Staying abreast with information published in digital sources

P.A. van Brakel
Department of Information Studies, Rand Afrikaans University
pavb@lw.rau.ac.za

N.C. Mafa
Kendal Power Station (Eskom)
carol.mafa@eskom.co.za

Contents

1. Introduction

In recent years there has been an exponential growth in the quantity of information and the number of information sources published digitally via the Web. These sources are typically published in various formats such as news wires, journal articles, conference proceedings and technical reports, reference works and even digital books. *Information overload* resulted because multiple digital sources were published, but without

- sophisticated access methods to those sources;
- retrieval mechanisms to guarantee a high factor of relevancy and precision of hits; and
- the end-user's knowledge of solutions available to assist him/her in effectively staying abreast of only the best digital Web sites and information sources published on a regular basis.

The basic issue addressed in this article is the categories and nature of the Web-based methods that can assist the end-user to address the information management problems resulting from information overload. These methods are described in more detail, supplemented with relevant links to follow and some of the multitude of approaches to eliminate the impact of information overload are explored.

An in-depth survey of the literature on this topic exposed various aspects thereof and especially that the potential use of the identified methods varies according to the nature of the information need each method can fulfil. For example, an economist who wishes to be informed of economic indicators throughout the day could benefit from using filtered news services because they deliver up-to-the-minute news in specific areas. The discussion that follows focuses on the nature of the methods and their application in the dissemination of current content published on the Web. Under each method, examples of systems are included to illustrate the procedure each method employs. The following approaches are discussed: Webcasting, also known as offline browsing, collaborative filtering, Web personalization, filtered news services and Web tracking services. A discussion of each method concludes with an evaluation of its potential role in facilitating access to current Web content of a scholarly nature.

2. Webcasting

Webcasting is an approach that, once activated, automatically downloads information to the end-user's workstation upon request and at specific times. Various authors refer to this method of retrieving information as Webcasting (e.g. Gustitus 1998:21). The idea behind Webcasting is that the user selects sites of interest and according to a specific information need. Once the specific site is updated, relevant information is downloaded automatically onto the computer at specific intervals: daily, weekly or monthly. A user can then launch his/her browser to view the downloaded Web content (Cohen 1997:14). Two systems, namely WebWhacker and WebSleuth are discussed below to illustrate the application of this offline method of accessing digital content.

WebWhacker

WebWhacker (Capture the wide world… 1999) is an add-on system developed by the Blue Squirrel Company to add Webcasting capabilities to Web browsers. Once installed on a Web browser, it creates toolbars and utilities for finding and downloading pages. Using these enhancements, end-users can specify sites they would like WebWhacker to monitor and then receive notifications when changes are made to those sites. It also provides a directory so that end-users can organize saved sites, choose how much information to retrieve from each site (specific pages or an entire site), and if the system should regularly check for updates (Cohen 1997:14). The Automatic Scheduler feature is used to control the way that WebWhacker performs a task. For example, it could monitor changes in content and also identify all the broken links in a particular Web site. Sites saved as Bookmarks or Favourites can be imported into WebWhacker so that changes to these sites can also be monitored.

WebSleuth98

WebSleuth98 uses a unique approach to Webcasting (Websleuth product description, n.d.). Instead of presenting notifications of changes made to sites, WebSleuth automatically analyses information pages found in the Web and then generates a complete navigable index of all the important words and phrases (each cross-referenced to the pages on which they appear). Each cross-reference includes an abstract of the referenced source. The user is consequently provided with information in a tabular, navigable, indexed format. This approach saves the time it would otherwise take to view page after page to establish precision.

WebSleuth98 supports most Web browsers. **Figure 1** shows the available options for activating a Webcasting method of retrieving content using Internet Explorer. Using this

subscription screen, a user can specify when pages should be downloaded (weekly, daily, monthly), how many levels into the site the system should check for new content, as well as the method of notification (e-mail or highlighting the updated site in the user's Favourites or Bookmarks).

As such, Webcasting software solutions should be ideal for scholars to monitor the content published on sites they consider authoritative in their respective research fields.

Figure 1 Webcasting set-up screen in Internet Explorer

3. Tracking services

Tracking services represent a recently developed method to track down and monitor changes in content *inside* specific sites. Relevant content could be in Web pages, news sites, Usenet news or indexes built by search engines (Notess 1999:75). Subscribers receive e-mail notifications when changes have occurred. The providers of these services use software tools that allow end-users to specify which sites they wish to monitor, what criteria of changes to be monitored, as well as the frequency of notifications of content changes.

