Design of Object-based Information System Prototype

Suhyeon Yoo*, Sumi Shin**, Hyesun Kim***

ARTICLE INFO

ABSTRACT

Researchers who use science and technology information were found to ask an information service in which they can excerpt the contents they needed, rather than using the information at article level. In this study, we micronized the contents of scholarly articles into text, image, and table and then constructed a micro-content DB to design a new information system prototype based on this micro-content. After designing the prototype, we performed usability test for this prototype so as to confirm the usefulness of the system prototype. We expect that the outcome of this study will fulfill the segmented and diversified information need of researchers.

1. Introduction

As information distribution technology develops, researchers become able to reach more diversified and richer information resources through the Web. At the same time, excessive supply of information makes them invest more effort and time to search and find proper information. Especially, as flow of information in R&D of science and technology field is fast and dynamic, it is essential to acquire quick and accurate information.

Many attempts have been made to guess and fulfill the information needs of researchers. One of the traditional methods to find information is the keyword based search presently provided by portal sites such as Google, Naver, and Daum. In particular, the science and technology field information services provide documents with huge amounts of terminologies, which renders it difficult for the users to obtain search results reflecting their intentions using primitive queries. Furthermore, even if the search results were obtained, the general users have to reconstitute their query to find the document they want, or find documents through links in the related web pages, consuming both time and effort (Lee, Kim, & Choe, 2013).

Studies on subject librarian or subject specialist services are also being performed in various aspects such as operation plan of subject librarian (Chung, 2009), study on how to introduce and how to upbring and educate (Chung, 2007; Ahn et al., 2009; Noh, 2009). Subject librarian or subject specialist services can provide user-tailored information. She can help her library services are directed toward the needs of users and also be instrumental in developing and implementing new services, which proactively address the changing user needs. However, it still has several difficulties in recruiting of staff and lacking of in-service training programs (Agyen-Gyasi, 2008).

Personalized recommendation system is slightly different from traditional information retrieval systems or search engines. Recommendation systems identify the knowledge about the similar user or the event and derive the favorable aspect based on it. As the review paper of Akshita and Smita (2013), the criteria of "individualized" and "interesting and useful" separate the recommender system from information retrieval systems or search engines.

Despite those studies that are focused on the researchers' interest, researchers' needs do not seem fulfilled. Researchers still have difficulty in acquiring information they need among the information overload. Then, what information service is to be provided to support R&D of researchers searching for science and technology information more efficiently and effectively?

In January 2012, a survey (2012) was made by KISTI (Korea Institute of Science and Technology Information) to understand what information service science and technology information users actually need. This survey was made to study the need the users actually feel for function and contents requirement needed in R&D activities. In the result of the survey, what the information service users actually needed was found to be able to use only the part they want, such as research method, conclusion, and images from traditional science and technology information like research papers. This means a service that provides parts of an article, rather than providing information at article level.

Therefore, based on the needs of researchers according to the results of the survey, science and technology information service organizations are required to support R&D activities of researchers by providing more segmented information service based on object, rather than existing article level information service. For this purpose, a prototype of an information system that segment the full-text of an article and excerpts the desired part was designed and developed. The prototype service was designed by dividing text and image.

The significance of this prototype is to suggest a new framework of information service by realizing an information service at a micro-content level. This prototype allows using segmented contents based on micro-content, different from the existing service at article level. This prototype is expected to fulfill more segmented and diversified information needs of researchers and further, to actively support R&D activity of researchers. After developing the prototype, we evaluated the information system prototype to confirm the usefulness by performing the usability test of the prototype. We draw the positive and negative opinion from the six participants of the usability test. It is expected that a more satisfactory system would be developed as improving the issues drawn from the test. Therefore the result of this study is expected to be a corner stone for development of a new value added service using micro-content.

2. Conceptual background

There are, to my knowledge, no studies on splitting the full text into meaningful contents and searching the splitted contents except deep indexing system of ProQuest. ProQuest has been awarded a patent for its deep Indexing technology by the U.S. Patent and Trademark Office. However, as it is hard for any individual information service centers to develop the similar system, there is no known case practically. It is because of the tremendous workload for extracting and indexing metadata from full text. Fortunately KISTI has developed full text of journal article as XML format. The fundamental concept of the object-based information system prototype which this paper designed is from deep indexing and micro-content.

