
SEARCHING THE GREEK WEB SPACE: AN INSTRUCTIONAL APPROACH FOR EFFECTIVE UTILIZATION OF SEARCH ENGINES

FOTIS LAZARINIS

Technological Educational Institute of Mesolonghi, 30200 Mesolonghi, Greece
lazarinf@teimes.gr

Abstract

This paper proposes an instructional approach aiming at helping students efficiently utilize search engines. This methodology considers the explanation of the internal search engine intelligence as a basic structural element, especially in spoken languages with inclinations and intonation like Greek. As demonstrated from the application of our teaching approach, students were able to successfully discover specific information on the Web in a short amount of time compared to the students who were not aware of the search engines' limitations.

INTRODUCTION

The ability to effectively search and locate information on the Internet is an important skill in the Information Society (Schlein, 2002) and an important tool in the process of learning (Large & Beheshti, 2000). To successfully utilize search engines one has to possess some knowledge of their internal intelligence. That is of how they operate and of their limitations, so not to get overwhelmed and confused by the possibly relevant retrieved documents (Henry, 2005).

Proper teaching of searching techniques and search engine potentials is important in creating truly Internet literate users (Eagleton & Guinee, 2002; Leu et al., 2004). This is even more crucial when the query terms are in a language with inclinations and intonation like Greek. A recent study showed that searching using Greek terms is more demanding than in English Web searching (Lazarinis, 2005). Greek searchers must be aware of the disability of international and even local engines to value the Greek language. It was shown that queries with capital words or words with intonation produce different results than in lower case query terms. Moreover stopwords (i.e. prepositions, articles, etc) are not removed and stemming (i.e. removal of prefixes of suffixes) is not supported. Therefore documents where

query terms appear in a different inclination are not retrieved. These factors substantially affect retrieval performance. Especially novice Web users most of the times get no results and they abandon their information hunting. Stopword removal, stemming and capitalization or more generally normalization of query terms are amongst the oldest and most widely used retrieval techniques (Baeza-Yates & Ribeiro-Neto, 1999), at least in English text retrieval.

Based on these observations we suggest and evaluate a 6-phase methodological instructional approach for supporting students in effective utilization of search engines. In the next sections we present an experiment based on our teaching approach. The experiment was unfolded in two phases. During the first phase the previously mentioned characteristics were not taken into account during the instructional process whereas the second is built on top of them, thus our approach was applied. These teaching practices exercised in two separate groups of students with the same characteristics though. Their performance is evaluated by an assignment at the end of the teaching procedure.

RELATED WORK

Learning to Search the World Wide Web or identifying the e-skills needed to effectively utilize web search engines is a subject negotiated in a few other studies as well. Even from the early days of search engines some researchers tried to identify the e-skills needed (Brandt, 1997). Lazonder (2001) suggests some efficient ways of using www search engines. His ideas are based on teaching the basic skills required to use a search engine and then let students discover information on the web. During their quest the assistance provided to the adult web searchers is gradually reduced. Another study (Eagleton & Guinee, 2002) concentrates on developing the

required e-skills through various examples and via evaluation of the returned results. Nachmias and Gilad (2002) examine the process of information search on the Internet and assess the participants' success in finding specific information on the Web. They try to identify the search strategies used and assess their effectiveness. In a more recent study the competencies for successful utilization of search engines are identified (Henry, 2005). Case studies which concern the utilization of search engines by students so as to discover specific information appeared in Greek studies as well (Lazarinis & Kougiourouki, 2003).

Although these studies contain didactic examples which could be taken into account when planning a relevant teaching activity they do not attempt to develop an instructional approach. Such a methodology should take into account the options and limitations of existing search engines and should be adaptable to non Latin spoken languages which have inclinations and intonation, so as to be broader.

INSTRUCTIONAL METHODOLOGY

Our approach is based on the following six steps: (i) Search engine utility explanation, (ii) Search engine access, (iii) Query explanation, (iv) Sample query execution, (v) Navigation of result set, (vi) Evaluation task. These steps are based on the classical instructional paradigm where an instructor explains the utility of a software tool and then through some examples explains the functions of the tool. Finally an evaluation task is given to the learners in order to measure the success of the teaching procedure and to spot potential problems so as to refine the teaching process in the future.

This instructional method was applied with two different forms. In the first type, called the condensed form, we applied the six steps to two classes. The first class was a high school class consisted of 20 students. The second adult class frequented by 26 trainees. In the condensed form of the methodology we gave some sample queries to the students without being analytical of the limitations of search engines and of their internal features which should they are utilized they could change the retrieved set of documents. So the explanation of phases iii and iv of the methodology was a straightforward procedure where students had to type some words in Greek and navigate through the results.

