

Effects of the Variety of Document Retrieval Methods on Interactive Information Access

– An Experiment in the NTCIR-9 VisEx Task –

Tsuneaki Kato
The University of Tokyo
3-8-1 Komaba Meguro-ku
Tokyo, Japan
kato@boz.c.u-tokyo.ac.jp

ABSTRACT

An experiment was conducted by participating in the NTCIR-9 VisEx task, which examines how different document retrieval methods provided by an information access environment influence the results and process of interactive information access. The experiment compared the baseline system, which only provides a common keyword-based retrieval method, with our experimental UTLIS system, which provides, in addition to keyword-based retrieval, narrowing-down of obtained documents by specifying related place names and publication dates, and similarity-based retrieval. A preliminary analysis of the data shows that in the VisEx task, these retrieval methods changed the process of interactive information access without significantly affecting the amount and quality of information obtained.

Categories and Subject Descriptors

H.3.0 [Information Storage and Retrieval]: General

General Terms

Experimentation

Keywords

evaluation, empirical user studies, interactive information access

Team Name

UTLIS

Subtasks/Languages

Event Collection, Trend Summarization/Japanese

External Resources Used

None

1. INTRODUCTION

The activities of seeking necessary information and acquiring knowledge from it are inherently interactive and explorative. In daily life, people employ several methods in combination as well as trial and error. This is also the case in document searches using information retrieval systems; keyword-based retrieval, which is attracting much attention in the field of information retrieval research, is just one of several possible methods. Moreover, such retrieval is used in a dynamic process of changing the keywords as the process proceeds based on the results of previous retrievals. Thus,

the analysis and evaluation of interactive and explorative information access are much more complex than one-shot keyword-based information retrieval which has been studied mainly so far.

The VisEx task, which was conducted as a pilot task of the NTCIR-9 workshop, is a trial to establish a framework for evaluating such interactive and explorative information access. In this term, various data were collected on the behavior of users engaging in information access and on the supporting environment, by conducting experiments involving human subjects. We are interested in the relation between the various document retrieval methods available to users and the process of interactive information access, and therefore participated in VisEx. In VisEx, the baseline system provided by the organizers for reference purposes supports only a common keyword-based retrieval method. Would a more versatile system that also offers narrowing-down of obtained documents and similarity-based retrieval change the effectiveness and/or process of information access? And if so, how do these characteristics change? We set out to examine these issues.

The structure of the paper is as follows. First, in Section 2 we discuss the motivation of our study in more detail. Section 3 then explains our experimental UTLIS submitted system to VisEx. The data obtained is analyzed in Section 4, and its meanings and implications are discussed in Section 5. Finally, some conclusions are drawn in Section 6.

2. MOTIVATION

Bates examined human information seeking behavior and showed that people used several methods that differ from keyword-based retrieval in document searches. They sometimes find a new document by following a list of references, or by looking at nearby books on the shelves in a library[1].

In document retrieval using information retrieval systems, the main method of information search has been keyword-based retrieval, and researches in this field have focused on achieving high precision and recall. Current information access environments, however, are more diverse. In web searches, since documents are related to each other through hyperlink connections, it is possible to browse or surf documents by following those connections. The latest web search engines consider the diversity of results and provide a feature for facet-based search. Meanwhile, one study proposed a document search method without using keyword-based retrieval [2].

We are interested in the effects of various methods of document retrieval in interactive and explorative information access environments. Especially, we are keen to examine the effects of introducing methods that do not depend on keyword-based document retrieval. Would such methods improve the effectiveness of information access? What would occur or change in the process of information access by introducing such methods? To clarify these issues, we participated in VisEx.

The VisEx task is a trial to establish a framework for evaluating interactive and explorative information access environments. This term, it is focusing on conducting experiments using human subjects to carry out explorative tasks, such as asking them to investigate a topic and compile obtained information into a short report. These experiments allow us to collect various data such as reports, direct products of the task, log records of users' actions and environment system behaviors, and subjective impressions through questionnaires. Because the same information retrieval engine and knowledge-editing environment are used in all the experimental environments, the resulting data can be compared in a uniform manner. These characteristics are suitable for our research objective, and by providing a system that offers additional retrieval methods to keyword-based retrieval, we can effectively compare its behavior with that of an ordinary keyword-based retrieval system.

