Context in Health Information Retrieval: What and Where

Carla Teixeira Lopes
Departamento de Engenharia Informática
Faculdade de Engenharia da Universidade do Porto
Rua Dr. Roberto Frias s/n
4200-465 Porto, Portugal
ctl@fe.up.pt

Cristina Ribeiro
Departamento de Engenharia Informática
Faculdade de Engenharia da Universidade do Porto/INESC-Porto
Rua Dr. Roberto Frias s/n
4200-465 Porto, Portugal
mcr@fe.up.pt

ABSTRACT

Researchers are aware that context affects information retrieval in general. The health area is no exception and is particularly rich in terms of context. To understand how context is used in health information research, we collected a sample of health information research papers that use context features. Papers were analyzed and classified according to the type of context features and to the stage of the retrieval process into which they were incorporated. Further, we also identified the specific context features used in each category of features and each stage of the process. Results show a weaker use of interaction context features than we expected and, as supposed, a large use of collective features. A considerable number of papers use context to query related activities. We also found that research is mainly aimed at health professionals, suggesting a gap in health consumers research that should be explored.

Categories and Subject Descriptors

H.3.3 [[**Information Storage and Retrieval**]]: Information Search and Retrieval

General Terms

Human Factors

Keywords

Context, Information Retrieval, Survey, Health

1. INTRODUCTION

Health Information Retrieval (HIR) focus on the application of IR concepts and techniques to the domain of healthcare. This field has largely evolved in the last few years. Habits of health professionals and consumers (patients, their family and friends) have been changing as a result of several factors like the increasing production of information in a digital format [27], the greater availability and the easier access to health information.

HCIR 2010 22 August 2010, New Brunswick NJ, USA

Several authors agree that context, often ignored, might be used to improve the retrieval process [21, 2]. A contextualised Information Retrieval (IR) could allow IR systems to learn and predict what information a searcher needs, learn how and when information should be displayed, present how information relates to other information that has been seen and how it relates to other tasks the user was engaged in and decide who else should be informed about new information.

According to Lin and Fushman, "the domain of clinical medicine is very well-suited for experiments in building richer models of the information seeking process" [23]. In fact, it's not difficult to foresee context features in this domain that could enrich HIR models. Similarly to any visit to the doctor, where the patient doesn't just say "itch", but explains the context of the "itch" to the doctor, context is relevant to HIR. Other examples of context features that can be used are the search scenario [24] and its specificities (e.g.: treatment of a disease), the searcher's personal health record, the clinical case in hands and the searcher's knowledge in the health domain.

We have done a review on the definition of context in a previous work [25]. To this work, context is considered an interactional problem, as defined by Dourish [8]. It not only includes the environmental features surrounding the user and his activities, but also the interaction in which he is involved. We believe context is dynamic and might change each time a new search is made, a new set of results is reviewed or a new document is viewed [14].

To understand how context is being used in health information research, we gathered a set of HIR research papers that use any kind of context features. These papers were analyzed and classified according to the type of used context features and to the stage of the retrieval process into which they were incorporated. Further, we also identified the specific features used in each context category and each stage of the process.

The following section presents the adopted methodology, specifying how the papers were selected and describing the taxonomies used in the classification. Section 3 presents the classification of the research papers and enumerates the specific context features used in each category and stage. Finally, in Section 4 we report the main conclusions of this analysis.

2. METHODOLOGY

To define the sample of papers, we considered all the documents classified with the tags context and health in CiteU-

Like[1], a social web service for management of bibliographic references. From this set we excluded papers not related with IR and papers in which IR was not the main focus. For example, papers on Information Extraction and papers proposing readability formulas for health documents were excluded from this analysis. In addition, papers without an innovative contribution (e.g.: literature reviews or comparisons of IR systems) were also excluded. The final set was composed of 27 papers.

To classify the research papers according to the used context features, we adopted the Ingwersen and Järvelin's nested model of contexts for Information Seeking and Retrieval (IS&R) [20] that is described in the next subsection. To analyze the usage given to the context features we adopted a taxonomy similar to the one defined by Lopes [25] for the "uses of context".

