Design factors affecting relevance judgment behaviour in the context of metadata surrogates

This item was submitted to Loughborough University's Institutional Repository by the/an author.


Additional Information:

- This article was published in the Journal of Information Science [© The author(s)] and the definitive version is available at: http://dx.doi.org/10.1177/0165551510386174

Metadata Record: https://dspace.lboro.ac.uk/2134/13182

Version: Accepted for publication

Publisher: SAGE Publications on behalf CILIP © The author(s)

Please cite the published version.
This item was submitted to Loughborough’s Institutional Repository (https://dspace.lboro.ac.uk/) by the author and is made available under the following Creative Commons Licence conditions.

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/
Design factors affecting user interaction and relevance judgment behaviour with metadata surrogates: the case of structure, highlighting and quantity of information

Abstract
This paper reports on the findings of a user study which investigated the effects of structure, highlighting and quantity of information on university students’ interaction with metadata surrogates of learning object repositories (LORs) during the relevance judgment process. The literature review revealed a lack of studies investigating how metadata surrogates should be designed to meet users’ needs and improve the effectiveness of relevance judgment. In order to address this issue, different versions of a prototype system, called META-LOR 2, were developed and evaluated in terms of the time needed for users to find relevant information, the accuracy of their relevance judgments and their satisfaction. To evaluate the prototype, a user study was conducted where test subjects were asked to complete a set of tasks and fill in a satisfaction questionnaire. The findings showed that participants performed better and were more satisfied with those versions of the prototype which highlighted metadata elements that included the query terms, used clusters or categories to organize metadata elements in the surrogates and included metadata elements which were relevant to the query or task at hand. Participants performed less efficiently and were less satisfied with metadata surrogates which lacked the use of a clear structure, did not highlight the query terms and contained a large number of metadata elements. It is anticipated that the findings of this research will enhance understanding of users’ interaction with metadata surrogates during the relevance judgment process and improve the design of the search result interfaces of educational information retrieval systems, such as LORs.

1. Introduction
Metadata surrogates are representations of information objects that provide a summary of an object’s contents, characteristics and use, such as the title, creator, subject, language and context (Zeng & Qin, 2008). In the context of information searching behaviour, metadata surrogates can play an important role in the process of relevance judgment (Barry & Schamber, 1998). The significance of their role can be attributed to their ability to provide users with a summary of an object’s characteristics as opposed to the full-text. When this summary describes appropriately an object’s content, characteristics and context of use, users can predict whether the full document is relevant, partially relevant or not relevant to their needs. In other situations, users can use surrogates for relevance feedback or refine their queries based on keywords and terms encountered in the context of the metadata surrogate (Joho & Jose, 2008; Johnson, 2007).

Previous studies have revealed that the presentation and the design of metadata surrogates in search result interfaces of query-driven information retrieval systems and search engines can have an impact on users’ performance and satisfaction when they judge relevance (Joho & Jose, 2008; 2006; Drori, 2003; Rele & Duchowski, 2005). Although these studies revealed some design factors which can affect user interaction with metadata surrogates (such as the use of contextual phrases, the highlighting of query terms as well as the use of thumbnails or typographical characteristics), the impact of the design of metadata surrogates has not been investigated in-depth. In addition, these studies were mainly focused on the context of short snippets of text presented in the search result interface of web search engines and only a few researchers have examined the design factors affecting user interaction with metadata surrogates in the context of structured and metadata-driven information retrieval systems (such as scholarly databases, learning object repositories or institutional repositories) (Balatsoukas, et al., 2009 give a review). The investigation
of user interaction with metadata surrogates provides an opportunity for researchers and evaluators of query-driven information retrieval systems to gain an insight into the different factors which affect users’ relevance judgment behaviour beyond the effectiveness of retrieval algorithms or the utility of the results retrieved. Furthermore, it can enhance understanding of the relevance judgment behaviour in information retrieval research and improve the design of metadata surrogates.

The aim of the research reported in this paper was to investigate the impact of the highlighting query terms, the quantity of information and the structure of metadata surrogates in search result interfaces of learning object repositories (LORs) on user interaction and relevance judgment behaviour in terms of the time needed to find relevant information, the accuracy of relevance judgment and user satisfaction. For this purpose, a user study was conducted which involved the recruitment of participants, completion of a set of tasks using different versions of a prototype database called META-LOR 2, and completion of satisfaction questionnaires. The research focuses on the context of learning object metadata because surrogates based on learning object metadata standards and schemas exhibit some unique features in the design of search result interfaces, such as an increased number of metadata elements per surrogate (semantics) as well as the different types and degrees of organisation and structure of these elements in the surrogates (for example, linear versus clustered design) (Balatsoukas, et al., 2007). However, the findings of the present research can be applied to other types of metadata-driven IR systems as well, such as institutional repositories based on the Dublin Core or AACR2-based online library catalogues. Furthermore, the findings can have implications in the design of the search result interfaces of web search engines.

The paper is structured as follows. In section two a description of the concept of a metadata surrogate is provided. A literature review on user interaction with metadata surrogates is presented in section three. Section four focuses on the three variables under examination (highlighting, quantity of information and structure) and makes a case about their importance in the examination of relevance judgment behaviour. The methodology and the results of the user study investigating user interaction with different versions of a prototype learning object metadata catalogue are presented in sections five and six respectively. The findings are discussed in section seven. Finally, section eight provides some conclusions and recommendations for further research.

2. Metadata surrogates
A metadata surrogate is a substitute for an information object which contains data describing it in terms of several aspects, such as the title, creator, date of publication or audience. Metadata surrogates can support several functions including information retrieval, relevance judgment, management and preservation of information objects as well as the personalised delivery of information based on the match between metadata describing people (users) and information (content).

