

Comparing and evaluating an interactive, subject-based search system with the traditional digital library's search functionality

Konstantinos Kyprianos¹

¹Department of Archival, Library & Information Studies, University of West Attica, Athens, Greece

kkyprian@teiath.gr

Article Info

Article history:

Received 01 March 2018

Received in revised form 30 May 2018

Accepted 15 June 2018

DOI: <http://doi.org/10.18780/jiim.v3i1.4269>

Abstract:

Purpose - The purpose of this paper is to compare and evaluate the usability, usefulness and effectiveness of an Interactive, Information Retrieval – IIR system with a DSpace-based digital library.

Design/methodology/approach – The proposed evaluation approach consists of two subcomponents. The first one refers to a log file analysis capable of revealing quantitative features of the systems' usage. The second part refers to a user survey that compares the new IIR system against the traditional subject-based search functionality provided by DSpace in terms of usefulness and effectiveness.

Findings - Based on the evaluation results, it seems that users are very interested in employing new methods and techniques in information seeking and retrieval, especially when such new tools and methods help them in fulfilling their information needs accurately and timely. The results also revealed that the users are more satisfied when employing the new search functionality and the search and retrieval process is improved.

Originality/value - A novel IIR system for subject-based browsing was evaluated and interesting results for the future of such tools are shown.

Index Terms — evaluation, questionnaire, user survey, IIR system, simulated work task scenario

I. INTRODUCTION

Nowadays, searching for information is beyond any doubt very common and it is exercised widely not only from search professionals, but also from average users [1] [2]. Yet, the effective use of the corresponding search technologies is still challenging [3][4].

The most recent Information Retrieval - IR systems provide functionalities and search capabilities that a few years ago would be beyond any imagination. However, in many cases they do not support searchers in finding the right tactic in order to satisfy their information needs with accuracy and in a short time [5]. Quite often, average users try to express their information needs as a search query that

may contain several meanings, thus failing in accurately specifying their requests and accordingly fulfilling their information needs [6]. In case of an unsuccessful search, average users reformulate their query by adding, deleting or replacing terms [7]. This could be interpreted that searchers may spend too much time in finding the right terms that will satisfy their information needs [8].

Average users rarely employ sophisticated search strategies, as compared to expert searchers [9]. Furthermore, they usually do not know how and when to use advanced search features in order to achieve the best result [10]. In this direction, IR systems evolved over time towards the direction of aiding their users in satisfying their information needs with accuracy in a short time. Thus, they do not only invest on providing advanced functionality, but they also strive in assisting and guiding users in finding the information they need through some kind of interactive process. Thus, it seems that old-fashioned IR systems are stepping back in favor of Interactive Information Retrieval – IIR systems.

In order to examine the effectiveness and usefulness of such systems, a number of evaluation measures and standards have been established. These evaluation measures and standards take into account not only the IIR system itself, but also the users' interactive process of information searching [11][12]. Generally, the more effective a system is, the less time a user needs in order to satisfy their information needs. The time a user spends using such a system includes the time that is spent learning it. Thus, it is important for an IIR system to help average users improving their searching capabilities over time [13].

Along these lines, Borlund [14] argues that “*the purpose of IR evaluation is twofold, i.e. to determine a) how well the system satisfies the information needs of actual and potential users; and b) how to improve the information retrieval process, both at a particular installation level and at a more general level*”.

Having the above thoughts in mind, a subject-based IIR system was created and accordingly integrated to the DSpace-based, digital library of University of Piraeus. The system gave the opportunity to its users to find the assets they were looking for by browsing through the subject headings of the underlying collection based on the syndetic structure of the subject headings (i.e. broader, narrower and

related terms).

In order to assess whether the aforementioned IIR system was useful, effective and satisfied the users' information needs, an evaluation procedure was performed. The evaluation procedure consisted of two parts. The first part contained the log files analysis of the system's use in a period of 6 months. The second part referred to a user survey that was based on a comparison of the traditional subject-based search functionality of DSpace against the new IIR system.

The rest of the paper is structured as follows: The next section contains a detailed description of the IIR system. Then, the measures and standards that should be kept in mind in order to create an effective and accurate evaluation procedure for an IR system are stated. The next section outlines a number of IR system evaluation frameworks. In the following sub-section, some individual efforts regarding the evaluation of IIR systems are presented. Section 4 describes the approach that is followed in order to evaluate the new IIR system. Finally, in section 5 the conclusions of the evaluation process that was followed and suggestions for future work are stated.

II. THE SYSTEM

In this section the IIR system under evaluation is presented. More specifically, the system enriches the subject-based search functionality of the DSpace-based Digital Library of University of Piraeus, Dione [15], which contains subject headings deriving from the Library of Congress Subject Headings – LCSH¹ vocabulary. It is comprised of an auto-suggest search box on the upper part of the screen where the users are prompted to type in the first letters of the words that best describe their information needs (see figure 1). The widget returns a list of subject headings that contain the string provided by the user.



Figure 1. Autosuggest search box

Upon selection of a subject heading, a box is sketched below representing their selection (see figure 2). The box contains the subject heading in English, possibly followed by its translation in Greek. The box also contains the relations of the subject heading as imposed by the underlying extended syndetic structure (i.e. namely broader, narrower, related and subdivision term).

