Chinese Information Retrieval Using Lemur: NTCIR-5 CIR Experiments at UNT

Jiangping Chen, Rowena Li, Fei Li
School of Library and Information Sciences
University of North Texas, P.O. Box 311068, Denton, TX 76203, USA
{jpchen, rll0099, fl0030}@unt.edu

Abstract

This paper describes our participation in NTCIR-5 Chinese Information Retrieval (IR) evaluation. The main purpose is to evaluate Lemur, a freely available information retrieval toolkit. Our results showed that Lemur could provide above average performance on most of the runs. We also compared manual queries vs. automatic queries for Chinese IR. The results show that manually generated queries did not have much effect on IR performance. More analysis will be carried out to discover causes behind hard topics and ways to improve the overall retrieval performance.
Keywords: *text retrieval, Chinese, Lemur, Chinese text segmentation, system evaluation.*

1 Introduction

Chinese Information Retrieval (CIR) research aims to find relevant Chinese documents for users' queries. CIR becomes an important and necessary component for some other Chinese information processing tasks such as Chinese Question Answering (QA), Cross-Language Information Retrieval (CLIR), and Cross-Language Question Answering (CLQA) involving Chinese. CIR is largely conducted following the same strategies as English monolingual information retrieval [2]. The only difference is that there is a pre-processing step prior to indexing and query processing – Chinese text segmentation.

This paper describes our experiments using Lemur – a freely available IR toolkit for experimental IR research – for Chinese Information Retrieval evaluation at NTCIR-5 Workshop. Our research purposes are two-fold: one is to evaluate the IR performance of Lemur for Chinese to help our deciding whether Lemur can be used as the search engine for our other IR evaluation tasks such as Chinese QA and CLQA; the other is to examine the effect of different query generation mechanism on retrieval performance to understand whether manually-generated queries improve the performance of information retrieval as compared to automatic generated queries.

In addition to Chinese IR, we had planned to participate in Chinese-English Cross-Language Information Retrieval at this workshop as well. However, we had to give up the plan due to the late release of the English document collection and the team members' involvement in other research. We only submitted our results on Chinese Information Retrieval using Lemur for CLIR task.

This paper is organized as follows. Section 2 introduces the basic functions of Lemur and the reasons behind our decision to use it as our IR search engine for various evaluations. Section 3 describes the procedures of using Lemur for CIR experiments. Section 4 reports our submissions and evaluation results. Section 5 provides analysis of the results. The paper concludes with possible research to improve IR performance using Lemur.

2 The Lemur Toolkit

As a newly established research group, we are considering adapting one of the available Information Retrieval (IR) systems for our research purposes, and Lemur becomes one of our candidates for IR search engines among others, such as Smart (ftp://ftp.cs.cornell.edu/pub/smart/) and Lucene (http://lucene.apache.org/java/docs/). The Lemur Toolkit (http://www.lemurproject.org/) is designed and developed by researchers from the Computer Science Department at the University of Massachusetts and the School of Computer Science at Carnegie Mellon University. The project is sponsored by the Advanced Research and Development Activity in Information Technology (ARDA) and by the National Science Foundation (NSF). There are several reasons to consider Lemur, including:

a. Lemur supports document indexing and several text retrieval models including the TFIDF retrieval model, the Okapi BM25 retrieval function, the KL-divergence language model, the InQuery (CORI) retrieval model, CORI collection selection, Cosine similarity model, and Indri structured query language (http://www.lemurproject.org/lemur/retrieval.html). Other systems focus on only one method.

b. The developer states that the toolkit is "under constant development for performance improvements as well as feature additions" (http://www.lemurproject.org/news.html). The latest version is Lemur 4.1 with certain additions such as UTF-8 document format parsing support and additional document structure support.
c. The system is designed as a research system, and is quite convenient to be used for TREC-type IR experiments because it accepts TREC document format and produces TREC-type results for evaluation.
d. The toolkit is expandable and adaptable with available source codes. We can adapt Lemur for various purposes such as Cross-Language Information Retrieval and Question Answering.

Before we decided to use Lemur as our primary search engine for Chinese text retrieval experiments, we tested Lemur using NTCIR-4 CLIR Chinese test collection. The steps and results are presented in the next section.

3 Chinese information retrieval experiments using Lemur

Our Chinese IR using Lemur generally includes following steps: Chinese text segmentation, query processing, and retrieval using Lemur.

