BibTutor – tutor assistance for research in information pools

Sandra Zilles

DFKI GmbH, Erwin-Schrödinger-Straße, 67663 Kaiserslautern, Germany,
e-mail: sandra.zilles@dfki.de

Abstract

Studies have shown that German students and scientists face massive problems in the efficient acquisition and effective utilisation of information relevant for their work. This can be ascribed to a deficiency in the users' ability to identify appropriate sources of information and to develop successful search strategies.

What would help the user would be a system supporting problem-oriented approaches (by adapting to the user's context) and making its suggestions transparent to the user. Thus the user can not only resolve search problems but also improve his or her capabilities in acquisition and utilisation of information. Below the idea for a new assistance system for literature research—called BibTutor—is described. BibTutor combines the idea of problem-orientation and transparency in providing availability, interactivity, and adaptivity. Considering systems for literature research as *tools*, a system featuring these properties would be more than a tool, namely a kind of *assistant*.

1 Motivation and goals of the BibTutor system

Although humans could not cope with most of their everyday work without using powerful computer systems as tools, often the full capabilities of such tools cannot be exploited just because their users do not know how to use them efficiently depending on their current needs. In particular, depending on a problem context, the user might fail to solve the problem with the given tool, though in general appropriate usage of the tool would lead to the desired solution.

A very prominent example illustrating this fact is how systems for searching in large information pools are used. Even with a quite simple architecture, it is possible to use search tools very efficiently in case proper queries are posed to the system. Which kind of queries suits best depends on the problem context. Unfortunately, most users tend to posing queries in such a way that either too many or too few results—or maybe just not the desired results—are returned by the system. Thus a lack of user expertise causes a

lack of quality of the system's output and consequently of the proposed solution to the user's problem.

An approach to tackling such deficiencies is to implement interactivity into computer systems and thus to develop systems which are no longer only *tools*, but rather *assistants* which communicate with the user in order to adapt to the current problem context and thus to solve the user's problem in *cooperation* with the user.

1.1 Information research systems and their users' needs

The growing amount of recorded information, the change of scientific information systems, but also the growing needs of gathering information account for a (feeling of) disorientation amongst users of information research systems. Studies have shown that German students and scientists face massive problems in the efficient acquisition and effective utilisation of relevant information, see, for instance, Klatt *et al.* [7]. These problems cannot be ascribed to a deficiency of algorithms for search, because the algorithms implemented can basically be used to effectively find the relevant information in any given context, as long as the relevant information is contained in the underlying data pools and the user poses the appropriate queries.

Instead, Klatt *et al.* [7] trace these problems back to a deficiency in the users' ability

- to identify appropriate sources of information (and thus to restrict the search to suitable data pools) and
- to develop successful search strategies.

Moreover, another problem is that users often don't realise that their search results are not optimal. Because of their lack of knowledge about the effect of different search strategies in different information sources they fail to identify cases in which a different method would be beneficial.

These studies concern only Germany, but similar aspects have been discussed in the United States earlier, see for example Mancall *et al.* [9] and Kuhltau [8]. In particular, by hierarchical goal decompositions of search tasks, Bhavnani and Bates [2] illustrate that the above two kinds of deficiencies are crucial in connection with information search. However, first approaches to user studies have been very much criticised, see, e. g., Hjørland [5], but recent approaches which also focus the user's context and task, seem to be more appropriate, see Wilson [10]. As will be discussed below, context and task are also the crucial parameters that should be taken into account when coping with the above two kinds of deficiencies in user skills.

This affects in particular the problem of literature research, which is among the important tasks students and scientists face to accomplish their scientific work. In Germany, up to now, the steps taken to enhance the skills needed for efficient literature research in (online) libraries have been to offer lectures by experts and to implement online tutorials, both of which have advantages and disadvantages.

What is good about lectures is that they offer the possibility of interactivity between the teacher and the learner. In particular, users can ask specific questions concerning their most recent special problems in literature research. Unfortunately, still these lectures have to be rather general and cannot adapt to the individual needs of all participants. Additionally, these lectures are not available on demand, i. e., when a research problem actually occurs. Consequently, not only because such lectures are often scheduled in collision with other (mandatory) lectures for students, they are not very well-attended—at least considered in relation to the number of students and scientists affiliated.

