Information Integration Agents: BargainFinder and NewsFinder

Bruce Krulwich

Andersen Consulting Center for Strategic Technology Research

3773 Willow Road, Northbrook, IL 60062

e-mail: krulwich@cstar.ac.com

Phone: (847) 714-2406

Abstract

This paper presents a class of intelligent agents called *Information Integration Agents.*   These agents are particularly well suited to application on the Internet, and can be used to satisfy a wide range of needs. We discuss two prototype Information Integration Agents that have been deployed on the Internet.   One, the BargainFinder agent, has been active for over 9 months amidst considerable interest from Internet users and the mass media.   BargainFinder performs comparison price shopping among a number of on-line CD stores. The second, the NewsFinder agent, is currently being tested internally.   NewsFinder retrieves on-line news articles, matches them against user profiles, and transmits them via portable ubiquitous displays (currently alphanumeric pagers).   We discuss these agents and the class of Information Integration Agents in general, and conjecture that agents of this sort will be extremely valuable to a broad spectrum of Internet users.

1. Introduction

The explosion in the use of the Internet's World Wide Web, and correspondingly in the volume of information to which it provides access, has led to a widely acknowledged need for intelligent tools and services to provide effective and efficient access.   The most common and least intelligent tools of this sort are the search engines that provide keyword search capabilities to a large segment of the information on the Web.   While search engines are proving critical to day to day use, they have significant limitations.

First, search engines require users to characterize their goals in terms of keywords or other search expressions. This is a very difficult task for many users, and is particularly difficult in domain areas with large vocabularies where documents may exist that are relevant to the user's goal but do not use the particular words selected. This limitation ha spurred a number of research initiatives that are developing intelligent agents to address this problem. The most common approach is to have an agent learn an interest profile for a user, relieving the user of the need to construct search strings [Holte and Drummond, 1994; Knoblock and Arens, 1994; Levy *et. al.,* 1994; Pazzani *et. al.,* 1995; Krulwich *et. al.,* 1996].

A second limitation with search engines, largely ignored by the research community, is that search engines and other so-called "spiders" can typically only access documents on the web that are published directly on the Web. A growing body of information and documents, however, are available on the Web through systems that access databases and respond to direct user queries. For example, many on-line stores and information services store their data in databases and provide access channels through forms that users fill out and submit. Additionally, a growing number of news providers are storing their articles in databases and retrieving them in response to user demand. Typically documents and data of this type are not available to search engines.

This paper presents a class of intelligent agents designed to address this problem. *Information Integration Agents* are designed to access heterogeneous information sources on the Web and integrate their results to fit user needs. This enables users to focus on their own tasks without needing to consider large numbers of access methods and procedures.

To motivate and introduce Information Integration Agents we will first discuss a particular such agent called BargainFinder [Krulwich, 1996; currently available at http://bf.cstar.ac.com/bf] that performs comparison price shopping among a number of on-line CD stores. BargainFinder has been active on the Web for over nine months, has been used over 200,000 times, and has been written up extensively in the mass media (Wall Street Journal, Information Week, Economist, Advertising Age, Wired, and others). We present BargainFinder and discuss how it integrates information from multiple sources to fit user needs.   We then discuss Information Integration Agents in general, and follow this with a discussion of another prototype agent called NewsFinder. We conclude with thoughts of how Information Integration Agents such
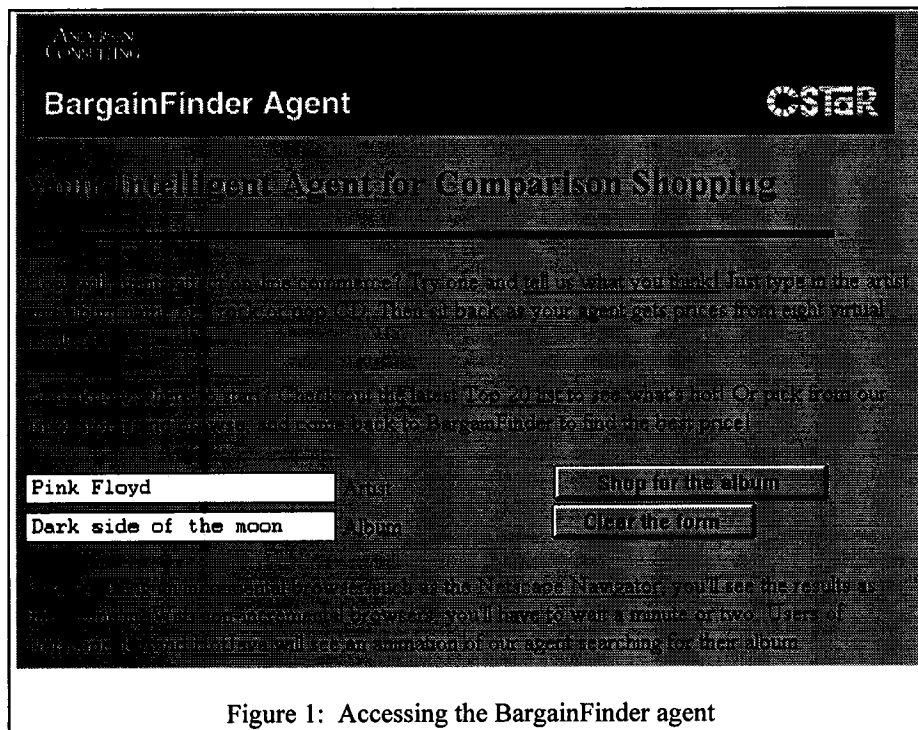
Figure 1: Accessing the BargainFinder agent

