

Abstract. This paper compares the utilization of ConceptNet and WordNet in query expansion. Spreading activation selects candidate terms for query expansion from these two resources. Three measures including discrimination ability, concept diversity, and retrieval performance are used for comparisons. The topics and document collections in the ad hoc track of TREC-6, TREC-7 and TREC-8 are adopted in the experiments. The results show that ConceptNet and WordNet are complementary. Queries expanded with WordNet have higher discrimination ability. In contrast, queries expanded with ConceptNet have higher concept diversity. The performance of queries expanded by selecting the candidate terms from ConceptNet and WordNet outperforms that of queries without expansion, and queries expanded with a single resource.

1 Introduction

Query expansion has been widely used to deal with paraphrase problem in information retrieval. The expanded terms may come from feedback documents, target document collection, or outside knowledge resources [1]. WordNet [2], an electronic lexical database, has been employed to many applications [9], where query expansion is an important one. Voorhees [14] utilized lexical semantic relations in WordNet to expand queries. Smeaton et al [13] added WordNet synonyms of original query terms with half of their weights. Liu et al [7] used WordNet to disambiguate word senses of query terms, and then considered the synonyms, the hyponyms, and the words from definitions for possible additions to a query. Roberto and Paola [10] utilized WordNet to expand a query and suggested that a good expansion strategy is to add those words that often co-occur with the words of the query. To deal with short queries of web users, Moldovan and Mihalcea [8] applied WordNet to improve Internet searches.

In contrast to WordNet, commonsense knowledge was only explored in retrieval in a few papers. Liu and Lieberman [5] used ConceptNet [4] to expand query with the related concepts. However, the above work did not make formal evaluation, so that we were not sure if the effects of introducing common sense are positive or negative. Hsu and Chen [3] introduced commonsense knowledge into IR by expanding concepts in text descriptions of images with spatially related concepts. Experiments

* Corresponding author.

showed that their approach was more suitable for precision-oriented tasks and for “difficult” topics. The expansion of this work was done at document level instead of query level. Document contributes much larger contextual information than query.

In the past, few papers have touched on the comparison of ConceptNet and WordNet in query expansion under the same benchmark. We are interested in what effects of these two resources have. If we know which resource is more useful in a certain condition, we are able to improve the retrieval performance further. In this paper, we design some experiments with evaluation criteria to quantitatively measure WordNet and ConceptNet in the aspect of query expansion. We employ the same algorithm, *spreading activation* [12], to select candidate terms from ConceptNet and WordNet for the TREC topics 301-450, which were used in TREC-6, TREC-7 and TREC-8. To compare the *intrinsic characteristics* of these two resources, we propose three types of quantitative measurements including discrimination ability, concept diversity, and retrieval performance.

This paper is organized as follows. In Section 2, we give brief introduction to WordNet and ConceptNet. The comparison methodology is specified in Section 3. Section 4 introduces the experiment environment and discusses the experimental results. Section 5 concludes the remarks.

2 WordNet and ConceptNet

In this section, we give brief introduction to the two resources to be compared. Frameworks and origins of the two knowledgebase are described. A surface comparison of their similarities and differences is also presented.

2.1 WordNet

WordNet appeared in 1993 and has been developed by linguistic experts at Princeton University’s Cognitive Science Laboratory since 1985. It is a general-purpose knowledgebase of words, and it covers most English nouns, adjectives, verbs and adverbs. WordNet’s structure is a relational semantic network. Each node in the network is a lexical unit that consists of several synonyms, standing for a specific “sense”. Such lexical unit is called as ‘synset’ in WordNet terminology. Synsets in WordNet are linked by a small set of semantic relations such as ‘is-a’ hierarchical relations and ‘part-of’ relations. For its simple structure with words at nodes, WordNet’s success comes from its ease of use [2][9].

2.2 ConceptNet

ConceptNet is developed by MIT Media Laboratory and is presently the largest commonsense knowledgebase [6]. ConceptNet is a relational semantic network that is automatically generated from about 700,000 English sentences of the Open Mind Common Sense (OMCS) corpus. Nodes in ConceptNet are compound concepts in the form of natural language fragments (e.g. ‘food’, ‘buy food’, ‘grocery store’, and ‘at home’). Because the goal of developing ConceptNet is to cover pieces of commonsense knowledge to describe the real world, there are 20 kinds of relations categorized as causal, spatial, functional, etc. ConceptNet has been adopted in many interactive

applications [4]. Hsu and Chen [3] utilized ConceptNet to expand image annotations and got improvement for some difficult topics. As commonsense knowledge is deeply context-sensitive, the suitability of ConceptNet for query expansion is still not clear.

