A Distributed Problem Solving Approach to Cooperative Information Gathering

Tim Oates, M. V. Nagendra Prasad, Victor R. Lesser and Keith Decker*
Department of Computer Science
University of Massachusetts
Amherst, MA 01003
Email: {oates,nagendra,lesser,decker}@cs.umass.edu

Introduction: The complexity of the modern information carrying landscape requires a sophisticated view in which information is *acquired* rather than simply retrieved; where the process must be dynamic, incremental, and constrained by resource limitations. Information Gathering (IG) is an activity involving pro-active acquisition of information, from possibly heterogeneous sources, in response to a complex query that may require the system to possess capabilities such as reasoning, representation and inferencing. In this paper, we present a model of information gathering designed specifically for such complex environments, a model of *Cooperative Information Gathering* (CIG). In addition to the complexity of query specification, control of the acquisition process may itself be complex and dynamic in IG systems. Traditional Information Retrieval (IR) is a limited sub-case of such information gathering systems in which queries generally map onto static, pre-specified retrieval plans. In this paper, we propose a cooperative agent-based solution for information gathering. Top level queries drive the creation of partially elaborated information gathering plans, resulting in the employment of multiple semi-autonomous, cooperative agents for the purpose of achieving goals and subgoals within those plans. Cooperation between agents implies management of interdependencies between their activities so as to integrate and evolve consistent clusters of high quality information from distributed heterogeneous sources. This paper draws upon a long history of thought in Distributed Problem Solving (DPS) to present a model of this type of cooperative information acquisition.

Distributed Processing vs. Distributed Problem Solving: The task of information gathering in a distributed setting can be viewed in general terms as either distributed processing or distributed problem solving. Distributed processing is appropriate when subproblems are independent and agents need nothing other than local information to arrive at a subproblem solution of the required quality that can be synthesized with other agent subproblem solutions to arrive at a global solution. Distributed problem solving, on the other hand, is characterized by the existence of interdependencies between subproblems leading to a need for the agents to cooperate extensively during problem solving. Lesser [Lesser, 1991] presents the functionally accurate, cooperative (FA/C) paradigm as an approach to distributed problem solving. In FA/C systems, various soft and hard constraining *goal/task interrelationships* among subproblems motivate agents to augment their local information with information about global problem solving activity in order to enhance the efficiency of the ongoing problem solving process. Once uncovered via communication of problem solving activities, such as receiving partial results or meta-information, these interdependencies can be exploited in a variety of ways to improve problem solving both locally and globally. In a CIG task, potentially useful constraints may exist between different pieces of information. The discovery and exploitation of such constraints is necessarily a dynamic and incremental process that occurs during problem-solving and entails communication of partial results among agents in a timely and selective manner, to augment each agent's local view with a more global view. Given the incomplete nature of the local views of the individual agents, another important aspect of FA/C is the explicit recognition of the role of solution and control uncertainty. Coupled with the fact that resources and time for conducting a search are limited in real-life problems, this leads to the notion of *satisficing search.* Another aspect of FA/C is the explicit recognition and exploitation or avoidance of *redundancy*, leading to increased robustness

*This material is based upon work supported by the NSF Center for Intelligent Information Retrieval, ONR contract N00014-92-J-1450, ARPA under ONR contract N00014-92-J-1698. The content of the information does not necessarily reflect the position or the policy of the Government and no official endorsement should be inferred.

or decreased resource demands depending on the context and the structure of the domain.

Cooperative Information Gathering: CIG can be viewed within the DPS framework as discussed above. In response to a query, one or more agents are released onto the network, each responsible for one or more corpora. Each agent treats its information seeking process as a cooperative planning activity. The global solution is the response to the query and it is a composition of the information retrieved and transformed appropriately by domain knowledge in the agents. Problem decomposition involves assigning subgoals to agents. The subgoals assigned to each agent involve seeking information relevant to global goals of the retrieval process. There may be interrelationships between the subgoals assigned to the agents and this may necessitate sharing partial results of their search to enhance the efficiency of the overall retrieval process. Subproblem composition involves combining the returned information into a coherent response to the original query.

