Scenario forms for web information seeking and summarizing in bone marrow transplantation

Margit Becher Brigitte Endres-Niggemeyer Gerrit Fichtner

Fachhochschule Hannover (University of Applied Sciences and Arts)
Dept. of Information and Communication
Ricklinger Stadtweg 120, D-30459 Hanover, Germany
{Brigitte.Endres-Niggemeyer, Margit.Becher, Gerrit.Fichtner}@ik.fh-hannover.de

Abstract

This paper presents the user-centered interface of a summarization system for physicians in Bone Marrow Transplantation (BMT). It serves both retrieval and summarization, eliciting the query and presenting multi-document summaries in a situation-specific organization.

Introduction: User-centered scenario forms for retrieval and summarization

This paper presents the user interface of a summarization system for physicians in Bone Marrow Transplantation (BMT). The interface has users state well-articulated questions that serve as summarization targets and basis for retrieval queries, and it displays summarization results in an organization that fits the user's situation. Although user interfaces have attracted almost no interest in summarization research so far, we think that a suitable user-oriented interface is important for a summarization system. This paper deals with such an interface, not with the summarization procedures the interface enables. In good user-centered design attitude (Norman and Draper, 1986), we developed the user interface first, and are still equipping it component by component with the intended functionality.

Our users are highly specialized physicians in Bone Marrow Transplantation (BMT), a life-critical field of internal medicine. They need answers to their questions that are fast, to the point, and prepared for direct application. Using question/answer scenario forms derived from empirical scenario descriptions, they can specify their current situation and the missing knowledge items with the help of a domain-specific ontology. The system accepts the filled out question scenario, projects it to a query for search engines and Medline (the most common medical reference retrieval engine), and starts the search. Retrieved documents are downloaded, preprocessed, and then checked for passages where question terms accumulate. These passages are examined by summarization agents that follow strategies of human summarizers (Endres-Niggemeyer 1998). Accepted statements enter the summary, under the heading given by the scenario element that asked for them. Thus the summary organizes new knowledge in a fashion that mirrors the user's situation. All the time, the user-centered interface keeps users in their own task environment.

To produce summaries that fit users' information needs, reasonably precise question statements are required. Questions (think of Who? Why? etc.) also have well-known qualities as text organizers, so that they can serve summary organization, the query items switching to headings for (partial) summaries when answers are delivered.

Well-structured queries are most easily elicited by a convenient form. With a suitable choice of real-life scenarios (ideas inspired by Carroll, 2000), users can formulate their search and summarizing requests by filling out such a form, simply stating what they know and what they are missing in a given situation. Where the user identifies a knowledge gap (a question), the system will feed in the respective summary items if possible.

In order to mediate between the users' and the system perspective, we equip the scenario forms with intermediary structures - a detailed interpretation for summarizing and an abridged one for IR. Within these interpretations, the

form itself is represented by constants, the user query provides variables.
In the following, we explain where our inspiration for the interface came from, how its design aims are met, ensued by empirical modeling and implementational details.

1 Background and related approaches

While graphical output of summaries was already addressed in the 80ies, interest in user interfaces of summarization systems is more recent. Aone et al. (1999) as well as Strzalkowski et al. (1999) and Ando et al. (2000) describe graphical user interfaces of their summarizers. White, Ruthven and Joemon (2001) positively evaluate a summarization function and interface added to Altavista and Google. Buyukkokten et al. (w.d.) summarize for hand-helds. Their small screens make them consider the user interface. Kan, McKeown and Klavans (2001) see summarization on top of an IR task, as we do. They intend to replace the common ranked output lists of retrieval or search engines by a multi-document summary structure. Our graphical user interface applies findings and principles of user-centered information seeking. We are not aware of any earlier approaches that deal with user-centered query formulation in summarization.
Human-computer interaction and interface development are well-trodden research areas in IR (overviews by Hearst 1999; Marchionini and Komlodi, 1998). Especially in the Digital Libraries context there is plenty of work and new ideas about how to improve user access and to make retrieval interfaces more intuitive, both on the Internet and in specialized collections such as videos (see e.g. Mackinlay, Rao and Card, 1995, or Geisler et al., 2001). In IR, templates are frequent, but we see no user interfaces that consistently derive the query from a description of the user need. One possible reason for this is that most approaches deal with less well-defined environments, while in the setting of clinical BMT, users are in well-circumscribed situations.