2.1. Deep indexing and Micro-content

Deep indexing system developed by ProQuest company is known as one of useful methods to surface relevant information that would be missed by other search methods. Sandusky (2008) defines deep indexing as an indexing system that supports discovery of information objects at levels of granularity beyond the abstract or article. Shortly deep indexing technique is an indexing method used after a journal article is published. More concretely speaking, it is to extract tables and figures from journal articles, index each table and figure, provide a retrieval method to locate tables and figures or complete articles that contain relevant figures or tables, and link them back to the article. Each table and figure extracted from a journal article is assigned index terms as appropriate for the type of table or figure (photograph, histogram, map, etc.), subject indexing, geographic indexing, taxonomic indexing, statistical indexing, and other relevant data using an automated indexing system. All tables and figures in an article are fully indexed and can be searched separately.

This study borrowed the concept of deep indexing system that extract tables and figures. This study extracted corpus of several sentences as well as table and figures. We define the extracted objects as micro-content.

The term of micro-content was first mentioned in a 1998 article of usability adviser Nielsen (1998). He referred to micro-content as small groups of words that can be skimmed by a person to get a clear idea of the content of a Web page. He included article headlines, page titles, subject lines and e-mail headings. Such phrases also may be taken out of content and displayed on a directory, search result page, bookmark list, etc. Another meaning of micro-content was defined by Anil Dash in 2002: "Today, micro-content is being used as a more general term indicating content that conveys one primary idea or concept, is accessible through a single definitive URL or perma-link, and is appropriately written and formatted for presentation in email clients, web browsers, or on handheld devices as needed. A day's weather forecast, the arrival and departure times for an airplane flight, an abstract from a long publication, or a single instant message can all be examples of micro-content". In summary, we can understand micro-content as a short content that delivers important idea or concept from the definition of Jakob Nielsen and Anil Dash.

In this study, we borrowed their concept of micro-content, however, the scholarly article was set up as the subject of contents, rather than a webpage. One paragraph of a scholarly article was

regarded as short phrases that deliver important idea or concept. Even though just one paragraph can deliver an important concept in a scholarly article, a group of paragraphs based on the table of content established by the author was defined as a micro-content. It is because context is important for researchers to understand the meaning of the article in scholarly articles. Along with micro-content focusing on the table of contents of scholarly article, tables and figures were also added in a category of micro-content.

2.2. Contents Clipping service

Contents Clipping service is the service name of the prototype designed to reflect the needs of information users. This means a service that allows users to excerpt, which is clip, only the parts they want to use from the article and use only parts of the contents, that is, micro-content. Especially, Contents Clipping service was realized for tables and images in this study. In particular, this prototype made micro-content DB from large amount of articles written in XML format. When the full-text is searched by a search engine, micro-content of the searched article is searched so that the optimal search result can be selected and compared. From the list of searched articles, items from table of contents of interest in similar articles can be compared, selected and exported.

The strength of the Contents Clipping service is that it can fulfill more segmented information needs of researchers through new information activities by extracting, that is, clipping by micro-content, which is the minimum unit of significance, comparing to other micro-content, and citing. In particular, in the case of deep indexing where figures or tables are used in search by extracting from the full text, its potential usability was found as follows.

Scientists identified many potential uses of tables and figures indexing to their work in both the observational sessions and diary entries. These potential uses include (Tenopir, Sandusky, & Casado, 2006).

- Teaching/lectures/presentations for which they would download figures directly into presentation software
- Locating and retrieving data in particular formats or particular object types
- Making comparisons between their work and the work of others
- Gaining faster and more precise understanding of the work reported in articles by direct examination of the tables and figures
- Assistance with writing of review papers, meta-analysis, proposals, and generating hypotheses
- Improving the efficiency of searching by providing more precise and smaller results sets
- Supporting the transformation of practice and supporting the learning of new skills and methods, including how to effectively present results in tables, figures, and graphs
- Use of objects by librarians to directly answer reference questions. Overwhelmingly, respondents said that the ability to search for specific types of objects would make a difference in their search and discovery processes.