We now introduce an example which will be used later to highlight various concepts of our approach. Imagine a user deciding to go on a vacation. She gives a travel planner program a few of her preferences for the vacation. Say she gives her specifications as a vacation for 3 or 4 days around July 20th, preferably in Massachusetts. In addition, either through user specifications or through user modeling, the travel planner knows that she prefers historical sites or nature spots. The travel planner has to plan for at least four different aspects of the vacation — places to visit, weather situation, accommodation and conveyance. So it sends off four agents to deal with the corresponding types of databases. During the process of query planning and information retrieval, the agents have to interact both among themselves and with the travel planner.

Subproblem Interaction: Various kinds of goal interrelationships like facilitates, enables, overlaps and subsumes that exist between subproblems can be detected and exploited in a variety of ways [Decker and Lesser, 1992; Decker and Lesser, 1994]. For example, uncertainty that may arise from incomplete local information can be reduced through detection and subsequent exploitation of overlaps and subsumes interrelationships. Carver et. al. [Carver and Lesser, 1991; Carver *et al.*, 1993; Carver and Lesser, 1994] address the problem of resolving uncertainty in the sensor interpretation domain. However, this process of detecting and exploiting goal interrelationships involves providing the agent with a more complete global view, and that entails communication costs. Hence an agent should communicate only relevant portions of its local view of the problem solving process to help form a more coherent view of the emerging global problem solving process in other agents. Partial solutions and meta-information received from other agents may facilitate (i.e. serve to focus or constrain or lend support to) a local solution or may point to an inconsistency in an agent's local processing or detect redundancies. The end effect is "better" or higher quality global solutions and reduced computational requirements. The constraints arising out of goal/solution interrelationships may also play a crucial role in exploiting parallelism among the agents. For example, an agent with a facilitates interrelationship from another agent can develop a plan in parallel with the facilitating agent. However, upon receipt of relevant results, it may need to iteratively repair or modify its partially developed plan. Alternatively, the agent could perform some other task while awaiting the receipt of facilitating information.

Figure 1 shows some of the subgoal interrelationships in the travel planning query. The planning of flight connections and car rentals has weather report data as an optional goal specification parameter. Acquisition of weather reports facilitates the planning process for car rentals and flight connections by eliminating or attributing low importance to retrieval of flight reservation and car rental availability on those days when the weather is not conducive to travel. Even though Agent 2 can plan for car rentals and flights without the weather data, if there is no time pressure then it is better off delaying planning for the flight schedule until the availability of weather information. In the meanwhile, it can search for airlines offering cheaper deals. Similarly, acquisition of weather reports facilitates planning for outdoor spots. On the other hand, a plan of the places to visit will enable the accommodation agent to start its work on planning and querying for lodging. The place at which accommodations are to be secured is a required goal specification parameter for the "Setup Accommodation" goal. Note that abstractions of plans are all that is needed for the accommodation agent to start its work. Thus, while Agent 2 is fleshing out the details of the abstract plan of the places to visit, Agent 1 can, in parallel, start its work. Similarly, a favors interrelationship says that once you have made the effort to design a plan to go to Concord, a plan for going to Lexington is obtained by minor modifications to the plan for Concord. An overlaps interrelationship says that the two agents involved may be doing similar work and can hence benefit by sharing their partial results.