3.1 Chinese text segmentation

The first task for any Chinese information-processing application, such as Chinese information retrieval, is text segmentation since there is no word boundary in Chinese text. So far, bi-gram text segmentation and word segmentation have been widely used for segment Chinese texts for CIR [4]. We have evaluated both bi-gram and word segmentation before applying Lemur for the NTCIR-5 Chinese information retrieval evaluation.

3.1.1 Bi-gram text segmentation. Bi-gram segmentation is quite straightforward. It segments Chinese sentences into overlapping bi-grams. The approach has been proved to be more effective than other n-gram approaches to Chinese text segmentation. The disadvantage of bi-gram segmentation is that it often leads to a large index term space, three times larger than that of short-word indexing [3]. Table 1 presents our IR experiments using bi-gram segmentation with various Lemur retrieval models.

3.1.2 Overlapping short word segmentation. Chinese word segmentation identifies word boundary within a Chinese text. There is a large body of literature on word segmentation [5, 6, 7, 8]. Various experiments have been carried out on different segmentation approaches including corpora or statistical approaches, lexical or dictionary based approaches, and hybrid approaches that combine lexical information and statistical information from training material. We applied dictionary based approach to segment the text using forward maximum matching between a Chinese sentence and the dictionary because it was fast and easy to implement.

We prefer to use a segmentation dictionary that is generated from the same document type as the test document collection. However, we could not find an appropriate one to use for CIR purpose. Therefore, we constructed a segmentation dictionary by combining Chinese lexical items from multiple resources. Table 1 denotes the lexical resources we used to compile the dictionary. The first three include Chinese words or phrases that are derived from current Chinese texts from mainland China. They are encoded in Simplified Chinese. An online Chinese encoding converter (http://www.mandarintools.com/zhcode.html) were used to convert them into Big5-encoding. The fourth resource is the website "Who's Who in Taiwan" retrieved on May 5, 2005. We compiled the person names, and names of high peaks and major rivers (http://www.gio.gov.tw/taiwan-website/5-gp/yearbook/P374.htm, http://www.gio.gov.tw/taiwan-website/5-gp/yearbook/P517.htm, http://www.gio.gov.tw/taiwan-website/5-gp/yearbook/contents.htm). The last resource is the document collection itself. It is used in two ways: to manually construct a list of named entities, and to automatically identify possible 2-character and 3-character short words. The process of constructing the dictionary is summarized below.

a. The lexical items (words and phrases) in the first three resources listed in Table 1 were compiled into a preliminary dictionary after the encoding conversion. The dictionary keeps the part-of-speech (POS) information of each term if it has one in the HIT-IRLab Synonym Dictionary or the PKU corpus. A new tag "ldc" was assigned to a term if it only occurs in the LDC bilingual dictionary. Also, the frequency of the term under each POS was also recorded. Table 2 shows the sample entries of the dictionary.
b. The person names and names of high peaks and major rivers on the website 'Who's Who in Taiwan" were automatically extracted and organized. A list of 1,082 non-duplicated person names and 56 mountain/river names were added to the dictionary generated in the previous step.
c. About 2,744 named entities (person names, organization names, and location) were manually extracted from the document collection and were added to the dictionary.

Table 1. Resources for constructing segmentation dictionary

Name	Owner/URL	Content Description	Sample Entries
LDC Chinese-English Translation Lexicon Version 3.0	Linguistic Data Consortium (LDC)	English-Chinese Bilingual word list containing about 54,170 Chinese head words	昂首 /hold one's head high/ 昂首阔步 /stride forward with one's chin up/stride proudly ahead/ 昂扬 /high-spirited/
HIT-IRLab Synonym Dictionary	Information Retrieval Lab, Harbin Institute of Technology, China (http://ir.hit.edu.cn/phpwebsite/)	About 77,343 Chinese words and their synonyms	Aa05B01= 别人 旁人 他人 人家 Aa05E01@ 克隆人 Aa06A01= 谁 孰 谁人 谁个 何人 哪个 哪位 何许人也
PKU Corpus	the Institute of Computational Linguistics, Beijing University (http://www.icl.pku.edu.cn)	About 39,057 Segmented and POS-tagged Chinese sentences and 56,534 unique Chinese words	19980107-05-006-003/m １９９７年/t １０月/t ３日/t 傍晚/t ，/w 在/p 淅淅沥沥/z 的/u 雨声/n 中/f ，/w 一/m 位/q 农业/n 战线/n 上/f 的/u 科技/n 工作者/n 去世/v 了/y 。/w
Who's Who in Taiwan	(http://www.gio.gov.tw/taiwan-website/5-gp/yearbook/contents.htm)	The site includes lists of person names and names of high peaks and major rivers in Taiwan	CHEN, JIN-DE (See CHEN, CHIN-DE 陳金德) CHEN, JIN-DING (See CHEN, CHIN-TING 陳進丁)
Document collection	NTCIR-5 CLIR Chinese document collection	About 901,446 news articles from 2000 to 2001 taken from UDN.COM	