2.1 Nested model of contexts for IS&R

The first version of Ingwersen and Järvelin's nested model of contexts has 6 dimensions [20]. The first and second dimensions represent the intra and inter object contexts and are the central component of the cognitive IS&R framework, proposed by the authors. The other four dimensions are: the interaction (session) context; the context provided by the remaining components of the framework; the societal infrastructures and, across the stratification, the historic context of all actors' experience. Later, and by the same authors, the social/organizational/cultural context dimension was divided in two subdimensions: an individual and a collective one [19].

This model may be centered on the information space, on the cognitive author (e.g.: searcher), on the interface, on the information technology (engines, logics, algorithms) or on the social/organizational/cultural context. This choice will affect the nature of the interaction context and the context of the individual and collective dimensions.

In this classification we decided to center the model on the information space as can be seen in Figure 1. The cognitive actor was another potential alternative but we felt the specificities of the information space in the health domain would be better described if placed in the first two dimensions of the model. Searcher's context is therefore included in the fourth dimension. We also felt the choice of the cognitive actor as the core would result in a more ambiguous model. In fact, depending on the use given to context features, the cognitive actor could be the searcher or another actor (e.g.: person contributing to the indexing process).

2.2 Uses of context taxonomy

To analyze how the context features are used, we adopted four categories, similar to the four top categories of the *uses of context* taxonomy proposed by Lopes [25]: Indexing and Searching, Query Operations, Ranking and Interface. The Query Operations category is more comprehensive than the *Relevance Feedback and Query Expansion* category initially proposed in Lopes's work because, in the health domain, it is frequent to have systems that generate queries and gather information resources from other systems. With this modification, papers describing this kind of research can fit into this category.

[1]`http://www.citeulike.org/search/all?q=tag\%3Acontext+\%26\%26+tag\%3Ahealth`

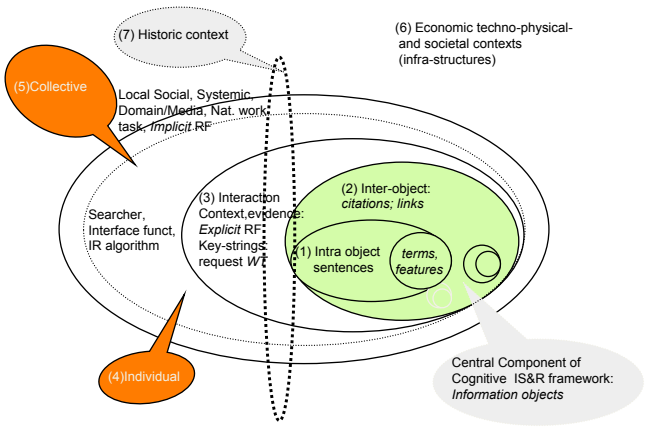

Figure 1: Ingwersen and Järvelin's nested model of contexts [20] with the information space as the central component.

In the IR process, the ranking phase is usually straight connected to the searching phase. Yet, we prefered to keep them as two distinct categories to help differentiate systems that have their own index and implement a retrieval model from systems that just reorder existing result sets based on some specific criteria.

3. RESEARCH ANALYSIS

The results of our analysis are presented in Figure 2 with the distribution of papers by categories. For convenience of representation, we switched the initial order [25] of the interface and ranking categories. Each paper is represented by its bibliographic reference and a letter (P, C or B) that represents the type of users to whom the system is targeted: professionals, consumers or both.

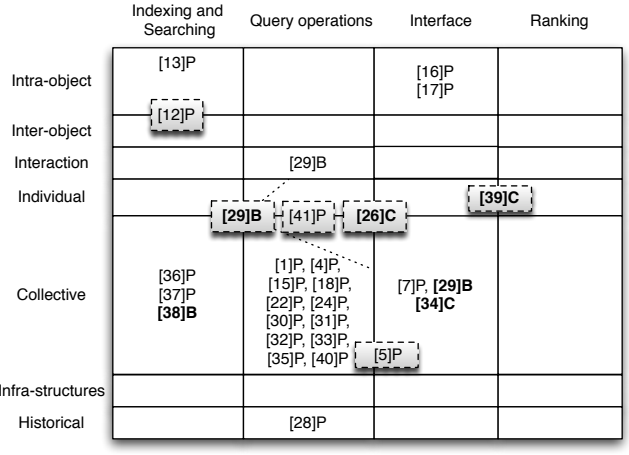

Figure 2: Classification based on the used context features and their specific use.

When a paper crosses more than one category, its reference is represented in the categories' intersection area. In some cases, it may also be connected with a dotted line to another cell of the matrix. For example, paper with refer-

Table 1: Context Features used in CHIR.