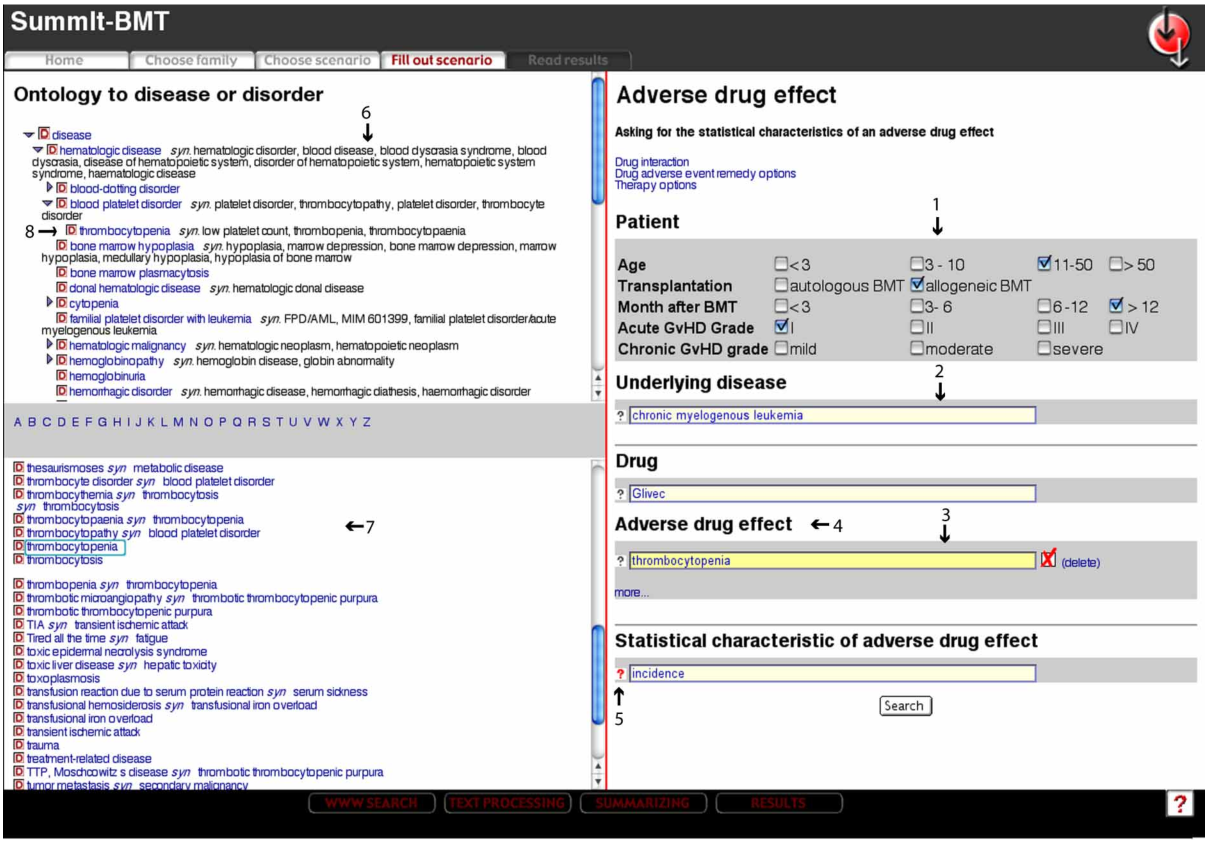

Figure 1. User-oriented query formulation scenario: Adverse drug effects (simplified)

