On the General Principles of Human-Computer Information Retrieval

Vesna Gega[1] and Pece Mitrevski[2]

[1]University for Information Science and Technology "St. Paul the Apostle" Ohrid
vesna.gega@uist.edu.mk
[2] St. Clement Ohridski University, Faculty of Technical Sciences, Ivo Lola Ribar bb
7000 Bitola, Republic of Macedonia
pece.mitrevski@uklo.edu.mk

Abstract. Nowadays, human-computer information retrieval (HCIR) is one of the most researched approaches, which applies human computer interaction (HCI) in information retrieval (IR), especially for text retrieval. The involvement of human activity (i.e. intelligence) in the process of retrieval is an interesting and challenging aspect of HCIR, thus it takes an interactive character with the purpose to improve user's needs. A number of researchers and research groups have conducted different experiments in this area. Innovative techniques for determining relevant documents and navigating in a large and complex set of full document results have been proposed. A novel approach based on focused retrieval for interactive identification of relevant document's parts has brought many improvements in this field. A special emphasis was placed on retrieving passages and XML elements from structured documents. Numerous techniques for focused retrieval based on snippets, facets and relevance feedback have been presented recently, bringing advantages and a more realistic scenario for text retrieval, but also a lot of open issues that need to be explored. This paper surveys the fundamental characteristics of the HCIR, as well as the focused HCIR, such as using passages and XML elements as retrievable units from structured or pure textual documents, thus providing a basis for development of novel HCIR approaches for interactive text retrieval, as our main goal in future research.

Keywords: HCIR, information retrieval, passage, XML element, evaluation

1 Introduction

Information retrieval is researched paradigm more than a half century and it is related to searching information and calculating how those are likely to be relevant for a user's asked query [4], [21]. It is an automated process that includes: indexing, searching, ranking, retrieval and presentation of the relevant information, as shown in Fig. 1. Usually, the searching is realized from very long and complex electronically stored sets of information (digital knowledge), such as World Wide Web, document collections, libraries, and relational databases. The results in different format, ranked by their degree of relevance are presented to the user. The term results refer to documents containing text, images, videos, audios and their parts. The degree of relevance is obtained as a result of a process implementing variety of probabilistic and statistical functions [19], [21], [33-35], [39].

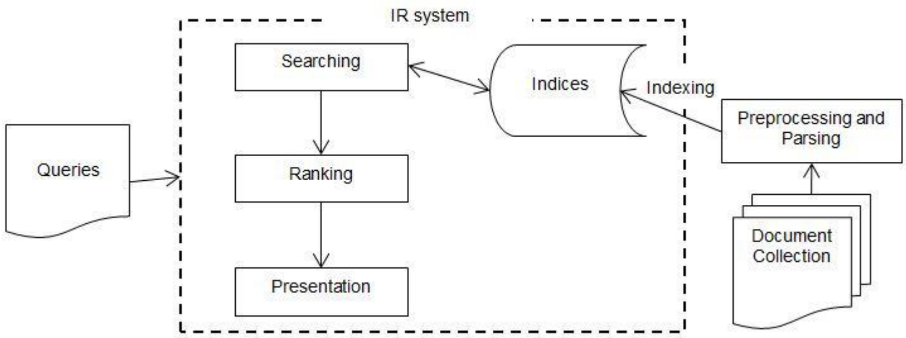

Fig. 1. IR process (inspired by: G. Marchionini, Toward Human Computer Interaction 2005 Lazerow Lecture, University of Washington)

Innovative techniques for information retrieval were developed through the years, tested on different large document collections, as part of English and non-English evaluation forums. One of the most challenging branches in the modern information retrieval is human-computer information retrieval (HCIR). In this paper we describe the progress of HCIR techniques, as well as focused HCIR techniques, for text retrieval in detail. We present the state of the art researches and explain future directions. This paper is organized as follows.

In Section 2, we survey the fundamental characteristics of HCIR, as well as various proposed techniques. The evaluation metrics used to measure effectiveness of HCIR systems are described in Section 3. Section 4 concludes the paper and shows future research directions.

2 Human-Computer Information Retrieval

An interesting and challenging aspect of HCIR is the involvement of human activity (HCI) in the process of retrieval (IR), thus it takes an interactive character and brings real scenario of searching, as shown in Fig. 2. With purpose to improve user's needs, the interaction continues until the user finds the right information of his/her interest by refining the originally proposed query [15]. HCIR concept has been introduced by *Marchionini G.* [24] for the first time between 2004 and 2006. Nowadays, it has evolved to stage where the user with action, perception and reflection is more involved in the relevant information seeking process [23], going beyond static queries and standard representation of results.

