search insider

In this third, and final, article on Internet searching (previous articles appeared in *IWR* 141, November 1998 and *IWR* 146, April 1999), David Green's Web wanderings reach the erudite world of intelligent agents.

Faced with the increasing difficulty of finding information on the Web, the larger and more established search engines – such as Excite or the search engine provider Inktomi – have been exploring new technologies to provide additional functionality.

Excite is looking at ways of providing 'serendipitous information': ways of suggesting links that the user may not have originally considered. Inktomi is looking at how it might give search engines the power to search by concept and not just by key word, although progress is slow as the huge amount of computer-processing that would be required to do this is similar to that needed to produce predictive global weather models.

While it is likely that search engines will, undoubtedly, eventually become better at the statistical analysis of search results and the refinement of search strategies, such developments do not go anyway towards enabling the search engine to develop an understanding of the user's requirements. But not to worry, there is an existing – and rapidly expanding – technology which complements search engines by learning to understand your requirements, even as they change over time.

Bots introduced

Intelligent agents, or bots as they also called (short for robot, derived from the Czech word robota meaning work), are generating an escalating interest that has resulted in the evolution of several different types, including datamining bots, search bots, shopping bots and chatter bots. Several of the key Web search technologies that have emerged in the last year or so incorporate intelligent agents. These include natural language searching (eg The Electric Monk), search engines that analyse the link structure of the web (eg Google) and some search utility programs such as Mata Hari and BullsEye.

So what exactly are intelligent agents? Aren't they just another software program? Well, no. Unlike a standard software program that will execute specific functions within clearly defined parameters, agents:

* are **adaptive**, meaning that they can interpret monitored events to make appropriate decisions;
* are **self-organising**, meaning that they assimilate both information and experience; and
* can **communicate** with both the user and other bots.

In order to function, agents need first to understand and then to organise data. 'Understanding' comes from using a statistical approach – such as latent semantic indexing. This enables the agents to draw metadata from information and to view separate information as 'objects'. This approach has many advantages, including greater language independence (BullsEye can search through information in 12 languages), resistance to 'noise' (such as spelling mistakes) and the support of mathematical analysis of results. Using this approach, agents can provide a coherent organisation of disparate information types and unstructured data, such as Word documents, as well as different types of structured data.

Bots defined

There is no universally accepted definition of an intelligent agent and this is, perhaps, a reflection of the wide range of agent types and applications in existence. Nevertheless, a common function of agents is that they allow the user to specify a high-level goal instead of issuing explicit instructions. In this way, the 'how' and 'when' decisions are left to the agent. A common limitation is that they are task-specific – they are incapable of dealing with broad expectations. This results in a need for different agents for different tasks.

The broad church of disciplines that contribute to agent/bot theory takes in artificial intelligence, psychology, mathematics and computing.There is a correspondingly broad range of technologies in use in agent development:

Function	Technologies
Profiling	neural networks/fuzzy logic
Matching	genetic algorithms/classification and indexing
Filtering	collaborative filtering

Whose bots?

Three of the leading developers of intelligent agents are Autonomy, Verity and Muscat. While each is developing agents to perform similar functions (information retrieval and organisation), the resulting agents are based on different technologies.

Autonomy uses a 'pattern matching technology' that was originally developed to assist fingerprinting and signature recognition for British police and intelligence services. This technology relies on identifying patterns, extracting concepts from a large body of information or providing a context for information. US competitor **Firefly**, which was acquired by Microsoft last year, bases its products on the same technological concept.

Verity's agents work using a 'knowledge mapping' technology that was originally developed for the CIA. This develops 'rules of »p26

→p25 evidence', generates 'families of topics' and then maps 'domains of knowledge'.

Muscat – which is 70 per cent owned by The Dialog Corporation – uses the concept of 'user-determined relevance'. It attaches different weightings to different words and allows the user to tick subsets from the main data set during a search to suggest related terms.

While we are used to hearing of US firms dominating Internet developments, the UK is actually an important player in this arena. One interesting UK company to look out for is **CK Ltd**. Named after its director, Chris Knowles, it was formed as a result of a management buy-out of a Mitsubishi Electric subsidiary which developed intelligent agents for searching databases on the invisible web.