	Indexing and Searching	Query Operations	Interface	Ranking
Intra-Object	Document contents and structure (e.g. abstract, conclusions, title, HTML structure).		Document images and captions.	
Inter-Object	Links between documents.			
Interaction		Browsing behavior.		
Individual	Authoring context.	Searcher's clinical data and user interest.	Searcher's clinical data and PHR.	PHR.
Collective	UMLS, domain categories, tasks, ontologies, taxonomies and patient data (age, sex and clinical context).	UMLS, MeSH, domain questions and terminologies, clinical practice guidelines, retrieval feedback, task context and patient data (clinical data, consult reports, exam reports, EHR).	UMLS, MeSH, domain questions, Gene Ontology and patient data (clinical data, EHR).	
Infrastructures				
Historical		Search history.		

ence [29] uses interaction, individual and collective context features in Indexing and Searching, Query operations and Interface stages.

Figure 2 shows that research is more intense on Query Operations using mainly context features from the individual and collective dimensions. We were surprised with the weak use of the interaction context. This might be explained by the preference to use context features more related to the health domain. Typically, interaction context is more generic and not so health-related as individual and collective context features. On the other hand, we already expected to have a large number of papers using collective context features since this category is exhaustive, covering the characteristics of all the components from the cognitive framework that are not at the center of the model.

In Figure 2 we highlight the papers dedicated to research on health consumers systems (letters C or B). As can be seen, research is mostly dedicated to health professionals. The small number of consumer dedicated research papers use interaction, individual and collective context features.

To show which exact context features are used, we built Table 1 where we included the features in a structure similar to the one in Figure 2. In this table, EHR stands for Electronic Health Record and PHR for Personal Health Record, to distinguish institutional data from the records managed by the patient. UMLS is a project from the National Library of Medicine (NLM) of the United States composed of three knowledge sources: the Metathesaurus, the Semantic Network and the SPECIALIST Lexicon and Tools. MeSH is also an NLM thesaurus.

As can be seen in the collective dimension of Table 1, the health domain is very rich in structured information. This dimension mainly consists of terminologies, thesaurus and ontologies. Note that in IR systems used by health professionals, the EHR and patient's clinical data is part of the professional work task. Therefore, in professional systems, these context features incorporate the collective dimension of context. In IR systems designed for patients, the use of clinical data or PHR about the searcher is considered individual context.

4. CONCLUSIONS

Most researchers are aware that context affects information retrieval. The health area is no exception, being particularly rich in terms of context. Results presented in the previous section show a weaker use of interaction context features than we expected. Also, research makes an extensive use of collective features. This was not a surprise because this dimension is very comprehensive, including several types of context features. In addition, it is the dimension where all the health-related structured knowledge sources (e.g.: thesaurus) are included. A considerable number of papers use context to query related activities.

We have noticed that research has been more focused on health professionals than on consumers. Of the 27 papers analyzed, only 3 are dedicated to health consumers and 2 are dedicated to both professionals and consumers. This difference may be explained by the longer tradition of information retrieval in health professionals when compared to consumers. Only recently, with the advent of the Web, has search become more popular among health consumers. Other possible reasons include the large number of medical knowledge sources, the possibilities open by the integration of search systems with clinical systems and the difficulties associated with user studies in consumer health retrieval.

The lack of research on the use of context in health IR by consumers, the growing number of health searches (61% of the American adults look online for health information [11] and so does 19,6% of the Portuguese population aged 15 or more [9]) and the importance of well-informed patients [10] suggest the importance of focusing research on health consumers.

5. ACKNOWLEDGEMENTS

This work is partially funded by Fundação para a Ciência e a Tecnologia under the grant SFRH/BD/40982/2007.

6. REFERENCES

[1] A. R. Aronson and T. C. Rindflesch. Query expansion using the UMLS Metathesaurus. *Proceedings of the AMIA Annual Fall Symposium*, pages 485–489, 1997.

[2] R. Bierig and A. Göker. Time, location and interest: an empirical and user-centred study. In *IIiX: Proceedings of the 1st international conference on Information interaction in context*, pages 79–87, New York, NY, USA, 2006. ACM.

[3] J. J. Cimino. Linking patient information systems to bibliographic resources. *Methods of information in medicine*, 35(2):122–126, June 1996.