d. The last step in the dictionary construction was to augment the dictionary with new terms extracted from the documents of the test collection. First, we segmented the document collection into overlapping bi-grams and trigrams separately. Then we calculated the frequency of each extracted unit and considered it as a legitimate word if the unit occurs more than 1000 times in the whole collection.

The Chinese word segmentation dictionary is in the format as shown in table 2, in which each entry contains the word, its POS, and the frequency of the term in that POS. Some of the terms may have multiple POSs. This dictionary was also used for our CLQA participation to provide preliminary annotation including part-of-speech tagging for the Chinese document collection and Chinese test questions.

Table 2. Sample Lexical items in the Chinese word segmentation dictionary

怪圈 n|1	機會 n|114
怪異 ldc|1	機遇 n|95
講究 v|16 a|3 vn|1	講排場 l|2 v|1
群策群力 i|8	機電 b|24
群雄逐鹿 l|3	繆青民 nr|1
歷史劇 n|7	繆紹強 nr|1

3.2 Query processing

Each of the NTCIR-5 test topics is composed of five portions: Title (T), Description (D), Narrative (N), and Concepts (CONC) as illustrated in Figure 1. The workshop required that each participant submit at least one run using only the title portion and one run using only the description portion, since they are short and more close to real-world user queries.

One of the purposes of our participation is to examine whether manually generated queries out of the topics will have positive effect on the retrieval performance compared with the automatic generated queries. Therefore, in addition to the required runs, we added a manual run to our submissions. In general, we have two ways to generate queries and send them to Lemur for retrieval.

```
<TOPIC>
<NUM>034</NUM>
<SLANG>KR</SLANG>
<TLANG>CH</TLANG>
<TITLE>賓拉登，美國，軍事手段</TITLE>
<DESC>查詢美國追捕反美恐怖份子首領賓拉登的策略。</DESC>
<NARR>
<BACK>美國急於逮捕賓拉登。</BACK>
<REL>詳細描述美國採取的軍事和外交手段和行動，並有列出提出陳述的主要負責機構或人員的報導為相關。沒有提到特定人名或只是一般活動的報導為不相關。反恐專家分析賓拉登為何要使用恐怖活動作為手段時，如有提到專家名字時為部分相關。賓拉登和蓋達組織以外的恐怖份子報導為不相關。</REL>
</NARR>
<CONC>賓拉登，蓋達組織，恐怖活動，阿富汗，美國中情局，CIA</CONC>
</TOPIC>
```

Figure 1. A sample NTCIR-5 test topic for Chinese information retrieval task

3.2.1 Automatic query generation. We wrote a parser to extract various portions from the test topics, such as Title or Description. Afterwards, the extracted Chinese sentences or phrases were sent to the two Chinese text segmentation programs for segmentation. The results were used by Lemur in two different types of IR experiments based on either bi-gram indexing, or short-word indexing. For retrieval based on bi-gram indexing, we segmented the queries into bi-grams, and for IR using short-word indexing, we segmented queries into short words applying the same segmentation dictionary as for the document collection. The segmented queries were then converted to Lemur format for retrieval.

3.2.2 Manual query generation. Among the five official runs that we submitted for evaluation, the third run UNTIR-C-C-D-03 was based on manually generated queries. The query creator was one of the authors who is a school librarian and is skillful in using the Web resources. She was given a short instruction on how to create the queries, as well as a file containing description portions of the 50 test topics in the format of the second column in Table 3. Below is the manual query generation instruction.

