Exploring Information Gathering Process in Networked Environments

Petek A_KAR[1], Arif ALTUN[1], Vildan ÇEV_K[1],
paskar@hacettepe.edu.tr, altunar@hacettepe.edu.tr, vildanc@hacettepe.edu.tr

Abstract

In this research, the steps under the process of information gathering are analyzed in the context of information literacy which is one of the most important skills required for students. Information gathering process in networked environments is modeled in this research by using qualitative methods. Case study is used with purposeful sampled three people who are studying for their graduate level education. Information gathering process which is starting with "reading and saving task" and ending with "evaluation" is modeled in a circular form. People can benefit from this model while developing instruction or training to support the learning of tasks (or steps), and also developing diagnostic tools and assessment forms in order to determine whether individuals have mastered and/or what to-be-learnt steps of information gathering processes are.

Keywords: Information literacy, information gathering, think-aloud.

1. Introduction

Information literacy is one of the essential skills for student on the way of becoming a lifelong learner. An information literate person is the one who can both benefit from library services and tools in the traditional sense and effectively use the sets of information presented on the internet and information searching tools [1]. Students are expected to navigate, comprehend and collect the specific data from networked environments by having the skill of information gathering and its sub-skills of accessing to accurate and reliable information. The transition from one-dimensional concept to multi-dimensional one [6], extended the available sources from physical to networked environments. Consequently, a bulk of texts has been transformed into hypertexts with links and connected these resources creating a new environment: networked information.

Information searching /seeking /gathering are used interchangeably in the literature referring to the information search and collection process in a broader context. While the former two correspond to searching, the latter can be thought as a combination of searching and collecting processes. In this study, the concept is used referring to "information gathering" (IG) in particular. Kellar, Watters, and Shepherd (2007) defined this process as the process of collecting information from many sources and may take a day to several days [10]. Since there is not a single correct answer at the end of this process, students cannot decide when the process of information gathering ends. In the meantime, many pages to review and too much time to spend make this process more complex. Writing a research paper for a biography, searching different car models before getting a new car, making plans for the upcoming summer vacation…etc. can be given as examples for understanding of the process of information gathering [10]

The search with the keywords of both "information gathering" and "searching" in the current literature reveal that studies on information gathering was explored participants' IG processes while using either the libraries [4, 8, 14], or the internet [5, 9, 10]. Emphasizing the fact that most research had been conducted 5-10 years ago, Kellar et al. (2007) have pointed out that the ways of accessing information varied in recent years; and information gathering skills have been redefined, especially with the use of web 2.0. tools (bookmarking, RSS .. etc..), web services and development of semantic search. Library searching skills compared to Internet is also considered as different in terms of the complexity of the tools and resources used in searches [10].

Brand-Gruwel, Wopereis, and Ywonne (2005) modeled the complex skill of information problem solving in their research [2] by starting out Eisenberg's Big6™. They classified information problem solving as a complex skill and adapted the process to the regulation component in Big6™. The sub-skills they revealed are as follows: defining the information problem, selecting information sources, searching and finding information, information processing, editing and presentation of information. This last component is a

[1] Hacettepe University, Dept of Computers Education and Instructional Technologies

structure that plays role at each stage. Then they analyzed the tasks by classifying these skills in sub-categories.

Some researchers emphasize the need for research focusing on the *role of micro-level factors in this process* [e.g., 12]. Examining the cognitive process or related skills alone could be too broad to develop interventional models. When the process and required cognitive skills are modeled at the micro level with the understanding of the nature of tasks, researchers will easily infer the skills and sub-skills expected from individuals.

In the light of these studies and arguments articulated above, it has been aimed to answer the following questions consecutively: "how information literate individuals collect information", "how can we model their process". The findings of this study would guide instructional designers to develop instruction or training to support the learning of tasks (or steps). Moreover, diagnostic tools and assessment forms can also be developed accordingly. Therefore, the primary goal of this study is to explore the procedural patterns in information gathering process from experts' point of view and to provide a working model of information gathering to be used in designing instruction.

