An Improved Corpus Comparison Approach to Domain Specific Term Recognition[*]

Xiaoyue Liu, and Chunyu Kit

Department of Chinese, Translation and Linguistics
City University of Hong Kong
Tat Chee Ave., Kowloon, Hong Kong
{xyliu0, ctckit}@cityu.edu.hk

Abstract. Domain specific terms are words carrying special conceptual meanings in a subject field. Automatic term recognition plays an important role in many natural language processing and knowledge engineering applications such as information retrieval and knowledge mining. This paper explores a novel approach to automatic term extraction based on the basic ideas of corpus comparison and emerging pattern with significant elaboration. It measures the termhood of a term candidate in terms of its peculiarity to a given domain via comparison to several background domains. Our experiments confirm its outperformance against other approaches, achieving an average precision of 83% on the top 10% candidates in terms of their termhood.

Keywords: Automatic term recognition, corpus comparison, emerging pattern

1. Introduction

Automatic term recognition (ATR) plays an important role in many natural language processing applications, e.g., information retrieval (IR) (Chowdhury, 1999; Zhou and Nie, 2005), information extraction (Yangarber et al., 2000), domain specific lexicon construction (Hull, 2001), and topic extraction (Lin, 2004). Its technological advancements can facilitate all these applications for performance enhancement.

Despite the large volume of literature on ATR, further significant success in the field still relies heavily on a sound resolution of two basic issues, namely, the unithood and termhood of a term candidate, as identified in Kageura and Umino (1996). The former quantifies the unity of a candidate (especially, a multi-word candidate), indicating how likely a candidate is to be an atomic linguistic unit. It works more like a filter to exclude non-atomic (thus unqualified) term candidates but has little authority to determine which atomic language unit is a true term. The latter measures how likely a qualified candidate is to be a true term in a subject field. It plays a decisive role in licensing a term.

Regardless of the previous progress in term extraction, termhood measurement remains the most critical problem to be solved. Novel technologies and methodologies are needed in order to bring up new insights into our understanding of this problem. This paper is intended to pre-

[*] The research described in this paper was supported by the Research Grants Council of HKSAR, China, through the CERG grant 9040861 (CityU 1318/03H) and by the City University of Hong Kong through the Strategic Research Grants 7002037 and 7001879. Correspondence concerning this research should be addressed to Dr. Chunyu Kit.

sent a novel statistical approach to domain specific term extraction from a collection of thematic documents following the basic ideas of corpus comparison and emerging pattern. It measures the termhood of a term candidate in a subject domain in terms of its peculiarity to this subject domain via comparison to several background domains.

In order to avoid unexpected interference from unithood issues, our study focuses on investigating the termhood measurement for mono-word terms. Nevertheless, this is by no means to imply that mono-word ATR can be any easier than multi-word ATR in any sense. It is pointed out in Daille (1994) that the automatic identification of mono-word terms is possibly more complex than that of multi-word ones. One of the reasons is that all structural information that can be utilized for multi-word ATR is not available for mono-word ATR. In this sense, the latter needs to tackle a fundamental issue, that is, how to differentiate between terms and non-terms without resorting to structure information.

The rest of the paper is organized as follows. Section 2 presents a brief review of previous work on ATR, to give a background for our research. Section 3 formulates our approach and the working procedure involved. A series of experiments are then reported in Section 4 for the purpose of evaluation. Section 5 concludes the paper with a highlight on the advantages of our approach.

2. Previous Work

Various approaches to ATR were developed in the past. From a methodological point of view, the existing approaches can be classified into the following categories.

