

RSLIS/AAU at CHiC 2013

Mette Skov¹, Toine Bogers², Haakon Lund², Maj Lauge Ward Jensen², Erik Wistrup² and Birger Larsen²

¹ Aalborg University, Department of Communication and Psychology
Nyhavnsgade 14, 9000 Aalborg, Denmark

² Royal School of Library and Information Science, University of Copenhagen
Birketinget 6, 2300 Copenhagen, Denmark

skov@hum.aau.dk, {tb,hl,blar}@iva.dk, {k09maje,k09erwi}@stud.iva.dk

Abstract. In this paper, we describe our participation in the interactive and ad-hoc tracks of CHiC 2013.

Keywords: Interactive IR, Europeana, cultural heritage, digital humanities, information retrieval, evaluation

1 Introduction

In this paper, we describe our participation in interactive and ad-hoc tracks of the Cultural Heritage at CLEF lab organized at CLEF 2013. The structure of this paper is as follows. We start in Section 2 by describing our participation in the CHiC Interactive track. Our participation in the CHiC Ad-hoc track is described in Section 3. We provide general conclusions to our results in Section 4.

2 CHiC Interactive track

This section describes a combined qualitative and quantitative study aiming at providing insight into users' non-intentional interaction with digital cultural heritage resources – in this case records from the European Digital Library (EDL). The following research questions guided this part of the study:

RQ1: What motivate participants' interaction with the EDL?

RQ2: How did participants react to the open and non-intentional work task situation?

RQ3: What grabbed the attention of participants in the browsing-based interface?

The first research question is motivated by numerous calls (e.g., Ingwersen & Järvelin, 2005) stressing the importance of including context dimensions in interactive information retrieval (IIR) together with a continued interest in understanding the motivation of museum audiences both online (e.g., Fantoni, Stein & Bowman (2012):

Goldman & Schaller, 2008) and in the physical museum (e.g., Ellenbogen, Falk & Goldman, 2008). Here focus is on exploring user study participants' motivation at two different levels. Firstly, we look at participants' motivation for visiting EDL compared to motivations for visiting museums in general. Secondly, we study what motivates participants' interaction with the EDL at session level.

The second research question concerns methodological aspects of using simulated work task situations. The concept of simulated work task situations is a component in a framework for IIR systems evaluation (e.g., Borlund, 2000). Since then simulated work task situations have been applied in numerous IIR studies (see review in Borlund & Schneider, 2010). However, only a limited number of studies have addressed task structure (exceptions include Toms et al. (2007)). Accordingly it is relevant to study how participants react to the non-intentional and open work task situation used in the present study. Finally, we investigate what parts of the browsing-based interface that grabbed the attention of participants, and how they switched between interface components. This is motivated by the fact that little is known about user behaviour in relation to this type of interfaces.

2.1 Methodology

Data from the online questionnaire was used to answer the two research questions in this section. We used data from a total of 200 respondents: 160 online respondents and 40 in-lab respondents (10 from Sheffield and 30 from Denmark). It should be noted that the questionnaire answered by online and in-lab participants was identical. We focused on the part of the questionnaire on participants' engagement, experience, and motivation. Three follow-up questions were asked to 10 in-lab participants (participant numbers 531, 551, 554, 558, 560, 565, 577, 578, 582, and 583) in order to further explore their motivations and search experience:

1. What motivated your interaction with the EDL?
2. Why did you add the chosen items to the Bookbag?
3. How realistic did you find this exploratory task? (Very realistic, partly realistic, not at all realistic)

The three follow-up questions expanded on the online survey responses and thus provided additional qualitative data.

The questionnaire data was analysed using mainly descriptive statistics. In addition chi-squared tests were used to calculate the statistical association between variables of experience and engagement (significance level α is set to 0.05). It was a hypothesis to find an association between a high level of experience with European culture and heritage and a high level of engagement. For example *if* a user is very interested in reading and examining things about European culture and heritage or frequently visits museums or galleries, *then* the user is expected to be very interested in the exploration task and to be absorbed in exploring etc.

