A Search Result Clustering Method using Informatively Named Entities

Hiroyuki Toda
NTT Cyber Solutions Laboratories, NTT
Corporation
1-1 Hikarinooka Yokosuka-Shi
Kanagawa, 239-0847 Japan
toda.hiroyuki@lab.ntt.co.jp

Ryoji Kataoka
NTT Cyber Solutions Laboratories, NTT
Corporation
1-1 Hikarinooka Yokosuka-Shi
Kanagawa, 239-0847 Japan
kataoka.ryoji@lab.ntt.co.jp

ABSTRACT

Clustering the results of a search helps the user to overview the information returned. In this paper, we regard the clustering task as indexing the search results. Here, an index means a structured label list that can makes it easier for the user to comprehend the labels and search results. To realize this goal, we make three proposals. First is to use Named Entity Extraction for term extraction. Second is a new label selecting criterion based on importance in the search result and the relation between terms and search queries. The third is label categorization using category information of labels, which is generated by NE extraction. We implement a prototype system based on these proposals and find that it offers much higher performance than existing methods; we focus on news articles in this paper.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Search process, Clustering*

General Terms

Algorithms, Experimentation

Keywords

Search result clustering, Named entity

1. INTRODUCTION

Since most users can't input the appropriate query to directly indicate the desired document[1], the typical retrieval system outputs the many results returned as a long flat list. The entries can be ranked by using criteria such as PageRank[2] or relevancy to the query[3]. This list is counterproductive because the user finds it extremely difficult to scroll through the many entries to find the desired information.

The preferred solution, forcing the user to input a larger number of highly specific keywords, reduces the number of results and yields somewhat better ranking accuracy. This solution is, unfortunately, not practical for the average user since most users tend to input no more than 3 keywords[4] .

To solve this problem, methods that cluster the search result have been proposed. These methods have two main goals. One is to make it easier for the user who has a clear search target to locate the desired document because the user can easily select the most appropriate cluster. The other is to assist the user, who would like to browse using just a few keywords or who has no clear search target, in understanding the outline of the search result through the labels of the clusters.

A secondary benefit is that the user often finds unexpected information related to the search target. If, for example, the user searches for information on the American presidential elections and he inputs "presidential election" to the search system, he can also find the cluster that describes the presidential elections of other countries.

Realizing these benefits, however, requires not only correct clustering but also labels clear enough to explain the clusters[5]. To realize these clustering and labeling goals, a lot of research is being directed towards methods which don't use conventional clustering methods and regard the clustering task as the task of selecting important terms(words or phrases) [5][6][7][8].

In this paper, we advance this idea; we regard the clustering task as the task of building an index of the search result. Here, index means a structured label list that makes it easier for the user to overview the labels and search results. To realize our proposition, we have three proposals.

The first one is to use Named Entity (NE) extraction tools to extract the terms as label candidates from documents because we consider that proper nouns are important for characterizing documents. Second is a new label selecting criterion that can select the labels from candidates. The criterion is based on the importance of the term in the search result and the relation between the label and search query. Third is a label presentation method that categorizes the labels using the category information identified in NE extraction, because we consider that the simple enumeration of labels does not support easy overview.

To summarize, we specify the important labels using NE extraction and a new label selecting criterion and construct the index, which makes it easier for the user to overview

the search results, from the labels and the label's category information.

The evaluations described in this paper use a newspaper corpora and an online news article corpora, because the main NE extraction tools were developed for newspapers.

The paper is organized as follows. The next section introduces related works. In Section 3, we formulate the problem and show our approach. In Section 4, we describe the details of our approach. Evaluation of the proposed methods is shown in Section 5. Finally, we conclude the paper in Section 6.

2. RELATED WORKS

2.1 Search result clustering

Many approaches are being researched for organizing the search results to improve the processes of overviewing and winnowing them. There are two main approaches: the document-based approach and the label-based approach.

Instances of the document-based approach include the many methods that employ document clustering. Such methods cluster the documents using the similarity of features such as the keyword vectors of documents. After that, they extract representative term(s) or sentence(s) as labels from each cluster, which are then presented to the user with the search result. Scatter/Gather[9] and the method of Leuski[10] adopt this approach.