The baseline system provided by the VisEx organizers for reference purposes is an ordinary keyword-based retrieval system for newspaper articles. It allows users to retrieve articles using a keyword-based retrieval method and to sort obtained articles by relevance score or publication date. The resulting articles are listed with headlines and snippets shown; by clicking the headline of an article, a user can examine its text body. One result list page contains ten articles, and a page of the next or previous ten articles can be displayed by clicking a button.

For our experiment, we constructed a system with two additional document retrieval methods. The first method is a kind of facet search and allows users to narrow down obtained documents by specifying a place related to the documents and publication date of the documents. The second method is a similarity-based search, which allows users to retrieve documents similar to a specified one.

These two methods are basic and general, and are supported by Apache Solr, an information retrieval engine that VisEx decided to use in its experiments. The methods are also interesting from the following perspectives. The approach of narrowing-down by place and time related to the articles could be developed into a visual interface that uses maps and time sliders[4]. Although our narrowing-down method in this experiment offered only a menu-based interface, not a visual interface, the results could be useful for considering ways of combining a visual interface with a traditional keyword-based search interface, and for providing guidance on designing the visual interface. A similarity-based search is based on connections among articles by their similarity, and allows users to move directly from one article to another without using indexes, and so is similar to browsing web documents. It is interesting to examine the effects of adding such a document structure and a browsing-like method on the processes of accessing information.

VisEx conducted two subtasks: an event collection subtask and a trend summarization subtask. Our system par-

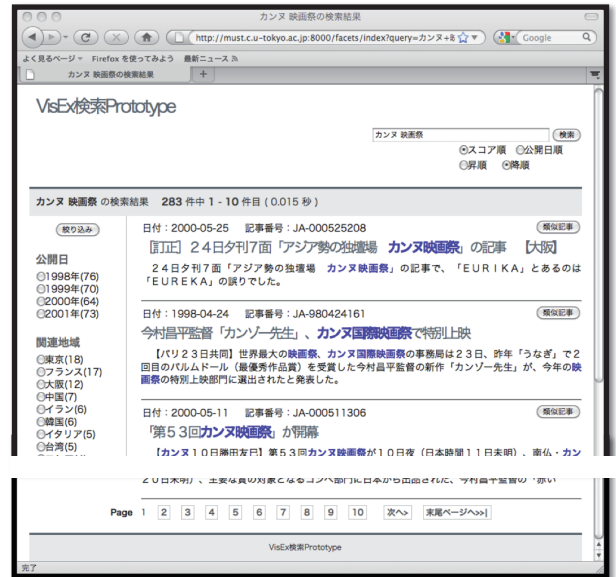


Figure 1: A snapshot of the UTLIS system

ticipated in both subtasks in order to investigate how differences in task characteristics would affect the use of the additional retrieval methods introduced to our system.

3. THE SUBMITTED SYSTEM

Figure 1 shows a snapshot of the submitted system, the UTLIS system. It has a similar look and feel to the baseline system. Menu lists and buttons for narrowing-down retrieval are shown in the left section of the window, where time periods are listed in the upper part, and place names are listed in the lower part. By selecting a radio button on the left of a list of items and then clicking the “narrowing-down” button at the top, the user can specify conditions to narrow down the current selection of articles. During the narrowing-down operation, the title of the result list shows its status, namely that the results have been narrowed down. In addition, there is a radio button for selecting whether to take over the current narrowing-down conditions to the next keyword-based retrieval located under the query box, which is located in the upper right of the window; the narrowing-down conditions are not taken over by default. A similarity-based retrieval is carried out by clicking the button on the right of each listed article. In the result page, the selected article is listed at the top, and retrieved articles are listed below it in order of similarity. The user can neither narrow down the results of similarity-based retrieval nor sort them by publication date. Other functions are the same as those of the baseline system; the user can retrieve documents by a keyword-based method, sort them by score or date, and examine the body of an article by clicking the headline in the result list page.

The UTLIS system is implemented using python and Django, the same as the baseline system. For the design, we referred to examples of Apache Solr client programs as shown in [5]. Narrowing-down retrieval was implemented using the faceting function of Apache Solr, and similarity-based retrieval using the MoreLikeThis component.

