Requirements Elicitation towards a Search Engine for Semantic Multimedia Content

Lydia Weiland
University of Mannheim
Mannheim, Germany
lydia@informatik.uni-mannheim.de

Felix Hanser
University of Mannheim
Mannheim, Germany
fehanser@mail.uni-mannheim.de

Ansgar Scherp
Leibniz Information Center for Economics
Kiel, Germany
a.scherp@zbw.eu

Abstract—We investigate user requirements regarding the interface design for a semantic multimedia search and retrieval based on a prototypical implementation of a search engine for multimedia content on the web. Thus, unlike existing image search engines and video search engines, we are interested in *true* multimedia content combining different media assets into multimedia documents like PowerPoint presentations and Flash files. In a user study with 20 participants, we conducted a formative evaluation based on the think-aloud method and semi-structured interviews in order to obtain requirements to a future web search engine for multimedia content. The interviews are complemented by a paper-and-pencil questionnaire to obtain quantitative information and present mockups demonstrating the user interface of a future multimedia search and retrieval engine.

Index Terms—Multimedia Search, Measurement, Semantic Integration, User Interface Design.

I. INTRODUCTION

Multimedia content provided by PowerPoint presentations or Flash documents are widely adopted and can be found in any domain. Despite the growing interest in multimedia web search, most research on search engines is dealing with single media types like images and videos [1], [2]. Regarding search and retrieval of *true* multimedia content, we find a gap in research. Thus, we have developed an early prototype of an engine for searching and exploring multimedia content shown in Fig. 1 [3] It supports keyword-based queries and filtering the results by audio, video, animation and interaction. The thumbnails are of different size, starting with larger ones in the upper position to smaller ones in the lower position of the results page, where pagination is used to limit the result list. Hovering over a thumbnail shows an animated preview of the document. Clicking on a document opens a detail view. The prototype has about 4000 multimedia documents in the database. On the basis of this prototype, we conducted a user study evaluating the features for a future multimedia search engine. The goal of study was to use this system to bootstrap requirement elicitation and detailed understanding of users' needs for semantic multimedia search.

II. RELATED WORK

Various *media retrieval systems* have been developed in the past like the MEMORAe project [4], where ontological knowledge is used for indexing and searching educational videos. Breaking the barrier of a single media modality,
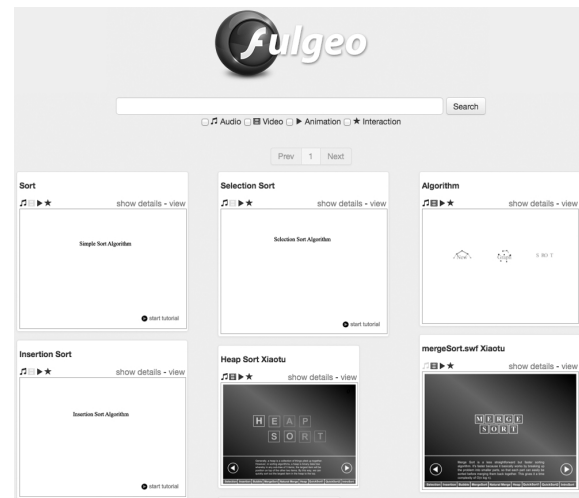

Fig. 1. fulgeo search interface and result view

there are approaches for semantic cross-media search and retrieval like the semantic search engine Squiggle [5] for images and audio. The FLAME framework (Flash Access and Management Environment) [6] is considered to be the so-far most comprehensive work on *multimedia search*. It supports retrieval based on some spatial and simple interaction constraints. Regarding the use of media retrieval systems, there have been some *empirical investigations* conducted in the past. Hearst [7] states that there are three main search behaviors in web search: fact finding (looking for specific facts or pieces of information), information gathering (the collection of information from multiple sources), and browsing (visiting web pages without particular goal). Kofler and Lux [2] conducted an evaluation of user intentions within image search. Maniu et al. [1] analyzed web server logs and all user actions during search sessions. Both conclude that current taxonomies, models and thus, interfaces, do not exactly represent the user's intent while searching for multimedia content.