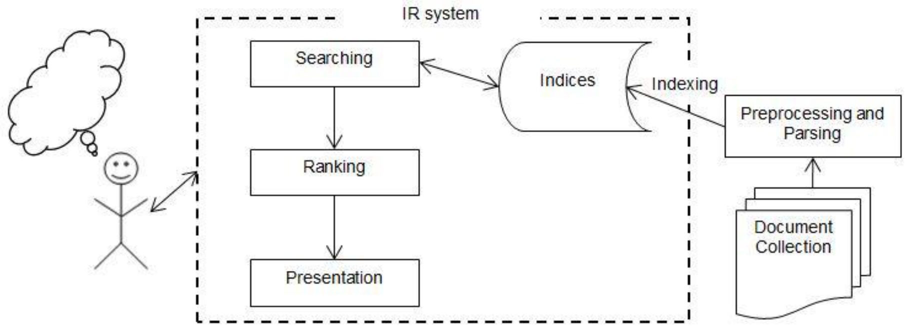

Fig. 2. HCIR process (inspired by: G. Marchionini, Toward Human Computer Interaction 2005 Lazerow Lecture, University of Washington)

2.1 HCIR Techniques

Intersection between HCI and IR is subject of many researchers and research groups, such as Workshop on Human-Computer Interaction and Information Retrieval [1], INEX[2], CLEF[3] and TREC[4]. New and powerful techniques for determining relevant documents and navigating in large and complex set of full document results have been proposed [22], [40]. A novelty approach based on interactive identification of relevant document's parts has brought many improvements in this field. A special emphasis was placed on retrieving passages and XML elements from structured documents, based on snippets, facets and relevance feedback/query expansion [5], [12]. All these techniques aim to effective and efficient interaction process without unnecessary costs of time, minimal mouse clicks or scrolling and context changing.

Faceted retrieval.

Given a query containing one or more terms, the information retrieval system returns an ordered list of ranked relevant results. Usually, that list is very long and user finds difficulties to navigate through it. Faceted search provides interactive hierarchical search and enables users to quickly determine the desired results. It means that user can browse in the categorized list of relevant results, refine the query and narrow down new different relevant results proposed by the system. The categories are also known as facets and the values they contain are facet-values [21]. The advantage of faceted search is the opportunity users to choose relevant results from those categories that are closest to their needs and way of thinking [3-5], [12]. The research challenge is to investigate, introduce and evaluate new effective and efficient approaches for facets and facet-values formulation.

[1] http://www.hcir.info/
[2] https://inex.mmci.uni-saarland.de/data/links.jsp
[3] www.clef-campaign.org
[4] trec.nist.gov

The Relational Browser called RB was for the first time introduced *Marchionini G. et al.* [22]. It is an approach that provides a list of relationships among attributes in an information space, which helps users in full text information seeking. For better and effective browsing, simple interactive information manipulation is implemented. It underwent several changes [40]. The authors have developed a novel interface, called RB++, where the faceted search plays the main role, allowing users explore relationships among facets and refine the original query by simple mouse actions on the facets and facet-values.

Ramirez G. [5] has experimented with large and highly structured document collection of IMDB[5] documents on Indri[6] search engine. Fixed set of facets were used, according to the collection DTD[7]. The facet-values were presented on hierarchical and non-hierarchical way. A non-hierarchical presentation means that all the facet-values are at the same level. As opposite, a hierarchical presentation is related to granularity, thus there are nested facet-values in three hierarchical levels.

Schuth A. et al. [5] also used the large and complex IMDB document collection. Limited numbers of facets covering not unique facet-values from a single document were defined. Each facet has name and an XPath[8] expression by which the facet-values can be retrieved. Two strategies for facets formulation were developed. The ranking of the facet-values in the both cases is based on hits that would be the result if the facet-values were selected.

There have been a number of studies that have examined faceted search on mobile web devices, such as that described by *Capra, R. et al.* [7]. Also different design patterns for mobile faceted search have been proposed[9]. Still there is an open question if faceted navigation is well-suited for the small screen.

Snippet retrieval.