These document-based approaches usually yield non - overlapped clusters and label quality is influenced by the accuracy of clustering. Though the number of clusters or similarity threshold generally controls clustering, it is difficult to select the value that suits the user's intention. As a result, labels are often unreadable, so it is difficult to adopt this approach in search engine applications.

The label-based approach, on the other hand, first extracts informative terms(words or phrases) from the search result as labels using a statistical analysis of components such as the frequency of appearance or the properties of clusters formalized by the terms(The cluster is formalized by the documents that include the term). The labels are presented to the user with the search result.

Zeng et al.[6] proposed a label based approach that uses phrases scored by a combination of some properties of labels and document clusters formalized by the labels. Kummamuru et al.[7] regard the label list as a taxonomy of the search result, and proposed a label selecting criterion based on taxonomy likelihood. Ohta et al.[8] proposed some label selecting criteria that improve on the TF-IDF[3] criterion which uses term position. The criterion is influenced by the ranking of the documents in which the term appears as well as the term appearance position in the documents. Unfortunately, empirical knowledge indicates that the TF-IDF criterion is influenced by term frequency and doesn't sufficiently reject high frequency terms[11] .

Vivisimo[1] and mooter[2] are web search engines that employ the label-based approach. Their problems are discussed in a following section.

While our method is based on the label-based approach, it also considers the readability of labels and the appropriateness of labels against the relevant documents.
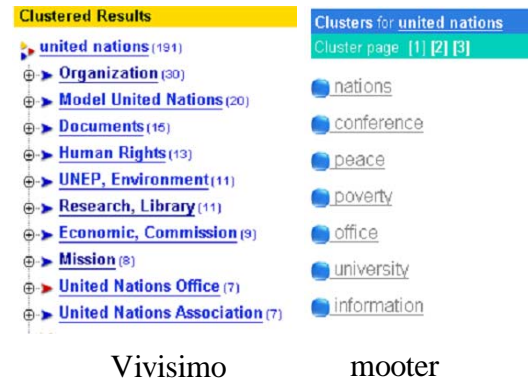
| Vivisimo | mooter |

Figure 1: Examples of Labels of Ordinary Systems

2.2 Named Entity Extraction

The NE extraction task was born in MUC[12] of the 1990's. It is the task of extracting the information units important to recognition like names, including people, organizations, and location names, and numeric expressions including time, date, money and percent expressions from unstructured text, for example newspapers[13]. Many NE tools and learning data have been developed for analyzing news articles but there are few NE categories. However, there are on-going efforts to extend NE extraction range. For example, one work expands NE category number from the original 8 kinds to 200[14], another work extracts the names of proteins or genes in the bio-informatics field[15]. Such enhanced NE extraction tools will broaden the application of our proposals.

Some research [16][17] tried to create a categorized vocabulary dictionary by using web mining. For example, one category is "programming language" and the vocabularies are "Java", "C", "Perl" and so on. If we can use these dictionaries, the coverage of NE extraction is widened which increases the application range of our proposals.

3. PROBLEM FORMULATION AND OUR APPROACH

3.1 Problem Formulation

This section takes the example of ordinary systems which employ the label-based approach (Vivisimo and Mooter) to formalize the problems common to such systems.

We challenged these systems with the query "United Nations" in English. The labels of the results are shown in Figure 1. Vivisimo allows labels to become mingled with general terms such as "Organization" or "Documents" and specific terms such as "Model United Nations" or "UNEP".

We examined the documents related to the above general terms, "Organization" and "Documents", to determine why these general terms were present in the labels.

First we address the example of "Organization". We found that this term is present in the names of various organizations. It seems that the frequency of such terms is increased for this reason, and so they are recognized as labels. The terms "Nations" and "Conference" presented by Mooter seem to be similar examples.

The above discussion leads to us to conclude that ordinary systems suffer from the following three problems.

