

SuperNews: Multiple feeds for multiple views

by S. Elo Dean
L. Weitzman

The IBM intranet hosts a range of news “feeds,” or distribution services, from internal and external sources. To read the news of the day, users need to visit a number of Web sites, each with a different categorization, navigational structure, and layout. Transfer of technology from the MIT Media Laboratory to IBM enabled the design of SuperNews, a prototype application that aimed to improve IBM employees’ experience in reading the news. SuperNews merged heterogeneous news feeds and presented a consistent interface to users. Users could choose to read news on the Web, in e-mail, or as engaging visualizations. Using text processing technology, SuperNews discovered meta-information in articles and created new ways to browse the news collection. SuperNews also allowed users to publish their own columns, as well as annotate and recommend articles to their colleagues. Starting with a range of uncoordinated feeds, SuperNews transformed the solitary reading of news into an engaging and visually appealing community experience.

How do individuals at IBM grasp what is happening in this company of more than 300,000 employees? How does one keep track of industry news, or the way the media portrays IBM? In the first years of the IBM intranet, employees needed to visit many Web sites with different navigation styles and categorization schemes to get the “big picture.” External and internal news, official reports, and opinions on one topic were spread across the intranet and difficult to find or compare.

Until recently, one reason for such a heterogeneous environment was the lack of standards. Content providers on the IBM intranet used proprietary data for-

mats and their own tools, without observing a common news article format or a standard meta-tagging schema. Since Extensible Markup Language (XML)¹ has emerged as the industry standard for content encoding, providers can begin to agree on a common vocabulary to describe the structure and content of news articles.

The lack of collaboration among content providers has further prolonged the disparate IBM news sites. Coordination among content providers to standardize formats and tools is difficult because most operate in a production environment. Their priority is to generate accurate and timely content, not to introduce changes that might disrupt critical daily operations. Long-term improvements easily become second priority.

Each content provider on the IBM intranet has also traditionally targeted a specific audience and delivery mechanism. Until recently, news was not considered a reusable commodity that could be repurposed for many audiences, end-user needs, or devices. In this environment, IBM employees were not getting the full benefit of the available resources.

Because of the difficulty in introducing new tools and standards to each content provider, a different approach was needed. By reformatting news articles

©Copyright 2000 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to *republish* any other portion of this paper must be obtained from the Editor.

after their creation, and without involving the content providers, SuperNews, an experimental application, was able to demonstrate what would be possible if all providers standardized their content.

SuperNews ran on the IBM intranet for about a year and a half, until December 1999. During the first six months, the application provided a personalized news pilot, after which the other features described here were added. The application gathered valuable feedback on features that users would like to see deployed in the official news application now running on the IBM intranet.

Background

Under the News in the Future (NiF) research consortium, the MIT Media Laboratory has produced a range of prototypes that investigate the role of news in the digital era. The consortium aims to enhance the efficiency of production, the timeliness of delivery, the convenience of presentation, and the relevance of editorial and advertising content to the consumer. NiF projects—FishWrap, PLUM, ActiveArticle, and VIA—have influenced the design of SuperNews.

The MIT campus-wide digital newspaper *FishWrap*² provides MIT students with a personal news page that includes hometown news, special topics, and on-and-off campus events. Each student provides a personal profile, which the system uses to filter the articles presented to him or her. In addition, FishWrap opens on a dynamic front page, which consists of the latest recommendations from the readers. Instead of an editor selecting the hot stories for the front page, the collective student body decides which news is interesting and important. On the IBM intranet, SuperNews implemented a similar personalized news page where readers subscribed to news categories. It also offered the People's Choice column, much like the FishWrap front page, where readers could recommend and annotate articles.

The PLUM system³ automatically generated augmented news. As disaster news arrived in FishWrap, PLUM analyzed each article and related facts reported in it to the reader's home community. The augmented article was linked to comparisons with the reader's home and references to historically significant disasters. These explanations helped readers understand the scope of the remote disaster. SuperNews used the same augmentation technique to link persons, places, organizations, and products detected

in an article to related resources, creating a meaningful context around the article.

ActiveArticle was a Media Lab research project built as a result of participating in a one-week assignment to "change the culture of the newsroom" at the *Chicago Tribune*. The project was to develop a story using different media reporting on the neighborhood surrounding Wrigley Field. Aimed at a diverse audience, two goals shaped ActiveArticle's implementation: to make text enjoyable to read on a computer screen and to reveal the dynamic nature of the contextual relationships between the text and the images. A working prototype is available.⁴ SuperNews used this visualization and created ActiveNews articles as one of its final presentation styles.