2.3 A Surface Comparison

WordNet and ConceptNet have several similarities: (1) their structures are both relational semantic networks; (2) both of them are general-purpose (that is, not domain-specific) knowledgebase; and (3) concepts in the two resources are both in the form of natural language. On the other hand, WordNet and ConceptNet differ from each other in some aspects: (1) as their processes of development differ (manually handcrafted vs. automatically generated), intuitively WordNet has higher quality and robustness; (2) while WordNet focuses on formal taxonomies of words, ConceptNet focuses on a richer set of semantic relations between compound concepts [6]; and (3) WordNet differentiates ambiguous meanings of a word as synsets, however, ConceptNet bears ambiguity of commonsense knowledge in its concepts and relations.

3 Comparison Methodologies

To compare the two knowledgebase in the aspect of query expansion intrinsically, we perform the same algorithm to expand queries. As WordNet and ConceptNet are both relational semantic networks, i.e., useful concepts for expansion in the network are usually those related to the concepts of the query, spreading activation [12] is adopted.

Figure 1 shows the overall procedure of the comparison. Given an original query, we perform spreading activation on WordNet and ConceptNet, respectively. Then, the two expanded queries are compared with three quantitative measurements. The first measurement computes the discrimination ability in information retrieval. The second measurement calculates the concept diversity in relevant documents. The third directly evaluates the performance of retrieval, including two typical evaluation criteria for ad hoc retrieval. All of these measurements are described in Section 3.4.

When we perform spreading activation in a semantic network to expand a query, the node of *activation origin* represents the concept of the given query. The activation origin is the first to be activated, with an initial activation score (e.g., 1.0). Next, nodes one link away from the activation origin are activated, then two links away, and so on.

Equation (1) shown below determines the activation score of node j by three factors: (i) a constant $C_{dd} \leq 1$ (e.g., 0.5), which is called *distance discount* that causes a node closer to the activation origin to get a higher activation score; (ii) the activation score of node i ; (iii) $W(i, j)$, the weight of the link from i to j . Different relations in the semantic network are of different weights. $Neighbor(j)$ represents the nodes connected to node j .

$$\text{Activation_score}(j) = C_{dd} \cdot \sum_{i \in Neighbor(j)} \text{Activation_score}(i) \cdot W(i, j) \quad (1)$$

Since most traditional IR systems are of bag-of-words model, we select the top N words with the higher activation scores as the expanded query. For a word w , its activation score is the sum of scores of the nodes (i.e., synsets in WordNet) that contain w .

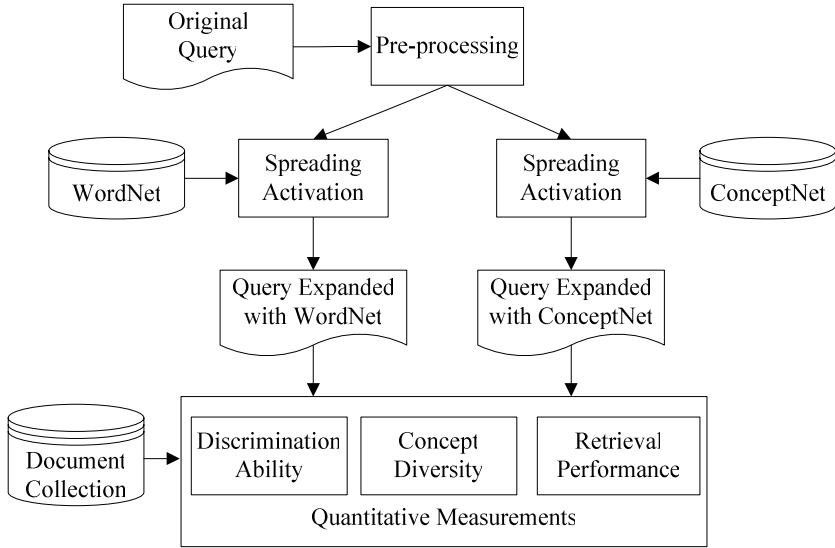


Fig. 1. Overall procedure in our approach

3.1 Pre-processing

Each query was lemmatized and POS-tagged by Brill tagger. Stop-words were removed and each of the remaining words with POS-tags was considered as a *concept* at the following stages.