2. The BargainFinder agent

The Internet's World Wide Web is growing in popularity as a forum for electronic commerce. New on-line stores open almost every day, infrastructure for electronic malls, marketing, and secure transactions are becoming increasingly available, and new companies are getting involved each day at every step in this process. Internet stores operate on the World Wide Web by presenting a "home page" to users entering the store's on-line location (URL). These home pages combine text, pictures, and in some cases animation or sound, to present the store's image and service options. In typical stores these options include the ability to search for particular items of interest, to see a list of new or notable items, or to get other information of interest to the various market segments targeted by the store. Many stores offer the ability to search for items by name, genre, price, and so on.

Due to the size and complexity of the Web, however, most users have a difficult time discovering on-line stores, and have no access to information about the relative merits of each store. Because of this, stores can flourish with relatively high prices if they publicize their availability well. This is especially true of stores that carry esoteric items not found in real "brick and mortar" stores, since

as these can address the needs of a broad range of Internet users.

customers will tend to quickly purchase a hard-to-find item without investing the effort to find it in other, harder-to-find, stores. If the store provides a lot of on-line information about the album, such as reviews and album cover pictures that will further entice the customer, the electronic point of sale is virtually locked.

If, however, perfect information were to exist about the stores and their prices and services, and were accessible easily and cheaply, customers would be free to buy only from low prices stores, and could in fact do so after having taken advantage of information provided by higher-priced stores. They could shop at one store, using of the store's system of reviews, ratings, detailed information, and recommendations, and then buy the albums they find at another store whose prices reflect less investment in top-quality systems.

