

An Analysis of Image Retrieval Behavior for Metadata Type and Google Image Database

Toru Fukumoto

Canon Inc., JAPAN

fukumoto.toru@canon.co.jp

Abstract: A large number of digital images are stored on the Internet. From an educational perspective, the fact that a vast number of images on a number of topics are readily available and are in some cases free is a very helpful thing. With the rapidly expanding nature of the net, however, it is becoming increasingly difficult to search out and retrieve relevant images. The aim of this study is to analyze users' behavior during image retrieval exercises. Results reveal that users tend to follow a set search strategy: firstly, they input one or two keyword search terms one after another and view the images generated by their initial search; they then navigate their way around the web by using the "back to home" or "previous page" buttons. These results are consistent with existing Web research. Many of the actions recorded reveal that subjects' behavior differs depending on whether the set task was presented as a "closed" or "open" task. In contrast, no differences were found in the time subjects took to perform a single action, or in their use of the AND operator.

Introduction

The number of images stored on the Internet is increasing due to an increase in affordable digital recording devices, such as digital cameras and scanners. In order to effectively manage these digital images, an image album or an image filing system has become the subject of study. As the Internet grows, so the task of image retrieval becomes more complicated. As a result, the need to devise an effective method for retrieving images from a database becomes more important.

From an educational perspective, the role of images as important educational aids is undeniable. Images not only serve to support curriculum and extracurricular subjects but can, if used properly, enhance students' understanding of concepts and issues, and, furthermore, in some cases provide the teacher with an alternative and inexpensive source of study materials.

There are three kinds of image database: feature type, sensitiveness type, and metadata type. Feature type (M. Flickner, 1995) is based on the colors or shapes of objects in the image. When the retriever puts in color or shape, the system immediately starts searching the database with color histograms or shape. In the case of sensitive type (Kimoto, 1999), the user inputs sensitive words and the system exchanges them for color information and searches the database. In the case of metadata type (Kiyoki, 1996), each image in the database is given a metadata tag which defines its characteristics through text or digits. In order to search the database the user has to use the corresponding metadata tag.

For the purpose of this study we focused on metadata image databases. In order to create a metadata type a database creator first needs to define the structure or framework of the metadata. From here, the database administrator needs to attach the metadata to images within the database. It is only at this point that a user hoping to retrieve images can do so by entering the appropriate terms. The database is searched using the user's search term and crosschecked against the metadata issued by the administrator. Examples of metadata are keywords, text, classification items and so on.

Moreover, another type of image retrieval technology has been developed: that is, automatically gathering images inserted in web pages and retrieval by users, for example, Google Image.

There are now accepted International Standards Organization (ISO) regulations governing the structure of contents description of metadata for multimedia known as MPEG-7. MPEG stands for "Motion Picture Expert Group", and although MPEG-7 puts little emphasis on the description of contents in the text, it can extract features from them (i.e., Video stream, Audio, and schema that store them).

MPEG-7 provides the distribution and utilization of multimedia contents using content-based retrieval. This media type will be especially beneficial for distance-learning, streaming databases, personal TVs, etc.

The Purpose of our Research

Over a number of years, a considerable number of studies have been conducted on information retrieval studies for information over the Internet. Many of these studies have concentrated on text retrieval or the analysis of logs. More recently, there has been a new method proposed for assessing text information searching behavior on the WWW (Miura, 2001), however this deals only with text information, not image-based information. There has been further research conducted on images stored from a database (Fukumoto, 2002). The Internet is rapidly expanding and the number of stored images on sites has increased. Because of this, we argue that effective retrieval methods aimed at retrieving stored images from the WWW have become significantly more important. However there have not been many studies that have attempted to assess users' behavior in regard to image information retrieval tasks. The aim of this study is to analyze users' behavior in relation to image retrieval tasks.

Method

Participants

Ten undergraduate students participated in this study. All were accustomed to using search engines on the Internet such as Google and Yahoo!

Materials

We developed the image database based on previous research (Fukumoto, 2000, 2001). The number of images in this database was 1700, with varied contents. The images in this database have linguistic metadata attached to them by specialists. The design of the system can be seen in Figure 1. Google Image was also used. The former, that is, the database of 1700 images, is named "Metadata environment", the latter, "Google environment".