Linguistic A linguistic approach played a dominant role in the early research on ATR. As early as twenty years ago, Ananiadou (1988) studied the effectiveness of theoretically motivated linguistic knowledge (e.g. morphology) in term recognition. A common procedure involved in this kind of approach is to carry out part-of-speech tagging first and then some pre-defined syntactic patterns, e.g., noun-noun compounds (Dagan and Church, 1994; Wu and Hsu, 2002) and base noun phrases (Justeson and Katz, 1995), can be applied to identify term candidates. All word combinations that match none of the predefined patterns are filtered out. This approach was reported to achieve good results on small scale corpora. However, its disadvantages include inadequate coverage of pre-defined syntactic patterns, low transplantability to other domains or languages, and incapability of excluding non-term candidates consistent with the pre-defined patterns.

Statistical Various kinds of statistical information can be utilized to support term extraction, e.g., frequency (Damerau, 1990), mutual information (Damerau, 1993), C-value (Frantzi and Ananiadou, 1996), NC-value (Frantzi et al., 1998), *imp* function (Nakagawa, 2001), KFIDF measure (Xu et al., 2002), standard deviation (Lin, 2004) and entropy (Chang, 2005), to name but a few. A multi-word term is assumed to carry a key concept and is thus expected to behave like an atomic text unit. Many of these statistical measures are applied to explore such unity or structural stability of a multi-word candidate, namely, its unithood. Besides, a bootstrapping approach is reported in Chen et al. (2003) to learn domain specific terms from unannotated texts on a subject. Wermter and Hahn (2005) identify multi-word terms among n-grams of words in a large biomedical corpus, measuring their termhood in terms of their paradigmatic modifiability. Although statistical approaches share an advantage, i.e., their language independency, they are far from reliable while working on small corpora.

Hybrid Linguistic knowledge is used in conjunction with statistical information in most hybrid approaches to ATR. For example, some syntactic patterns are first applied to identify term candidates, by filtering out those unqualified ones, and then a statistical measure is applied to validate the true terms among them. Daille (1994) presents an integrated approach to ATR that works this way. In addition, a hybrid approach can also be applied to combine several independent term recognizers for a better performance than any of them alone, as reported in Vivaldi et al. (2001).

Corpus Comparison This approach is a popular direction in recent ATR research. Its basic idea is to utilize the distinct distributions of terms and non-terms in different corpora to facilitate term extraction. That is, true terms are more prominent in their own subject field than in others. The original idea of this approach can be traced back to Yang (1986) that attempts to identify scientific terms by their statistical distributional difference between science and general texts. The statistics in use for this purpose include document frequency, average frequency, relative standard-deviation, etc. Ahmad et al. (1994) quantify similar contrasting distributions of terms in different corpora by means of the ratio of relative frequencies of a word in a domain corpus and a background corpus. The words with a score larger than 1.0 are then identified as the most potential terms. Chung (2003) applies a similar scheme called normalized frequency ratio to extract single-word terms in anatomy, reporting a performance of about 86% overlap with the results from a manual rating approach. In Uchimoto et al. (2001), more statistical characteristics (e.g., term frequency, document frequency and field frequency) of a term candidate are explored, achieving an F-score of 58.49%. Kit and Liu (2007) propose to measure mono-word termhood in terms of a candidate's rank difference in a domain and a background corpus.

The approach of corpus comparison can also be accomplished by statistical tests based on the null hypothesis that there is no difference between the observed frequencies for the same word in different corpora. Words with large testing values indicate a statistically significant difference between corpora and hence are more likely to be terms. The statistical tests include log-likelihood ratio (Rayson and Garside 2000), χ^2-test, Mann-Whitney ranks test and t-test (Kilgarriff 2001). Drouin (2003) bases a term extraction process on a statistical test, which uses a normal distribution as an approximation to words' binomial distribution, obtaining an overall precision of 81% on term recognition. Based on Drouin's work, Lemay et al. (2005) examine various corpus comparison approaches in different terminological settings. Among them, one is to use a general corpus as background and another is to break down the specialized corpus into six topical sub-corpora for comparison to the entire specialized corpus.