The following variables were tested:

- 3 variables of experience:
 - o How often do you visit museums or art galleries, either in person or on the web?

- How familiar are you with European culture and heritage?
- How interested are you in reading and examining things about European culture and heritage?
- 18 variables of engagement:
 - Engagement related to endurability (5 variables)
 - Engagement related to focused attention (7 variables)
 - Engagement related to finding involvement (3 variables)
 - Engagement related to novelty (3 variables)

Hypothesis testing did not include the experience variable on how often EDL is searched since the majority of test participants (81 %) has never used the EDL. Further, engagement related to aesthetics variables and perceived usefulness variables was considered out of scope in relation to the research questions. Before chi-squared tests were calculated answer choices (agree – disagree) related to the engagement variables were reduced from 5 to 3 categories to avoid too small subgroups.

Finally, a qualitative categorization of free text answers to the question “can you elaborate on why you looked for these particular objects?” was mapped to 7 categories of motivation. Participants were asked to answer this question for each of the objects added to the Bookbag. A total of 291 answers were identified. If a participant had specifically explained his motivations for looking at, for example, 3 objects then this answer was split into 3 parts. Out of the total 291 answers 222 answers were qualitatively analysed. The remaining 69 answers were omitted because they did not relate to user motivation or gave too little information (“no”, “out of interest”, “I did not use the function”, “I liked them” etc.). A literature review by Goldman and Schaller (2004) served as starting point for developing the categories of motivation in the present study. They characterize the most common motivations from museum web site visits as:

1. Gathering information for an upcoming visit to the physical museum
2. Engaging in very casual browsing
3. Self-motivated research for specific content information
4. Assigned research (for job or study) for specific content information

Lately, a fifth motive has been added by Fantoni, Stein and Bowman (2012):

5. Make a transaction on the web site

The 5 categories of motivations were not directly applicable to the present study. Firstly, based on the free text answers to why objects were added to the Bookbag it was not possible to distinguish between casual browsing and research for specific content information (motivation number 2 and 3 above). Secondly, the fifth motivation category is not relevant to this study. Instead the above categories served as a starting point and were expanded based on a bottom-up approach (see the 7 categories in Table 2). In this way we aim to answer a call by Ellenbogen et al. (2008). They suggest that motivations of visitors to museum web sites differ significantly from the motivations of visitors to physical museums. Therefore they call for further studies to elaborate our understanding of online visitors’ motivation. They especially point to the importance of identity-related motivations, which has

proven to strongly influence the learning and behaviour of physical museum visitors (Ellenbogen et al, 2008, p. 193).

To answer research question 3 we tracked the gaze behaviour and mouse clicks of a subset of 10 of the participants from Denmark using a binocular RED-M remote eye tracker from SensoMotoric Instruments (SMI) set at a sampling rate of 120Hz. The eye tracker allows for a head movement of approximately 32 by 21 cm at a distance from the monitor of 60 cm. According to the producer the accuracy is 0.5 degrees. The study was executed at the Royal School of Library and Information Science, University of Copenhagen in a quiet setting. The test leaders were present during tests. We analysed the eye tracking results by defining a number of Areas of Interest (AOIs) following the logical division of the CHIC experimental interface. Furthermore we did a qualitative inspection of the gaze patterns as well as extraction of data describing key metrics as time on AOIs and number of fixations on AOIs.

2.3 Results

2.3.1 Motivation and reaction to open-ended work tasks

As described earlier the first research question on user motivation was studied at two different levels. Firstly, we look at motivation at a general level. Table 1 compares frequency of visits to the EDL to frequency of visits to museums and art galleries. Table 1 shows that the majority (81 %) of the respondents has never visited the EDL and 15.5 % of the respondents visit EDL less frequently than a few times per month. In contrast only a single respondent (0.5 %) has never visited a museum or art gallery and 81 % visit museums less frequently than few times per month. Further, Table 1 compares the main reason for visiting the EDL and museums in general respectively. Table 1 shows that personal interest and enjoyment is the main reason for visiting museums and art galleries (69.1 %), whereas participants' motivation for visiting the EDL is more sided with personal interest, research for study and for work as top-3.

