Digital libraries and software agents

Digital libraries will need digital, or software, agents to perform many important tasks such as communication with the user, acquisition of new information and maintenance of the links to that information (assuming that it is external to the structure of the digital library). Fortunately software agents are a diverse breed and there are many different types of agents which can be used to perform those tasks. Personal agents can be used for the user interface, along with metacrawlers and perhaps searchbots. Autonomous agents will also be needed for the acquisition of information and maintenance of links to external information. Each agent, or type of agent will need some form of intelligence as well, so it can recognize important or relevant information when it encounters it.

User Interface:	Information Acquisition:	Maintenance:
Personal Agent Metacrawlers Searchbots (CIG) Autonomous Agents	Autonomous Agents Communicative Agents	Autonomous Agents

Metacrawlers are very similar to web crawlers, or other search engines, which search for information and return results to the user. Metacrawlers contact other crawlers, or metacrawlers and then use them to search. The difference is that metacrawlers can take less specific search strings and manipulate them into the correct queries for different crawlers for the user. The metacrawler may also decide which search engines are the most likely to have the information that you asked for.
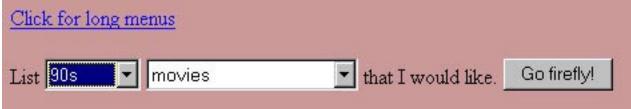
Metacrawlers share many similarities with an initiative called Cooperative Information Gathering (CIG) which uses agents to reformulate the query entered by the user and contact search engines with the new query. The agents can then take the results from the new query and analyze them, using them to reformulate another query or decide which other search engines to contact. While CIG is in a rather primitive stage it shows

much promise and along with metacrawlers will make information gathering significantly less painful in the future.
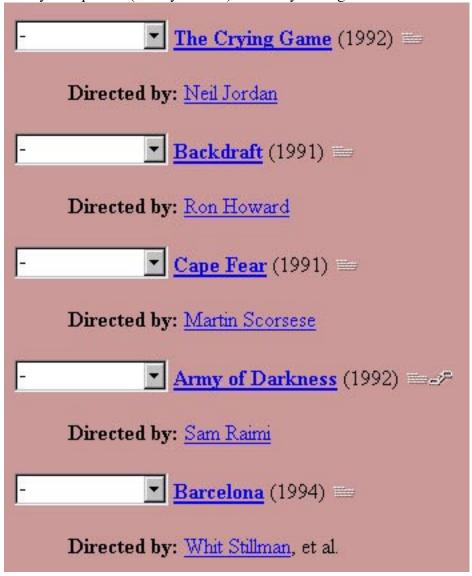
The user interface presents many issues, some of which have been addressed by the draft of UIUC's digital library specifications. Most of the issues center around the user being able to customize their interface and be able to define their own mini digital libraries. This is a perfect example of where a personal agent would fit into the digital library. The personal agent would be what interacts with the digital library, through its customized interface. The user would be hidden from the library and vice versa. Agents Inc. has developed this sort of personal agent interface already.

Firefly, what Agents Inc. has made available to the public, is "your own personal agent" which can figure out what movies or songs you might like the most, determine which users you might want to interact with and has a fully customizable interface. In addition to those abilities firefly also saves a listing of all the sites recently visited and displays them. You can create your own page inside of firefly and add links or images to it. In a very real sense your own page is very similar to your own digital library with access points that you define. Firefly demonstrates the power of a personal agent as an interface.

A test of Firefly's guessing ability:

Click for long menus

List [90s ▼] [movies ▼] that I would like. [Go firefly!]

Firefly's responses (mostly correct) after only having been trained on 80's movies:

The Crying Game (1992)

Directed by: Neil Jordan

Backdraft (1991)

Directed by: Ron Howard

Cape Fear (1991)

Directed by: Martin Scorsese

Army of Darkness (1992)

Directed by: Sam Raimi

Barcelona (1994)

Directed by: Whit Stillman, et al.

In addition to a personalized interface a personal agent could act as a agent for multiple people so that, for example, all staff members might have the same interface. The restriction would be that the interface would not be customizable. If, however, the user wished to customize that interface they could give it a unique name and then customize it.

Autonomous agents, often also called knowbots, will be needed for the user interface. Knowbots are often found in the service of people who want to stay up to date

on a variety of subjects but do not have the time to read through all the available news resources to find what they are looking for. They are let loose in the database of news articles or on the Internet, and allowed to roam around looking for articles or objects of interest to the user. When an article is found its importance is determined and then it is sent to the user, depending on how important it might be.

There are several knowbots available which, for a small fee, will search the news services for you and return those articles which it determined were of interest to you. Depending on how you specified your interests to the knowbot they can be quite handy. Some services have users "train" the knowbot before it is set loose. Instead of relying on rules, or other specifications given to it, the knowbot develops its own specifications after watching the user wade through the news articles and select which ones were of interest.

