Changes in Users’ Knowledge Structures Before and After Web Search on a Topic: Analysis Using the Concept Map

Hitomi Saito  
Aichi University of Education  
1 Hirosawa, Igaya-cho Kariya-shi, Aichi 448-8542, Japan  
hsaito@ueecc.aichi-edu.ac.jp

Yuka Egusa  
National Institute for Educational Policy Research  
3-2-2 Kasumigaseki, Chiyoda-ku, Tokyo 100-8951, Japan  
yuka@nier.go.jp

Hitoshi Terai  
Nagoya University  
Furo-cho Chikusa-ku Nagoya-shi, Aichi 464-8601, Japan  
terai@cog.human.nagoya-u.ac.jp

Noriko Kando  
National Institute of Informatics  
2-1-2 Hitotsubashi Chiyoda-ku, Tokyo 101-8430, Japan  
kando@nil.ac.jp

Ryo Nakashima  
Sumitomo Electric Information Systems Co., Ltd.  
2-1-3 Nishimiyahara, Yodogawa-ku, Osaka-shi, Osaka 532-0004, Japan  
ryonaka@sei.co.jp

Masao Takaku  
National Institute for Materials Science  
1-2-1 Sengen, Tsukuba, Ibaraki 305-0047, Japan  
TAKAKU.Masao@nims.go.jp

Makiko Miwa  
The Open University of Japan  
2-11 Wakaba, Mihama, Chiba, 261-8586, Japan  
miwamaki@ouj.ac.jp

ABSTRACT
In this study, we investigate the changes in users’ mental representations of a topic during their exploratory search on the Web. We use concept maps to compare the users’ knowledge before and after a search. Participants are required to gather information on the Web as preparation for a regular feature of a magazine on environment and trip topics. They are divided into two scenario groups: divergent and convergent. The divergent scenario requires the gathering of Web pages for an entire series of articles in a regular feature while the convergent scenario requires the gathering of Web pages for a single article of a regular feature. A comparative analysis between the pre- and post-search concept maps indicate that the participants in the divergent scenario change their knowledge widely while the participants in the convergent scenario change their knowledge deeply. The results also show that there were more nodes, links, and link labels in the trip topic than in the environmental topic.

Keywords
Concept Map, Knowledge Structure, Information Seeking on the Web, Exploratory Search.

INTRODUCTION
As information becomes more ubiquitous and the demands that searchers have on search systems grow, there is need to support search behaviors beyond a simple lookup (White & Roth, 2009). Marchionini (2006) defined such searches as exploratory, and he suggested that search systems should be made to support exploratory searches.

In this study, we examine a method for evaluating the exploratory search system. Traditional evaluations of IR systems have centered on evaluating the documents found during searches based on certain metrics, such as recall and precision. However, interest in user-centered evaluations has increased in the IR community. Previous studies on user-centered evaluations and evaluations of exploratory search systems analyzed whether users could effectively seek information and whether they could conduct an exploratory search through interaction with the systems. Moreover, some studies have focused on what users could acquire from the information sources they found (White, Marchionini & Muresan, 2008). However, how to detect changes in the users’ knowledge was not established. Our focus is on evaluating the users’ cognitive changes during exploratory searches. In this study, we use a concept map to evaluate the knowledge that the users acquire and how their knowledge structure changes as a result of their searching for information on the Web.

The Concept map is a graphical representation to allow people to explicitly represent their knowledge (Novak & Gowin, 1984). The concept map consists of concept words, arrows that connect the concept words, and linking words on the arrows:

• Concept words (nodes): Nouns that represent events or objects, such as a car, cleaning, a dog, learning, a chair, and a birthday party. In the concept map, concept words are enclosed in circles or boxes.

• Linking words (link labels): Verbs, adjectives, and conjunctions that represent the relationships among the concept words in the concept map, such as have, like, and is. The linking words are written on the arrows as labels.
• Arrows (links): Relationships among the concept words. Connected concept words and link words make up sentences such as “Plant has a flower.” In this case, the arrow is drawn from “plant” to “flower” and labeled “has a.”

Concept maps have been used as measures to assess the learners’ knowledge and understanding. Meagher (2009) reported that the graph structures of the concept maps became more complex from the first class in a course through the final exam. Rebich & Gautier (2005) also showed that the total number of useful items on the post-course maps increased while the total numbers of weak items and misconceptions decreased.