Concerning online tutorials, an advantage is their availability. Nevertheless, first approaches implemented in German research systems have mainly dealt with local library catalogues and have neglected problem-oriented strategies. In particular, the manifold possibilities e-learning offers—as for instance implemented in the e-learning system DaMiT [6, 12] (a data mining tutor)—have not been exploited. Improved tutorials concerning strategies for the selection of catalogues and for the formulation of research queries are currently used, such as for example DISCUS [4, 13], FIT [14], and LOTSE [15]. Here one main disadvantage is that the structure of the learning material offered is rather prefabricated and has little relation to the user's individual needs. The reason is that these systems do not support interactivity.

What would help the user better would be a system combining the advantages of both lectures and online tutorials and moreover supporting problem-oriented approaches (by adapting to the user's context). This requires the provision for *availability*, *interactivity*, and *adaptivity*. Considering systems for literature research as *tools*, a system featuring these three properties would be more than a tool, namely a kind of *assistant*. Such an assistant is currently being developed in the tutor system BibTutor [11] under grant of the German Federal Ministry of Education and Research (BMBF).

1.2 Intelligent assistance in research systems

The scope of assistance in literature research must be to support the user in those tasks in which deficiencies have been observed, namely in the problem of

- identifying appropriate sources of information and thus restricting the search to suitable catalogues and
- developing successful search strategies and queries.

These are the main goals of the tutor system BibTutor. To guarantee availability, BibTutor will run on web servers of library institutions. To achieve interactivity and adaptivity, BibTutor will guide the user during an actual research process using the existing catalogues and research systems provided by the library institutions. Aiming at the desired support functionalities, BibTutor will be realised in two major components, as illustrated in Figure 1:

1. the *context acquisition* component, which is designed to suggest appropriate catalogues depending on the user's current context;

2. the *online research assistant* component, which is designed to help the user in developing appropriate research strategies depending on the user's context and current search actions.

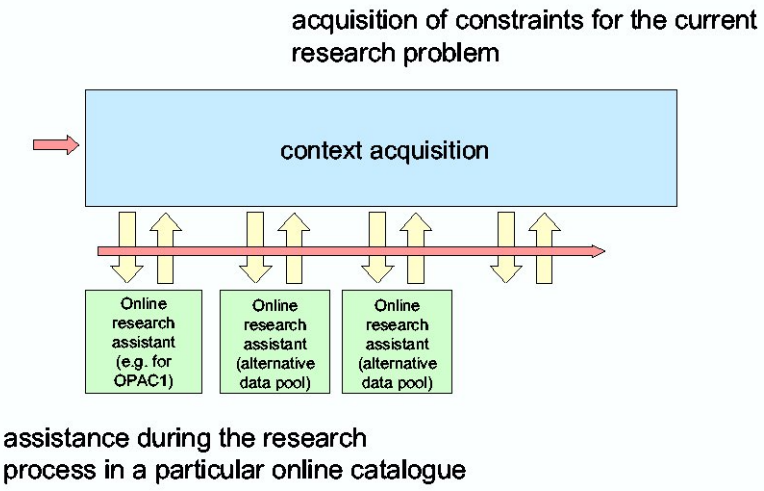

Figure 1: The components of BibTutor.

Context acquisition takes into account special research preconditions given by the user. These might include for example the desired topic and language of the literature searched for, the kind of scientific work the literature is needed for (e. g., writing a diploma thesis or a scientific paper, preparing a talk or just getting an overview on a subject), or the time constraints concerning the availability of the desired documents. These preconditions then affect the choice of catalogues BibTutor suggests for the start of the search.

The online research assistant then guides the user during the following search process. First, the user will formulate queries to the research system provided by the library. Then the online research assistant is supposed to detect whether or not the number and quality of the results is adequate for the user's needs (taking into account the constraints recorded by the context acquisition component). Depending on its diagnosis, it suggests alternative queries to the user in order to improve the results. Moreover, if the user is registered in the system, then his or her research problems recorded in previous sessions shall be taken into account when proposing search strategies.

