Personalcasting:
Tailored Broadcast News

Mark Maybury, Warren Greiff, Stanley Boykin, Jay Ponte[1]*, Chad McHenry and Lisa Ferro*
Information Technology Division
The MITRE Corporation
202 Burlington Road
Bedford, MA 01730, USA
{maybury, greiff, boykin, red, lferro}@mitre.org
www.mitre.org/resources/centers/it

ABSTRACT

Broadcast news sources and newspapers provide society with the vast majority of real-time information. Unfortunately, cost efficiencies and real-time pressures demand that producers, editors, and writers select and organize content for stereotypical audiences. In this article we illustrate how content understanding, user modeling, and tailored presentation generation promise personalcasts on demand. Specifically, we report on the design and implementation of a personalized version of a broadcast news understanding system, MITRE's Broadcast News Navigator (BNN), that tracks and infers user content interests and media preferences. We report on the incorporation of Local Context Analysis to both expand the user's original query to the most related terms in the corpus, as well as to allow the user to provide interactive feedback to enhance the relevance of selected news stories. We describe an empirical study of the search for stories on ten topics from a video corpus. By personalizing both the selection of stories and the form in which they are delivered, we provide users with tailored broadcast news. This individual news personalization provides more fine-grained content tailoring than current personalized television program level recommenders and does not rely on externally provided program metadata.

Keywords: broadcast news, story selection, personalization, user modeling, query expansion, relevance feedback

[1] Formerly of MITRE, Jay Ponte's current affiliation is Google, Inc., ponte@google.com.

1. PERSONALCASTING

People are offered vast quantities of news in the form of multiple media (text, audio, video). For the past several years, a community of scientists has been developing news-on-demand algorithms and technologies to provide more convenient access to broadcast news (Maybury 2000). Applications promising on-demand access to multimedia information such as radio and broadcast news on a broad range of computing platforms (e.g., kiosk, mobile phone, PDA) offer new engineering challenges. Unlike earlier systems which require television content to be manually annotated (e.g., Bove 1983), more recent systems have been developed that automatically index, cluster/organize, and extract information from news. Synergistic processing of speech, language and image/gesture promise both enhanced interaction at the user interface, and enhanced understanding of electronic media such as web, radio, and television sources (Maybury 2000).

Unlike traditional broadcasts, which are created in a standard format delivered from a single source (broadcaster) with general content to address a wide range of audience and interests, we define a *personalcast* as a custom created, interactive sequence of stories that are selected based upon specific, individual user interests from a variety of sources and presented in a form tailored to user preferences. Whereas a broadcast is disseminated from one to many, a personalcast is one to one.

In separate research, investigators have developed many different ways to adapt presentations (Brusilovsky 1996, 2001) including adapting the navigation support (e.g., sorting links, adding/removing/disabling links) or adapting the presentation itself. It has been shown empirically that adaptation can help, in both the speed of navigation/search (Kaplan et al. 1993) as well as in enhancing text understanding (Boyle and Encarnacion 1994). One would hope the same benefits could accrue with tailored news. Finally, personalization of electronic programming guides (EPGs) (e.g., Ardissono et al. 2001) promises more rapid and custom access to shows of interest to the user.

We claim that a valuable scientific and technological advance would be the integration of methods from user modeling and user-adapted interaction together with news understanding technologies. This combination could enable new services such as the delivery of story alerts to interested users as well as interactive, individual news programs which we call personalcasts. This research is distinct from personalized EPGs in several primary ways:

1. *No metadata*. Metadata about content (programs and stories therein) is unavailable and must be automatically extracted from video sources via speech, text, and image processing.
2. *Linguistic analysis of user query*. User queries are analyzed linguistically, considering the available large broadcast news corpus.
3. *Story-level access*. Tailoring is performed at the story level, and not the program level.
4. *Analytic user tasks*. User tasks are primarily focused on information analysis, and not on entertainment or enjoyment. Users primarily work individually and not in groups.

In addition, this research adds to our knowledge of interactive search and query refinement. Salton and Buckley (1990) showed that automated relevance feedback improves search. Koenemann and Belkin (1996) and Koenemann (1996) showed that relevance feedback that is used for query refinement improves search of text news articles. While we also explore interactive relevance feedback, this research differs in several principal ways.

1. *Short and errorful video stories*. Users are searching short, errorful video broadcast news stories as opposed to longer, edited newspaper articles. The average length of a video story in our corpus is 51 seconds or only 122 words per story. On average, stories contain 5.9 named entities (i.e., proper names such as people, organizations, locations) per story, 4.7 of which are distinct names. Automated speech or manual human transcription introduces significant errors into our corpus which can reduce the performance of automated story segmentation, retrieval algorithms, and human relevance judgments. Stories are also interspersed with (irrelevant) commercials.
2. *Multiple topics.* Whereas prior research tested user performance on 20-minute search tasks on two TREC topics (Topic #162 - Automobile Recalls and Topic #165 - Tobacco

Advertising and the Young) from 75,000 Wall Street Journal articles, we explored ten topics selected from a multiyear corpus of nine news broadcasts. As a consequence, while we used fewer subjects (four), each subject was carefully measured in a fully instrumented environment across many topics.