So far there are only a few tracking services; those that have appeared recently include Mind-It, javElink, The Informant, TracerLock, Reference.com, and NewsIndex (Notess 1999:75-78). Mind-it and javElink are discussed below to illustrate the concept.

Mind-It

The Mind-It service is offered free of charge by NetMind Company. Besides monitoring sites, Mind-It posesses a subsystem that could be used to monitor content from diverse electronic sources (this approach is outlined later in this section). To monitor specific sites, end-users first register with Mind-It, input the URLs of pages that the system should monitor

and then choose a notification delivery mechanism. The updated page or pages could be sent as attachments to an e-mail message if the user prefers this option (Notess 1999:75). **Figure 2** depicts the screen used to register for Mind-It. Users can specify the frequency of notifications, section of the page to be evaluated, kind of information to be delivered and whether to receive notifications when Web pages are moved or deleted. An important breakthrough is that Mind-It can also track content using keywords, text, images, links, bounded text, surrounded text and form results.

Figure 2 Mind-It tracking system
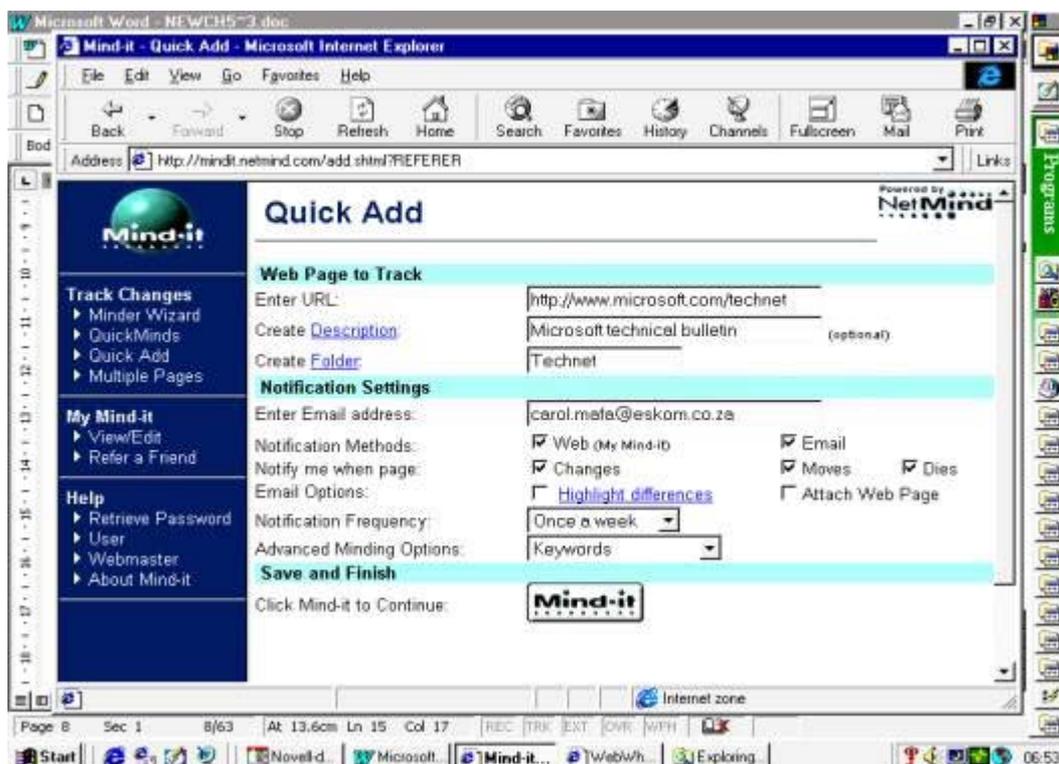
After submitting the sign-on page, a new subscriber receives an e-mail message that confirms registration with the Mind-It service.

Using the *Import Bookmarks* feature, users can import and view their browser's Bookmarks or Favourites, and can observe how they change over time. As more sites are monitored, a user can organize the sites being monitored by describing and categorizing the pages into a Mind-It folder. The ability to group together pages with similar subject areas is what makes Mind-It an important tool for managing multiple current awareness profiles (Notess 1999:75).