3. Prototype development of Contents Clipping service

3.1. Micro-content Creation

   The subject of this prototype service was research paper in journal literature, the most basic type of traditional science and technology information. The prototype of Contents Clipping service was designed by dividing into article clipping service based on text, and image clipping service based on table and figure. Overall procedure as below was followed to construct the micro-content DB. The module that comprises article DTD parsing and import process consists of three parts. The first part realizes original text XML input module, the second part realizes XML parsing module, and the third part realizes the module that imports parsing result into DB. These three parts were realized purely based on java to construct a database. The relation between these modules can be schematized in Figure 1. To link micro-content database and to enable to service images consisting of micro-content in the best condition, following four modules were realized.
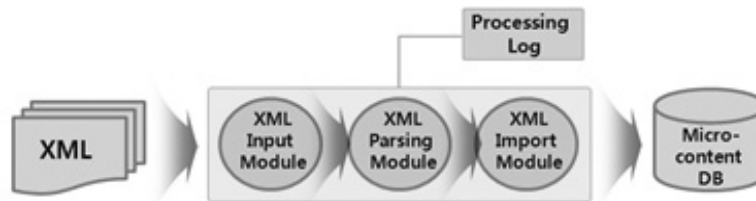
Fig. 1. Article DTD parsing model

   Firstly, to make the prototype in an expendable structure, external service link protocol was designed and achieved. The structure was designed so that the service request parameter can be defined and a new protocol is easily realized once the data structure to deliver is defined, as needed. Currently Oracle 11g is used for database, and link module was designed and completed to enable data inquiry or transaction data processing linked to this. In Article Clipping service, a module that resizes and provides image in a form right for service was designed and realized to make smooth image search service. Since one article is broken down and saved in a database as micro-article or image, these components should be combined as requested and provided when a service request is made. A module was designed and realized to do so.

3.2. Development of Article Clipping service

   The Article Clipping service is focused on the text-oriented micro-content. Main features of the Article Clipping service are like follows. First, at brief viewing of the search result, table of contents of each article is listed up, and if an item in the table of contents is selected, the full text in the item, the micro-content, is displayed like Figure Second, by listing up the table of contents of each article, researchers can figure out the overview of the article faster, and by directly seeing the full text by selecting a part of the table of contents, they can understand the part they need more intuitively.
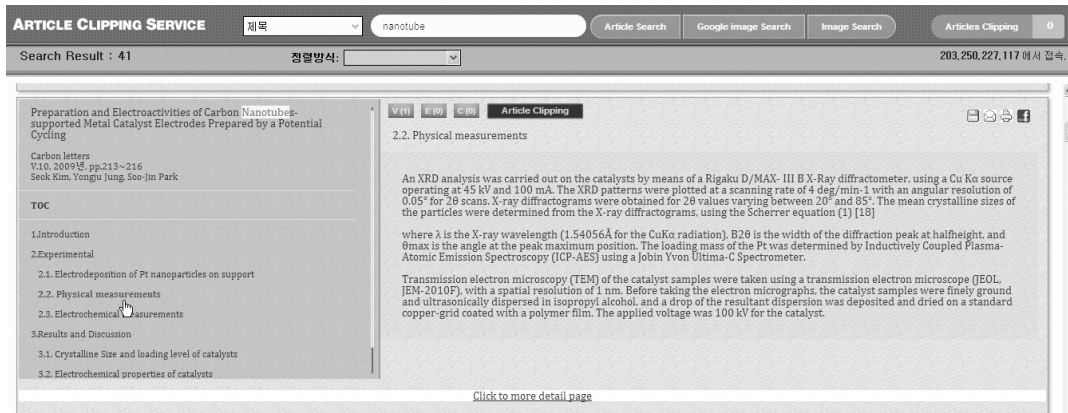
Fig. 2. Table of contents list up and full text viewing screen per item

