The Past, Present and Future of Information Retrieval: Toward a User-Centered Orientation

Makiko Miwa

Professor, Faculty of Liberal Arts, The Open University of Japan

Summary

Initially, the differences in the information retrieval process when looking for information as a searcher on behalf of a client and when looking for information for the purpose of writing one's own article will be described. Major differences are that, when looking for information as a searcher, the clearly defined information needs of a user remain constant to the very end, whereas when looking for information as an author, one's vague information needs gradually evolve. This is the difference between system-centered information retrieval and user-centered information retrieval.

My presentation will then move on to the two approaches to information retrieval: the system-centered approach to information retrieval, and the user-centered approach to information retrieval. Focusing on the differences between the two approaches in terms of information needs, the information search process, and relevance, it is imperative that when building new databases, or when revamping existing databases, support needs to be given to learning and investigating processes, supporting gradually evolving information needs and taking into account the subjective determination of relevance by users.

In conclusion, the importance of metadata, and about the merits of conforming to metadata standards will be described. When searching for image, audio or video content that does not contain text data, it would be impossible to find content unless metadata was properly provided. The role played by metadata standards in the interoperable databases containing all forms of media content—namely, characters, images, audio and video—and in the cooperation among museums, libraries and archives, will be considered.

Makiko Miwa

Biography

<Education>

2000 Ph.D. in Information Transfer, Syracuse University, U.S.A.

1994 Enrolled in the Ph.D. Program, School of Information Studies, Syracuse University

1983 Completed the Doctoral Program, School of Library and Information Science, Keio University

1978 MLS, School of Library and Information Science, University of Pittsburgh, U.S.A.

1973 BA in History, Japan Women's University

<Career History>

2009-present Professor of Informatics, Faculty of Liberal Arts, Open University of Japan

2001-2009 Professor, National Institute of Multimedia Education
Concurrently taught at several universities including the Graduate University of Advanced Studies

1983-2001 Executive Officer, Epoch Research Corporation

<Major Publications>

Miwa, Makiko and Shizuko Miyahara, eds. *Quality Assurance in LIS Education: An International and Comparative Study*. New York: Springer, 2015. (Miwa, Makiko. "GlobaLIS: Toward the Regional Cooperation in the Education of Library and Information Professionals in the Asia-Pacific Region," pp. 3-24.)

Miwa, Makiko. "*Kaigai-ni-okeru 'Toshokan Joho Senmonshoku' no Shitsuhosho-to Konpitenshi* (Quality Assurance and Competency of Library and Information Professionals Overseas)", Joho-no Kagaku-to-Gijutsu (*The Journal of Information Science and Technology Association*), 66(2), pp.71-78, 2014. [in Japanese]

Miwa, Makiko. *Johokodo: Shisutemu Shiko-kara Riyosha Shiko-e* (Information Behavior: From a System-Centered toward a User-Centered Approach). Tokyo: Bensei Shuppan, 2012. [in Japanese]

Miwa, Makiko, et al. "A Method to Capture Information Encountering Embedded in Exploratory Web Searches," *Information Research*, 16(3), 2011.
http://www.informationr.net/ir/16-3/paper487.html

The Past, Present and Future of Information Retrieval: Toward a User-Centered Orientation

Makiko Miwa

Professor, Faculty of Liberal Arts, The Open University of Japan

Introduction

Since the 1980s, I have been involved in information retrieval as both a professional and a researcher. In the early days when computers were introduced into information retrieval, the limitations of storage capacity meant that searches could only be done on bibliographical data, that is, on short character strings, such as catalogs of books and indexes of academic papers. Following that, with the advent of electronic journals, academic texts emerged in searchable digitized form. In other words, it was now possible to conduct searches on the full texts. In the 1990s, the Internet spread rapidly worldwide. There was a simultaneous increase in the capacity of computers facilitating the creation of multimedia databases of not only character data, but also image, audio and video data. Nowadays, we are able to search databases of diverse media online. Recently, electronic books—in which whole books are digitized—have also begun to grow in popularity.

Today, I will discuss the historical developments and future prospects for information retrieval, focusing on the following points, while also recounting my own personal experiences. First, I will talk about my experience as an information retrieval specialist—commonly known as a "searcher"—here I became aware of the difference between the system-centered process of information retrieval and the user-centered process of information retrieval. Next, I will trace the changes in the history of information retrieval. In particular, I will look at the differences in searches of textual information and searches of non-textual information, such as solid objects and multimedia, that is, images, audio and video. Following that, I will introduce the differences between the system-centered and the user-centered approaches to information retrieval, while interspersing the discussion with specific examples. I will discuss the recent situation in which we have been transitioning from an information retrieval environment (built as single database systems) to federated searches and interoperable multiple databases. Put simply, I will talk about the growing collaboration between museums, libraries and archives. Finally, I will consider why metadata—an alternative expression of information that is key to searching information effectively and efficiently—is important amid this expansion of databases.

My Experience in Information Retrieval

I will begin with an account of my own experiences. After graduating university, I got a job as a type of editorial assistant at a foreign publisher. It was at this time that I learned about how characters can be coded and handled on computers, leading to my interest in databases. I could not find a graduate school in Japan that would teach me about character processing on computers, and so I enrolled on a master's course at the University of Pittsburgh School of Library and Information Science in the United States. Here, I learned about the digitization and computer processing of characters—in other words, about building databases and conducting searches—leading to becoming a Master of Library and Information Science. On returning to Japan, I enrolled on a PhD program in library and information science which had just started at Keio University, where I had an opportunity to examine research and development related to constructing and searching bibliographic databases in Japanese. I gained enough credits to graduate, but unfortunately, did not complete a doctoral thesis. During the doctoral program, I was in charge of a service at the University of Tsukuba Science Information Processing Center—that is now called the Research Facility Center for Science and Technology—providing an academic database online to universities and other institutions throughout Japan. My boss at that time had an attitude of, "Computers work 24 hours a day, and so it's only natural that humans do likewise." So I spent my days working the whole time—providing database services, planning and running workshops during the day, conducting R&D and writing papers at night.