As mentioned in the preceding paragraph, Mind-It can also be used to monitor the content indexed by search engines. The premise is that a search conducted today might not provide the same output when conducted tomorrow because new information is being generated and added daily onto Web sites. Mind-IT stores the search query together with its criteria, queries the same search engine at a specified time and then notifies the end-user of changes it finds in the search results. In this way, Mind-IT can be used to monitor how the search output of a particular search engine changes over time.

javElink

javElink (eoMonitor 2000) is provided by InGenius Technologies as a free service for

monitoring changes in Web pages. Unlike Mind-It, this service does not send e-mail notifications about changes to users. Instead, one has to log into the javElink Web site to observe changes on the pages specified (Notess 1999:76). However, javElink enhances the monitoring of Web pages by also indicating which aspects of the page or pages have changed. This includes details of what was deleted or added. The discussion so far has highlighted advanced options that end-users could use to monitor Web sites. Though tracking services are still in the early stages of development, their contribution towards keeping track of information available via the Web could be invaluable for researchers who would like to follow in detail the developments of a specific topic. Tools such as javElink could be used to indicate where changes have occurred on the site, and also to provide extracts and graphic representations of changes. This represents a formidable step in making digital content easier to assimilate.

4. Web personalization

Web personalization allows users to choose what content should appear on the home page each time a particular site is accessed. Personalization is being used in the market as a strategy to attract customers, especially when promoting certain products. The fee-based information vendors too, for example, Dow Jones Interactive, have incorporated personalization as a strategy to provide end-users with information of specific interest. Locke (1997:369) quotes Yelvington who emphasises '…the market demands personalization. Technology makes personalisation possible. Personalisation is, therefore, not an option. It's an imperative.'

O'Leary (1999:80) states that end-users can use the following three models to personalize their access to Web sites:

- Customization model: Users can convert their individual searches in certain databases into interest profiles (or search strategy).
- Registration model: When accessing a site for the first time, the visitor is asked to complete a questionnaire about his/her personal, financial and consumption preferences. This information is used to identify the user and 'design' a Web site that will reflect the preferences.
- Creeping personalization: This model uses the same technique as the registration model to personalize Web sites but, instead of requesting users to respond to a list of questions, creeping personalization uses computerised tracking of individual preferences and usage patterns. The computer system collects these preferences and usage patterns and applies them where necessary (O'Leary 1999:80).

Sites supporting Web personalization are on the increase, for example MyZDNet and Siteseer. Both demonstrate varying levels of Web personalization.

MyZDNet Alerts

MyZDNet is a solution provided by ZDNet to enable end-users to choose a range of subsections of the site that they want displayed as the opening page. A selection is made through ticking sections of interest among the many categories of information made available. These range from general news, product reviews, stock markets and technical tips to site recommendations and appear as the opening page each time the end-user accesses the ZDNet site (see **Figure 3** for an example).

ZDNet has expanded Web personalization by incorporating a service called MyZDNet

Alerts. MyZDNet Alerts is a free service that alerts the user about new content appearing on the personalised ZDNet page. It is closely integrated with the personalised ZDNet page and as such uses the same personal preferences as those defined in the personalised page. The software required to run MyZDNet Alerts can be downloaded from the ZDNet site. When installed on a workstation, MyZDNet Alert shows an icon in the Windows Task bar and it flashes when there is new content on the MyZDNet page.

Figure 3 Personalized opening page at ZDNet site

Clicking on the flashing icon will display the latest content that matches the preferences defined in the personalised page (see **Figure 4**). This approach ensures that users are regularly notified of new content on their personalised pages. One limitation is that the effectiveness thereof relies on the user regularly checking if new content has been added to the site.