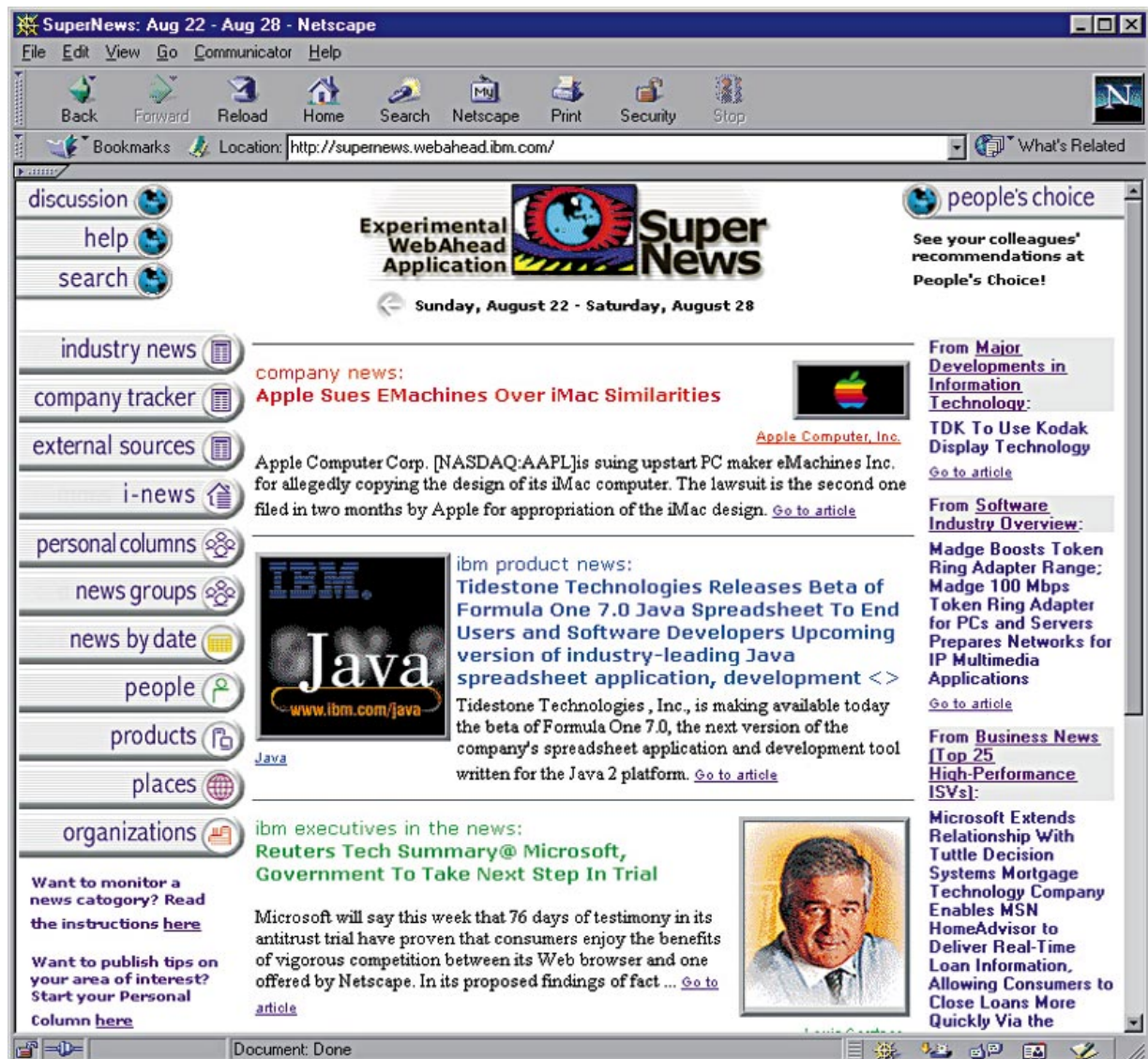
The separation of form and content has long been championed by the SGML (Standard Generalized Markup Language) community and has recently come to the Web in the form of XML and XSL (Extensible Style Language) style sheets.⁵ SuperNews also drew inspiration from the *Visual Information Architecture* (VIA) project⁶ from the Media Lab. VIA uses formal visual languages to support the design process, from interactive tools to support the creation of an article, to the automatic presentation of an article in dynamic environments. VIA personalizes the delivery of information based on user profiles, the hardware context, and the current use of the information being delivered. SuperNews built on this tradition, delivering three different presentations of the same article.

IBM SuperNews project

SuperNews unified the heterogeneous news environment on the IBM intranet. It brought together IBM internal news, news from the external press, and news published by the IBM employee community into a consistent interface. The SuperNews front page is shown in Figure 1.

SuperNews also summarized the most prominent people, places, organizations, and products in the weekly news. These new views complemented the categorization of the various content providers. For example, those familiar with the internal news taxonomy could still check their preferred categories for new stories. But for those who did not know in which category to look, articles relevant to their work could go unnoticed in the unfamiliar taxonomy. In SuperNews the articles were easier to find because

Figure 1 SuperNews front page



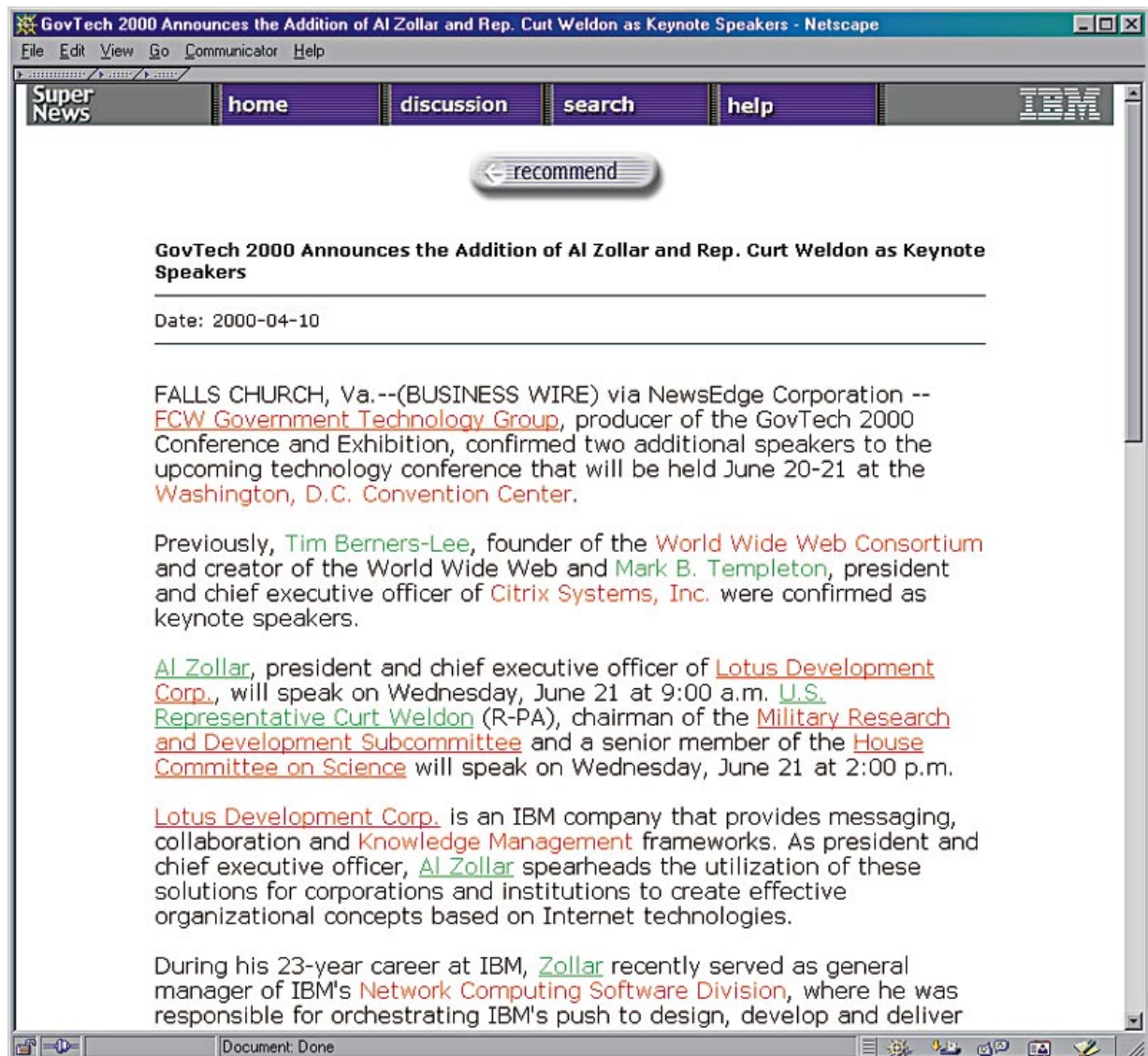
the users could browse by the name of a person, a place, an organization, or a product.