3.2 Spreading Activation in ConceptNet

In addition to commonsense concepts and relations, ConceptNet also provides a set of tools for reasoning over text [6]. One of these tools, *get_context(concepts)*, performs spreading activation on all kinds of relations in ConceptNet, to find contextual neighborhood relevant to the concepts as parameters. Different relations are set to different weights in default setting. For example, the weight of ‘IsA’ relation is 0.9 and the weight of ‘DefinedAs’ is 1.0, etc. We adopt this tool directly for collecting expanded terms. In our experiments, each word in a compound concept has the same activation score as that of the compound concept. More details about the reasoning tools in ConceptNet please refer to [6].

3.3 Spreading Activation in WordNet

Since each node in WordNet is a synset that contains synonyms of certain sense, spreading activation in WordNet is surely performed on the unit of synset. Because ConceptNet covers most relations in WordNet, we determine the weights of relations in WordNet, shown in Table 1, by referring to the settings in ConceptNet. For each concept (a word with POS-tag) in the query, we choose its most frequent sense (synset) as the activation origin from the corresponding POS. In other words, we do not disambiguate the sense of query terms in this paper for simplicity.

Table 1. Relations in WordNet and their weights for spreading activation

Relation Type	causes	holonyms	hypernyms	hyponyms	meronyms	Pertainyms
Weight	0.8	0.8	0.8	0.9	0.9	0.7

3.4 Quantitative Measurements

The following proposes three types of measurements to investigate the intrinsic differences between WordNet and Concept. They provide different viewpoints for the comparison.

(1) **Discrimination Ability (DA).** Discrimination ability is used to measure how precisely a query describes the information need. In IR, the inverse document frequency (IDF) of a term denotes if it occurs frequently in individual documents but rarely in the remainder of the collection. For IR systems as the state of the art, discrimination ability of a term can be estimated by its IDF value. Hence the discrimination ability of a query is measured and contributed by the IDFs of the terms in the query. For a query q composed of n query terms (q_1, q_2, \dots, q_n) , we define its discrimination ability (DA) as follows.

$$DA(q) = \frac{1}{n} \sum_{i=1}^n \log\left(\frac{N_C}{df(q_i)}\right) \quad (2)$$

where N_C is the number of documents in a collection, and $df(q_i)$ is the document frequency of query term q_i .

(2) **Concept Diversity (CD).** This measurement helps us observe the concept diversity of an expanded query, relative to the relevant documents. That is, we measure how much an expanded query covers the concepts occurring in the relevant documents. Let $tm(\cdot)$ denote the function that maps the parameter (a document or a query) to the set of its index terms. Let $\{d_{q(1)}, d_{q(2)}, \dots, d_{q(m)}\}$ denote the set of m documents which are relevant to the query q in the collection. The concept diversity (CD) of a query q is defined as follows.

$$CD(q) = \frac{1}{m} \sum_{i=1}^m \frac{|tm(q) \cap tm(d_{q(i)})|}{|tm(d_{q(i)})|} \quad (3)$$

(3) **Retrieval Performance:** This type of measurements includes two typical evaluation criteria for ad hoc retrieval, i.e., average precision (AP) and precision at top 20 documents (P@20).

4 Experiments and Discussion

We adopted the topics and the document collections in the ad hoc track of TREC-6, TREC-7 and TREC-8 as the experimental materials for the comparison. There are 556,077 documents in the collection of TREC-6 and 528,155 documents in TREC-7 and in TREC-8. Only the “title” part was used to simulate short query, since web

users often submit short queries to search engines. There are totally 150 topics with identifiers 301-450. However, 4 of them (i.e., topics 312, 379, 392 and 423) are unable to be expanded by spreading activation either in WordNet or in ConceptNet, so that these 4 topics are neglected in the experiments. For each short query, the top 100 words with the higher activation scores form the expanded query. The IR system adopted for the measurement of retrieval performance is Okapi's BM25 [11]. The retrieval performance is measured on the top 1000 documents for each topic.