Methodologically, a corpus comparison approach takes advantage of the intrinsic statistical characteristics of true terms in different corpora and thus has a preferable theoretical grounding over others that utilize only a special domain corpus. Our research reported in this paper falls into the category of corpus comparison approach with necessary elaboration for further enhancement.

3. Term Extraction

Terms are linguistic representations of domain specific concepts to encode our special knowledge about a subject field. Emerging pattern (EP) (Dong and Li, 1999) presents a similar idea to corpus comparison in the field of database for knowledge discovery. EPs are defined as itemsets whose growth rates, i.e., the ratios of their supports[1] in one dataset over those in another, are larger than a predefined threshold. When applied to datasets with classes (e.g., cancerous vs. normal tissues, poisonous vs. edible mushrooms), EPs can capture significant differences or useful contrasts between the classes in terms of their growth rates. In principle, the larger the growth rates, the more significant the patterns. This approach has been successfully deployed in several applications of data mining, e.g., Li and Wong (2002) on identification of good diagnostic gene groups from gene expression profiles.

While the EP approach works on well-structured databases, corpus comparison deals with unstructured texts. Following the essential principle shared by the two, we consider domain specific term recognition from a thematic corpus of documents an issue of identifying words and expressions as EPs in the form of string highly peculiar to their own subject fields than to any others. In this sense, the higher peculiarity of a term candidate to a particular domain but lower to the others, the more likely it is to be a true term in that domain.

Accordingly, we opt to quantify the peculiarity of a term candidate to a subject field in terms of its emerging difference, which is to be scored according to statistical information such as fre-

[1] In database studies, the term *support* refers to the frequency of an itemset in a dataset.

quency difference in its own subject field and another field as background. This illustrates the basic idea of corpus comparison. To our knowledge, however, very few existing approaches of corpus comparison to ATR use more than one background corpus. When comparing a thematic corpus with multiple background corpora, we need to find an appropriate way to sum up the comparison results into the final termhood scores for the term candidates in question. This would pay off if such comparison and summing up can make the termhood scoring more reliable. We will follow this idea to derive the termhood for a term candidate in a target subject field.

Given a collection of *n* documents (or corpora) each representing a subject field, henceforth referred to as thematic documents, we follow the following working procedure to extract mono-word terms from each corpus.

1. Extraction: Extract all mono-words as term candidates, including those appearing only once, via stop word filtering, and then assign to each of them a weight in terms of the statistical measure in use (e.g., frequency, or *tf*-*idf* score).
2. Normalization: Normalize a weight in each subject field according to the sum of all weights in that field.
3. Computing emerging difference: For a candidate *w* in a target subject field *i*, calculate its emerging difference d_{ij} via comparison with another subject field *j* as

$$d_{ij}(w) = s_i(w) - s_j(w) \qquad (1)$$

   where $s_i(w)$ and $s_j(w)$ are *w*'s normalized weights in fields *i* and *j* ($0< i, j \leqslant n$), respectively. Consequently each candidate in each field will have *n*-1 emerging difference scores corresponding to the *n*-1 background fields involved in the comparison.
4. Ranking: Rank all candidates in each field *i* in terms of their emerging differences $d_{ij}(\cdot)$, resulting in *n*-1 ranking lists for each *i* accordingly to *n*-1 background corpora in use. In each ranking list, candidate *w* has a rank $r_{ij}(w)$ corresponding to its score $d_{ij}(w)$.
5. Sorting: Sort all candidates in each field *i* in terms of their termhood defined as

$$\tau_i(w) = \sum_{j=1}^{n} r_{ij}(w) \qquad (2)$$

   where $j \neq i$.
6. Evaluation: Examine the sorted list for each subject field to check how true terms are pushed to the top of the list by their termhood.