Table 1. Frequency of visit to and motivation for visiting the EDL and museums and galleries (n = 200)

	Frequency of visit (%)			Motivation for visiting (%)	
	EDL (n = 200)	Museums and gal- leries (n = 200)		EDL (n = 49)	Museums and galleries (n = 269)
A few times a week	0.5	1.5	To follow up on an in-person visit	0.0	6.3
A few times a month	3	17	For personal interest and enjoyment	36.7	69.1
Less frequently	15.5	81	To prepare for a visit to a museum in person	6.1	9.3
Never	81	0.5	To do research for studies	36.7	8.6
			To do research for work	20.4	3.7
			Other	0.0	3.0

Secondly, we look at motivation at session level. Table 2 shows a categorization of questionnaire respondents' answers to the following question when their chosen Bookbag objects were displayed: "Can you elaborate on why you looked for these particular objects?". The top three motivations (see Table 2) relate to personal interest, followed by motivations related to the participant's own country or family etc., and motivations based on an aesthetic or visual experience. Table 2 shows examples of the different categories of motivation.

Table 2. Categorization of participants' motivation: why did participants' look for particular objects and add them to Bookbag?

	User motivation related to:	n	%	Examples of user motivation:
A	Personal interest	91	41.0	"An interest in Engineering and Technology history." "I like very much Titian and I consider this painting one of his master pieces".
B	My... family/university/city/ country/...	33	14.9	"Minery is a very important activity in my country so the Bilston Glen coal mine picture represents this interest" "I live in Cornwall, and so was interested to see the areas I know as they used to look".
C	Aesthetics or visual experience	30	13.5	"Aesthetically pleasing - nice hues." "While exploring, I saw the first one and it just looked beautiful so I saved it".
D	Curiosity, imagination, serendipity, surprise, etc.	24	10.8	"I was just browsing and found these, and they happened to catch my attention". "Hunting for old objects is interesting"
E	Place that I have visited/are going to visit	20	9.0	"The second image because slate mines in Wales look like this when it rains and it reminded me of a happy but very wet holiday". "The images of India I chose because they reminded me of my trip to India."
F	Study or work related interest	17	7.7	"My research and teaching area is the early medieval period, and I am always on the lookout for useful images/resources..." "Potential topic for an essay"
G	Other	7	3.2	"I wanted to show my son" "I am a Bowie fan and thinking of going to the exhibition in London that is on at the moment so wondered if there were any good pictures of him".
	Total	222	100.1	

The second research question concerns how users react to the open and exploratory work task situation. To answer this question we first look at participants' level of agreement with 18 statements about exploring the EDL website (see Figure 1). The 18 statements on engagement relate to endurance, focused attention, involvement, and novelty. It is interesting that the 3 variables that participants disagree with the most are all related to focused attention (When I was exploring, I lost track of the world around me; I was so involved in this experience that I lost track of time; I lost myself in this experience). In contrast, the 3 most agreeable statements relate to novelty (I continued to explore this website out of curiosity; The content of the website incited my curiosity; I felt interested in my exploration task) together with a statement: "This experience did not work out as I had planned".