There are also a number of products, such as Brightware, NeuroStudio and Agent Builder, that allow programmers to develop intelligent agents for their own specific purposes.

Bots in action

Agents can provide a uniform interface for search queries across different sources and are true 'infomediaries', in that they can identify and search appropriate resources which may or may not be known to the searcher. The adaptive element of intelligent agents is central to the functionality of many search products that incorporate agents.

Autonomy's Agentware Daily Briefing monitors hundreds of Web sites, newsfeeds and chat areas. It seeks news and information that match the interest profile of the user and automatically collates this information to produce a personalised results page, including hyperlinks, for each licensed user. Shell UK is one of the companies that uses Autonomy products, in its case to track what others are saying about it on the Web.

IBM have a product called Webby that can help you find sites of interest from your surfing habits while WebWatch by Canadian firm eWatch will, at least twice a day, check competitor sites for you, allowing you to enter important words or phrases to avoid being notified of mundane changes. Such functionality is far more useful than the basic 'change notification' feature on Microsoft Internet Explorer as it produces a report of exactly what the changes are.

The following search utilities are all available both as freebies and as more comprehensive 'pay-for' versions:

Mata Hari can learn one set of power-search commands and then automatically translate these for each search service/database that it queries.

BullsEye Pro incorporates 11 different intelligent agents to conduct what it calls 'Web mining', a term which covers its information analysis, filtering and automatic analysis, and refinement of search statements functions. The different agents are used to target specific types of information in over 450 sources on both the visible and invisible Web. It allows users to view results by site type, relevance or in 'concept clusters'. It also provides automatic document summaries, live highlighting, and active hyperlinking of query words in the retrieved documents. Like most agents it can search across a wide range of document types and formats. It will run searches automatically, allows the import and export of searches to other users and allows users to chose to receive change alerts by HTML email, pager or other hand-held data devices.

As a result of the huge success of its search utility **Copernic**, Canadian company Agent Technologies Corp changed its name to Copernic Technologies Corp! Like Mata Hari, Copernic can translate a search statement for different services and will simultaneously submit the query to these search engines, Web directories and databases. It incorporates about 20 categories – such as Business and Finance or Science – with relevant sites for each category already defined.

Bots in KM

The ability of intelligent agents/bots to search across data in unstructured format, to automatically learn and adapt to user preferences and even to identify patterns makes them ideally suited to knowledge management. Agents are integral to the technology aspect (as opposed to the people or process aspects) of KM in helping organisations to 'know what they know' and in integrating complementary or relevant data from external sources, thereby providing a more comprehensive picture of the topic being investigated.

All the search utilities discussed above can be used with intranets. Indeed Autonomy's Agentware Knowledge Server provides automatic XML tagging, categorisation and hyperlinking of documents.

The increasingly widespread integration of XML-support into Web publishing products, browsers, search engines, agents/bots and a wide range of other software makes Dialog's LiveIntranet product look not only remarkably, and speedily, redundant but also increasingly incompatible with current Web trends. Why pay for seamless categorisation of internal data with external data from one publisher when, if you wait a little while, you can have it for free from everyone? XML is becoming a useful, widely adopted standard, just like HTML.

Conclusion

Intelligent agents represent the most advanced approach yet to searching the ever-burgeoning and increasingly complex Web. Even though XML, by providing a uniform and structured data format for the visible and invisible Web, will allow faster, more accurate and easier searching, the role of intelligent agents will continue to grow, as organisations increasingly rely on the Web as their primary means of communication.

Finally, while databases and scripts are used to dynamically *produce* Web pages, there is further scope for intelligent agents to allow Web sites to dynamically *restructure* their pages to match the user's requirements more closely, by analysing previous and current request patterns and anticipating what they may request next. One agent would assist with the identification of relevant Web sites for your information requirements. Once you visit a site, another agent could assist with the speedy presentation of information to you.

Very intelligent.

Useful links:

• Autonomy	www.autonomy.com
• Muscat	www.muscat.com
• Verity	www.verity.com
• Botspot	www.botspot.com
• Mata Hari	www.thewebtools.com
• Bullseye	www.intelliseek.com
• Copernic	www.copernic.com