SuperNews turned the reader into an active participant. Readers were encouraged to create their own personal columns, to help others in the company find out who is an expert on what topic. Many IBM employees send e-mail to close colleagues with tips and pointers to interesting Web pages and add their own commentary or viewpoint. These employees are ex-

perts on a topic and serve as “filtering agents” for their peers. By encouraging these experts to publish a personal column, SuperNews began to build a knowledge bank of IBM internal experts. To make publishing as easy as possible, the experts could simply send an e-mail message to their personal columns.

Going through daily news in search of relevant articles is difficult. To help with this task, SuperNews encouraged employees to recommend articles. The

Figure 2 Augmented article on the Web

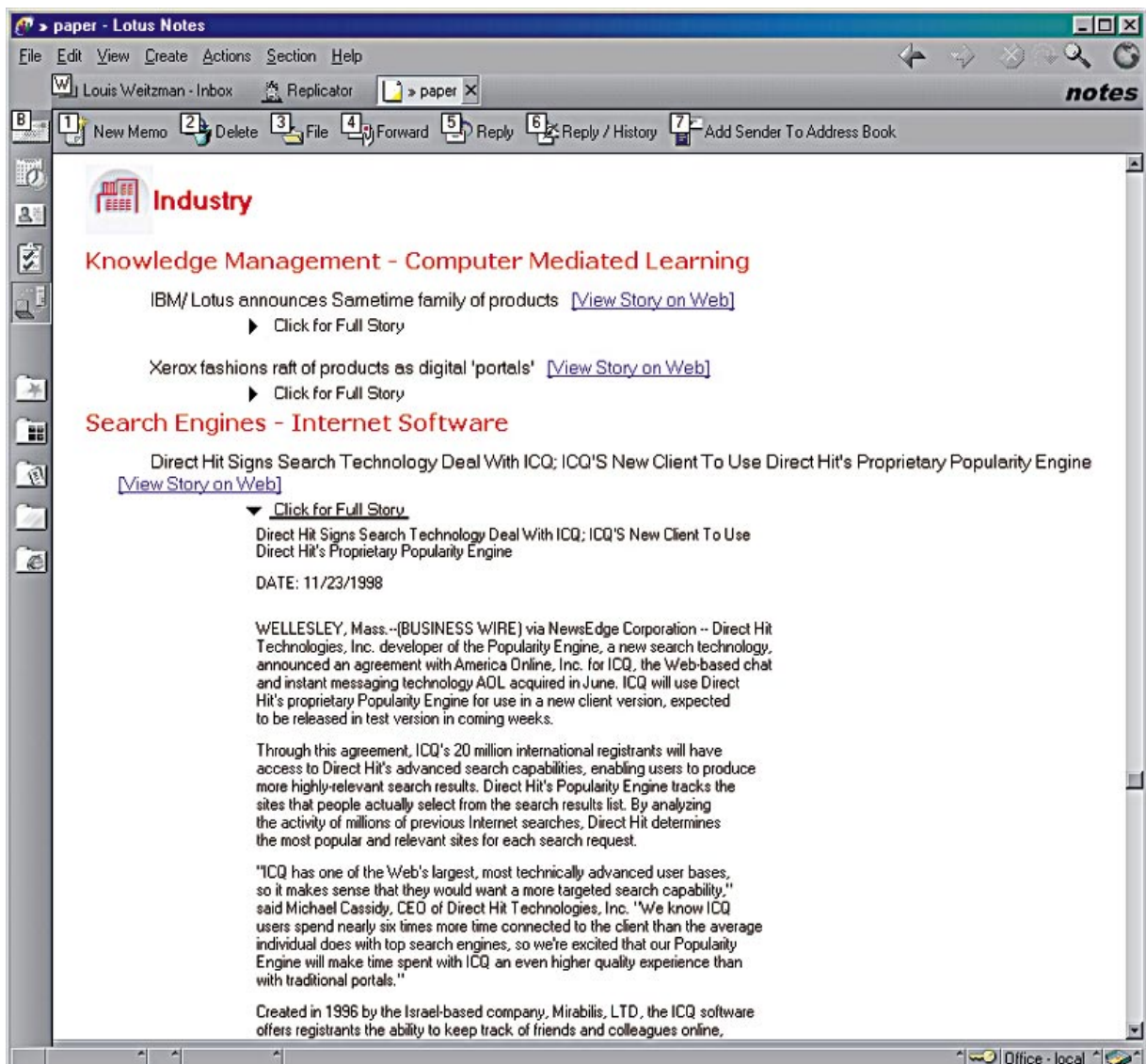


employee could target the recommendation toward one or more job types, to produce more focused recommendations. The People's Choice column then displayed the readers' recommendations, presenting yet another view of the week's news. It reflected the voice of the employees, and what they were reading and thinking.

