Competive Intelligence, Software Robots and the Internet: The NewsAlert Prototype

David King
Comshare, Inc.
Austin TX
dave@comshare.com

Kirk Jones
Sterling Information Group
Austin TX
kirk@zilker.net

Abstract

The Internet offers managers and executives a vast array of untapped data that can be applied to issues of competitive intelligence. The exponential growth in the Internet has made it virtually impossible for managers and executives to track these data with even the best Windows-based Internet tools. What is required is a proactive system that automatically and routinely monitors Internet data for patterns of interest and only notifies the end user when these patterns appear. At the core of these systems are software robots. In this paper we explore a robot-based “detect and alert” system called NewsAlert that delivers Internet alerts to the desktops of managers and executives in the form of a personalized electronic newspaper.

1: Introduction

For years pundits heralded the coming of the networked society or global village. They imagined an interconnected web of networks linking virtually every computer and database on the planet, a web that sci-fi writer William Gibson dubbed the matrix. [1] Today, the web, which now goes by a number of names including the information superhighway, infobahn, and digital highway to name a few, is rapidly becoming a reality. While construction of the National Information Infrastructure (NII) is waiting in the wings, growth of existing webs is booming. It has been estimated, for example, that the Internet is growing at least 10% per month, regardless of which metric is used (i.e. growth in number of networks, hosts, users or network traffic).
While much of the growing business interest in the Internet revolves around its use for enhanced customer service and low cost marketing and advertising [5], the focus in this paper is on its potential use by business executives, managers and analysts as a source of competitive intelligence.

From the standpoint of competitive intelligence, the Internet offers a vast array of data. Unfortunately, from a managerial and executive viewpoint, these data are virtually impossible to tap even with the present generation of GUI-based tools like Mosaic. The problem is that no direct manipulation interface can sift through the glut of data in a timely fashion. Instead, a more proactive, detect and alert system is required. At the heart of these systems are software agents or robots. In this paper we describe a robot-based product prototype for delivering competitive intelligence from the Internet to the desktops of executives and managers in the form of a personalized electronic newspaper. This system is called NewsAlert.

2: Competitive Intelligence

As Benjamin Gilad reminds us, "competitive intelligence is not competitive intelligence." Instead,

Competitive intelligence is based on learning, which is based on the ability to listen to customers, to consumers, to partners, to competitors, to industry experts, and most importantly, to one's own enterprise. The competitive environment sends messages all the time: signals about change, trends, prospects, threats, and weaknesses. Early on, these signals are weak, ambiguous and hidden. Tapping them and then learning from them is an art that requires open eyes, ears and minds. [6]

Until recently, most managers and executives have relied on "hard data" for their competitive intelligence, using various management reports to compare bottom line performance against various competitive benchmarks. But, most competitive messages and signals come in the form of soft data from both inside and outside the enterprise -- rumors, speculation, field intelligence, press releases, market research, industry publications and government are some examples.

Many enterprises have simply chosen to ignore these soft data, rationalizing that the data are "too soft," or "too expensive," or "too public to be of any competitive advantage." Besides, "we barely have enough time or staff to deal with the data we do collect." Ignoring data is not the solution. This can result in loss of market share, missed opportunities, customer dissatisfaction and technological obsolescence.

Other enterprises, whose numbers are growing, have come to recognize the importance of soft data and have developed systematic methods for its collection, dissemination and, most importantly, assessment.[7] After all, it's the assessment that leads to competitive advantage, especially when the messages and signals are open for all to see. In many instances, the tasks of collecting and disseminating soft data within an enterprise are relegated to a librarian whose job it is to query public sources (e.g. Dialog, Nexis, Dow Jones, etc.) in response to requests from managers, executives and analysts who have neither the access, skill, nor time to deal directly with the various data sources. Once delivered, it's up to the recipient to interpret or assess the data.

The Internet is a vast, untapped reservoir of competitive intelligence data that has substantial benefits. It offers:

1. Direct access to many of the commercial online data providers. For example, the Internet has long provided direct (for fee) access to Dow Jones. In recent months this service has not only been enhanced but the fee has been reduced substantially.