This paper is focused on the textual metadata surrogates displayed in search result interfaces of a query-driven information retrieval system. These surrogates can help users decide whether a retrieved information object is relevant to their needs or not. Even though the bibliographic records displayed in online library catalogues constitute a form of metadata surrogates, the term ‘metadata surrogate’ or ‘surrogate’
is more frequently used for the description of digital information in the web (Zeng & Qin, 2008). For example, when a user performs a search in the Yahoo! search engine a list of retrieved results is presented in the form of short surrogates that include the title, a few sentences (with the query terms inline) and the URL of the webpage. Information retrieval systems (such as digital libraries or LORs), however, make use of more structured and complex metadata surrogates than the automatically generated snippets of the search result interfaces of common search engines such as Yahoo! or Google. Usually, the metadata surrogates presented in search result interfaces of query-driven information retrieval systems include many metadata elements that correspond to different characteristics of an information object and they provide users with the opportunity to customise the sorting of the results retrieved or the presentation of metadata surrogates. These surrogates often appear in two different states, an abbreviated version which contains only basic information and a detailed version that includes all metadata elements assigned to a specific information object. Basic information about a retrieved metadata surrogate is presented in the search result overview interface, while detailed information can be accessed in a preview interface. The overview interface represents the list of metadata surrogates retrieved after a search has been performed. In this case all metadata surrogates retrieved are aggregated and presented in an alphabetical, chronological or ranked order and are characterized by the presence of a few metadata elements per surrogate. This is similar to the search result interface of many web search engines. In most IR systems each metadata surrogate presented in the overview interface has a corresponding metadata preview interface where the full details of the metadata surrogate are presented. The terms preview and overview were coined from Green, et al (2000) who used them in the context of visual search result interfaces.

Learning object metadata can be defined as “structured data used for the efficient description of learning objects and the effective support of educational functions related to the described learning objects” (Morris, et al., 2007). In order to address the multiplicity of educational functions for which a learning object can be used, learning object metadata surrogates contain many metadata elements that document various characteristics of a learning object, such as its contents, educational value, presentation and use. This often results in lengthy metadata surrogates that can be presented in different formats. The LOM (Learning Object Metadata) standard (IEEE, 2002) defines a set of approximately 80 metadata elements for the description of educational resources (also called learning objects). These are divided into nine categories of General, Life-Cycle, Meta-metadata, Relation, Rights, Educational, Technical, Annotation, and Classification metadata. Although LOM is the dominant standard used for the description of learning objects other metadata schemas used for educational purposes include the Dublin Core Education, the ARIADNE schema (which is an application profile of LOM), the GEM schema or the metadata schema used in MERLOT online system.

3. User interaction with metadata surrogates

3.1. Relevance judgment behaviour
Research on relevance judgment behaviour has implications for the design of metadata surrogates in search result interfaces of IR and the understanding of user interaction with metadata surrogates (Barry, 1998). Studies on relevance judgment
behaviour have been focused on two main research questions: firstly, *What are the effects of the presentation of certain metadata elements on users' relevance judgment (such as the type, order, length and the number of metadata elements)?* and secondly, *What type of relevance criteria do users employ when judging the relevance of metadata surrogates?*

The first question has been investigated by many researchers (such as Saracevic, 1969; 1971; Marcus, et al., 1978; Eisenberg & Barry, 1988, Jane, 1991, Liddy, et al., 2003; White et al., 2005) in a series of experiments using a variety of query-driven IR systems including scholarly databases, online library catalogues, search engines and news databases. Apart from the methodological differences and differences in scope between these studies (a comparison is out of the context of this paper), they provided some useful insights about user relevance judgment behaviour such as the importance of an abstract for judging relevance, the significance of designing surrogates which include sentences with keywords in context as well as participants’ preferences for content-related metadata elements (such as title, abstract, keywords and subject categories). In addition, from a methodological point of view, many researchers who investigated the impact of the presentation order of metadata surrogates on users’ relevance judgment found that experimenters in relevance behaviour research should present metadata surrogates randomly to users in order to eliminate bias from order effects (see Eisenberg & Barry, 1988; Huang & Wang, 2004; Xu & Wang, 2008).

An interest in the second research question was developed during the early 1990s by information scientists at Syracuse University and still remains a significant research problem among information scientists who investigate user-centred relevance behaviour in IR systems. This field of research is focused on the criteria users employ to judge the relevance of various types of documents and document representations (i.e. bibliographic records and metadata surrogates). Researchers have investigated relevance criteria in the context of both web search engines (i.e. the unstructured surrogates or snippets presented in the search result interface of web search engines) and structured information retrieval systems (i.e. metadata surrogates and bibliographic records presented in the search result interface of scholar databases and other types of metadata-driven IR systems). This type of research helped the understanding of the multi-dimensional and dynamic nature of relevance (Borlund, 2003; Spink et al, 1998). For example, it is widely reported in the literature that users do not only judge relevance based on the subject matter or the topic of a resource. Other criteria used for relevance judgment may include the currency, quality, authority and availability of the resource as well as users’ subject background and characteristics. Lists of such criteria have been proposed by several researchers (see for example, Barry & Schamber, 1998; Wang & Soergel, 1998; Crystal & Greenberg, 2006). Researchers in this field have also argued that the importance of relevance criteria may change throughout the information seeking and searching process exposing its dynamic dimension (Rieh, 2002; Tang & Solomon, 2001; Crystal & Greenberg, 2006). This change can be attributed to several individual characteristics, such as topic familiarity, level of interest with the topic and confidence in assessing relevance (Ruthven, et al., 2008).

Although these studies provided a good overview of the impact of the content of metadata surrogates on user relevance judgment behaviour, they did not examine issues related to the presentation-design of metadata surrogates in search result
interfaces (for example, is there an effect of metadata surrogate design on user relevance judgment?). The design of metadata surrogates can refer to several factors, such as positioning of metadata elements, the structure of the surrogate or the quantity of information presented in the surrogate (Balatsoukas, et al., 2009). The remainder of this section summarises some attempts made by researchers to address these design issues in the context of various types of IR systems and web search engines.