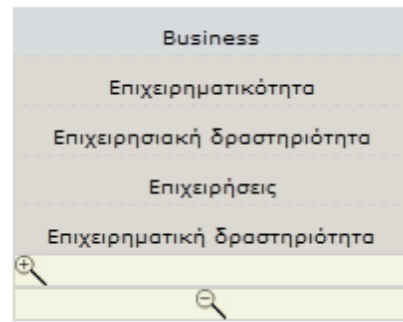


Figure 2. Box with the selected subject heading "Business"

If the user clicks on a relation, a list of subject headings appears, each of which is associated with the sketched one through the selected relation (see figure 3).



Figure 3. Broader terms for the selected subject heading "Business"

Then, by clicking on one of the subject headings presented in the list, another box is sketched next to the first one (see figure 4). The two boxes are connected with a labeled line containing the description of the selected relation.

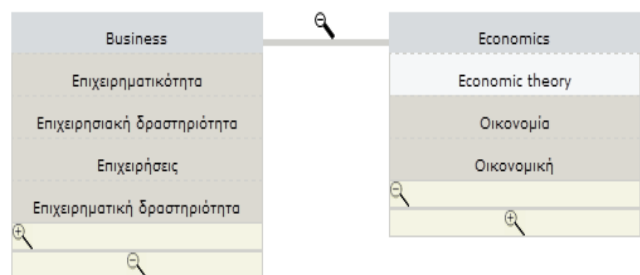


Figure 4. Depiction of the broader term relation for the two subject headings "Business" and "Economics"

On the lower part of the screen, the user is presented a list containing the assets of the digital library that are assigned to the selected subject heading (see figure 5). The whole process can be repeated until the user locates a subject heading that satisfies his information needs.

¹ Library of Congress Subject Headings, available at: <http://authorities.loc.gov> Date retrieved: 15/5/2018

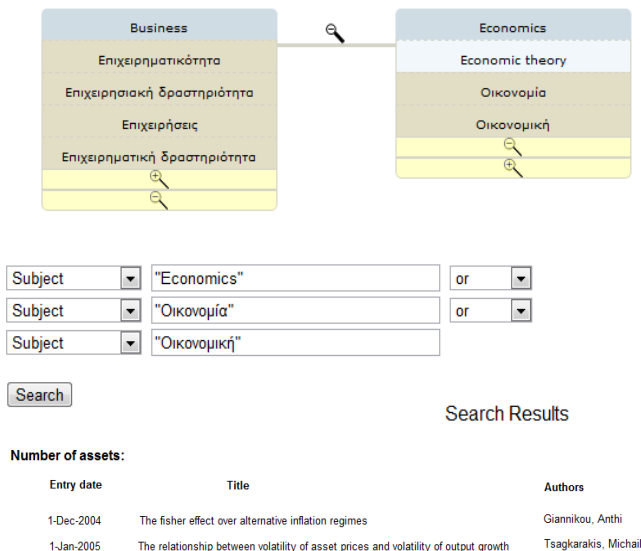


Figure 5. Screenshot from the IIR system depicting the whole functionality

A detailed description of the system is provided in Papadakis et al. [15].

In order to evaluate this IIR system, the corresponding evaluation literature is examined. In this context, the following section presents some important measures for an effective usability evaluation.

III. IMPORTANT MEASURES FOR AN EFFECTIVE USABILITY EVALUATION

Effective and accurate evaluation of IIR systems usability should be based on the examination of a number of adequate measures. IIR systems are usually evaluated in terms of three main aspects of usability: effectiveness, efficiency and user satisfaction [16]. These measures are defined by ISO 9241-11², as:

- **Effectiveness** is the “accuracy and completeness with which users achieve specified goals”. In other words, a tool is effective if it helps users accomplish particular tasks.
- **Efficiency** is the “resources expended in relation to the accuracy and completeness with which users achieve goals.” A tool is efficient if it helps users complete their tasks with minimum waste, expense or effort.
- **Satisfaction** is the “freedom from discomfort, and positive attitudes of the user to the product”. Satisfaction can be understood as the fulfillment of a specified desire or goal. It is often the case that when people discuss satisfaction they speak of the contentment or gratification that users experience when they accomplish particular goals.

Based on the aforementioned ISO standard for usability, a number of researchers have proposed various evaluation measures and processes that should be considered for the evaluation of an IIR system.

The most common evaluation measures seem to be “Precision”, “Recall” and the deriving “F-measure”. These measures have been employed for quite a long time [11]. Precision is the fraction of retrieved documents that are relevant to the query. Recall is the fraction of documents that are known to be relevant to the query and have been successfully retrieved. In order to balance between these two measures, the F-measure was introduced. According to traditional IR algorithms, the more documents an algorithm retrieves, the more likely is to increase recall. But on the other hand, this augmentation to the search results could bring more irrelevant documents. In order to address this issue, F-measure is defined as the combination of precision and recall as shown in the equation below.

$$F = 2 * \frac{precision * recall}{precision + recall} \quad (1)$$

Some researchers believe that the above measures are not the best option in order to make an accurate evaluation of an IIR system. Borlund [17], for example, endorses the idea that precision and recall are insufficient for evaluating IIR systems. The above two measures cannot quantify the “informativeness” of interaction which is exhibited in the case of users wishing to modify or develop their initial queries and strategies during a search process.

In order to come up with more suitable measures for the evaluation of IIR systems, other researchers propose alternative solutions. Some of these measures are stated below:

According to Su [18], users consider “Task Completion Time” as critical to successful IIR. In the same line of thoughts, Dunlop [19] proposes a measure called “Expected Search Duration” and creates an interface-based predicted-time model, which measures the time that a user needs in order to view a set of assets and concludes to a relevant asset.