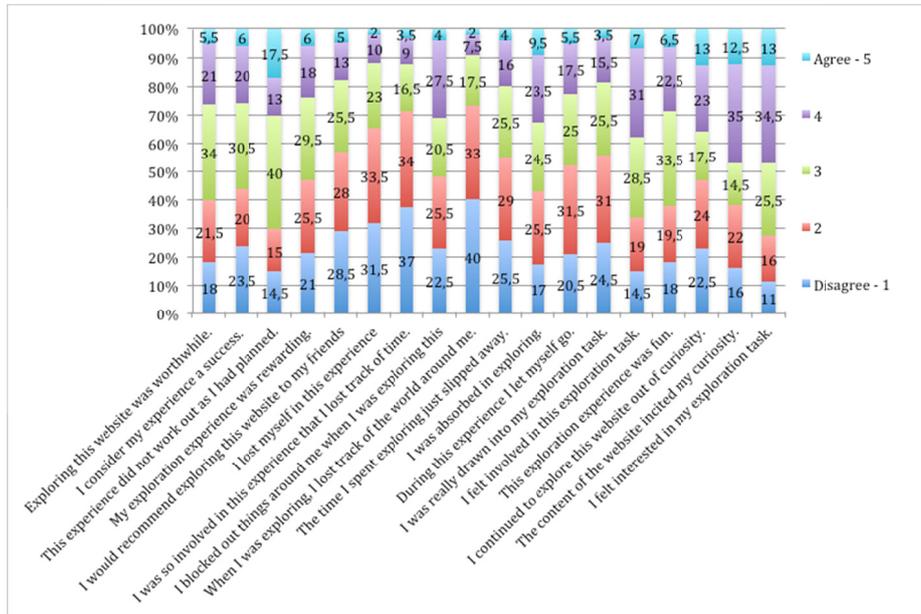


Figure 1. Participants' level of agreement with statements about exploring the EDL website. The statements relate to 18 variables of engagement (n = 200).

Next we tested the statistical association (chi-squared tests) between the 18 variables of engagement and participants' 1) level of familiarity with European culture and heritage, 2) level of interest in reading and examining about European culture and heritage, and 3) how often they visit museums or art galleries. In general results of the 54 chi-squared tests show only few statistical significant associations between the tested variables. However, some few, interesting associations were found. Not surprisingly participants' level of interest in reading and examining about European culture and heritage showed a strong statistical association with the following two engagement variables related to novelty:

- The content of the website incited my curiosity (p-value = 0.001)
- I felt interested in my exploration task (p-value = 0.048)

A weaker association was found with two variables related to focused attention:

- I blocked out things around me when I was exploring this website (p-value = 0.068)
- I was absorbed in exploring (p-value = 0.082)

In addition two weaker associations were found between participants' level of familiarity with European culture and heritage and the following variables of endurance and novelty:

- I blocked out things around me when I was exploring this website (p-value = 0.054).
- The content of the website incited my curiosity (p-value = 0.125)

No statistical significant associations were found between participants' frequency of visits to museums and their agreement with the 18 engagement variables.

Finally, answers to one of the three follow-up questions (How realistic did you find this exploratory search task?) also provided insight into how participants' experience with the non-intentional task. The follow-up questions were asked to 10 in-lab participants. 5 participants found the task highly realistic, 1 participant found it partial realistic, and 4 participants found the non-intentional task unrealistic.

2.3.2 Analysis

Understanding user motivation is a key variable in understanding their experiences and interaction online (Fantoni et al., 2012). The results from the questionnaire about participants' motivation for visiting the EDL (see Table 1) show that user study participants rarely use the EDL to follow up on or prepare for an in-person visit to a museum. Likewise, only one single user statement (out of 222) on motivations concerns visit to a physical museum or art gallery (see the Bowie-fan example in Table 2). This contrasts with previous research (e.g., Fantoni et al., 2012) where trip planning is the primary motivation for visiting a site. A possible explanation to this is that the EDL portal covers collections across multiple cultural heritage institutions and is thus not closely linked to in-person visits.