3.2. Studies on user interaction with metadata surrogates

In one of the few studies that investigated users’ interaction with metadata surrogates, Fraser and Gluck (1999) found that participants preferred metadata elements to be grouped into semantically relevant categories. They found that participants also expected the metadata elements that provided or arranged access to the information object to follow content related information in a sequence. In terms of the quantity of information included in the surrogate, participants were interested in surrogates that included all necessary metadata elements for judging relevance, such as title, description, keywords, location and technical characteristics. Therefore, a balance should be maintained as far as concerns the quantity of metadata that should be presented in a surrogate. The study was focused on the context of Geographical Information Systems (GIS), and justified the use of both quantitative and qualitative data collection techniques in order to investigate users’ interaction with metadata surrogates.

Other researchers have examined user interaction with specific components of surrogates such as the summary or snippet, keywords and categories (Drori, 2000; 2003; Paek et al., 2004), the organisation of metadata elements (Rele & Duchowski, 2005), the display of thumbnail previews (Dziadosz & Chandrasekar, 2002; Joho & Jose, 2006) and the quantity of information presented (Joho & Jose, 2008). These studies, however, were conducted in the context of search engines rather than structured (in terms of metadata) information retrieval systems, such as scholarly databases and LORs.

Only three studies have been focused specifically on user interaction with learning object metadata surrogates. The first was based on a set of interviews held for the evaluation of the Search LT repository (Search LT evaluation report, 2002). The results revealed that participants preferred metadata elements that were easy to understand and grouped into categories. In this case, it was recommended that tabs or category labels should be made visible for users to identify and access the components of the surrogate (Search LT evaluation report, 2002). In the second study, Balatsoukas et al (2007) investigated the impact of two different presentations of LOM-based learning object metadata surrogates (linear and clustered) on users’ performance and satisfaction during the relevance judgment process. The findings revealed that participants were more satisfied and performed the tasks faster using the interface that presented metadata surrogates in a clustered format rather than as a linear list of metadata elements. The third study was focused on the usefulness of GEM-based metadata elements presented in metadata surrogates. The GEM (Gateway for Educational Material) is a US based portal of learning material that applied the GEM metadata schema (a Dublin Core application profile) for organising them. The study (based on interviews and eye tracking data) revealed that users preferred viewing the description metadata element. When a description was not present then
users tended to focus on other metadata elements and specifically the title, subject, rights and creator elements. Users also suggested some further cues that could be included in metadata surrogates for learning objects as a means for facilitating relevance judgment. These included: relevance ratings, reviews, and comments from others who had done similar searches (Liddy, et al., 2003; Diekema, 2006).

Although these studies provided some useful findings about users’ interaction with learning object metadata surrogates, they lacked an explicit focus on the presentation. For example, in the case of the Search LT evaluation, the purpose of the study was to evaluate the usefulness of the system as a whole without systematising its focus on the presentation and design of metadata surrogates, while in the case of the GEM study the researchers emphasised the importance of the content of the metadata surrogate for relevance judgment and did not investigate the impact of interface design on users’ interaction and relevance judgment behaviour. The research reported in the remainder of this paper extends the one conducted by Balatsoukas et al (2007) and provides an in-depth interpretation of users’ performance and satisfaction with learning object metadata surrogates during the relevance judgment process.

4. Highlighting, Structure and Quantity of information

The selection of the variables under examination (highlighting, quantity of information and structure of metadata surrogate) was based on the results of a preliminary study which involved interviews with 19 students, studying different disciplines, at Loughborough University. The data collected from the interviews formed part of a larger study which examined the usability of the metadata surrogates (see Balatsoukas, et al., 2008) and provided input for the design of the user study described in the next section of this paper (Section 5). The present section summarises the main findings of the interviews and the reasons why the three variables (highlighting, structure and quantity of information) were selected for further examination.

The purpose of the interviews was to investigate the perceptions and preferences of university students regarding the presentation of metadata surrogates in the search result interface of three LORs: the ARIADNE Knowledge Pool, the MERLOT system and the JORUM repository. The selection of the three systems was based on the fact that they differed in terms of their metadata data structure and the design of the metadata surrogates presented in the search result interface. It was anticipated that the presentation of metadata surrogates with different characteristics would motivate participants to provide more accurate and rich responses.

The analysis of the interview protocols showed that eight main design factors influenced participants’ perceptions about metadata surrogates. These were:

1. The quantity of information in the surrogate.
2. The structure of the metadata surrogate.
3. The positioning of metadata elements in the surrogate.
4. The highlighting of query terms and key metadata elements.
5. The indication of the quality of the resource.
6. The typographical characteristics of the surrogate (font type and size, graphics and colour).
7. The use of hypertext in the surrogate.
8. The content of the metadata surrogate (for example, metadata element semantics).

The findings showed that the structure, highlighting and quantity of information appeared to be the most critical factors of user performance and satisfaction when judging relevance. The remainder of this section gives an overview of participants’ responses regarding the three variables (highlighting, structure and quantity of information) and justifies their inclusion in the user study.

**Highlighting of query terms:**
Almost half of the participants in the interviews disliked the fact that query terms were not highlighted within the context of the metadata surrogate. Since none of the LORs used in this study employed highlighting of query terms it is worth investigating further whether the addition of this component in the case of structured learning object metadata surrogates could improve user performance and satisfaction during the relevance judgment process. Previous studies in the context of web search engines revealed that participants performed better and were more satisfied with search result interfaces that included highlighting of query terms rather than with their non highlighted counterparts (Drori, 2000; 2003). This happened because participants in Drori’s studies needed less time and spent less attentional and cognitive effort to identify parts of the surrogate that were relevant to their queries during the relevance judgment process. There are no studies, however, investigating the effects of highlighting in structured IR systems, such as in the case of LORs. For example, in the case of web search engines, highlighting was used in the title and the abstract displayed in the search result interface, while in the case of structured IR systems, such as LORs, highlighting should be applied to all metadata elements that include terms or phrases that are relevant to a user’s query. Psychologists and cognitive scientists have investigated for a long time the effects of different forms of presentation of various types of stimuli on human visual search behaviour and attention and they have agreed that the stimuli that stand out from the background (for example, due to different shape, text colour or background colour) are perceived immediately by humans without spending too much visual and attentional effort (Sternberg, 2003).