Figure 4 MyZDNet Alerts displays selected stocks
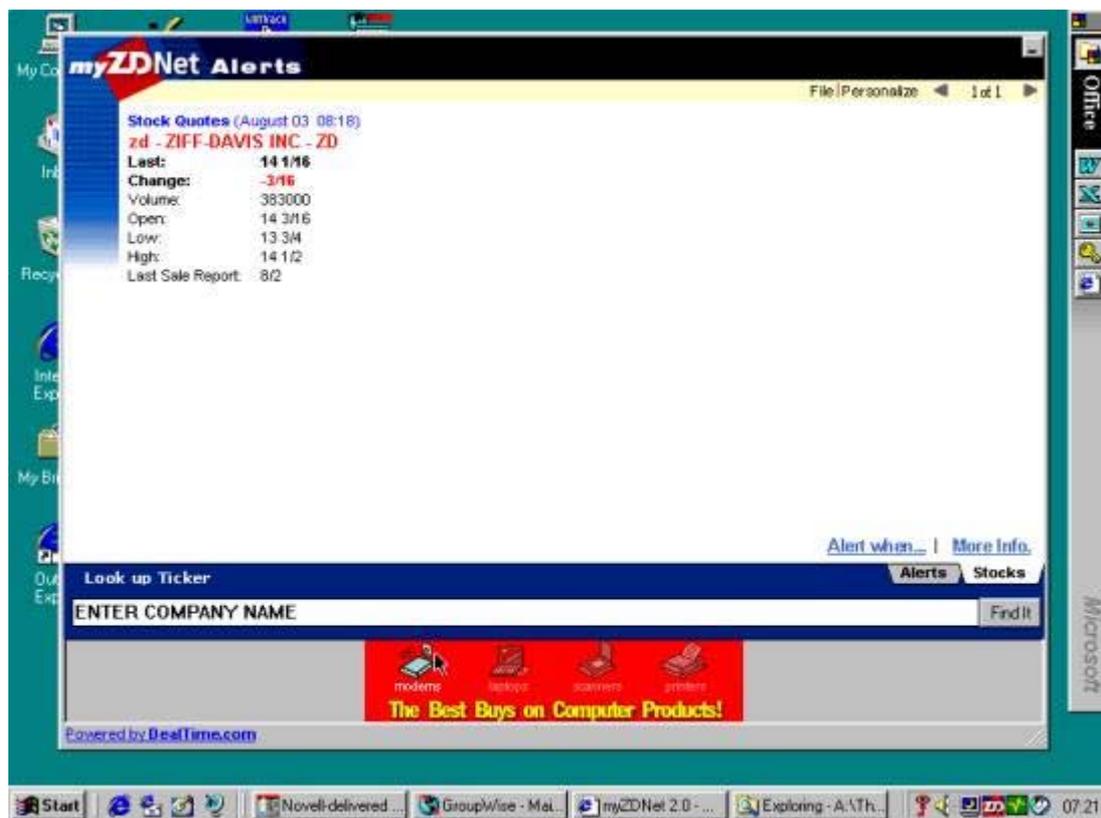
Siteseer

Early efforts to personalize the Web can be found in Siteseer, a Web recommendation system that uses bookmarks or favourites to predict and recommend relevant pages. Siteseer uses bookmarks to make an implicit declaration of interest in the underlying content. It does this by using folders in which related bookmarks have been grouped together as an indication of semantic coherences or relevant groupings between subjects. These folders are then used by the system to contextualize Web page recommendations (Rucker and Polanco 1997:73). In this way, users are provided with information that matches the content of their bookmark lists.

Indications are that more commercial information providers and products plan to incorporate personalization features. A survey conducted by Jupiter Communications reports that 40% of the top 25 Web retailers offer a personalization option and that nearly all had plans to offer it in 2000 (O'Leary 1999:80).

Web personalization uses individual preferences to present a personalised view of the content covered in the site. This method would be ideal if end-users were reluctant to spend time on navigating sites to check the latest content, but preferred to view specific content in a manner they would find comprehensible. Web personalization may therefore be viewed more as a Web content presentation facility than a method of keeping abreast of recently added Web content.

5. Collaborative filtering

Collaborative filtering is based upon the principle of sharing information between users with similar information needs. Filtering is usually done by means of a weighing scheme that assigns higher weights to items other users have selected. On receiving a number of

information items, users have provided what Balabovic and Shoham (1997:67) calls 'relevance feedback' by means of rating procedures. Other users with similar interests are then notified of items that have received high ratings from others.

According to Goldberg *et al.* (1992:61), 'collaborative filtering simply means that people collaborate to help one another perform filtering by recording their reactions to documents they read'. It creates a sense of community for like-minded users. The same function is now performed by software that manages the profiles of a specific group of end-users. The software functions as an intermediary between the database, in which the desired information is stored and the user. Interest profiles in the repository are used in the selection of information contained in databases that have information of potential interest to a group of users. In the Web environment, collaborative filtering is being applied to resources such as Web sites, Usenet news, electronic journal articles and movies. Two examples of systems that use collaborative filtering are Alexa Internet and Kenjin.

Alexa Internet

Alexa Internet (Alexa n.d.) is a collaborative filtering system that returns a single set of recommendations from ratings done by multiple users. According to Bates (1999a:88), Alexa is based on the theory that if individuals compile their collective experience and evaluation of Web sites, they could more effectively separate the useful from the useless. Alexa applies collaborative filtering in two ways:

1. Keeping track of how users move from one site to the next; and
2. Allowing them to vote on how beneficial a particular site is found.