2. Method

2.1. Research Design

This research is a qualitative study designed for revealing the information gathering processes of individuals in networked environments. Qualitative research contains deep, rich description and is more concerned with process than specifying outcomes or products [3].

In this research "case study" which is one of the research design methods is used. According to Yin (1984), case study is a kind of research method in which the focus is on the actual phenomenon within its real-life context, boundaries between phenomenon and its context are not clearly evident, and there are more than one proof and data sources in that situations [15].

For providing construct validity of the relevant research more than one type of data (triangulation) was recorded in data collection process. It can be said this study has both internal and external validity because of giving sample sentences from findings (internal) and external validity with the analytical generalization to the studies in related literature. To provide reliability the researchers has defined the all paces of the study clearly and saved all documents and data electronically.

2.2. Participants

This research was carried out with 3 people who are studying for graduate level education. Two of these people are studying on Computer Education & Instructional Technologies for master's degree and one is studying on English Language and Literature for integrated PhD education. The first Participant (P1) is in 5th term and has finished taking lessons and also writing his master's thesis. The second participant (P2) is taking lessons of 3rd term and the 3rd participant (P3) is in 3rd term of integrated PhD education. P1 is male and 25 years old, P2 and P3 are both female and 23 years old.

Qualitative researches allowing people to work in depth of the cases that are thought to be rich with information so in this research people are selected by purposive sampling (Yıldırım & _im_ek, 2006). Purposeful sampling is generally used in case study research; therefore, explain sampling procedures and case selection, and the defining characteristics and typicality or atypicality of the case (TESOL, 2009). The process of information gathering was only been limited to the internet sources it has been considered of participants to be able to use computer and internet well. Participants were asked whether they could reach easily to the information they searched on the internet or not and all three participants expressed they had any trouble in this regard. Just as participants are continuing educations for graduate-level and they already carrying out researches and preparing reports about the issues they interested in. Overall, it would not be wrong to claim for these individuals that "information gathering" process has become almost a part of their life.

2.3. Data collection and analysis

For preparing a presentation on "How developments in genetics will affect human life", participants were given a task of thinking aloud while gathering information on the internet. This method is widely used while studying cognitive processes, such as problem solving, learning, decision making, human–computer interaction, and cognitive task analysis [7].

Thinking aloud is explained to individuals by providing examples. Researchers reminded each participant to think aloud when he/she becomes silent in the process of gathering information. While choosing the task to be searched, it has been ensured that participants did not have any prior knowledge about the topic chosen. Also, it is assumed that with low or non-existing prior knowledge, participants could be more reflective while thinking aloud in the process.

In case studies more than one data collection method is usually chosen for trying to achieve rich data diversity and the data which can confirm each other [15]. While information gathering processes were being recorded by video-camera, the speech of participants were recorded by sound-recorder and an observation record was noted by the researcher. Thus, triangulation was ensured so that it is aimed to increase the construct validity and credibility of this research [15]. After the first interview with P1, it was seen that the sound was recorded clearly while video camera was capturing screen, so the sound recorder was not needed and then the speech data of P2 and P3 are written from video recordings. All data were taken electronically and saved in a directory so, in addition to the verbal data, electronic data is also kept.

While transcribing the recorded data, Microsoft® Office Word 2003 software is used. The screen captures are saved by using computer software that comes with the video camera and in the process of data encoding, when the text is inadequate, current screen image was considered. In observation report, computer and internet self efficacy perceptions of participants are noted, and also the reminder tips about participants are written. General information summarizing participants is presented in Table 1.