**Structure of the metadata surrogate:**
Categorisation or grouping is an important function of the human mind and it is vital in the processing of information. There is a need, however, to identify what type of organization or categorization of information in the metadata surrogate could improve user interaction during the relevance judgment process. For example, although the results of the interviews showed that participants liked the grouping of metadata elements into semantically relevant sections (as in the case of ARIADNE), many of these participants proposed alternative ways of structuring and presenting metadata surrogates, such as the use of clusters of metadata elements. A specific characteristic of these clusters would be the presentation of one group of metadata elements in a single screen. In this case the user could select which group of metadata elements was displayed in the screen. Some benefits of this type of design might be:
The focus of attention on a set of specific metadata elements as opposed to the whole information contained in the surrogate. This could minimize the risk of ignoring metadata information that is useful for relevance judgment. For example, some participants in the interviews agreed that it was difficult to read all the information displayed in a metadata surrogate of JORUM at the preview level. This happened because users had to interact with a long list of metadata elements (approx. 50-60) displayed in a linear format.

The decrease in the need for users to scroll down long lists of metadata elements.

The presentation of metadata elements into clusters or categories has been also suggested by participants in the evaluation of the Search LT system (SearchLT evaluation report, 2002) and by Wang (1997) as a part of the document selection model, but it has not been empirically tested.

**Quantity of information in the surrogate:**
In the case of the quantity of information included in the surrogate, some participants liked the fact that the metadata surrogates in the MERLOT system contained a fair amount of metadata elements (i.e., less information than the ARIADNE and JORUM, but not as little information as in the case of web search engines, such as Google). However, most participants suggested that metadata surrogates should include only metadata elements which were relevant to a specific query. This would minimize the amount of information in the surrogate, decrease the time needed by users to scan the surrogate and increase the relevance of the information contained in metadata surrogates to users’ tasks at hand.

These three factors or variables (highlighting, structure and quantity of information) were embedded in the design of eight versions of a prototype learning object metadata catalog, called META-LOR 2 (see Section 5). In addition to these variables, the research design involved the investigation of the effects of task complexity on user performance and satisfaction with learning object metadata surrogates during the relevance judgment process.

**Effects of task complexity:**
Although research focused on the usability of search interfaces has found significant effects of task complexity on users’ performance during query formulation (Chan, et al., 1997; Topi, et al., 2005), there are no studies that have investigated the effects of task complexity on users’ performance and satisfaction with metadata surrogates during the relevance judgment process. **Component complexity**, which is based on the number of different information cues that need to be processed by a user to complete the task, has direct implications for the evaluation of metadata surrogates. For example, tasks of high component complexity require users to process a larger number of metadata elements in order to find a relevant resource. It is reasonable to test the component complexity of tasks, because the intention of this study is in the judgment of relevance where complexity is usual. Therefore, it is worth investigating further whether the level of task complexity affects user performance and satisfaction during the relevance judgment process.
5. Research design

A total of 24 university students participated in the study. None of these had participated in the interviews (see Section 4). Participants were recruited by means of emails and announcements on University notice-boards. Before the experiment a background questionnaire was completed by the candidate participants. The background questionnaire facilitated the final selection of the participants in the study based on their familiarisation with Electronic Information Services (EIS) and the web. Due to the exploratory nature of this study the selection of participants was limited to experienced web users only. It was anticipated that this group of participants would be more motivated to perform the tasks than less experienced users and therefore, the data collected from their interaction would be more accurate (Rubin and Chisnell, 2008). In addition, the homogeneity of this sample could increase the rigidity and validity of the statistical analysis.

Of the 24 participants, 18 were taught postgraduate students and the remaining six were PhD students. All participants, who were students in a variety of disciplines, were familiar or very familiar users of the World Wide Web. Participants used the web for a variety of reasons, such as reading textual information, communicating with peers, listening to music and watching videos. They were also frequent users of various types of EIS, such as the library online catalogue, the library portal (Metalib), scholarly databases and search engines. They were not familiar with the use of LORs or digital libraries of educational material.

During the experiment, participants were asked to complete a set of two relevance evaluation tasks using eight versions of the META-LOR 2 prototype (2 x 8). Each version of the prototype included a different representation of metadata surrogates typical of search result interfaces (see Section 5.1 for a detailed presentation of the META-LOR 2 system). For each task, participants were instructed to evaluate the relevance of the first 10 metadata surrogates of the result page of the META-LOR 2 prototype. This procedure was repeated for the eight versions of the prototype using two tasks per version. Since this study was focused on the effectiveness of metadata surrogates, participants did not have to evaluate the relevance of the learning object itself. However, they were required to access the full-text of the corresponding learning object they perceived to be relevant to their needs, by clicking on a relevant link provided in the metadata surrogate. Then, participants had to return back to the results page and continue the evaluation of the remaining metadata surrogates. It is worth mentioning, that by accessing the full text of a learning object, participants indicated which metadata surrogates were relevant to their needs. Participants’ selections were matched against relevance judgments made by the authors of this paper before the initiation of the experiment in order to calculate the accuracy of participants’ relevance judgments.

A within-subjects design was employed that required all participants to perform the same set of task across the eight interfaces. The sequence with which the interfaces and the tasks were presented to the subject was randomly altered for counterbalancing the effects of ‘learning transfer’. The two tasks performed with each interface differed
in terms of component complexity (low and high complexity). Table 1 presents two examples of low and high complexity tasks used in this experiment.

<table>
<thead>
<tr>
<th>Task</th>
<th>Complexity level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find a lecture on the use of PowerPoint</td>
<td><strong>Low complexity</strong> (Chunks to evaluate for relevance: lecture and power point)</td>
</tr>
<tr>
<td>Find a lecture of high interactivity about the use of METALIB for</td>
<td><strong>High complexity</strong> (Chunks to evaluate for relevance: lecture, high interactivity,</td>
</tr>
<tr>
<td>Higher Education students in HTML format.</td>
<td>METALIB, Higher Education, students and HTML format).</td>
</tr>
</tbody>
</table>

The inclusion of both low and high complexity tasks in the research design was designed to enhance the generalisability of the results of the present study. All tasks reflected general themes, such as ‘find a lecture about statistics’ or ‘find a tutorial about how to write an essay’.