Belkin, Cole and Liu [20] and Hienert and Mutschke [21] propose another measure for evaluation of IIR systems, namely “Usefulness”. Usefulness can be used to evaluate system support from the aspects of both outcome and process in the accomplishment of a task.

In another approach, Cheng, Hu and Heidom [13] suggest two new measures to evaluate IIR systems, the “Normalized Task Completion Time” and the “Normalized User Effectiveness”. These two measures take into account the familiarity of users with the use of such systems, the capability of the user to retrieve information with the use of IIR systems and the expertise in the domain of the given task and thus the ability to create good queries.

Lastly, Borlund and Ingwersen [17] introduce the concept of “Simulated Work Task Situation” or “Scenario” and the involvement of real end users as test persons. Their method is designed to collect two types of data, the cognitive data

² Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability (1998). Available at:

http://www.iso.org/iso/catalogue_detail.htm?csnumber=16883 Date retrieved: 15/5/2018

and the system-oriented data. The former refers to the collection of qualitative and quantitative data from the user's experience with the system. The latter refers to the collection of IR performance data. The whole process requires the involvement of real users who are performing searches of their own and for simulated tasks.

The next section contains several frameworks that are widely used in order to evaluate IR and IIR systems.

A. Previous IR system evaluation frameworks

Evaluation of IR systems has intrigued researchers for many years, since evaluation is considered an integral part of system development. In order to perform such evaluations, a number of well-known test collections are employed.

During the 60's, the Cranfield model was introduced. The indexing experiments of the Cranfield model are often considered as the beginning of the modern era of computer-based IR system evaluation [22]. In the Cranfield studies, retrieval experiments were conducted on a variety of test databases. In the second series of experiments, known as Cranfield II, alternative indexing languages constituted the performance variable under investigation. The aim of the research was to find ways to improve the relative retrieval effectiveness of IR systems through better indexing languages and methods [23]. A small test collection of documents, a set of test queries, and a set of relevance judgments (i.e. a set of documents judged to be relevant to each query) were the components of the Cranfield experiments. End users, their interaction with the system, their interpretations of the query were not calculated and taken into account in the experiments [24]. The measures used in the Cranfield experiments were recall and precision. Nowadays, the Cranfield model is still in use for the most elementary pilot experiments [25].

The Cranfield model inspired in some way the Text Retrieval Conference – TREC³. Within this model, there have been many tracks over a wide range of different test collections. Nevertheless, the marquee task of TREC is the ad-hoc retrieval track, in which systems compete in ranking documents according to relevance judgments [26]. Participants over the years have examined a wide variety of retrieval techniques and retrieval environments, including cross-language retrieval, retrieval of web documents, multimedia retrieval, and question answering. Recently, the interactive TREC – iTREC was introduced in order to develop better methodologies for evaluation of IIR systems [14].

During the last two decades, the Cross-Language Evaluation Forum – CLEF⁴ emerged, aiming in developing an infrastructure for the testing, tuning and evaluation of information retrieval systems operating on European languages in both monolingual and cross-language contexts.

As far as the evaluation of digital libraries is concerned, one major evaluation model is the Distributed Agents for

User-Friendly Access of Digital Libraries - DAFFODIL model. DAFFODIL is a system for integrated search within the heterogeneous digital libraries of a scientific community, with merging of results. At this time, a prototype for the area of Computer Science exists that allows searching within more than ten different digital libraries and other sources of information [27]. The DAFFODIL framework consists of two major parts: the graphical user client and a set of agent-based services in the back-end [28]. The DAFFODIL framework also provides an integrated questionnaire tool and a logging facility to help gathering the data.

In a more recent approach, Wei, Zhang and Gwizdka [29] proposed YASFIIRE as a system that is capable of supporting IIR user studies on the Web. The system supports user and task management, for processing web-based task specific interfaces and for web-event logging.

To sum up, all these efforts have a specific aim; to measure the effectiveness of an existing IR or IIR system via test collections.

In the next section, individual evaluation efforts are presented and the most suitable is selected for the evaluation of the IIR system that is described earlier in this paper.

B. Individual evaluation efforts

In order to evaluate the specific IIR system, the aforementioned frameworks are considered. However, none of them can be applied as-is for a number of reasons.

The Cranfield framework cannot be used because of its inflexibility to deal with dynamic information needs. Cranfield treats information needs as a static concept entirely reflected by the search statement (query) [25].

The next options are the iTREC and CLEF evaluation frameworks. Both approaches adopt a methodology that provides a set of predefined queries for which the corresponding relevant results are known in advance. When an IR system addresses such queries to the underlying dataset, precision and recall are measured and accordingly compared against the pre-calculated metrics. However, in an IIR system like the one that is under evaluation in this paper, users are prompted to choose a predefined query that best suits their information needs, which acts as an entry point to the IIR system. Trained personnel have already determined the relevant assets that correspond to such a query. Thus, the quality of the proposed approach does not depend on precision/recall [30] but on its ability to provide the users with an entry point that is as closer as possible to their information needs.

The last option is the adoption of the DAFFODIL framework, which requires the integration of the system under evaluation with the DAFFODIL User Interface. However, such an approach would result in the assessment of the integrated system, not the original one.