Next, micro-contents of each table of contents can be compared and exported (Figure 3). With this feature, once the researcher checked the full text of the micro-content by table of contents, he or she can select the related contents from table of contents of other articles and clip them, compare clipped multiple micro-contents, display and export them as needed. Export options are file saving, sending an email, printing and sending to the representative SNS media or Facebook.
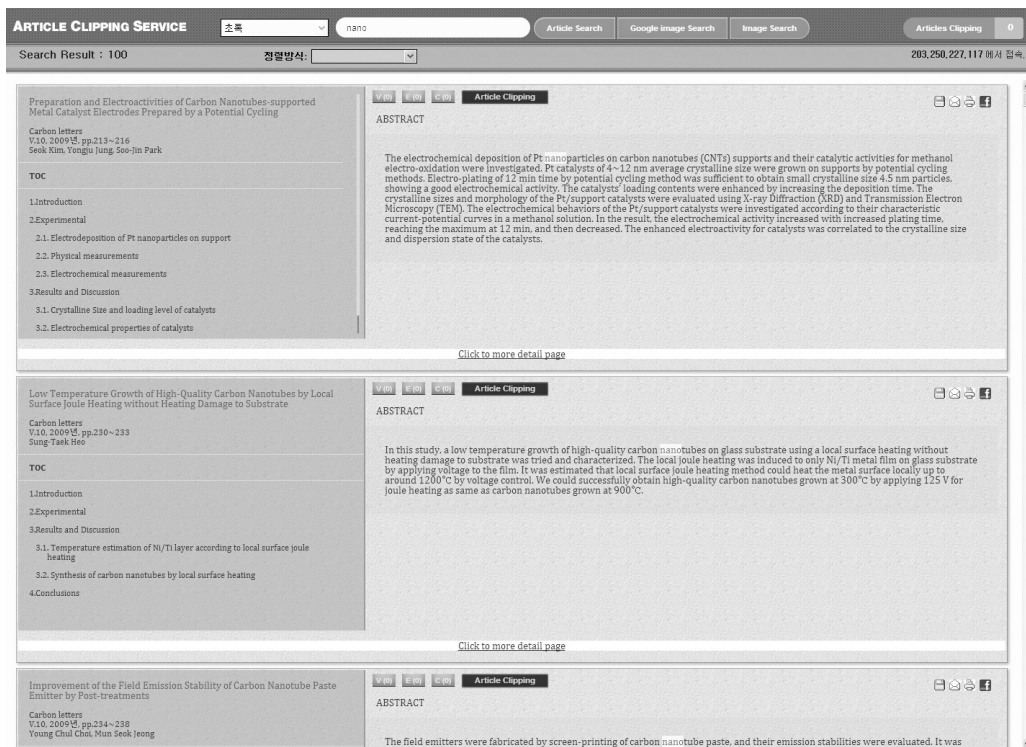

Fig. 3. Screen of comparison and export of micro-contents by table of contents

The third feature is to display by micro-content on detail viewing screen of one article. When an article shown as a search result is clicked, a detailed viewing screen of the relevant article is shown. On the detail viewing screen of the article, full text can be displayed by table of contents, and separated from full text, while figures and tables can be displayed in order of the table of contents. When a figure or table is selected, the full text to which the figure or table belongs is shown, so that the researcher can figure out more detail context. Figure 4 shows the detail-viewing screen of an article.

Fig. 4. Detail viewing screen of an article

3.3. Development of Image Clipping service

Image Clipping service is oriented to the tables, figures and images included in a scholarly article. Beside deep indexing system of ProQuest focuses on various components such as figure, tables, graphs, and so forth, Image Clipping service regards all kinds of components as equal object or image.