1.3 Helping people to help themselves

At this point, one important feature of the BibTutor system must be stressed. In contrast to many other assistance systems, BibTutor does not aim at easing the user's work as much as possible, but at helping users to help themselves. Instead of hiding much

of the assistance functionalities behind the user interface, the system will be implemented to act in a more transparent way. In particular, all suggestions directed to the user are supposed to be explained in a way that the user can comprehend the reasons for these suggestions. Hopefully, this enables the user to improve his or her own skills in literature research. Thus, the user's deficiencies are not tackled by solving the tasks the user has to solve but by showing the user how to solve these tasks. According to Buckland and Florian [3], this is one of the important criteria for providing assistance for users of complex systems (such as information systems).

In general, as Buckland and Florian [3] state, assistance in complex systems can be designed along two dimensions,

1. the location—relative to the system—in which assistance functionalities are implemented:
   - *outside* the system core (like, e. g., in tutorials),
   - *within* the system core (like, e. g., in on-line help);
2. the role of the assistance for the user when handling complex problems:
   - *reducing* the user's need for expertise (by resolving parts of the complex problem),
   - *increasing* the user's expertise (by helping the user to self-dependently resolve the complex problem).

This leads to four categories of assistance for complex systems. BibTutor clearly focusses on the category of assistance within the system by increasing the user's expertise.

The degree of assistance from which the user benefits most is also dealt with by Bates [1].

2 Framework for a conceptual development

The BibTutor project is funded by the German Federal Ministry of Education and Research (BMBF) and processed by six partners in Germany: the university libraries of Kaiserslautern, Heidelberg, Darmstadt, and Hamburg-Harburg, the company Brainbot Technologies, as well as the German Research Center for Artificial Intelligence (DFKI).

The project is currently in its initial phase in which concepts are being developed especially for the context acquisition component and the online research assistant. A main task of this phase is to work out prototypical scenarios for research processes, thereby focussing on the possible benefit from interactivity and adaptivity. Of course this implies that the following sections deal with needs and requirements more than with concepts.

The required scenarios must be developed together with information scientists with sufficient experience in typical problems in literature research and suitable chains of activities for resolving these. The knowledge of these chains of activities will then help

to implement the desired assistance functionalities in the BibTutor system. For that purpose, it is helpful to distinguish between different levels of search activities, as are proposed by Bates [1], see Table 1.

Level	Name	Definition
1	Move	An identifiable thought or action that is a part of information searching.
2	Tactic	One or a handful of moves made to further a research.
3	Strategem	A larger, more complex set of thoughts and/or actions than the tactic (consists of multiple tactics and/or moves, all designed to exploit a particular search domain).
4	Strategy	A plan, which may contain moves, tactics, and/or strategems, for an entire information search.

Table 1: Levels of search activities according to Bates [1].

2.1 Context acquisition

During the context acquisition, the user's context and parts of the user's task are recorded. The aim is to identify a category of search strategy according to Table 1 which fits the user's needs (examples for such strategies may be 'locating information for research term paper', 'locating information for book report', or 'verify or complete a reference for a bibliography', see Bates [1]) and to suggest an appropriate catalogue in which to start the search.

In order to be able to adapt to the user's needs, a user model is developed. Parts of the information are recorded during registration of users, parts after the login. Although some of the information must be prompted from the user, some is collected unintrusively by logging the user's activities every time he or she uses the BibTutor system.[1] Thus users can be classified before the actual context acquisition process starts.

Profile of user and context

For the described purpose, for each user, the following kinds of information are requested:

- a topic relevant for the current search problem,
- the category of research product the user currently is supposed to develop (e. g., seminar paper, bachelor, master, diploma or PhD thesis),
- the category of the current research problem (e. g., searching a document with a known title or theme, documents by a known author or from a known series) as well as the form of the searched document (e. g., book, article, encyclopedia, glossary

[1] Obligations of data protection will be taken into account.

entry, collection of tables) depending on the category (e. g., 'glossary entry' will not be a choice suggested by BibTutor, if a document in a known series is searched for),

- the desired language of the required documents,
- the desired time range for the publication date of the required documents,
- the amount of time and money the user is willing to spend for getting access to the required document.