[4] J. J. Cimino, A. Aguirre, S. B. Johnson, and P. Peng. Generic queries for meeting clinical information needs. *Bulletin of the Medical Library Association*, 81(2):195–206, April 1993.

[5] J. J. Cimino, S. B. Johnson, A. Aguirre, N. Roderer, and P. D. Clayton. The MEDLINE button. In *Proc Annu Symp Comput Appl Med Care*, pages 81–85, 1992.

[6] J. J. Cimino and J. Li. Sharing infobuttons to resolve clinicians' information needs. In *AMIA Annu Symp Proc.*, page 815, 2003.

[7] A. Doms and M. Schroeder. GoPubMed: exploring PubMed with the Gene Ontology. *Nucl. Acids Res.*, 33(suppl_2):W783–786, July 2005.

[8] P. Dourish. What we talk about when we talk about context. *Personal Ubiquitous Comput.*, 8(1):19–30, February 2004.

[9] R. Espanha and F. Lupiáñez Villanueva. Health and the Internet: Autonomy of the User. Technical report, LINI - Lisbon Internet and Networks, 2008.

[10] T. Ferguson. e-patients - how they can help us save health care. Technical report, e-patients scholars working group, 2007.

[11] S. Fox and S. Jones. The social life of health information. Technical report, Pew Internet & American Life Project, June 2009.

[12] M. E. Frisse. Searching for information in a hypertext medical handbook. In *HYPERTEXT '87: Proceedings of the ACM conference on Hypertext*, pages 57–66, New York, NY, USA, 1987. ACM.

[13] C. W. Gay, M. Kayaalp, and A. R. Aronson. Semi-automatic indexing of full text biomedical articles. *AMIA Annual Symposium Proceedings*, pages 271–275, 2005.

[14] D. J. Harper and D. Kelly. Contextual relevance feedback. In *IIiX: Proceedings of the 1st international conference on Information interaction in context*, pages 129–137, New York, NY, USA, 2006. ACM Press.

[15] Z. Hashmi, T. Zrimec, and A. Hopkins. CQGF: Context specific query generation framework from computerized clinical practice guidelines. In *Second International Conference on the Applications of Digital Information and Web Technologies - ICADIWT '09*, pages 288–293, October 2009.

[16] M. A. Hearst, A. Divoli, H. Guturu, A. Ksikes, P. Nakov, M. A. Wooldridge, and J. Ye. Biotext search engine: beyond abstract search. *Bioinformatics*, 23(16):2196–2197, August 2007.

[17] M. A. Hearst, A. Divoli, J. Ye, and M. A. Wooldridge. Exploring the efficacy of caption search for bioscience journal search interfaces. In *BioNLP '07: Proceedings of the Workshop on BioNLP 2007*, pages 73–80, Morristown, NJ, USA, 2007. Association for Computational Linguistics.

[18] W. Hersh, S. Price, and L. Donohoe. Assessing thesaurus-based query expansion using the UMLS metathesaurus. *AMIA Annual Symposium Proceedings*, pages 344–348, 2000.

[19] P. Ingwersen. Context in information interaction revisited 2006. Presentation at Prolissa 2006: Proceedings of the Fourth Biennial DISSAnet Conference, 2006.

[20] P. Ingwersen and K. Järvelin. *The Turn: Integration of Information Seeking and Retrieval in Context (The Information Retrieval Series)*. Springer, 1 edition, September 2005.

[21] P. Ingwersen, K. Jelin, and N. Belkin, editors. *Proceedings of the ACM SIGIR 2005 Workshop on Information Retrieval in Context (IRiX)*, Royal School of Library and Information Science. Denmark., August 2005.

[22] L. C. Kingsland, A. M. Harbourt, E. J. Syed, and P. L. Schuyler. Coach: applying UMLS knowledge sources in an expert searcher environment. *Bull Med Libr Assoc.*, 81(2):178–183, April 1993.

[23] J. Lin and D. D. Fushman. Representation of information needs and the elements of context: A case study in the domain of clinical medicine. In *ACM SIGIR 2005 Workshop on Information Retrieval in Context (IRiX)*, 2005.

[24] Liu, Zhenyu, Chu, and Wesley. Knowledge-based query expansion to support scenario-specific retrieval of medical free text. *Information Retrieval*, 10(2):173–202, April 2007.