Thus, in order to evaluate the IIR system that is described

³ Text Retrieval Conference. Available at: <https://trec.nist.gov> Date retrieved 15/5/2018

⁴ Cross Language Evaluation Forum. Available at: <http://www.clef-campaign.org/> Date retrieved: 15/5/2018

earlier in this paper, some hybrid procedures and efforts are considered. The first effort is the evaluation procedure of the Concept-based Information Retrieval Interface – CIRI [31]. The authors use the “Simulated Work Task Situation” to make their searching situations realistic. Their system contains an ontology-based query interface, which is constructed for searching a digital newspaper archive. In order to assess their system, they use a similar search interface without ontology support.

The second effort that is taken under consideration is the evaluation procedure that is followed by Suomela and Kekalainen [31]. They aim to evaluate their system through searches based on three different task types and accordingly study how college users interact with highly structured data. This experiment is applied to the Initiative for the Evaluation of XML - INEX interactive track (iTrack⁵). The overall goal of INEX [32] is to experiment with the potential of using XML to retrieve relevant parts of documents. During the evaluation process, the searchers are given brief online questionnaires in order to support the analysis of log data.

Lastly, Kriewel and Fuhr [33] evaluate an adaptive search suggestion system that is based on case-based reasoning techniques. They develop a suggestion tool for the DAFFODIL system to support users with useful strategic search advice. The aim of the evaluation is to learn whether an adaptive search suggestion system could help users in searching and whether it could teach users how to use the advanced capabilities of a complex search system. The corresponding evaluation approach dictates the assignment of a number of complex search tasks to the participants that should be carried out separately in two systems. Both of the systems are identical and based on DAFFODIL. The only difference is that one of them does not include the suggestion tool. The DAFFODIL logging framework is then used to capture all users’ activities during the task.

IV. OUR APPROACH

Having the aforementioned approaches in mind, we concluded that the most appropriate method for the evaluation of the IIR system would be a comparison against the traditional, subject-based browsing functionality of the DSpace’s digital library of the University of Piraeus in Greece, Dione⁶. More specifically, the IIR system provides another option to the library’s users that wish to perform subject-based search within the underlying resources.

Following the steps of Borlund [25][34], the comparative evaluation process consists of a quantitative log file analysis regarding a period of 6 months and a user survey employing an adequately designed questionnaire. The former is anticipated to measure the overall performance of the IIR system and the later aims in estimating its impact to the end users.

A. Quantitative evaluation – The log files analysis

The log files that were analyzed during the quantitative evaluation process refer to the IIR system’s usage from 20.10.2016 to 20.3.2017. The specific period of time was selected because it refers to a fully functional semester at the University of Piraeus. Log files in general are a valuable resource for understanding the kinds of information needs that users have, for improving ranking scores, for showing search history, and for attempts to personalize IR [26]. Within the context of the specific log files analysis, the term “*search session*” refers to the sequence of requests made by one user for a single navigation purpose [35].

According to the log files of both the DSpace and the IIR system, 54,7% of the total number of users that selected the subject-based browsing option, preferred to employ the IIR system instead of the typical functionality of DSpace. Such a measure does not imply any particular prevalence of the one option over the other.

Moreover, the log files reveal that the vast majority of the users did not spend much time with the IIR system (see table 1). In fact, 164 out of 230 users spent less than a minute. This could be attributed to the fact that the provided GUI consists of two widgets (namely the autosuggest search box and the boxes traversal), with different average usage time. More specifically, the autosuggest search box provides rapid suggestions in accordance with the typing speed of the user.

Table 1. IIR system’s usage

| no. of user sessions | no. of minutes per session |
|----------------------|----------------------------|
| 164 | Less than 1 |
| 34 | 1-2 |
| 17 | 3-5 |
| 9 | 6-10 |
| 1 | 11-20 |
| 3 | 21-40 |
| 2 | 41-59 |

On the contrary, the boxes traversal requires mental effort from the users since they have to select the most appropriate choice from a number of semantically relevant options [36]. From such a viewpoint, table 1 could be interpreted as that users having a specific subject in mind employed the autosuggest search box whereas users with vague information needs employed the boxes traversal.

The above scenario is further justified from the findings stated at table 2. Table 2 indicates that the autosuggest search box was employed 171 times, whereas the boxes traversal was employed 55 times.

⁵ INEX 2010 Interactive Track (iTrack). Available at: <http://www.inex.otago.ac.nz/tracks/interactive/interactive.asp> Date retrieved: 15/5/2018

⁶ Dione. Available at: <http://dione.lib.unipi.gr> Date accessed: 15/5/2018

Table 2. IIR's widgets usage

| widget | usages | % |
|------------------------|--------|------|
| Autosuggest search box | 171 | 75,6 |
| boxes traversal | 55 | 24,4 |

From the 55 box traversals (see figure 6), 15 times users concluded their subject-based browsing session in their third hop. This is the most popular number of hops from one box to another according to the semantic relations of the subject headings that each box corresponds to (see figure 6). Following that, in a descending order, users concluded their subject-based browsing session in one, two and four hops respectively. The fact that 33 out of 55 box traversals were concluded in one, two or three hops leads to the conclusion that the majority of the users conclude their subject browsing process in a rather short period of time.