To suit different readers' needs and preferences, SuperNews allowed them to subscribe to categories of

interest. It also allowed readers to choose among three types of visualizations of an article. SuperNews provided a Web version in HTML, an e-mail version in plain text, and an ActiveNews article. The version for the Web was an augmented article, where all proper nouns were highlighted and linked to related resources, shown in Figure 2. The e-mail version contained the full body of the article for off-line reading, as shown in Figure 3. The ActiveNews article, an interactive presentation, is described later.

Figure 3 E-mail version of an article in Lotus Notes



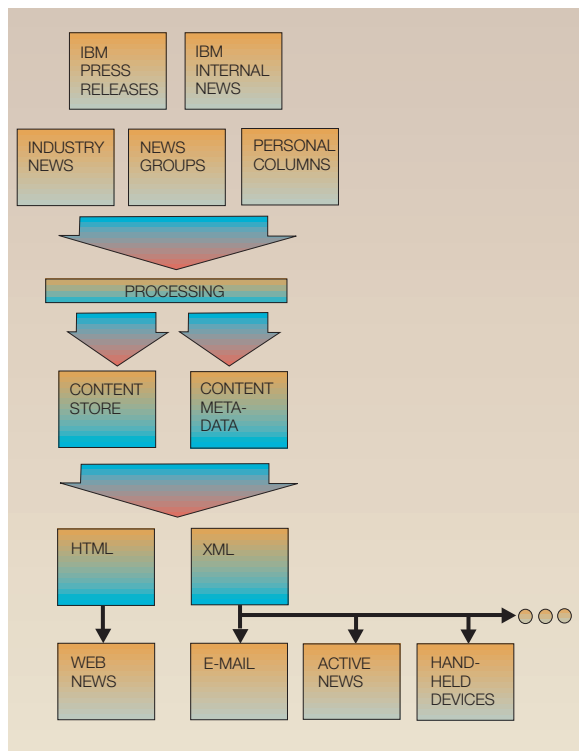
Life of an article. An article begins with an author at a news source. SuperNews retrieves articles from five sources, shown in Figure 4. Each incoming article is transformed into a common XML format. An example is shown in the appendix.

Next, Textract,⁷ designed and implemented by the Information Retrieval and Analysis Group at the IBM Thomas J. Watson Research Center, processes each article. Among Textract's natural language processing algorithms, SuperNews uses Nominator, special-

ized for detecting and distinguishing among different types of proper nouns.⁸ Nominator can detect person, place, and organization names. If a dictionary is imported, Textract detects the terms and their synonyms listed in the dictionary. To detect all product names that occur in articles, SuperNews runs Nominator with an IBM product dictionary.

Nominator uses rules that define how proper nouns appear in text, how to distinguish individuals' names from company names (e.g., James Smith vs James

Figure 4 The life of an article in SuperNews



Software Inc.) and how to recognize variants of the same name (e.g., James Smith, Jr., Mr. Smith, Smith). If the algorithm relied simply on dictionaries of names, it would fail on any text with new names. The Nominator technology runs autonomously, discovering valuable structure without requiring the help of an editor. However, since language in text is ambiguous, Nominator is not 100 percent accurate and sometimes makes even comical mistakes (e.g., it considered Ginger Chicken in IBM's internal cafeteria news to be a prominent "person" in the news.)

Next, a program augments the article by generating links from the discovered proper nouns to other articles and related Web sites, such as the IBM employee directory "BluePages," IBM's e-commerce site "ShopIBM," and search engines and taxonomies on the IBM intranet and the Internet. Figure 5 illustrates a typical augmentation on the Web. Rules that depend on the source of the article and the type of proper noun determine the appropriate augmentations to generate.

SuperNews then publishes the article as a static HTML (HyperText Markup Language) page to provide speedy retrieval for users with older browsers. The application also publishes it in the XML format for delivery to browsers that support more recent client-side technology. Microsoft's Internet Explorer V5.0, for example, can combine content in XML format with an XSL style sheet to display the article to the user. As new client applications become more prevalent, SuperNews could move completely away from storing articles in HTML and use XML instead. The XML format is also used to generate plain text for e-mail delivery and for the transformation into an ActiveNews article. This XML format would easily allow rendering for other devices, such as a hand-held computer, a phone, or a pager.

Since all proper nouns discovered in an article are stored as meta-data about the article, SuperNews can construct new index pages. The system generates index pages, based on the proper nouns, to complement the index pages based on the meta-data supplied by the content provider. For efficiency, index pages are updated after processing a new batch of articles from a source.

During the six-month personalization pilot, SuperNews also sent out a personalized e-mail to subscribers after processing the daily morning news distributions. The full text of articles that matched the selected categories in the user profile was included in the message.

ActiveNews visualization. In order to support different types of visualizations, articles in SuperNews are stored as XML documents. Although many different visualizations of a news article are possible, this paper focuses on the ActiveNews article visualization.⁹

The ActiveNews implementation of ActiveArticle can be seen in Figures 6 and 7. On the right, a few lines of the main text of the article are in focus and set in large type. The type size of the other lines of the article is inversely proportional to their distance from the text in focus, creating a fish-eye lens effect.¹⁰ To the left of the text is a scaled-down version of the complete article. This unconventional "scroll bar" tracks the reader's progress through the article by highlighting the lines of the article in focus. One way the reader can control focus is by manipulating this scroll bar. A similar use of scroll bars for fish-eye views has also been explored in the context of col-

Figure 5 Typical augmentation of an automatically identified proper noun



laborative views.¹¹ The user can also scroll the article by clicking and dragging on the main text.

Located on the left-hand side of the display are images linked to sections of the text through defined "elastic" relationships. Each section of the article is represented by a different graphical layout that specifies how images scale and position themselves. As the reader enters a new section, the ActiveArticle software triggers a smooth animated transition of the image layouts, visually representing the shift in context. The user can also select an image and bring it into focus with its caption displayed.