To test the applicability of these hypotheses on the Internet, and to measure the reactions of consumers and merchants to such perfect information, we built and deployed an electronic intelligent agent for automatic price-gathering and comparison shopping. This agent, called BargainFinder, takes the name of a particular album, searches for the album at nine Internet stores, and returns to the user a list of the prices found. After the search, the user can select one of the stores and be taken electronically into that store directly to the album. Options then exist to get more information, look for other albums, or to actually buy the CD.
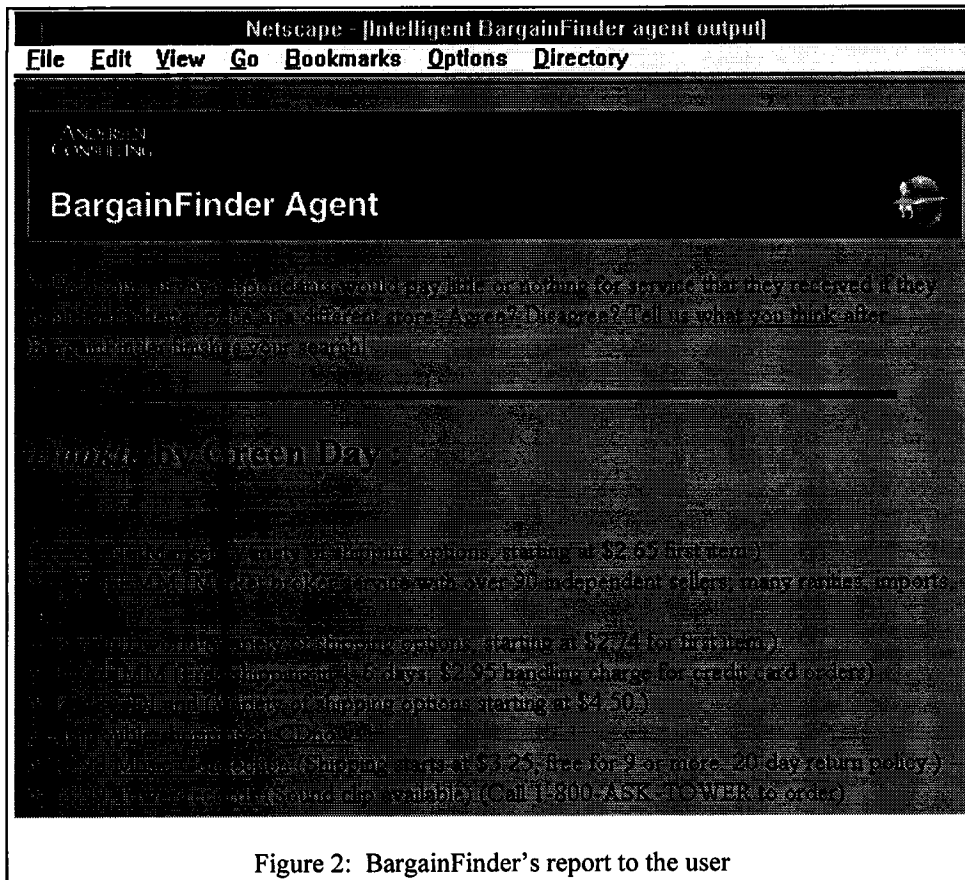
73

Figure 2: BargainFinder's report to the user

Figure 1 shows the World Wide Web entry point to the BargainFinder agent. Customers enter the band and album names in which they are interested, and BargainFinder searches for the album. If the customer wishes to browse prior to price-comparison, they can access a list of the on-line stores that are available, each of which can be accessed directly from the store list. At the bottom of the page (not shown) the customer has access to a survey, an on-line bulletin board, and a short article discussing the issues that BargainFinder raises.

After the customer fills in the desired band and album names and pushes the "shop for the album" button, BargainFinder initiates its search of the available on-line stores. As the agent receives information back from the stores, it extracts the prices and displays them, one line at a time, on the customer's screen. A complete set of returned prices is shown in Figure 2. If the user selects any of the highlighted store names (shown in blue) they are taken directly to their desired album at the particular store, automatically completing the particular store's entry and search forms.

BargainFinder facilitates the customer' ability to comparison shop in a number of qualitative ways. First, and most obviously, the agent searches all of the stores in parallel, and reports all the prices in roughly the same time it would take the customer to search a single store. Second, the agent knows about and searches a larger number of stores than most on-line shoppers are aware of, and in this ways the agent serves as a form of marketing. Third, the agent knows enough about each store's Internet presence to bypass the introductory graphics and explanatory text of the stores and go straight to the search systems, and it can then take the users straight to their album without taking time for other pages.

3. BargainFinder experimental results

The BargainFinder experiment began publicly on June 30, 1995, and was used over 100,000 times in its first two months. It was released on the Internet with significant attention from the media[1] and on-line Internet referral

---

[1] *The Wall Street Journal* (June 30, 1995, page B2; August 16, 1995, page B1), *Information Week* (July 24, 1995, page

groups, and was being used over 2,000 times per day within a week of its release. No advance warning was given to the stores that the agent searched, enabling us to measure their reactions as well as that of consumers. Users of the system were offered a survey to fill out, and an electronic bulletin board was established for discussing the agent and its ramifications.

The most significant measure of the effects that agents like BargainFinder will have on the market is how customers responded. Our hypothesis, as discussed above, was that customers would utilize the information and other services offered by higher-priced stores, and then use BargainFinder to find a better price at a cheaper store with that offered less service. Unfortunately, the limitations of user tracking on the Internet kept us from proving this conclusively, and while our user base included a large number of people interested in seeing an agent like BargainFinder but not actually interested in shopping for a CD. Nonetheless, we feel that we have demonstrated this phenomenon taking place.