III. USER STUDY

We conducted a user study with 20 subjects (eight female) with an average age of 26 years (SD=2.87), ranging from 22 and 34 years. Special consideration was given to the subjects'

diversity regarding educational background and profession in order to avoid bias [8]. The multimedia documents cover terms related to the topic climate change and global warming. Every subject had to solve four tasks of finding specific multimedia documents. For solving the tasks, the subjects had to use the different features offered by the engine. The tasks were motivated by a task scenario, wherein the subject is supposed to explain climate change to pupils using multimedia. We roughly explained the system and asked the subjects to write down their first impressions of the prototypical multimedia search engine. While then conducting the tasks, the subjects were encouraged to think aloud. After that, the subjects were asked to fill in a paper-and-pencil questionnaire. Here, a set of 19 closed-questions taken from the IsoMetrics [9] questionnaire were chosen and adapted to assess the prototype. The subjects answered the questions on a 7-point Likert scale. In addition, a semi-structured interview was conducted in order to explore emergent meanings and intentions of the subjects in context [10].

IV. RESULTS

The mean duration for all sessions was 39 minutes (SD = 9 minutes). All subjects were able to successfully accomplish their tasks. From the IsoMetrics, we focused on questions regarding the suitability for executing the tasks (T.1-T.2), suitability for learning the application (L.1-L.4), and controlability (C.1-C.2). In addition, we have asked specific questions regarding search and exploration of multimedia content. These features are about filtering by medium type (F.1), thumbnail preview of multimedia documents (F.2), the detailed view of a selected document (F.3), and the ranking by size to support the prediction of relevance (F.4). Most subjects predominantly agree to the statement L.1: "The interface of the search engine is understand-able at first glance." (M = 6.2; SD = 1.12)(cf. Fig. 3). Also most of the subjects predominantly agree to L.2: "The search engine is designed in such a way, that functionality not yet known could be learned by trying out." (M = 6.15; SD = 0.91) and to L.4: "I don't have to remember a lot of details to operate the search engine." (M = 6.35; SD = 0.96). A higher deviation and lower agreement is received for L.3: "It did not take long time before I learned to operate the search engine." (M = 5.75; SD = 1.68). Overall, the interface is easy to understand, but there are still some features which require learning or explanation.
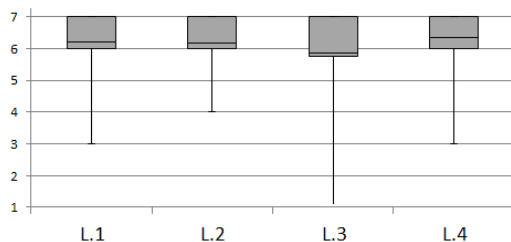
Fig. 2.   User ratings regarding learnability.

The results for the L-items are supported by C.1: "Handling the multimedia search engine is easy." (M = 5.8; SD = 0.98) and C.2: "The engine can only be used in a rigid way." (M = 4.05; SD = 1.66). T.1: "The search engine permits to enter queries just the way it is necessary for searching multimedia content." (M = 5.7; SD = 1.35) confirms the initial keyword as approach for searching multimedia content. Nevertheless, representing results need some rework: T.2: "The results found by the search engine match my queries."(M = 4.85; SD = 1.46).
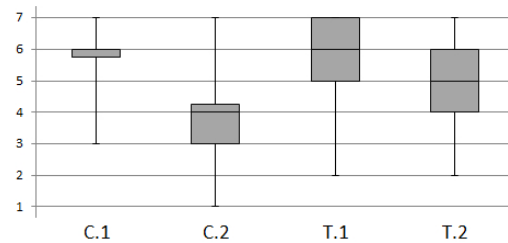
Fig. 3.   Ratings on controlability and suitability.