In SuperNews, an ActiveNews article enhances the viewing of a static textual article with a dynamic presentation enriched with images. The process starts with the XML representation of an article, as shown in the appendix. The XML contains standard news markup tags, such as title and author, and the automatically generated references to proper nouns and associated images.

In order to produce an ActiveNews article, a process first examines the XML file to determine acceptability. The criteria for acceptability of an ActiveNews article are:

Figure 6 An automatically constructed ActiveNews article in the style of ActiveArticle. The main text is presented as a fish-eye view on the right. Relevant images appear on the left. In between, a scroll bar of the complete article illustrates the portion of the article currently visible.

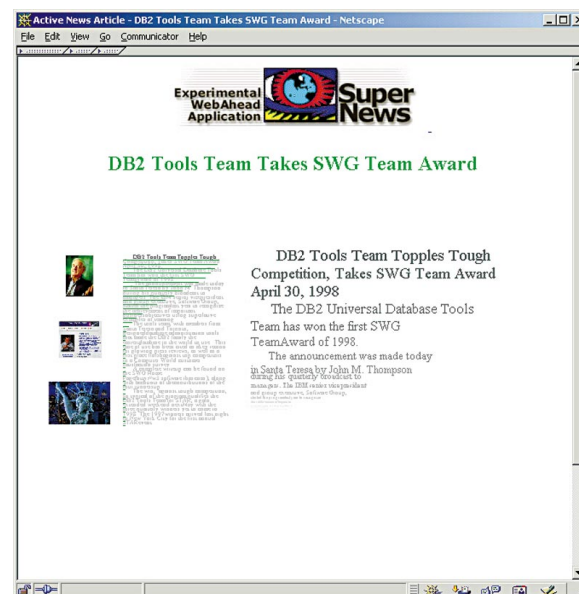
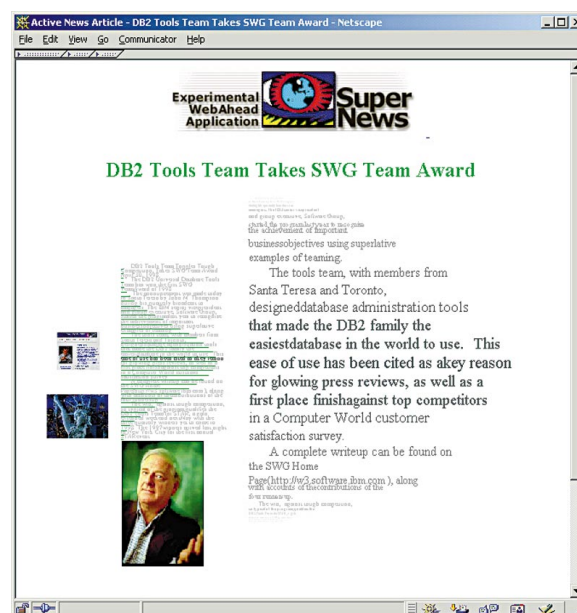


Figure 7 As the user scrolls through the article in Figure 6, images change size and position to reveal a shift in context.



1. Identification of proper nouns, or names, in the article.
2. Existence of images in the media directory corresponding to the names identified, whether the name be that of a person, product, place, or organization. A minimum of three images was required for a valid ActiveNews article.
3. A minimum number of occurrences of the identified names. Those with the highest occurrence were favored.
4. Proximity, based on paragraphs, of the identified names to each other.
5. Display of a maximum of five images. Those distributed across the paragraphs of the article were favored.

Based on these five criteria, an ActiveArticle is created for only the suitable candidate articles. ActiveNews articles are not created for articles without images, articles that refer to a name only in passing, or articles in which the identified images are not distributed across the paragraphs of the article.

To produce an ActiveNews article, an HTML file is created from the XML article and an XSL style sheet. This HTML file defines the input for a Java** applet, which parses transition markers placed within the text identifying the contextual sections to be highlighted by the different images.

Implementation. The SuperNews system consists of the following components:

1. A DB2* (DATABASE 2*) Universal Database* stores all meta-data about articles, categories and sources.
2. The DFS (Distributed File System) stores the content of the articles, the index pages, the front page, the user recommendations, and the different visualizations of the articles.
3. The “news feed” processing module consists of a series of feed-specific modules that transform an article from a content provider’s format to the generic XML format in SuperNews:
 - The News Group module, written in Java code, uses the Java API (application programming interface) to NNTP (Network News Transfer Protocol) to export updates to all monitored news groups.
 - The IBM Press Release module, written in Java

code, uses the Java API to Lotus Notes* to export updates to the Press Release database.

- The Industry News module, written in Perl, retrieves by FTP (File Transfer Protocol) the daily update from NewsEdge.¹²
 - The Personal Column module, written in Perl, processes a mail spool file to retrieve updates from employees publishing to their Personal Columns via e-mail.
 - The INews module, written in Perl, exports updates to the IBM internal news feed from a remote DB2 database and file system.
4. The postprocessing module augments an article. This module runs the Textract and Media modules to identify the proper nouns and media references, respectively. It then writes the XML and HTML versions of the article into DFS and the meta-data into DB2. This module also generates the index pages that point to the articles.
 5. The Textract module⁸ consists of language processing algorithms. SuperNews uses Nominator to detect the proper nouns in the articles.
 6. The Media module checks for images in an LDAP (Lightweight Directory Access Protocol)-based media directory and returns an XML object with references to the images in DFS.
 7. The ActiveNews module, implemented in Java code, generates the input required by the ActiveNews Java applet.
 8. The front page module monitors incoming articles for categories that have been selected to appear on the front page because of their strategic importance to IBM. The module refreshes the SuperNews front page every hour.
 9. The personalization module, implemented in Perl and Java code, allows users to select categories for their news profile. The module generates a Web page or a daily e-mail with all articles that match the profile.
 10. A Lotus Notes discussion database and reader survey gather user feedback.
 11. A Lotus Go Web Server serves all pages to end users on the Web and runs the People’s Choice recommendation CGI (common gateway interface) scripts.