Place names that related to an article are determined by tokenizing and POS tagging the headline and first 50 characters of the body using a Japanese morphological analyzer, MeCab¹, and then extracting words tagged as the name of a country other than Japan and the name of a prefecture. Some place names that show the local version of the article or source of information are eliminated, which appear in a special kind of parentheses. Names of regions, cities, mountains, and oceans are not included; only country names and prefecture names are extracted. No place name or more than one place name may be extracted from and related to one article. The menu list consists of the top 20 place names arranged in descending order of number of articles that include that place. Narrowing-down can be performed multiple times, and the process is the same as the first time. Thus, if a user narrows down by specifying the name “Russia” and then narrows down again by specifying the name “United States,” the result would be articles that contain both “Russia” and “United States” in the above sense.

Narrowing down based on publication date is carried out by referring to the publication date field of the articles. Since the document set consists of newspaper articles of four years, 1998 to 2001, the first menu list shows these four years as the menu items. When a user narrows down the articles to those of a specific year, then the list items change to the four quarters of that year. The finest-level list, which is shown when one more narrowing-down is carried out, consists of three months of a specified quarter.

4. DATA ANALYSIS

Data obtained through the VisEx experiments consist of reports, which are made by the subjects as direct products of the tasks, log records of information access behaviors of the subjects, and subjective evaluations collected by questionnaire survey of the subjects. This section discusses the results of the preliminary analysis of each type of data.

As shown in the overview report of the organizers[3], an analysis of the quantity of contents in the reports shows no apparent difference between the baseline system and the UTLIS system. Table 1 shows cosine measures between sets of articles collected and referred to in the reports, which were proposed by the organizers as a measure of similarity between systems in terms of their retrieval tendency. The similarity to the baseline system of the UTLIS system is the second highest next to the TM2011 system in the event collection subtask, and the highest in the trend summarization subtask, on average, though different tendencies are observed in some topics. The result is reasonable, since the TM2011 system provides just a keyword-based retrieval as the baseline system does, while the Grid system and the KN system provide unique retrieval methods respectively: one based on a two-dimensional array of keywords and one utilizing charts as a visual interface. This result indicates that the introduction of two additional document retrieval methods to the baseline system did not make a large difference, though some changes probably occurred.

Our log analysis is based on a simple model of the interactive information access shown in Figure 2. The result list is shown whenever a user conducts a retrieval related action, such as sorting, moving to the next/previous page of the result list, and going backward/forward, in addition

¹<http://mecab.sourceforge.net/>

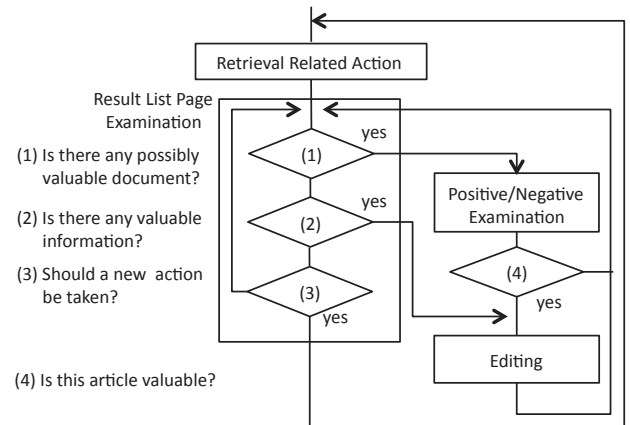


Figure 2: A model of interactive information access

to the provided retrieval actions of keyword-based retrieval, narrowing-down, and similarity-based retrieval. The user examines the result list for the following three points. (1) Is there any listed document that appears to contain valuable information? (2) Is there valuable information on the result list page itself? (3) Should a new retrieval related action be taken? In case of (1), the user accesses that article, examines the body, and then judges whether that article contains valuable information. If it does, she starts editing to record and compile the information found, and then returns to the result list page to continue examination. Otherwise she returns to the result page directory without editing. The body examination of the former case is called positive examination, while that of the latter is called negative examination. In case of (2), the user starts editing, moving directly from the result list page, and returns to the result list page after editing. Sometimes she may find aggregative information, such as few documents obtained through some keywords. In the experiment, article IDs were often copied from the result list page and added to the reports. In case of (3), she may decide to move to the next page after checking all items listed on that page, to try other keywords if the current result looks wrong, to conduct some narrowing down if too many articles were retrieved, to conduct a similarity-based retrieval for an article found in that page, and so on. As a result, another result list page is shown which she starts to examine. This loop is repeated until a satisfactory report is written or time runs out. A user is regarded as being in a corresponding state when she is carrying out time-consuming actions such as examinations and editing. Retrieval related actions are contrasted to those actions, and cause state transition. Another state is visiting external pages not provided by the environment or editor; this state is not shown in the figure since users can transit to that state from any state and then return to the same state.