The median values regarding the specific features range between 4 and 6 (Fig. 4), but the detail view needs improvement.
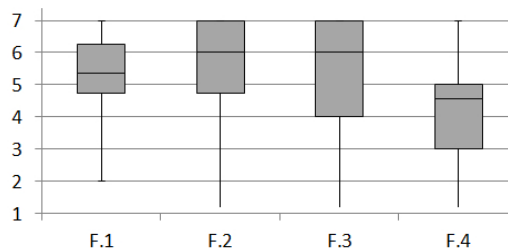
Fig. 4.   Ratings of multimedia search features.

V. REQUIREMENTS ELICITATION

The user feedback obtained in the semi-structured interviews was analyzed in a bottom-up process and concepts were obtained from the detailed analysis of the collected information (codification) [11]. In addition, we analyzed the results of the questionnaire and studied existing literature. From this, we derive the following requirements for a future multimedia search engine.

Search Bar: The search interface of a multimedia search engine should offer an error tolerant autocompletion. Queries should be easy to formulate, adapt and change according to the user's needs and preferences [12]. More filters, which cover metadata, like author or date of creation, should be offered. To avoid an overload of the interface these additional filters can be hidden in an advanced search interface.

Results Page: Two possibilities to style the result page were suggested: (1) Arrange results strict from top to bottom and representing the relevance from top to bottom in ascending order, without the representation of relevance by thumbnail size. (2) Results are presented in a grid layout with the same thumbnail sizes, but users can arrange their appearance on

basis of adjustable relevance options or rearrange via drag-and-drop.

Thumbnail View: The aim of snippets is to support the user in predicting the relevance of result items. Therefore, the representation should be clear, self-explanatory, and provide enough information. The thumbnail view should provide icons for filters, which are more visible and, where the function, if a certain media type is included or not, is clear. As known from other search engines, the headlines should provide a link to the URL of the original document. Besides providing a link, a download button might also be useful. These requirements are supported by half of the subjects: they nearly ignored the icons for the different filters. As there is a difference in "show-details" and "view", five users suggested to state the difference between those more apparently.

Detail View: The suggestion for the detail view is to enable the user to scroll through the whole text, search for further keywords within that document and to use the highlighted keywords as markers for navigating between them. When a user clicks on the highlighted keyword and then jumps to the next highlighted keyword, the preview image of the document will also change according to the text. Customization by resizing or moving the detail view window should be supported. Also adding buttons for navigating between detail views of several multimedia documents is recommended.

Request for New Features: Among the different requests were browser-based features like a top menu bar, e.g., to save search sessions, a back-button, a right-click menu, should be added. Interesting was also the comment of three subjects who liked to customize the results by drag and drop the results order.

VI. Search Engine Mockup

Based on the user study and the derived requirements (cf. Section III, IV, and V), we created mockups for a future search engine for multimedia content, which will then also consider the semantic integration of time, space and, interaction. Overall we can state, that there are no requirements for totally new functions.

The mockup provides an overview of the result list and presents additional information for each single result such as an advanced search, more filters, customization with choosing amount of shown data and thumbnails sizes and menu options like storing the current status (cf. Fig 5). The search by media type supports image, audio, and video. Additional filters for animation and interaction are added and explained when hovering over. The option for an advanced search is included in a dropdown list next to the textbox for entering the query. The advanced search enables to search for media in context, e.g. music, science, maps, or file type. The interface enables the users to customize the presentation of the result items, such as: changing scalability of the preview thumbnail and the amount of metadata by a slide control in line with the textbox.

The results page enables users to view the result list by several aspects such as the overall relevance represented by
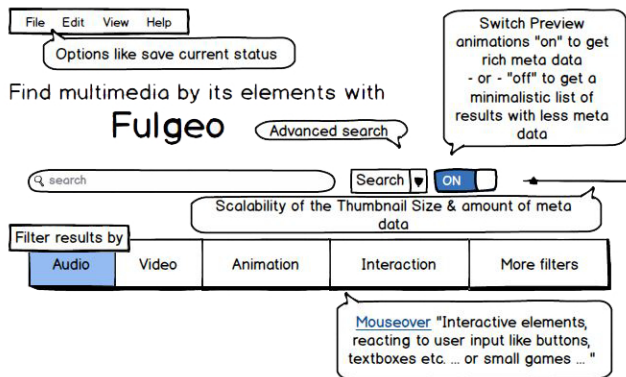