In the IR community, there have been several studies using concept maps as a means of measuring the change in an actor’s knowledge. Pennanen and Vakkari (2003) explored how a student’s conceptual structure is related to the search tactics and the success of searching. They reported that, between the beginning and end of the overall task, different features of the student’s conceptual structures were connected to a search success in terms of the useful documents they found. Cole et al. (2007) focused on how the students’ mental model diagrams for a topic were represented in an early exploration stage of the information-seeking process, and they suggested a 12-category classification schema of the mental models. Egosu et al. (2010) investigated how a user’s concept map differs before and after a search and how the differences between the topics, scenarios, and browser types influence the user’s concept map. A comparative analysis of the concept maps between the pre- and post-search maps indicated that users significantly changed their knowledge structures on a topic through an exploratory search.

In the educational setting, many studies have shown evidence of the utility of concept maps as an assessment of the learner’s knowledge and understanding. In the IR setting, however, there have been very few research works using concept maps as a measurement of how the user’s knowledge has changed through the search process. The goal of this study is to use concept maps to measure the knowledge a user acquires through exploratory searching.

**METHODS**

**Experimental Design**

In the experiment, we focused on the influences of the topics (environment and trip) and scenarios (divergent and convergent). The participants were assigned to a factorial experiment that included two topics as within-subject factors and two scenarios as between-subjects factors. The two within-subject factors were counter-balanced.

**Participants**

Thirty-two undergraduate students aged 20 to 23 years participated in this study. Sixteen were male and sixteen were female. They were recruited from various departments and universities in the Tokyo area. They pursued various academic majors. They used the Internet and search engines on a daily basis.

**Topics**

The participants were instructed that they were working as a magazine editor and were required to gather information on the Web as preparation for a regular feature of a magazine on environment and trip topics. The environment topic required them to introduce various issues of the environment, while the trip topic was about various destinations for a one-day trip from Tokyo.

**Scenarios**

There are two scenarios, divergent and convergent. In the divergent scenario, the participants were required to gather Web pages for a series of articles in a regular feature; in the convergent scenario they were required to gather pages for only a single article of the regular feature. We prepared tasks for each scenario of the two topics.

**Procedures**

The participants answered questions about their experience in using Web search engines and the Internet on a search-experience questionnaire. They were instructed on the writing method of the concept map and practiced it, and then they were instructed on the task and drew a concept map for the assigned topic for 10 minutes. A blank sheet of paper with a single node of the topic, either environmental issues or one-day trips from Tokyo, was provided to the participants for drawing the concept map.

After drawing the concept map, they conducted a search task for 15 minutes. After completing each search task, they were required to draw a concept map about the assigned topic again and to answer a questionnaire that asked about prior knowledge of the task, their interest in the topic, the difficulty of the task, the difficulty of gathering information, and their comments on the task. Then they repeated the task for the other topic, from the instruction stage to answering the questionnaire.

After the two main tasks, they answered a questionnaire on comparing the two tasks, the changes in their knowledge after completing a task, and comments on how they felt about the changes between the two concept maps, i.e., before and after web search.

In the final session, they were asked to check whether there existed the same concept between the two concept maps, and if they found corresponding concepts, these were assigned the same number.

**RESULTS**

We compared the two concept maps for each topic and scenario, that is, a pre-search concept map drawn by a participant before conducting a search task and the corresponding post-search concept map drawn after conducting the search task. Below, we explain the method of analysis and its results.

**Numbers of Nodes, Links and Link Labels**

We analyzed the numbers of nodes, links, and link labels of concept maps drawn by participants in the pre-search and post-search stages.
Table 1 shows the average numbers of nodes, links, and link labels for each topic and scenario. From the perspective of nodes and links, the 3-way mixed ANOVA with one between-subjects factor (scenario: divergent and convergent) and two within-subject factors (topic: environment and trip, test: pre and post) revealed significant interactions between the topic and test (nodes: $F(1,30)=5.77, p<.05$; links: $F(1,30)=5.16, p<.05$). These results show that the numbers of nodes and links in the topic of trip are greater than those in the topic of environment (nodes in the pre-search: $F(1,30)=30.80, p<.01$; nodes in the post-search: $F(1,30)=9.85, p<.01$; links in the pre-search: $F(1,30)=22.09, p<.01$; links in the post-search: $F(1,30)=8.01, p<.01$). These results also show that the number of nodes and links in the post-search are greater than those in the pre-search (nodes: $F(1,30)=8.15, p<.01$; links: $F(1,30)=7.25, p<.05$). There are no differences between scenarios.

Regarding nodes and links, an ANOVA revealed that the main effects of topic, scenario, and test were significant (topic: $F(1,30)=3.19, p<.10$; scenario: $F(1,30)=13.25, p<.05$; test: $F(1,30)=8.52, p<.05$). These results show that the number of link labels in the topic of trip is greater than those in the topic of environment, the number of link labels in the divergent scenario is greater than those in the convergent scenario, and the number of link labels in the post-search is greater than those in the pre-search.