Although this model is useful for preliminary research, it still needs some sophistication in order to overcome some problems, such as handling parallel examination using multiple windows and/or tabs, and comparative examination of more than one document.

The log analysis based on this model allows us to characterize the achievement of each task by drawing a chart such as the one shown in Figure 3. This clearly shows when the user resides in each state and for how long. A sequence

Table 1: Similarity to the baseline system in terms of retrieved articles

		1	2	3	4	average
Event	Grid	0.48	0.28	0.50	0.57	0.46
	TM2011	0.44	0.41	0.70	0.75	0.57
	UTLIS	0.54	0.28	0.67	0.67	0.54
Trend	Grid	0.76	0.60	0.20	0.70	0.56
	KN	0.50	0.35	0.41	0.78	0.51
	UTLIS	0.72	0.66	0.38	0.79	0.64

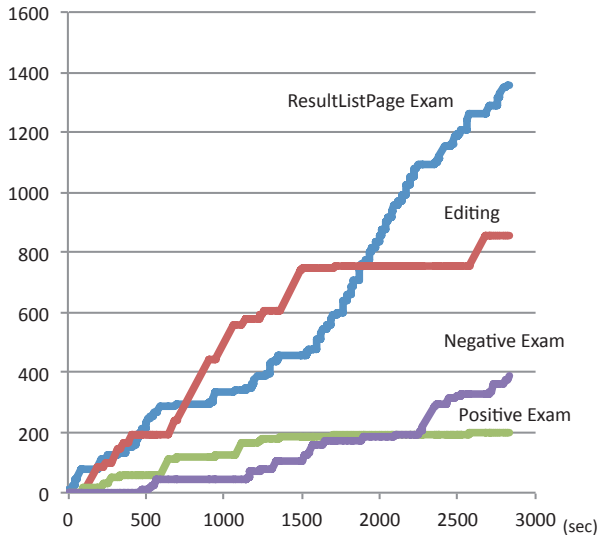


Figure 3: An example of state staying pattern

of retrieval related actions arranged on a time line could be overlaid on this chart, though this figure does not show those actions. It is worth noting that the editing time obtained based on this model differs from that defined in the overview paper, which is the duration for which the editor tab is active. The editing period of our model starts when a user activates the editor tab after positive examination, and ends when she moves back to the result list page. During this period, she comes and goes between the article and the editor. This definition is more precise than the one shown in the overview report.

There are several likely expectations. For example, if a given environment has a high precision of search and provides a more effective result list page such as one with more informative and easy-to-read snippets, the time of staying on the result list page examination and negative examination will be shorter and the editing time will be longer. A similar analysis may reveal the characteristics of topics. For example, if it is easy to find documents relevant to a given topic but the number of such documents is small, most of the time at the beginning of a search will be used for editing as many relevant documents are found, but thereafter more time will be spent on result list examination and negative examination, as the user has to search for documents that do not actually exist.

However, a comparison between the baseline system and the UTLIS system did not show a significant difference between those state transitions. As an example, Table 2 com-

pares the result list examination and editing. It shows the minimum, maximum, average and median values of those times grouped by each topic. These results correspond to the fact that there was no clear difference in the resulted reports. On the other hand, the distributions of retrieval related actions show large differences. Table 3 shows the number of times of keyword-based retrieval and page moves, that is, the sum of the number of times of clicking the next/previous button and specifying the number of pages. Both are much smaller for the UTLIS system than for the baseline system. Table 4 shows the number of times of using narrowing-down and similarity-based retrieval in the UTLIS system. While narrowing-down is used in both subtasks, similarity-based retrieval is used mainly for the event collection subtask.

Several kinds of information, such as transition patterns between states, changes of keywords used, and conditions of narrowing-down were obtained from the log records. Although the results are still being analyzed in detail, some preliminary impressions gained from the examination are discussed in the next section.