Figures 2, 3, 4, and 5 show the results of the quantitative measurements, where the x-axis represents topic number, and 301-350, 351-400, and 401-450 are topics of TREC-6, TREC-7 and TREC-8, respectively. To compare the differences between WordNet and ConceptNet, the result presented for each topic is the difference between the two expanded queries, i.e., the measurement of the WordNet-expanded query subtracts that of the ConceptNet-expanded query.

4.1 Preliminary Analysis

Figure 2 shows the differences of two kinds of expansions in discrimination ability (DA). The DA averaged over the 146 experimental topics is 5.676 and 4.191 for the WordNet-expanded and ConceptNet-expanded queries, respectively. From Figure 2, it is obvious that the terms in the WordNet-expanded queries have higher discrimination ability. In other words, they are more specific than those terms in ConceptNet-expanded queries. A specific term highly relevant to a topic can be considered as one of the *kernel words* of that topic. Figure 2 shows that the queries expanded by WordNet are more probable to contain the kernel words.

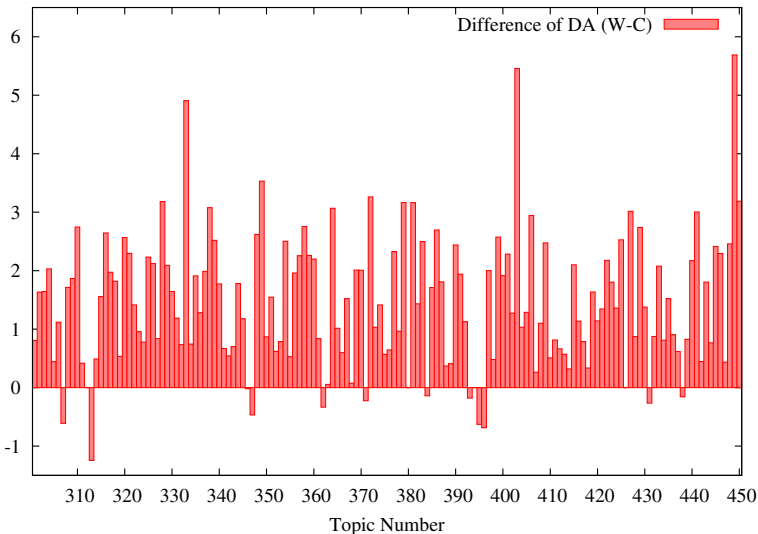


Fig. 2. Differences between WordNet and ConceptNet in discrimination ability (DA)

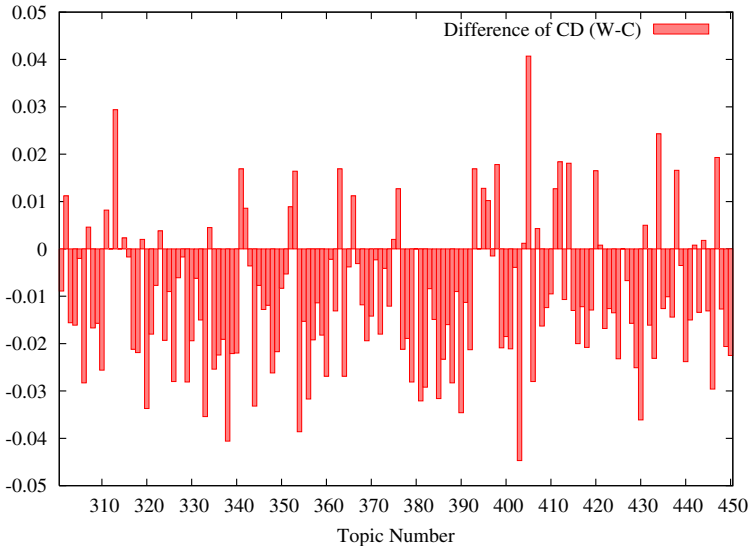


Fig. 3. Differences between WordNet and ConceptNet in concept diversity (CD)