Alexa provides summarised statistics on each site. These include the number of Alexa users who have visited the site, when and by whom the site was registered with InterNIC, how many links point to that site, the number of pages on the site and how frequently the site is updated. It also provides a 'links button' to related sites that are often visited by those who have visited the site being accessed. The input about the site is displayed on the frame at the left of the screen (see Figure 5). One can read reviews of the site from other Alexa users and also write one's own review of a particular site. This is useful as it gives an indication of the perceived value of the site (Bates 1999:88). Alexa software can be downloaded from its Web site.

Kenjin

Kenjin (Kenjin n.d.) is freely available and developed by Autonomy to facilitate the collaborative filtering of Web content. Kenjin uses pattern-matching technology that can analyse any piece of text by identifying and ranking the main ideas, and then display a list of related Web sites or news relating to the subject (Autonomy Kenjin 2000).

Figure 5 Alexa Internet as applied within IE 5.0

Kenjin (http://www.kenjin.com/) uses an auto suggest feature to provide links related to the content currently being read or typed in any Windows application, such as Word, Excel, PowerPoint, as well as browser and e-mail clients). The related links automatically appear in the window as depicted in **Figure 6**. This shows a paragraph of a Word document and based on the subject of the paragraph, Kenjin has provided some links to related Web content in the bottom screen. As alternatives to the auto suggest feature, text can be dragged from a document to the toolbar to identify links. Alternatively, a Kenjin icon can be dragged from the toolbar and dropped on text to bring up related links.

When users access the Kenjin site, their usage patterns are analysed, which are then stored as user profiles. As a service users can request a list of e-mail addresses of users with interests matching the subject of the document they are reading or typing.

Figure 6 Site recommendation feature in Kenjin
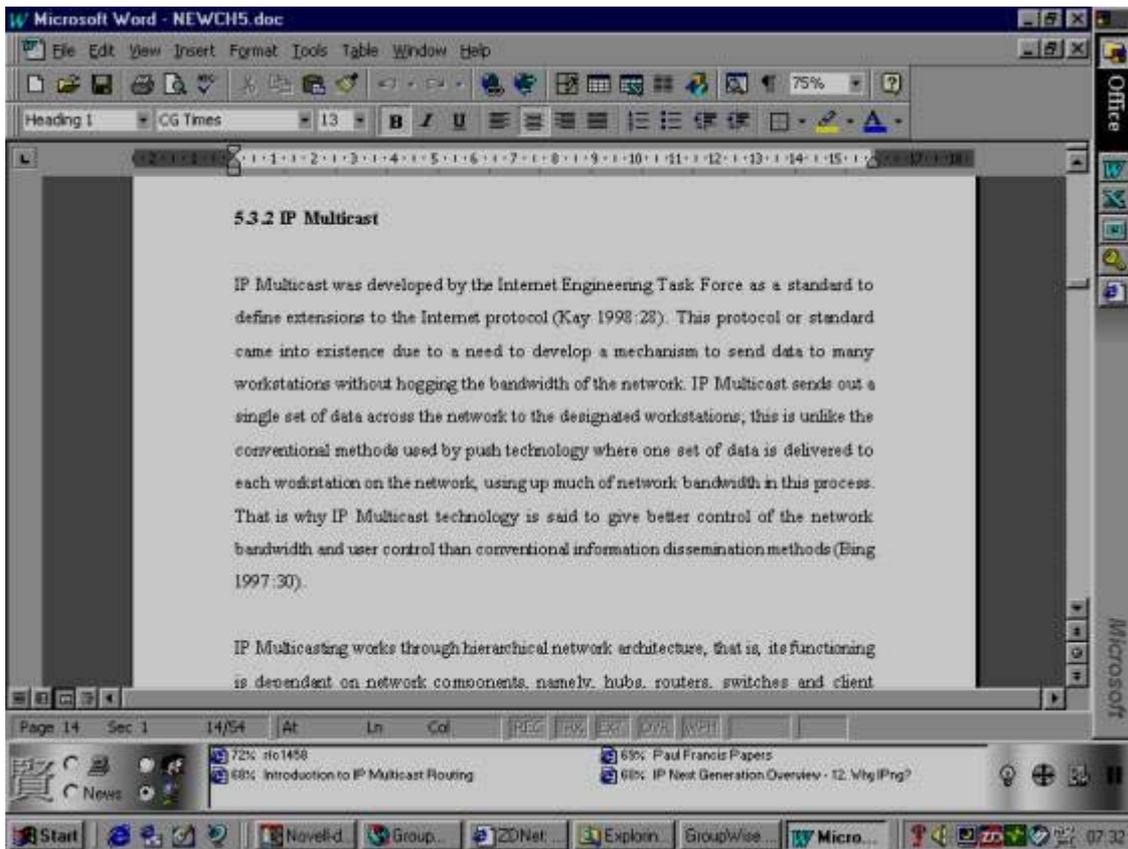
User profiles and details will only become available if users have elected to be included in the online community of Kenjin users. Kenjin can be downloaded from its Web site and can run either automatically when starting Windows, manually from the Start Menu or from the icon optionally installed on the desktop (Autonomy Kenjin 2000).

