Student Work   Vol.4(1) March 2002

Intelligent agents (bots) - are they trustworthy?

S. de Wet
Postgraduate Diploma in Information Management
Rand Afrikaans University
infosci@rau.ac.za;
sdewet@mj.org.za

Contents

1 Definition of an intelligent agent (bot)

'Bots' (short for robots, from the Czech word 'robota' meaning 'work') are computer programs that run automatically as stand-in workers in the place of humans (What's a bot? 2000). Pallman (Proffit 2001) defines a bot as equivalent to an automated program, or 'any software that doesn't require constant human intervention.'

Bots are sometimes called agents, and the two terms are used interchangeably. Most writers seem to agree that agents are bots with specific features. Webster's New World Dictionary defines an agent as 'a person or thing that acts or is capable of acting or is empowered to act for another'. The term has become popular in computer science, although there is as yet no common definition:

- Delicato et al. (2001) define agents as software that performs tasks for the user, usually with autonomy, playing the role of personal assistants. They quote the definition given by Caglayan and Harrison, saying that a software agent is a 'computer entity that performs tasks delegated by the user in an autonomous way.'
- Pallman (Profitt 2001) defines an agent as software that serves a specific body, whether it is a person, department or organization. Agents are aware of the needs and wishes of their masters and are skilled in specialized areas.
- Botspot (What's a bot? 2000) considers an agent to be a bot that goes out on a mission, usually to find information and report back, rather than operating in one place, for example a bot in Microsoft Front Page that automates work on a Web page.

Intelligent agents have further distinguishing features such as the ability to analyse their

users' choices and to adapt their responses. Intelligent agents are the result of research into artificial intelligence (AI). AI is an advanced form of computer science that aims to develop software capable of processing information without the help of human direction. The ultimate goal is to make computer programs with a 'human'-like capacity for problem solving and goal achievement, but this does not necessarily imply anthropomorphism in the software.

Intelligent agents are distinguished by their ability to automate tasks with a minimum of human intervention. Pallman (Profitt 2001) considers intelligence to be a necessary adjunct to agents because the public responds badly to automation without intelligence.

Stenman (1998) offers this definition of an intelligent agent: 'An autonomous agent is a system situated within and part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future.'

'Bot' is therefore the broader term, with agent as one of its subcategories that may be further subdivided into intelligent agents. In short, an agent will always be a bot and may or may not be intelligent, but not all bots are agents.

2 Operation of an intelligent agent

There is some consensus (but also differences) in the references consulted about the general features of intelligent agents. Table 1 lists the features as discussed by three writers: Delicato et al. (2001), Stenman (1998) and Foner (1999).

Table 1 Features of intelligent agents

Feature	Delicato et al. (2001)	Stenman (1998)	Foner (1999)
Autonomy	Autonomy: The agent's capacity to control its own actions.	Autonomous: The agent must have control over its own actions and be able to work and launch actions independent of the user. Agents should not be started or stopped for explicit tasks, but rather be a continuously running process.	Autonomy: Without it the agent is just a glorified front-end, irrevocably fixed to the actions of its user. To pursue independent action requires periodic action, spontaneous execution and initiative.
Intelligence	Intelligence: The degree of reasoning and learned behaviour; the agent's ability to accept the user's statement of goals and carry out the task delegated to it via mechanisms such as relevance feedback.	Reactive: The agents can detect changes in its environment and react to it by answering to events and initiating actions. Agents learn and change their behaviour based on previous experiences.	Personalization: The purpose of an agent is to enable people to do different tasks better. An agent must be educable in the task at hand. There should be components of learning and memory.
Communication	Social capacity: The agent's ability to	Communicative: The ability to interact and	

	communicate with others agents and users.	communicate with users and other agents.	
Mobility	Mobility: The agent's ability to move in an environment. A mobile agent is capable of being transported from one machine to another during its execution.	Mobility: The agent's ability to move itself from one machine to another and across different architectures and platforms.	
Purpose		Goal-driven: Agents have a purpose and act in accordance with that purpose until it is fulfilled.	Discourse: The assurance that the agent shares our agenda and can carry out the task the way we want it done. This generally requires a discourse with the agent and something resembling a contract about what is to be done.
Cooperative		Agents can be separated in user-oriented agents that focus on the individual and collaborative agents - the cumulative knowledge of several cooperative agents.	Cooperation: The user and the agent are essentially collaborating as peers. The user specifies what actions should be performed and the agent specifies what it can do and provides results.