- *problem 1:* Term extraction errors
  When the system extracts terms as label candidates, it extracts many ineffective terms while failing to extract important terms; the appropriate terms may not be extracted.
- *problem 2:* Label selection errors
  Inappropriate terms are selected as labels. Meaningless terms with high frequency tend to be extracted as labels.
- *problem 3:* Ineffective overview of search results
  If the labels are simply enumerated in the label list, we can't efficiently find the desired label and can't overview the labels and the results effectively.

3.2 Proposed Method

3.2.1 Solutions

To solve the *first problem*, we propose to use the NE extraction results as label candidates. An NE extraction tool can accurately extract proper nouns from unstructured text [3]. This technology was developed to mirror the characteristic that "people noticed that it is important to recognize information units" from sources such as newspaper articles[13] and they are also important when searching for news articles. As mentioned in Section 2, the coverage of NE extraction is widening, so our method is not just for news articles retrieval but for retrieving any other articles, for example weblog searching, EC site searching, and Web searching.

For *problem 2*, we consider the properties of efficient labels and propose new criteria to select the labels. We detail this in Section 4.1.

For *problem 3*, we propose to categorize the labels in the index. In this way, we can group the kinds of labels using the categories given by NE extraction, which allows the user to easily overview the search results. Furthermore, to ensure that the index provides not only easy overviewing but also efficient document location, we define a category ranking criterion in Section 4.2.

3.2.2 Algorithm

The algorithm that makes an index of the search result is shown here. We consider that all documents are first registered with our system. Next, the terms(NEs) are extracted in pre-processing. When our system accepts a query, the system uses the following algorithm.

1. Fetch search result
2. List the NEs in the search result
3. Select the labels from listed NEs
4. Organize the labels by NE category

At first, we fetch the search result. Second, we list the terms(NEs) that are extracted from the documents in the documents list. Each NE has its category information.

In the third process, we first calculate the score of each NE using the label selecting criterion. The terms that have high score of this criterion are then selected as labels. More precisely, the labels that construct similar clusters are combined using the similarity of clusters and labels; we will elaborate this point in another paper.

In organizing the labeling process, the labels are organized by the categories given by NE extraction.

---

[3]Though NE extraction tool extracts numeric expressions as well as proper nouns, we use only proper nouns(person, organization, location, artifact name) in this paper.

4. DETAILS OF OUR APPROACH

4.1 Label Selecting Criterion

Labels that can provide a good overview of the results and document location are terms that are important in the search result. Accordingly, criteria based on "Frequency" and "TF-IDF" are often used.

The criterion of "TF-IDF" is given by the next equation.

$$I_i^{TF-IDF} = TF_{R,i} \times \log(\frac{|D|}{DF_{D,i}})$$

I_i^A is the importance of term i, defined using the A method, and $TF_{R,i}$ is the frequency of term i in the search results $|R|$. $DF_{D,i}$ is the frequency of documents that include term i in document collection $|D|$. Document collection $|D|$ consists of the documents registered with the search system. A preliminary investigation showed that when $TF_{R,i}$ is replaced by $DF_{R,i}$, which is the frequency of documents that include term i in search result $|R|$, the same result is achieved according to the evaluation using the method mentioned in Section 5. Accordingly, in this paper we use $DF_{R,i}$ instead of $TF_{R,i}$. We call $DF_{R,i}$, RDF(Retrieved Document Frequency), and the name of this criterion is "RDF-IDF".

TF-IDF is basically the criterion that weights the index terms in the full text search index and is based on the idea that a term that is common in the document but rare in the document collection is important. We call the former the "Local Factor(LF)" and the latter the "Global Factor(GF)".

Hereinafter, we consider the significance of the labels and propose new Local and Global Factors. We evaluate the proposed label selecting criterion and consider its properties in Section 5.

4.1.1 Local Factor

According to empirical knowledge, the TF-IDF criterion is strongly influenced by term frequency and does not adequately reject high frequency terms[11]. In RDF-IDF, RDF has the same tendency as TF in TF-IDF. Therefore, we propose to use either of the following factors instead of simple frequency.

One factor is based on logarithmic frequency. To avoid TF's high frequency problem, logarithmic TF is often used in index term weighting. It is calculated by the following equation.