Our multi-agent Case Based Reasoning (CBR) system called CBR-TEAM[Nagendraprasad *et al.*,] is another example of a sophisticated system where a consistent case is iteratively assembled from distributed case bases through the exchange of locally

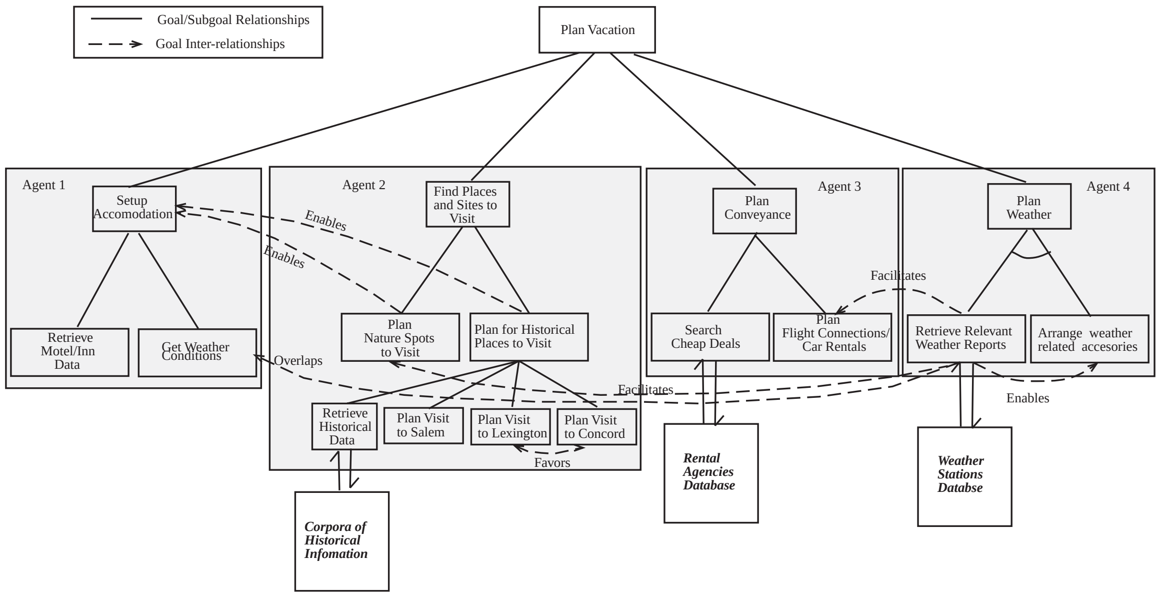

Figure 1: Goal Tree for the Vacation Planning Example

violated requirements by the set of agents involved in the process. CBR-TEAM is directly relevant to the information gathering model we propose. Information acquired by an agent can be related to the requirements of information acquisition in another agent. Viewing partial results as information relevant to a query opens up a rich set of possible subproblem interrelationships that may be exploited. Figure 2 shows an example that highlights the same issues in the document retrieval domain (modified from [Decker, 1994]). For a given query, there may be many sources of relevant information. Product reviews often exist on-line, or may obtained from publishers for a fee in paper or electronic format. Relevant reviews may be found on-line in the review section of the TidBits newsletter, in the Info-Mac archives, or in discussions about the product in Usenet news groups. The query may be satisfied by dispatching agents to locate the required review and then retrieve it. Each agent may employ different access methods (such as WAIS, FTP, HTTP, telnet, etc.), and the access methods may have recourse to the same information at a variety of physical locations (such as the main TidBits archive ftp.tidbits.com or its various mirrors). Interrelationships exist between some of the goals of the agents involved. Locating a paper review "enables" its retrieval, i.e. paper reviews may be obtained by first finding a citation, and then either finding the actual article or obtaining it from the publisher. Finding a citation via Uncover "facilitates" the goal of getting the article faxed to the user. An overlaps interrelationship exists between Agent1's "Get from Seller" goal and Agent 2's "Use Uncover" goal. This is due to the fact that once an agent accessing the seller's archive finds a particular citation, Agent 2 can avoid the search for that same citation at the Uncover database.

