Information Agents:
a new challenge for AI

Daphne Koller and Yoav Shoham
Computer Science Department
Gates Building 1A
Stanford University
Stanford, CA 94305-9010
{koller,shoham}@cs.stanford.edu

April 2, 1996

The goal of the Stanford University Digital Libraries Project (see `http://www-diglib.stanford.edu/`) is the creation of the *Information Bus,* an open-ended infrastructure that supports multiple protocols for search, information exchange, payment, and other mechanisms necessary to effect transactions in a networked information environment. The infobus and its associated services draw on multiple areas of computer science and beyond. Here, we focus on one service-level functionality, and argue that it provides a new challenge to AI research.

One of the striking characteristics of AI is the spectacular breadth of the field, and one aspect of this breadth is the scope of problems tackled. These problems run the gamut from the very specific to the epic. This breadth is, to us, one of the main attractions of AI. The ultimate grand challenge, creating machine intelligence that rivals that of humans, is AI's *raison d'etre;* it lies at the heart of our excitement about the field. At the same time, the focused technical work is an essential ingredient in achieving quantifiable progress and laying solid foundations; in other words, making AI a science.

Recent years have seen significant advances in specialized areas of AI, including reasoning under uncertainty, knowledge representation and reasoning, planning, and machine learning, to mention but a few. At the same time, it has been suggested that these advances have sometimes come at the expense of "keeping an eye on the prize" (Nilsson, 1995). The potential danger is both cultural — the field might lose its sense of direction — and pragmatic — it is often difficult to integrate the results of different isolated threads of specialized research.

So-called "grand challenges" are a common device used to reconcile detailed technical work with long-term vision. A grand challenge is useful for this purpose if it is not so specialized as to be of interest only to a sub-community (for example, solving the frame problem in the Knowledge Representation sub-community), but not so ambitious as to be "AI-complete" (such as passing the Turing test). The domain of digital libraries offers a new spectrum of useful challenges, one of which is the creation of robust *information agents.* Broadly speaking, an information agent is a piece of software that operates in an ongoing fashion in the domain of networked information services and clients, represents the interests of some entity (typically, an end user or a service provider), and does so with a high degree of autonomy, exposing the human only to task-level decisions and information. We believe this challenge to be at the right level for stretching the boundaries of AI techniques. On the one hand, the relatively circumscribed domain, and the ability to test out new ideas in software without worrying about physical sensors and actuators, serve to bring the solution to this problem within reach. On the other hand, attaining this degree of autonomy in a real-life domain — one which is large, complex, uncertain, dynamic, and involves multiple human and artificial agents — poses a significant challenge to AI technology.

In fact, the complexity of the task seems overwhelming at first. Even the most fundamental task of an information agent — to go out and gather information that is of interest to the user — seems hopeless. After all, we do not even have a reasonable approach for specifying the user's wishes, let alone a methodology for deciding whether a document (or a collection of documents) satisfies those needs. And even if the agent could evaluate the value of a document, how should it go about finding and retrieving 'good' documents? We believe that recent advances in various subfields of AI can come together and provide a solution to this problem.

Consider the problem of evaluating the value of a given text document to a given user. Clearly, the straightforward approach of first understanding the content of the document, and then matching the content with a faithful model of the user's interests, is fraught with daunting problems, not the least of which is the age-old problem of natural language understanding. However, recent work indicates that we can do pretty well at this task without even attempting to understand the document. We can simply treat our lack of understanding of the document and the user preferences as a source of uncertainty, and act accordingly. This perspective allows us to reformulate our goal in a more tractable fashion: The agent should build a probabilistic model of the user's preferences (one which can be adapted over time using *relevance feedback*) and uses that model to evaluate different documents. Recent advances in machine learning, derived from a closer interaction and synergy with probability theory and statistics, allow us to construct such models and reason with them.

Of course, such a model can only be used to estimate the *probability* that a given document is relevant to the user. Thus, the information agent must deal with uncertainty when deciding how information should be gathered. This is particularly relevant given the fact that information agents will be exchanging real money, and we do not want our information agent to pay a lot of money for a low-relevance document. Nor do we want the agent to have to obtain the user approval before each document is retrieved. This type of problem calls for the user of decision theoretic notions and tools such as expected utility. The same techniques can be brought to bear on the more complex problem of deciding where to go and look for the information in the first place. To do that, the agent would need to maintain some representation of the content of the different information sources. Here, uncertainty about the documents contained in the information source augments the uncertainty about the content of each of them. Thus, we need a knowledge representation language that can adequately represent this type of information. This seems to require the expressive power of (a sublanguage of) first-order logic as well as of probability theory, calling for an integration between the two competing, yet complementary, thrusts of work in knowledge representation.

Given a probabilistic representation of the different data sources, the agent still needs to decide how they should be accessed in order to answer a particular query. By viewing each information access as a possible action, whose results depend on the information content of the source, we can represent this task as a problem of planning under uncertainty. The recent work on *decision-theoretic planning* seems to provide a promising starting point for this kind of reasoning.

The discussion so far presents the domain of digital libraries from the perspective of a single information agent interacting with its environment. This is clearly a simplified model. In the world of digital libraries, we will have multiple agents, representing both consumers and producers of online information. These information agents will have to interact autonomously on behalf of their users. As we mentioned, these agents will be exchanging real money, so that we want them to make economically sound decisions. Again, decision-theoretic agents will be able to apply economic and game-theoretic models for dealing with multi-agent interactions. These tools include various mechanisms such as auctions, bidding schemes, and negotiation strategies. Finally, there will be the need and opportunity to conduct system-level analyses of the multiple agents, investigating the roles of system-wide constraints on behavior, whether these are imposed by the library administration ("artificial social laws") or arrived at dynamically through gradual adaptation ("artificial social conventions").

In short, digital libraries offer a wonderful novel intellectual playground, and within it information agents are a powerful new challenge, combining elements of several areas in AI, including knowledge representation, reasoning about uncertainty, and multi-agent systems. We believe AI has the right set of tools to tackle this challenge, and that the next few years promise to be a lot of fun.

N. Nilsson (1995), "Eye on the prize (artificial intelligence)," in *AI Magazine* (Summer 1995) **16**(2), pp. 9–17.