At about this time in the United States, "searchers" were getting actively involved at places like libraries and research institutions, as professionals searching online databases on behalf of users, and attention was being drawn to "information brokers," ―companies that engage in the collection and analysis of information. This inspired me to start up an information broker business in Japan, and in 1983 I founded a company named the Epoch Research Corporation. At the beginning, it was difficult to get people understand what the business involved. Initially when I approached banks to raise funds, they were reluctant to lend to me, as there was no precedent for such a type of business. Gradually, the idea gained attraction in the mass media, allowing me to proceed. My company's business included doing investigative work for other businesses and public institutions, designing databases, digitizing library catalogs and undertaking international surveys in collaboration with fellow brokers overseas. While the nature of my business covered a wide range of activities, the most successful one was the training course for searchers—that is, professionals in information retrieval. I provided practical training in performing searches in a variety of databases, in the fields of medicine, patents, science and technology, and corporate business. The course became the

precursor for the subsequent “Searcher Course” offered by the Information Science and Technology Association (INFOSTA). On a personal front, I also published a book through Maruzen called *Saachaa no Jidai*[1] (which translates as “The Age of the Searcher”). It became a best seller as a technical book, and helped the “searcher” profession gain recognition in the wider community. At the same time, I also had opportunities to teach “information retrieval” and other classes part-time at a number of universities, including Keio University.

In the 1990s, with the advent of the Internet, the information retrieval environment underwent a number of changes. The biggest difference was that online databases, which previously could only be accessed by professional searchers, could now be searched directly by users. Faced with such a change, I could no longer hold much hope for the future of information brokerage, and so I resolved to turn my hand to research and get that doctoral degree that had eluded me at Keio University. With a goal of writing my doctoral thesis, I enrolled in the PhD program offered by the School of Information Studies of Syracuse University in 1994. My initial plan was to stay for three years, but being involved in research and education as a full-time student was so much fun that I ended up leading a student life in Syracuse for an extended period, just shy of six years. Syracuse is a city located in the U.S. state of New York. Being so close to the Canadian border and to the Great Lakes, the city receives a lot of snow, and so I spent many days snowbound from November to March. I don’t think you could find a better environment anywhere else to work on research.

After obtaining a doctoral degree in 2000, I returned to Japan. Then, in 2001, I was appointed as a professor at the National Institute of Multimedia Education (NIME). At the time, NIME had been an inter-university research institute, but it subsequently took the form of an incorporated administrative agency, before being merged—or should I say absorbed—into the Open University of Japan in 2009. Currently, I am a professor at the Open University of Japan.

When working as a searcher at the Epoch Research Corporation, I was asked to write a four-page piece entitled “The Adventure of a Searcher” for a monthly magazine published by KDD called *On the Line*. Each time, I would decide on a theme in consultation with the editor, and then I would search databases and summarize my findings. I enjoyed writing those articles. The series of articles were later published as a book titled *Amerika Kaidoku Shinsho* (which translates as “A New Book on American Deciphering”).[2] When composing this series of articles, I became aware of a major difference between looking for information as a searcher for others, and looking for information as an author for one’s own compositions.

When looking for information as a searcher, first, I would interview the client to clarify their information needs and what they desired in terms of the search results. Then, I would select a

database suited to the purpose, and map out an appropriate search strategy. Only then would I begin the search. If the envisaged result was not forthcoming, I would re-map the search strategy and continue searching until a satisfactory result could be obtained. However, since the databases were not free to use, it was important that I perform the search efficiently in a short period of time. Once I got search results corresponding to the client's information needs, I would look over them, extract the necessary information, add my interpretation, put it all together into a report, and deliver it to the client. The whole process would take about a week at the most. Sometimes, though, I would get a request from a medical professional marked "Very Urgent," and would have to respond in only two or three hours.

On the other hand, when looking for information as an author for the purpose of writing an article, I would then look for information in databases in order to decide on the article topic. Information needs at this time were vague and I would start by selecting a provisional topic and then consult with the editor. Next, I would perform a search focused on the provisional topic, looking for potential articles and statistical data, before composing a story while interpreting the information I had obtained. At this point, if a coherent narrative did not emerge, or if I could not get enough data, I would change the topic with the editor's permission. Then I would write up the story I had composed using a certain number of words and then convey my ideas for the graphic to the illustrator.

There are significant differences between these two information retrieval processes. When looking for information as a searcher on behalf of a client, the information needs defined at the outset and the anticipated output requirements remain constant to the very end. Also, whether or not the search results are relevant the information needs is objectively determined by looking at the outline or summary. In contrast, when looking for information as an author, the information needs defined at the start—in other words, the article topic—gradually changes depending on the knowledge gleaned during the information retrieval process. Consequently, the initial topic changes almost all the time. Also, when scrutinizing the research results, whether or not a certain text is likely to be useful as an information source for the article, is determined subjectively from reading it.

Let us now take a step-by-step look at the search process involved in developing an article while searching for information, using the three aspects of: cognition, emotion and action. In the first step: the cognitive aspect is the generation of ideas, while the emotive aspect is the mood of welcoming information, and the active aspect is reading newspaper articles, yearbooks and trade journal articles, as well as talking with people. In the second step: the cognitive aspect is defining a broad topic, while the emotive aspect is being adventurous, and the action is searching databases within the topic domain and looking for the latest articles. In the third step: the cognitive aspect involves composing

a story, while the emotive aspect is a state of flow, in other words, a state of stopping other things from intruding, due to being totally engaged. The action here is reading through potential documents acquired as search results and checking statistical data. In the fourth step, the cognitive aspect is putting together the data and information and connecting them into a story. Since this involves making decisions, the emotive aspect involves being critical. The action is extracting and selecting text and data that matches the story. In the fifth step: the cognitive aspect is integration, the emotive aspect is being logical, and the action is creating charts and putting your interpretation into words. In the final sixth step: the cognitive aspect is making an announcement, the emotive aspect is imagination, and the action is writing the story.

Now, let us take a more detailed look at the differences in searching when mediating user information needs as a searcher and when looking for information yourself as a user. I will use the term “intermediary searches” for searches performed as a searcher, and “self search” for searches performed by the user. First, with respect to topics, in the case of an intermediary search, the topic is clearly defined, whereas in the case of a self search, the topic is vague. In terms of the search process, intermediary searches are systematic, whereas self searches are not. With regard to search strategy, in intermediary searches, the strategy is configured as step-by-step procedures, whereas in self searches, the strategy is based upon trial and error. As for emotion, in intermediary searches, emotion is stable, or hardly moves at all, whereas in the case of self searches, emotion is unstable, a mixture of pleasure and anxiety depending on the search results. In terms of knowledge structure, in intermediary searches, the structure gradually deepens in a fixed direction, but in self searches, there is a sense of expansion as the direction changes. Finally, as for the theme that runs throughout the search process, in the case of intermediary searches, the aim is for achievement, whereas in the case of self searches, the aim is for creation, or to be more precise, the creation of results.

Without doubt, these differences are the differences between system-centered information retrieval and user-centered information retrieval. In other words, in system-centered information retrieval, information needs do not change, and relevance can be evaluated objectively, whereas in user-centered information retrieval, information needs change in the middle of the search, and the relevance of the information is evaluated subjectively. When I became aware of this difference, I thought to myself that I could write a doctoral thesis on this.