Procedure

Subjects are divided into two groups: members of the first group retrieve first with Metadata environment, then with Google environment; the other group vice versa. They were shown an image and asked to retrieve it so that they could become accustomed to the system. We repeated this procedure twice with a different image on each trial (see Figure 3). These initial tasks were labelled

Task 1-1 and Task 1-2, respectively. We then issued instructions to the subjects asking them to search for images that would be suitable for a “Summer Greetings Card” and to retrieve an image/s from the database. We did not impose a time limit on this exercise, nor did we restrict the number of images that subjects could retrieve; this task was labelled Task 2. During both tasks subjects’ behavior was recorded on video.

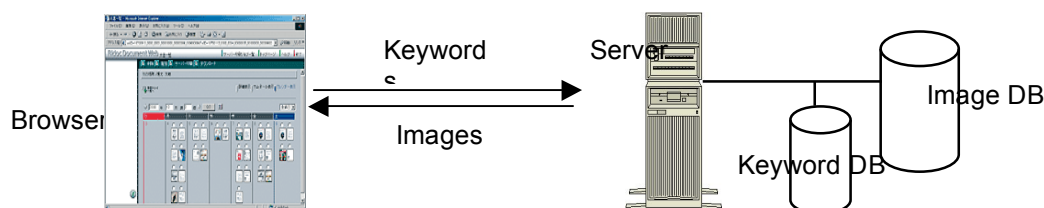


Figure 1: System design

Results

Categorization of Actions

Table 1 shows the ratio of subjects’ actions. Subjects’ actions were classified using an existing classification method for Web Search behavior (Miura, 2001), the main features being 1) keyword operation, 2) Web page operation, 3) browser operation, and 4) other (see Table 1). In both environments all tasks under keyword operation occupied from 40% to 78% of the subjects’ time. In contrast, during subjects’ Web search, Web page operations occupied about 10% of subjects’ time. Keyword operations on all input tasks occupied 40% of subjects’ time. “Back” operations occupied from about 2% to 16% of their time. There were a few “add”, “delete” or “replace” actions. Subjects tended to act as follows: first, they input one or two keywords one after another and allow the system to show images as the result of their search; then they operate the “back to home” or “previous page” button. These results concur with existing Web behavior research.

Table 1: Categorization of subjects’ actions

		Google			Metadata		
		Task1-1	Task1-2	Task2	Task1-1	Task1-2	Task2
Keyword Operation	Input	40.06%	9.84%	41.67%	77.50%	24.43%	54.26%
	Add	3.10%	17.18%	4.17%	0%	3.24%	0.83%
	Delete	9.87%	6.38%	4.17%	2.5%	6.92%	2.56%
	Replace	16.03%	20.00%	7.50%	7.5%	6.56%	7.14%
Web Page Operation	Magnify	10.25%	5.07%	10.42%	10.00%	8.25%	9.49%
	Select Link	0%	0%	0%	0%	0%	0%
	Select Pull-down	0%	0%	0%	0%	0%	0%
Browser Operation	Forward	11.85%	34.59%	18.33%	0%	37.52%	9.62%
	Back	8.83%	6.44%	3.75%	2.50%	13.08%	15.66%
	Home	0%	0%	0%	0%	0%	0%
	Jump	0%	0%	0%	0%	0%	0.43%
	Subtotal	100%	100%	100%	100%	100%	100%
Other	Other	0%	0%	0%	0%	0%	0%



Task 1-1

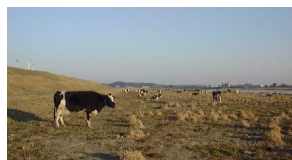


Task 1-2

Metadata Type



Task 1-1



Task 1-2

Google Image

Figure 2: Images of Task 1-1 and Task 1-2

Process of Search

Table 2 shows the ratio of subjects' search processes, that is, page, action, time, and average time per action. We counted the averages for each process and examined the differences between all tasks.

The results of F-test generally show that the three processes—page, action, and time—for Task 2 were used significantly more than those for Task 1-1 and Task 1-2. But there was no significant difference between all processes in Task 1-1 and those in Task 1-2. Details of page function use show that: in Google environment Task 1-2, it was 5% significantly more used than in Task 2; and in Metadata, in Task 2 it was 6% significantly more used than in Task 1-1. Details of actions and times use show that: in Google environment, for Task 1-2 they were 5% significantly more used than for Task 2; and in Metadata, for Task 2 they were 6% significantly more used than for Task 1-1 and Task 1-2. These results illustrate the differences in the nature of the tasks. Task 1-1 and Task 1-2 were presented as “closed task” questions. That is, subjects had to find images that they had already been shown. In comparison, Task 2 was based on an “open task” question and subjects were able to choose whichever images they felt best matched their interpretation of the question. For this reason, we suspect that subjects took longer to search databases for suitable images.