2. Cheaper access to many of the same sources provided by the commercial online data providers. For instance, Clarinet -- a data provider whose offerings are only available on the Internet -- delivers Reuters, AP and business news at lower rates than other commercial services.[8] In the same vein, the SEC EDGAR database of quarterly and annual financial filings from all public companies is "free" on the Internet, while the same data cost $11 per report on CompuServe. [9]
3. **Timely access to many governmental data sources.** The daily procurement announcements from Commerce Business Daily, the Internet Patent News Service weekly listings of patents issued, and the nightly updates from the U.S. Supreme Court (while in session) are all cases in point. [10]

4. **Broad (and often unique) access to discussions and assessments of products, competitors, customers and market, social and technical trends.** Outside of e-mail, the most popular venue of discussion and assessment is the Usenet News groups. While News groups mirror other bulletin boards or forums, the range of topics is far greater (i.e., it has been estimated that there are 10,000 or more News groups). Unlike other services, the Internet also offers 5000 or more mailing lists or list servers that deliver focused commentary and discussion via e-mail. In some instances, these lists provide the same functionality as a News group. In other instances, a list serves the same role as a newsletter.

But, it also has substantial drawbacks, especially for managerial and executive users: [11]

1. **It's hard to use.** Most users don't have PPP or SLIP access. This means that they're faced with a myriad of tools (telnet, ftp, Archie, Gopher, Web, etc.), most of which have a basic "Unix" interface. This is a barrier for most managers and executives. It's the same barrier that corporations faced when they tried to deliver "legacy" data to executive desktops before the advent of executive information systems. In part this has been remedied by the rising number of PPP and SLIP accounts and by the emergence of new Windows-based tools (especially gopher and web "browsers") which can provide "one stop shopping" for multiple Internet services.

2. **There's too much data and it's hard to find things.** With 20,000 networks and 2.5 million hosts, the amount of data on the Internet is beyond comprehension. While "yellow pages" do exist, they only cover a small segment of the data and given the rapid growth of the net are quickly out of date. Although there are search tools like Archie, Veronica and WAIS, most of these tools are beyond the ken of most executives and managers. These means that a human intermediary (like the librarian) is required to locate and retrieve data of interest.

3. **The tools are too interactive.** While the new breed of Internet tools (e.g., Mosaic) are much simpler to use than previous generations, most employ a "point and click" or "fill-in-the-blank" interface. As Alan Kay points out:

   ... the icon-and-mouse based interfaces are good for a few hundred objects. Now imagine a trillion objects. At some point the whole metaphor breaks down, just as it does when you go into the Library of Congress and the card catalog is larger than your whole hometown library. Most of us believe that when you're dealing with that many resources, any hand-tool metaphor [like the mouse] is not going to work... So you're going to require something that's like the people in the Library of Congress who try to find out what your [research goals are, and then can work on your behalf even when you're working on other things.[12]

These electronic functionaries go by a number of names. Knowbots, daemons, smarifacts, dynadots, Olivers, and spiders are some examples. Although the most common term is software agent, we have chosen software robot instead.

### 3: Detect and Alert with Software Robots

Detect and Alert applications aim to routinely and automatically filter data before it reaches the desktop, according to rules established by the user, and to alert the user only when patterns of interest arise. Detect and Alert applications transform business intelligence applications from passive reporting applications to proactive, personalized alerting applications.

Detect and alert is a response to information overload coupled with a reduction in the time and staff to address this overload. At the core of Detect and Alert are software robots. A software
robot is a (background) program or application that:

- utilizes detection rules (i.e., formal statements describing patterns or trends of interest)
- to routinely monitor or search a large-scale or frequently changing data source for those patterns or trends
- and to alert interested persons whenever the patterns appear.

**Figure 1. Software Robot**

We have specifically chosen the term robot to distinguish it from its more "intelligent" counterpart, the software agent. Many agents have learning capabilities.[13, 14, 16] Our robots are basically watcher agents in the Verity [17] or Hoover [18] sense and lack learning capabilities. For instance, suppose you ask a human assistant to look on a daily basis for stories about competitors in the Wall Street Journal and to send you a copy any time a story appears. If you substitute the term software robot for human assistant, then you've got the basic idea.

There is more to a software robot than simple alerting. When a software robot delivers an alert, it is not enough to notify the interested party of impending disaster (or opportunity). For instance, telling someone their behind plan (no matter how elegantly displayed) doesn't help understand the problem or the actions that must be taken to correct it. At a minimum the alert should be packaged with enough intelligence to place users in a context to better understand the problems or opportunities and to communicate their findings to others.