Satisficing: Although the amount of information available on the Internet is seemingly boundless, the resources like time, money and computational resources available to search that information typically are not. Rather than being able to develop an exhaustively complete and accurate response to a query, intermediate results from disparate sources must be pieced together to form consistent islands of information that can be incrementally refined to form a more accurate solution depending on the extent of available resources and time. That is, the information gathering process must be *satisficing* along various dimensions like precision, quality, etc. If communication is slow, we may access nearby data with low expected quality first, rather than trying distant data sources of higher quality that may require more time than is available. When more time is allocated to the search process, the scope of the search can be broadened to include higher quality sources, while retaining some amount of effort on inexpensive low quality sources.

Redundancy: Redundancy in distributed search, either in the form of replication of data at multiple sites or the possibility of deriving the same conclusion from different sets of data, raises a host of issues. Advantages of redundancy include increased robustness of the system in environments with failure-prone components and increased flexibility in responses. Redundancy can play a role in the reduction of uncertainty when dealing with erroneous

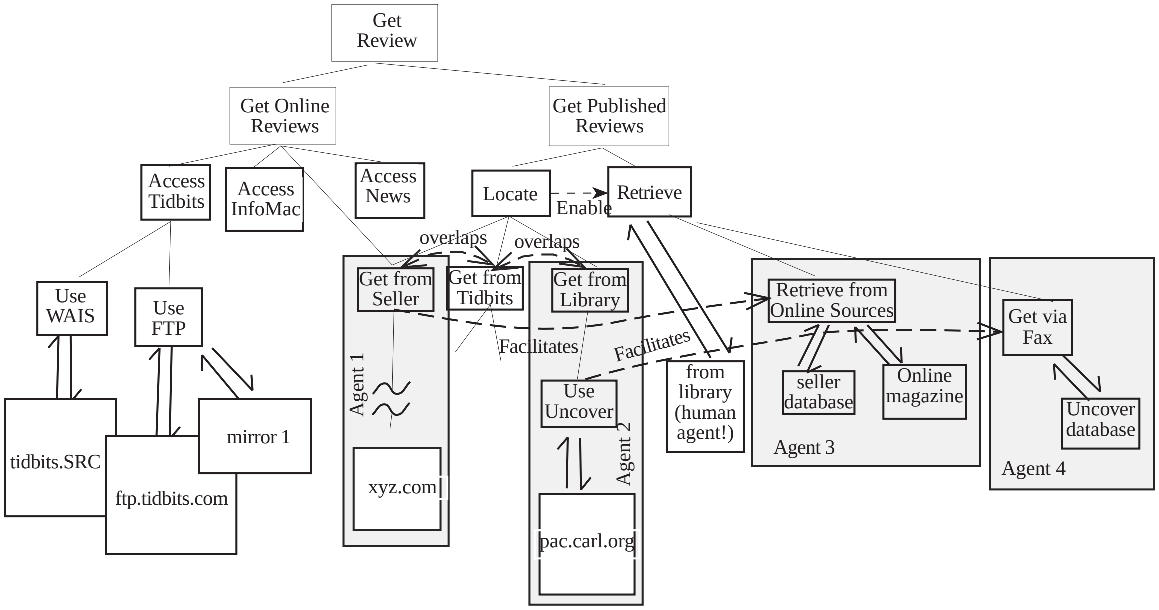

Figure 2: A goal tree for retrieving Macintosh related product reviews

or incomplete information. On the other hand, redundancy has the disadvantage of increased resource usage and possibly increased total processing times. For example, the Internet may contain “mirror” sites for certain data repositories or it may contain redundant data from different sources for the same task. Data from different sources may be of different quality or may be differently organized. A particular task could possibly do with low quality data which perhaps could be locally acquired. Thus, recognizing the role of redundant data and computation could be important for exploiting the possibilities such redundancy offers in a CIG system. Redundancy could be permitted if the control costs outweigh the benefits of avoiding it. Alternatively, if we are dealing with faulty systems or poor quality data, redundancy could help resolve the uncertainty in the retrieved data by providing additional constraints.