For set actions such as inputting a keyword or going back to a previous page, no significant time differences in the execution of these tasks were found. We believe that this characteristic is not dependent on the task but is instead a personal characteristic.

Table 2: Ratio of search processes

	Google			Metadata			F value
	Task1-1	Task1-2	Task2	Task1-1	Task1-2	Task2	
Pages	7.70	16.40	2.90	1.10	5.40	4.20	F=8.50**
Actions	11.60	21.00	4.90	1.60	10.00	9.20	F=4.55*
Time (MM:SS)	03:38	06:25	01:11	00:28	02:57	02:45	F=4.17*
Time/Actions	00:18	00:22	00:22	00:20	00:19	00:22	F=0.01

**p = .01 *p=.05

Keywords of Search

Table 3 shows the ratio of subjects' keyword operations, that is, the number of input keywords, the number of unique keywords, and the depth of the AND operation. We counted the averages for each process and examined whether there were any significant differences between the tasks.

In Google environment, inputting operations at Task 1-1 and Task 1-2 are significantly more used than for Task 2 (at 5% level); there were no significant differences between Task 1-1 and Task 1-2. In Metadata environment, inputting operations at Task 1-1 are significantly more than those of Task 1-2 and Task 2 (at 5% level); there were no significant differences between Task 1-2 and Task 2. This result suggests that the differences in the nature of the tasks (closed or open questions) significantly affected how subjects performed. However there were no significant differences found in the fixed depth of the AND operation for either of the conditions, or between subjects.

Table 3: Ratio of keyword operation

	Google			Metadata			F value
	Task1-1	Task1-2	Task2	Task1-1	Task1-2	Task2	
Inputting Keywords	12.20	13.40	2.90	1.70	4.00	4.70	F=4.07*
Unique Keywords	5.20	7.10	4.90	1.70	2.90	4.30	F=3.98
Depth of AND	1.80	2.60	1.60	1.10	1.90	1.40	F=1.18

Discussion

There are two kinds of processes described in Table 1. One is a search function that includes keyword operations such as input, add, delete and replace; the other is browse, which includes Web-page operations and browser operations such as magnify, forward, back and so on. According to Table 1, in Metadata environment the ratio of these two processes for Task 1-1, Task 1-2, and Task 2 are 8:2, 5:5, and 6:4, respectively. In Google environment the ratio of these two processes for Task 1-1, Task 1-2, and Task 2 are 7:3, 5:5, and 7:3, respectively

Task 1-1 and Task 1-2 were closed task questions so the search strategy used was more important relative to browsing. In contrast, Task 2 was the reverse of this situation with browsing being more important than the search strategy. This result suggests the following: that a system should be developed that has a fast-browse interface to allow users to search for images that users are already familiar with. This kind of interface would allow the user to choose between options much like the

comparison interface found on on-line shopping carts. A comparison interface would allow users to view at a glance the images retrieved on the basis of their input-query, and, furthermore, give them the opportunity to quickly view a selection of images before deciding if they wish to either revise their search, or continue browsing the images they have already been presented with.

Many of the processes in Task 2—pages, actions, time, inputting keywords, and keyword uniqueness—in Metadata environment were found to differ significantly when compared to the same processes in Task 1-1 and Task 1-2. In Google environment, Task 1-2 was found to differ significantly from Task 1-1 and Task 2.

In Google Help, “Google analyzes the text on the page adjacent to the image, the image caption and dozens of other factors to determine the image content. Google also uses sophisticated algorithms to remove duplicates and ensure that the highest quality images are presented first in your results.” That is to say, the keywords attached to an image that Google automatically extracts, are less accurate than those attached by users. In short, ease of retrieval differs according to these environments. This is shown in Figure 3.