Table 1. General preferences of participants

Participant	Age/ Sex	Department	Participation period	Data size	The numbers of participants get/download about given issue			
					Total links	ppt files	Documents (pdf, doc)	keywords
P1	25, M	CEIT	10:45 - 11:45	376 line	27	12	9	32
P2	23, F	CEIT	15:38 - 16:20	193 line	6	-	3	18
P3	23, F	ELL	14:57 : 16:00	367 line	6	-	3	9

3. Findings

We observed the expert performance which is video recorded and coded by using of think aloud protocols and created the model of information gathering process as shown in Figure 1. It can be seen from the model that it is cyclic. Finding, accessing, and storing the information had various steps and runs in circular form, not hierarchical or linear [13].

The model that thought to be ideal has been created by bringing together the information gathering process of three participants. The arrows in the model show the direction of the relationship. Bi-directional arrows symbolize the two way movements between related steps, and one way arrows symbolize only one way movement. In evaluation level, there is a condition, and the process will either continue or halt according to the answer; Yes or No. Because of related model operates on the internet while gathering information, it is the pre-condition of individuals to be able to use computer and internet at basic level. The steps of model are explained below.

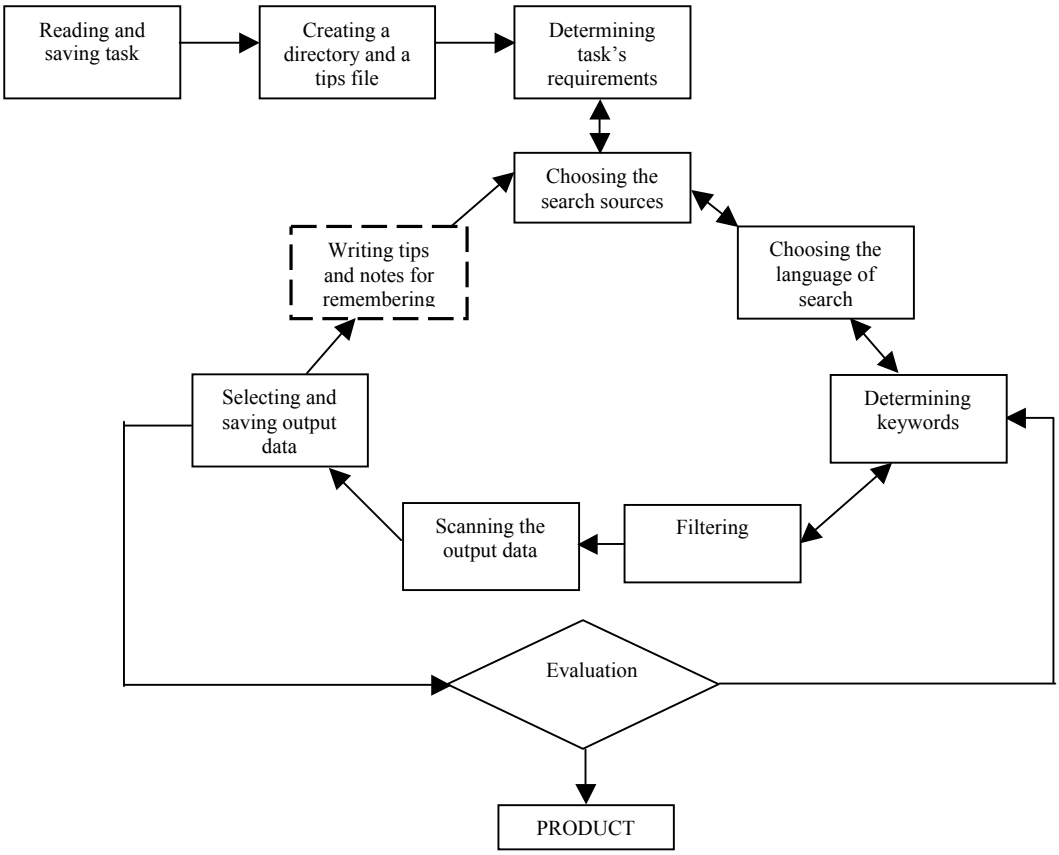

Figure 1. Modeling of information gathering process

Reading and saving task:

The initial state of information gathering starts with reading. In other words, the input for the process is initiated by encountering the task. Individuals read the task and develop strategies, such as writing it down or saving it as a file in their computers. Saving it as a file can be considered as a note-taking strategy for further reference.