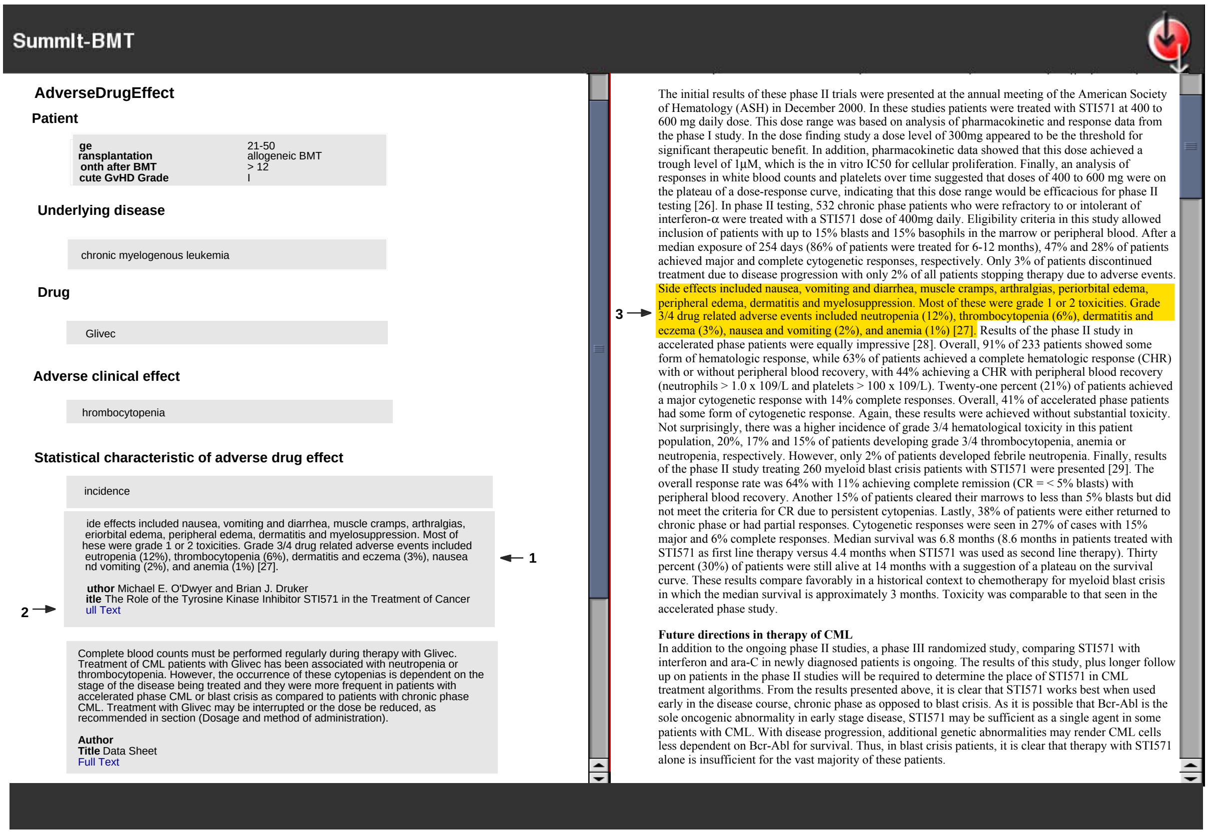

Figure 2. Answering state of the scenario form – source document at the right side. Numbered arrows point to items explained in section 2.1.

User-centered research on information seeking (Belkin, Oddy and Brookes 1982; Bates, 1989; Ingwersen, 1992; Marchionini, 1995; Borgman, 1996; Marchionini and Komlodi, 1998) claims that users are entitled to state their information needs in their own thinking and working context. The research history reflects a long struggle for user-oriented information seeking as opposed to machine-driven query formulation in IR. The ASK (Anomalous State of Knowledge) hypothesis of Belkin, Oddy and Brooks (1982) gives the classic formulation of the problem: Users who ask are in difficulty, they lack a piece of knowledge and are busy restructuring their convictions. This is a bad moment to additionally confront them with an IR system, instead of adapting to their needs. Bates' (1989) landmark "berrypicking" model of information seeking pinpoints the advance in understanding of how human users seek and process information in natural environments. During an information seeking process, Bates (1989) observes many rounds of retrieving documents and learning with changing goals. Empirical data have confirmed her analysis (see Hearst, 1999). When Marchionini and Komlodi (1998) wrote their overview of information seeking interfaces, they put user-centered query interfaces on the research agenda.

It is commonplace knowledge in IR that end users are not very proficient in IR tasks, because of missing background knowledge and practice. They have poor chances to overcome an "eternal novice" condition (Borgman, 1996). Therefore, the traditional work organization in IR includes intermediaries (cf. Ingwersen, 1992). For better retrieval results, they help with problem definition, structure the question strategy in a presearch interview, develop queries that promise retrieval success, execute the search, and check the retrieval result in a post-search interview.