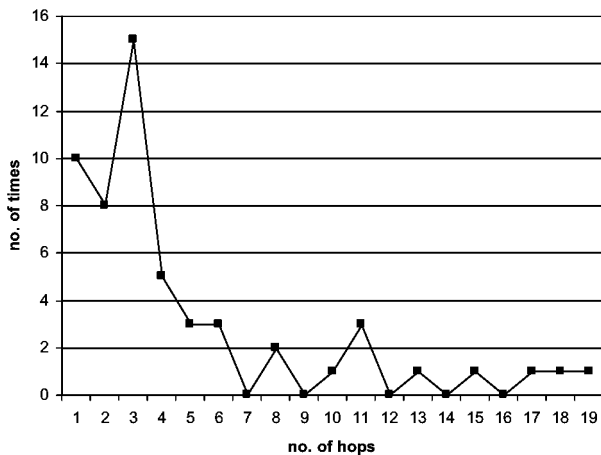


Figure 6. Number of hops in a boxes traversal

According to the log files (see table 3), the most popular semantic relation between subject headings is “contains” followed by a subdivision.

Table 3. Semantic relations

| Semantic relation | No. of times |
|-------------------|--------------|
| “contains” | 178 |
| “is part of” | 14 |
| “in context of” | 1 |
| checkbox | 13 |
| subdivision | 27 |

At this point, it should be noted that “checkbox” is not a semantic relation per se. It refers to the user action of clicking on a box in order to erase its subsequent boxes and thus make the GUI more readable. Moreover, “subdivision” refers to the common subdivisions of two subject headings, defining this way the extended syndetic structure [15] of the underlying subject headings’ collection. The relations “contains”, “is part of” and “in context of” referred to the

semantic relations such as narrow term, broader term and related term.

The predominance of the semantic relation “contains” could be justified from the fact that users tend to think from broader concepts to narrower ones when trying to satisfy their information needs. Additionally, the fact that the second most popular semantic relation refers to subdivisions, underpins the importance of the extended syndetic structure during a subject browsing process.

As far as users’ satisfaction is concerned, safe conclusions cannot be extracted from the log files analysis. Some users may have started using any of the two systems and concluded their sessions without shifting to the other one. From another point of view, some users may have started using any of the two systems and, due to their dissatisfaction, concluded their sessions by using the other one. The satisfiability of the users employing the IIR system is measured through the circulation of a suitable questionnaire [37] that will be described later in this paper.

Thus, the qualitative part of the evaluation refers to a user survey where real users performed specific search tasks and then they filled-in a pre- and a post- questionnaire that were based on the principles of Kelly [16]. According to the proposed approach, the questionnaires consist of a number of questions where a specific set of accordingly weighted predefined responses is provided for each one of them.

B. Qualitative evaluation – Search scenarios, Questionnaires

The qualitative evaluation of the IIR system was implemented through a user survey. More specifically, the participants were initially asked to perform two subject search task scenarios and fill-in adequately designed pre- and post- questionnaires [34][38].

1) The procedure

The purpose of the survey was to measure the impact of the new subject-based IIR system that was recently integrated to the DSpace digital library of the University of Piraeus. In order to achieve this, a comparison between the traditional subject-based search functionality of the digital library and the new IIR system was performed.

Initially, the participants were gathered in a computer lab and they were accordingly informed about the purpose of the survey. Then, they were asked to fill-in a pre-questionnaire in order to record their search profiles. More specifically, their educational level, their familiarity with web search engines and DSpace’s subject-based searching functionality were logged. After that, they were asked to perform two subject-based search task scenarios. Both of the scenarios should be implemented with both of the systems under comparison. The participants were asked to start with the traditional DSpace subject-based searching functionality and then repeat the same scenario with the new IIR subject-based search system. The whole process was supervised by the evaluators in order to ensure that there would be no technical difficulties in conducting their tasks.

After they had completed both scenarios, they were asked to fill-in an accordingly designed post-questionnaire in order to express their opinion about the two systems under comparison.

2) Participants

A total number of 16 users participated in the survey. They were all members of the academic society of the University of Piraeus. The participants were all familiar with the University's library. However, as indicated from the questionnaire's statistics, some of them had never used the DSpace digital library before.

3) Search task scenarios

Each participant was asked to perform two subject-based search task scenarios. The first scenario was simple in order to give to the novice participants the opportunity to get acquainted with the two systems. The second one was more complicated due to the fact that the described information need was more general and required a more detailed search. Finally, it should be noted that both search task scenarios correspond to existing subject headings within the digital library.

The participants were asked to perform the first task to the traditional subject-based search functionality and then to the new IIR system. After completing the first task, they were asked to perform the second task following the same sequence.

The two search tasks were the following:

- a) You are looking for information about "Stress management". Try to satisfy your information needs using the traditional subject-based search provided by the DSpace digital library. Then, try to satisfy the same information needs by employing the new IIR service.
- b) You are looking for information about "Computer network protocols". Try to satisfy your information needs using the traditional subject-based search provided by the DSpace digital library. Then, try to satisfy the same information needs by employing the new IIR service.

4) Questionnaire

Upon completion of the two search tasks, the participants were prompted to fill in a post-task questionnaire. In this questionnaire the users were asked to answer questions referring to the usability, the satisfiability and the effectiveness of the new IIR system as compared to the traditional subject-based search provided by the digital library.

5) Results

The whole process lasted between 10 to 20 minutes for each participant. A completed search process could either end up with successful or unsuccessful search results.

As indicated in figure 7, most of the participants were students in an under- or post-graduate level.