The primary measurement of user activity that relates to this hypothesis is the user's selecting to proceed from BargainFinder's output to one of the stores. These selections indicate that an overwhelming percentage of the users selected to go to the cheapest store on the list, without influence from other factors such as position of the store in the output list. Moreover, a majority of the users that proceeded to the second-lowest priced store did so only after having gone to the cheapest. Also interesting in this regard is user response to a survey taken after BargainFinder's output is presented. Out of over 4,200 survey responses, 76% seem to say that they had no moral or ethical compunctions about benefiting from the services provided by one on-line store and then purchasing the items from another on-line store from whom they got a better price. While the selection base of this survey is arguably self-selecting, both as Internet users interested in agents like BargainFinder, and further as those interested in answering a follow-up survey, nonetheless this result makes a statement about the relevant ethical feelings of a significant number of Internet users.

Another goal that we had was to measure the response by the on-line stores themselves. Their options include cooperating with BargainFinder, combating it (by refusing to answer its queries), or remaining neutral. After eight months of operation, one of the eight original stores has stopped operating, three are refusing queries from the agent, one is cooperating actively, and three are remaining

89), *Advertising Age* (July 31, 1995, page 8), *Financial Review* (July 30, 1995), and others.

neutral. Additionally, seven stores have approached us to be added to the system and offered to cooperate. One of these has been added, and is cooperating actively, and the others are pending.

There are two reasons that stores have given for blocking BargainFinder's requests. The first is that they don't want their store evaluated on price alone. In particular, stores that invest heavily in their infrastructure and pre-sale service will, as we have discussed earlier, tend to have higher costs and thus have to charge higher prices. Stores of this sort are motivated to block out BargainFinder's queries.

The other reason that stores have blocked BargainFinder queries is due to the computational or economic costs imposed. Because of the agent's popularity, it gets more queries than many of the stores do, effectively doubling the number of queries handled by the participating stores. Furthermore, many of BargainFinder's users are more interested in seeing the agent than in purchasing a CD, so the value to the stores may be low. The cost can be high, however, in terms of the computation required to handle the agent's queries and the resulting slowdown of the rest of their systems.

4. Information Integration Agents

BargainFinder is an example of what we are calling *information integration agents*. These are intelligent agents that help users manage large amounts of information by integrating it into a form that is more useful. They typically get information from a large number of sources, integrate it into a format that is useful to the user, and give the user the means to get more information from the original sources as necessary. In the case of BargainFinder, the agent gets prices from each of the on-line stores, integrates it into the price listing shown in Figure 2, and presents it to the user. The user can then take follow-up actions such as going to one of the stores.

The general design of an information integration agent like BargainFinder is shown in Figure 3. As we have discussed, the agent takes a user query and formulates nine corresponding queries to make to the individual stores. The stores all return web pages with large amounts of information on product price, availability, and detailed information, and the agent simply extracts the prices and integrates them into a single output page. The agent then enables the user to follow-up by transporting directly to the particular album at his choice of stores, as we've discussed.

Figure 3: Information integration agent
architecture
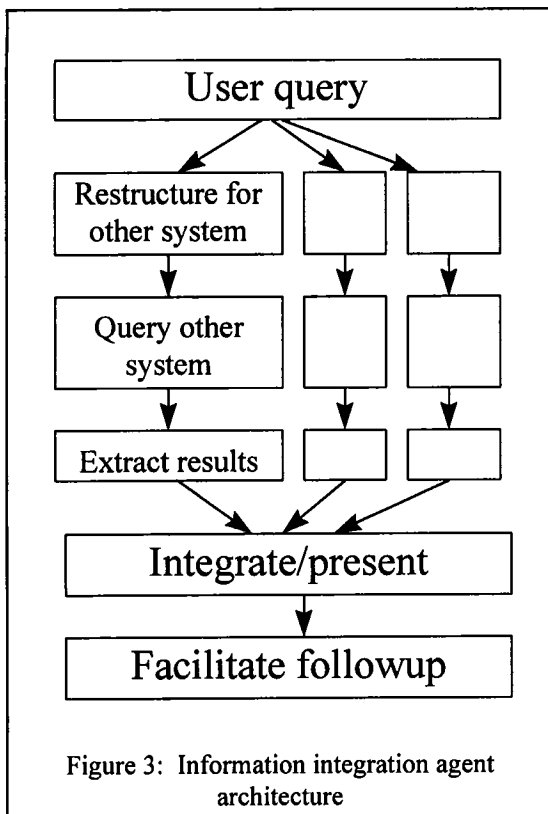
In general, information integration agents offer wide opportunities to help users manage the large amounts of information and on-line services available on the Internet today. While most approaches to solving this problem have been in the form of improved and more intelligent search engines, an information integration approach offers several advantages. First, information integration agents can be designed specifically for user tasks, and not require that the user transform the specific problem at hand into a search query. Second, as we discussed above, information integration agents can use interactive information sources who's information can not be searched using standard means. BargainFinder has both of these advantages, in that it is designed around the task of CD shopping, not web page searching, and that it gets its information from interactive stores who's price data is not open to standard search engines.