Two Approaches to Information Retrieval

Information retrieval systems have been developed since the 1950s. Up until the 1980s, prior to the advent of the Internet, it was normal for the retrieval of information to be managed by

professional searchers, and the system-centered approach to information retrieval was the norm. With the arrival of the Internet in the 1990s, users were able to search for information themselves using a web browser. This led to the emergence of the user-centered approach to information retrieval and now both approaches vie for supremacy.

The System-Centered Approach to Information Retrieval

First, I will introduce the model of system-centered information retrieval (Figure 1). At the far left of this diagram is a group of documents contained in a database. The indexer attaches an index, or metadata, to each document. I will explain about metadata in more detail shortly, but for now, in the case of information retrieval, it refers to the titles of a document, the name of the author, the name of the publisher and subject keywords, which alternatively work as its access points. On the right side of the diagram is the user who has information needs. The database is queried by expressing the information needs as an enquiry combining keywords. Thesauri, classification tables and subject headings are used as translation tools for matching the keywords used by the user in the query against the keywords attached to documents by the indexer. To put it another way, the indexer refers to these translation aids when attaching indexes, and the user refers to these translation aids when selecting keywords.

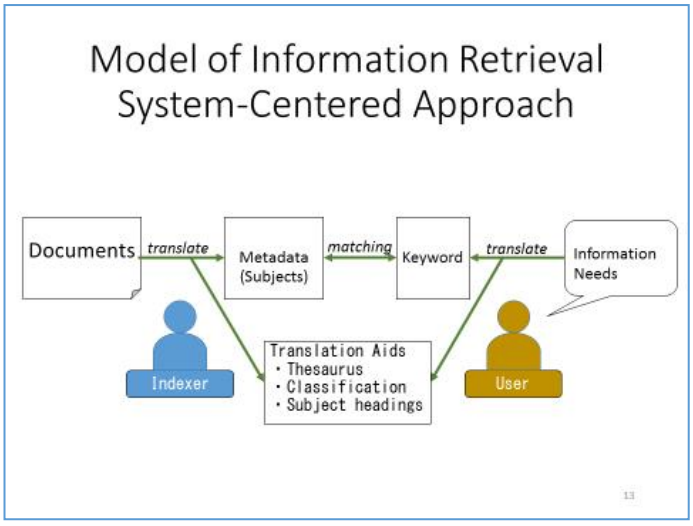

Figure 1. The System-centered information retrieval model

The user represents information needs in queries using logical operations on keywords. There are three types of logical operations that can be used here: logical sum, logical product and logical difference.

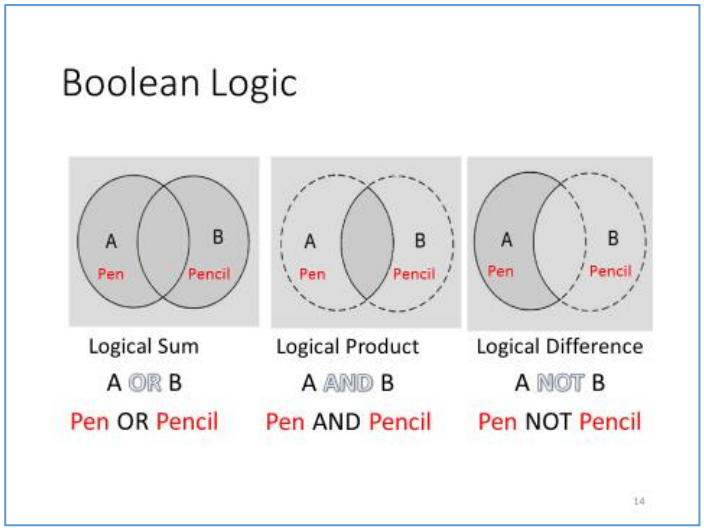

Figure 2. Boolean operations

For example, let us consider logical operations for the two keywords “pen” and “pencil.” In the case of a logical sum, the search will retrieve documents that contain either “pen” or “pencil” or both. In the case of a logical product, the search will retrieve documents that contain both “pen” and “pencil.” In the case of a logical difference, the search will retrieve documents that contain the keyword “pen” but which do not contain the keyword “pencil.” In real-life searches, complex queries can be expressed by combining logical sum, logical product and logical difference operations.

In information retrieval, the first query that is created not necessarily lead to search results that match the information needs. For this reason, a technique was devised called relevance feedback. This is a method of leading to better search results by revising the initial search results. There are three types of relevance feedback techniques: explicit feedback, implicit feedback and pseudo relevance feedback.

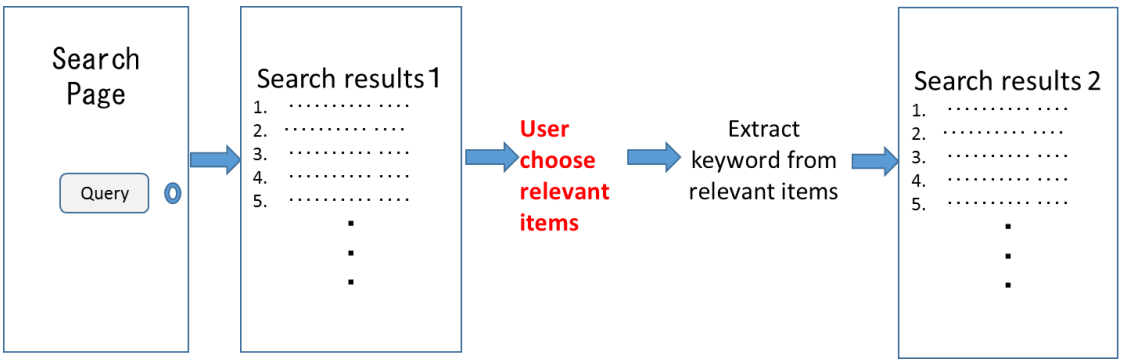

Figure 3. Explicit feedback

Explicit feedback (Figure 3) is the technique of getting the user to decide on the relevance of the documents retrieved as a consequence of the initial query, and forming the next query using the metadata of those documents that are determined to be relevant.