Thereby, if the user is registered to the system, then further knowledge recorded during registration and previous sessions (such as, e. g., his search history from previous BibTutor sessions and whether he is a bachelor, master, or PhD student) can be used to adapt the above choices suggested by the system. For instance, if a user's history is non-empty, then the subject area in which he searched in the preceding session may be on top of the list of suggested topics. Similarly, if the user is a PhD student, the system can fill in the category 'PhD thesis' as a default, which may be changed by the user.

In order to administrate user and context profiles a meta-data concept is being designed, in which attributes and possible values for all kinds of information recorded as described above can be stored. Here the IMS Learner Information Package [16] will serve as a basis.

Profile of data pools

With each data pool (catalogue) the following types of information are stored:

- the forms of the included publications (e. g., books, articles),
- the subject areas of the included publications,
- the languages of the included publications,
- the range of publication dates of the included publications,
- the time constraints and financial constraints for access to the included publications.

Again, in order to administrate data pool profiles, a meta-data concept is being designed, in which attributes and possible values for all kinds of information described above can be stored.

Suggestion of data pools

Depending on a user/context profile BibTutor is supposed to suggest suitable data pools for literature research. This can be implemented in, for instance, a rule-based approach. Since the system has to be generic and the integration of further data pools into the system

must be possible at any time, generality is an important aspect for this implementation and therefore also for the meta-data concept.

To guarantee the assistance functionalities, it is important that the user first is explained in a clear and succinct way, why the particular data pools have been chosen and second can interact with the system and ask for an alternative catalogue suggestion in case of any discontent with the first choice. Here perhaps a specification of the context profile model might be helpful. In particular, if none of the integrated data pools matches the user's requirements, the system has to provide suggestions concerning which constraints the user could or should mitigate in order to find suitable information sources. Again, explanations on how mitigating certain constraints affects the choice of data pools might be helpful for the user.

Note that for the functionalities of the online research assistant, it is necessary to be able to interpret user and context profiles in such a way that at least certain constraints for the following search process and maybe a category of search strategy (as in Table 1) can be deduced.

Problems we face

When developing more detailed concepts for the context acquisition component, we face several problems that have to be solved.

For instance, how can the system make the user understand that even if he/she is content with the catalogues frequently used before, there might be other, maybe better suitable, catalogues? In particular, other catalogues might require different modes of usage, which—though possibly harder to get used to—can provide a benefit for the user in the current context. So how can the user be convinced to put more effort into changing their usage modes for the sake of better search results?

Another important aspect is the subject area/topic the user needs information about. Usually topic hierarchies in the user's personal perception will differ from the internal taxonomies or ontologies the system choices build upon. How can this problem be tackled?

Moreover, it is not straightforward to formulate the explanations appropriately. In particular, it would be nice to extract enough information from the user profile to be able to adapt the presentation mode of suggestions and explanations (i. e., to provide different formulations depending on whether the user is a novice or an expert, whether he/she needs examples or rather general rules to understand suggestions and explanations, etc.). How should user histories be recorded and exploited to achieve such functionalities?

Can we deduce some kind of helpful feedback for redesigning the system from comparing different user profiles and in particular the corresponding search histories? In general, how can the scenarios of interaction between the system and the user be modelled to achieve an optimal outcome?

All these and similar questions have to be addressed during system development, since the respective functionalities would very much contribute to making BibTutor an assistance system.

2.2 Online research assistant

The online research assistant guides the user's research in the selected data pool. It has to be designed to support interactivity and transparency, thereby in particular respecting the directive of 'helping people to help themselves'. Its main function is to identify search problems, to support the user in resolving them, to give individual and general hints and explanations concerning individual and general search problems, and to impart knowledge on general strategies, components, and tools in the subject area of literature research.

In order to allow the user to keep the 'look and feel' of the research system provided by the library, BibTutor offers a graphical user interface appearing in two frames: (i) a search frame, in which the known web pages of the libraries (and their research systems) are displayed with the usual modes of interaction and (ii) a guidance frame, in which the individual suggestions, hints, explanations, ... are displayed.