Among the list of relevant returned results, there are also some irrelevant results that don't satisfy user's needs. Showing short summary of the result is another way for helping users to easily decide whether or not the result is relevant, without needing to view the result itself, because the summary is in more human-consumable form. The summary is called snippet. Usually, snippets are automatically generated from the result. They can be extracted as a block of text containing one or more sentences [6], [8] and [13], and a paragraph or several successive paragraphs from the result. In terms of size, the snippet can have fixed or variable length where the length is expressed as number of characters, words, lines or bytes. If the results have logical structure, the snippets can be formed with different kind of grouping elements. There are two types of snippet generation, static and dynamic [21]. The first one is not related to the query. The second one is query dependent and follows the users' needs [36], [38]. Also, metadata associated with the document as result can be used for snippet

[5] http://imdb.com
[6] http://www.lemurproject.org/indri/
[7] http://www.w3schools.com/dtd/default.asp
[8] http://www.w3schools.com/xpath/
[9] http://www.uxmatters.com/mt/archives/2010/04/

formulation. The research question is how to generate snippets to provide sufficient information for maximizing effectiveness and efficiency of the search process.

The preliminary results in [5] showed that poor snippets cause the users to miss over than 50% of relevant results. Also, the percentage of non-relevant documents that are correctly assessed was very high, indicating that snippets are useful in non-relevant documents determination.

The evaluation in [36] and [38] have shown that use of query dependent snippets brought many improvements in terms of speed, precision and recall than the use of query independent snippets.

Efficiency was main goal in [37], where also query dependent snippets were generated from results pages presented by search engines. The authors have proposed document compression method that significantly speeds up the snippets generation. Also, it was shown that snippets are generated faster if documents were previously cashed in the RAM.

Huang Y. et al. [14] presented eXtract, a system that efficiently generates snippets from XML structured documents. They identified that a good snippet should be a self-contained meaningful information unit of a small size that effectively summarizes the query result and differentiates it from others, according to which users can quickly assess the relevance of the query result. For that purpose, they have developed a new effective and efficient algorithm.

Leal L. et al. [5] experimented on Wikipedia[10] document collection and snippet generation was based on selecting highly ranked sentences, ranked according to the frequency of query terms. Also the title of the document was concatenated to the sentences in the snippets. Snippets had limited length of maximum 300 characters, depending to the document length. They expanded the original queries with query expansion technique, and main conclusion is that a large number of extra terms negatively influences on sentences selection.

Rongmei Li et al. [5] ranked the Wikipedia documents with the Language Model [39] and snippets are generated from the ranked list of relevant results. The snippets were term-extracted. Each term was weighed according its relative occurrence in the document and in the entire document collection. The top K scoring terms in two different clusters were chosen to form the snippet. One is a cluster containing only terms (words) that don't have any relationship between each other. Another is a cluster that contains semi-sentences related to the query.

Wang S. et al. [5] used Vector Space Model [19] for searching and ranking Wikipedia documents, based on structure and content. The authors also expand the query with query expansion techniques. For snippet generation they used query relevance, significant words, title/section-title relevance and tag weight to evaluate the relevance between sentences and query. The sentences with higher score formed the snippets.

Relevance feedback.

Query refinement is core process in relevance feedback task, realized in order to improve retrieval performances. Information retrieval system fully automatically or

[10] http://www.mpi-inf.mpg.de/departments/d5/software/inex/

with user's help can refine originally proposed query, depending of ranked results proposed by the system [1], [30]. There are three types of feedback: explicit, implicit and blind or "pseudo" feedback [21]. The explicit feedback is characterized by user's involvement in the retrieval process. It is an interactive process where user marks relevant and non-relevant returned results for his/her asked query and the system re-ranks unseen results and returns new list of results, closer to his/her interests. The implicit feedback is related to user's behavior, thus results that are more often viewed, scrolled, clicked, saved and printed are considered to be more relevant [18], [25]. The main difference between implicit and explicit feedback is that the user doesn't asses the results and give any addition effort, but he/she on indirect passive way influences on the feedback. The blind or "pseudo" feedback is a fully automated approach where information retrieval system simulates user's activity and does standard information retrieval. It retrieves a number of relevant results, and supposes that top K results of them are most relevant to the user's asked query, influencing on the feedback. Mechanisms for improving relevance feedback retrieval are interesting to explore.

Rocchio algorithm is method that implements relevance feedback using Vector Space Model [19], proposed by *Rocchio, J. J.* [29], around 1970. By query refinement an arbitrary percentage of relevant and non-relevant results tends to increase the information retrieval system's recall, and possibly precision. This scenario is optimal when the angle between the refined query vector and result vector is smaller, thus the similarity is maximized with relevant documents and minimized with non-relevant documents [21].