Main features of Image Clipping service using tables and images of an article are as follows. The first feature of the Image Clipping service is that the search results are provided from the keyword extracted from captions of images or tables of an article and searched images are grouped by their subject. Grouping of micro-content by subject suggests a pathway to access to contents by subject of interest of researchers (Figure 5). In this prototype, all articles were divided in advance into 10, using classes of DDC (Dewey Decimal Classification), according to the subject of the journal.
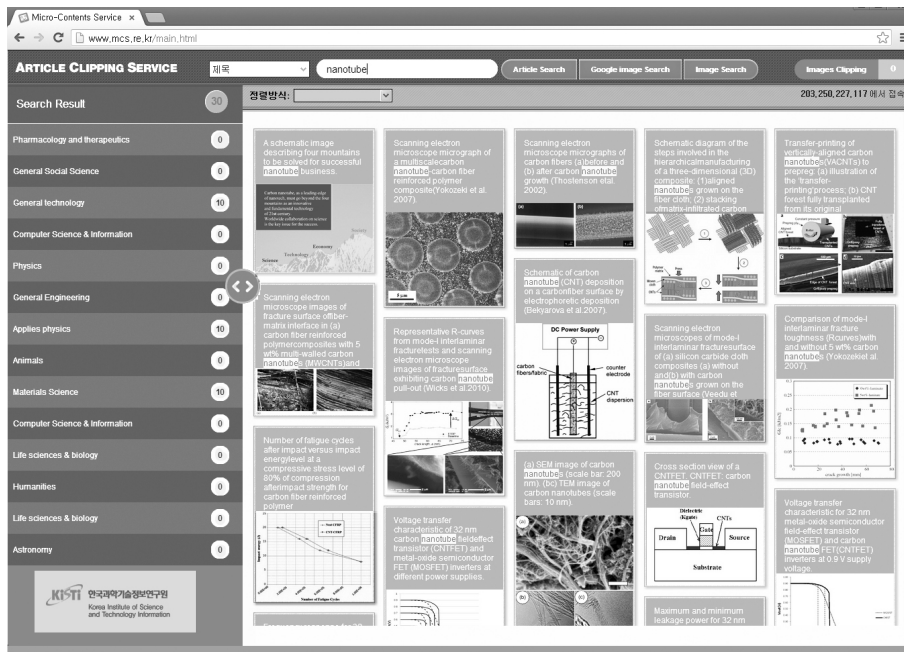
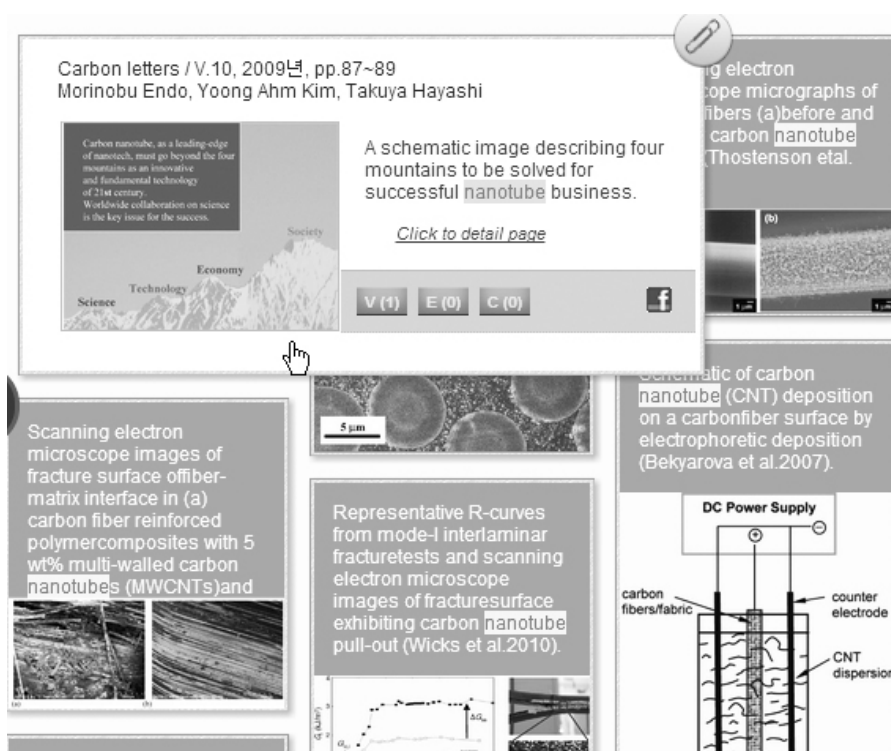**Fig. 5.** Feature of grouping by subject of image clipping service