> *Suppose you were a librarian or consultant who is helping a client to locate relevant documents to her information needs from a large Chinese document collection and there were 50 topics presented to you by the client. You need to formulate a query (a list of keywords) for each topic based on your understanding of what the client really wants. You can use the Internet or other references available to you to expand your queries with important terms. There is no right or wrong answers in this experiment. Please try your best to formulate the queries and include only the most important terms (less than 20) in the queries. Then generate the queries in the same format as the provided sample query. But separate the keywords with white spaces.*

Table 3. Samples given to the query creator

Original sample topic given to the creator	<TOPIC> <NUM>001</NUM> <DESC>查詢台灣勞工秋鬥遊行的訴求內容以及政府在 1998 年所提出的勞工政策。</DESC> </TOPIC>
Required output	<TOPIC> <NUM>001</NUM> <DESC>秋鬥 訴求 勞工 抗議 台灣 勞工政策 勞委會 秋鬥大遊行 1998 回應重點</DESC> </TOPIC>

The required output format is the same as the input format shown in Table 3. Generally speaking, we asked the creator to provide segmented Chinese keywords as the query for each topic. Also, we asked her to provide the average time she spent on each topic, the resources that she had used for creating the queries, and any comments on queries. As a result, the creator reported that she spent 5-20 minutes for each topic, and solely used Google to help her to formulate the queries. Google was used to find important keywords to the topics. The retrieval results are reported in section 4.

3.3 Indexing and Retrieval using Lemur

After the documents were segmented and the queries were generated, we used lemur to index the documents and to conduct the retrieval. We tested Lemur on NTCIR-4 CLIR Chinese IR test collection before we carried out the experiments for NTCIR-5 Chinese IR.

The test collection used in NTCIR-4 CLIR task is composed of one document collection, 59 test topics, and two relevance judgment files. The document collection consists of 249,203 news articles taken from United Daily News and 132,172 articles from China Times, China Times Express, Commercial Times, China Daily News, and Central and Daily News spanning from 1998 to 1999. The test topics were in the same format as NTCIR-5 CLIR Chinese IR subtask, as shown in Figure 1. Originally there were 60 test topics. Topic 025 was removed from the relevant judgment file due to too few rigid relevant documents [4]. Unlike TREC, NTCIR workshop has applied four degrees of relevance in the judgment processes: highly relevant, relevant, partially relevant, and irrelevant. Thus, two relevance judgment files were generated out of the human evaluation process: the rigid relevance judgment file, which assigns relevancy only to highly relevant and relevant documents; and the relax relevance judgment file, which also assigns relevancy to partially relevant documents. The two relevance judgment files facilitate the automatic evaluation of Lemur retrieval performance.

We applied both bi-gram segmentation and word segmentation to the NTCIR-4 Chinese documents. Two types of indexing were generated. One is overlapping bi-gram indexing, and the other is Chinese word indexing. IR experiments were carried out using various retrieval methods in Lemur. Table 4 presents the IR performance with various retrieval models in Lemur for both indexing approaches. Please refer to Lemur documentation (http://www.lemurproject.org/lemur/retrieval.html) for the retrieval methods in the table. The Mean Average Precision (MAP) scores reported in the two tables are based on the rigid relevance judgment file. It turned out that bi-gram segmentation slightly outperformed word segmentation on most of the runs. The highest performance was achieved by using bi-gram segmentation and Okapi BM25 retrieval model with pseudo-relevance feedback, as shown in bold in Table 4. However, the word segmentation runs using Okapi with feedback returned more relevant documents from the top 1000 retrieved ones than the corresponding bi-gram runs. Some runs failed to return any results due to either query segmentation errors or unknown reasons, which is indicated as "/" in Table 4.

Table 4. IR performance based on rigid relevance judgment

Retrieval Method	Mean Average Precision				Total number of retrieved relevant docs.			
	D run		T run		D run		T run	
	Bi-gram	Word	Bi-gram	Word	Bi-gram	Word	Bi-gram	Word
Okapi	0.1590	0.1438	0.1881	0.1809	940	894	958	985
Okapi with feedback	**0.1948**	0.1718	**0.2200**	0.2192	1011	1023	1024	1079
tfidf	0.1437	0.1193	0.1751	0.1418	877	846	971	930
tfidf with feedback	0.1699	0.1491	/	0.1556	1000	898	/	1011
rerank_simple_tfidf	0.1437	/	/	0.1418	877	/	/	930
rerank_fb_tfidf	0.1653	0.1469	/	0.1530	877	846	/	930
mixfb_kl	0.1742	0.1757	0.1818	0.1829	977	1057	989	1027
simple_kl_abs	0.1450	0.1425	0.1800	0.1839	901	895	948	964
simple_kl_jm	0.1386	0.1418	0.1708	0.1739	896	931	938	957
simple_kl_dir	0.1423	0.1511	0.1753	0.1731	892	953	956	984
mixfb_kl_dir	0.1659	0.1711	0.1787	0.1783	955	1039	977	1011
mixrelfb_kl_dir	0.1659	0.1711	0.1787	0.1783	955	1039	977	1011
rerank_simple_kl_dir	0.1420	0.1472	/	0.1694	877	846	/	930
twostage	0.1333	0.1309	0.1728	0.1732	771	857	908	955