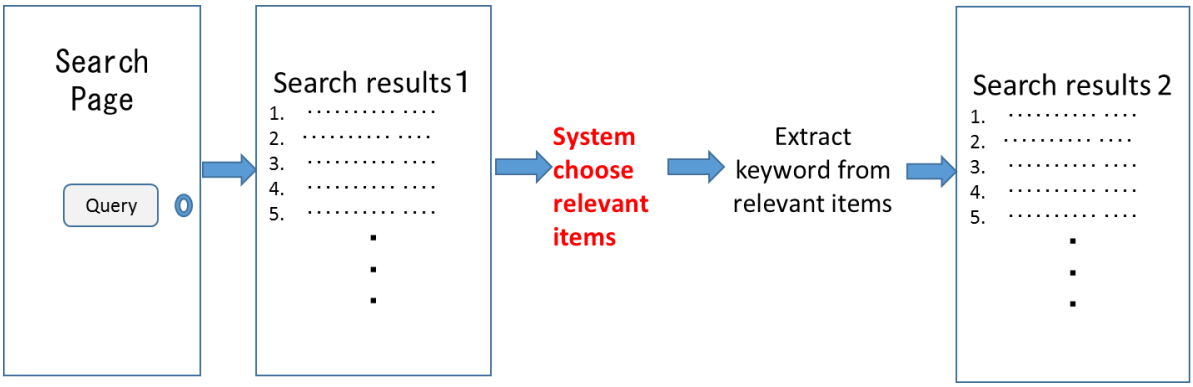

Figure 4. Implicit feedback

In implicit feedback (Figure 4), the system detects whether the user has looked at each of the documents appearing in the search results of the initial query, and if so for how long and detects any browsing or scrolling action. Based on this, the system then determines relevance. The next query is then automatically formed using the metadata of those documents determined to be relevant. This technique is called “implicit” feedback because relevance is reckoned without the user knowing.

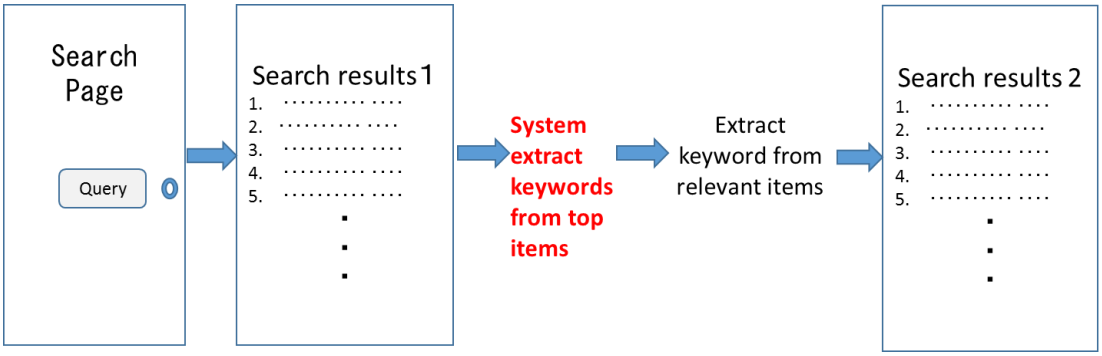

Figure 5. Pseudo relevance feedback

In pseudo relevance feedback, once the set of relevant documents has been searched, the top few documents in the list of search results shown in order of relevance are judged as being highly relevant. The next query is then automatically formed using the metadata of these documents.

In this way, relevance feedback is used for deriving more relevant search results, by revising an initial query based on the content of relevant documents. In system-centered information retrieval,

various other techniques are also used in order to increase the relevance of search results, such as by attaching weight to certain keywords.

————

The User-Centered Approach to Information Retrieval

Next, let us look at the thinking on the user-centered approach to information retrieval. In the user-centered approach, information needs are seen as gradually evolving. Figure 6 shows the development process for information needs.

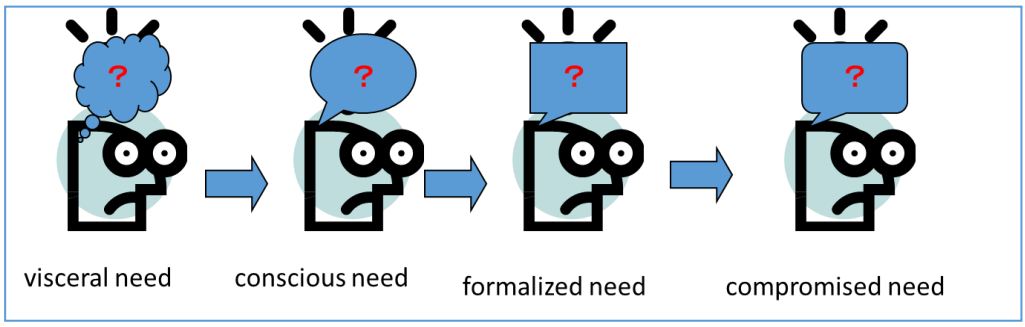

Figure 6. Evolving information needs[3]

As shown in this figure, information needs evolve in stages. At first, there is visceral need, or true need that cannot be expressed in words. Put another way, it is the situation where, even though a person feels a vague sense of unease, they are unable to explain it to others because they are not sure what it is themselves. The next stage is conscious need. This is the situation where a person has an undetermined, ambiguous need that cannot be adequately defined, but whose ambiguity may be resolved through talking with others. The next stage is formalized need. This is the situation where a person can clearly express their own information needs. The last stage is compromised need. This is the situation where a searcher explains a person's information needs in a way that can be understood, such as by forming queries combining keywords in accordance with the characteristics of the information retrieval system, so as to obtain a response from the system. Since information needs develop in this way, in the early stages, users might find satisfactorily expressing their information needs difficult. The idea in user-centered information retrieval that "information needs change and cannot be defined clearly" differs from the premise in the system-centered information retrieval model that "information needs can be defined clearly and do not change."

Next, I will present a model of information behavior that represents the overall picture of user behavior and cognition in the information search process (Figure 7).