**Number of Nodes at Each Distance from the Center Node**

We analyzed the number of nodes at each distance from the center node, which is placed at the center of the concept map. Nodes at distance 4 or more were counted together. A node linked from other nodes and that had more than one distance was counted at each distance.

Table 2 shows the numbers of nodes at distances 1, 2, 3, and 4 or more for each topic in each scenario. At distances 1 and 2, which are near the center node, a 3-way mixed ANOVA revealed that the main effect of the topic (distance 1: $F(1,30)=4.71, p<.05$; distance 2: $F(1,30)=9.33, p<.01$) and the interaction between scenario and test (distance 1: $F(1,30)=27.90, p<.01$; distance 2: $F(1,30)=14.23, p<.01$) were significant. The main effect of the topic showed that there are more nodes in the topic of trip than the topic of environment. The interaction showed the differences between two scenarios. In the divergent scenario, there are more nodes in the post-search than in the pre-search (distance 1: $F(1,30)=4.52, p<.05$; distance 2: $F(1,30)=3.61, p<.10$). In contrast, in the convergent scenario, there are more nodes in the pre-search than in the post-search (distance 1: $F(1,30)=28.56, p<.01$; distance 2: $F(1,30)=11.79, p<.01$).

At distance 3, the ANOVA revealed that the main effect of the topic were significant ($F(1,30)=8.42, p<.01$). The result showed that there were more nodes in the topic of trip than in the topic of environment.

At distance 4 or more, which is far from the center node, the ANOVA revealed that the interaction between scenario and test ($F(1,30)=16.93, p<.01$) was significant.

In the divergent scenario, there are more nodes in the pre-search than in the post-search ($F(1,30)=16.85, p<.01$). In contrast, in the convergent scenario, there are more nodes in the post-search than in the pre-search ($F(1,30)=22.58,$
p<.01). These results were opposite to the results of distances 1 and 2.

**Numbers of Common, Lost and New Nodes**

The participants were asked to circle nodes that have the same meaning with a red pen. We define such a node as the common node. We define a node existing only in pre-search as a lost node. We define a node added at post-search as a new node. We analyzed the numbers of same nodes, new nodes, and lost nodes.

Table 3 shows the average numbers of common, lost, and new nodes in each scenario. A 3-way ANOVA revealed that the number of common nodes was not significant. From the perspective of the lost and new nodes, the main effects of topic were significant (lost node: $F(1,30)=21.98$, $p<.01$; new node: $F(1,30)=8.81$, $p<.01$). These results show that the numbers of lost and new nodes in the topic of trip are greater than those in the topic of environment.

In total, there were few common nodes and relatively many lost and new nodes. These results suggest that the concept maps greatly changed after searching on the Web.

**DISCUSSION AND CONCLUSIONS**

From the analysis results, we found that searchers' knowledge of the topics changed dramatically between before and after searching on the Web. The results also showed the differences between the two topics and between the two scenarios.

First, we discuss the differences between the topics. The numbers of nodes, links, and link labels in the trip topic were greater than those in the environment topic. Moreover, the results of distances 1 and 2 from the center node also showed that the trip topic had more nodes than the environment topic. These results indicate that there are differences in the knowledge structures of the topics and that the concept map could represent these differences.

Next, we discuss the differences between the scenarios. The scenarios have no influence on the number of elements in the concept maps. However, they influenced the construction of the concept maps. In the divergent scenario, the nodes that were near the center node increased while the nodes that were far from the center node decreased before and after search. In the convergent scenario, the nodes that were near the center node decreased and the nodes that were far from the center node increased before and after the Web search. These results indicate that the scenarios could influence the searchers' information-seeking and that the concept map could reflect information that the searchers found during their search.

To confirm these speculations, further investigation needs to reveal the relationship between the changes in the searchers' concept maps and the searchers' information-seeking behaviors.

**ACKNOWLEDGMENTS**

A part of this research was supported by the Grant-in-Aid for Scientific Research (B) No.21300096 and the National Institute of Informatics joint research grant.

**REFERENCES**


<table>
<thead>
<tr>
<th></th>
<th>Environment</th>
<th>Trip</th>
<th>Environment</th>
<th>Trip</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common Nodes</td>
<td>7.13</td>
<td>3.67</td>
<td>8.75</td>
<td>4.56</td>
</tr>
<tr>
<td>Lost Nodes</td>
<td>13.69</td>
<td>8.08</td>
<td>20.81</td>
<td>6.96</td>
</tr>
<tr>
<td>New Nodes</td>
<td>17.38</td>
<td>7.47</td>
<td>22.13</td>
<td>6.81</td>
</tr>
</tbody>
</table>

Table 3. The average number of common, lost, new nodes.