In our approach, a subject corpus is compared with more than one background corpus and all comparison results for each candidate are summed up together to represent its termhood (or peculiarity) in a subject field in question. However, a term candidate may get a negative score from (1) above, which will, unreasonably, weaken the total sum of scores from the comparison to all other background corpora. This certainly would not help to achieve the right ranking outcomes in an ATR output list sorted by the termhood scores so resulted. To alleviate this problem, we opt to rank term candidates first according to their emerging differences before the summing-up is conducted. The ranking in this way keeps the relative position (or relationship) of term candidates in the candidate list while avoiding the unexpected problem. Finally, the termhood of a term candidate is measured by the sum of its ranks that are derived from the comparison to a number of other domain corpora.

4. Evaluation

4.1.Data

A number of experiments following the above working procedure are carried out on the BLIS corpus (Kit et al., 2005) to extract legal terms in HK laws. The corpus consists of all ordinances and subsidiary legislation to prescribe laws and regulations involving almost every aspect of livings in HK. Excluding those repealed, ceased, expired, or not adopted ordinances, there are a total of 503 chapters in the current version of the corpus. Each ordinance can be regarded as a sub-corpus for a “field”, e.g., “Public Finance Ordinance”, “Forest and Countryside Ordinance”, “Hospital Authority Ordinance”, and “Marriage Ordinance”. We use the English texts from the corpus as data for our experiments.

In order to get reliable statistical information about individual words during the extraction procedure, a series of preprocessing steps are carried out, including word tokenization, lemmatization, and filtering of stop words (including words with only digits or without letters, mono-character words and function words). The evaluation of our approach focuses on the precision of ATR output, for our main concern is how to capture as many true terms as possible at the high end of a candidate list sorted by termhood.

Table 1: Distribution of precision score

Range of precision	$1 \geq p \geq 0.9$	$0.9 > p \geq 0.8$	$0.8 > p \geq 0.7$
Number of chapters	6	16	8

Table 2: Samples of ATR output

Chapter	Terms in the output
Probate and Administration Ordinance	executor, probate, administrator, caveat, testator, *box*, decease, affidavit, estate, renter
Import and Export Ordinance	export, import, manifest, cargo, article, transhipment, trader, smuggle, validate, customs
Theft Ordinance	steal, deception, burglary, theft, deceit, inducement, indictment, menace, robbery, cheat
Adoption Ordinance	adoption, infant, *litem*, accredit, adopt, guardian, parent, adoptive, wedlock, applicant

Table 3: Performance vs. corpus size

Precision	Size	Precision	Size	Precision	Size
1.0000	15	0.8684	38	0.8052	77
0.9655	29	0.8649	148	0.8000	45
0.9474	19	0.8594	64	0.7980	99
0.9320	103	0.8571	42	0.7826	92
0.9048	21	0.8293	82	0.7805	41
0.9000	30	0.8250	40	0.7798	109
0.8889	27	0.8228	79	0.7714	35
0.8727	55	0.8167	60	0.7419	93
0.8718	117	0.8078	26	0.7391	46
0.8710	31	0.8065	31	0.7246	69

4.2.Weighting with *tf-idf*

Applying the ATR procedure given in Section 3 above to the BLIS corpus with *tf-idf* (Salton, 1992) scoring for the initial weighting in the first step, in total we get 503 output lists sorted by termhood. Among them 30 lists are randomly picked for evaluation. For each selected list, the top 10% candidates are manually checked. Chang (2005) presents a modified *tf-idf* model based on inter-domain entropy calculation, giving an average precision of 65.4% on the top 10% of ATR output. It is assumed that around 10% words in each sub-domain are domain specific terms. The exact percentage of true terms among words, however, may vary significantly in different domains. Table 1 presents a distribution of precision scores over the 30 BLIS chapters manually evaluated, each of which corresponds to an ATR output list. The precision score varies from above 70% to 100%, giving an average of 82.98%. As a whole, twenty-two chapters, i.e., 73% of the evaluated lists, have a precision greater than 80%. A few output samples are presented in Table 2, illustrating the top 10 term candidates in the ATR output for four BLIS chapters, with non-terms highlighted in italic font.