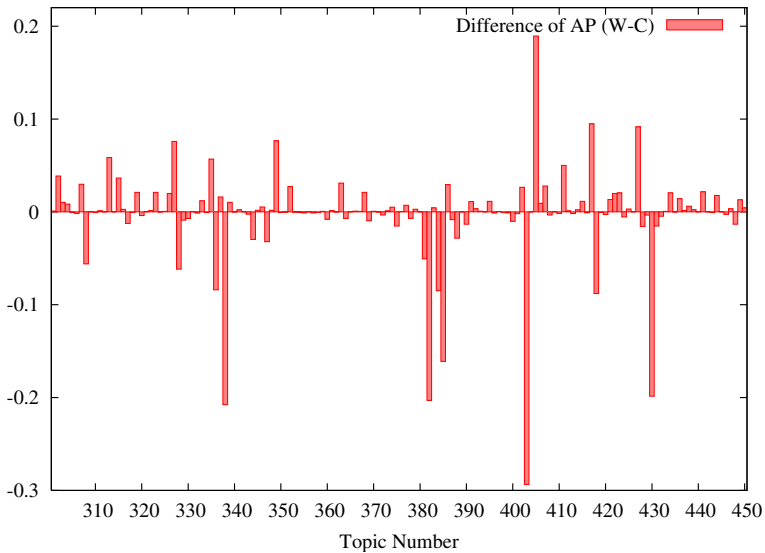


Fig. 4. Topic-by-topic differences of retrieval performance in average precision (AP)

The average concept diversity (CD) is 0.037 and 0.048 for the WordNet-expanded and the ConceptNet-expanded queries, respectively. In contrast to the result of discrimination ability, Figure 3 shows that ConceptNet-expanded queries have higher

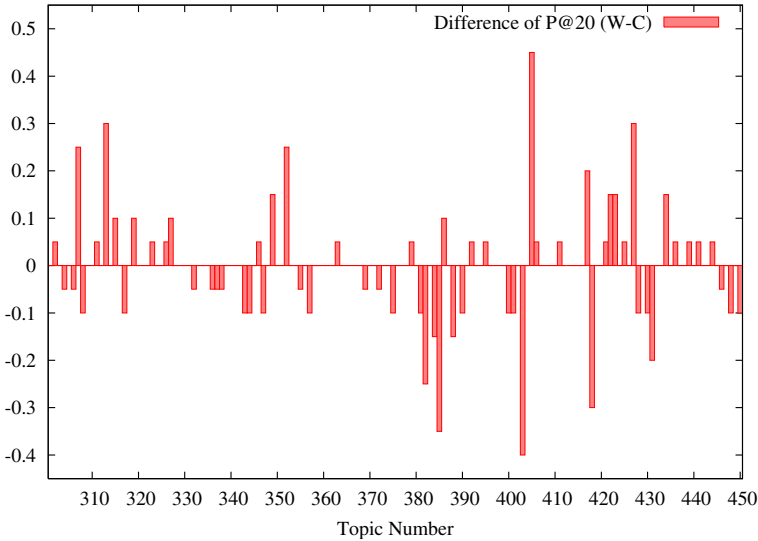


Fig. 5. Topic-by-topic differences of retrieval performance in P@20

concept diversity than WordNet-expanded ones do. Note that the concept diversity of an expanded query is computed according to the relevant documents. As the terms in ConceptNet-expanded queries are usually more general than those in WordNet-expanded queries, Figure 3 shows that ConceptNet-expanded queries cover more of the concepts that would usually co-occur with the kernel words in the relevant documents. We call the concepts that co-occur with the kernel words as *cooperative concepts*.

Expanding a short query with the kernel words will help IR systems to find more relevant documents. On the other hand, co-occurrence of the cooperative concepts and the kernel words will help the IR system to rank truly relevant documents higher than those containing noise. Here we take topic 335 (“adoptive biological parents”) as an example to illustrate this idea. The kernel words of this topic may be “pregnancy”, “surrogate”, etc, and the cooperative concepts may be “child”, “bear”, etc. Of these, “surrogate” and “child” are suggested by WordNet and ConceptNet, respectively. The pair (child, surrogate) has stronger collocation in the relevant documents than in the irrelevant documents. The detail will be discussed in Section 4.2.