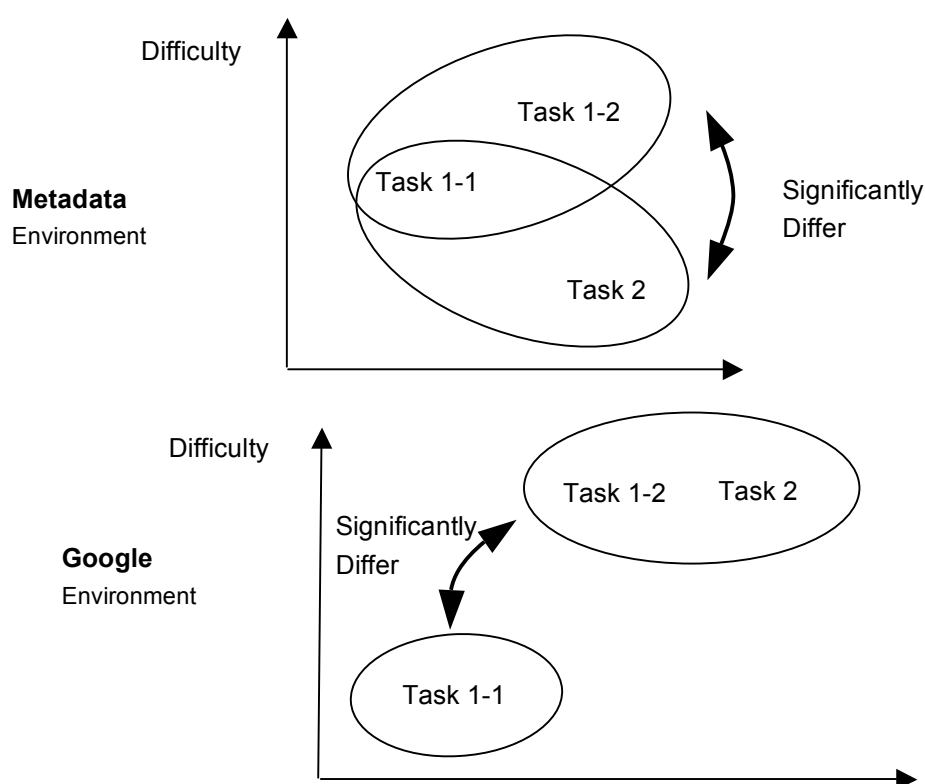


Figure 3: Range of difficulty between tasks

These findings suggest that subjects tend to change their strategy depending on the kind of task set. Subjects tend to follow a set pattern of behavior: first, to input one or two keywords one after another and view the results of their search, and then to operate the “back to home” or “previous page” button. Input actions occupied 40% of users’ time and “back” operations occupied from about 2% to 15%. It was also found that subjects took a fixed amount of time to execute a function regardless of the task, and this remained true for their use of the AND function. We believe that the findings generated by the single function and the AND operator were the result of the users’ own personal characteristics and were therefore not affected by the tasks.

Conclusion

In this paper, we sought to analyze users' behavior in relation to image retrieval tasks. Users' behavior tended to follow a set pattern: firstly, to input one or two keywords one after another and to view the results generated by their search, and then to operate the "back to home" or "previous page" button. These results concur with existing Web search research. More actions were executed when the task was presented as an "open" question as compared to a "closed", and these differences were found to be significant. These results lead us to suggest that users require a choice between a "quick-browse" interface, or a "choice and comparison" interface, depending on the aim of their search. No significant time differences were found in users executing an action, nor were there any significant differences found in their use of the operation AND.

In the future, we hope that multimedia retrieval using metadata will be easier for all individuals. When viewed from an educational viewpoint, we believe that image retrieval would aid teachers in allowing them to retrieve and use more accurate images in order to support their subjects' course materials. Furthermore, it would benefit students by allowing them to easily retrieve relevant images from the Internet.

References

- Flickner, M., et al. Query by image and video content: the QBIC system, *IEEE Computer* 28 (9) (1995) 23–32.
- Fukumoto, T. and Akahori, K. The criteria and evaluation of metadata/keywords in image retrieval, *ICCE/ICCAI* (2000) 542–546.
- Fukumoto, T. and Akahori, K. An analysis of image cognition structure for metadata type image retrieval, *ICCE/SchoolNet* 2001 367–371.
- Fukumoto, T. and Akahori, K. An analysis of image retrieval behavior for metadata type image database, *ICCE* 2002 1470–1471.
- Kimoto, H. An image retrieval system using impressional words and the evaluation of the system, trans. *IPS Japan* 40 (3) (2001) 886–898.
- Kiyoki, Y., et al. A semantic search method and its learning mechanism for image databases based on a mathematical model of meaning, trans. *IEICE Japan* J79-D(4) (1996) 509–519.
- Miura, A., et al. Experimental study of information searching behavior on WWW: categorization of behaviors and examination of difference between tasks, trans. of Japanese *Society for Information and Systems in Education* 18 (1) (2001) 121–128.
- MPEG Web Page: <http://drogo.cselt.it/mpeg/> (accessed on March 1, 2002).