P1: May I write the subject (title of given issue)? (And opens a .txt file for writing)
P3: Ok, I'm writing on a paper.

Creating a directory and a tips file:

Individuals create a directory in which they will save all content they gathered and then rename it. And create a tips file or a notepad file in that directory for entering notes and/or keywords as a brainstorming area.

P1: Ok, Let me write the framework of the given issue. I'll talk about what is genetic, what it is trying to do. And, the studies on what has been done so far, what is being done currently with two perspectives; in our country and in the world. Then, in what areas it can affect human life, I try to detect what are these such as food, health…etc. . Then I'll give examples about these and in the future how can be affected in these areas?

P1: For example, here it says it could be used of the production of human organs. I can look something about this and then get information to my presentation. This is also an important thing in terms of developments in genetics. Let me write this to my tips file, Hmmm, because it is an interesting topic, production of human organs. This would be the biggest effect of genetics I suppose.

P3: I'm copying (selects all and then copies and pastes into MS Word.) Then, this is my own personal habit of writing where I get this from. Here the address is genetikvebilim.blogcu.com. I usually do it at that moment because if I use it for a research later I do not need to go back and look its source again. Let me write, on 5th December 2008.

Determining the requirements of task:

This step involves the determination of requirements, to which they will need in order to accomplish the task by the end of the process. For example if the task requires them to prepare a presentation, individuals should search audio-visual data (video, audio, images, graphics ... etc..) in addition to text. In a similar way, the expected presentation would be less intense than a critiquing task. If the task requires them to prepare an academic presentation, or to write a term report, they would determine different requirements for that particular task.

P1: If I'm going to prepare a presentation I look directly to the ppt sources. For example with the keywords of "genetics, how will affect human life". I'm designing the frame of topics, right?

P1: While designing the framework of given issue, I've taken the advantage of my prior knowledge about preparing presentations rather than the searches I made up to now. I created the framework by asking myself how can I explain this subject. Now I will identify keywords according to this framework.

P1: If I'm preparing for presentation I benefit from the interesting points for attracting people. I determine the interesting points while I'm reading and search something related them specifically. For example, I have given as an example of cloning sheep and try to find some visual graphics. What was the name of that friend, Dolly? I search with keywords of Dolly.

P3: If I'll make a presentation I search more pictures and graphics. Now I'll search visual graphics by looking Google Images with keywords of genetics and human and then copy the pictures I've reached.

P3: For example, Times. Maybe I can say something to audiences "it is written like this in Times", "according to Times this is the recently found" ...etc.

P3: And, I know BBC has documentaries and I want to look if they have anything about the given issue.

Choosing the search sources:

Where will an individual gather the information from? Search engines, databases and online libraries, e-books or video and/or photo sharing sites? In this step, the nature of the task is again a determinant factor for experts to decide what source they need to choose:

P1: For example I'm taking this all and searching from Google at first.

P1: I look at images for preparing my presentation more attractive. And then search Wikipedia and read what it says about this issue. And I search with the keyword of "genetics"

P1: For example I try to find something striking from Youtube. I want to make my presentation more visual and show people the information with a different point of view. I'm busy with this idea now. I'm searching now for making the presentation more visual. I want to prepare attractive and well designed one which has less texts.
P1: I don't search the databases like Science Direct, because now I need general information more.

P2: At first I'm opening Google which is the most common search engine I've used. Then I'm writing the issue as a whole rather than separated by commas like genetics, human life and developments in genetics.

P2: I can open academic articles from here for getting other researches. I search from Scholar Google again.

P2: In fact, I can search from normal Google search page with English keywords.