3 Web research with intelligent agents

Intelligent agents have many uses, such as the ability to design other bots and intelligent agents, to 'chat' with humans or to make commercial transactions. They operate in different subject domains, such as news, finance and education. Botspot (The list of all bots 2000) lists 15 such categories of bots.

Bot activities mostly capitalize on their ability to 'dig' through data. As such, they are relevant to anyone interested in information search and retrieval. Users give a bot directions and it brings back answers. Although not restricted to the Web, bots have gained great importance there, for example the crawlers that Web search engines send to compile their databases. Bots may perform similar functions for individuals and organizations. The Internet offers a large amount of information to a wide range of users, making it difficult to deal with it. By locating and filtering information on the user's behalf, intelligent agents claim to be a useful tool in reducing the amount of information with which the user has to deal. It also claims to obtain better results over time.

Botspot (The list of all bots 2000) categorizes such agents as 'search bots' and includes many intelligent agents in its list. These agents are capable of the most efficient, state-of-the-art searches on the Internet, which they conduct autonomously, as directed by a user, adapting as they search.

Delicato et al. (2001) identify two types of search bots:

- Information retrieval agents are used to satisfy short-term information needs in a single session.
- Information filtering agents are used in repeated interactions over multiple sessions for long-term goals. They assist users by filtering the data stream and delivering the relevant information. Botspot (What's a bot? 2000) refers to filtering as data mining - the process of finding patterns in enormous amounts of data through persistent searches.

Delicato et al. (2001) and Stenman (1998) both supply lists of the processes that the user and agent follow, illustrating the specific abilities of an intelligent agent in providing Web information to a user. These are described in Table 2. Delicato et al. (2001) is describing a specific system (Fenix) while Stenman is generalizing.

Table 2 Specific abilities of an intelligent agent in proving Web information to a user

Task	Delicato et al. (2001)	Stenman (1998)
Interact	A graphic interface to interact with the user for registration, as well as a choice of agent from three options namely to create a new agent, to load an existing one or to activate the autonomous mode.	Interface agents add presentation ability to systems and may also add features such as speech and natural language understanding.
Profile	When creating a new profiling agent, the user must choose a name, provide the search parameters (keywords linked by 'AND') and the maximum number of documents to be shown.	Profiling agents are used to build dynamic sites with information and to provide recommendations tailored to each visitor's individual taste and need.
Search and retrieve	The agent is responsible for starting the execution of search tasks, one for each search profile. Using different search engines, it searches Web pages for documents containing the user's keywords. It saves documents in a local database.	Retrieval agents search and retrieve information and serves as information brokers or documents managers.
Filter	The documents obtained undergo a filtering process. The agent selects documents with the highest degree of similarity to the profile, classifies them, eliminates repetitions and presents them to the user.	Filtering agents are used to reduce information overload by removing unwanted data (i.e. data that does not match the user's profile) from the input stream.

Recommend	The learning method adopted by the system is relevance feedback. A successful first search is difficult and it is common to perform interactive searches and reformulate query statements. The relevance feedback method automatically generates improved query formulations in the form of a modified profile, based on positive or negative feedback from the user on the documents delivered.	Recommender agents are usually collaborative, as they need many profiles to be available before an accurate recommendation can be made based on the user's previous behaviour.
Navigate		Navigation agents are used to navigate external and internal networks, remembers short cuts, preloads caching information and automatically bookmarks interesting sites.
Monitor		Monitoring agents provide the user with information when particular events occur, such as information being updated, moved or erased.

4 Criteria for intelligent agent Web search products

Three products, namely Bullseye (http://www.intelliseek.com/prod/bullseye/bullseye.htm), Copernicus (http://www.copernic.com/products/index.html) and LexiBot (http://www.lexibot.com/index.asp) were evaluated against the above seven criteria to review existing intelligent agent features in the marketplace. The products were also compared in functionality.

4.1 Interface
The function of an interface is to assist the user. All three products had graphical user interface (GUI) front-ends with familiar menu, toolbar and help features. All three became intuitive with use, although the initial learning curve with Bullseye was quite steep. All three catered to novice as well as advanced users, with preset quick-start features and customizable options and configurations. This allowed more complicated approaches for advanced users. LexiBot excelled at assisting users with tutorials, on-line help, user messages and right-click menus for context-sensitive choices. Single interface design, whether used with desktop databases or the Internet, meant there was no need to move between programs. The products had built-in browsers, as well as the choice of using one's default browser. Only Copernic appeared to work on a platform other than Windows, allowing installation for Macintosh computers. ILIAD (http://prime.jsc.nasa.gov/iliad.html) provided intelligent, selective access to Internet information through a simple low-cost e-mail interface for users who had no Web access or were vision impaired. Searches were submitted and then performed off-line. The search results were e-mailed to review at one's

convenience.