3. *Rich annotation***.** Our corpus includes stories that are richly annotated with hierarchically organized topics and named entities.

2. BROADCAST NEWS NAVIGATOR

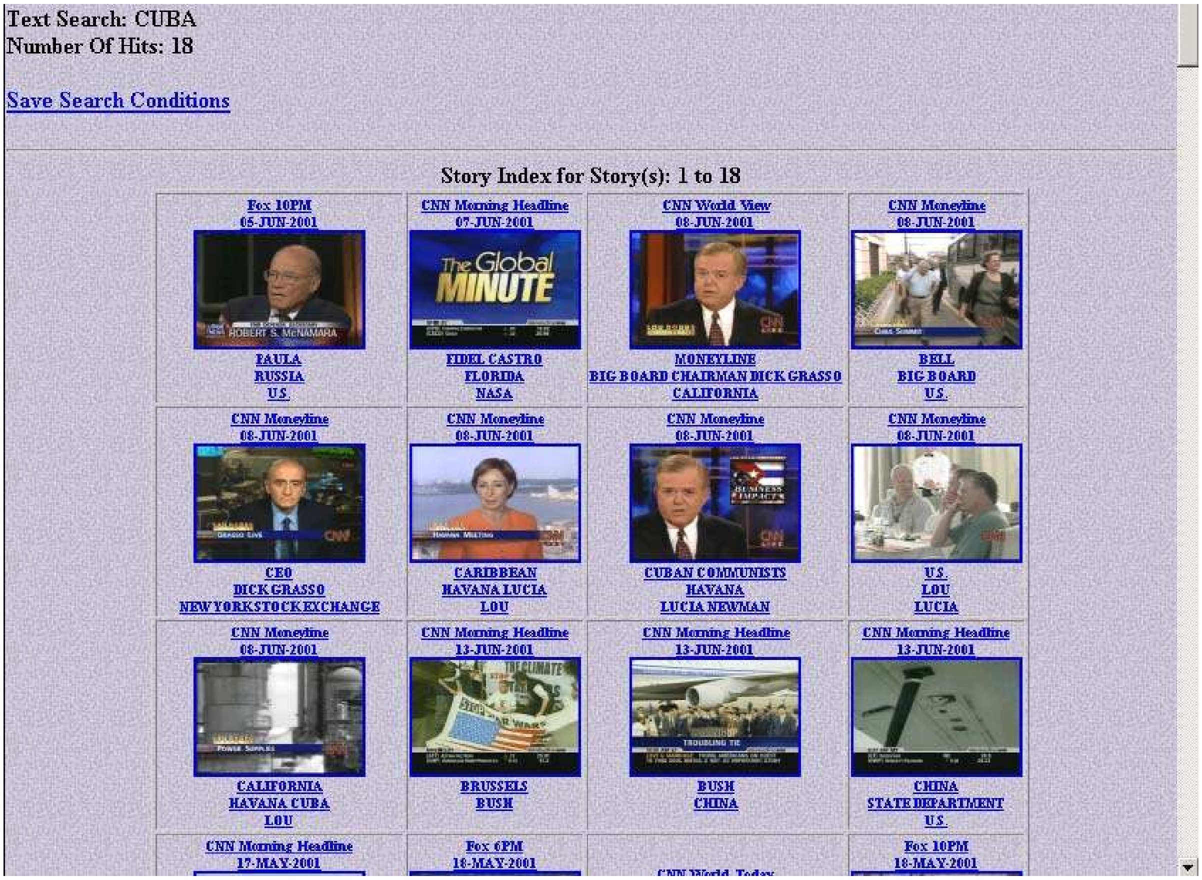
To illustrate personalcasting as defined above, we describe the Broadcast News Navigator (BNN). In our research, we have created BNN, a system that exploits video, audio, and closed-caption text information sources to automatically segment, extract, and summarize news programs (Maybury et al. 1997). Figure 1a shows the results of BNN responding to a user query requesting all reports regarding "Cuba" between May 17 and June 16, 2001. For each story matching the query, the system presents a key frame, the three most frequent named entities within the story, and the source and date of the story. This display is called a "Story Skim".

This, in essence, provides the user with a "Cuba" channel of information, personalizing the channel to their information interests. Moreover, the user can create arbitrarily complex queries combining key words, named entities (e.g., people, organizations, locations), sources (e.g., CNN, MS-NBC, ABC), programs (e.g., CNN Prime News vs. CNN Headline New vs. CNN Moneyline, etc.), and time intervals (e.g., specific days, weeks or years). These queries result in selected video stories specific to their interests.