P3: At the same time I want to search with the databases from our university, and opening hacettepe.edu.tr, from students, libraries. In fact, this is interesting. Interesting for me. I always used for the social sciences databases previously, I will look now to medical databases or those related to biology at the moment. I'm searching. Dermatology, Pediatrics, Neurology…etc.

P3: hmmm, I intend to search from Google scholar before this. What kinds of books are there related to this issue and what kind of books I can get. I want to look at them and I'm opening Google scholar.

Choosing the language of search:

An individual can reach rich data by gathering information with different languages which individual feel himself/herself competent enough.

P1: Let me search by writing "Genome Project". And look what is called in English. Human Genome Project information, impact on human health. These are important keywords for me.

P1: The human genome Project and its impact on society. I'm searching on a page in Turkish and another page in English. In fact I couldn't find enough information in Turkish. Future medicine, for example I couldn't find these titles, so I'm looking interesting titles in English. Let me take this, and save.

P2: I can look in English, Is it written as "genetics", let me look it. (opens a free dictionary for looking how is genetics' spelling). Maybe I can reach more sources when searching in English.

P3: hmmm, I decided to change my research language because of getting no more interesting results from Google.

Determining Keywords:

An individual should produce keywords from the given task for further exploration. While gathering information, keyword production is an ongoing process and after producing new keywords, all other steps in cycle continue.

P1: Hmmm… Let me add "future" and look like that. "future genome research" I'm saving this. Save link as. The future of genome Project for medical research. This is related to the thing I'm searching. And now, I'm adding "future" and "impact" to my keywords.
P1: Look I still am looking for new keywords. Yes, we found the thing we searched within this page. It was important for me. Future, impact on society. These are my keywords.

P1: I'm searching with "Genome Project", "drug". And adding "impact". "impact on drug". I'm looking this. ***Why did you add drug?*** *I've seen it in a research which I didn't save about diabetics. I've written tips and when I've seen now, I remembered "drug" again. When I see an interesting word, I'm looking and searching it. Because I will add this to my presentation.*

P2:I usually see "human development" so I'll change it with "genetic development".
P2:Let me search with genetic cloning in English. Genetics cloning, human clonning
P2: It has been written here about genes and as a science of genetics. I wonder if I get genetics as science. Genetic science, developments.

P3: Now there is a sentence as "genetics and danger" in this study. And it gave me another keyword; "genetic risk".

Filtering:

Individuals limit his/her search according to file or data types. This process is called "filtering". For example, if an individual wants to search some particular file extensions (such as .pdf, .doc, .ppt …etc.), he/she can filter the search during information gathering process.

P1: Now I'll search the same keywords for looking to pdf files, but I won't read content of them I'll look just titles and decide if there was an interesting one. Let me search, future impact of human genetics filetype: pdf

P1: Human genome Project information, and let me search adding of it filetype:ppt

Scanning the output data:

The next step is scanning. Individuals must have skills of reading on computer screen, scanning the text on the monitor and using the preferences of internet browser (such as under Edit menu – Find). In addition to these computer-related skills, individuals are expected to be able to answer the following questions about the data they accessed:

- Is this information related to the given issue?
- Is it dependable?

- Is it actual?

P1: This resource is good, let me save it. I'm looking for if there were any other keywords in this text. Because it's a good resource. I'm looking to its bibliography. For example there is a given web address here, I can look what it has.

P1: Hmmmm.. for example I can look if there was a word of "gene" in the lines of this text? (Opens menu of Explorer and run "Find") (find -> gene)

P1: Ethical values under ethics, I can take this, oh no, I don't think it is useful because of being k12.edu.tr.

P1: I found a thing here, I'm looking the source of this, who has done, it's belong to tip.hacettepe.edu.tr. We can say this is a reliable web site because of its education connection.

P2: Human cloning will be the most important development. Genetic modification. It's written here deeply. And in this page it is more biological. It says genetics engineering here.

P3: If it's a blog then it's a simple source. Because, what people usually do with blogs? Iuummm, they usually make brainstorming. Such as, they usually share their ideas with others. And it takes place in an informal way. It's different from journal articles because they usually are not analyzed totally. They are written in daily language.