Collaborative filtering can be a very functional concept when used by a peer group representing a particular subject area. Dragan (1997) said that 'collaborative filtering takes some of the things human beings do best – thinking and making critical judgements'. The application of this method via the Web makes it possible for a vast number of scholars across the world to participate in the selection and recommendation of relevant Web content.

6. Filtered news services

Filtered news services supply the latest news releases in areas such as politics, stock markets, sports and weather information. They apply authoring tools, editing facilities and systems that generate interactive and customized news to meet the needs of specific end-users. A filtered news service is formed on a three-tier relationship among news providers, also referred to as content providers (collect news and build databases), the vendor (provides end-user with the retrieval mechanism and creates transmission channels), and the end-user or information customer (Watters *et al.* 1998:142-144).

In this interrelationship, the content provider's role is to acquire, produce and prepare, on a daily basis, current news for immediate distribution. When prepared, the news stories are published as newswires, business newspapers and magazines, trade journals and Web sites such as ZDNet (Curle 1998:16). The vendor acts as intermediary between the content provider and the end-user. The vendor owns and maintains the servers that are used to store end-user profiles and consequently filters news within the databases and disseminates only

news that matches specific end-user profiles. Vendors have an obligation to choose appropriate sets of content providers that could supply information matching the interests of the end-users. There is an enormous array of news providers from which vendors can choose from, for example, ABC News, CBS Sportline, CNN, Knight-Ridder, Money Magazine and ZDNet (Finnie 1997:28). At the end-user site, the client software is used to manipulate the news delivered.

Many filtered news services exist today, some available at a fee though others can be accessed free of charge. PointCast and NewsAlert are discussed below as examples. Both are available free of charge, while no additional software or site licensing is required to use them.

PointCast Network

PointCast was the first filtered news service to deliver current news directly to desktop. The system uses large news wires such as CNN, Wire, New York Times and ZDNet as its content providers (Gustitus 1998:23). Notifications of the latest news can be delivered onto the desktop as ticker messages, which, when opened are direct hyperlinks to multiple sites on the Web. PointCast has a daily delivery quota of approximately 450 million news items. 'News feeds', that is news from content providers, disseminate into PointCast through satellite and leased telecommunications lines (Lawton and Stevens 1998:29).

When end-users register with the service, a client application that enables the workstation to receive news from PointCast news servers is downloaded to the end-user's computer. End-users have to complete a form requiring details such as name, gender, e-mail address and occupation. Once they have been registered they can select any channels of interest, which have been organised according to broader fields such as news, sports, technology and business. Within each of the selected channels users can narrow down the type of content they would like to receive, for example, with the ZDNet channel, the selection could include news, products, Web catalogues, a software library and game spotting. Depending on the preferences of the users, they receive notifications according to the selected categories. Figure 7 shows the latest content in the software library category that has been delivered to desktop through the ZDNet channel.