[25] C. T. Lopes. Context features and their use in information retrieval. In *Third BCS-IRSG Symposium on Future Directions in Information Access*, September 2009.

[26] G. Luo and C. Tang. On iterative intelligent medical search. In *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 3–10, New York, NY, USA, 2008. ACM.

[27] P. Lyman and H. R. Varian. How much information. Available from: http://www.sims.berkeley.edu/how-much-info-2003 [cited 2008-07-10], 2003.

[28] R. S. Marcus. An experimental comparison of the effectiveness of computers and humans as search intermediaries. *Journal of the American Society for Information Science. American Society for Information Science*, 34(6):381–404, November 1983.

[29] D. S. Martins, L. H. Z. Santana, M. Biajiz, Antonio, and W. L. de Souza. Context-aware information retrieval on a ubiquitous medical learning environment. In *SAC '08: Proceedings of the 2008 ACM symposium on Applied computing*, pages 2348–2349, New York, NY, USA, 2008. ACM.

[30] S. M. Maviglia, C. S. Yoon, D. W. Bates, and G. Kuperman. Knowledgelink: Impact of context-sensitive information retrieval on clinicians' information needs. *J Am Med Inform Assoc*, 13(1):67–73, Jan 2006.

[31] E. A. Mendonça, J. J. Cimino, S. B. Johnson, and Y. H. Seol. Accessing heterogeneous sources of evidence to answer clinical questions. *Journal of biomedical informatics*, 34(2):85–98, April 2001.

[32] R. A. Miller, F. M. Gieszczykiewicz, J. K. Vries, and G. F. Cooper. CHARTLINE: providing bibliographic references relevant to patient charts using the UMLS Metathesaurus Knowledge Sources. *Proceedings of the Annual Symposium on Computer Application in Medical Care.*, pages 86–90, 1992.

[33] S. M. Powsner and P. L. Miller. Linking bibliographic retrieval to clinical reports: Psychtopix. In *Proc Annu Symp Comput Appl Med Care*, November 1989.

[34] W. Pratt, M. A. Hearst, and L. M. Fagan. A knowledge-based approach to organizing retrieved documents. In *AAAI '99/IAAI '99: Proceedings of the sixteenth national conference on Artificial intelligence and the eleventh Innovative applications of artificial intelligence conference innovative applications of artificial intelligence*, pages 80–85, Menlo Park, CA, USA, 1999. American Association for Artificial Intelligence.

[35] S. L. Price, W. R. Hersh, D. D. Olson, and P. J. Embi. Smartquery: context-sensitive links to medical knowledge sources from the electronic patient record. *Proc AMIA Symp*, pages 627–631, 2002.

[36] G. P. Purcell, G. D. Rennels, and E. H. Shortliffe. Development and evaluation of a context-based document representation for searching the medical literature. *International Journal on Digital Libraries*, 1(3):288–296, December 1997.

[37] G. Quellec, M. Lamard, L. Bekri, G. Cazuguel, B. Cochener, and C. Roux. Multimedia medical case retrieval using decision trees. In *IEEE Engineering in Medicine and Biology Society*, volume 2007, pages 4536–4539, ENST Bretagne, GET-ENST, Brest, F-29200 France. gwenole.quellec@enst-bretagne.fr, 2007.

[38] S. Sakji, A. D. D. Dibad, I. Kergourlay, S. Darmoni, and M. Joubert. Information retrieval in context using various health terminologies. In *Research Challenges in Information Science, 2009. RCIS 2009. Third International Conference on*, pages 453–458, 2009.

[39] J. M. Silva and J. Favela. Context aware retrieval of health information on the Web. In *LA-WEB '06: Proceedings of the Fourth Latin American Web Congress (LA-WEB'06)*, pages 135–146, Washington, DC, USA, 2006. IEEE Computer Society.

[40] P. Srinivasan. Retrieval feedback in MEDLINE. *Journal of the American Medical Informatics Association : JAMIA*, 3(2):157–167, 1996.

[41] Q. Zeng and J. J. Cimino. Linking a clinical system to heterogeneous information resources. In *Proc AMIA Annu Fall Symp.*, pages 553–557, 1997.

[42] Q. Zeng-Treitler, S. Goryachev, T. Tse, A. Keselman, and A. Boxwala. Estimating consumer familiarity with health terminology: A context-based approach. *J Am Med Inform Assoc*, 15(3):349–356, May 2008.