Our summarization follows a knowledge-based human-like style of argumentation (Endres-Niggemeyer, 1998). The main knowledge source is a BMT ontology comprising about 4400 concepts of the domain. Since users state their information need with concepts from this ontology, the ontology is also an important component of the user interface (see figures 1, 4). From an IR point of view, ontologies are legitimate offspring of thesauri. Many thesauri are used for indexing documents and formulating queries (see Lutes, 1998). By applying a domain ontology like a classic thesaurus, we use a practice that has stood the test of decades.

2 The scenario interface

A scenario reflects a concrete situational context. When an information need comes up in a situation, situation characteristics inspire and restrict information seeking: if the patient is a toddler, specific therapies for geriatric patients are out of scope. Situational features of this type are in an obvious way useful for finding out what information is desired to fill a current knowledge gap. They are recorded in our inductively developed scenarios. The graphical user interface supports interaction supported by scenarios: navigation, selection and transfer of items, etc. in a way that is easy and safe to use. A movie helps novice users to get started.

2.1 User view and use

The claim of user-centered information seeking is to enable users to remain integrated in their own domain and think about their own issues when stating their information needs.

We serve physicians in high-duty situations. Well-designed scenarios help them to structure their questions for better retrieval and summarization performance. They incorporate knowledge about the important features of an information need situation.

To fill out the query template of Medline, their standard information source, our users have to deal with features of IR systems, with databases and database fields to be searched, search terms, document types, publication dates and the well-known Boolean operators.

The query scenario presented in figure 1 avoids these problems. With this interface, the retrieval and summarization apparatus remains behind the scenes: no databases to choose, no database fields to select, no Boolean operators to apply.

Instead, we have some unobtrusive technical controls at the screen: check buttons (pointer 1), simple text fields (pointer 2), multiple text fields (pointer 3), field headings procuring the appropriate ontology classes when clicked (pointer 4), question marks (pointer 5). Text fields are labeled and prestructured. They receive an ontology concept that fits the class indicated by the label. If a text field is left empty with the question button selected, this means a question about the ontology class that is prespecified in the slot. Entries for the slots hop in from the system ontology (left side of the screen) by means of a mouse-click. By clicking on the question mark, the user flags a slot as conveying a question. All other slots and checkboxes state the given features of the current situation.

Related scenarios are linked to each other by hypertext links in the head of the forms.

We illustrate how the interaction with scenario forms works with “Adverse drug effect”, a scenario serving situations with drug-related complications. In Figure 1, the physician’s problem is to clear up an adverse drug event: a patient’s platelet count has dropped (trombocytopenia) when Glivec was administered.

Figure 1 shows the question scenario screen with the question already formulated: In the patient block, the user has ticked checkboxes for stating the patient’s age, the type of transplantation (s)he has had, the time that has gone by since the transplant, and the mild (grade 1) Graft-versus-host disease of the patient. After that, the user has clicked on the field label “Adverse drug effect” (pointer 4). The respective class of the ontology showed up on the left of the screen, in systematic order (upper half - pointer 6) and in alphabetical order (lower half – pointer 7). There, the user found “thrombocytopenia”. In both displays a little “D” icon in front of an entry (pointer 8) provides a description of the concept when clicked on. With a click on the “thrombocytopenia” entry, the concept has slipped into its target field “Adverse drug effect”. The user has stated the statistical value of interest, which happens to be “incidence”, and activated the question mark of the slot (pointer 5). This is his information need. The question form is ready for search.

The system returns with the answer screen (Figure 2). The answer (pointer 1) is given as an excerpt from the original. It is inserted under the question field, having here the content “incidence”. This field functions now as a sort of subheading, in line with standard ideas of text organization. Each text clip is linked to its position in the source document (pointer 2). By a click on the hypertext link, the source document is displayed with the extract highlighted (pointer 3).

2.2 Predicate logic interpretation and queries for search engines

Scenarios are equipped with a predicate logic interpretation that sets up the target structure for summarization. It restates the scenario by means of propositions contained in the ontology. Possible user input is introduced by variables.