Participants’ behaviour was captured by the use of screen recording software (Camtasia Studio, v. 4). After the completion of the low and high complexity task in each version of the search result interface, participants were asked to complete a short subjective satisfaction questionnaire. This measured users’ satisfaction against a set of measures, such as ease of use, stimulation, frustration and satisfaction with the presentation of metadata surrogates. Each variable was measured using a 7-point Likert scale (for example, dull vs stimulating, frustrating vs satisfying, difficult vs easy and did not like vs liked the presentation of the metadata surrogate). By the end of the experiment each participant had completed a total of eight questionnaires (one questionnaire for each version of the META-LOR 2 prototype). The research instruments were piloted and then testing took place during July – August 2008.

The data analysis included estimation of the means and statistical analysis for the time needed for users to find relevant documents in the retrieved set (two way within subjects ANOVA tests) as well as for users’ ratings of subjective satisfaction (Friedman tests).

5.1. The META-LOR 2 system

For the needs of this study, eight versions of a prototype learning object metadata database, called META-LOR 2, were created using HTML, JavaScripts and XML technology. The underlying system stored and provided access to 160 learning object metadata records coded in XML. Metadata records included an identifier that provided access to the learning object itself. The data structure of the META-LOR 2 system was based on 20 elements that were derived from the UK LOM Core schema.
(an application profile of LOM). The number and terminology of the elements was finalised after pilot testing.

Each version consisted of three main interfaces. The first provided a ready list of queries to users (for example, “Find a lecture about digital divide”) (Figure 1). Since the focus of this study was on the investigation of user relevance judgment behaviour during the interaction with metadata surrogates and not query formulation or the interaction between the two (i.e. search and search result interaction), the queries were developed in advance.

The second interface was the ‘search result overview’ that provided a list of the retrieved results. Each metadata surrogate in the list included only the title of the learning object, the name of the author of the learning object, a human generated abstract of the contents of the learning object and a link to the third interface (Figure 2). The third interface was the ‘metadata surrogate preview’ which included a detailed description of a learning object (all 20 metadata elements) (Figures 3-8).
The eight versions of the META-LOR 2 system represented different types of metadata surrogates in the search result interface. The eight versions were:

1. **Linear (non-highlighted).** This version included a standard search result overview interface and a preview interface with all metadata elements displayed in a list (Figure 3).

2. **Linear (highlighted).** As opposed to its non-highlighted counterpart, this version highlighted any query terms or phrases presented in the metadata surrogate. The highlighting was applied both at the overview and preview level using a yellow background colour (Figure 4).

3. **Linear with categories (non-highlighted).** This version included a standard search result overview interface and a preview interface which presented all metadata elements in a list which were divided into three categories. The three categories were: General, Educational and Technical (as in the case of the clustered version). However, as opposed to the clustered version, the Linear with categories version presented all three categories of metadata elements in the same interface (Figure 5).

4. **Linear with categories (highlighted).** In this version, highlighting of query terms was applied at both the overview and preview level (Figure 6).

5. **Clustered (non-highlighted).** This version included a standard search result overview interface and a preview interface which presented groups of metadata elements in clusters. The clusters represented three different types of metadata elements. These were: General (Title, Author, Subject, Description, Date, Language, Cost, Copyright), Educational (Learning resource type, Interactivity type, Interactivity level, Audience, Context, Typical Age Range, Difficulty and Typical Learning Time) and Technical (Format, Identifier, Size, Technical Requirements). Only one cluster of metadata elements appeared each time in the preview interface (Figure 7).
6. **Clustered (highlighted)**. The highlighted version of the clustered interface applied highlighting both at the overview and preview level (Figure 8).

7. **Query metadata (non-highlighted)**. The Query metadata version included only a search result overview interface. However, as opposed to the standard version, the overview interface included only the metadata elements which were relevant to the task at hand (Figure 9).

8. **Query metadata (highlighted)**. As opposed to its non-highlighted counterpart this version applied highlighting for query terms which appeared in the metadata surrogate presented in the search result overview (Figure 10).

![Figure 3. Linear interface (preview level).](image1)

![Figure 4. Linear highlighted interface (preview level).](image2)
Figure 5. Linear with categories (preview level).

Figure 6. Linear with categories highlighted (preview level).

Figure 7. Clustered (preview level).

Figure 8. Clustered highlighted (preview level).
Table 2 provides a summary of the metadata elements and element categories used in the 8 versions of the META-LOR 2 prototype. Furthermore, the numbering within the three columns of Table 2 demonstrates the sequence with which metadata elements were presented in six out of the eight versions (this does not apply to Query metadata highlighted and non-highlighted versions since both versions presented only metadata elements relevant to the tasks at hand).

<table>
<thead>
<tr>
<th>General</th>
<th>Educational</th>
<th>Technical</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Title</td>
<td>9. Learning Resource Type</td>
<td>17. Format</td>
</tr>
<tr>
<td>2. Author</td>
<td>10. Interactivity Type</td>
<td>18. Identifier</td>
</tr>
<tr>
<td>5. Date</td>
<td>13. Context</td>
<td></td>
</tr>
<tr>
<td>7. Cost</td>
<td>15. Difficulty</td>
<td></td>
</tr>
<tr>
<td>8. Copyright</td>
<td>16. Typical Learning Time</td>
<td></td>
</tr>
</tbody>
</table>

The selection of the 20 metadata elements was based on the following factors:

1. These elements were defined as mandatory by the developing UK LOM CORE (Campbell, 2004).
2. These elements were proposed by the CanCore, the Canadian application profile of LOM, to be more appropriate for presentation in search result interfaces. The CanCore is accompanied by detailed guidelines and it is the
only application profile that refers to the type of metadata elements that should be presented in search result interfaces (Friesen, 2004).