An illustration of the performance of the *tf-idf* scoring vs. corpus size is presented in Table 3, resulted from the experiments on the same 30 chapters as above. The number of recognized terms (i.e., 10% of the total number of word types in a corpus) in those chapters varies from 15 to 148. The best performance, a precision of 100%, is achieved on a small chapter of only 15 candidates, whereas a precision of 93% is on a chapter of 103 candidates. This seems to suggest that our approach is able to achieve an excellent performance even on a very small corpus.

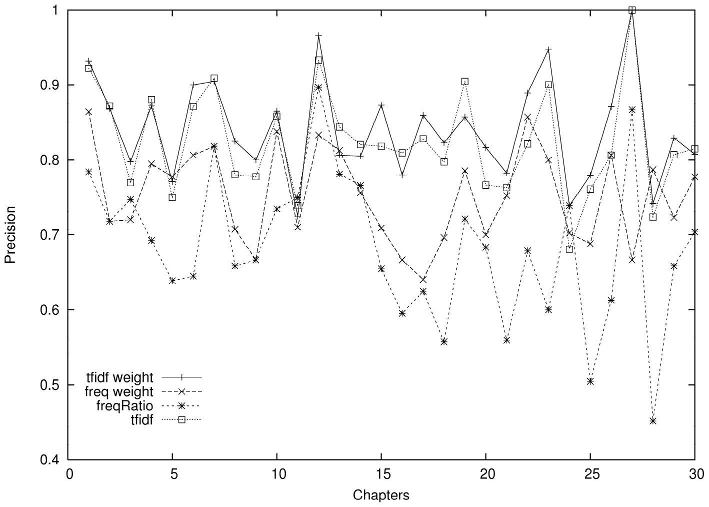

Figure 1: Performance comparison

Table 4: The average precision

Approach	*tf-idf* as weight	freq as weight	freq Ratio	*tf-idf* as termhood
Average precision	82.98%	75.34%	66.93%	81.46%

4.3.Comparison with Other Approaches

On the same data set, several other scoring schemes have also been tested for a comparison with the *tf-idf* scoring above. The comparison is carried out on the same 30 chapters.

1. Frequency for initial scoring: Use frequency as the initial scoring measure, and then follow the above working procedures in Section 3 to derive a term list for each sub-corpus.
2. Corpus comparison using frequency ratio: For each mono-word term candidate, compute the ratio of its frequencies in its own subject sub-corpus and a general background corpus.[2] Candidates with a ratio greater than 1.0 are recognized as true terms.
3. Using *tf-idf* as termhood: Sort the term candidates in a sub-corpus (i.e., an ordinance as a document in our case) in terms of their *tf-idf* scores as given by (3) below, where *freq*(*w*) is the frequency of a candidate *w* in the sub-corpus, *N* the total number of documents in the entire corpus in use and *d*(*w*) the number of documents containing *w*.

$$tf\text{–}idf(w) = freq(w) \cdot \log \frac{N}{d(w)} \tag{3}$$

[2] British National Corpus is used as the background corpus here for comparison. See its official site at http://www.natcorp.ox.ac.uk/ for more information.

Figure 1 presents the performance of the four approaches given above in terms of their precision scores on the 30 chapters, and the average precision achieved by each approach is presented in Table 4. From this figure and table we can see that corpus comparison using frequency ratio directly is outperformed by our working procedure with frequency for initial scoring. Similarly, the working procedure with *tf-idf* for initial scoring outperforms that using *tf-idf* directly as termhood, and performs the best among all the four ATR procedures formulated above. All these verify the significance and effectiveness of the improved corpus comparison approach that we have implemented for enhancing the current ATR technology.