Kelly D. et al. [18] have continued the work of *Morita M. et al.* [25], experimenting with implicit feedback, in context of these hypotheses: Users will spend more time reading those documents that they find relevant, Users will scroll more often within those documents that they find relevant, and Users will interact more with those documents that they find relevant.

Lv Y., et al. [20] have proposed new effective pseudo relevance feedback approach, named positional relevance feedback, based on positions of query terms in feedback documents. Their experiments were conducted using full document feedback and passage feedback.

Allan J. [2] has developed a hybrid approach for relevance feedback retrieval from large documents. The idea was to use passages in relevance feedback task only where necessary. It means that full short documents in combination with passages from long documents in relevance feedback, produce for more effective retrieval.

Query reformulation may be a difficult and time consuming activity [21], thus fully automatic relevance feedback information systems were proposed and evaluated on [5], [21]. They have developed simulated exhaustive incremental user feedback [1]. Many proposed algorithms for focused retrieval based on passages and XML elements feedbacks went far beyond Rocchio feedback algorithm [5], [12].

3 On the Methodologies for HCIR System Evaluation: Issues and Prospects

Usually, the metrics for evaluating the effectiveness of a standard information retrieval system [27] use the widely accepted Cranfield methodology [10], including document collection, queries and relevance judgments in the evaluation process [21]. They are based on a set of measures, such as precision, precision @ K, recall, F-Measure, Mean Average Precision (MAP) and Discounted Cumulated Gain (DCG) [4], [9], [27].

In faceted search, the relevance of the data covered by facet-value can be measured with Normalized Discounted Cumulated Gain (NDCG) as a DCG variant for calculating relevance for all queries [5]. For the first time, NDCG was introduced by *Järvelin K. et al.* [16], and modified in two variants: normalized NDCG and recursive NDCG by *Schuth A. et al.* [32]. Also, Interaction cost is used to measure effectiveness of a faceted retrieval system, based on user-computer interaction evaluation in information seeking process [5]. It is measured as the number of results and facets that the user examined before he/she encounter the first relevant results. Usually, the interaction process is simulated in order to avoid expensive user study and make user study repeatable [17, 32].

There is a pallet of measures for snippet retrieval evaluation. *Overwijk A. et al.* [26] have employed measures based on the standard precision and recall, modified in terms on spans and nuggets. Comparison of relevance assessments based on whole documents vs. short snippets was introduced by *Bellot P. et al.* [5]. They have proposed a measure for calculating how effective snippets were in the retrieval process satisfying user's needs, called Mean Prediction Accuracy (MPA). Its averaged variant over all topics is called Mean normalized prediction accuracy (MNPA). Negative recall (NR) is the percentage of irrelevant documents correctly assessed, averaged over all topics. Also, *Bellot P. et al.* [5] have used Geometric Mean of recall and negative recall (GM) for snippet retrieval evaluation. Different approaches for evaluation of web search snippets at Yandex, a powerful Russian web search engine were surveyed and experimented by *Savenkov D. et al.* [31].

Ruthven I. et al. [30] gave a review of evaluation methods for relevance feedback, intended to measure the effect of feedback on the unseen relevant documents. The first one is Residual ranking, where the documents used in relevance feedback system, are removed from the document collection before evaluation. That includes the relevant and some non-relevant documents. Precision and recall are calculated on the remaining (residual) document collection. The second one is Freezing, where the initially top ranked documents are "frozen", the ones used to modify the query. The relevance feedback system performs ranking again the remaining documents, and the precision/recall evaluation is conducted on the entire document collection. The third one is Test and control groups, where the document collection is randomly partitioned into two equal groups, the first used to train the relevance feedback system and the second used to evaluate the system. With simulated exhaustive incremental user feedback proposed by *Bellot P. et al.* [5] and *Geva S. et al.* [12], the approach for evaluation was extended in several ways, relative to traditional ad-hoc evaluation with standard precision/recall measures based on Cranfield methodology. It gives exhaustiveness on the evaluation process, thus it becomes reliable and less dependent on the specific search engine in use. Here, there is an evaluation platform where evaluation is performed over executable implementations of relevance feedback algorithms rather than being performed over result submissions.