The categorization of participants' motivations based on a bottom-up approach resulted in an elaborated categorization. Table 2 shows that 'Personal interest' is by far the most frequently identified motive. Motives related to 'My...family/university/city/country' and 'Places that I have visited/are going to visit' can be seen as sub-categories to personal interest. However, earlier research (Ellenbogen, Falk & Goldman, 2008; Falk, 2009) stresses the importance of uncovering identity-related motivations for visiting museums and other cultural heritage organizations. As such the sub-categories provide an additional level of information and can to some extent be mapped to identity-related motivations. For example, user statements representing an aesthetic or visual experience can reflect a *recharger* (using Falk's (2009) terminology) motivated by the yearning to emotionally and intellectually recharge in a beautiful and refreshing environment. Likewise, an *explorer's* (again using Falk's (2009) terminology) interaction is driven by a need to satisfy personal curiosity and interest in a challenging environment. For example illustrated by the example "Hunting for old objects is interesting" in Table 2. The research design of the present study does not fully support identifying identity-related motivations and it could be interesting to further explore in future research.

The second part of the study addresses how participants reacted to the open and non-intentional work task situation given. Firstly, we look at the positive reactions:

5 out of 10 in-lab participants answered that the task was highly realistic. This is supported by participants' high level of agreement with variables on "The content of the website incited my curiosity" and "I felt interested in my exploration task" (see

Figure 1). Further, it is inspiring to see the huge variety in patterns and directions of user interaction.

Secondly, looking at the challenges the high level of disagreement with 3 variables related to focused attention (see section 2.3.1) indicates that participants are not fully absorbed in the non-intentional task. 4 of 10 in-lab participants found the task unrealistic. They explained that they lacked a search motivation and direction. A quote from a questionnaire response reflect this view: “I would have to be wanting to look and research a subject to want to use this website. It got boring within about 30 seconds due to no desire to research anything at the time”.

Future research design could include asking the users whether (s)he finds the non-intentional task realistic. In this way we can analyse whether this variable significantly influence the search experience. Further, given the high percentage of non-European users (81 %) in the present experiment, it would be very interesting to study if the search experience differs depending on whether real users or test persons are included.

2.3.3 Investigation of participant attention

We present the results as an analysis of sequence charts, average number of fixations over AOIs as well as a tiled view AOI analysis.

The sequence charts such as the sample one shown below in Figure 2 shows the order in which users inspect individual Areas of Interest (AOIs). We defined six AOIs according to the logical division of the CHIC user interface. Overall the sequence charts show a common pattern between users and how they navigate the CHIC interface. The first AOI accessed is the *assignment* given to the participant followed closely by an inspection of the *result* AOI and the *category browser*. The *item detail* AOI, *book bag* AOI and *search box* AOI tended to be visited more frequently later in the sessions when participants have chosen an item for closer inspection, or in the case of the *search box*, trying to locate items by searching after the have tried using the *category browser*.

The sequence charts also show how participants jump between the AOIs or revisit them over the session. The *result* AOI is the one showing the highest number of visits (and revisits) and also the AOI were most time is spent followed by the *item detail* AOI and *category browser* AOI. Users also return to the *assignment* AOI at various intervals during the test. Table 3 shows the average number of fixations for the 10 participants distributed over AOIs. We see that by far the most time is spent on *results* followed by *item detail*.

Table 3. The average number of fixations distributed over AOIs for the 10 eyetracked participants.

	Fixations (avg.)
book bag	55,8
item details	443,3
results	1320,8
search box	45,2
category browser	150,5
assignment	37,4