Figure 3 shows how the intermediary structure buffers users from IR machinery. It comprises the code of the user-oriented scenario and its interpretation: the structured question for summarization formulated in first-order statements including user-produced values, and the query strategy derived from it.

Figure 4 presents the interpretation of the "Adverse drug effect" scenario whose user view is shown in figure 1. It consists of Prolog-style first-order assertions. The generic state of the scenario interpretation is given in the upper left corner of the figure. From this representation, an abridged generic scenario query set has been developed and tested for retrieval capacity (upper right corner). As soon as the user has entered values, they replace in the assertions the surface variable names (tagged with "#") that serve as their placeholders. Thus a specific scenario interpretation (middle left in Figure 4) is built up. Retrieval queries need fewer items than summarization target specifications, but they are derived likewise by replacing variables with their current content. We use up to four queries, relaxing the constraints if the answer set is too small.

Summarization targets are first-order contexts (McCarthy and Buvac, 1993) as used in the ontology. They equip first-order statements with additional context propositions. This is useful in medicine where many assertions are limited in scope. Restrictions must be stated in order to make up true expressions.

When building the target context expression for summarization (see Figure 4, lower left corner), propositions without user input are disregarded. If the respective slot is given the question property, the proposition goes into the core of the context, otherwise it is put into its context section. Like this, different instances of the same scenario give rise to different query sets and trigger different summaries.

2.3 Empirical modeling

Scenario Acquisition. During scenario acquisition, we provide broad scenario families and ask physicians to describe scenarios that might fit into them, drawing from their experience during the last week or month. We obtain empirical scenarios such as:

> "Patient with clearly progressing MDS and isochromosome 13. Autologous Tx?"
> (MDS = Myelodysplastic Syndrome, Tx = Therapy, here bone marrow transplantation)

So far, we have collected 131 empirical scenarios.

Content Engineering. Almost all scenarios of our users need reworking (splitting, restructuring, reformulation etc.) in order to provide the basis for scenario forms of general usability. First, scenario descriptions are filled up with items left out by the physicians, but needed for successful retrieval. Next, empirical scenario descriptions are generalized by separating parameters of the individual situation from its more general and recurring features. If a physician asks for therapies against a herpes simplex infection, the same framework will apply to questions about other infections, for instance about a fusarium infection. A user will enter fusarium instead of herpes simplex. The scenario form codes the reusable structure, individual content is filled in from the ontology.

On the whole, we manage to remodel empirical scenarios.

3 Implementation

Nearly the whole user interface is coded in JavaScript. The scenario interpretations and the ontology are stored in an XML database. For processing, scenarios are transferred with the user input in an XML structure defined by a general DTD describing the common pattern of all scenarios. After processing the user input, for every field marked as a question, extracts of the retrieved documents as well as links to the corresponding full text documents are inserted in the scenario structure. In this way, exactly the same data structure is used both for the question scenario and for the answer scenario.