12. IBM's Network Dispatcher in front of the Web server is a safeguard to allow routing the SuperNews domain name to the replica system running on the staging server in case of a system outage on the production server. Network Dispatcher can also perform load balancing of user requests if more than one Web server were to serve SuperNews content.
13. A series of monitoring scripts notify the system administrator in case of database, file system, or feed processing failures.
14. All feed processing in the SuperNews system is scheduled using UNIX** crontab jobs. The jobs run each feed processing module based on the estimated frequency of updates to the feed.

When live, the SuperNews application ran on two IBM SP2* nodes under the AIX* (Advanced Interactive Executive) operating system. One node hosted the production version of SuperNews while the other hosted the exact replica as a staging version where further application development could take place. The staging server used the Distributed File System with read-only and read-write partitions, with changes released to the read-only partition used by the production system.

Evaluation. SuperNews ran on the IBM intranet in order for employees to experiment and evaluate new features for the official IBM intranet news application. Therefore, a discussion database that gathered user feedback and a reader survey were a critical part of the SuperNews application.

The discussion database served two purposes: users posted "bug" reports and made suggestions for features they wanted to see in the future. In the case of bug reports, the discussion database allowed the reader community to resolve some of the problems without any help from the developers. Since the prototype application had no official support mechanism or help line, readers who had already encountered the same problem could sometimes answer the questions directly.

The suggestions in the discussion database, along with a reader survey, gave the SuperNews team valuable feedback, in particular on the personalization pilot that ran from June 1998 to December 1998. The pilot started with 200 subscribers and grew to 4000 during the six-month period. The best-liked feature was the ability to receive daily news by e-mail. IBM

employees use Lotus Notes for e-mail and they preferred to read news in the same application instead of launching a Web browser. For mobile workers, replicating the news onto their local machines for off-line reading was critical. Since the body of the article was included in the e-mail, many readers found it the most convenient way to read news.

Users also liked the ability to subscribe to categories. Readers found the industry categorization from NewsEdge useful, but also liked the additional automatically generated SuperNews categories on specific persons, places, products, and organizations. Subscribing to these categories "discovered" by Texttract gave the readers the ability to track IBM products they might be working on, their IBM executives, or their IBM organization. Such items were not tagged by NewsEdge, nor by the internal news provider, which observes a fixed subject-matter taxonomy. A "discovered" category gave a reader a view across all sources, internal and external, and the ability to view the difference in the coverage of the topic.

The most common criticism of the personalization pilot was that readers could not define the categories themselves. Instead of subscribing to two categories, for example "e-business" and "WebSphere," users wanted to specify a Boolean expression, such as "e-business AND WebSphere." A Boolean expression would have given better focus and allowed a person to receive a smaller number of more relevant articles.

The reader survey gave useful feedback to the IBM intranet team which, at the time, was implementing the official intranet news application, MyNews. Meanwhile, the SuperNews team began to "roll out" the later features: People's Choice, personal columns, and ActiveNews articles. The People's Choice column had a few active participants, but overall did not catch on in the same way as the personalized news function. The personal columns attracted about 30 publishers who submitted e-mails to their columns with varying frequency. If these functions had been offered as part of the Lotus Notes e-mail delivery of news, the success rate inside IBM would probably have been higher. Until recently, only a minority of the IBM population spent much time on the Web.

Although neither the SuperNews team nor the Media Lab has conducted a formal evaluation of ActiveNews, useful comments were collected from SuperNews users. The general consensus was that these visualizations were initially very engaging, and aug-

menting the articles with relevant images was a big improvement. However, users found that this particular visualization was difficult to use for reading the news on a day-to-day basis.

The iterations of the SuperNews prototype, trying out new features and gathering user feedback, have been fruitful. The prototype has helped the IBM intranet news team make decisions about the direction to take before investing in full implementation.

Future work

SuperNews was built in an environment where distributed content providers suffered from a lack of cooperation and standardization of formats. The system demonstrated how to add value to news feeds *after* the editorial process. However, automated meta-data extraction and augmentation are not always accurate, due to the ambiguity of free text. If these steps in SuperNews were inserted into the editorial process, to support the editor in categorization and building of relevant links, the results would be more accurate, due to human intervention. Demonstrating the capabilities that add value to a news feed in an automated mode gives momentum to integrating this functionality earlier in the editorial chain.

Since SuperNews is now a stable information repository, it can be used as a test bed for various applications. One possible application is the multicasting of news, experimenting with alternative delivery strategies to reduce the load on the internal network at IBM.

Another project, called People Web, pushes ahead toward knowledge management. By discovering named relationships between people, products, and organizations in the news using Textract¹³ (e.g., Scott McNealy <-> CEO <-> Sun), a knowledge base of facts can be constructed. When a reader subscribes to or searches within a category, the knowledge base could suggest closely related information in SuperNews if strong relationships with the item of interest have been recorded.

A number of issues remain open in the ActiveNews visualization. Providing a database with a rich enough set of images to aid in the production of these presentations is a challenge. Another issue centers on usability. Although ActiveNews articles are engaging, readers find them hard to use as the main mode for reading news. Therefore, an interesting next step

would be to explore ways to increase the usability of ActiveNews articles, while still retaining their engaging visual quality. Finally, building a style sheet for an ActiveNews article is a tedious process. Currently, each image must be given a position and scale factor for each state within the presentation. An important extension would be to provide a tool to aid in the layout of these templates. This tool could also support the construction of handcrafted presentations by allowing the user to define the template by direct manipulation.

Related work

Many on-line news services that complement a print version present the reader with fixed categories such as politics, business, sports, and travel. For example, on-line versions of *The New York Times*¹⁴ and *San Francisco Chronicle*¹⁵ take advantage of the manually edited sections in their print version. Some services, such as Yahoo!**,¹⁶ use automated keyword lists to group articles into predetermined categories. Others, such as NewsPage¹⁷ use automated means combined with an editorial eye to categorize news from several feeds into predetermined hierarchies. SuperNews differs from all of them, because it complements the categories created by editors with four types of automatically created categories: persons, products, places, and organizations. Because these categories reflect the topics in the news, their number varies from week to week.