Fig. 6. Zoom-in feature of micro-content

The next feature of the Image Clipping service is that, when a micro-content shown in a search result is moused over, the micro-content is zoomed in and related contents are shown, then it becomes a selectable, clippable condition (Figure 6). This allows faster and easier browsing by reducing procedures of clicking, checking contents, and closing the window. A micro-content in Image Clipping service consists of the figure or table, its publication information such as journal title, and caption of the figure or table. The search keyword is highlighted in the all micro-contents so as to discern the precision of the search results.

Third, zoomed in micro-content can be selected, that is, clipped, and multiple contents can be compared, displayed, and exported (Figure 7). This feature enables researchers to see multiple related images at the same time, and to export the micro-content. File saving, email sending, printing, and Facebook sending are available just like the Contents Clipping service.
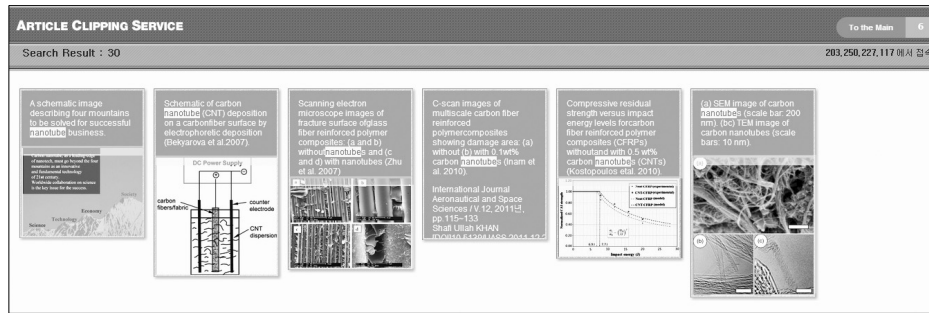

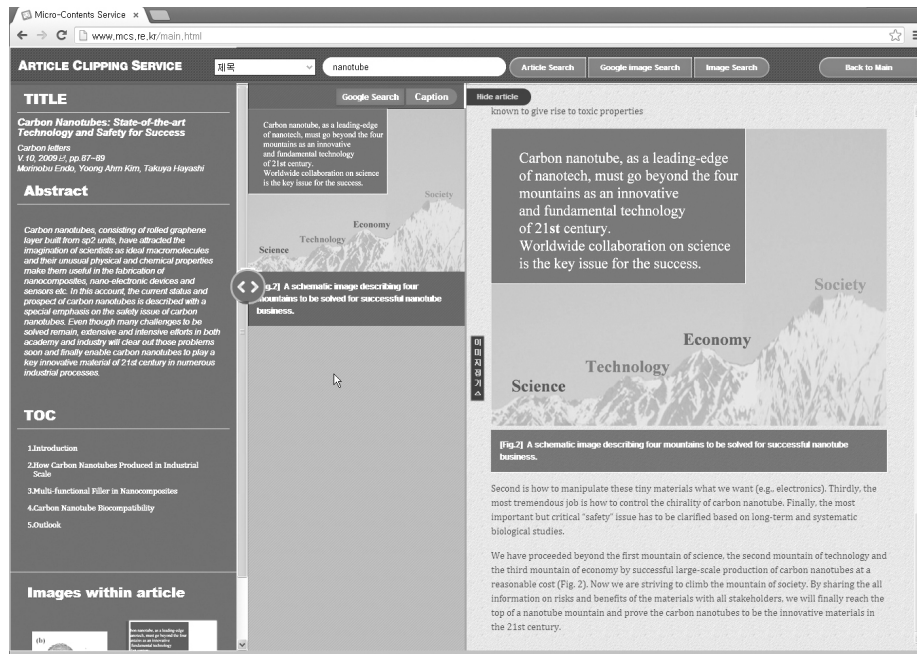**Fig. 7.** Screen of micro-content comparison and export

Fig. 8. Detail viewing screen of image micro-content

Finally, Image Clipping service shows contents like images related to a micro-content in full text (Figure 8). When a micro-content is clicked, screen changes to the detail view screen of the relevant content. Here, the caption that explains the contents is displayed and the screen changes to the location to which the content belongs in the article body. On the same screen, bibliographic data including abstract and TOC of the article to which the micro-content belongs are shown, and "Images within article" feature is provided so that the researcher can move to other images or tables within the article. Through this feature, researchers can browse other images in the same article together with the micro-content of interest.