- Logarithmic Retrieved Document Frequency (LRDF)

$$LF_i^{LRDF} = \log\left(1 + DF_{R,i}\right)$$

Since we consider that a term that is useful in overviewing the results and efficiently locating the desired documents will not be too rare or too common, we propose the following criterion.

- Original LF(OLF)

$$LF_i^{OLF} = DF_{R,i} \times \log(\frac{|R|}{DF_{R,i}})$$

These functions are graphed in Figure 2. Both of the proposed factors assign less weight to high frequency terms than RDF based on TF. Furthermore, the latter factor's weights decrease strongly at higher frequencies.
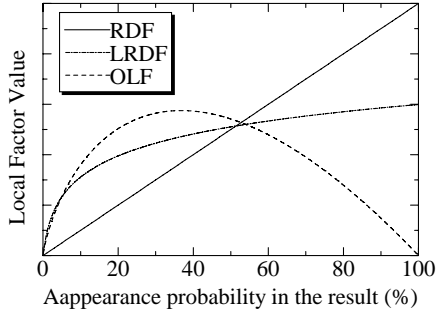
Figure 2: Form of Each Local Factor Function

4.1.2 Global Factor

Next we introduce the Global Factor. TF-IDF uses document frequency in the collection and its criterion is based on the idea that "important terms are rare in the collection".

We consider here the difference between index term weighting and weighting for label selection. In index term weighting, we consider the entire text collection and each document of the collection, and we weight the terms in each document. On the other hand, for weights used for label selection, we consider the entire collection and the document set returned by the search query; we weight the terms in the document set filtered by the search query.

Accordingly, we consider that "terms that are related to the query are useful as labels".

According to this argument, our Global Factor uses the following term appearance ratio.

- Original GF(OGF)

$$GF_i^{OGF} = \frac{DF_{R,i}/|R|}{DF_{D,i}/|D|}$$

In Section 5, we evaluate the criterion that uses these factors.

4.2 Category Ranking Criterion

This criterion ranks label categories in terms of their usefulness in not only easy overviewing of the labels but also efficient document location.

Though category ranking is influenced by labels, in this paper, we decide that labels are selected by the above criterion. We assume that the number of labels is about 20, because it's difficult for users to handle more than 20 labels.

Takata et al.[18] proposed the following criteria which rank categories (Each category is single layer.). Their research focused on categorizing the search result by using a category dynamically selected from a set of predefined categories.

- *Clearness of category:* Paucity of overlapping documents between each label.
- *Equality of category size:* Equality of the number of the documents related to each label.

The above criteria were developed to suit non-overlapped clustering. We, however, are tackling overlapped clusters, and so they can't be used as is. Hence we change the meaning of "Clearness of category" to "fewest overlapping documents between each label" and the following criterion is added.

- *Exhaustiveness of category:* The percentage of the documents in the search result that can be labeled.

In this paper, we integrate these criteria to realize category ranking. The above criteria are given by the following equations.

$$p_j^{Clearness} = |D_j|/\sum_{i \in C_j}(|D_{j,i}|)$$

$$p_j^{Equality} = \sum_{i \in C_j}(-\frac{|D_{j,i}|}{\sum_{i \in C_j}(|D_{j,i}|)} \times \log(\frac{|D_{j,i}|}{\sum_{i \in C_j}(|D_{j,i}|)}))$$

$$p_j^{Exhaustiveness} = |D_j|/|R|$$

Here, p_j is the ranking score of category j, C_j is the label list in category j, D_j is the aggregation of documents related to any label in C_j,and $D_{j,i}$ is the aggregation of documents related to label i in

5. EVALUATION

In this section, we evaluate the above label selecting criterion and the efficiency of label presentation method.

5.1 Prototype System for Evaluations

We implemented a system that applied the above propositions. We used LISTA[19] as a full text search system and Isozaki's NE extraction tool[20].

5.2 Evaluation Resource

In evaluating the label selecting criterion, we used the Japanese newspaper collection of IREX[21]. The collection covers 2 years(1994 and 1995) and holds about 200,000 articles. We used 30 search topics and the relevance judgment data for the topics as defined by IREX.