**Figure 4. Alert Object**

4: NewsAlert: General Vision

NewsAlert is a prototype detect and alert application that ensures that the data that reaches the manager's or executive's desktop is of paramount importance. It does this by:

- putting software robots to work to routinely monitor data according to individual, personalized rules.
- automatically, immediately delivering alerts produced by these robots to the user's desktop. Users see their alerts in their personalized newspaper, which they can easily navigate, understand and customize.
- organizing alerts by user specified subject areas
- providing a set of smart tools to allow users to investigate the context of an alert and to communicate their findings to others.

While there are several components to the NewsAlert prototype, the key components include: software robots, alert objects and a newspaper client.

4.1: Robot for the Internet

Virtually any process or application can serve as software robot for NewsAlert, as long as (1) it is capable of producing alerts in the form of a metafile that adheres to our "alert object" standard and (2) (at a minimum) provides end users with the ability to "subscribe/unsubscribe" to alerts of interest from within the NewsAlert (client) newspaper. We have developed robots for Lotus Notes, for a multi-dimensional data...
mining tool (called Commander Exception Monitor), for a Dow Jones news feed and for a Reuters news feed. We have also prototyped a robot for Internet News groups and e-mail. It’s this latter robot that is of interest here. These simple sources -- New groups and e-mail -- were chosen for two reasons: (1) these are the most widely used tools on the net; and (2) the data or content from virtually any other tool (ftp, Archie, Gopher, WAIS, etc.) can be delivered via e-mail (see 18).

The Robot for Internet takes its lead from earlier work done on e-mail filtering agents (e.g. the Information Lens [19]). While these agents are rather rudimentary in nature, it is important to note that the participants at a recent workshop on Interface Agents at MIT [20] all concluded that “(these) and database daemons were the only successful agents in widespread use.” Even the more intelligent agents with learning capabilities watched over the shoulders of e-mail users.

In an e-mail filtering agent, users specify keywords or phrases for the various fields in the e-mail message (e.g. the Subject field). Users can also specify cascading keywords (ala backward chaining inference) that subsume lower/higher level concepts or group membership (e.g. if From is VicePres then ... where VicePres includes a specified set of names). The agent then uses these detection rules to send, forward or file the mail according to the users specifications.

Our robot for Internet News groups and e-mail works a little differently. First, the end user specifies the data sources to which the detection rule is to be applied (Figure 3). Here, the user can choose e-mail, All News groups (on the server) or some subset of News groups. The user then selects Add to create a new detection rule. At this point the user is provided with a simple form for entering keywords or phrases. Unlike other e-mail agents, no distinctions are made among the fields. So, if the user specifies “Comshare” as a keyword, then the robot for the Internet will search for this keyword anywhere in a News group item or an e-mail message. Users can also select an advanced mode for entering keywords and phrases (see Figure 4). This form is similar in structure to the form used in free-text searching in Lotus Notes [21] and is a simplified way of entering Boolean queries.

At the present time no attempt has been made to utilize the “fielded” information in the messages. This was done so that the user interface for specifying search criteria could be uniform across all text sources including Notes, Dow Jones, Reuters, and the Internet. It was also done in order to simplify the input for less experienced users.

While the fielded information is useful (and will appear in subsequent releases for all our text sources), the approach we’re using is the same strategy used by the Stanford NetNews Filtering server [22] which employs a WAIS search engine [23] that runs against an index of the entire message, not the specific fields in the message.

Once a user has specified a detection rule, the rule is shipped to the (Unix) server where the Robot for Internet is located. The Robot for Internet is written in Perl and supports searches on “regular” expressions, plus the Boolean operators “and,” “or,” and “not.” There is a robot for each user. The robot runs (automatically) on a scheduled basis - typically
twice a day for a morning and afternoon edition of the paper.

On a given run, the robot loops through the rules applying the search criteria against the messages in the specified News groups and against e-mail messages. When a match occurs, the message is extracted and converted into a news item. News items have a (subject) title, a date, an author, a body and a type (text or image). News items also have associated exploratory tools and files. These tools and files are used to help the reader better understand the context of the story. Presently, all Internet news items have two exploratory tools. The first tool points to either a PC News reader or Mail reader depending on the source of the item. The second points to an archive of stored news items (in our case the stories are archived into a Lotus Notes database). The archive allows the end user to bring up stories on related topics.