Conclusion: We believe that distributed information retrieval tasks characterized by complex, heterogeneous and unstructured data environments, can be viewed as a distributed problem-solving task within the FA/C paradigm. The benefits of such a view not only stem from the fact that it provides a comprehensive conceptual model for the myriad of methods being proposed for intelligent information retrieval (IIR), but also from the fact that the view provides a direct map from the wealth of existing methods in Multi Agent Systems (MAS) to the distributed information retrieval domain.

We hope that this paper encourages IR system designers to take a radically new view of information gathering as a distributed problem solving activity and to develop IR systems that are appropriate for that model. Such IR systems must be able to concurrently, asynchronously, and incrementally gather information from a variety of sources, employing a range of access and search methods; they must be able to handle new constraints, revising decisions based on the arrival of new information or the status of a partially completed search. Information Gathering systems proposed in the literature typically either do not fully exploit the potential of knowledge-intensive methods for the task [Bowman *et al.*, 1994; Callan, 1994; Huhns *et al.*,] or tend toward distributed processing, failing to exploit the dependencies between agents working on different aspects of an information gathering task[Knoblock and Arens, ; Vorhees,]. Cooperating to enhance efficiency of a resource-limited information acquisition process or negotiating to dynamically resolve conflicts and inconsistencies in the acquired data, leading to further search or retrieval, may be important aspects of IG systems in the future. Our model is an initial step in this direction. We also suggest that existing methods in MAS can serve to leverage future implementations of IG systems based on this view.

References

Bowman, M. C.; Danzig, P. B.; Manber, U.; and Schwartz, M. F. 1994. Scalable internet resource discovery: Research problems and approaches. *Communications of the ACM*.

Callan, J. 1994. Personal communication.

Carver, Norman and Lesser, Victor 1991. A new framework for sensor interpretation: Planning to resolve sources of uncertainty. In *Proceedings of the Ninth National Conference on Artificial Intelligence.* 724–731.

Carver, Norman and Lesser, Victor 1994. A first step toward the formal analysis of solution quality in FA/C distributed interpretation systems. In *Proceedings of the Thirteenth International Workshop on Distributed AI*, Seattle, WA.

Carver, Norman; Lesser, Victor; and Long, Q. 1993. Resolving global inconsistency in distributed sensor interpretation: Modeling agent interpretations in DRESUN. In *Proceedings of the Twelfth International Workshop on Distributed AI*, Hidden Valley, Pa.

Decker, Keith S. and Lesser, Victor R. 1992. Generalizing the partial global planning algorithm. *International Journal of Intelligent and Cooperative Information System s* 1(2):319–346.

Decker, Keith S. and Lesser, Victor R. 1994. Designing a family of coordination algorithms. In *Proceedings of the Thirteenth International Workshop on Distributed A I*, Seattle, WA. AAAI Press Technical Report WS-94-02. 65–84. Also UMass CS-TR-94-14.

Decker, Keith S. 1994. *Environment Centered Analysis and Design of Coordination Mechanisms.* Ph.D. Dissertation, University of Massachusetts.

Huhns, M.; Mukhopadhyay, U.; Stephens, L. M.; and Bonnell, R. up . Dai for document retrieval: The minds project. In *Distributed Artificial Intelligence.*

Knoblock, C. A. and Arens, Y. ens. An architecture for information retrieval agents. In *Working Notes of the AAAI Spring Symposium on Software Agents.*

Lesser, V. R. 1991. A retrospective view of FA/C distributed problem solving. *IEEE Transactions on Systems, Man, and Cybernetics* 21(6):1347–1363.

Nagendraprasad, M. V.; Lander, S. E.; and Lesser, V. R. up . Retrieval and reasoning in distributed case based systems. Unpublished working paper.

Vorhees, E. M. ees. Software agents for information retrieval. In *Working Notes of the AAAI Spring Symposium on Software Agents.*