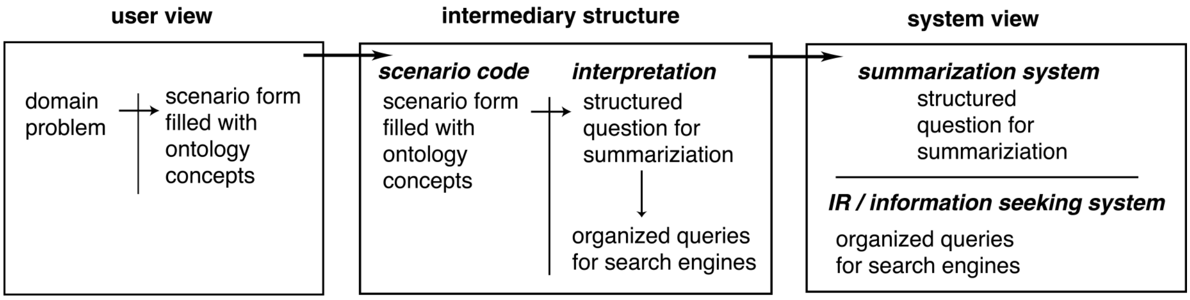

Figure 3. Scenarios mediating between user and system

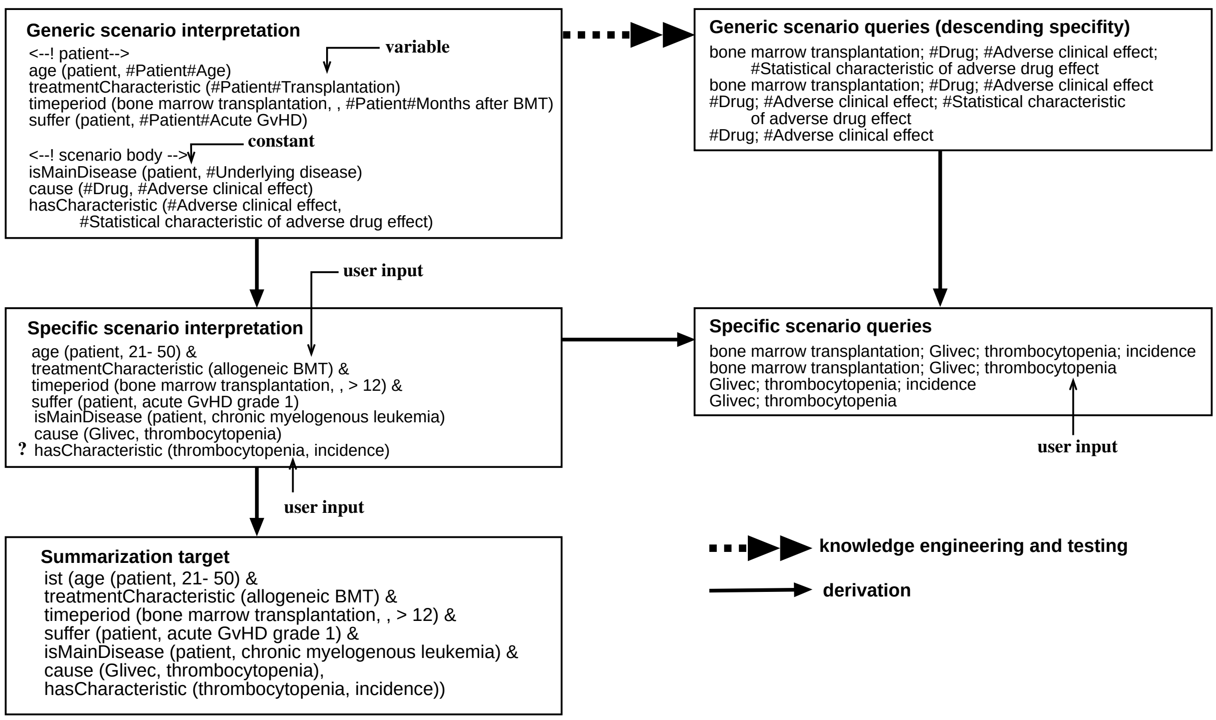

Figure 4. The example scenario's interpretation

4. Current state of development

At the time of writing, our current scenario library comprises 40 scenarios. We expect to have a manageable set (less than 100 of them) when we finish. They serve recurrent situational structures. A catch-all scenario deals with cases where no prefabricated scenario is appropriate. It passes users to PubMed and Google.

Our BMT ontology currently comprises about 4400 concepts, 2800 propositions, and some 1400 contexts. This should be mentioned, since the ontology is crucial for the usability of the system and the user interface.

Scenario engineering is carried out in cooperation with a BMT specialist. We adhere to formative evaluation (Scriven, 1967). Currently, we are starting retrieval tests. Summarization trials will follow. A serious restructuring that will change the usability features of the interface is envisaged for later when we shall move it to a mobile hand-held device.

Conclusion

In this article we have shown how user-based scenario forms support users' interaction with information during information seeking. Thus we comply with a core requirement for user-centered information-seeking interfaces, and we manage to obtain structured questions that set up a reasonable target structure for multi-document summarizing.

Acknowledgements

The thanks of the authors go to Bernd Hertenstein, Claudia Villiger and Juliane Topp for

their help with scenario development, and to Arnold Ganser and Michael Stadler for testing the interface.
Implementation of SummIt-BMT is supported by the German Science Foundation (DFG) under grant EN 186/6-1 and HE 2927/2-1, by the German Federal Ministry of Education and Research (bmbf) under grant 1701200, and by the Ministery of Science and Culture of Lower Saxony under grant 1999.384.