Queries to the system were constructed using DESCRIPTION[4]. We also created the "Useful Keywords List", a collection of useful keywords to overview the relevant documents for each topic[5].

In addition, we used the latest online Japanese news article collection to evaluate the label presentation method. This collection is not fixed and consist of the last one month of articles. The size of this collection is about 15,000 articles.

5.3 Evaluation Method

5.3.1 Evaluation of Label Selecting Criterion

We consider that the label selecting criterion should be estimated from two viewpoints. One is whether users would select the labels. This is related to label readability and meaning. The other is whether the labels are related to the relevant documents. Accordingly, our evaluation basically examined the accuracy of the search results using the labels that can be assumed to be selected by the user. The labels that also exist in the "Useful Keywords List" are specified as labels selected. The accuracy is calculated by the average precision of the top p results when the labels are selected. Here, we set $p = 10$ because not so many results are examined by most users.

---

[4]DESCRIPTION is defined for each topic; each expression consists of 2 or 3 nouns. We extract the meaningful terms from DESCRIPTION and link them by the "or" operator.

[5]To make this list, we showed each set of relevant documents and a topic to 5 subjects, and each subject selected the most suitable keywords. Keywords that were selected by more than 3 subjects were added to the keyword list. The result was a "Useful Keywords List" for each topic.

Table 1: Label Selecting Criteria

Method ID	Local Factor	Global Factor
FREQ	RDF	—
RDF-IDF	RDF	IDF
LRDF-IDF	LRDF	IDF
OLF-IDF	OLF	IDF
RDF-OGF	RDF	OGF
LRDF-OGF	LRDF	OGF
OLF-OGF	OLF	OGF

In this evaluation, we estimate the criteria shown in Table 1. Criteria are combinations of Local Factor and Global Factor, which were proposed in 4.1. Each criterion is calculated using the next equation.

$$I_i = (LocalFactor) \times (GlobalFactor)$$

When we generate labels, we analyze the top m ranked documents of the search result, calculate the label candidate's score using one of the above criterion, and select the top 20 (according to score) label candidates as labels.

We compared the proposed method to a simple method that uses only RDF(FREQ) and a method based on TF-IDF(RDF-IDF).

We use macro average precision as the evaluation metric.

In this evaluation, we evaluate the contents(labels) of "index" when we use NE extraction.

5.3.2 Evaluation of Label Presentation Method

To evaluate the new label presentation method using NE category, we prepared a system that generated two kind of labels. One adopted the approach proposed in this paper (System 1), this system divides the labels into categories, while the other did not group labels using category (System 2). Both used the OLF-OGF label selecting criterion described in Section 5.3.1.

We conducted experiments with two collections and different subjects. One experiment used 12 adult subjects experienced in using web search Engines and the IREX collection. Each subject searched for 10 topics using each system. In each search trial, the user searched until he found all results or 30 minutes had passed. The subjects were divided into 3 groups and each group had a different set of topics. In more detail, each group was divided into 2 sub-groups; each used one half of the topics (a different half) to challenge each system. Each subject was required to search for 5 topics on one system in one day. After this experiment, the 12 subjects answered the following questionnaire.

- Which method is more useful?
- What are the advantages and disadvantages of each system? (Answer is free text.)

The other experiment was run on the Internet. The contents were the latest Japanese news articles. We slightly modified a search system to be able to activate either System1 and System2 by one button and set the questionnaire on the interface. The questionnaire was just the first of the two above questions, "which method is more useful?" The response required the use of one radio button and one submit button. We used Cookies to eliminate duplicate voting.

In this evaluation, we evaluated the structure of "index" when we used NE extraction.
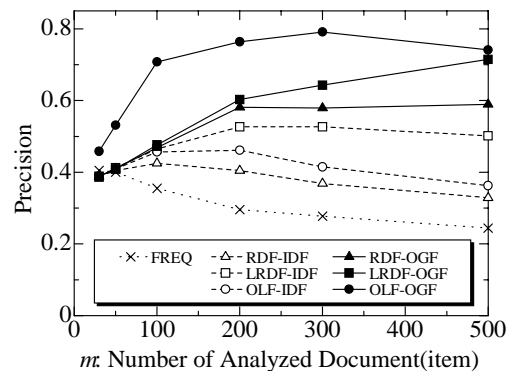
Figure 3: Evaluation of Label Selecting Criterion

5.4 Results

5.4.1 Evaluation of Label Selecting Criterion

The results are shown in Figure 3. The vertical axis plots the average precision and the horizontal axis plots the number of analyzed documents (m) used in selecting the labels.