Research problems and possible solutions

One main task in the conception of BibTutor is to identify (i) categories of user and problem context that can be recorded and (ii) categories of intended suggestions provided by the online research assistant. These categories have to allow for a generic method of mapping the acquired context on the one side to the intended help on the other side. For such a mapping, again rule-based approaches might be useful.

The top-level categories of search problems might be distinguished as follows:

- The number of hits upon a query is too low.
- The number of hits upon a query is too high.
- The number of hits upon a query is appropriate, but the quality of the hits is not optimal.
- The user does not know how to further handle the list of hits.

Additional categories are conceivable; in particular it will be necessary to further specialise these into sub-categories and to propose a method for classifying a user and context profile into the obtained categories.

Top-level categories of individual suggestions might be:

- a reformulation of the search query (e. g., use a synonym, more or less search terms, or different boolean operators),
- a modification of the criteria used in the search query (e. g., add a year of publication, an author, etc.),
- a change of the underlying catalogue,
- a change of the underlying thesauri or subject catalogues used to form queries,

- a list of possible suggestions from which the user is supposed to pick the most appropriate.

Here again further categories are conceivable and sub-categories are absolutely inevitable. For instance, types of query reformulation must be distinguished depending on whether they generalise, specialise, or just generally readjust the former query. In a next step, sub-categories of the intended methods for readjustment have to be determined. These might for instance include readjustments by

- alignment with subject indices,
- alignment with thesauri,
- modified usage of boolean operators.

In case a mapping of a user/context profile to a bottom-level sub-category of intended suggestion is accomplished, it is moreover necessary to determine one or more promising individual specifications of this suggestion (i. e., one or more concrete query formulations).

Note that an important aspect in an optimal assistance might be the cooperation between the user and the system. That means in particular that sometimes assistance also has to be given by the user in order to allow for the best possible result. For instance, in case a mapping between a user/context profile and an intended help becomes too fuzzy, it might be advisable to let the user choose between several options of concrete query readjustments or several options of research moves, tactics, or strategems.

Learning modules and glossary

In addition to the concrete suggestions on how to modify a search query or a search tactic, BibTutor will offer learning modules depending on the current context. Moreover, the user will have the option to look up special notions related to information science in a permanently accessible glossary.

The learning modules can explain different prototypical search tactics, strategems, and strategies, provide information on different data pools and their usage, or give hints on prototypical ways of utilisation of research results. On a more basic level, they can also explain the general components and functions of catalogues, thesauri, etc. As a generic system, BibTutor will allow the integration of learning modules different libraries would like to provide their users.

Problems we face

Again, the development of the online research assistant yields many interesting questions which can become crucial in the design of an assistance system.

For example, one of the most important questions is how to explicate that much better results might be obtained with a different search strategy even in case the user is absolutely

content with his/her results. How can we make the user realise that different methods can be beneficial in case he/she doesn't feel any need for improvement of the research results?

Another problem concerning the suggestions and explanations to be made is that they have to depend not only on the user and context profile but also on the particular data pool. And again the formulation of appropriate moves, tactics, strategems, and strategies—particularly different presentation modes depending on the user's expertise—will be a crucial aspect of design. For instance, it might be useful (especially when the user has found an appropriate number, but not an appropriate choice of results) to illustrate suggestions for alternative tactics with prototypical examples. The latter may be thought of as some kind of best practice cases in a learning module showing a problem, a sequence of successful search activities and their positive effect. The question is not only when and how to display such best practice cases, but also whether there is a generic procedure for displaying them depending on the user and context profile.

As above, also the question of how to exploit search histories of different users for redesign is relevant. And in general, how can the scenarios of interaction between the system and the user be modelled to achieve an optimal outcome? Especially, how can the system identify when assistance from the user is needed to accomplish a successful search?

Such questions will be crucial in the development of the desired functionalities.

3 Conclusions

The observed problems students and scientists have in acquiring and utilising information, especially literature, which is relevant for their current needs has to be ascribed to deficiencies in crucial skills in information science, particularly in the choice of adequate information sources and in the development of successful search strategies. BibTutor is a system supporting the user in these two tasks during a research process accomplished with existing systems for literature research. By modelling the user and the context, BibTutor is supposed to adapt to individual needs and thus to realise a problem-oriented approach. Because of the transparency BibTutor is supposed to provide, the user will not only resolve search problems but also improve his or her respective skills in information science.