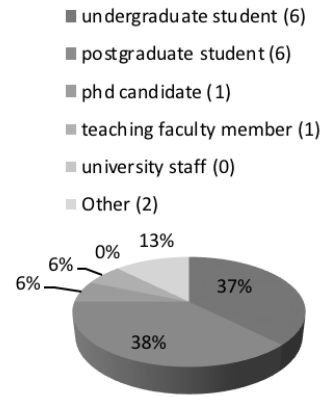


Figure 7. Educational level of the participants

Moreover, they had all, more or less sufficient experience in searching for information online. However, it seems that the digital library of the University's Library is not popular among the participants of this survey. 94% of the participants (see table 4) had never or very seldom used the digital library in the past. This could be attributed to the fact that the digital library was quite recently incorporated within the overall Library's infrastructure.

Table 4. Frequency of DSpace digital library's visits

| Frequency of visits | participants | % |
|--------------------------------|--------------|--------|
| one time each semester or less | 5 | 31.25% |
| one to three times a month | 0 | 0.00% |
| once a week | 1 | 6.25% |
| more often than once a week | 0 | 0.00% |
| never | 10 | 62.50% |

When combining the answers of question 3 with the answers of question 4, it becomes apparent that participants that have used the traditional DSpace functionality in the past were more reluctant in finding the new IIR system easy to use, as compared to users that employed both of the systems for the first time. More specifically, 2 out of 6 users that have used DSpace before found the traditional DSpace's functionality easier to use than the new one (see table 5), in contrast to 10 out of 10 users without any prior experience with DSpace that found the new system easier to use.

Table 5. Comparison of the two systems based on ease of use for users with prior experience with DSpace

| System | participants | % |
|--------------------------------|--------------|--------|
| DSpace subject-browsing system | 2 | 33.33% |
| New IIR system | 4 | 66.67% |

The same conclusions apply when combining the answers of question 3 with the answers of question 5. Participants with prior experience with DSpace found the two systems

equally easy to understand (see table 6). Again, 10 out of 10 users without any prior experience with DSpace found the new system easier to understand than the traditional one.

Table 6. Comparison of the two systems based on understandability for users with prior experience with DSpace

| System | participants | % |
|--------------------------------|--------------|--------|
| DSpace subject-browsing system | 3 | 50.00% |
| New IIR system | 3 | 50.00% |

When combining the answers of question 6 with the answers of question 8 it becomes apparent that users that succeeded in satisfying their information needs with both of the systems found that the new IIR system aided them in completing their search tasks faster than the traditional one. More specifically, 7 out of 10 participants satisfied their information needs faster with the new IIR system.

Answers to question 9 indicate that the new metaphors in subject-based browsing introduced by the new IIR system (i.e. box traversal) helped DSpace-non-experienced participants in satisfying their information needs while at the same time such new metaphors did not discourage the DSpace-experienced participants in using the system (see figure 8). More specifically, 15 out of 16 participants found that browsing through the subjects by employing the new IIR system helped them in satisfying their information needs.

The browsing through subjects functionality was usefull

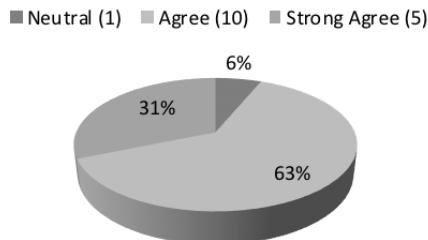


Figure 8. The participants' opinion about the browsing through subjects' functionality provided by the new IIR system

Finally, answers to question 10 indicate that the vast majority of the participants enjoyed using the new IIR system in order to satisfy their information needs (see table 7). In fact, 14 out of 16 participants would prefer to browse the digital library of the University of Piraeus by subject through the employment of the new IIR system instead of the traditional DSpace functionality.

Table 7. Comparison of the two systems based on the users' intention of reusing one system or another

| System | participants | % |
|--------------------------------|--------------|--------|
| DSpace subject-browsing system | 2 | 12.50% |
| New IIR system | 14 | 87.50% |

The remaining 2 participants are DSpace-experienced users that seem to be reluctant in changing their information seeking habits.

During the questionnaire procedure many participants asked questions concerning the usage of the new IIR system. stating that they will use it again in the future as indicated from the answers to question 10.

V. CONCLUSIONS

In this paper a novel IIR system for subject-based browsing was evaluated. The system is integrated to the DSpace-based digital library of the University of Piraeus in Greece.

The evaluation process is twofold: quantitative and qualitative. The former refers to the log file analysis of the IIR system's usage for a period of 6 months. The latter refers to the results of a user survey that compares the traditional subject-based search of DSpace against the new IIR system. More specifically, 16 participants were asked to perform two subject-based search task scenarios employing both systems. Then, they were asked to fill-in an adequately designed questionnaire in order to record their impressions concerning the two systems.

The log file analysis reveals that end users usually do not spend too much time searching for information by subject. Moreover, they seem to prefer the new functionality that is provided by the IIR system, despite the fact that it introduces new metaphors as far as the user interface is concerned.

Finally, the extended functionality of the new IIR system as compared against the traditional DSpace's functionality seems to outweigh the fact that the new IIR system has greater learning curve due to the new metaphors it introduces. Thus, users are very interested in employing new methods and techniques in information seeking and retrieval, especially when such new tools and methods help them in fulfilling their information needs accurately and timely.

Future work could be targeted towards the alteration and modification of the service under evaluation so as to provide personalized results based on the user's preferences and/or to give the opportunity to the users to provide tags in order to enrich the subject headings. Such an enrichment could facilitate and improve search and retrieval process.