5. The NewsFinder agent

To explore the application of the Information Integration Agent approach to information management (as opposed to electronic commerce), we have developed a prototype agent called NewsFinder. NewsFinder maintains profiles of the interests of its users and routinely sends them new

gathered from on-line sources that match their interests. Because these articles are typically stored on the Web for a very short amount of time, and because they often have complex access protocols, they are most often not accessible through standard search engines. Furthermore, timely access to these articles requires that users navigate a large number of articles, at a large number of media provider sites, each with a different access method, to see the current news. For these reasons an Information Integration Agent approach appears promising.

NewsFinder retrieves all available news on a regular basis (currently hourly) by spawning processes for each of the on-line news sources with which the system is familiar. Within the architecture from figure 3, NewsFinder structures queries for each of the on-line news sources, invokes the query, and extracts individual articles from the responses. Next, the system identifies those articles that have not been seen previously. These articles are then matched against user profiles, and those that match are sent to the users directly. This allows users to have access to these articles virtually as they appear on the Web, and to have them through a channel that can get to them easily without navigating on-line access protocols.

To maximize the effectiveness of this approach, NewsFinder is exploring the use of ubiquitous display devices, currently two-way alphanumeric pagers, for information communication. These pagers, available from SkyTel and other sources, allow wireless communication of up to 500 characters of each message to users located anywhere nationwide. This allows NewsFinder to truly keep users up-to-date with currently breaking news. Users that want additional information can use the Web to access NewsFinder directly, and can be directed to the full multimedia contents of the article of their interest.

Most importantly, users can use the two-way feature of the pagers to tell NewsFinder which messages fit their interests and which do not. NewsFinder then learns improved profiles of the user's interests using the approach embodied in our InformationFinder agent [Krulwich and Burkey, 1996]. In this way the user can refine the selection of news delivered and continue to receive up to date news messages from on-line sources, all through the ubiquitous device.

While NewsFinder is still in early testing, initial results appear promising. Our continuing research is investigating how the user can initiate NewsFinder directly from a pager using the 2-way response feature. Further research and testing will determine the effectiveness of NewsFinder for keeping users informed of up-to-date news. This is made

possible by the Information Integration Agent approach that we have adopted.

6. Summary and conclusions

We have presented a new class of intelligent agents for the Internet called Information Integration Agents. This approach appears promising for providing easy access to a wide range of Internet resources, both in the realm of information and of electronic commerce. We conjecture that as the volume of information on the Web expands, and as the use of databases and programmatic access to information continues to increase, agents of this sort will prove increasingly critical to effective use of the Web. Future research and prototype development will demonstrate the degree to which this is true, and the areas of application for which Information Integration Agents are most valuable.

References

Holte, R. and Drummond, C., 1994. A learning apprentice for browsing. In *Working Notes of the 1994 AAAI Spring Symposium on Software Agents,* Stanford, CA, pp. 37-42.

Knoblock, C. and Arens, Y., 1994. An architecture for information retrieval agents. In *Working Notes of the 1994 AAAI Spring Symposium on Software Agents,* Stanford, CA, pp. 49-56.

Krulwich, 1996. The BargainFinder agent: Comparison price shopping on the Internet. In *Agents, Bots and other Internet Beasties,* SAMS.NET publishing (a division of Macmillan USA), to appear 1996.

Krulwich, B. and Burkey, C., 1996. Learning user information interests through the extraction of semantically significant phrases. In *Working Notes of the 1996 AAAI Spring Symposium on Machine Learning for Information Access,* Stanford, CA.

Levy, A., Sagiv, Y., and Srivastava, D., 1994. Towards efficient information gathering agents. In *Working Notes of the 1994 AAAI Spring Symposium on Software Agents,* Stanford, CA, pp. 64-70.

Pazzani, M., Nguyen, L., and Mantik, S., 1995. Learning from hotlists and coldlists: Towards a WWW information filtering and seeking agent. http://www.ics.uci.edu/~pazzani/Coldlist.html.