4. Evaluation of the Prototype

4.1. Usability test

Generally usability test refers to evaluate effectiveness, efficiency, and satisfaction of the system. Jeong et al. (2013) summarized the service usefulness in view of efficiency, effectiveness, and satisfaction. They subdivided the service usefulness into functional quality and information quality. Zin and Yue (2013) divided the test of the History Digital Game Based Learning Software into effectiveness evaluation and usability evaluation.

In this study, we provided the four tasks to the six participants: task 1) Article Clipping service - screen of comparison and export by search, task 2) Article Clipping service - detail viewing screen of an article, task 3) Image Clipping service - screen of image comparison and export by search, task 4) Image Clipping service - detail viewing screen of an image. Positive feedback was presented while the participants of each task suggested several improvements. Their brief profile (Table 1) and testing site photo of each participant is as below (Figure 9).

Table 1. Participants profile

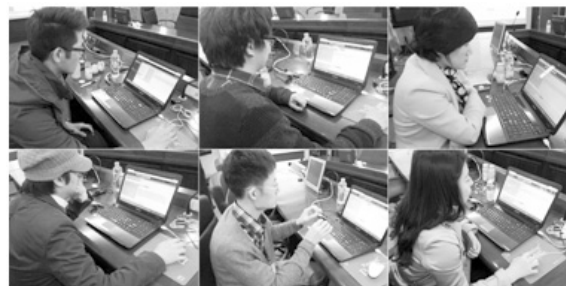| Participant	No.1	No.2	No.3	No.4	No.5	No.6
Gender	Male	Female	Male	Male	Male	Female
Age	25	31	31	35	38	43
Job	Student	Researcher	Research Assistant	Businessman	Research professor	University faculty

Fig. 9. Photo of usability testing site

4.2. Result of the usability test

As a result of the usability test, participants provided several positive opinions and suggested improvement issues. The positive opinions for each task follow. For the task 1, there were opinions that rough content can be figured out quickly through abstract during data search. Also, the participants said that it was convenient to be able to directly move from the table of contents page of the scholarly article to the relevant page of the article and be able to easily repeat a search later with the feature of bookmark provided through clipping. For the task 2, the screen composition was generally evaluated to be easy to understand and neatly arranged. Participants presented the opinions saying that contents of main text were clearly organized so that the viewers could browse the screen conveniently and a page that extracts the images of the article was useful to search the article. For the task 3, this feature was found to show searched images at a glance and it increased the interest on the relevant article. Also, caption of an image accompanied with the image made viewers understand the image immediately. Finally for the task 4, there were responses that it was convenient to see all images in the article at once.

On the other hand, participants proposed several issues for improvement of the prototype. First, participants suggested that V, E, and C on the green screen should be displayed in full name so that the users can directly figure out what feature they were for the task 1). They recommended enhancing the understanding of article clipping by introducing the meaning of article clipping using something like speech bubble (explanation displayed at mouse-over). For the task 2, participants said that readability was required to be improved with the size of letter compared to the composition of full screen. Participants suggested to increase the convenience of the search by segmenting the fields of theme further in searching by theme for the task 3. For the task 4, participants presented opinions that when image on the left area is provided, brief explanation should be added to help understanding the meaning of the image.

5. Conclusions

As information distribution technology and general information environment developed, information need of users became more segmented. It was found that researchers do need full text but they don't need the whole body of the text in every stage of R&D through the preliminary survey. In particular, for commonly needed information throughout the R&D life cycle, what the researchers needed were parts such as introduction, abstract, purpose of the study, experiment method, image, and table, not the whole text each time they sought information. Therefore we developed a prototype of a new information service that divides the full text and allows users to use only the parts they want, overcoming the traditional information service that shows full text. After developing the prototype system, we got evaluation from six participants who seek scholarly information for R&D. Generally they responded that the system was useful to catch up an article in outline and convenient to see the whole images in an article at one sight. However, more detailed explanations and functions as well as intuitive user interface should be considered for the user-centric system.