4.2 Profiling
This is where search and display parameters are set, as well as tracking schedules.A bot must be able to address specific user needs. New users should be assisted with features like the Search Wizard from Copernicus. For more detailed searching, all three products provided topical categories that were subdivided into actual resources within the subject field - the equivalent of directory listings in the major search engines. The more choices users have, the better they can direct their search. Bullseye offered 14 subject areas; LexiBot had 60 with over 600 individual Web sources, while Copernic had 93 in the Pro version. All three allowed users to add their own resources for searching. It should also be possible to remove sources, which it seemed Bullseye could not do.

Products must allow sophisticated query options with simple text (keyword) or structured (Boolean) queries. Products must support all Boolean operators, contextual searches with phrases in quotation marks, morphing for similar words, stemming, etc. There should be a universal query format for all searches, whether of the Internet or the desktop database. A history of queries should be kept. LexiBot accepted natural language queries and offered a spell check.

Display options should include choices on how to display (e.g. total time of downloading), what to display (e.g. document size and number retrieved per engine), where to put the search results (e.g. estimated size of database created) and the language of pages.

Tracking options should allow the user to set scheduled automatic searches and specify the notification method.

4.3 Retrieval
The issues here are where and what the product is capable of searching, what volume of the Internet is covered and the speed and accuracy with which the product operates. The products were all meta-search engines, accessing multiple search engines simultaneously. Bullseye used a collection of some 700 search engines, Copernic retrieved from all the most important search engines on the Web - 80 in the free Basic version and up to a 1000 search engines in the Plus version - and LexiBot categorized 600 search engines into 60 topic areas for easy selection.

Both Bullseye and LexiBot were able to search the invisible or deep Web with LexiBot claiming access to 2200 deep Web searchable databases. Most bots could search more than just html text while others just covered document types such as PDF format or multimedia.

Excalibur Internet Spider (www.excalib.com/products/ispi/ispi.html), a member of the Excalibur RetrievalWare product family and known as the industry's first multimedia Web crawler, could also deal with text documents.

EuroSeek (www.euroseek.net), considered Europe's premier agent and bot-based search engine, offered 24 languages and was the first true multilingual search engine.

GIF Runner (www.softwaresolutions.net/freewareplus/gr.htm) located targeted animated GIFs on the Internet with a keyword search and then displayed and saved them.

Isearch (www.cnidr.org/ir/isearch.htm) searched through large amounts of text, not with keywords or an abstract but by searching every word of every document. This allowed greatly improved chances of discovering new information in old collections.

Simultaneous searching of one or many search engines should be possible with groups. LexiBot was able to run multiple searches concurrently. Depending on the speed of the user's connection, searches could be very quick through simultaneous links, as in Copernic. In LexiBot, a simple graphical control (a slider bar) set searches to be either fast and superficial or slower with higher quality results. The bot also claimed the ability to access up to 150 sites simultaneously.

4.4 Filtering
This feature allows information to be presented in a useful way and should allow for both agent and user manipulation. Filters are automatically and dynamically applied to results. Duplicate retrievals, sites that are not responding and files that are too big are automatically filtered out. A status bar tells the user how the search is progressing, how many sites have been found and how many discarded. Results of all the sources searched are consolidated into one list and relevance-ranked according to the profile set, for example by term frequency. The best bots combine more than one scoring method. LexiBot offered five. Results should be displayed in informative lists, including title, URL, search engine used, etc. LexiBot offered a 'terms folder' that listed all the terms in all the documents retrieved to help users evaluate the suitability of the keywords they used.

When a user chose to see a particular hit, Copernic and LexiBot both offered a preview of the page from a cached source, saving the time of going to the live site. In BullsEye, search terms were highlighted in the text, which made it easy to view a term's context and to locate key paragraphs. LexiBot translated between languages while Copernic had a translation tool that triggered a link to www.gist.com, an on-line translator site.