News items produced by the same rule are grouped into a single alert (metafile) which itself has a title and a priority (set when the user defined the rule). Once a run has been made, all the alerts files generated during that run are shipped to the PC where they alerts and their accompanying news items are extracted into the NewsAlert database. Any item that is tagged as an image rather than a text story is uuencoded before it is stored. An edition of the NewsAlert newspaper is then generated.

4.2: NewsAlert Newspaper

Alerts are delivered to the executive or managerial desktop in the form of a personalized electronic newspaper. The electronic newspaper combines the features of a paper newspaper -- stories come from multiple sources; stories are prioritized (e.g. front page stories more important than back page stories); and the format is familiar to the intended audience -- with those of an electronic paper -- its easily navigated; its contents can be personalized; stories can be augmented with other tools for understanding and communication.

A NewsAlert paper has a front page (Figure 5) followed by a series of user defined sections (Figure 6). The front page (shown below) is reserved for those alerts and news items that are of highest priority. The priority of an alert is established when the rule producing the alert was defined. In this case the priorities are similar to e-mail priorities (top, high, medium, and low). On the front page the alerts headlines and associated story headlines are displayed in the “Top Stories” area. This serves as a table of contents that can be used to bring the story into view on the front page. The idea behind the front page is to give the reader a taste of the most important items. While it is possible to zoom a front page news item to a fuller view, most of the reading takes place in the user defined sections.

**Figure 5. NewsAlert Frontpage**

They are called “user defined” sections because it’s up to the user to create the sections in his or her paper, to determine which robots will report alerts to those sections, and to define the rules that the robots will use in generating stories. [Note: In NewsAlert it is possible for an administrator to (remotely) do this work for an end user].

**Figure 6. User Defined Section**
The layout of each section is the same. On the left is a table of contents listing the alert headlines for that section along with the accompany news item headlines. When a particular item is selected, the item comes into view on the right hand side. The large arrow at the top left of the news item provides a way for the end user to zoom the news item to a full display.

Above the news item are two icons representing the tools that the end user can employ to: (1) clip and save selections of news items that they read; and (2) access the exploratory tools associated with the particular news item being read. The ClipPad is a standalone application that can be accessed without actually activating the newspaper. The exploratory tools are also separate applications that are invoked to help the user better understand or communicate about the story with others.

5: The Future

The primary purpose in developing the Robot for Internet News groups and e-mail was to determine whether the robot could proactively deliver items of "competitive intelligence" interest to managers and executives. Given the relatively low costs of obtaining data that many enterprises already receive from other sources at much higher costs (e.g., Reuters, EDGAR, stock quotes, etc.), there is no doubt that a robot for Internet can serve important competitive intelligence functions. Even though this is the case, we're not really interested in enhancing this particular robot.

When we designed NewsAlert we made it easy enough for virtually any process or application to act as a robot reporter to the paper. Currently, there are several research and commercial projects well underway whose goals are to develop software robots and agents for the Internet. The Stanford NetNews Filtering project (mentioned earlier) is a case in point. InterAp by California Software [24] is another. This commercial product, which is due at the beginning of 1995, provides a Net Scripting language which is compatible with Visual Basic, can be used to define software agents for various Internet processes, and can schedule the running of these agents on a periodic basis. Our aim is to simply modify these processes so that their output is consistent with our alert object (i.e. news item) standard and to provide managerial and executive end users with a uniform rule building interface so that they can define detection rules for these robots.

There are other initiatives aimed at defining and developing distributed discovery and retrieval tools for the Internet. The Clearinghouse for Networked Information Discovery and Retrieval (CNIDR) is one of these. [18] The Softbot Interface to the Internet services developed by Etzioni and Weld [25] is another. Similarly, the work done on various spiders and web walkers [26, 27] for the World Wide Web constitutes another set of initiatives. While it is clear that the Web represents the future of the Internet, the import of these Web robots for managerial and executive support is less obvious. Most of these initiatives are aimed at indexing and retrieving information on the Web in particular and the Internet in general. Just like WAIS, Archie and Veronica, these are basically reactive or user initiated processes. While these robots can certainly assist with finding sources of interesting competitive information, they are not really aimed at delivering up-to-date competitive intelligence on a routine basis. A better metaphor for this type of information is a robot focused on mailing lists or listservs where the news is delivered on a proactive basis.

References