Many information service centers have been devoting a lot of efforts to fulfill the information needs of researchers. However, little research has been done for the micro-content, using deep indexing technique. Therefore, in this study, we suggest a new information service plan that allows users to select only the parts they want using micronized contents, different from the existing article level service. The information service based on micro-content is expected to support more particular and segmented R&D activities of researchers more closely. In addition, it is expected to develop a value added service like an analysis service customized for users, by constructing various statistical data based on the information service applying micro-content.

References

Agyen-Gyasi, K. (2008). The need for subject librarians in Ghanaian academic Libraries. *Electronic Journal of Academic and Special Librarianship, 9*(3). Retrieved from http://southernlibrarianship.icaap.org/content/v09n03/agyen-gyasa_k01.html

Ahn, I. J., Noh. D. J., Noh. Y., & Kim. S. J. (2009). Competency based curriculum development of subject specialist librarians. *Journal of the Korean Library and Information Science Society, 43*(1), 333-361.

Akshita, Smita. (2013). Recommender system: review. *International Journal of Computer Applications, 71*(24), 38-42.

Chung, J. Y. (2007). Focusing on the role of the subjects related to library: a study on the cooperation model of subject specialist upbringing plan. *Journal of the Korean Library and Information Science Society, 41*(1), 391-409.

Chung, J. Y. (2009). A study on operational plan of subject specialist librarian at academic libraries: focus on case analysis of three academic libraries. *Journal of Korea Library and Information Science Society, 40*(3), 119-136.

CSA. *What do we mean by "deep indexing?" CSA FAQs*. Retrieved from http://www.csa.com/factsheets/supplements/OBNATS_FAQ.php

Dash, Anil. (2002). *Introducing the micro-content client.* Retrieved from http://dashes.com/anil/2002/11/introducing-microcontent-client.html

Jeong, D. H., Kim, J., Hwang, M., Song, S. K., Jung, H., & Kim, D. W. (2013). Analytics service assessment and comparison using information service quality evaluation model. *International Journal of Information Processing and Management, 4(*4), 32.

Kim, K. Y., & Kim, H. M. (2013). A study on developing and refining a large citation service system. *International Journal of Knowledge Content Development & Technology, 3*(1), 65-80.

Korea Institute of Science and Technology Information, Dept. of NDSL Service. (2012). User survey on R&D lifecycle supporting functions and service development of domestic researchers, Internal Document.

Lee, S. H., Kim, H. M., & Choe, H. S. (2013). An evaluation of applying knowledge base to academic information service. *International Journal of Knowledge Content Development & Technology, 3*(1), 81-95. doi: 10.5865/IJKCT.2013.3.1.081

Nielson, J. (1998). *Micro-content: How to write headlines, page, titles, and subject lines*. Retrieved from
http://www.nngroup.com/articles/microcontent-how-to-write-headlines-page-titles-and-subject-lines

Noh, Y. (2009). A study on how to introduce subject-oriented service to university libraries based on their size in Korea. *Journal of the Korean Biblia Society for Library and Information Science, 20*(1), 101-117.

Sandusky, Robert J. (2008). *Deep indexing and discovery of tables and figures*. NISO Discovery Tools Forum 2008. Retrieved from
http://www.niso.org/news/events/2008/discovery08/agenda/sandusky.pdf.

Tenopir, C., Sandusky, R., & Casado, M. (2006). *The value of CSA deep indexing for researchers (executive summary)*. School of Information Sciences Publications and Other Works, 1. Retrieved from http://trace.tennessee.edu/utk_infosciepubs/1

Zin, N. A. M., & Yue, W. S. (2013). Design and evaluation of history digital game based learning (DGBL) software. *Journal of Next Generation Information Technology, 4(*4), 9.