4.5 Learning
Intelligent search agents should be able to adapt to a user's input. The most common way to achieve this is through the ability to search within a previous search and through user relevance feedback, which triggers re ranking of retrieved results. For instance, LexiBot allowed the user to select a particularly good reference and re-ranked results based on his or her profile. It also offered a 'more like this' option, generating a new query profile based on the selected document. Searches should be saveable for future reference and re-runs.

4.6 Navigating
This allows the users some functionality in managing the hit list and information or documents (pages) retrieved.

In LexiBot, all results could be sorted, annotated, deleted and re-ranked. Status pictures showed whether pages were unread, read, annotated or used in re-ranking. Information or URLs could be saved in various formats, for example as text databases or auto-generated Web pages for off-line viewing and sharing. These could be indexed separately for easy off-line retrieval. Partial or complete listings of document results could be shared. Copernic offered no management utility for downloaded files while Bullseye had a partnership with the WIM Surfsaver.

4.7 Monitoring
The 'freshness' of information is a key concept in information management, and bots can contribute to this. Bots must be able to update themselves and the user's information. Copernic could quickly update its search engine list and categories, and had an automatic link checker. Users could update their information by scheduling auto re-runs of searches and specifying how they should be notified of results, for example by e-mail. Bots should be programmed to react to changes in text, images, links or other data.

5 Conclusion

Intelligent agents have distinct advantages in end-user information retrieval for finding, tracking and managing information on the Internet. Agents are designed to make a Web researcher's life easier and more productive. Their great advantage lies in their time-saving capabilities and reach. Users can find relevant information without having to actively search. Bots can save labour because they persist in a search, refine it as they go along and repeat set searches at scheduled times without user intervention.

These products have all the advantages of meta search engines, but go a lot further, for example their reach into the invisible Web and multimedia formats, as well as their delivery of the results as a single ranked list. Deep Web databases can only be retrieved through a direct query and bots are able to make dozens of queries simultaneously, rather than laboriously one by one. They are both faster and smarter.

However, search results are still only as good as the query and the search strategy of the user and intelligent agents do not change this. Bots can assist by providing subject categories and verified resources. Their ranking algorithms can try to wean out irrelevant information and assist in modifying searches. Tools such as translators and link filters spare the user some frustration. Information management and sharing tools are a bonus and there are many good products available. Bots are not an unqualified success. On the negative side, bots raise serious privacy issues and can interfere with Web server operations. They can also be computer viruses. As a result, search engine developers have formed standards on how robots should behave and how they can be excluded from Web sites. However, many positive factors favour the development of bots, particularly as they are desperately needed to contend with the huge amount of information and data that exist on the Internet (Proffit 2001). Foner (1999) discusses a number of criteria for determining whether or not the information retrieval task should be entrusted to a bot. The idea of an agent is intimately tied up with the notion of delegation, which implies relinquishing control of a task. Thus we open ourselves up to certain risks, weighed against trust. The domain of interest is crucial. If the domain is a game or a social pursuit, most failures of the agent have relatively harmless consequences. On the other hand, one may think twice about using such a system, say, for the control of a nuclear reactor. If most of a task can still be accomplished (as opposed to failing to accomplish any of the task), there is generally a better outcome, and it raises the user's trust in the agent's performance.

In a world where tasks, goals and means constantly change, perfect operation could nevertheless lead to disappointment. Agents are most useful in domains where graceful degradation and the correct balance between risk and trust can be obtained (Foner 1999).

6 References

Delicato, F.C. et al. 2001. Fenix: personalised information-filtering system for WWW pages. *Internet Research: Electronic Networking Applications and Policy*, 11 (1): 42-48. [Online]. Available WWW: http://www.emerald-library.com.

Foner, L. 1999. What's an agent, anyway? [Online]. Available WWW http://foner.www.media.mit.edu/people/foner/Julia/Julia.html.

Proffitt, B. 2001. Copernic 2001 hits shelves. [Online]. Available WWW: http://botspot.internet.com/news/news032001 2.html.

Proffitt, B. 2001. LexiBot provides powerful tools. [Online]. Available WWW: internet.com/news/news051401.html http://botspot.

Proffitt, B. 2001. Defining and creating bots. [Online]. Available WWW: http://www.botspot.internet.com/news.news02060.html.

Stenman, D. 1998. Information agents for the Web. [Online]. Available WWW: http://w3.informatik.gu.se/~dixi/agent/class.htm.

The list of all bots. 2001. [Online]. Available WWW: http://www.botspot.com/search/.

What's a bot? 2000. [Online].Available WWW: http://www.botspot.com/bot/what_is_a_bot.html.

Disclaimer