Among the questionnaire results, which indicate the impressions of users, evaluations of the systems are shown in Table 5, which are the averages for all the subjects on a seven-point Likert-type scale. The UTLIS system is clearly evaluated more highly than the baseline system, suggesting that the additional methods were well accepted.

5. DISCUSSION

The results of the experiments can be summarized as follows. The additional document retrieval methods of narrowing-down and similarity-based retrieval were used as alternatives to keyword-based retrieval and sequential examination of result list pages. The functions were well accepted, but had little influence on the quantity and quality of information obtained.

Upon closer examination, consider an example of searches in which the user makes much use of narrowing down. In the E2 topic in the event collection subtask, the user first conducts a keyword-based search using “nuclear power plant” and “accident” as the keywords, and then conducts narrowing down by specifying the names of related prefectures, such as “Ibaraki,” “Fukui,” and “Shizuoka,” in turn. As she returns to the original results after examining the result of each narrowing-down, this sequence of actions means sequential examination of the retrieved articles grouped by related places. After finishing this series of examinations, the user tries another keyword-based retrieval using different keywords.

Another example is as follows. In topic T2 in the trend summarization task, after conducting a keyword-based search using “the Cabinet” and “approval rating” as keywords, the user examines the result list in order of publication date

Table 2: Comparison of state staying time

(1) Editing (sec)

		1				2				3				4			
		min	max	avg	med	min	max	avg	med	min	max	avg	med	min	max	avg	med
Event	Baseline	1300	1721	1445	1433	419	1013	799	917	753	1781	1307	1252	724	1480	1149	1249
	UTLIS	1061	1639	1461	1557	854	1465	1050	1010	566	1667	1112	1146	1127	1702	1332	1269
Trend	Baseline	951	2223	1690	1758	890	2607	1568	1273	924	2277	1537	1422	545	2037	1232	1314
	UTLIS	955	2084	1736	1890	1184	1841	1553	1522	1344	1835	1617	1667	973	1564	1168	1099

(2) Result List Page Examination (sec)

		1				2				3				4			
		min	max	avg	med	min	max	avg	med	min	max	avg	med	min	max	avg	med
Event	Baseline	439	976	670	591	736	1535	1090	850	758	1420	1035	1050	705	1131	935	931
	UTLIS	355	975	615	593	874	1357	1066	994	1003	1713	1322	1431	589	1119	790	728
Trend	Baseline	141	849	551	653	41	833	426	462	385	947	764	844	551	1818	1087	1074
	UTLIS	213	637	416	377	432	834	569	522	309	970	619	693	580	1499	908	773

Table 3: Comparison of the number of retrieval related actions

(1) Keyword-based Retrieval

		1				2				3				4			
		min	max	avg	med	min	max	avg	med	min	max	avg	med	min	max	avg	med
Event	Baseline	4	24	11.2	5	22	37	30.6	34	1	35	13.2	12	5	17	11.8	15
	UTLIS	1	5	3.6	4	2	10	7.2	9	1	7	4	4	1	8	2.6	1
Trend	Baseline	6	17	9.8	8	1	14	6.4	2	2	29	13	13	10	40	24	24
	UTLIS	4	21	9.6	8	1	15	8.6	11	1	10	3.8	2	8	17	12	11

(2) Page Movement

		1				2				3				4			
		min	max	avg	med	min	max	avg	med	min	max	avg	med	min	max	avg	med
Event	Baseline	13	33	21.6	20	14	104	49.2	43	49	101	70.4	72	11	52	35.4	37
	UTLIS	0	26	9	6	0	62	29.6	36	14	125	59	39	10	39	22	17
Trend	Baseline	1	25	9	5	1	54	13	3	4	46	23	13	2	109	34	17
	UTLIS	1	7	3.6	2	3	17	8.6	7	0	45	17	8	6	68	24.8	14

using narrowing-down by publication date. In an extreme case, the user may conduct the keyword-based retrieval only once, and all other actions are narrowing-down.

If the user wants to search for the same pattern as the above using the baseline system, she repeats keyword-based searches with the keywords of “nuclear power plant,” “accident,” and “X”, where X is the name of a prefecture, for the first case, and she sequentially examines articles sorted by publication date using the next page button, page by page, for the second case. These correspondences explain the observation that narrowing-down was used as an alternative to keyword-based retrieval and sequential examination of result list pages. In fact, patterns similar to those were observed in the log records of the baseline system users. By using narrowing-down functions, users can carry out these patterns more easily. Especially in the former case, as options for narrowing-down are listed explicitly in the UTLIS system, users do not need to find them. Also, those actions using narrowing-down are, in a sense, active, which is surely preferable to just clicking the next page button. This ease of using the UTLIS system is probably why it received a higher evaluation.

