Proposal for a Planning Approach for Information Seeking

César Carranza & Héctor Muñoz-Avila & Frank Weberskirch & Ralph Bergmann
University of Kaiserslautern, Dept. of Computer Science
P.O. Box 3049, D-67653 Kaiserslautern, Germany
e-mail: {carranza|munioz|weberski|bergmann}@informatik.uni-kl.de

Abstract

The information seeking process which includes several phases is a dynamic, iterative, and interactive process where the user plays an important role. Because the first query given by the user normally may be vaguely stated or ill-defined, it often must be reformulated. For avoiding the generation of a new plan when the user modifies the query we propose the use of a planning architecture that combines partial-order planning with dependence maintenance.

The Problem of Information Seeking

Today, information, usually in form of text, is highly distributed and managed by a variety of systems with different retrieval functionalities. Thus, seeking for information in this *dynamic heterogeneous space* becomes a complex activity. Usually information seekers are confronted with the following problems: first, they cannot specify exactly their information needs. The known information changes as a result of the interaction with the world and so does their understanding of the information problem. Their first requests may be vaguely stated or ill-defined and often must be reformulated. Second, the user have difficulties with the different query languages. The search for information is not a one-step process, it is a *dynamic, iterative and interactive process* (Belkin 1996).

Traditional information retrieval systems (IRS) - among other things- search for information in unstructured text-based documents. These systems are normally concerned with access to a single information source with a fixed retrieval technique such as vector-based retrieval (Salton & McGill 1983) and probabilistic-based retrieval (Fuhr 1992). These retrieval techniques are very *efficient* but the *quality* of the results often does not satisfy the user needs. An important reason for this is that they lack of background knowledge that can supply semantic matching between the search terms and the index terms and that they have a fixed problem solving process.

In addition, satisfying the information needs of the user cannot be answered by just accessing one information source but by the combining partial results from different information sources. Moreover, user's experience and knowledge also affects the interaction with the IRS. Traditional IR systems don't consider these issues because they are not *adaptive*.Due to the diversity of information services and sources needed to solve an information problem, the incorporation of an information mediator is needed.

Examples of applications domains are the search for information in heterogeneous bibliography-related sources of mathematical logic (Lenski & Wette-Roch 1996) or another example on the Web could be the search for public-domain PC-Software. In the first application a mediator with additional background knowledge structures for entities inside the documents in the domain of mathematical logic can be developed and in the second a mediator with technical background on public-domain software and the components of PC is required. In recent projects, we have started to investigate the information seeking process in both domains.

Process for Information Seeking

The information seeking process is composed of the following subphases (Carranza & Lenski 1996; Carranza & Lenski 1997):

- Query Formulation Phase. The user expresses the query in the language offered by the information mediator.
- Query Interpretation Phase. The query is analyzed and transformed to the mediator's internal representation based on a domain ontology. Modifications based on this ontology are offered, thus the search terms finally found may not necessarily be identical to the terms given by the user.
- Information Service/Source Selection Phase and the Retrieval Phase. In these two phases the internal representation of the query is analyzed to capture aspects which permit the selection of the suitable information services and sources. A procedural representation of the query equation is obtained, including the abstract data manipulation operators and its execution order by the underlying information services. This so-called retrieval plan is dynamically executed.