Fig. 5.   Mockup for the multimedia searchbar.

the ranking position of the items, explicit headlines with a short and contextualizing sub-line (approx. snippets), a preview animation by mouse-over, metadata based on media type specifications, e.g., file size and duration, and metadata of common web content specifications, e.g., publisher, URL and release date (cf.Fig. 6). The result list can be scrolled down to (in principle) infinity. Thus, no pagination is needed for users. The result view is divided into two parts: A vertical bar splits the list with the thumbnails of all relevant results from the detail view (cf. Fig. 7). Using drag and drop one can move a document from the overview side to the detail view side. Alternatively, a user can also use the "Show details" button below a thumbnail.
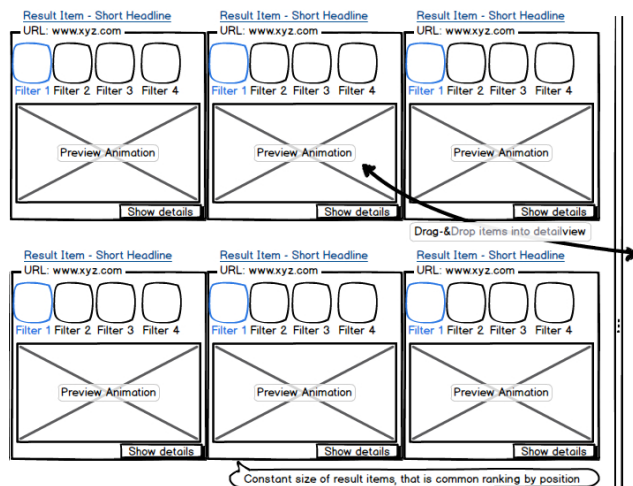

Fig. 6.   Mockup of the result list view.

The preview text is scrollable as most users expected it to be rich in quality. When the users scroll through the preview text, the animation of the presentation changes, too. Likewise, when the "play"-button is pressed to render the presentation also the text below changes. This allows to search in the whole multimedia document and has a strong focus on the visual media types. The affiliated text, which changes accordingly,

supports the search for visual media types and represents the relation of visual media types and text. A user is able to navigate through the result list via the detail view by clicking on the previous or next item buttons (cf. Fig. 7).
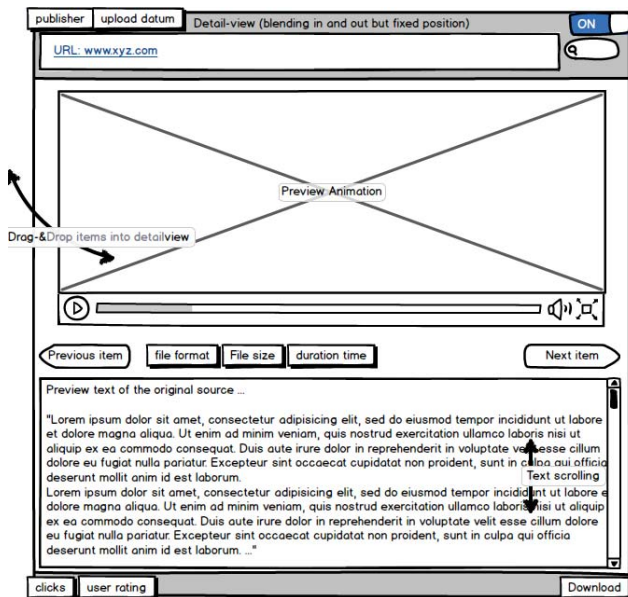

Fig. 7. Mockup of the details view.

VII. DISCUSSION AND LIMITATIONS

Even though, the ergonomic quality of the prototype as assessed by users can be considered quite high, most users prefer a straight structure like a simple grid-layout of the results. This is in particular important as naturally for a multimedia search engine the result set will be full of colorful animations.

The subjects made clear that they did not want to have advanced features like query-by-example. This finding is in line with the research by Jaimes [13] who states that in many real world applications it is hard to find an example to describe the user's information need.