3. One further element of LOM was used in the design of the metadata surrogates. This was the Difficulty element. This element was not proposed for mandatory use by the UK LOM Core. It was not one of the elements which, according to the CAN Core guidelines, should be included in the search result interface. However, the decision to include the Difficulty element in this study was made in order to enhance the number of educational metadata elements of the surrogates, as well as because the particular element was identified as an important criterion for relevance judgment by previous studies (Small et al., 1998; SearchLT evaluation report, 2002).

All interfaces included the same font type and size (Times New Roman, 12 pt), colour of links (blue), font colour (black) and background colour (white). In addition, due to the controlled nature of this study, no added functionality was included in the search result overview and preview interfaces, such as the opportunity for users to print, save or email the results retrieved. Also the positioning of metadata elements did not change across the interfaces. Furthermore, no advertisements or links to non relevant resources were included. Finally, for all versions of the META-LOR 2 prototype participants used a 17” screen desktop pc.

6. Results

6.1. Impact of interface design on time
As shown in Table 3, participants needed more time to complete the tasks using the ‘linear non highlighted’ interface (mean time = 290secs), but they performed the tasks more efficiently in the case of the ‘query metadata highlighted’ interface (mean time = 70.5 secs).

<table>
<thead>
<tr>
<th>Interface type</th>
<th>Mean</th>
<th>Std.Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear (highlighted)</td>
<td>158.5</td>
<td>5</td>
</tr>
<tr>
<td>Linear (non highlighted)</td>
<td>290</td>
<td>13.3</td>
</tr>
<tr>
<td>Clustered (highlighted)</td>
<td>159</td>
<td>5.2</td>
</tr>
<tr>
<td>Clustered (non highlighted)</td>
<td>257</td>
<td>10</td>
</tr>
<tr>
<td>Linear with categories (highlighted)</td>
<td>156</td>
<td>3.5</td>
</tr>
<tr>
<td>Linear with categories (non highlighted)</td>
<td>226.6</td>
<td>9.6</td>
</tr>
<tr>
<td>Query metadata (highlighted)</td>
<td>70.5</td>
<td>2.1</td>
</tr>
<tr>
<td>Query metadata (non highlighted)</td>
<td>94</td>
<td>2.6</td>
</tr>
</tbody>
</table>
The results of the two-way within subjects ANOVA revealed that there were significant differences at the 0.01 level in the time needed to identify relevant learning objects between the 8 interfaces. In particular, the Bonferroni pairwise comparison test revealed that the interfaces that presented a small quantity of metadata elements relevant to the query (i.e. ‘query metadata highlighted’ and ‘query metadata non highlighted’) differed significantly from the other interfaces in terms of time at the 0.01 level. Furthermore, the four interfaces with query terms highlighted in the metadata surrogate were significantly more time efficient that the non highlighted interfaces (difference significant at the 0.01 level). Finally, non significant differences in time were observed between the linear, clustered and linear with categories interfaces.

In order to investigate whether there was an impact of task complexity on users’ performance it was necessary to examine first whether there were true differences between the two tasks. The identification of differences between the two tasks suggests that the division of tasks in the two categories of low and high complexity is valid. The average mean time revealed that participants needed less time to complete low complexity task, but they spent more time in the case of the high complexity tasks (low complexity = 168.5 seconds and high complexity task = 184.3 seconds). The results of the two-way within subjects ANOVA revealed that this difference in mean time between the two tasks was significant at the 0.01 level. Therefore, it is suggested that there were differences between the two tasks that validate their categorisation as low and high complexity.

Table 4 summarises the mean time needed for users to perform the two tasks across the eight interfaces. Participants performed the low complexity task faster using the Query metadata highlighted interface (mean time = 66 secs) and slower using the Linear non highlighted interface (mean time = 277). The second most time efficient interface for the low complexity task was the Query metadata non highlighted interface (mean time = 86 secs). The highlighted interfaces performed better than the non highlighted interfaces in the low complexity task. The findings were similar in the case of the high complexity task. In the case of this task, the Query metadata highlighted interface was the most time efficient interface (mean time = 75 seconds) followed by the Query metadata non highlighted interface (mean time = 101 seconds). The Linear non highlighted interface was the least time efficient (mean time = 303 seconds), while in the case of the highlighted interfaces, participants performed better than the non highlighted ones (see also Table 4).

<table>
<thead>
<tr>
<th>Interface type</th>
<th>High Complexity</th>
<th>Low Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear (Highlighted)</td>
<td>174</td>
<td>153</td>
</tr>
<tr>
<td>Linear (Non Highlighted)</td>
<td>303</td>
<td>277</td>
</tr>
<tr>
<td>Clustered (Highlighted)</td>
<td>145</td>
<td>133</td>
</tr>
<tr>
<td>Clustered (Non Highlighted)</td>
<td>266</td>
<td>248</td>
</tr>
<tr>
<td>Linear With Categories (Highlighted)</td>
<td>172</td>
<td>150</td>
</tr>
<tr>
<td>Linear With Categories (Non Highlighted)</td>
<td>239</td>
<td>215</td>
</tr>
<tr>
<td>Query Metadata (Highlighted)</td>
<td>75</td>
<td>66</td>
</tr>
<tr>
<td>Query Metadata (Non Highlighted)</td>
<td>101</td>
<td>86</td>
</tr>
</tbody>
</table>

Table 4. Difference in time across the two tasks
Two one-way within subjects ANOVA tests were performed. The first was focused on the identification of differences between the interfaces for the low complexity task, while the latter was focused on the high complexity task. The ANOVA test showed that there were significant differences between the eight interfaces for both Low and High complexity tasks at the 0.01 level. In the case of the low complexity task, participants needed significantly less time to judge relevance using the two Query Metadata interfaces (both the highlighted and non highlighted) rather than the six interfaces that included a larger quantity of metadata elements in displayed in different structures (for example, Linear, Clustered and Linear with categories). In addition, the highlighted interfaces (where the query terms were highlighted in yellow) were more time efficient than their non highlighted counterparts. However, non significant differences were observed between the Clustered and Linear interfaces (Linear and Linear with categories) for both levels of task complexity.

Finally, the data analysis showed no impact of the interaction between level of task complexity and interface type on the time needed by participants to judge the relevance of learning object metadata surrogates (for example, the interface types that were more or less time efficient were the same for both types of tasks).