Figure 7 Content delivered though ZDNet Channel in PointCast
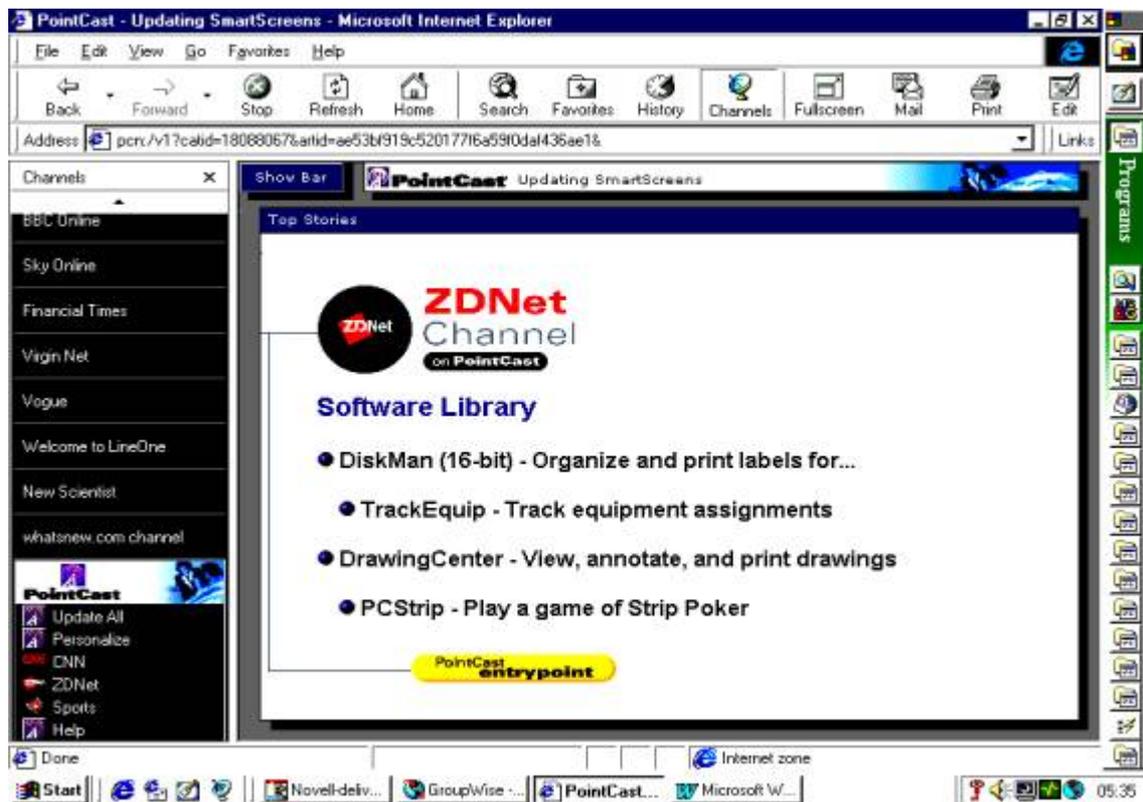
News Alert Service

News Alert (NewsAlert 1999) provides online access to news and market data via the Web. Users subscribe to the service by entering their own unique search profiles which News Alert Service saves in the subscriber's account. Subscribers are then presented with customised updates as soon as they log onto the News Alert site. A subscribers can also create a personal 'watch list,' that is, real-time monitoring of changes in, for example, specific stock markets and foreign exchanges statistics. Subscribers have an option to create their own personalised home pages that will launch every time they access the News Alert Service.

Microsoft and Netscape have seized the opportunity by being among the first to incorporate this facility into their browsers. IE developed the Active Desktop push facility and Communicator used a push package called Castanet to develop Netcaster. These two serve as clients to the existing content providers such as ABC News, CBS Sportsline, CNN, Knight-Ridder and Money Magazine, Yahoo! and ZDNet (Herther 1998:113). End-users have a choice between using the pre-set channels that come with these browsers or subscribing to channels of their own choice. IE and Navigator have popularised the use of filtered news services to keep users informed of news disseminated via the Web, thus making push technology widely available to a far larger audience than before. Until recently most providers of filtered news services have mainly covered popular news, for example CNN.com has covered the news also covered on its television network. However, multipe services have been developed to fill the various gaps in the type of news, mainly business and financial, that could be filtered to various categories of users, such as those in politics, economics and technology.

Source-based alerts

Some electronic journals and magazines, or many important Web sites for that matter, have also embarked on a type of alerting feature. On the Web site hosting the source, users can register with an alerting service provided by the site. Once registered, they receive, by e-mail, regular notifications of the contents of the subsequent issues of the journal or

magazine. Detailed information as in Figure 8 may be required though some publications only require an e-mail address to register a user.

Figure 8 Sign-on screen for source-based alerts

The publishers of technical magazines and electronic journals can also utilize another form of source-based alerts. These are integrated with sections reporting on the reviews of the latest products for Web applications, PC software tools and hardware. For example, *PC Magazine* has developed an alerting service called 'Product Alert' for its product reviews section. Users can sign up to receive reviews from its product reviews section three times a week (Subscribe to Machrone's… 2000).