If tags are shown next to the documents, which describe categories, e. g., persons or events, the user's assessment about the relevance of a document can be supported. This approach is inspired by Voxalead[1]. Altogether, the use of user generated annotations, comments, ratings, colorized tags etc. should be encouraged to enrich the metadata and include relevance feedback for the multimedia search [13], [14], [15].

The fact, that models in search engines often do not fit to the intentions of users [2], [1], implicates a need for more psychological research to build models of human activity during multimedia information search and retrieval, e. g., connected to common dual channel theories [16] or concrete versus abstract thinking [17]. The tendency to request customization of the search engine suggests that there is a higher level of perceptual

[1]http://voxaleadnews.labs.exalead.com/, last access: 06/12/2013

gap between content and current representation of content in contrast to personal requirements of users regarding the representation of search results to estimate relevance of results [13]. Despite the results obtained from the study, the evaluation also needs to be seen in context of the methodological limitations based on self-report using a standardized questionnaire.

REFERENCES

[1] S. Maniu, N. O'Hare, L. M. Aiello, L. Chiarandini, and A. Jaimes, "Search behaviour on photo sharing platforms," in *ICME*, 2013, pp. 1–6.
[2] C. Kofler and M. Lux, "An exploratory study on the explicitness of user intentions in digital photo retrieval." in *I-KNOW*, Sep. 2009.
[3] J. Tingvold, D. Stohr, D. Schneider, and A. B. Amundsen, "Interactive multimedia search and exploration," Tech. Rep., 2013. [Online]. Available: https://ub-madoc.bib.uni-mannheim.de/34753/
[4] G. Merzougui, M. Djoudi, and A. Behaz, "Conception and use of ontologies for indexing and searching by semantic contents of video courses," *FJCSI*, vol. 8, no. 3, 2012.
[5] I. Celino, E. D. Valle, D. Cerizza, and A. Turati, "Squiggle: a semantic search engine for indexing and retrieval of multimedia content," in *SEMPS*. CEUR-WS.org, 2006.
[6] J. Yang, Q. Li, L. Wenyin, and Y. Zhuang, "Content-based retrieval of FlashTM movies: research issues, generic framework, and future directions," *Multimedia Tools Appl.*, vol. 34, no. 1, pp. 1–23, 2007.
[7] M. A. Hearst, *Search User Interfaces*. Cambridge University Press, 2009.
[8] M. Miles and A. Huberman, *Qualitative Data Analysis: An Expanded Sourcebook*. Sage, 1994.
[9] G. Gediga and K.-C. Hamborg, "Isometrics: An usability inventory supporting summative and formative evaluation of software systems." in *HCI (1)*. Lawrence Erlbaum, 1999.
[10] D. M. Myers, *Qualitative research in business & management*, 1st ed. Sage, 2009.
[11] M.-L. Young, F.-Y. Kuo, and M. D. Myers, "To share or not to share: a critical research perspective on knowledge management systems." *EJIS*, vol. 21, no. 5, 2012.
[12] P. Morville and J. Callender, *Search Patterns: Design for Discovery*, 1st ed. O'Reilly Media, Inc., 2010.
[13] A. Jaimes, M. Christel, S. Gilles, R. Sarukkai, and W.-Y. Ma, "Multimedia information retrieval: what is it, and why isn't anyone using it?" in *MIR*. ACM, 2005.
[14] D. Tjondronegoro, Y.-P. P. Chen, and A. Joly, "A scalable and extensible segment-event-object-based sports video retrieval system." *TOMCCAP*, vol. 4, no. 2, 2008.
[15] J.-C. Wang, Y.-C. Shih, M.-S. Wu, H.-M. Wang, and S.-K. Jeng, "Colorizing tags in tag cloud: a novel query-by-tag music search system," in *MULTIMEDIA*. ACM, 2011.
[16] D. Kahneman, *Thinking, fast and slow*. Farrar, Straus and Giroux, 2011.
[17] Y. Trope and N. Liberman, "Construal-level theory of psychological distance," vol. 117, pp. 440–463+, 2010.