Selecting and saving output data:

Individuals select, save or eliminate data according to the answers to the questions asked in the previous step. Also in this step an individual must record bibliographies and create references. Individuals can use software or if he/she does not have an adequate level of technical knowledge they can index saved pages manually. At the end of this step individual can choose two paths. One is evaluating all data he/she had another is keeping on searching.

P1: Hmmm, it says here "developments in medicine". And, it is written here "developments in medicine" too. And there is another one. The three of them has the same texts I think. Let me look and compare each other. It says "The effects of developments in medicine to humanity". I'm looking now if all of them are the same. Yes I'm closing two others.
P2: I can use this while deciding the titles of presentation. So I save this.

P2: I've seen a title named "life that is reduced to the genetics". It requires membership to save the file, so I directly save the page. There is an article about genetics here. Hmmm.. There is a sentence I interested in here, ...so there is a relationship with human. I'm saving this page too.

P2:These are the resources I've reached before. There is something here about recent developments and it says more about technology. I probably can use this. So I'm saving. There are academic studies here. G evaluation, N evaluation, genetic technology.. Let me save this file too.

P3: And,, genetics and cell biology. I'm opening this as a pdf document and I'm saving this by renaming of "genetic1" to the folder of "genetic" which I opened on my desktop. And save it.

Writing tips and taking notes for remembering:

Individuals write down and keep notes for further use, such as "return here later", "look at later", "… is a page with too much detail about …", "you definitely need to find this source" etc. In their cyclic process, they do not always follow this step; therefore, this process was drawn with cut point in Figure 1.

P1: Ethics and the point of genetics science has reached. This can be a good title for me either. Hmmm, I'll add an item here under the title of health. Illnesses and their reflections on Drug Industry. Because drugs are also related with genes in future (Writes the titles as a framework of his presentation in tips file).

P2: In fact I can write these one by one. Maybe by opening a new document from here. Generally "probable developments" and "their effecs on human life". Cancer, making longer of human-life, human. It says here the concept of "personality" will be solved. Gene teraphy, I've taken these related with cancer. Skin cancer,

"sleepiness gene" and arranging the biological clock. In fact this is the general title, I can move this to top. Second title is human cloning.

P3: But I want to copy the name of this journal because if I've had to focus on it at the end of my searches I want to look it again. And copying it to my word document and taking note as "there are more researches studied in 2008 so go back again!"

Evaluation:

Evaluation is the exit point of the information gathering process. At the end of the evaluation step, if individuals believe that they had gathered enough information and they make meaning out of data, then they are ready for preparing the output –whatever the task requires them to prepare. Otherwise, model operates in its cycle from the beginning, which is determining new keywords for further gathering.

P1: Normally, I have a break after searching and evaluate what kind of sources I've.

P2: Generally, I design a framework maybe in a word document. And I start to prepare the presentation at the end of my searches. When I totally searched and get everything I start to prepare my presentation. When I believe that was enough and got everything about that issue? If I get enough data to fill the titles and subtitles according to my framework, I can decide to finish.

4. Conclusion

In this research, information gathering as a process was modeled. To understand and explain the steps of searching and collecting information, there is a need for investigation this process in a context [12]. Researchers have created a context based on a given task of information gathering process and explored this process from experts' point of views.

Three participants in this study are from two different departments (CEIT & ELL) and they tend to reflect their background and study strategies during the information gathering process. For example, when "reading and saving the task", the two participants from the department of CEIT recorded and saved the given task on their computers while the other participant, who is from the department of ELL preferred to write it down on a paper. The same difference between individuals can be seen on the strategy of "creating a directory and tips file" such as the way they preferred. While two participants from CEIT created a directory and saved the shortcuts of the web pages in it, the other participant from ELL created a word document and then copied and pasted the pages and images in it.