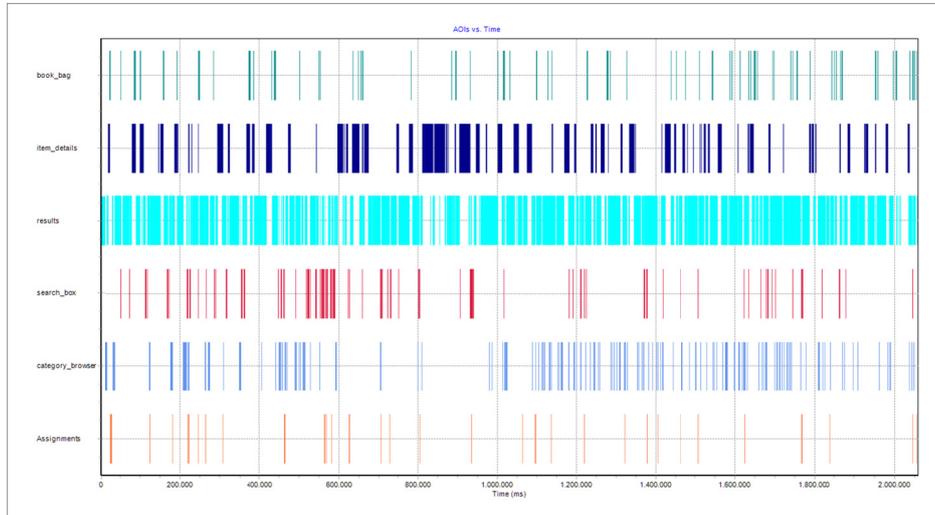


Figure 2. Sample AOI sequence chart. The chart shows the distribution of visits over time. The X-axis is time in milliseconds (0 to 2,000,000 ms), and the Y-axis is the defined AOIs (top to bottom): 1) *book bag*, 2) *item details*, 3) *results*, 4) *search box*, 5) *category browser*, and 6) *assignment*.



Figure 3. Sample distribution of fixations using tiled view. Red indicates the highest number of fixation, and dark blue the lowest number of fixations. Diamonds indicate mouse clicks.

Figure 3 show a sample tile view overlay on the CHIC experimental interface. It indicates that the *results* and *item detail* in the middle of the interface receive a lot of attention - including the links to the next page of results.

3 CHiC Ad-hoc track

This section describes our participation in both the monolingual and multilingual tasks of the CHiC ad-hoc retrieval track.

3.1 Methodology

Monolingual retrieval

In all our monolingual retrieval experiments, we used the language modeling approach with Jelinek-Mercer (JM) smoothing as implemented in the Indri 5.1 toolkit. We set λ to 0.4 and did not perform stemming or stopword filtering. For each topic we retrieved up to 1000 documents. We indexed each language collection separately and ran the topic translations for each language against these 13 indexes in turn.

Multilingual retrieval

For our participation in the multilingual task, we explored two different approaches. In the first approach we ran topics in each language against all 13 indexes at the same time. For example, first we ran the English topics against all 13 indexes combined. Then we ran the French topics against all 13 indexes combined, and so on, until we had 13 different runs, one for each topic language. The rationale behind running monolingual topics against a multilingual index is that Europeana content is divided over the 13 different indexes by country of origin, not by the actual language used in the metadata descriptions. By running the topic set in a particular language against all 13 indexes at the same time instead of just one, it is possible that more relevant documents from the other language indexes will be retrieved. Each of these 13 runs were performed using Jelinek-Mercer smoothing with λ set to 0.4, no stemming, and no stopword filtering, similar to the monolingual runs.

For our second approach we investigated the benefits of *fusing* multiple retrieval runs into a single run. We specifically focus on *collection fusion*, where the results of one or more algorithms on *different* document collections are integrated into a single results list (Voorhees *et al.*, 1995). In our case, we fuse the 13 different monolingual runs together. As different retrieval runs can generate wildly different ranges of similarity values, so we apply normalization to each retrieval result to map the score into the range $[0, 1]$. We normalize the original retrieval scores $score_{original}$ using the maximum and minimum retrieval scores $score_{max}$ and $score_{min}$ according to the formula proposed by Lee (1997):

$$score_{norm} = \frac{score_{original} - score_{min}}{score_{max} - score_{min}}. \quad (1)$$

Over the past two decades, many different fusion methods for retrieval runs have been proposed. In our experiments, we restrict ourselves to two of the most effective unweighted combination methods proposed by Fox & Shaw (1994): CombSUM and CombMNZ. The CombSUM method fuses runs by taking the sum of similarity values for each document separately; the CombMNZ method does the same, but boosts this sum by the number of runs that actually retrieved the document. Because of a limit on

the number of runs that could be submitted, we fused together the English, French and German runs from the first approach to multilingual retrieval using both CombSUM and CombMNZ.