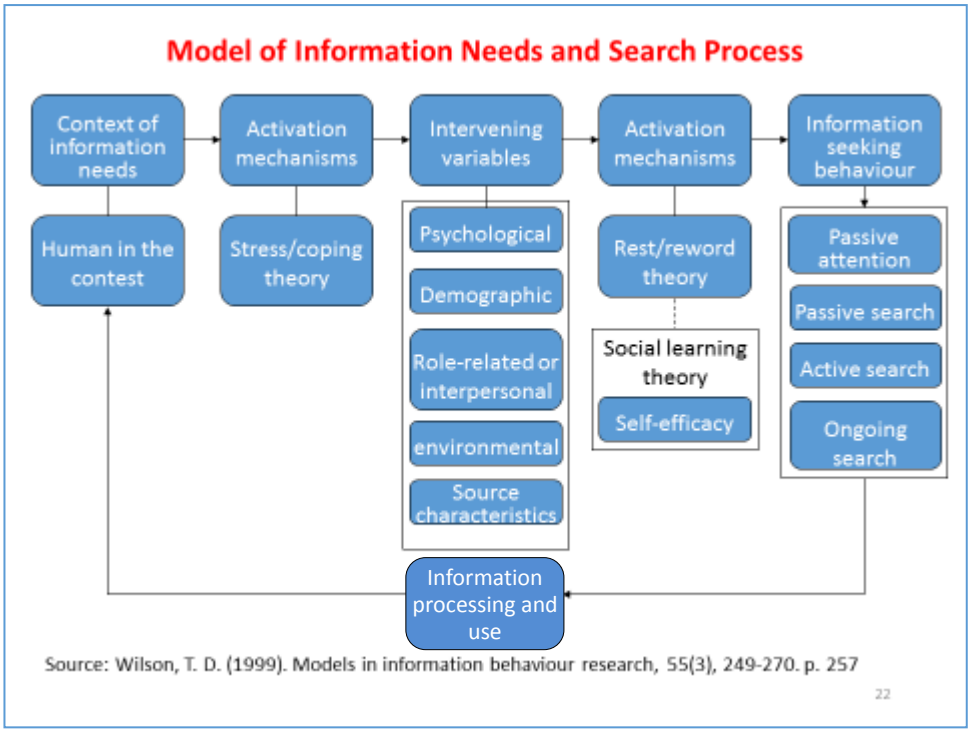

Figure 7. Model of information behavior[4]

This model shows that, when a user with information needs engages in information seeking behavior, the user will decide whether to take action, taking into account the stress felt when searching for information based on stress-coping theory. The intervening variables when engaging in information behavior are: the user's psychological factors, socio-economic factors, roles and interpersonal factors, environmental factors and the characteristics of information sources. The model also shows that, as an activation mechanism for information seeking behavior, the user will, based on social cognition theory, decide whether to seek information depending on the degree of self-efficacy with respect to looking for information, such as by weighing up the relative merits from obtaining information and from not obtaining information based on risk compensation theory. The model also shows that there are four types of subsequent information seeking behavior: passive attention, where an individual keeps a watchful eye on the surrounding environment and constantly checks that there are no unusual changes; passive search, where an individual reads newspapers and listens to the news in order to keep track of world events; active search, where an individual satisfies his or her information needs by eliminating anomalous states of knowledge; and ongoing search, where an individual continuously tries to obtain new information on a specific theme. Finally, processing, or interpreting, the obtained information completes the chain of information seeking processes.

In this way, the user-centered approach to information retrieval presents a number of important findings. First, since knowledge is anomalous, the user may not be able to clearly state his or her own information needs. Next, since the approach provides search results that satisfy information needs, the method of expressing information needs is crucial. The approach also reveals that certain types of information needs do not lead to information seeking behavior. For example, it is known that some people do not go to see a doctor despite being in poor physical condition, and some patients do not seek information on their own disorder or the best method of treatment despite being diagnosed as having a serious illness. The approach also brings to light that the fact information seeking behavior can be passive, active or ongoing.

In information retrieval, relevance is used to determine whether search results satisfy the user's information needs. In system-centered information retrieval, relevance is regarded as an objective indicator determinable by a third party, and is used in judging search results. Figure 8 describes the two measures for determining retrieval performance based on relevance: precision and recall.

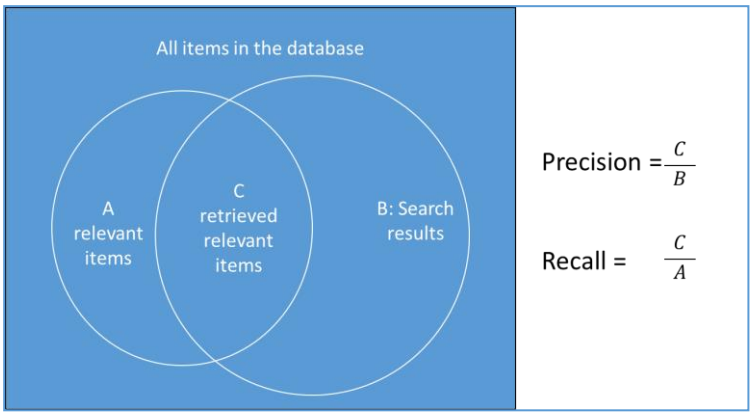

Figure 8. Precision and recall

In this diagram, the blue area indicates all documents contained in a database. Suppose that there are 20 documents (A) in the database which match specific information needs. Performing a search in this database generates 40 documents in the search results. Ordinarily all of A should be included within B, but if there are residual items, these constitute "omissions"; and if documents unrelated to A are mixed in with B, these constitute "noise." Recall and precision are measures of omissions and noise. Suppose the search results contain 12 relevant documents (C): recall R = C/A = 12/20 = 0.6 (omissions = 20-12 = 8); and precision P = C/B = 12/40 = 0.3 (noise = 40-12 = 28). Furthermore, if both R and P equal one (the maximum value), then we have a perfect search, with neither omissions nor noise.

This model assumes that a third party can objectively determine whether the search results are relevant or not. In contrast, in the user-centered approach to information retrieval, relevance can only be determined by users themselves. In system-centered information retrieval, whether the subjects of search results are relevant or not is determined objectively, and so relevance can be measured; whereas in user-centered information retrieval, whether search results are relevant to the context is determined subjectively, and so measuring relevance in this case is considered difficult. This leads us to the question: How do users judge search results? Users make evaluations from many different perspectives:

Novelty: Did the user obtain previously unknown information?

Recency: Is the information the latest information?

Quality: Is the information obtained of high quality?

Comprehensiveness: Are there any omissions in the information obtained?

Simplicity: Can the user read and digest the information easily?

Appeal: Does the information engage the user's interest?

In the information search process, users seek information with some purpose in mind. They then evaluate the obtained information from many different aspects, and if they acquire that information, that user's own knowledge structure will change. For this reason, it is difficult to map out in advance the information retrieval process of a user seeking information.

Studies on user-centered information retrieval have produced a number of models that depict the information search process. I will now introduce some of them.

First is the berrypicking model (Figure 9).