Some publishers have even added a citation facility onto source-based alerts so that users can subscribe to receive notifications whenever a new article published in a specific journal cites a chosen article. For example, at the end of all articles published in the *British Medical Journal (BMJ)* there is an option for users who need to receive notifications when new articles cite the article they have just read (BMJ.com 2000). In generally, source-based alerts are on the increase and more scholarly digital publications are incorporating this feature.

This outline of Webcasting or push technologies indicates that especially commercial information providers are very active in facilitating methods to address typical information overload problems.

7. **Conclusion**

The purpose of this research was to establish recent trends in the methods currently available to keep abreast of information published via the Web. A framework for these methods

includes the nature of the information need, the procedure for fulfilling the need, the electronic system and consequently the type of Web content.

The trends indicate that these methods perform two basic functions:

• Monitoring new content published at selected sites. The methods facilitating this function include Webcasting, Web personalization and tracking services. Usually, these methods do not build a profile of the content (represented by keywords) that a particular user needs to monitor, but merely automate the monitoring of new content at given sites.

• Collaborative filtering, source-based alerts and filtered news services that first search for sites, select content of potential interest and then send subscribers notifications of content found. A unique feature is that end-users can specify their information needs in the form of interest profiles. Some of these methods closely resemble a typical SDI service, in which case interest profiles are structured by means of Boolean and proximity operators, truncation and field limitations.

8. References

1. Alexa. n.d. [Online]. Available WWW: http://www.alexa.com.
2. Autonomy Kenjin. 2000. [Online]. Available WWW: http://www.zdnet.com/products/stories/pipreviews/0,8827,195262.htm.
3. Balabovic, M. and Shoham, Y. 1997. Fab: content-based, collaborative recommendation. *Communications of the ACM*, 40(3):66-72.
4. Bates, M.E. 1999. Alexa Internet. *Database*, 22(2):88.
5. BMJ.com. 2000. [Online]. Available WWW: http://www.bmj.com.
6. Capture the wide world of the Web with a touch of a finger! 1999. [Online]. Available WWW: http://www.bluesquirrel.com/products/whacker/whacker.html.
7. Cohen, S. 1997. Surf n' go. *Training & development*, 51(3):14-15.
8. Curle, D. 1998. Filtered news services: solutions in search of your problem? Online, 22(2):15-24.
9. Dragan, R.V. 1997. Advice from the Web. [Online]. Available WWW: http://www5.zdnet.com/pcmag/features/advice/_intro.htm.
10. EoMonitor. 2000. [Online]. Available WWW: http://www.jevelink.com/cat2main.htm.
11. Finnie, S. 1997. Not just browsing: Netscape's Communicator 4.0 brings together e-mail, groupware, and browsing. *PC magazine*, (SA) 5(8):23-28.
12. Goldberg, D. et al. 1992. Using collaborative filtering to weave information tapestry. *Communications of the ACM*, 35(12):61-70.
13. Gustitus, C. 1998. The push is on: what push technology means to the special librarian. *Information outlook*, 3(1):21-24.
14. Herther, N.C. 1998. Push and the politics of the Internet. *The Electronic library*, 16 (2):109-116.
15. Kenjin. n.d. [Online]. Available WWW: http://www.kenjin.com.
16. Lawton, S. and Stevens, L. 1998. IP multicasting – tell it on the mountain – with multicasting, a single voice can be heard by all with minimum impact on network activity. *LAN times*, 15(17):29.
17. Locke, C. 1997. Intelligent agents create dumb users. *Online & CDROM review*, 21 (6):369-375.
18. NewsAlert. 1999. [Online]. Available WWW: http://www.newsalert.com.
19. Notess, G.R. 1999. Internet current awareness. *Online*, 23(2):75-78.
20. O'Leary, M. 1999. Web personalization does it your way. *Online*, 23(2):79-80. Rucker, J. and Polanco, M.J. 1997. Siteseer: personalised navigation for the Web. *Communications of the ACM*, 40(3):73-75.
21. Subscribe to Machrone's new product alert! it's free! 2000. [Online]. Available

WWW: http://www.zdnet.com/pcmag/lists/pcmalert/subscribe.html.
22. Watters, C.R. et al. 1998. Electronic news delivery project. *Journal of the American Society for Information Science*, 49(2):134-150.
23. Websleuth product description. n.d. [Online]. Available WWW: http://www.promptsoftware.com/products/wsleuth.htm.

Disclaimer

top