SuperNews uses Nominator's knowledge of proper nouns to augment articles and to group articles with the same names together. Among natural language processing techniques, more sophisticated ways exist to deduce the meaning of text or to group similar articles together. The approaches can be roughly divided into two schools of thought.

Some researchers, such as Gerald Salton of Cornell University, use statistical techniques in processing text.¹⁸ For example, word frequencies and co-occurrences help determine if one document is a summary of another. They also help discover clusters of related documents. Clustering differs from "supervised" classification methods in one significant way: a clustering algorithm does not require predefined categories with sample documents in order to learn. Instead, it automatically discovers appropriate groupings using the vocabulary of the articles. For example, if several news categories were to carry stories on the naming and promoting of IBM employees to new positions, clustering would group these

articles together because of their similar vocabulary. When a new subject emerges in the news, editors have to fit the articles into existing categories. While they may later decide to add a new category, clustering would automatically identify any new prominent subject and group relevant articles within it.

While statistical techniques generalize, to some extent, across language boundaries, the second approach to natural language processing relies on the grammar and vocabulary of a specific language. This approach grew out of the artificial intelligence community of the 1970s and 1980s and aims to understand the detailed story told in text: who did what, when, and where? Roger Schank's early work¹⁹ exemplifies this approach. Such story understanding is evaluated by asking questions to see if the system can draw inferences. These techniques often incorporate a parser or a part-of-speech tagger to recognize the grammatical role of words.

Other work in story understanding seeks to accomplish even more specific tasks. A "text interpreter" aims to accurately extract predefined facts from text on a given topic. Ideally, such a system could enter facts directly into a database from large collections of text on a constrained domain. Text-interpreting systems compete against each other in conferences on message understanding. The systems have to interpret texts on topics such as terrorism or military reports as fast and accurately as possible.

SuperNews takes a slightly simpler approach to natural language processing. Instead of trying to "understand" the meaning of an article, it detects who, what organizations, locations, and products are mentioned, and builds links to relevant resources. SuperNews creates a meaningful context and turns a disconnected piece of text into a gateway to related information.

Recommendation systems have been available on the Web since the early 1990s when one of the first music recommendation systems demonstrated collaborative filtering.²⁰ This technique requires users to first rate items within a narrow subject area, such as music CDs, movies, or books. Using the ratings of other users with similar tastes, the recommendation system then suggests items that the user has not yet heard, seen, or read. To be accurate, collaborative filtering requires users to rate as many items as possible. In addition, it assumes that at least some of the users rate the same items. Since several thousand news articles appear in SuperNews daily, two

readers may rarely read the same article. Therefore, SuperNews implements a much simpler technique: the recommenders explicitly target a job category for each recommendation.

Flyswat,²¹ a free service on the Internet, incorporates an augmentation technique similar to that of SuperNews. A user installs into a Web browser or a Microsoft Windows** application a plug-in that automatically analyzes retrieved Web pages or selected text in the application. Flyswat detects items such as companies, sports, movies, music, persons, places, health, cars, computers, and products, and links them to related Web sites. Because Flyswat relies on a predefined database of names and terms, it detects fewer items than Textract does. If a new useful name or term appears in the text, Flyswat does not recognize it until it is added to its database.

Conclusion

SuperNews brings into one consistent interface a range of disparate news sources offered on the IBM intranet. By transforming each article into a generic XML format, SuperNews can efficiently generate many presentations of the same article. By applying XSL style sheets, the article is delivered as an augmented Web page, as plain text for e-mail delivery, or as an engaging interactive ActiveNews article. The process of augmenting an article links proper nouns discovered in the article to related information on other Web sites. SuperNews can turn passive readers into active participants by encouraging them to publish their personal columns. Employees can also recommend and comment on articles they come across, for their colleagues.

SuperNews is a result of influences from text processing, digital news, and visualization research, and the practical need to provide relevant news on the IBM intranet. The main research influences originate from IBM Research (Textract) as well as from the MIT Media Lab (ActiveArticle, FishWrap, PLUM, VIA).

SuperNews ran as an experimental application on the IBM intranet from June 1998 to December 1999 to solicit feedback from the company's employees. The objective was for users to evaluate new features before investing in full development in the production news application. SuperNews also demonstrated what is possible if content providers abide to common data formats and categorization schemes.

Acknowledgments

Over the 18 months of creating SuperNews and ActiveNews visualizations, many people have contributed to implementing or providing crucial input and design ideas. We thank Maria Arbusto, Marcus Beattie, Bill Brockner, Peter Davis, Jim Doran, Chuck Dorris, Christopher Fry, Greg Geiselhart, Ed Geraghty, Dave Grossman, Maria Hernandez, Dan LaDore, Kieran Lal, Raymond Lee, Cliff Liang, Allandel Manipon, Dipen Mehta, Jeff Mueller, Julius Quiaot, Yael Ravin, Melissa Sader, Dana Spiegel, Andy Stanford-Clark, Matthew Stokes, Bill Sweeney, Ning Tang, Phillip Tionson, Lawrence Wang, Lauren Zack, and Nianjun (Joe) Zhou.

Appendix

Here is the XML representation of a SuperNews article with tags for proper names identified in the text body.

```
<?xml version="1.0" standalone="yes"?>
<SuperNewsArticle>
  <title>
    XML - the future markup language?
  </title>
  <source>NewsEdge Corp</source>
  <date>1999-06-11</date>
  <body>
    <paragraph>JUST AS the world is coming to terms
      with the electronic commerce revolution,
      the computing world really seems to be
      excited about a new language called
      <proprname
        type ="product"
        cname ="XML"
        info ="http://supernews.webahead.ibm.com/
          Summaries/product/X/XML.html"
        Image ="http://media.webahead.ibm.com/
          what/(cn=XML)(whattype=Product).list?
          imageurl"/>
      XML
      </proprname>
      or Extensible Markup Language that many
      people look upon to fulfill the promises of
      E-commerce.
    </paragraph>
    <paragraph>
      <proprname
        type ="organization"
        cname ="IBM"
        info ="http://supernews.webahead.ibm.com/
          Summaries/product/I/IBM.html"
        Image ="http://media.webahead.ibm.com/
          what/(cn=IBM)(whattype=Company).list?
          imageurl"/>
      International Business Machines Corp
      </proprname>
      has recently launched a web site
      to bring together
      <proprname
        type ="product"
        cname ="XML"
        info ="http://supernews.webahead.ibm.com/
          Summaries/product/X/XML.html"
        Image ="http://media.webahead.ibm.com/
          what/(cn=XML)(whattype=Product).list?
          imageurl"/>
```

```
XML
</proprname>
resources.
<proprname
  type ="organization"
  cname ="IBM"
  info ="http://supernews.webahead.ibm.com/
    Summaries/product/I/IBM.html"
  Image ="http://media.webahead.ibm.com/
    what/(cn=IBM)(whattype=Company).list?
    imageurl"/>
  IBM
</proprname>
's new site can be viewed at
http://www.ibm.com/xml.
</paragraph>
</body>
</SuperNewsArticle>
```