The overall retrieval performances of the WordNet-expanded queries in AP and P@20 are 0.016 and 0.0425, respectively. For the ConceptNet-expanded queries, the performances are 0.019 and 0.0438, in AP and P@20, respectively. These retrieval performances are low because the expanded queries are formed by the top 100 words with the higher activation scores. The simple expansion method introduces too much noise. Figure 4 and Figure 5 show the differences of AP and of P@20 for each topic. We observed that WordNet-expanded queries perform better for some topics, but ConceptNet-expanded queries perform better for some other topics. While WordNet and ConceptNet are different in discrimination ability (Figure 2) and in concept diversity (Figure 3), the two resources can complement each other in the task of ad hoc retrieval. Hence, we made further experiments in the following subsection.

4.2 Further Analysis

In the next experiments, we performed manual query expansion by selecting some of the top 100 words proposed by spreading activation in WordNet or in ConceptNet, to expand the original query. Two research assistants, each of whom dealt with half of the topics of TREC-6, performed the process of manual expansion. They read the topic description in advance, and he/she had no idea about the real content or vocabulary in the relevant documents. This manual selection process was performed separately on the words proposed by WordNet and by ConceptNet. These manually selected words for expansion are called WordNet-consulted (WC) and ConceptNet-consulted (CC) terms, respectively. In this way, we compared four expansion strategies, i.e., original (no expansion), WordNet-consulted, ConceptNet-consulted, and combination of WC and CC. We also increased the weights of the original query terms with different degrees to observe how the performances vary with the degrees. In the experiments, we only used the topics of TREC-6 for analyses.

Figure 6 shows the performances in mean average precision of the four expansion strategies. The x-axis represents the degrees (times) by which the weights of the original query terms are increased. The performance of the original query is 0.221 and doesn't vary with the degrees since there is no expansion. CC slightly performs better than the original when the degree is larger than 3, as well as WC with the degree larger than 6. The slight improvement of CC only or WC only shows that without information about the real content of the relevant documents, effective query expansion for IR is a challenging task even for humans. The best performance is 0.2297, which is obtained with combination of WC and CC at degree 6, and 3.94% increase to the

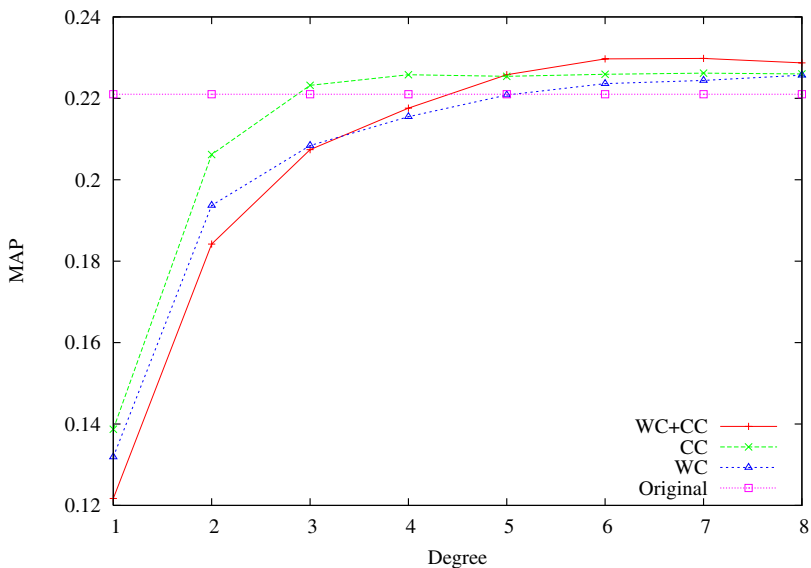


Fig. 6. Performances of the four expansion strategies vs. the weights of the original query term

baseline. This performance improvement on no expansion is examined as significant by a t-test with a confidence level of 95%. The corresponding p-value is 0.034.

Figure 6 also shows that a careful weighting scheme is needed no matter whether WordNet only or ConceptNet only is adopted. With an unsuitable weighting scheme, combination of WC and CC performs even worse than WC only or CC only. In Figure 6, CC performs stably when the degree increases larger than 3, but WC performs stably only after the degree is larger than 6. As the degree stands for how much the weights of original query terms are increased, in the aspect of ranking documents, the degree also stands for how much the weights of CC or WC terms are lightened. While the words proposed by WordNet are usually more specific and influence more heavily on the rankings of retrieved documents, it is shown that an appropriate weighting scheme is more important for WC than for CC.