Based on the testing results of NTCIR-4, we decided to apply Okapi BM 25 with relevance feedback for NTCIR-5 Chinese IR.

4 NTCIR-5 submissions and results

We employed the default parameter setting for Okapi for all our Chinese IR experiments. Five runs were officially submitted to NTCIR-5 for evaluation including one T run (01), two D runs (02 and 03), and two other runs (04 and 05). Run UNTIR-C-C-04 was a DN run, and UNTIR-C-C-05 was a run that employed all the components of the topics. The results are shown in Table 5 and Table 6, along with the minimum, maximum, median, and average MAP scores of all participating systems for the same type of runs. Table 5 presents the results based on the rigid relevance judgment, and Table 6 shows the results based on the relax relevance judgment. Our results are all above average. Most of our scores are also above median except one run (UNTIR-C-C-D-02) in terms of rigid evaluation.

Table 5. Official rigid evaluation results

	C-C-T	C-C-D	C-C-O
UNTIR runs	**0.3180** (01)	**0.3146** (02) **0.3279** (03)	**0.4044** (04) **0.4048** (05)
min	0.0086	0.0061	0.1876
max	0.5049	0.4826	0.4419
med	0.3069	0.3223	0.3772
ave	0.2874	0.3279	0.3520

Table 6. Official relax evaluation results

	C-C-T	C-C-D	C-C-O
UNTIR runs	**0.3752** (01)	**0.3852** (02) **0.4025** (03)	**0.4783** (04) **0.4777** (05)
min	0.0112	0.0113	0.2175
max	0.5441	0.5249	0.5095
med	0.3576	0.3839	0.4399
ave	0.3319	0.3523	0.4131

The results show that Lemur could provide acceptable performance on Chinese IR without any extra effort on query expansion or applying NLP techniques. All the runs achieved above average performance, and most of them were higher than the median. However, the results show that significant differences exist between our system and the best performed ones. There is big room for performance improvement.

It is surprising that the official MAPs of the two D runs were quite close to each other. We expected that the manually generated queries would have positive effects on the performance, but it turned out that it might not be the case. In the next section, we will perform some analysis to our results.

5 Analysis

We first extracted the average precision scores for each query of our five runs from the official evaluation results. We tried to understand what types of queries were causing trouble by adapting the Retrieval Outcome Analysis Framework in [1].

The ROA Framework was originally proposed for analyzing Cross-Language Information Retrieval results. Here we take one of the components – the query categorization scheme as illustrated in Table 7 for our topic classification.

We classified the 50 test topics into six categories based on the faceted classification in Table 7. Table 8 shows the classification results. The thresholds used in the classification, such as Ptop (0.5) and Pbottom (0.17), is determined based on the distribution of the actual average precision scores of all queries in the five runs.

Table 7. Query categorization scheme in [1]

Facet	Category	Categorization Approach
Difficulty	Easy	Queries whose average precision is above certain threshold P_{top}
	Moderate	Queries whose average precision is between thresholds P_{top} and P_{bottom}
	Hard	Queries whose average precision is below certain threshold P_{bottom}
Stability	Stable	Queries of which the differences in average precision scores between experimental groups are within a predefined range
	Unstable	Otherwise

Table 8. Topic classification results

Category	Classification Criteria	Topics in the Category	Total Number of Topics
Hard & Stable	AV (average precision) score was below 0.17 for all runs	26, 31, 34, 38, 45	5
Hard & Unstable	AV score was below 0.17 for most runs, but one run got higher scores	25, 33, 39, 49	4
Moderate & Stable	AV score was between 0.17 – 0. 5 for all runs	1, 3, 4, 8, 12, 22, 46, 50	8
Moderate & Unstable	Most AV score was between 0.17 – 0. 5 for most runs, but one or more other runs got higher or lower scores	2, 5, 7, 9, 10, 11, 13, 14, 15, 19, 20, 21, 28, 29, 32, 35, 36, 37, 40, 42, 43, 44, 47, 48	24
Easy & Stable	AV score was above 0.5 (include 0.5) for all runs	6, 16, 17, 18, 23, 24, 30	7
Easy & Unstable	AV score is above 0.5 (include 0.5) for most runs but one run got lower scores	27, 41	2