The criteria that use OGF offer generally higher precision than those that use IDF. On the other hand, LRDF and OLF have higher precision than ordinary RDF without regard to the Global Factor. This result indicates that the hypothesis proposed in Section 4.1 is correct.

We also note that OLF offers higher maximum precision with relatively fewer processed documents than LRDF, when we use OGF. On the other hand, LRDF offers higher precision than OLF when we use IDF.

To conclude this evaluation, OGF, which we proposed as a Global Factor, greatly increases the accuracy with which the search results can be winnowed and OLF and LRDF, which we propose as Local Factors, yield higher accuracy than ordinary RDF. Thus it can be said that OGF-OLF and OGF-LRDF are efficient as label selecting criteria. Moreover, the criterion based on LRDF requires more documents to be analyzed than is true with OLF.

The criterion based on TF-IDF(RDF-IDF) yielded lower accuracy in the evaluation; its performance approached that of simple frequency(FREQ). It seems to basically match empirical knowledge as shown in Section 2.

5.4.2 Evaluation of Label Presentation Method

The result of this questionnaire "Which method is more useful?" is shown in Table 2. In both results, the proposed method was said to be more useful than the ordinary label list by more than 80% subjects/users.

In response to the question, "What are the advantages and disadvantages of each system?", 5 subjects replied that the advantages of system 1 were "the fact that the labels were categorized by NE" while the remaining 7 subjects said "the fact that labels were provided". On the other hand, the noted disadvantages of system1 included "it is hard to identify conceptual context", "no solution is possible if I can't

Table 2: Result of Questionnaire

experimant condition	percentage of the people who said proposed method is better
subject experiment	83.3%(10/12)
experiment on the Internet	85.6%(113/132)

Figure 4: Examples of index of search result

third is label categorization using NE category information of labels. Evaluations indicated that the proposed methods are more useful than the current methods. Our evaluation used only Japanese newspaper articles, but our method is not language specific and is not content specific so it could be used to handle other languages and contents.

Note that we did not evaluate the Category Ranking Criterion and this remains as a future task. Furthermore, we will investigate the problems with search result clustering. First is that we should try to combine NE labels with other labels(not proper nouns). Second, we should consider other NE tools and dictionaries as argued in Section 2.

comprehend label meaning", "the same labels are presented in several categories (mis-extraction by the NE tool)" and "Sometimes, the label desired does not exist". Common complaints about systems 1 and 2 were "I prefer the result to be ordered by time (the system sorted the document by relevancy of query.)" and "a previously selected link shows no visual change".

The disadvantages of system 2 included "the fact that the labels were not categorized" (6 subjects) while 4 subjects criticized label quality. The remaining 2 subjects indicated the common problems. As to its advantages, 11 subjects noted "the fact that labels are provided"; the remaining user entered a negative comment in response to this question.

To summarize, 8 of 12 subjects said that "System 1 is better at facilitating searches because it categorizes labels" or "System 2 is is worse at facilitating searches because the labels are not categorized".

According to questionnaire results, we find that the subjects/users feel that label categorization is useful.

5.5 Examples

In Figure 4, we present an index for the search result gained when we use the keyword "president bush" to access the latest news collection[6]. The left list represents the index generated by our system, The right one represents the labels generated by Clusty[7] which uses vivisimo's clustering technique. According to these examples, the index generated by our system provides easy overviewing and efficient document location. However, current NE approaches do not cover all topics, so we should try to mix NE and other terms as labels.

6. CONCLUSION

In this paper, we regard the clustering task as constructing an index of search results. To realize this goal, we make three proposals. First is to use NE Extraction for term extraction. Second is a new label selecting criterion. The

---

[6] The search target of our system is Japanese news articles and that of Clusty is English news articles. So, when we use our system, the queries are translated to Japanese.