REFERENCES

- [1] Liu, J., Zamir, H., Li, Y. and Hastings, S.K. (2018). Search systems and their features: What college students use to find and save information. *Library & Information Science Research*, 40 (2), pp. 118-124.
- [2] Kelly, D. and Sugimoto, C.R. (2013). A Systematic Review of Interactive Information Retrieval Evaluation Studies, 1967-2006. *Journal of the American Society for Information Science and Technology*, 64 (4), pp. 745-770. DOI: 10.1002/asi.22799.
- [3] Drabentstott, K. M. (2003). Do non-domain experts enlist the strategies of domain experts. *Journal of the*

- American Society for Information Science and Technology*, 54(9), pp. 836–854.
- [4] **Wildemuth, B.M. (2004).** The effects of domain knowledge on search tactic formulation. *Journal of the American Society for Information Science and Technology*, 55(3), pp. 246–258.
- [5] **Brajnik, G., Mizzaro, S., Tasso, C. and Venuti, F. (2002).** Strategic help in user interfaces for information retrieval. *Journal of the American Society for Information Science and Technology*, 53(5), pp. 343–358.
- [6] **Pollock, A. and Hockley, A. (1997).** What's wrong with internet searching. *D-Lib Magazine* <Available at: <http://www.dlib.org/dlib/march97/bt/03pollock.html>>
- [7] **Rieh, S.Y. and Xie, H.I. (2001).** Patterns and sequences of multiple query reformulations in web searching: a preliminary study. In: *Proceedings of the 64th Annual Meeting of the American Society for Information Science and Technology*, vol. 38, pp. 246–255.
- [8] **Fields B., Keith S., Blandford A. (2005).** Designing for Expert Information Finding Strategies. In: Fincher S., Markopoulos P., Moore D., Ruddle R. (eds) *People and Computers XVIII — Design for Life*. Springer, London, pp. 89–102
- [9] **Carstens, C., Rittberger, M. and Wissel, V. (2009).** How users search in the German education index - tactics and strategies. In: *Proceedings of the workshop Information Retrieval at the LWA*.
- [10] **Markey, K. (2007).** Twenty-five years of end-user searching, part 1: Research findings. *Journal of the American Society for Information Science and Technology*, 58(8), pp. 1071–1081.
- [11] **Frakes, W. and Baeza-Yates, R. (eds) (1992).** *Information Retrieval, data structures and algorithms*. Prentice Hall, New York: Englewood Cliffs.
- [12] **Kelly, D., Erguello, J., Edwards, A. and Wu, W.C. (2015).** Development and Evaluation of Search Tasks for IIR Experiments using a Cognitive Complexity Framework. In: *Proceedings of the 2015 International Conference on The Theory of Information Retrieval*, Northampton, Massachusetts, USA, September 27 - 30, pp. 101-110.
- [13] **Cheng, J., Hu, X. and Heidorn, P.B. (2010).** New Measures for the Evaluation of Interactive Information Retrieval Systems: Normalized Task Completion Time and Normalized User Effectiveness. *ASIST*, October 22-27, Pittsburgh, PA, USA.
- [14] **Borlund, P. (2000).** Experimental components for the evaluation of interactive information retrieval systems. *Journal of Documentation*, 56(1), pp. 71-90.
- [15] **Papadakis, I., Kyprianos, K., Mavropodi, R. and Stefanidakis, M. (2009).** Subject-based Information Retrieval within Digital Libraries Employing LCSHs. *D-Lib Magazine*, 15(9/10) <Available at: <http://www.dlib.org/dlib/september09/papadakis/09papadakis.html>>
- [16] **Kelly, D. (2009).** Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends in Information Retrieval*, 3(1-2), pp. 1-224.
- [17] **Borlund, P. and Ingwersen, P. (1997).** The development of a method for the evaluation of interactive information retrieval systems. *Journal of Documentation*, 53(3), pp. 225-250.
- [18] **Su, L.T. (1992).** Evaluation measure for interactive information retrieval. *Information Processing and Management*, 28(4), pp. 503-516.
- [19] **Dunlop, M. (1997).** Time, relevance and interaction modeling for information retrieval. In: *Proceedings of the 20th ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.206-213.
- [20] **Belkin, N., Cole, M. and Liu, J. (2009).** A model for evaluation of interactive information retrieval. In: *Proceedings of the SIGIR 2009 Workshop on the future of IR Evaluation*, pp. 7-8.
- [21] **Hienert, D. and Mutschke, P. (2016).** A Usefulness-based Approach for Measuring the Local and Global Effect of IIR Services. In: *Proceedings of the 2016 ACM on Conference on Human Information Interaction Retrieval*, Carrboro, North Carolina, USA, March 13-17, 2016, pp. 153-162.
- [22] **Cleverdon, C.W., Mills, J. and Keen, E. M. (1996).** *Factors determining the performance of indexing systems*. Cranfield. UK: Aslib Cranfield Research Project, College of Aeronautics.
- [23] **Cleverdon, C.W. (2000).** The effect of variations in relevance assessments in comparative experimental tests of index languages. Cranfield, UK: Cranfield Institute of Technology.
- [24] **Hildreth, C.R. (2001).** Accounting for users' inflated assessments of on-line catalogue search performance and usefulness: an experimental study. *Information Research*, 6(2) <Available at: http://InformationR.net/ir/6-2/paper_101.html>
- [25] **Borlund, P. (2003).** The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. *Information Research*, 8(3).
- [26] **Hearst, M. (2009).** *Search User Interfaces*. Cambridge: University Press.
- [27] **Fuhr, N., Klas, C.P., Schaefer, A. and Mutschke, P. (2002).** Daffodil: An integrated desktop for supporting high-level search activities in federated digital libraries. In: *Research and Advanced Technology for Digital Libraries. 6th European Conference on Digital Libraries*, pp. 597-612.
- [28] **Klas, C.P., Kriewel, S. and Fuhr, N. (2007).** An experimental framework for interactive information retrieval and digital libraries evaluation. In: *Proceedings of the 1st international conference on Digital libraries: research and development (DELOS'07)*. Springer-Verlag, Berlin, Heidelberg, pp. 147-156.
- [29] **Wei, X., Zhang, Y. and Gwizdka, J. (2014).** YASFIIRE: yet another system for IIR evaluation. In: *Proceedings of the 5th Information Interaction in Context Symposium*, Regensburg Germany, pp. 316-319.
- [30] **Agirre, E., Di Nunzio, G.M., Mandl, T. and Otegi, A. (2009).** CLEF 2009 ad hoc track overview: robust-WSD task. In: *Proceedings of the 10th cross-language evaluation forum conference on Multilingual*