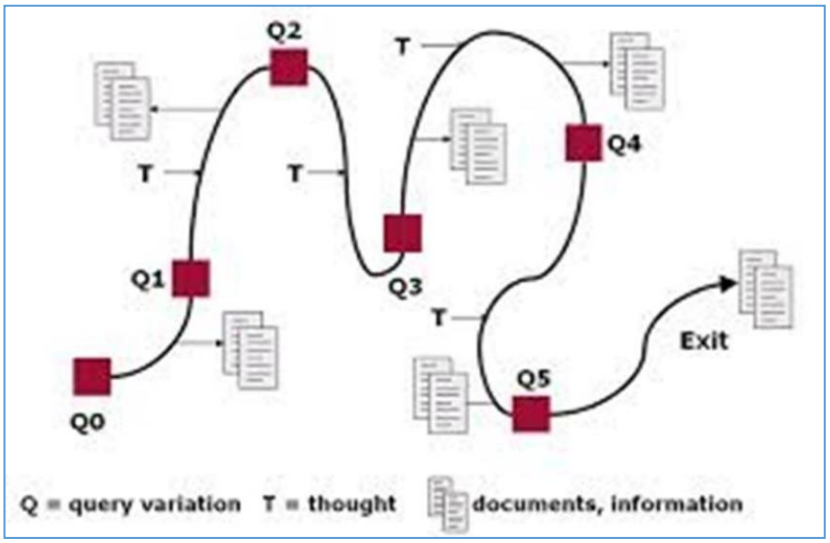

Figure 9. Berrypicking model[5]

This model applies the method of picking berries in the wild to the information search process. Wild berries do not grow in bunches, they grow scattered among bushes. Therefore, they must be picked one at a time. Similarly, a user seeking information collects information bit by bit, not as a single complete search result. The model also uses diverse search techniques beyond that provided in document databases. In general, the following six search techniques are used.

Footnote chasing is the technique of following up on documents indicated in the footnotes or endnote citations of an academic paper, or some other material.

Citation searching is the technique of finding new documents by tracing the citation index of an academic paper.

Journal run is the technique of finding documents by regularly reading through the table of contents of a journal in the user's area of interest.

Area scanning is the technique of reading through academic papers and conference presentations that are published in the user's area of interest.

Abstract and index searching is the searching of databases by reading through abstract journals and index journals.

Author searching is the method of looking for academic papers and books published by a specific author. This includes searching for the author's own website and SNS profiles.

Next is the information search process model which depicts the information search processes of a high school student writing a report for the first time (Figure 10).

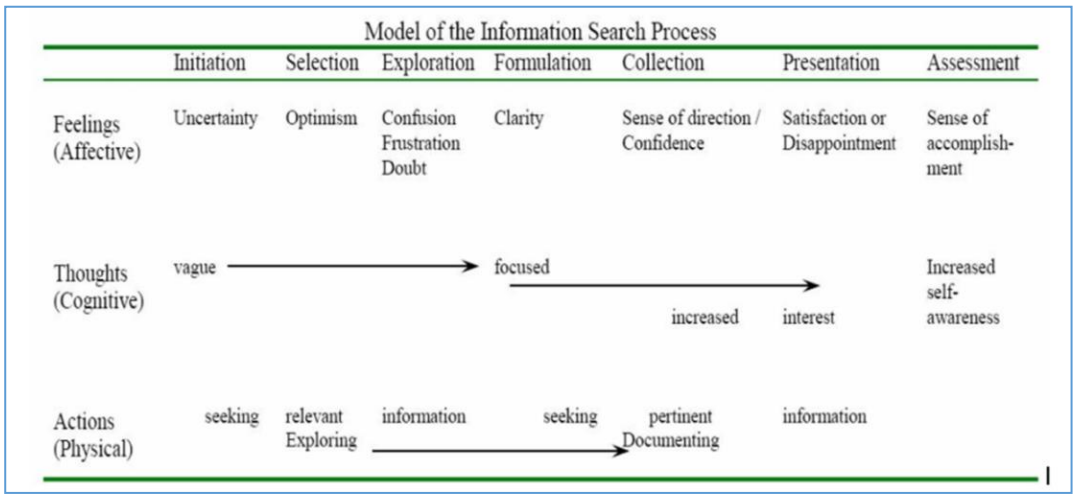

Figure 10. Model of the information search process[6]

During the first stage, the teacher explains the report to the student. Unsure what to do, the student feels uncertain and seeks advice from people about topic selection and how they should approach the report. In the second stage, the student decides on how to approach the report topic. She selects a topic, envisaging the requirements of the report as well as the available sources of information and her allotment of time. Once the topic has been decided, feelings of uncertainty fade, giving way to vague feelings of hope. During the third stage, the student deepens her understanding of her chosen topic by collecting information on the topic. Information from different sources can be inconsistent, making the student feel discouraged and frightened, and causing her frustration to rise. During the fourth stage, as the knowledge structure becomes more definite, the focus of the topic is formed. At this stage, as the student reads documents on the topic and writes down themes and ideas, her feelings of uncertainty diminish and confidence increases. In the fifth stage, the student picks out information linked directly to her topic focus, that is, information that she could use in her report. She makes detailed, focused notes. As her interest in the topic focus grows, her feelings of uncertainty fade away and her confidence increases. During the final sixth stage, the student completes her search for information, checks that she has not missed anything, and then begins preparing to write her report. If the topic has been synthesized from a personal point of view, then as the student's reflection reaches a climax, her thoughts take shape and she concludes her information search, she feels a sense of release. This model suggests that thoughts and feelings change during the information search process, and that teachers and other mediators need to envisage which stage of the information search the student is at before providing appropriate information.

Next is the information seeking behavior model for professionals, which depicts the information search processes of an expert in information gathering, such as a researcher or executive (Figure 11).

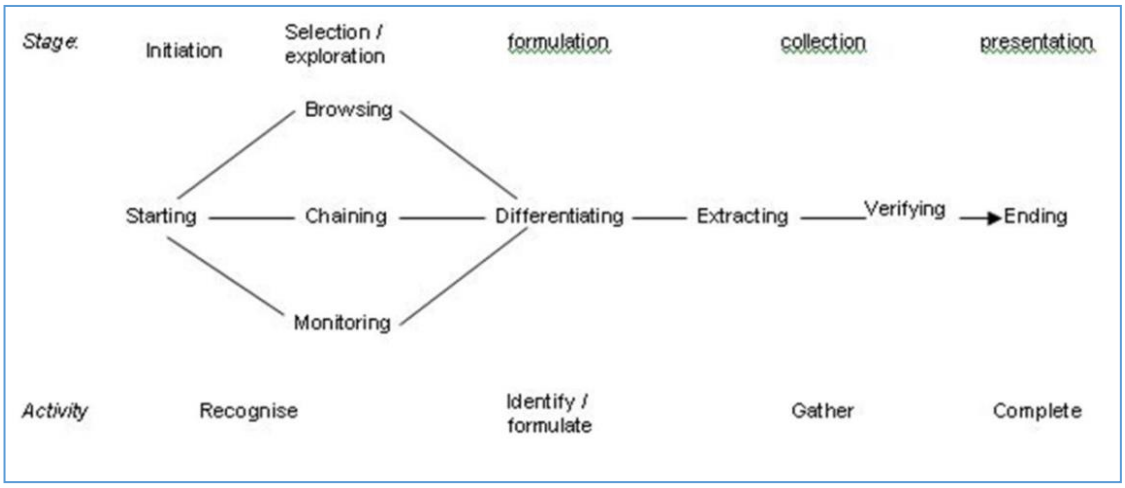

Figure 11. Model of the information seeking behavior of professionals[7]