With HCIR popularization the need of new and different measures has occurred. Despite the information retrieval evaluation, user's effort in the interaction process should be considered and evaluated. Our point of view is that, despite described automated measures, the complex process of HCIR evaluation could be solved with a range of different methods based on user behavior, such as eye-tracking [11, 31], manual evaluation [31], and click-through mining [28]. The aim of all the developed and effectively evaluated approaches in the HCIR field is the improvement of searching quality. Nevertheless, we note that they can be used in the process of discovering how people search, thus different information retrieval systems should be built.

4 Conclusion

The major work realized so far in the field of HCIR is investigated in this paper. We have surveyed and explained several techniques that involve human effort in the searching process, such as facets, snippets and relevance feedback. Existing approaches for effective and efficient retrieval have been summarized and various research challenges examined in detail, such as retrieval of document's parts rather than whole documents. Thus, our contribution is to present altogether the problems and solutions in the domain of the HCIR paradigm as a base for further research steps, such as development of novel improved HCIR approaches for interactive text retrieval and experimenting with them on both English and Macedonian document collection.

References

1. Aalbersberg, I.J.: Incremental Relevance Feedback. In: Proceedings of SIGIR, pp.11-22 (1992)
2. Allan, J.: Relevance Feedback With Too Much Data. In: University of Massachusetts Amherst, MA, USA (1995)
3. Athenikos, S.J., Lin, X.: Search As You Think AND Think As You Search: Semantic Search Interface for Entity/Fact Retrieval". Presented at the Fifth Workshop on Human-Computer Interaction and Information Retrieval (HCIR) (2011)
4. Baeza-Yates R., Ribeiro-Neto B.: Modern Information Retrieval New York. In: ACM Press (1999)
5. Bellot, P., Chappell, T., Doucet, A., Geva, S., Kamps, J., Kazai, G., Koolen, M., Landoni, M., Marx, M., Moriceau, V., Mothe, J., Ramírez, G., Sanderson, M., SanJuan, E., Scholer, F., Tannier, X., Theobald, M., Trappett, M., Trotman, A., Wang, Q.: Report on INEX 2011. In: SIGIR Forum 46(1), pp. 33-42 (2012)
6. Brandow, R., Mitze, K., Rau, L.F.: Automatic condensation of electronic publications by sentence selection. In: Information Processing & Management 31(5), pp.675–685 (1995)

7. Capra, R., Raitz, J.: Diamond Browser: Faceted Search on Mobile Devices. Presented at the Fifth Workshop on Human-Computer Interaction and Information Retrieval (HCIR) (2011)
8. Chuang, W.T., Yang, J.: Extracting Sentence Segments for Text Summarization: A Machine Learning Approach. In: SIGIR '00 Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval (2000)
9. Clark, C.L.A, Kolla, M., Cormack, G.V., Vechtomova. O., Ashkan, A., Büttcher, S., MacKinnon, I.: Novelty and diversity in information retrieval evaluation. In: SIGIR 2008, pp.659-666 (2008)
10. Cleverdon, C.W.: The Cranfield tests on index language devices. In: Aslib Proceedings, 19, pp.173–192 (1967)
11. Cutrell, E., Guan, Zh.: What Are You Looking For? An Eye-tracking Study of Information Usage in Web Search. In: CHI'07 (2007)
12. Geva, S., Kamps, J., Schenkel, R., Trotman, A.: Comparative Evaluation of Focused Retrieval. In: 9th International Workshop of the Inititative for the Evaluation of XML Retrieval, INEX 2010, Vugh, The Netherlands. December 13-15, (2010). Revised Selected Papers Springer (2011)
13. Goldsteiny, J., Kantrowitz, M., Mittal, V., Carbonelly, J.: Summarizing Text Documents: Sentence Selection and Evaluation Metrics. In: Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval. (1999)
14. Huang, Y., Liu, Z., Chen,Y.: eXtract: a snippet generation system for XML search. In: PVLDB 1(2), pp.1392-1395 (2008)
15. Ingwersen, P.: Information Retrieval Interaction. Taylor Graham (1992)
16. Järvelin. K., Kekäläinen, J.: Cumulated gain-based evaluation of IR techniques. In: ACM Trans. Inf. Syst., Vol. 20, No. 4., pp. 422-446 (2002)
17. Kashyap, A., Hristidis, V., Petropoulos, M., FACeTOR: Cost-Driven Exploration of Faceted Query Results. In: CIKM (2010)
18. Kelly, D., Belkin, N.J.: Reading time, scrolling and interaction: exploring implicit sources of user preferences for relevance feedback. In: Proceeding SIGIR '01 Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, pp.408 - 409, NY, USA (2001)
19. Lee, D., Chuang, H. & Seamons, K.: Document ranking and Vector space model. In: IEEE Software, 14(2), pp.67–75 (1997)
20. Lv, Y., Zhai, C.: Positional relevance model for pseudo-relevance feedback. In: Proceeding SIGIR '10 Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval, pp.579-586, ACM New York, NY, USA (2010)
21. Manning, C.D., Raghavan, P., and Schütze, H.: Introduction to Information Retrieval. Cambridge, UK: Cambridge University Press (2008)
22. Marchionini, G., Brunk, B.: Towards a General Relation Browser: A GUI for Information Architects. In: J. Digit. Inf. 4(1) (2003)
23. Marchionini, G.: From Information Retrieval to Information Interaction. In: ECIR 2004: 1-11 (2004)
24. Marchionini, G.: Toward Human-Computer Information Retrieval Bulletin. In: June/July 2006 Bulletin of the American Society for Information Science (2006)
25. Morita, M., Shinoda, Y.: Information Filtering Based on User Behaviour Analysis and Best Match Text Retrieval. In: SIGIR 1994, pp.272-281 (1994)
26. Overwijk, A., Nguyen, D., Hauff, C., Trieschnigg, D., Hiemstra, D., Jong, F.D.: On the evaluation of snippet selection for WebCLEF. In: Proceeding CLEF'08 Proceedings of the