There are other kinds of autonomous agents, however, which are not knowbots. There are "spiders" which search a specified part of the WWW and find links which have moved. These agents will also be needed in the digital library, searching out dead links and either fixing them, if possible, or sending mail to an administrator. These agents could even send a knowbot out to look for whatever resource used to be at the other end of the dead link.

Agents similar to those mentioned above are already being used both commercially and in the public domain. As time has passed agents have become more and more popular as information filtering services for those who do not have time to wade through the increasingly chaotic Internet, and now the WWW.

In the commercial sector AT&T has a service called PersonaLink which is essentially a knowbot service where users can configure knowbots which then watch the news feeds or whatever it is that they are configured to watch and report back to the user when something interesting is found. PersonaLink is a service that is accessible through AT&T as well as through the personal communicators of both Sony and Motorola.

AT&T is able to provide this service through the use of Telescript |, which is a product of General Magic. Telescript allows for interactive, autonomous, communicative agents. The agents have to exist on servers which are running the Telescript API however. Telescript is a highly integrated package, including security, control over system command, etc. and has been around for quite a while.

There is another company which provides a complete agent package as well. Personal Library Software (PLS) has recently released a suite of software which incorporates distributed searching, metacrawlers, autonomous, communicative agents and full text natural language searches. PLS, like Telescript, can and does operate on the WWW and uses the WWW for information resources. To be part of the PLS network, it is of course necessary to be running the PLS server.

Both General Magic and PLS have well integrated agent - based systems which are on the WWW. PLS, however, offers much more functionality than Telescript. In the public domain there is even less functionality, but some interesting software development is being done by the TACOMA project. There is also some promise for a Java| agent which would be autonomous.

The TACOMA project is a joint venture between Cornell University and the University of Troms→, Norway. The only type of agent that can currently be sent anywhere for processing is a weather agent which is sent off to Norway (from the client browser) to be processed. Users who are "trusted" can create their own agents to be sent of to be processed in Norway. The only way to be process a TACOMA agent is to be running the TACOMA package on your web server. This system seems very much like the Telescript system by General Magic except less refined and totally free.

The Java agent template, currently in version 2.0 is also free. This agent template allows for autonomous agents which can communicate using a KQML, a messaging standard developed as part of the ARPA information sharing effort. This shows great promise in that Java applets do not need to have a server running some sort of package

for them to be executed. Since Java applets can be executed anywhere this gives them much more freedom of movement than any of the other systems. This of course limits what they can do on the server or client that they end up on, as there would have to be security restrictions. According to the readme there is however a bug with the agent template "Netscape will not currently allow Agents to create ServerSockets. Therefore, you can only execute AgentFrame and ANSFrame objects as stand alone applications or via the classes AgentApplet and ANSApplet using the AppletViewer." This means that there are serious problems with the network part of the agent.

Most of the agent functionality which is needed for a digital library is available. The problem is that there are competing agent systems and all of the functional systems require a specific type of environment for the agent to be able to exist. This is unfortunate but perhaps the only way that agents will be able to live on the WWW.   The user interface requirements appear to be able to be met by a system similar to Firefly, searching via metacrawlers is already implemented on the WWW. Knowbots are prevalent as information finders and information filters. MOMSpider is a working utility that will check links to see if they have moved. The only element that would seem to be missing would be the glue that
puts all of these elements together into a whole. Unfortunately Java is not developed enough to be that glue and the only systems which are require a monetary contribution and use of proprietary software.

Some agent resources for the digital library

UMBC, Intelligent Software Agents	http://www.cs.umbc.edu/agents/
SICS, Intelligent Software Agents	http://www.sics.se/isl/abc/survey.html
NetWatch	http://www.pulver.com/netwatch/topten/tt9.htm
MetaCrawler Parallel Web Search Service	http://metacrawler.cs.washington.edu:8080/
SavvySearch	http://www.cs.colostate.edu/~dreiling/smartform.html
CIG Searchbots	http://dis.cs.umass.edu/research/searchbots.html
Firefly	http://www.ffly.com/
Telescript	http://www.genmagic.com/Telescript/index.html
AT&T PersonaLink Services	http://www.att.com/personalink/
Personal Library Software News Release	http://www.pls.com/news/@1plweb.html
Overview of TACOMA project	http://www.cs.uit.no/DOS/Tacoma/Overview.html
TACOMA Agents in a WWW Context	http://www.cs.uit.no/DOS/Tacoma/Webdemo.html
Java(tm) Agent Template	http://cdr.stanford.edu/ABE/JavaAgent.html