When starting out looking for information, a professional might ask a friend or colleague who looks knowledgeable, or may set about doing research in a library or online. Chaining is the method of using information already acquired as a lead in extracting other information, such as by using citations or getting a referral from an acquaintance. Browsing is a nebulous search for information without any clear direction. An example is going to a library or bookstore and scanning the shelves of books and journals in the relevant area. Another example of browsing is clicking on a relevant category online and scanning to see what kind of information is there. Monitoring is the act of gathering new information on a specific topic. This also includes current awareness searching, such as asking acquaintances or experts to notify you if they come across any new information on a certain topic. It can also mean registering for a specific topic on a database, in order to receive automatic notifications if any new articles or academic papers appear on that topic. Next is differentiating. This is the differentiation of information sources based on the nature of those sources. It includes selecting academic papers published in a prestigious journal, selecting books based on reviews printed in a newspaper, and selecting information sources recommended by experts. Extracting refers to the extraction of useful information from information sources. Prominent cases in point are taking notes while reading an article, paper or book, and copy-and-pasting important sections while browsing web pages. Verifying validity refers to checking the accuracy and appropriateness of information. Comparing related information obtained from multiple sources fits into this category. If inconsistencies are found between different sets of acquired information, the user follows up to determine which is correct. A search ends when no more new information can be obtained, when the desired information cannot be found, or when time runs out.

These models reveal that the user-centered information search process has the following characteristics:

- The information search process is a linear process that includes loops and trial and error.
- The process starts with a broad topic, and gradually narrows down to a more focused topic.
- Information needs evolve during the information search process.
- Users continuously evaluate their own decisions.
- Knowledge structure changes during the information search process.
- Users have some strategy to end the information search process.

Figure 12 shows the outcome of reflecting on this kind of user-centered information search onto the model of the web search process.

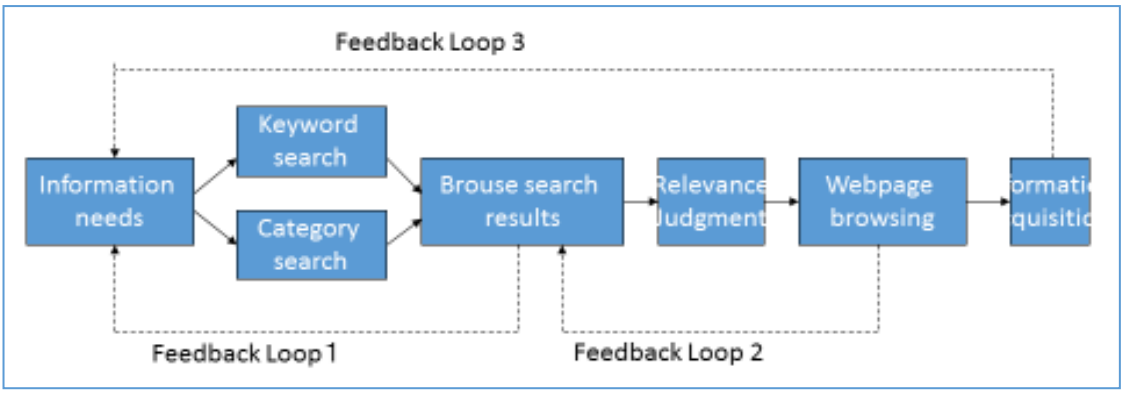

Figure 12. Web search process

When a user feels a need for information, he searches online. I think that if his information needs are well-defined, then he will do a keyword search; but if his information needs are vague, then he will perform a category search. Browsing the list of search results, he clicks on a result relevant to his information needs. If he does not find any relevant results, he rethinks his information needs and changes the keywords, category or search engine. Browsing the displayed web page, he acquires his desired information.

If his desired information is not available, he returns to the list of search results, and clicks on a separate result that matches his information needs. If he draws a blank again, he rethinks his actual information needs. Then, as shown in the berrypicking model, the acquired information leads to a new query, that is, it generates a new information need, and so the next information search begins.

User-centered information retrieval uses the concept of “exploratory search” (Figure 13). This figure compares and contrasts user-centered information retrieval against conventional information retrieval. It classifies information retrieval into three categories—Lookup, Learn and Investigate—and regards searches performed for the purpose of learning and investigating as exploratory searches. “Lookup” searches involve system-centered information retrieval. It is thought that current information retrieval systems support searches performed for the purpose of looking up information, but do not adequately support searches performed for the purpose of learning and investigating. Therefore, when building new databases in the future, or when revamping existing databases for user-centered information retrieval, consideration will need to be given to supporting learning and investigating.

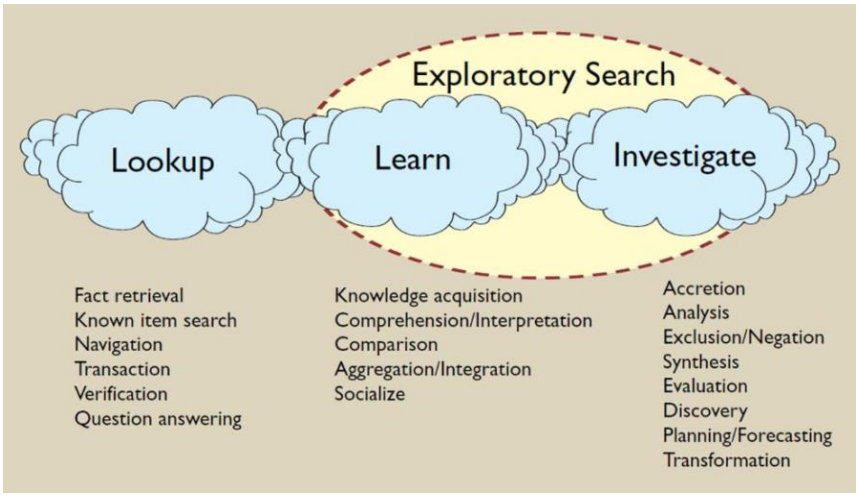

Figure 13. Exploratory search[8]

Importance of Metadata

Metadata is an important element in information retrieval, and in particular, in subject searches. Metadata means "data about data," and in information retrieval, it is regarded as a substitute for content, that is, the keywords attached by indexers.

Let us look at a visual representation of the function of metadata (Figure 14).

Figure 14. Illustration of metadata

In this way, metadata are the labels that show what the content of the information is. When searching electronic objects, characters can be used to perform the search as long as the content is

character data; but if the content is a photo, sound or video that does contain characters, the search can only be performed if metadata is properly provided.