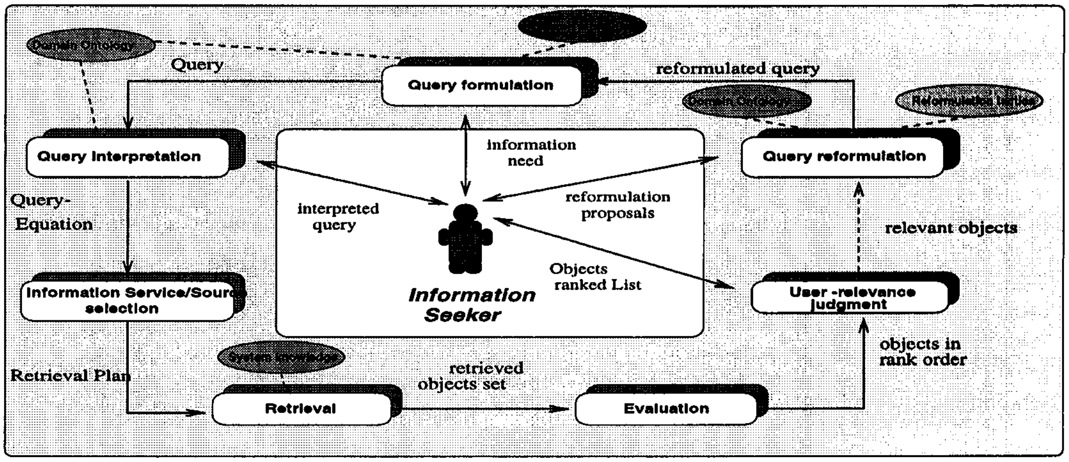


Figure 1: Knowledge-based Information Seeking Process

- Evaluation Phase. The results are combined and ranked according to application-oriented criteria.
- User Relevance Judgment Phase. The user analyse the results according to his information need.
- Query Reformulation Phase. The query is reformulated if the results do not satisfy his information need.

In this context, *information gathering* could be seen as a subpart of the information seeking process (see Figure 1). It begins at Query Formulation, and finishes at the Retrieval Phase. Information gathering considers the interaction between these phases; a plan is dynamically generated in which the information gained from accessing information sources is used to plan further. Information gathering excludes the phases of Evaluation, User Relevance Judgment and Query Reformulation based on user interactions.

As suggested by other authors, the use of planning techniques for the development of such mediators play an important role and is a focus of research of the planning community as reflected in the different approaches for planning for information gathering (Golden, Etzioni, & Weld 1996; Knoblock 1996; Kwok & Weld 1996).

A Model for Information Seeking

Our model for information seeking consists of three levels: *the strategic level*, *the conceptual level* and *the operational level* (Carranza & Lenski 1997). At the conceptual level for each given query, a plan is dynamically generated at the operational level. The results obtained after executing the generated plan may not satisfy the information needs of the user. As a result, he may decide to *reformulate* the query, which requires the generation of a new plan at the operational level (whether the query reformulation is done by the user alone or with the support of the mediator are not exclusive choices.). The user is the only who knows if the mediator's interpretation/reformulation of the query corresponds to his information needs. Notice that the query modification does not necessarily implies the generation of a new plan but only the modification of parts of the original plan that are affected by the changes in the query at the conceptual level.

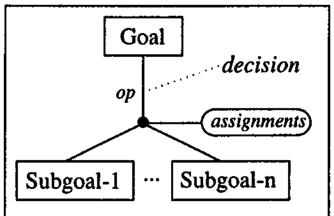
Clearly, planning concepts such as handling incomplete information and integration of planning and execution are required for the development of a planner for information seeking, but in this paper we concentrate on dynamic replanning due to query reformulation.

Dependency Maintenance for Replanning due to Query Reformulation

To avoid planning all over again after the query is reformulated a kind of dependency maintenance is required. Dependency maintenance allows to perform replanning without having to plan from the scratch. In CAPLAN, an architecture combining partial-order planning with mechanisms for explicit representation and maintenance of dependencies between planning decisions has been implemented (Weberskirch 1995; Weberskirch & Paulokat 1995). This architecture allows the user to guide the planning process and change conditions of the problem during the planning process. It has been used also to developed an adaptation paradigm, complete decision replay, that exploits the advantages for replanning of the architecture (Muñoz-Avila & Weberskirch 1996).

For representing knowledge about plans and contingencies that occur during planning, CAPLAN is built on the generic REDUX architecture (Petrie 1992). Key concepts of REDUX are goals, constraints, and contingencies. Planning proceeds by applying operators to goals, which may result in subgoals and in assignments (see Figure 2).

Applying an operator is called a *decision* and represents a backtracking point as different operators might be applicable to a goal. Assignments originally are



(a) decision

Figure 2: The subgoal graph

thought to assign values to variables, more generally, they stand for modifications made in the plan (addition of steps/orderings/constraints). Goals and subgoals build the *subgoal graph*. It represents basic dependencies between goals and subgoals as well as between subgoals and decisions. Basically, REDUX represents validity and local optimality of decisions and dependencies among them (Petrie 1992). In the case of rejecting a decision, this structure is traversed to reject dependent decisions also.

