Human Evaluation of Keyword Extraction System Using Lexical Chains

Bo-Yeong Kang° Sang-Jo Lee
Department of Computer Engineering in Kyungpook National University
comeng99@hotmail.com sjlee@bh.knu.ac.kr

어휘 체인을 이용한 키워드 추출 시스템 성능 평가

강보영° 이상조
경북대학교 컴퓨터공학과

Abstract

In Information Retrieval or Digital Library, one of the most important factors is to find out the exact information which users need. Exact keywords which represent the content of a document can be much help to find the exact information. In this paper, we evaluate an efficient keyword extraction system by recall and precision. The results presented here are based on the human evaluations of the quality and the appropriateness of keywords.

1 Introduction

In Information Retreival and Digital Library, It is very important factor to extract the information which matches with users' need. If there are good indicators or keywords to represent sementic content of documents, it can be much help to research the areas like web content mining, text mining, document summarization or document clustering/classification. We presented an efficient keyword detection system using lexical chains in the earlier paper[1], and in this paper we will evaluate the system by well known metrics, recall and precision.

Broadly speaking there are two indexing methods, *extraction indexing* and *assignment indexing*[2][3]. Our system which we suggested is to select keywords directly from a document text.

Moens[2] said that the disadvantage of extraction indexing method is that the words in a document is potentially ambiguous. Our system performs sense disambiguation roughly in the process of finding keywords. This enables us to deal with lexical ambiguities better. There is a further research of extracton indexing that discourse structures and their signaling linguistic phenomena can help in selecting terms to reflect content of a text [3][4]. Morris and Hirst[5][6] suggested that the discourse structure of a text might be found out by lexical chains. These gave us a clue to employ lexical chains for detecting keywords.

This paper is organized as follows: in section 2, we will review the algorithm of Lex, and in section3, we evaluate our system. In section 4. we conclude this paper.

2 Keyword Extraction System

We assumes that nouns used importantly in

documents can indicate the semantic content of documents well. To find out important nouns in documents, we used lexical chains. Proposed system works as the following:

1. Extract nouns from a document
2. Extracted nouns have multiple senses. Therefore, we perform rough sense disambiguation. We specially present the semantic_window, the method to search the semantic relations of peripheral nouns using window for sense disambiguation.
3. Generate lexical chains with nouns whose senses are determined and rank each nouns and chains in lexical chains.
4. After ranking each chain in the lexical chain, we find out strong chains among many chains by revised Barziay's formula.
5. Extract keyword in the strong chains. We extract keywords not from the strongest chain, but from a few strong chains because they can repersent the semantic content of a document better.

The criteria for *strong chain* follows the revised formula of Barzilay[5]: constant C can be varied depending on the number of keywords. we extract keywords from a few strong chains.

$$\text{Score(Chain)} > \text{Average(Scores)} + C*\text{StandardDeviation(Scores)}$$

3 Human Evaluation

In this section, we will evaluate the quality of our system. For extracting keywords, 11 abstracts of 11 English papers were randomly selected from on-line ACM paper archive. The contents of 11 abstracts were related with Computer Science. And keywords were manually selected by 5 subjects. Subjects are graduate students in the Department of Computer Engineering at Kyungpook National University and Korea Advanced Institute of Science and Technology.

We measured agreement among judges using *percent agreement*, a measure defined by Gale[6]. Percent agreement is the ratio of oberved to possible agreements with the majority opinion. Following table 1 shows the percent agreement according to the number of keywords selected by subjects. For example, the number of keywords selected by subjects is 2, the percent agreement about keywords is 93.6%.

Table 1. Percent agreement among subjects

	Percent agreememt
Manully extracted keyword = 2	93.6%
Manully extracted keyword = 3	87.8%
Manully extracted keyword = 4	80%

We manually extracted 2, 3 and 4 keywords from each abstract. We used recall, precision and F-value as evaluation metrics[2]. We can calculate precision and recall as a single value, F-value.

Following tables are the results of our system according to the number of keywords selected by our system and the number of manually selected keywords by subjets.

Table 2. The number of selected keywords by subjects are 2

	Recall	Precision	F-value
Extracted keyword = 2	.27	.27	.27
Extracted keyword = 3	.46	.30	.36
Extracted keyword = 4	.59	.30	.40
Extracted keyword = 5	.77	.20	.31
Extracted keyword = 6	.82	.26	.39

Table 3. The number of selected keywords by subjects are 3

	Recall	Precision	F-value
Extracted keyword = 2	.24	.36	.28
Extracted keyword = 3	.39	.39	.39
Extracted keyword = 4	.55	.48	.51
Extracted keyword = 5	.67	.40	.50
Extracted keyword = 6	.76	.38	.50

Table 4. The number of selected keywords by subjects are 4

	Recall	Precision	F-value
Extracted keyword = 2	.18	.36	.24
Extracted keyword = 3	.30	.39	.34
Extracted keyword = 4	.43	.43	.43
Extracted keyword = 5	.52	.42	.46
Extracted keyword = 6	.61	.41	.49

We examined the system accuracy under the condition where our system gave us the best F-value. In other words, we got the system accuracy when the number of manually extracted keywords by subjects was 3 and the number of keywords selected by system was 4. The overall system accuracy is 0.86.

5 Conclusion and Future Work

In this paper, we evaluated the presented efficient keyword extraction system. Sense disambiguation result was average 69.2%. But it can be enhanced if we support the control of compound nouns, and proper nouns and unknown nouns which aren't registered on WordNet. In an experiment with 11 abstracts of papers, when the number of manually selected keywords was 3 and the number of keywords by system was 4, our system gave us the best F-value, 0.51. Then, percent agreement among subjects was 87.8%, and the system accuracy was 0.86. But, there are some limitations in our system. We extract the keywords only from nouns which exist in a document. And our system was developed to work in general domains, but our test documents have many terminologies of Computer Science. These factors interrupted constructing exact lexical chians. In addition, keywords are often more informative when they compose a compound noun. The improvements of these problems will enable us to implement a robust keyword extraction system.

References

[1] Kang, B.Y., and Lee, S.J., "Efficient keyword extractin method using lexical chains", the 2nd International Symposium on Advanced Intelligent System, 2001.

[2] Moens, M.-F., *Automatic Indexing and Abstracting of Document Texts*, Kluwer Academic Publishers, 2000.

[3] Lewis, D.D., and Sparck Jones, K., "Natural language processing for information retrieval," *Communications of the ACM*, 39(1), 92-101, 1996.

[4] Morris, J., and Hirst, G., "Lexical cohesion computed by thesaural relations as an indicator of the sturcture of text," *Computational Linguistics*, 17(1), pp. 21-43, 1991.

[5] Barzilay, R. and Elhadad, M., "Using lexical chains for text summarization," *In the Proceedings of the ACL'97 Workshop on Intelligent Scalable Text Summarization*, 1997.

[6] Gale, W., Church, K., and Yarwsky, D., "Estimating upper and lower bounds on the performance of word-sense disambiguation programs," *In Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics(ACL-92)*, 1992