The primary purposes of metadata are to assist in detecting relevant information, to systematically organize information sources for information management, and to provide unique IDs. Traditional examples of metadata include library catalogs, business cards and directories. The metadata of electronic objects is created using metadata standards for specific genres. Doing so not only assists in information retrieval and interoperability, but it also has the potential to increase the utility of the objects. Examples of metadata for electronic objects include the title and language described in the meta tags of web pages as well as the creation tools. Wikipedia article categories, titles, information sources and access methods are also all metadata.

Table 1 shows the main metadata standards in each genre.

Table 1. Metadata standards in each genre

genre	Metadata Standards
Library	Dublin Core, MARC (Machine Readable Catalog)
Digital government	e-GMF (e-Government Metadata Framework)
Museum	CIDOC CRM
Geography	FGDC (The Federated Geographic Data Committee)
Education	IEEE LOM (IEEE Learning Object Metadata)
Rights	Indecs, XrML (eXtensible rights Metadata)
Multimedia	MPEG-7, MPEG-21
News articles	NewsML
Broadcasting	TVAF, SMPTE, P/META, Jmeta

In the libraries genre, Dublin Core is used as the standard for metadata. Dublin Core is also used as a basic element in e-GMF, which is the metadata standard for e-government, or to be more precise, for archiving, and in IEEE LOM, which is the metadata standard for learning resources.

The Dublin Core metadata scheme is comprised of 15 elements shown in Table 2. “Title” is the name of the object; “Creator” is the entity primarily responsible for creating the object; and “Subject” is the keywords representing the topic. “Description” is a sentence describing the outline of the object; “Publisher” is the entity responsible for making the object available; and “Contributor” is the relevant person, other than the creator, who contributed to creating the object. “Date” is the date of a major event in the lifecycle of the object, and could describe the date at which the object was created,

made public or modified. “Type” is an element representing the nature of the object, and could describe a journal article, photograph, statistical data or such like. “Format” is the physical form of the object, and is described using MIME types. “Identifier” is the unique ID given to the object, using either URI or URL if there is no URI. “Source” is the original document or other resource from which the object is derived. “Language” describes the language in cases where the object contains characters or voice. “Relation” describes the vertical relationship or close relationship between objects. “Coverage” is a representation of the location or time of the object: for a photo, it describes where it was taken, and for a historical object, it describes its age. Finally, “Rights” describes information pertaining to the copyright. Metadata descriptions are particularly important in cases of images, audio or video which do not contain characters, but at the same time, this is a time-consuming task. Still, finding objects would be impossible unless metadata was described, and so it is essential that metadata be added in as much detail as possible. In addition, when adding metadata, it is also important to observe the metadata standards of the area or genre.

Table 2. Dublin Core metadata elements[9]

ELEMENT	DEFINITION
Title	A name given to the resource.
Creator	An entity primarily responsible for making the resource.
Subject	The topic of the resource.
Description	An account of the resource.
Publisher	An entity responsible for making the resource available.
Contributor	An entity responsible for making contributions to the resource.
Date	A point or period of time associated with an event in the lifecycle of the resource.
Type	The nature or genre of the resource.
Format	The file format, physical medium, or dimensions of the resource.
Identifier	An unambiguous reference to the resource within a given context.
Source	A related resource from which the described resource is derived.
Language	A language of the resource.
Relation	A related resource.
Coverage	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.
Rights	Information about rights held in and over the resource.

I think I have already adequately explained why metadata is important, but another reason is because it makes it possible to perform federated searches across different databases. In other words, when using a single query to search multiple databases, the query is passed on to multiple search

engines, and search results obtained from them all are displayed in one list. If the metadata was wrong, the searches would not go well. Recently, collaboration among museums, libraries and archives is being promoted, but a common metadata scheme needs to be used in order to align the content of each of the facilities. The same goes for international alliances of databases in the same genre. Moreover, recently, local governments and community museums, libraries and archives are also independently pushing ahead with the digitization of their local resources, but again, unless the metadata is standardized, users will end up not being able to conduct bulk searches of all data.

Conclusion

I began my presentation by describing the differences in the information retrieval process when looking for information as a searcher on behalf of a client and when looking for information for the purpose of writing your own article. In the course of my discussion, I stated how I became aware of the fact that, when looking for information as a searcher, the clearly defined information needs of a user remain constant to the very end, whereas when looking for information as an author, your vague information needs gradually evolve. This is the difference between system-centered information retrieval and user-centered information retrieval.

My presentation then moved to the two approaches to information retrieval: the system-centered approach to information retrieval, and the user-centered approach to information retrieval. I explained about the differences between the two approaches, in terms of information needs, the information search process, and relevance. I also pointed out that, when building new databases in the future, or when revamping existing databases, support needs to be given to learning and investigating processes, supporting gradually evolving information needs and taking the subjective determination of relevance by users into account.

Finally, I talked about the importance of metadata, and about the merits of conforming to metadata standards. I pointed out that, when searching for image, audio or video content that does not contain character data, it would be impossible to find content unless metadata was properly provided. I also spoke about the role played by metadata standards in the interoperability of databases containing all forms of media content—namely, characters, images, audio and video—and in the cooperation among museums, libraries and archives. My hope is that this knowledge will be utilized in realizing efficient and effective searches of archives and multimedia databases.

---

[1] Miwa, M. (1986). *Saachaa no Jidai: Koudo Deetabeesu Kensaku*. Maruzen, 224p. Miwa, M. (1992). *Saachaa no Jidai: Koudo Deetabeesu Kensaku (Dai 2-Ban)*. Maruzen, 240p.
[2] Miwa, M. and Ichinowatari, K. (1991). *Amerika Kaidoku Shinsho*. Sekai Bunka Publishing, 248p.
[3] Taylor, R.S. (1968). Question-negotiation and information seeking in libraries. *College and Research Libraries*, vol.29, no.3, p.178-194. (Figure is original by author)

[4] Wilson, T.D. (1999). Models in information behavior research. *Journal of Documentation*, vol.5, no.3, pp.249-279. (Figure is Figure C3.3, p.34)
[5] Bates, M.J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Review, vol.13, no. 5, p.410, Figure 2.
[6] Kuhlthau, C.C. (1994). Seeking meaning: A process approach to library and information services, Norwood, NJ. Ablex Publishing, p.82.
[7] Ibid. 4, p.160, Figure 25.1.
[8] Marchionini, G. (2006). Exploratory search: From finding to understanding. Communication of ACM, vol. 49, no. 4, Figure 1.
[9] Dublin Core Metadata Initiative. (1995). Dublin Core Metadata Element Set, Version 1.1: Reference Description. <http://dublincore.org/documents/1999/07/02/dces/>