A restriction of CAPLAN is that it considers only classical AI planning concepts. That is, basic concepts for planning for information gathering like run-time variables or functional predicates (Knoblock 1996) have not been represented yet. Currently, we are extending CAPLAN to incorporate these capabilitites and to represent them in CAPLAN and more concretelly in the subgoal graph. From the point of view of the architecture, the following functionality is required:

- the removal of facts, and
- the addition of new facts.

CAPLAN can handle the first one. That is, when replanning has to be done because conditions taken into account to build a plan are declared as invalid. In these situations, by traversing the subgoal graph, decisions made that depend on the invalid conditions are identified and the corresponding parts of the plan are removed. CAPLAN is capable of completing the remaining parts of the plan. The second one is matter of current research.

Discussion

Information gathering is usually concerned with the search for information in structured sources such as databases. We consider search for information in a wider environment, particularly in semi-structured sources containing text. As the user is usually not aware of all the facts necessary to obtain the required information, the information seeking process is a dynamic, iterative and interactive process in which the queries are reformulated. Our proposal is to add dependency maintenance capabilities to the planner so that the plans at the operational level that solve previous queries can be modified without having to plan from the scratch.

The SIMS project studied query access planning and query reformulation for dynamic information integration (Arens, Knoblock, & Shen 1996). In the SIMS project the user's query expressed in domain terms is automatically refined into queries to specific information sources following the mapping between global and local schemes. In information seeking, however, the user changes the information request (i.e., the query) based on the results of the previous queries.

References

Arens, Y.; Knoblock, C. A.; and Shen, W. 1996. Query Reformulation for Dynamic Information Integration. *Journal of Intelligent Information Systems* 6(2/3):99–130.

Belkin, N. 1996. Intelligent Information Retrieval: Whose Intelligence? In *Proceedings of 5th Intern. Symposium on Information Science (ISI-96)*.

Carranza, C., and Lenski, W. 1996. A Planning-based Approach to Intelligent Information Retrieval in Text Databases. In *Proc. of the 20th Annual Conference of the Gesellschaft für Klassifikation*. Springer.

Carranza, C., and Lenski, W. 1997. Planning the Information Seeking Process in heterogeneous Bibliography-related Sources. In *Workshop KI-Techniken für intelligente Mensch-Maschine-Schnittstellen, 21. Deutsche Jahrestagung für KI (KI-97)*.

Fuhr, N. 1992. Probabilistic Models in Information Retrieval. *The Computer Journal* 35(3):243–255.

Golden, K.; Etzioni, O.; and Weld, D. 1996. Planning with Execution and Incomplete Information. Technical Report, Dept. of Computer Science and Engineering, University of Washington.

Knoblock, C. 1996. Building a Planner for Information Gathering: A Report from the Trenches. In *Proceedings of the 3nd International Conference on AI Planning Systems (AIPS-96)*, 134–141.

Kwok, C., and Weld, D. 1996. Planning to Gather Information. In *Proceedings of AAAI-96*, 32–39.

Lenski, W., and Wette-Roch, E. 1996. Foundational Aspects of Knowledge-based Information Systems in Scientific Domains. In *Proc. of the 20th Annual Conference of the Gesellschaft für Klassifikation*. Springer.

Muñoz-Avila, H., and Weberskirch, F. 1996. Planning for Manufacturing Workpieces by Storing, Indexing and Replaying Planning Decisions. In *Proc. of the 3nd International Conference on AI Planning Systems (AIPS-96)*. AAAI-Press.

Petrie, C. 1992. Constrained Decision Revision. In *Proceedings of AAAI-92*, 393–400.

Salton, G., and McGill, M. 1983. *Introduction to Modern Information Retrieval*. Mac Graw Hill Book Company.

Weberskirch, F., and Paulokat, J. 1995. Caplan- a Cooperative Planning Assistant. In Ghallab, M., ed., *Preprints of the 3rd European Workshop on PLanning*.

Weberskirch, F. 1995. Combining SNLP-like Planning and Dependency-Maintenance. Technical Report LSA-95-10E, Centre for Learning Systems and Applications, University of Kaiserslautern, Germany.