In the second expanded version of the method we revised steps iii and iv to take into account differences resulting from points such as capitalization, stemming and stopwords. This form was tested again into two classes; one high school class of 19 students and on an adult class of 28 learners.

In both cases each participant had its own computer so as to practice the given examples and to perform the assignments asked at the end of the teaching activity. Each instructional activity lasted two hours.

FIRST TEACHING APPROACH

In the 20 high school student class and in the 26 adult group teaching went through the following steps:

(i) *Search engine utility explanation*: Initially the usefulness and the significance of search engines were explained to the students.

(ii) *Search engine access*: Google (www.google.gr) was accessed and the structural parts of its standard localized interface were explained. We used Google as it is considered the most powerful search engine while having a simple and localized interface.

(iii) *Query explanation*: We explained the steps needed to search for information using Google. Additionally an explanation on "broad" and "narrow" queries was given to the participants. Broad queries are those composed by one or two general terms which have several meanings in different context or do not adequately describe the information need (e.g. Athens). Narrow queries are those who consist of more than one, usually specialized, terms which better describe the information request (e.g. Athens Georgia United States).

(iv) *Sample query execution*: Four sample requests executed in Google. Queries suggested by the participants. All terms were written in lower case letters with accentuation.

(v) *Navigation of result set*: Participants visited the first URLs of the result set to realize how the result navigation mechanism operates.

(vi) *Evaluation task*: A task was distributed to each participant. The task composed of two sets of queries in Greek. The first set consisted of four one-word or two-word queries.

These words were specialized computers related terms (e.g. binary system). Two queries were in capital letters without intonation and the other two were in lower case letters with intonation. The second group of queries consisted of descriptions of the information need rather than specific queries. So participants had to construct their own queries.

SECOND TEACHING APPROACH

The second teaching approach practiced in the 19 high school student class and in the group of 28 adults. The former 6-step teaching approach altered to cover the additional requirements posed when searching the Greek Web space. The steps of this modified version are (only the altered phases are explained):

(iii) *Query explanation*: In addition to the previously analyzed practices (broad, narrow queries) we explained the following factors which affect retrieval in Greek Web searching:

- Difference between upper and lower case typed terms.
- The importance of accent and other intonation marks in searching.
- The role of stopwords which are not automatically removed as in English www searching.
- The significance of suffixes and especially of the final sigma (which is not used only in the plural form as in English, e.g. both "Υπολογιστής" and "Υπολογιστή" mean Computer and "Υπολογιστές" means Computers).

Search engines extract the important terms from web pages and they use them as an index. Then they match the query terms to the index terms to retrieve relevant documents. As mentioned, there is no support for stemming, stopword removal or any other form or normalization of query and index terms in Greek web searching (Lazarinis, 2005). Thus documents where accent marks are absent or where the index terms appear in a different conjugation than the query terms are not retrieved, resulting in lower success.

All international and most local Greek search engines produce different results when the same query is typed in upper and lower form. For instance Google retrieves 122000 documents when we seek information about "Ολυμπιακοί αγώνες Αθήνας" (Olympic games in Athens) while it retrieves only 288 in the query "ΟΛΥΜΠΙΑΚΟΙ ΑΓΩΝΕΣ ΑΘΗΝΑΣ". Obviously some of the results are common but in any case the number of retrieved web documents is quite different. Most important is the observation that the top results, which are the results with the highest probability to be viewed, are different.

(iv) *Sample query execution*: Six examples run in Google demonstrating the above mentioned elements. Each query had multiple versions. For example the query about "Olympic Games in Athens" executed in the following variations: "Ολυμπιακοί αγώνες στην Αθήνα", "Ολυμπιακοί αγώνες Αθήνας", "Ολυμπιακοί αγώνες Αθηνών", "Ολυμπιακοί αγ. ώνες Αθήνα", "ΟΛΥΜΠΙΑΚΟΙ ΑΓΩΝΕΣ ΑΘΗΝΑ". All these variations express the same user information need; information about the Greek Olympic Games organized in Athens. Sample queries can provide feedback to the query explanation phase as well.

EVALUATION

Estimation of the success of each teaching technique relied on an evaluation task. As explained earlier each participant had to run two sets of queries, each consisting of 4 queries. Tables 1 and 2 summarize the results of this assessment.

Table 1 refers to the first set of queries. Participants were asked to return at least 6 relevant URLs for each query. They had 12 minutes in order to complete this task. As seen in table 1 for queries typed in lower case letters with intonation, all participants were able to discover relevant documents. Since the query terms referred to focused and uniquely identifiable computer science terminology Google ranked highly the truly relevant results. For the rest two queries, typed in capital letters, some participants could not successfully complete the task. This is because some of the first ranked results were not relevant, so students had to devise ways of refining their query in order to discover relevant pages. Some of them were unable to formulate and execute alternative queries.