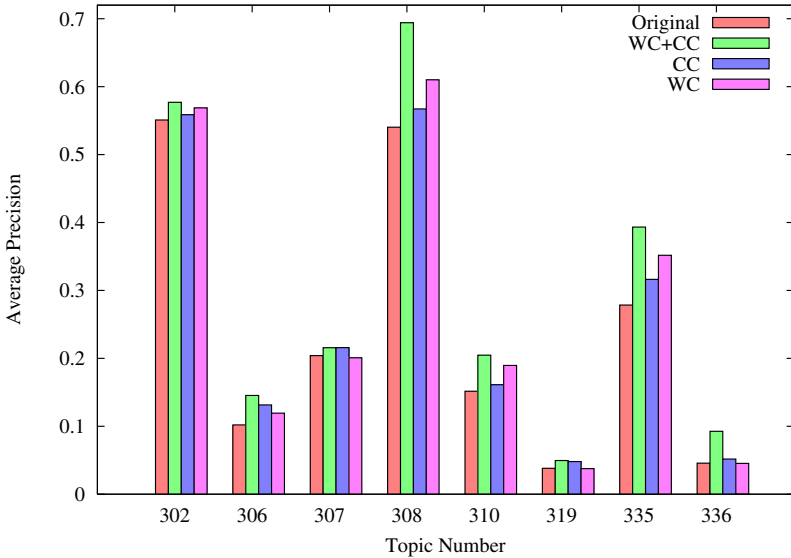


Fig. 7. The performances of 8 topics using different expansion strategies

While Figure 6 shows the result averaged over all topics of TREC-6, Figure 7 shows some more strong evidences supporting the argument that WordNet and ConceptNet can complement each other. Using different expansion strategies with the degree 6, the performances (AP) of eight topics are presented. While CC only or WC only improve the performance of each of the eight topics, it’s obvious that all the eight topics benefit from the combination of WC and CC. Therefore, the overall improvement (refer to Figure 6) of combination of WC and CC is mostly exhibited in the eight topics.

We verify the complementary of WordNet and ConceptNet, i.e., frequent co-occurrence of WC terms and CC terms in relevant documents, by the following way. For each pair of CC term t_c and WC term t_w , we calculate $LRP(t_c, t_w)$ using Equation

(4). $LRP(t_c, t_w)$ is a value of logarithm of the ratio of two conditional probabilities: (1) the co-occurrence probability of t_c and t_w in the relevant documents; and (2) the co-occurrence probability of t_c and t_w in the irrelevant documents.

$$LRP(t_c, t_w) = \log \frac{P(t_c, t_w | R)}{P(t_c, t_w | IRR)} \quad (4)$$

where R and IRR represent relevant and irrelevant documents, respectively.

Table 2 shows the title, the CC terms, the WC terms and the term pairs having high LRP values for each topic in Figure 7. The term pairs with high LRP in the eight topics are the major evidences to support that combination of WC and CC is effective as shown in Figures 6 and 7. Note that the pairs (child, surrogate), (life, surrogate) and (human, surrogate) of topic 335 have high LRP values. They also confirm the idea of kernel words and cooperative concepts mentioned in Section 4.1.

Table 2. Illustration of the complementarity of WordNet and ConceptNet in 8 topics

Topic	Title of Topic	CC (t_c)	WC (t_w)	(t_c, t_w): LRP(t_c, t_w)
302	poliomyelitis and post-polio	affection flame global	disease place paralysis	(global, paralysis): 5.967 (global, disease): 5.043 (affection, paralysis): 4.983
306	african civilian deaths	war kill nation	megadeath event casualty killing	(war, killing): 3.715 (kill, casualty): 3.598 (nation, killing): 3.408 (nation, casualty): 3.357
307	new hydroelectric projects	state exploration station	proposition risky examination	(station, risky): 3.906 (station, examination): 3.598 (exploration, examination): 3.428
308	implant dentistry	medical reproductive function	artificial prosthesis specialty	(medical, prosthesis): 9.400 (function, prosthesis): 9.009 (medical, artificial): 6.980
310	radio waves and brain cancer	cause carcinogen state produce	corpus radiation	(cause, radiation): 5.234 (produce, radiation): 3.958 (state, radiation): 3.081
319	new fuel sources	find energy natural	material head	(energy, material): 3.744 (energy, head): 3.217 (natural, material): 3.071
335	adoptive biological parents	child human life	surrogate married kinship	(child, surrogate): 8.241 (life, surrogate): 6.840 (human, surrogate): 5.867
336	black bear attacks	animal claw battle	fight strike counterattack	(claw, fight): 6.255 (animal, fight): 3.916