The classification provides a clear picture of what topics the system did well and what topics the system did poorly. It is easy to understand that many topics are unstable because each run uses different portions of the topics. However, there are some hard topics that are persistently difficult no matter what portions were used for generating the queries, such as topics 26, 31, 34, 38, and 45. Further analysis will first focus on hard topics to discover the causes for the low scores, then on unstable topics to find out better query expansion solutions. Alternative retrieval methods may need to be employed for hard queries. Yang et al. [9] applied ontology in addition to other query expansion techniques to Chinese Information Retrieval and achieved quite big improvement, which shows the power of linguistic knowledge and techniques in handling hard topics.

We also analyzed the two D runs by looking at the queries and the IR results. UNTIR-C-C-D-02 employed automatically generated queries by our query processing program, and UNTIR-C-C-D-03 uses manually generated queries as described in Section 3.2.2. Even though the run using the manual queries achieved slightly better performance for about 35 queries, the difference is rather small. As a result, the two MAPs are very close to each other. A further inspection of the queries shows that the manual queries contain rather limited new terms that were not occurred in the original description where the automatic run was derived. The terms added by the human subject did improve the IR performance, but not in a dramatic way.

6 Conclusions and future work

The purposes of our participation in NTCIR-5 Chinese IR include evaluating Lemur and gaining insights into the effects of manually generated queries on retrieval performance. Our results proved the usefulness of Lemur for Chinese IR though there is room for further improvement. The manually produced queries did not improve much of the IR performance, which indicated that other ways may need to be explored to improve query generation.

Future research will focus on exploring the reasons behind hard topics and integrating NLP techniques to automatically identify these hard topics in order to improve their IR performance.

References

[1] J. Chen. Toward a Unified Retrieval Outcome Analysis Framework for Cross Language Information *Online proceedings of 2005 annual conference of the American Society for Information Science and Technology*, 2005. http://max.lis.unt.edu/publication/ASIST2005_Chen.pdf.

[2] K. L. Kwok, L. Grunfeld and J. Xu. TREC-6 English and Chinese Retrieval Experiments Using PIRCS. Paper presented at *the Sixth Text Retrieval Conference (TREC-6), NIST Special Publication 500-240, Gaithersburg: National Institute of Standards and Technology*, pp. 207-214, 1998.

[3] K. L. Kwok. Lexicon effects on Chinese information retrieval. *Proceedings of 2nd Conference on Empirical Methods in NLP*, pp. 141-148, 1997.

[4] K. Kishida, K. Chen, S. Lee, K. Kuriyama, N. Kando, H. Chen, S. H. Myaeng and K. Eguchi. Overview of CLIR Task at the Fourth NTCIR Workshop. *Online Proceedings of NTCIR-4 Workshop*, 2004. Retrieved Oct. 15, 2005, from: http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings4/CLIR/NTCIR4-CLIR-TaskOverview.pdf.

[5] D. D. Palmer. A trainable rule-based algorithm for word segmentation. Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL'97), 1997.

[6] J. Y. Nie, J. F. Gao, J. Zhang, & M. Zhou On the use of words and n-grams for Chinese information retrieval. Proceedings of the Fifth International Workshop on Information Retrieval with Asian Languages (pp. 141 – 148), 2000.

[7] N. Xue, & L. Shen. Chinese word segmentation as LMR tagging. In. Proceedings of the Second SIGHAN Workshop on Chinese Language Processing. 2003. Available at: http://www.cis.upenn.edu/~libin/paper/sighan03.pdf

[8] C. Goh, M. Asahara, & Y. Matsumoto. Chinese word segmentation by classification of characters. Proceedings of 3rd ACL SIGHAN Workshop on Chinese Language Processing. 2004. Retrieved September 16, 2004, from http://www1.cs.columbia.edu/~rambow/master-cd-rom/sighan/PDF/13-Ling-2col.pdf.

[9] L. Yang, D. Ji, & L. Tang. Chinese Information Retrieval Based on Terms and Ontology. Working Notes of NTCIR-4, Tokyo, 2-4 June 2004.