3.3 Results & Analysis

Table 4 below shows the results for the 13 different monolingual runs. The language for which we obtained the best performance was German with a MAP score of 0.0613. Other performances with relatively good performance are Dutch, Norwegian, and French. Furthermore, the results also show that most of the languages with smaller groups of native speakers, such as Finnish, Greek, Hungarian and Slovenian—with the exception of Norwegian—fare less well with MAP scores between 0.0030 and 0.0079. This could be due to smaller numbers of available relevant documents in Europeana for these languages. Overall, there does seem to be room for improvement: a MAP score of 0.0613 is not likely to translate into satisfactory search engine performance for real-world users of Europeana.

Table 4. Results of the monolingual runs for the 13 different Europeana languages.

Language	MAP
ENG (English)	0.0249
FIN (Finnish)	0.0030
FRE (French)	0.0301
GER (German)	0.0613
GRE (Greek)	0.0063
HUN (Hungarian)	0.0070
ITA (Italian)	0.0293
NLD (Dutch)	0.0422
NOR (Norwegian)	0.0369
POL (Polish)	0.0156
SLV (Slovenian)	0.0079
SPA (Spanish)	0.0221
SWE (Swedish)	0.0133

The results of our multilingual runs show improvements compared to the monolingual runs. Our three multilingual runs for English, French, and German, where we ran these three topic sets against the complete multilingual index achieved MAP scores of 0.0370, 0.403, and 0.679 respectively. These increases in performance ranging from 11.1% to 48.5% appear to confirm our hypothesis that collections in different languages also contain many documents in other languages.

Our two fusion runs with the CombSUM and CombMNZ methods show an additional increase in performance at 0.0836 and 0.0837 respectively. This shows that combining different retrieval runs indeed results in better performance. It is likely that combining runs for all 13 different languages would have resulted in additional performance increases.

4 Conclusions

We described the Royal School of Library and Information Science and Aalborg University participation in the CLEF 2013 CHIC interactive and ad hoc tracks.

References

- Borlund, P. (2000). Experimental components for the evaluation of interactive information retrieval systems. *Journal of Documentation*, 56(1), 71-90.
- Borlund, P. & Schneider, J.W. (2010). Reconsideration of the simulated work task situation: evaluation of information retrieval interaction. In: Belkin, J. & Kelly, D. (eds.). *IiX'10 Proceeding of the Third Symposium on Information Interaction in Context*, New Brunswick, NJ, USA. New York: AC, 155-164.
- Ellenbogen, K., Falk, J. & Goldman, K.H. (2008). Understanding the motivations of museum audiences. In P.F. Marty & K.B. Jones (Eds.) *Museum informatics: people, information, and technology in museums* (pp. 187-194). New York, NY: Routledge.
- Falk, J. (2009), *Identity and the museum visitor experience*. Walnut Creek, CA: Left Coast Press
- Fantoni, S.F., Stein, R. & Bowman, G. (2012). Exploring the relationship between visitor motivation and engagement in online museum audiences. In J. Trant and D. Bearman (Eds.). *Museums and the web 2012: proceedings*. Toronto: Archives & Museum Informatics.
- Ingwersen, P. & Järvelin, K. (2005). *The turn: Integration of information seeking and retrieval in context*. Dordrecht, The Netherlands : Springer Verlag.
- Toms, E.G., O'Brien, H., Mackenzie, T., Jordan, C., Freund, L., Toze, S., Dawe, E. & MacNutt, A. (2007). Task effects on interactive search: The query factor. In: Fuhr, N., Kamps, J., Lalmas, M. & Trotman, A. (eds.). *Proceedings of the 6th International Workshop of the Initiative for the Evaluation of XML Retrieval (INEX)*, Dagstuhl, Germany. Berlin/Heidelberg : Springer Verlag, 359-372.