*Trademark or registered trademark of International Business Machines Corporation.

**Trademark or registered trademark of Sun Microsystems, Inc., Lotus Development Corporation, The Open Group, Yahoo! Inc., or Microsoft Corporation.

Cited references

1. World Wide Web Consortium, Extensible Markup Language, <http://www.w3.org/XML>.
2. W. Bender, P. Chesnais, S. Elo, A. Shaw, and M. Shaw, "Enriching Communities: Harbingers of News in the Future," *IBM Systems Journal* **35**, Nos. 3&4, 369–380 (1996).
3. S. Elo, *PLUM: Contextualizing News for Communities Through Augmentation*, master's thesis, MIT, Media Arts and Sciences, Cambridge, MA (1996).
4. P. Tionson, ActiveArticle (1998), available at <http://ic.media.mit.edu/~phillip/activearticle>.
5. World Wide Web Consortium, Extensible Stylesheet Language, <http://www.w3.org/Style/XSL>.
6. L. Weitzman, *The Architecture of Information: Interpretation and Presentation of Information in Dynamic Environments*, Ph.D. thesis, MIT, Media Arts and Sciences, Cambridge, MA (1995).
7. Information Retrieval and Analysis Group, Texttract description available at <http://www.research.ibm.com/irgroup/talent.html>.
8. Y. Ravin and N. Wacholder, *Extracting Names from Natural-Language Text*, IBM Research Report 20338 (1996), available at <http://www.research.ibm.com/people/r/ravin/>.
9. P. Tionson, L. Weitzman, and L. Wang, "ActiveNews: Automatic Creation of Dynamic News Articles," *IEEE Visualization '99*, San Francisco, CA (October 27–29, 1999).
10. G. Furnas, "Generalized Fisheye Views," *Proceedings of the Conference on Human Factors in Computing Systems*, Boston, MA (April 13–17, 1986), pp. 16–23.
11. S. Greenberg, C. Gutwin, and A. Cockburn, "Awareness Through Fisheye Views in Relaxed-WYSIWIS Groupware," *Proceedings of the Graphics Interface Conference*, Toronto, Canada (May 21–24, 1996), pp. 28–38.
12. NewsEdge, see <http://www.newsedge.com>.
13. R. Byrd and Y. Ravin, "Identifying and Extracting Relations in Text," *Proceedings of the Applications of Natural Language to Information Systems Conference*, Klagenfurt, Austria (June 17–19, 1999), pp. 149–154.
14. *The New York Times*, see <http://www.nytimes.com>.
15. *San Francisco Chronicle*, see <http://www.sfgate.com>.
16. Yahoo!, see <http://my.yahoo.com>.

17. NewsPage, personalized news service, <http://www.individual.com>.
18. G. Salton, *An Introduction to Modern Information Retrieval*, McGraw-Hill, New York (1983).
19. R. Schank and R. P. Abelson, *Scripts, Plans, Goals, and Understanding*, Lawrence Erlbaum Associates, Inc., Mahwah, NJ (1977).
20. U. Shardanand, *Social Information Filtering for Music Recommendation*, master's thesis, MIT, Cambridge, MA (1994).
21. Flyswat, see <http://www.flyswat.com/>.

Accepted for publication May 11, 2000.

Sara Elo Dean *IBM Advanced Internet Technology Group, 590 Madison Avenue, New York, New York 10022 (electronic mail: saraelo@us.ibm.com)*. Ms. Elo Dean joined IBM's Advanced Internet Technology group as information architect for the 1996 Atlanta Olympic Games Web site. Since then, she has contributed to the IBM patent site and the Intranet Taxonomy project, and designed and implemented several technology prototypes on the IBM intranet, such as SuperNews news service, BluePages employee directory, MyIBM personalized portal, Virtual Resource Locators, and most recently Franklin XML publishing system. Before joining IBM, she completed her bachelor's and master's degrees in computer science at the University of Geneva, and a master's degree in media arts and sciences at the MIT Media Lab. She has also worked as a software consultant in the jewelry industry in Switzerland and the health industry in India. Her interests include natural language processing, content representation and XML, electronic publishing, adaptive design, and personalization.

Louis Weitzman *IBM Advanced Internet Technology Group, 150 Kettletown Road, Southbury, Connecticut (electronic mail: louisw@us.ibm.com)*. Dr. Weitzman has worked at the intersection of design and computers for over 25 years. Interested in the computational support for the design process, he has investigated tools and techniques to support designers both interactively and automatically. While a student at MIT's Media Lab, he investigated the use of visual languages as a formalism to represent and enable design. At the Media Lab's Visible Language Workshop, he was Muriel Cooper's first Ph.D. student. Dr. Weitzman also holds a bachelor of architecture degree from the University of Minnesota and a master of architecture degree in advanced studies from MIT's Architecture Machine Group. He is currently a member of IBM's Advanced Internet Technology Group, building applications that try to leverage the power of computers while maintaining focus on the principles of good design. Recent projects include the Franklin XML publishing system, MyIBM personalized portal, Virtual Resource Locators, and interactive visualizations for SuperNews articles and ActiveProduct browsing.