As has been illustrated in the previous section, we will face many crucial questions in the design of the desired assistance system. These concern the interpretation of user and context profiles and their mapping to data pools and problem categories, the mapping of problem categories to intended suggestions and learning modules, the presentation of suggestions, explanations and learning modules, the modes of interaction between the system and the user, and the amount of assistance the user has to provide to the system. Nevertheless, these questions can only be answered after a detailed analysis of the typical user behaviour of standard literature research systems as integrated in BibTutor.

For the conception of BibTutor it is first of all important to identify typical search problems and solutions as well as typical scenarios in which problems occur. Exactly these points in the scenarios, in which problems occur, must be seized and acted upon by the

desired assistance system. The better the scenarios and problem categories are analysed, the better the assistance functionalities can be implemented. On the one hand, in order to keep the system generic, it is very important not to specialise these scenarios too much, but to aim for classifications. On the other hand, in order to be able to adapt to the user's individual needs, the scenarios must be specialised enough.

The identification of the relevant problem and solution categories must be the next step before the design of the components of the system. Since—with four libraries—information scientists are involved in the project the required expertise can be used. After implementation of a first system prototype, hopefully, field-testing in libraries will help to redesign the system such that finally students and scientists will benefit from an assistant guiding them in the use of their well-known tools.

Acknowledgments

Many thanks are due to Bertin Klein, Martin Memmel, and Steffen Lange for their valuable comments, to Stevan Agne and Markus Junker for designing Figure 1 and for providing further background material on the topic, as well as to Gunter Grieser for supporting the presentation of this paper at the PISA workshop.

References

[1] M. J. Bates. Where should the person stop and the information search interface start? *Information Processing and Management*, 26:575–591, 1990.

[2] S. K. Bhavnani, M. J. Bates. Separating the knowledge layers: cognitive analysis of search knowledge through hierarchical goal decompositions. In: *Proc. American Society for Information Science and Technology Annual Meeting*, vol. 39, pp. 575–591, 2002.

[3] M. Buckland, D. Florian. Expertise, task complexity, and the role of intelligent information systems. *Journal of the American Society for Information Science* 42(9):635–643, 1991.

[4] T. Hapke, O. Marahrens. Spielen(d) lernen mit DISCUS: Förderung von Informationskompetenz mit einem E-Learning-Projekt der Universitätsbibliothek der TU Hamburg-Harburg. In: M. Ockenfeld (Ed.) *Proc. Online-Tagung der DGI*, DGI e. V., 203–217, 2004 (in German).

[5] B. Hjørland. Domain analysis in information science: eleven approaches – traditional as well as innovative. *Journal of Documentation* 58(4):422–462, 2002.

[6] K. P. Jantke, M. Memmel, O. Rostanin, B. Rudolf. Media and service integration for professional e-learning. In: *Proc. E-Learn 2004*, pp. 725–732, Washington DC 2004.

[7] R. Klatt, K. Gavriilidis, K. Kleinsimlinghaus, M. Feldmann *et al. Nutzung elektronischer wissenschaftlicher Information in der Hochschulausbildung.* Leske + Budrich, 2001 (see also http://www.stefi.de/).

[8] C. C. Kuhltau. *Information Skills for an Information Society: A Review of Research.* Syracuse, NY: ERIC Clearinghouse on Information Resources, 1987.

[9] J. C. Mancall, S. L. Aaron, S. A. Walker. Educating students to think: the role of the school library media program. *School Library Media Quarterly* 15(1):18–27, 1986.

[10] T. D. Wilson. Information needs and uses: fifty years of progress? In: B. C. Vickery (Ed.) *Fifty Years of Information Progress: A Journal of Documentation Review*, Aslib, London, pp. 15–51, 1994.

[11] http://www.dfki.uni-kl.de/KM/bibtutor/

[12] http://damit.dfki.de/

[13] http://discus.tu-harburg.de/

[14] http://www.ub.uni-heidelberg.de/helios/fachinfo/www/psycho/psyik/

[15] http://lotse.uni-muenster.de/

[16] http://www.imsglobal.org/profiles/index.html