information access evaluation: text retrieval experiments (CLEF'09). Springer-Verlag, Berlin, Heidelberg, pp. 36-49.

- [31] **Suomela, S. and Kekalainen, J. (2006).** User evaluation of ontology as query construction tool. *Information Retrieval*, 9(4), pp. 455-475 DOI: 10.1007/s10791-006-6387-3.
- [32] **Pharo, N., Nordlie, R., Fuhr, N., Beckers, T. and Fachry, K.N. (2009).** Overview of the INEX 2009 interactive track. In: Proceedings of the Focused retrieval and evaluation, and 8th international conference on Initiative for the evaluation of XML retrieval (INEX'09), Springer-Verlag, Berlin, Heidelberg, pp. 303-311
- [33] **Kriewel, S. and Fuhr, N. (2010).** Evaluation of an adaptive search suggestion system. In: *Proceedings of the 32nd European Conference on Information Retrieval Research (ECIR 2010)*, Springer-Verlag, pp. 544-555.
- [34] **Borlund, P. (2016).** A study of the use of simulated work task situations in interactive information retrieval evaluations: A meta-evaluation, *Journal of Documentation*, 72 (3), pp. 394-413.
- [35] **Huang, X., Peng, F., An, A. and Schuurmans, D. (2004).** Dynamic Web Log Session Identification With Statistical Language Models. *Journal of The American Society for Information Science and Technology*, 55(14), pp. 1290-1303.
- [36] **Avouris, N. (2000).** *Introduction to Human-Computer Interaction*. Athens: Diavlos Publications.
- [37] **Covey, D.T. (2002).** *Usage and usability assessment: Library practices and concerns*. Technical report, Digital Library Federation.
- [38] **Li, Y. & Hu, D. (2013).** Interactive retrieval using simulated versus real work task situations: Differences in sub-facets of tasks and interaction performance. In: *Proceedings of the American Society for Information Science and Technology*. DOI: 50.10.1002/meet.14505001092.

| | | |
|--|--|--|
| | order to satisfy your information needs? | |
|--|--|--|

| Post-task questionnaire | | |
|-------------------------|--|---|
| | Question | Answers |
| 4 | Which system did you find easier to use? | - DSpace subject-browsing system - New subject-browsing system |
| 5 | Which system was easier to understand? | - DSpace subject-browsing system - New subject-browsing system |
| 6 | Did you manage to satisfy your information needs with both of the provided systems? | - Yes - No |
| 7 | Which system did not provide any results? | - DSpace subject-browsing system - New subject-browsing system - Both |
| 8 | Which system was the fastest in satisfying your information needs? | - DSpace subject-browsing system - New subject-browsing system |
| 9 | When employing the new subject system, browsing through subjects (top of the page) helped me in satisfying my information needs? | - Strong disagree - Disagree - Neutral - Agree - Strong agree |
| 10 | Which of the two systems would you prefer in reusing in the future for satisfying your information needs? | - DSpace subject-browsing system - New subject-browsing system - None |

APPENDIX

In this section the pre- and post- questionnaire that was given to the participants of the survey are given.

| Pre-task questionnaire | | |
|------------------------|---|--|
| | Question | Answers |
| 1 | What is your educational level? | - Undergraduate student - Postgraduate student - PhD candidate - Teaching faculty member - University staff - Other |
| 2 | I have experience in searching/ browsing the web | - Strong disagree - Disagree - Neutral - Agree - Strong agree |
| 3 | How often do you employ the digital library of the University of Piraeus in | - One time each semester or less - One to three times a month - Once a week - More often than once a week - Never |

AUTHORS



Konstantinos Kyprianos was born in Athens. He received his bachelor diploma in Librarianship from ATEI of Athens in 2002. He attended his MSc in Computer Science at University of Piraeus. He obtained his Ph. D. from the Department of Archives, Library Science and Museology, at the School of Information Science and Informatics of Ionian University with the topic "Information services based on controlled vocabularies and the semantic web". Currently, he is occupied as adjunct lecturer at the Department of Archival, Library Science and Information Studies at the University of West Attica. His research interests include the semantic web, linked data and digital libraries in general.