5. Conclusion

We have presented in the above sections an improved approach of corpus comparison to automatic extraction of domain specific terms from thematic corpora, which extends the basic principle of corpus comparison and emerging pattern effectively for performance enhancement. Different from the previous approaches, our approach compares a subject corpus with more than one background corpora and sums up the respective comparison results for each term candidate as its termhood score in the subject field in question.

Accordingly, we have implemented a novel working procedure for term extraction to examine the effectiveness of this approach. The experiments we have carried out on the BLIS corpus of HK laws show that the proposed approach outperforms other approaches of direct corpus comparison and achieves an average precision of 82.98% on the top 10% of candidates according to their termhood in each domain corpus. Also, a nice advantage of this approach is that its performance seems sustainable on a small corpus.

References

Ahmad, K., A. Davies, H. Fulford and M. Rogers. 1994. What is a Term? The Semi-Automatic Extraction of Terms from Text. In M.S. Hornby, F. Pochhacker and K. Kaindl (eds), *Translation Studies: An Interdiscipline*, pp. 267-278. Amsterdam: John Benjamins Publishing Company.

Ananiadou, S. 1988. *A Methodology for Automatic Term Recognition*. Ph.D. thesis, University of Manchester Institute of Science and Technology.

Chang, J. S. 2005. Domain Specific Word Extraction from Hierarchical Web Documents: A First Step toward Building Lexicon Trees from Web Corpora. *Proceedings of the Fourth SIGHAN Workshop on Chinese Language Learning*, pp. 64-71. Korea.

Chen, W. L., J. B. Zhu, T. S. Yao and Y. X. Zhang. 2003. Automatic Learning Field Words by Bootstrapping. *Proceedings of the Joint Seminar on Computational Linguistics 2003*, pp. 67-72. Beijing: Tsinghua University Press.

Chowdhury, G. G. 1999. *Introduction to Modern Information Retrieval*. London: Library Association.

Chung, T. 2003. A Corpus Comparison Approach for Terminology Extraction. *Terminology*, 9(2), 221-246.

Daille, B. 1994. *Approche Mixte pour l'extraction Automatique de Terminologie: Statistique Lexicale et Filtres Linguistiques*. Ph.D. thesis, University Paris 7, France.

Dagan, I. and K. Church. 1994. Termight: Identifying and Translating Technical Terminology. *Proceedings of the 4th Conference on Applied Natural Language Processing*, pp. 34-40. Stuttgart, Germany.

Damerau, F. J. 1993. Generating and Evaluating Domain-Oriented Multi-Word Terms from Texts. *Information Processing & Management*, 29(4), 433-447.

Damerau, F. J. 1990. Evaluating Computer-Generated Domain-Oriented Vocabularies. *Information Processing & Management*, 26(6), 791-801.

Dong, G. and J. Li. 1999. Efficient Mining of Emerging Patterns: Discovering Trends and Differences. *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 43-52. San Diego, CA: ACM Press.

Drouin, P. 2003. Term Extraction Using Non-technical Corpora as a Point of Leverage. *Terminology*, 9(1), 99-115.

Frantzi, K. T., S. Ananiadou and J. Tsujii. 1998. The C-value/NC-value Method of Automatic Recognition for Multi-Word Terms. In C. Nikolaou and C. Stephanidis (eds.), *Proceedings of the 2nd European Conference on Research and Advanced Technology for Digital Libraries*, pp. 585-604. Heraklion, Crete, Greece.

Frantzi, K. T. and S. Ananiadou. 1996. Extracting Nested Collocations. *Proceedings of the 16th International Conference on Computational Linguistics*, pp. 41-46. Copenhagen, Denmark.

Hull, D. A. 2001. Software Tools to Support the Construction of Bilingual Terminology Lexicons. In D. Bourigault, C. Jacquemin and M.C. L'Homme (eds), *Recent Advances in Computational Terminology*, pp. 225-244. Amsterdam: John Benjamins Publishing Company.