6.2. Impact of interface design on the accuracy of relevance judgments.
The analysis of the data about the accuracy of participants’ relevance judgments showed that the non highlighted interfaces were less effective for relevance judgment than the highlighted interfaces. All participants made accurate relevance judgments in the case of the highlighted interfaces. In the case of the non highlighted interfaces, however, seven participants did not identify the correct learning object in the Linear interface, six participants in the Linear with categories interface and three participants in the Clustered and Query metadata interfaces respectively. In addition, relevance judgments were less accurate in the case of the high complexity than the low complexity tasks. It is interesting that the Clustered non highlighted interface was as accurate as the Query metadata non highlighted interface. The latter, however, was significantly more time efficient than the former. Table 5 presents a summary of the accuracy of the relevance judgments across the non highlighted interfaces.

<table>
<thead>
<tr>
<th>Interface type (non highlighted)</th>
<th>High complexity tasks</th>
<th>Low complexity tasks</th>
<th>Total (n/24)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>5</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Linear with categories</td>
<td>4</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Clustered</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Query metadata</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

* The Total represents the total number of incorrect relevance judgments across the 24 participants in the study.

6.3. Impact of interface design on satisfaction
Table 6 presents a summary of user satisfaction with the eight interfaces. Participants were less frustrated, more stimulated and more satisfied with the presentation of the “query metadata highlighted” interface. The clustered highlighted interface, however,
was perceived to be easier to use. Furthermore, the clustered interfaces (highlighted and non highlighted) scored higher than the linear or linear with categories interfaces across the four satisfaction measures. In addition, participants were more satisfied with the highlighted interfaces than with the non highlighted ones. Also, the metadata surrogates which presented only information relevant to the user query or tasks at hand, (‘query metadata highlighted’ and ‘non highlighted’) scored higher than the longer surrogates which contained a detailed list of metadata.

A set of Friedman tests were performed to examine the impact of interface design on learners’ subjective satisfaction. The findings revealed statistically significant differences for all satisfaction measures at the 0.01 level.

The subjective satisfaction questionnaire also measured participants’ perceived difficulty in completing the two tasks (low and high complexity tasks) using the eight interfaces. The results are summarised in Figure 11. In particular, this figure shows that participants found it easier to complete both types of tasks using the highlighted interfaces rather than the non-highlighted ones. The results of the Wilcoxon tests showed that this difference was significant for both tasks at the 0.0005 level. Furthermore significant differences, at the 0.01 level, were observed between the Clustered interface and the Linear and Linear with categories interfaces for both the low and high complexity tasks. Also, participants found it significantly easier to complete the low and high complexity tasks using the Query metadata interfaces rather than the interfaces that included more metadata elements (Linear, Clustered and Linear with categories). The differences between interfaces were observed at the

<table>
<thead>
<tr>
<th>Interface type</th>
<th>Frustrating vs Satisfying</th>
<th>Dull vs Stimulating</th>
<th>Difficult vs Easy</th>
<th>Did not like vs Liked the presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear (highlighted)</td>
<td>4.33</td>
<td>3.67</td>
<td>5</td>
<td>3.83</td>
</tr>
<tr>
<td>Linear (non highlighted)</td>
<td>2.96</td>
<td>2.38</td>
<td>3.42</td>
<td>2.67</td>
</tr>
<tr>
<td>Clustered (highlighted)</td>
<td>3.75</td>
<td>3.92</td>
<td>5.79</td>
<td>4.33</td>
</tr>
<tr>
<td>Clustered (non highlighted)</td>
<td>3.67</td>
<td>3.58</td>
<td>4.17</td>
<td>4.13</td>
</tr>
<tr>
<td>Linear with categories (highlighted)</td>
<td>3.67</td>
<td>3.29</td>
<td>3.79</td>
<td>3.42</td>
</tr>
<tr>
<td>Linear with categories (non highlighted)</td>
<td>3.08</td>
<td>2.88</td>
<td>4.33</td>
<td>2.96</td>
</tr>
<tr>
<td>Query metadata (highlighted)</td>
<td>5.54</td>
<td>4.88</td>
<td>4.88</td>
<td>5.29</td>
</tr>
<tr>
<td>Query metadata (non highlighted)</td>
<td>4.53</td>
<td>4.63</td>
<td>5.04</td>
<td>5.13</td>
</tr>
</tbody>
</table>
0.01 level. Finally, although differences between interfaces were observed within the two levels of task complexity, no significant differences were observed between the two types of tasks. This means that there was no impact of the interaction between task complexity and interface type on user satisfaction.

Figure 11. Differences in perceived difficulty (7-point scale)

7. Discussion

The results of this study suggest that human information processing constraints can influence the way people approach the learning object metadata surrogates particularly in terms of the use of highlighting, quantity of information and structure. The remainder of this section discusses the findings of the present study in the light of other studies conducted in information seeking research and human computer interaction.

Quantity of information
The quantity of information in the metadata surrogate had an effect on users’ performance and satisfaction. For example, participants performed the tasks faster, made more accurate relevance judgments and were more satisfied with the Query Metadata version of the META-LOR 2 prototype which presented only metadata elements that were relevant to users’ queries or tasks at hand. Although these findings suggest that metadata surrogates with fewer metadata elements can increase the efficiency of relevance judgment as well as user satisfaction, the results of the relevance judgment behaviour research have shown that participants judge relevance using a large number of relevance criteria that can be mapped to several metadata elements. In accordance with these findings some researchers have suggested that the design of surrogates should be rich in metadata elements (Crystal & Greenberg, 2006; Barry, 1998). The differences between these studies can be attributed to the nature of the tasks involved (pre-determined tasks versus real information needs). For example, in the evaluation of META-LOR 2 participants had to perform a set of pre-defined tasks. In this case participants had to match or evaluate the relevance of a surrogate against the criteria specified in the simulated task at hand (such as Audience, Topicality and Format). The presence of information in the surrogate that matched the relevance criteria triggered by the task at hand indicated the presence of relevant content. The absence of a complete match between metadata elements and relevance criteria indicated the lack of relevance. The findings of other studies that employed fixed tasks and known item tasks for the investigation of the impact of different types of search result interfaces and metadata surrogates on search result interfaces have been similar (see for example, Paek, et al., 2004; Drori, 2000; 2003; Sebrechts, et al., 1999). These studies revealed that participants performed the tasks faster and were more satisfied with shorter metadata surrogates and search result interfaces.