We consider that narrowing-down supports a kind of systematic search in which the user sequentially checks articles obtained by keyword-based searches, according to the re-

lated place or date. It also seems to encourage users to conduct systematic searches. The fact that users using the baseline system sometimes changed the keywords in an un-systematic way supports this hypothesis. For example, one user used date or statistical value alone as a keyword after retrieving articles related to “birth rate,” while another user tried various keywords at random. However, a more extensive and quantitative analysis is required, since the discussion here might be based on subjective impressions.

It is also worth pointing out that the VisEx task allowed users to collect relevant information through a systematic search like the one mentioned above. Other roles of narrowing-down may be found in other types of tasks such as those that have few relevant documents and ask users to perform trial-and-error with various keywords.

As for the similarity-based search, the reason why it was used frequently only in the event collection subtask is probably that the unit of information in this subtask, that is, an event, corresponds to the unit of document, that is, one article. In this kind of task, the similarity-based search provides a way of obtaining more suitable documents. However, it is still unclear whether this retrieval method achieves a substantial improvement.

There is another aspect of introducing new methods to an information access environment. There were some com-

Table 4: The number of retrieval related actions of the UTLIS system users

(1) Narrowing-down retrieval

	1				2				3				4			
	min	max	avg	med	min	max	avg	med	min	max	avg	med	min	max	avg	med
Event	1	14	5	3	2	24	14.4	12	0	31	14.6	15	0	18	7.4	7
Trend	0	13	7.6	9	0	37	14.4	12	1	26	14.2	16	0	21	12	17

(2) Similarity-based retrieval

	1				2				3				4			
	min	max	avg	med	min	max	avg	med	min	max	avg	med	min	max	avg	med
Event	2	22	10.4	9	5	12	8.8	8	1	8	4.6	5	1	9	5.8	6
Trend	0	8	2	0	0	4	0.8	0	0	1	0.2	0	0	4	1.2	0

Table 5: Evaluations of the systems

		Usability	Functionality	Efficiency
Event	Baseline	5.8	4.0	4.0
	UTLIS	5.8	4.2	5.2
Trend	Baseline	4	3.4	3.2
	UTLIS	5	4.0	4.8

plaints in the final questionnaire about the low precision of additional document retrieval methods, and the inconvenience that the system does not support sorting or narrowing-down of the results of similarity-based retrieval. These findings suggest that introducing a new function does not necessarily improve the system as a whole. Especially, the second comment shows the importance of achieving a balanced overall environment. This is one of the most difficult parts of designing information access systems.

6. CONCLUSION

By participating in the VisEx task, we examined how the introduction of additional document retrieval methods changes the process of interactive information access and its results. Although findings on the situation and purpose of use were obtained, a more detailed analysis is needed. We will continue to analyze the data and may develop a new analysis method, in order to draw more convincing and quantitative conclusions. Our findings may well depend on the settings of VisEx, in which users could collect sufficient amounts of information using a systematic search approach for given topics. Similar studies in more explorative tasks are needed to clarify this point.

7. REFERENCES

- [1] M.J. Bates: The Design of Browsing and Berrypicking Techniques for the Online Search Interface. In *Online Information Review*, Vol. 13, No. 5, pp. 407-424, 1989.
- [2] D.R. Cutting, D.R. Karger et.al.:Scatter/Gather: A Cluster-based Approach to Browsing Large Document Collections In *Proceedings of 15th Annual International ACM SIGIR Conference*, pp. 318-329, 1992.
- [3] T. Kato, M. Matsushita, and H. Joho: Overview of the VisEx task at NTCIR-9. In *Proceedings of the Ninth NTCIR Workshop Meeting*, 2011.
- [4] B. Shneiderman: Dynamic Queries for Visual Information Seeking. In *IEEE Software* Vol.11, No.6., pp. 70-77, 1994.

- [5] H. Sekiguchi, J. Ootani, et.al.:*Introduction to Apache Solr (In Japanese)*, Gijutsu-Hyoron-sha, 2010.