5 Conclusions and Future Works

In this paper, we used the technique of spreading activation to investigate the intrinsic characteristics of WordNet and of ConceptNet. Three types of quantitative

measurements, i.e., discrimination ability, concept diversity, and retrieval performance are used to compare the differences between the two resources. With the preliminary analysis and the verification of manual expansion, we have shown that WordNet is good at proposing kernel words and ConceptNet is useful to find cooperative concepts. With an appropriate weighting scheme, the two resources can complement each other to improve IR performance.

In future work, we will investigate an automatic query expansion method to combine the advantages of the two resources. Commonsense knowledge in ConceptNet is deeply context-sensitive so that it needs enough context information for automatic query expansion. Using existing methods such as pseudo relevance feedback or resources such as WordNet to increase the context information can be explored. We also intend to investigate whether complex techniques of word sense disambiguation (WSD) in WordNet are necessary for IR, under the existence of ConceptNet.

Acknowledgements

Research of this paper was partially supported by National Science Council, Taiwan, under the contracts NSC94-2752-E-001-001-PAE and NSC95-2752-E-001-001-PAE.

References

1. Baeza-Yates, Ricardo and Ribeiro-Neto, Berthier: *Modern Information Retrieval*, Addison-Wesley (1999).
2. Fellbaum, C. (ed.). *WordNet: An Electronic Lexical Database*. MIT Press (1998).
3. Hsu, Ming-Hung and Chen, Hsin-Hsi: *Information Retrieval with Commonsense Knowledge*. In: *Proceedings of 29th ACM SIGIR International Conference on Research and Development in Information Retrieval* (2006).
4. Lieberman, H., Liu, H., Singh, P., and Barry, B. *Beating Common Sense into Interactive Applications*. *AI Magazine* 25(4) (2004) 63-76.
5. Liu, H. and Lieberman, H: *Robust Photo Retrieval Using World Semantics*. In: *Proceedings of LREC2002 Workshop: Using Semantics for IR* (2002) 15-20.
6. Liu, H. and Singh, P.: *ConceptNet: A Practical Commonsense Reasoning Toolkit*. *BT Technology Journal* 22(4) (2004) 211-226.
7. Liu, S., Liu, F., Yu, C.T., and Meng, W.: *An Effective Approach to Document Retrieval via Utilizing WordNet and Recognizing Phrases*. In: *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (2004) 266-272.
8. Moldovan, D.I. and Mihalcea, R.: *Using WordNet and Lexical Operators to Improve Internet Searches*. *IEEE Internet Computing* 4(1) (2000) 34-43.
9. Morato, Jorge, Marzal, Miguel Ángel, Lloréns, Juan and Moreira, José.: *WordNet Applications*. In: *Proceedings of International Conference of the Global WordNet Association* (2004) 270-278.
10. Navigli, Roberto and Velardi, Paola: *An Analysis of Ontology-based Query Expansion Strategies*. In: *Proceedings of Workshop on Adaptive Text Extraction and Mining at the 14th European Conference on Machine Learning* (2003).

11. Robertson, S.E., Walker, S. and Beaulieu, M.: Okapi at TREC-7: Automatic Ad Hoc, Filtering, VLC and Interactive. In: Proceedings of the Seventh Text Retrieval Conference (1998) 253-264.
12. Salton, G. and Buckley, C.: On the Use of Spreading Activation Methods in Automatic Information Retrieval. In: Proceedings of the 11th CM-SIGIR Conference on Research and Development in Information Retrieval (1988) 147-160.
13. Smeaton, Alan F., Kelledy, Fergus, and O'Donnell, Ruari: TREC-4 Experiments at Dublin City University: Thresholding Posting Lists, Query Expansion with WordNet and POS Tagging of Spanish. In: Proceedings of TREC-4 (1994) 373-390.
14. Voorhees, Ellen M.: Query Expansion Using Lexical-Semantic Relations. In: Proceedings of the 17th Annual ACM SIGIR Conference on Research and Development in Information Retrieval (1994) 61-69.