9th Cross-language evaluation forum conference on Evaluating systems for multilingual and multimodal information access, pp.794-797 Springer-Verlag Berlin, Heidelberg. (2009)
27. Pehcevski, J., Piwowarski, B.: Evaluation Metrics for Semi-Structured Text Retrieval. In: Encyclopedia of Database Systems, Editors-in-chief: Liu, Ling; Özsu, M. Tamer, Springer (print and online) (2009)
28. Radlinski, F., Kurup, M., Joachims, T.: How Does Clickthrough Data Reflect Retrieval Quality? In: CIKM'08 (2008)
29. Rocchio, J. J.: Relevance feedback in information retrieval. In: G. Salton, editor, The SMART Retrieval System: Experiments in Automatic Document Process-ing, Prentice-Hall Series in Automatic Computation, chapter 14, pp.313–323. Prentice-Hall, Englewood Cliffs NJ (1971)
30. Ruthven, I., Lalmas, M.,: A survey on the use of relevance feedback for information access systems. In: Knowl. Eng. Rev., 18(2), pp.95-145 (2003)
31. Savenkov, D., Braslavski, P., Lebedev, M.: Search Snippet Evaluation at Yandex: Lessons Learned and Future Directions. In: CLEF, Vol. 6941 Springer, pp. 14-25 (2011)
32. Schuth, A., Marx. M.J.: Evaluation Methods for Rankings of Facetvalues for Faceted Search. In:Proceedings of the Conference on Multilingual and Multimodal Information Access Evaluation (2011)
33. Singhal, A., Buckley, C. & Mitra, M.: Pivoted Document Length Normalization. In: ACM SIGIR'96 (1996)
34. Sparck Jones, K., Walker, S. & Robertson, S.E., 2000a.: A Probabilistic model of information retrieval: development and comparative experiments. In: Part 1. In Information Processing and Management 36, pp.779-808 (2000)
35. Sparck Jones, K., Walker, S. & Robertson, S.E., 2000b.: A Probabilistic model of information retrieval: development and comparative experiments. In: Part 2. In Information Processing and Management 36, pp.809-40 (2000)
36. Tombros, A., Sanderson, M.: Advantages of query biased summaries in information re-trieval. In Proceeding SIGIR '98 Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval Pages 2 – 10 ACM New York, NY, USA (1998)
37. Turpin, A., Tsegay, Y., Hawking, D., Williams, H.E.: Fast generation of result snippets in web search. In: SIGIR 2007, pp.127-134 (2007)
38. White, R.W., Jose, J.M., Ruthven, I.: A task-oriented study on the influencing effects of query-biased summarisation in web searching. In: Information Processing & Management 39(5), pp.707–733 (2003)
39. Zhai, C., Lafferty, J.: A Study of Smoothing Methods for Language Models Applied to Information Retrieval. In: ACM Transactions on Information Systems (TOIS), Volume 22, Issue 2 (2004)
40. Zhang, J., Marchionini, G.: Coupling browse and search in highly interactive user interfac-es: a study of the relation browser++. In: JCDL, pp.384 (2004)