[7] http://www.clusty.com/

7. REFERENCES

[1] Belkin, N. J.: "Anomalous states of knowledge as a basis for information." *Canadian Journal of Information, Vol. 5, pp.133-143,* 1980.

[2] Brin, S. and Page, L.: "The anatomy of a large-scale hypertextual(Web) Search Engine." *Proceedings of WWW7, pp.107-117,* 1998.

[3] Salton, G. and Yang, C. G.: "On the Specification of Term Values in Automatic Indexing." *Journal of Documentation 29,* 1973.

[4] Baeza-Yates, R. and Ribeiro-Neto, B.: "Modern Information Retrieval." *ACM Press,* 1999.

[5] Zamir, O., Etzioni, O. and Grouper, A.: "Grouper: A Dynamic Clustering Interface to Web Search Results." *Proceedings of WWW8, pp.1361-1374,* 1999.

[6] Zeng, H. J., He, Q. C., Chen, Z., Ma, W. Y. and Ma, J.: "Learning to Cluster Web Search Results." *Proceedings of SIGIR'04, pp.210-217,* 2004.

[7] Kummamuru, K., Lotlikar, R., Roy, S., Signal, K. and Krishnapuram, R.: "A hierarchical monothetic document clustering algorithm for summarization and browsing search results." *Proceedings of WWW'04, pp.658-665,* 2004.

[8] Ohta, M., Narita, H. and Ohno, S.: "Overlapping Clustering Method Using Local and Global Importance of Feature Terms at NTCIR-4 Web Task." *Working Notes of NTCIR(NII-NACSIS Test Collection for IR Systems)-4 Vol. Supl. 1, pp.37-44,* 2004.

[9] Hearst, M., and Pedersen, J.: "Reexamining the cluster hypothesis: scatter/gather on retrieval results." *Proceedings of SIGIR'96, pp.76-84,* 1996.

[10] Leuski, A.: "Evaluating Document Clustering for Interactive Information Retrieval." *Proceedings of CIKM'01, pp.33-40,* 2001.

[11] Hisamitsu, T., Niwa, Y. and Tsujii, J.: "Measuring Representativeness of Terms." *Proceedings of IRAL'99, pp.83-90,* 1999.

[12] Grishman, R. and Sundheim B.: "Message Understanding Conference - 6: A Brief History." *Proceedings of COLING'96, pp.466-471,* 1996.

[13] Sekine, S.: "Named Entity: History and Future." *http://cs.nyu.edu/s̃ekine/papers/NEsurvey200402.pdf,* 2004.

[14] Sekine, S. and Nobata, C.: "Definition, Dictionary and Tagger for Extended Named Entities." *Proceedings of LREC'04,* 2004.

[15] Kim, J. D., Ohta, T., Tsuruoka, Y., Tateisi Y. and Collier, N.: "Introduction to the Bio-Entity Recognition Task at JNLPBA." *Proceedings of JNLPBA-04. pp.70-75,* 2004.

[16] Shinzato, K. and Torisawa, K.: "Extracting Hyponyms of Prespecified Hypernyms from Itemizations and Headings in Web Documents." *Proceedings of COLING'04,* 2004.

[17] Pasca, M.: "Acquisition of Categorized Named Entities for Web Search." *Proceedings of CIKM'04, pp.137-145,* 2004.

[18] Takata, Y., Nakagawa, K. and Seki, H.: "Flexible Category Structure for Supporting WWW Retrieval." *Proceedings of 2nd International Workshop on the WWW and Conceptual Modeling, pp.165-177,* 2000.

[19] Hayashi, Y., Tomita, J. and Kikui, G.: "Searching text-rich XML documents." *ACM SIGIR Workshop on XML and Information Retrieval, pp.27-35,* 2000.

[20] Isozaki, H. and Kazawa, H.: "Efficient Support Vector Classifiers for Named Entity Recognition." *Proceedings of COLING'02, pp390-396,* 2002.

[21] Sekine, S. and Isahara, H.: "IREX Project Overview." *Proceedings of the IREX Workshop, pp.7-12,* 1999.