Justeson, J. S. and S. M. Katz. 1995. Technical Terminology: Some Linguistic Properties and an Algorithm for Identification in Text. *Natural Language Engineering*, 1(1), 9-27.

Kageura, K. and B. Umino. 1996. Methods of Automatic Term Recognition: A Review. *Terminology*, 3(2), 259-289.

Kilgarriff, A. 2001. Comparing Corpora. *International Journal of Corpus Linguistics*, 6(1), 97-133.

Kit, C. and X. Liu. 2007. Mono-Word Termhood as Rank Difference in Domain and Background Corpora. *Proceedings of International Conference: Keyness in Text*, pp. 41-45. Pontignano, Siena, Italy.

Kit, C., X. Liu, K. K. Sin and J. J. Webster. 2005. Harvesting the Bitexts of the Laws of Hong Kong from the Web. *Proceedings of the 5th Workshop on Asian Language Resources*, pp. 71-78. Jeju Island, Korea.

Lemay, C., M.-C. L'Homme and P. Drouin. 2005. Two Methods for Extracting "Specific" Single-Word Terms from Specialized Corpora: Experimentation and Evaluation. *International Journal of Corpus Linguistics*, 10(2), 227-255.

Li, J. and L. Wong. 2002. Identifying Good Diagnostic Gene Groups from Gene Expression Profiles Using the Concept of Emerging Patterns. *Bioinformatics*, 18(5), 725-734.

Lin, S. C. 2004. Topic Extraction Based on Techniques of Term Extraction and Term Clustering. *Computational Linguistics and Chinese Language Processing*, 9(2), 97-112.

Nakagawa, H. 2001. Automatic Term Recognition Based on Statistics of Compound Nouns. *Terminology*, 6(2), 195-210.

Rayson, P. and R. Garside. 2000. Comparing Corpora Using Frequency Profiling. *Proceedings of the Workshop on Comparing Corpora, the 38th Annual Meeting of the Association for Computational Linguistics*, pp. 1-6.

Salton, G. 1992. The State of Retrieval System Evaluation. *Information Processing & Management*, 28(4), 441-449.

Uchimoto, K., S. Sekine, M. Murata, H. Ozaku and H. Isahara. 2001. Term Recognition Using Corpora from Different Fields. *Japanese Term Extraction: Special issue of Terminology*, 6(2), 233-256.

Vivaldi, J., L. Màrquez and H. Rodríguez. 2001. Improving Term Extraction by System Combination Using Boosting. *Proceedings of the 12th European Conference on Machine Learning*, pp. 515-526. Freiburg, Germany.

Wermter, J. and U. Hahn. 2005. Finding New Terminology in Very Large Corpora. *Proceedings of the Third International Conference on Knowledge Capture*, pp. 137-144. Banff, Alberta, Canada.

Wu, S. H. and W. L. Hsu. 2002. SOAT: A Semi-Automatic Domain Ontology Acquisition Tool from Chinese Corpus. *Proceedings of the 19th international conference on Computa-*

tional linguistics, pp. 1-5. Taipei.

Xu, F. Y., D. Kurz, J. Piskorski and S. Schmeier. 2002. A Domain Adaptive Approach to Automatic Acquisition of Domain Relevant Terms and Their Relations with Bootstrapping. *Proceedings of the Third International Conference on Language Resources and Evaluation*, Spain.

Yang, H. Z. 1986. A New Technique for Identifying Scientific/Technical Terms and Describing Science Texts. *Literary and Linguistic Computing*, 1(2), 93-103.

Yangarber, R., R. Grishman, P. Tapanainen and S. Huttunen. 2000. Automatic Acquisition of Domain Knowledge for Information Extraction. *Proceedings of the 18th International Conference on Computational Linguistics*. Germany.

Zhou, G. and Y. Nie. 2005. Improving Retrieval Effectiveness by Using Key Terms in Top Retrieved Documents. *Advances in Information Retrieval*, 3408.