References

Ando, R. K., Boguraev, B. K., Byrd, R. J., & Neff, M. S. (2000) Multi-document Summarization by Visualizing Topical Content. *Workshop on Automatic Summarization*, 2000, 79-88.

Aone, C., Okurowski, M.E., Gorlinsky, J. & Larsen, B. (1999) A scalable summarization system using robust NLP. In I. Mani and M. Maybury (eds), *Advances in automated text summarization*. Cambridge, MA: MIT Press, 71–80.

Bates, M. J. (1989) The design of browsing and berrypicking techniques for the online search interface. *Online Review* 13, 407-424.

Belkin, N., Oddy, R.N., & Brooks, H.M. (1982) ASK for information retrieval: Part I. Background and theory. *Journal of Documentation* 38, 2, 61-71.

Borgman, C. L. (1996) Why are online catalogs still hard to use. *Journal of the American Society for Information Science* 47, 7, 493-503.

Buyukkokten, O., Garcia-Molina, H. & Paepcke, A. (w.d.) *Seeing the whole in parts: Text summarization for web browsing with handheld devices.* http://www-db.stanford.edu/~orkut/papers/www10b/index.html

Carroll, J.M. (2000) *Making use. Scenario-based design of human-computer interfaces*. Cambridge MA: MIT Press.

Endres-Niggemeyer, B. (1998) *Summarizing information*. Berlin: Springer.

Geisler, G., Marchionini, G., Nelson, M., Spinks, R., & Yang, M. (2001) Interface concepts for the Open Video Project. *Proc. of the 64th Annual Meeting of the American Society for Information Science and Technology* (pp. 58 – 75). Washington DC.

Hearst, M. (1999) User Interfaces and visualization. In R. Baeza-Yates & B. Ribeiro-Neto (Eds.) *Modern Information Retrieval*. (pp. 257–324). New York: ACM Press.
http://www.sims.berkeley.edu/~hearst/irbook/10/node1.html

Ingwersen, P. (1992) *Information retrieval interaction*. London: Taylor Graham.

Kan, M.-Y., McKeown, K.R. & Klavans, J.L. (2001) Domain-specific informative and indicative summarization for information retrieval. *Workshop on Text Summarization / DUC 2001 Meeting on Summary Evaluation.* New Orleans. http://wwwnlpir.nist.gov/projects/duc/duc2001/agenda_duc2001.html

Lutes, B. (1998) Web thesaurus compendium. http://www-cui.darmstadt.gmd.de/~lutes/thesauri.html

Mackinlay, J. D., Rao, R. & Card S. K. (1995) An organic user interface for searching citation links. http://www.acm.org/sigchi/chi95/Electronic/documnts/papers/jdm_bdy.htm

Marchionini, G. (1995) *Information seeking in electronic environments*. New York: Cambridge University Press.

Marchionini, G., & Komlodi, A. (1998) Design of interfaces for information seeking. In *ARIST 33, 1998 Annual Review of Information Science and Technology* (pp. 89-130). Medford, NJ: Information Today.

McCarthy, J. & Buvac, S. (1993) Notes on formalizing context. *Proceedings of IJCAI'93* (pp. 555–560). Chambery, France. http://www-formal.stanford.edu/jmc/context3/context3.html

Norman, D. A. & Draper, S. W. (eds.). (1986) *User-centered system design*. London: Erlbaum.

Scriven, M. (1967). The methodology of evaluation. In R. Tyler, R. Gagne & M. Scriven, (Eds.) *Perspectives of curriculum evaluation.* (pp. 39-83). Chicago: Rand McNally. Strzalkowski, T., Stein, G. , Wang, J. & Wise, B. (1999). A robust practical text summarizer. In I. Mani and M. Maybury (eds), *Advances in automated text summarization*. Cambridge, MA: MIT Press, 137–154.

White, R., Ruthven, I. & Jose, J. M. (2001). Web document summarisation: a task-oriented evaluation. International Workshop on Digital Libraries. *Proc. of the 12th International Database and Expert Systems Applications Conference (DEXA 2001).* Munich. http://www.cs.strath.ac.uk/~ir/papers/