Table 1: Evaluation data of first query set

	Completion of lower case queries with accents	Completion of upper case queries without accents	Mean task completion time (minutes)
SG1	100%	70%	11.4
SG2	100%	100%	8
AG1	100%	52.63%	11.9
AG2	100%	96.43%	8.5

SG1, SG2: Student Groups - AG1, AG2: Adult Groups

Mean task completion time is another important issue. It is clear that the modified teaching approach helped our survey subjects to complete the assignment in less time. The second student group accomplished the task in approximately 30% less time and the respective adult group showed a 28.6% decrease in time.

Table 2. Evaluation data of second query set

	Full task completion	Completion of at least one query	Mean task completion time (minutes)
SG1	15%	30%	23.6
SG2	94.74%	100%	19.5
AG1	7.69%	19.23%	23.9
AG2	85.71%	100%	21

Our students were asked to complete the second task in at most 24 minutes and to return at least 3 relevant URLs. Table 2 shows that a high percentage of the second participating group accomplished the task in approximately 20 minutes. All of them were able to discover 3 relevant Web locations for at least one query. On the other hand, the vast majority of adolescent and adult students who had no idea about stopwords, suffix removal and capitalization were unable to fully complete their assignment within the given time. A small number of them were able to discover relevant www sites only for one query and most of them about none. Another problem is that the mean task completion time is obviously increased. Nearly every one of the first groups could not discover relevant information so the average time spent approximates the maximum allocated time, since almost everyone spent all the time allotted to them.

In both cases high school students performed better than the adult participants, as they were more competent computer users and could type queries and assimilate and evaluate the retrieved web pages faster.

SYNOPSIS

In this short paper we propose a methodological teaching approach for information searching on the WWW applicable both to English and to other spoken languages with inclinations and intonation. This 6-phase teaching approach (Search engine utility explanation, Search engine access, Query explanation, Sample query execution, Navigation of result set, Evaluation task) allowed our student population to successfully accomplish their assignments. The average time required was less than the maximum allocated time and the success of the evaluation task was higher in the groups taught with the expanded version of the methodology, i.e. in the version where the internal characteristics of search engines were explained and demonstrated through examples.

As a general conclusion it can be argued that Greek users need to be more creative and knowledgeable than English searchers when utilizing search engines. This conclusion may apply to other spoken languages with similar characteristics. Therefore international search engines need to become adaptable to be truly multilingual and until that happens teachers should redefine their teaching methodology to take into consideration some of the basic internal search engines characteristics.

BIBLIOGRAPHY

- Baeza-Yates R., & Ribeiro-Neto B. (1999). *Modern Information retrieval*. New York: Addison Wesley, ACM Press.
- Brandt, S.D. (1997). Constructivism: Teaching for understanding on the Internet, *Communications of the ACM*, 40, 112-116.
- Eagleton, M.B., & Guinee K. (2002). Strategies for supporting student Internet inquiry. *New England Reading Association Journal*, 38(2), 39-47.
- Henry, L. (2005). Information search strategies on the internet: a critical component of new literacies. *Webology*, 2(1), Retrieved May 1, 2006 from <http://www.webology.ir/2005/v2n1/a9.html>
- Large, A., & Beheshti, J. (2000). The Web as a classroom resource: Reactions from the Users. *Journal of the American Society for Information Science*, 51(12), 1069-1080.
- Lazarinis, F. (2005). Evaluating user effort in Greek web searching. *10th PanHellenic Conference in Informatics* (pp. 99-109). Volos, Greece: University of Thessaly.
- Lazarinis, F., Kougiourouki, O. (2003). Discovering environmental pollution forms in the Web. *8th PanHellenic Conference: Perspectives, Progress and Teaching of Natural Sciences* (in Greek)
- Lazonder, A. W. (2001). Minimalist instruction for learning to search the world wide web. *Education and Information Technologies*, 6(3), 161-176.
- Leu, D. J., Kinzer, C. K., Coiro, J., & Cammack, D.W. (2004). Toward a theory of new literacies emerging from the Internet and other communication technologies. In R. Ruddell and N. Unrau (Eds.) *Theoretical models and processes of reading* (5th ed.) (pp. 1568-1611). Newark, DE: International Reading Association.
- Nachmias, R., & Gilad, A. (2002). Needle in a hyper stack: Searching for information on the World Wide Web. *Journal of Research on Technology in Education*, 34, 475-486.
- Schlein, A. (2002). *Find it Online* (3rd ed.). New York: Facts on Demand Press.