The findings indicate that P1 is the one who had the most effective and productive result in terms of the numbers of files, documents, links and keywords gathered. P1 has also the most efficient computer skills (for example, filtering strategy used only by him). These findings might indicate that "having effective computer skills" enabled this person to be successful on information gathering process in networked environments.

Another finding of this research indicated that participants preferred different search engines and changed the search engines according to their needs. For example, they preferred Wikipedia™ at first for getting general information. If they needed more academic sources, they preferred to look the databases such as Science Direct™ and Google Scholar™. For getting information about scientific developments they preferred to look at BBC™ and Times™. They preferred Youtube™ for video search and Google Image™ to get visual materials (like photo, image…etc.). Apparently, knowledge about search engines and deciding on which one to choose become an important competency to be developed for novices.

Expert individuals are attended to this study by purposeful sampling. The results can be helpful for novices as a road map and the suggestions below can be provided based on the findings of this study:

- This model can be applied to teaching information gathering process to novices
- Assessment forms can be designed to determine how individuals perform through the information gathering process
- Certain diagnostic tools can be developed to identify the problems and issues for each individual to pinpoint the pitfalls in information gathering process
- The effect of computer expertise in information-technology based search could be explored with more diverse participants in computer experience.

Acknowledgement

Researchers give special thanks to three voluntary participants.

References

[1] Altun, A. (2005). *Geli_en teknolojiler ve yeni okuryazarlıklar (Emerging technologies and new literacies)*. Ankara: Anı Yayıncılık.

[2] Brand-Gruwel, S., Wopereis, I., & Ywonne, V. (2005). Information Problem solving by experts and novices: analysis of a complex cognitive skill. *Computers in Human Behavior, 21*, 487-508.

[3] Campbell, T. (1996). Technology, Multimedia, and Qualitative Research in Education. *Journal of Research on Computing in Education, 28*(5), 1-6.

[4] Chang, N.-C., & Perng, J.-H. (2001). Information Search Habits of Graduate Students at Tatung University. The *International Information & Library Review, 33*(4), 341-346.

[5] Gerjets, P., & Hellenthal-Schorr, T. (2008). Competent information search in the World Wide Web: Development and evaluation of a web training for pupils. *Computers in Human Behavior, 24*(3), 693-715.

[6] Glenn, M. (2008). *The future of higher education: how technology will shape learning* (No. http://www.nmc.org/pdf/Future-of-Higher-Ed-(NMC).pdf): The Economist Intelligence Unit (N. Publications o. Document Number)

[7] Gwizdka, J., & Spence, I. (2007). Implicit measures of lostness and success in web navigation. *Interacting with Computers, 19*, 357-369.

[8] Hart, R. L. (1997). Information gathering among the faculty of a comprehensive college: Formality and globality. *The Journal of Academic Librarianship, 23*(1), 21-27.

[9] Hyldegård, J. (2006). Collaborative information behaviour--exploring Kuhlthau's Information Search Process model in a group-based educational setting. *Information Processing & Management, 42*(1), 276-298.

[10] Kellar, M., Watters, C., & Shepherd, M. (2007). A Field Study Characterizing Web-Based Information-Seeking Tasks. *Journal of the American Society for information science and technology, 58*(7), 999-1018.

[11] Nisbett, R., E. , & DeCamp-Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231-259.

[12] Pharo, N., & Järvelin, K. (2004). The SST method: a tool for analysing Web information search processes. *Information Processing & Management, 40*(4), 633-654.

[13] Rumsey, S. (2004). *How to find information*. Berkshire, GBR: McGrawHill Education.

[14] Sadler, E. B., & Given, L. M. (2007). Affordance theory: A framework for graduate students' information behavior. *Journal of Documentation, 63*(1), 115-141.

[15] Yıldırım, A., & _im_ek, H. (2006). *Sosyal bilimlerde nitel ara_tırma yöntemleri.* (*Qualitative Research Methods for Social Sciences*) .Ankara: Seçkin Yayıncılık.