Summarising, it appears that the quantity of information in a surrogate is related to the type of task (i.e. controlled or naturalistic). The results showed that metadata surrogates with a small quantity of metadata can meet the needs of fixed or controlled tasks. In the case of less defined and naturalistic tasks, however, metadata surrogates should include a large quantity of information. In the latter case, the design of rich metadata surrogates, in terms of content, can support several functions, such as relevance judgment, query modification or even serendipity.

Structure of metadata elements
The results of the study showed that participants performed better and were more satisfied with the clustered and linear with categories metadata interfaces rather than with the linear only interface. This finding shows that the former interface made the presence of some kind of organisation of metadata elements explicit through categorisation in semantically discrete groups. This however, was not evident, at least at a perceptual level, in the case of the linear interface. Categorisation or grouping is an important function of the human mind and it is vital in the processing of information. The principles of cognitive economy and perceived world structure have been used for a long time in psychology to indicate two complementary functions of human mind for organisation and categorisation tasks (Rosch, 1978). Cognitive economy involves the aggregation of similar entities into groups and the differentiation of these entities from the entities included in other groups. The principle of world structure is based on the human perceptual ability to identify structure in physical [or digital] materials. This categorisation serves the need for not
only identifying similar information in the group but also for differentiating metadata included in other groups. Of course, other categories could arise as indicated by some other participants, such as the division between important and less important metadata elements or based on their frequency of use for relevance judgment. These types of categorisation have been proposed by researchers in human computer interaction and menu design research who argued that items should be grouped either in terms of similarity, importance or frequency of use (Mehlenbacher, 1989; Shneiderman & Plaisant, 2004; Liu, et al., 2002).

Besides structure and organisation of metadata surrogates, the use of clear labels improved participants’ visual search process. For example, in the case of the clustered and linear with categories interfaces, participants relied on the labels which accompanied each cluster or category of metadata elements in order to decide whether a group included useful information for relevance judgment. This finding aligns with the systematic visual search process according to which the user first inspects the labels of each group and then selects the target label in order to view its contents (Hornof, 2004).

**Highlighting**
The highlighting of query and key terms in the metadata surrogate can have an impact on users’ performance and satisfaction. This was evident in the user study, the results of which showed that participants performed the tasks faster (identified relevant learning objects) and were more satisfied with the highlighted rather than non highlighted interfaces. This was also suggested, but not empirically tested, by other researchers who investigated the perceptions of users about metadata surrogates in search result interfaces, such as Fraser and Gluck (1999) Su (2003) and Wang and Soergel (1998). In particular, they suggested that key metadata elements (such as the title and abstract) should be highlighted for visibility. Other researchers have found that the highlighting of query terms in the surrogates of search engines can speed up the relevance judgment process (Drori, 2000; 2003). This finding is not surprising given the fact that human perception and attention are influenced by the form and presentation of stimuli or items in the environment (Sternberg, 2003). In the case of the highlighted metadata surrogates, participants paid immediate attention to the (yellow) highlighted text of the surrogates rather than the other non highlighted information.

**8. Conclusions**
The results of our research showed that participants’ interaction with metadata surrogates in search result interfaces of LORs can be influenced by design factors, such as the structure of the surrogate, the highlighting of query terms and the quantity of information included in the surrogate. The discussion of the findings showed that these factors are influenced by human mental mechanisms or information processing constraints. Since this is the first study to examine the effects of the design of metadata surrogates on users’ relevance judgment behavior, any conclusions should be made with caution. Further research is needed on the generalisation of these findings in different contexts (for example, the search result interfaces of search engines, scholarly databases and digital libraries) and with different types of tasks. In addition, the repetition of this study with different types of users (such as novice users
of the WWW or university teachers) and real information needs could enhance the understanding of user interaction with metadata surrogates.

Some implications of this research include the user-centred design of learning object metadata surrogates in search result interfaces as well as knowledge about users’ interaction with metadata surrogates. Based on the results of the study reported in this paper some recommendations for improving the presentation of metadata surrogates can be made. Some of these include:

- The application of clear structure for the contents of the metadata surrogate. Such structure should include grouping of semantically related metadata elements, as well as the inclusion of clear labels to indicate the contents of each group.
- The highlighting of query terms or key metadata elements in the surrogate.
- The inclusion of metadata in the surrogate elements that are necessary for users to judge the relevance of learning objects.

Finally, except for the design of user-centred metadata surrogates, the implications of this research can be extended to the evaluation of IR systems. In particular, the results suggest that other factors beyond the effectiveness (operationalised as recall and precision) or utility of search results can account for user performance and satisfaction with search result interfaces during the relevance judgment process. Therefore, this finding suggests the incorporation of new criteria or variables (such as the structure of metadata surrogate or the highlighting of query and key terms) in the investigation of user satisfaction and performance with search result interfaces. This is of particular importance for frameworks examining the multi-dimensional evaluation of information retrieval systems (for example, Su, 2003 and Johnson, et al., 2003). Novel research questions and hypotheses could be formulated in order to investigate the relationship between perceived recall - precision and the structure of a metadata surrogate, the highlighting of query terms, or the quantity of information in the surrogate. The examination of these research questions could provide further insight about the factors that affect user relevance judgment behaviour.

8. References


Campbell, L. (2004). UK LOM Core update. CETIS. Available at: [http://metadata.cetis.ac.uk/sig_meetings/glasgow2004_presentations/lorna.uklomcore.ppt](http://metadata.cetis.ac.uk/sig_meetings/glasgow2004_presentations/lorna.uklomcore.ppt), [accessed 15.06.2006]. [power point presentation].