The user can then select any of the key frames to get access to details of the story, such as shown in Figure 1b. In this "Story Details" presentation, the user has access to all people, organizations and locations mentioned in the story, an automatically extracted one-line summary of the news (the sentence with the most frequently named entities), a key frame extracted from the story segment, and a pointer to the full closed-caption text and video source for review. An empirical evaluation previously reported in Merlino and Maybury

(1999) demonstrated that users could enhance their retrieval performance (a weighted combination of precision and recall) by utilizing BNN’s Story Skim and Story Details presentations. User satisfaction in that study was 7.8 for retrieval and 8.2 for mixed media display (e.g., story skim, story details, such as those shown in Figure 1a and 1b), on a scale from 1=dislike to 10=like.

The BNN system provides navigation support, so that the user can select named entities and find stories including them. Further, by employing a clustering algorithm, the system enables the user to select stories similar to the current story.



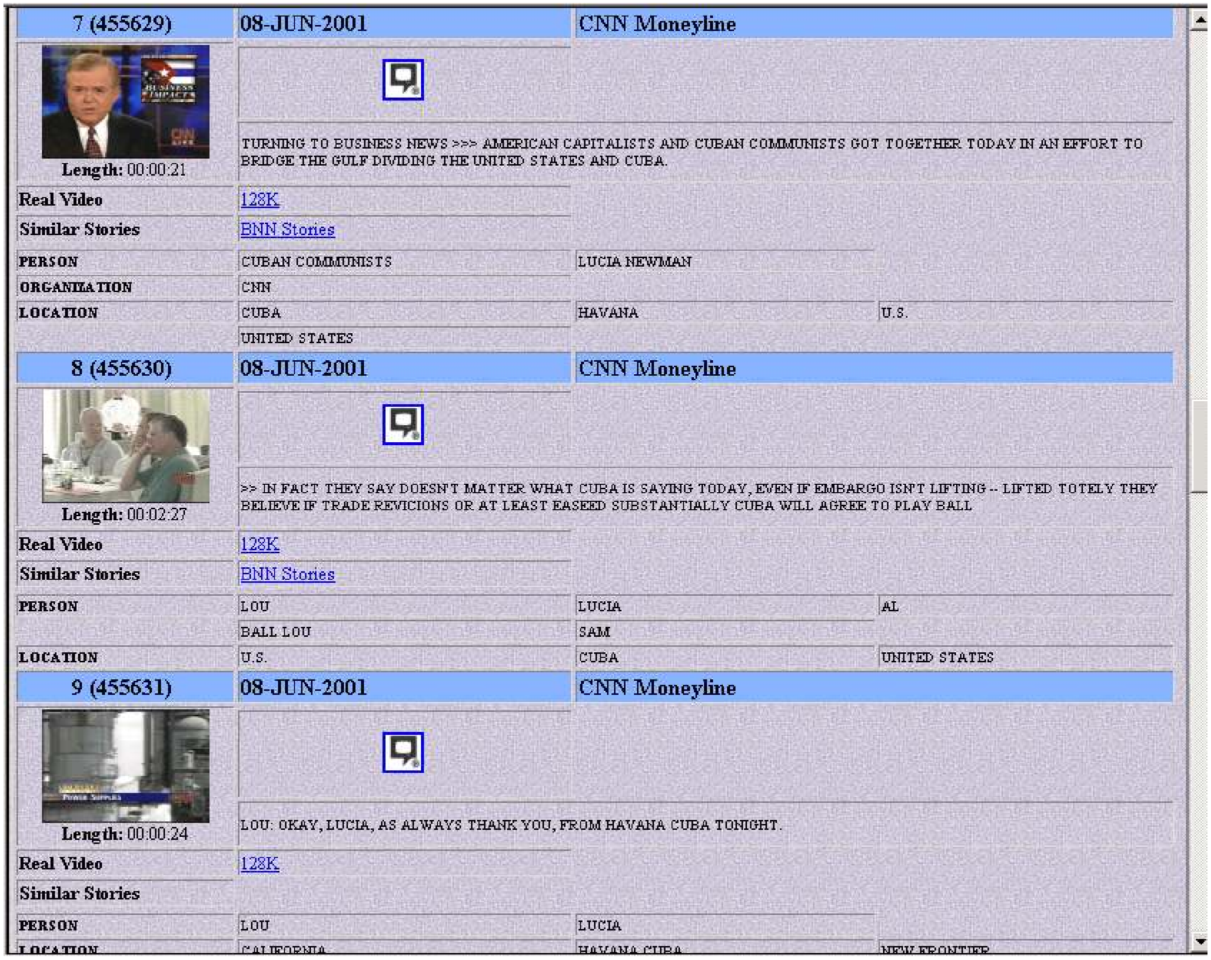
Text Search: CUBA
Number Of Hits: 18

Save Search Conditions

Story Index for Story(s): 1 to 18

Fox 10PM 05-JUN-2001 PAULA RUSSIA U.S.	CNN Morning Headline 07-JUN-2001 FIDEL CASTRO FLORIDA NASA	CNN World View 08-JUN-2001 MONEYLINE BIG BOARD CHAIRMAN DICK GRASSO CALIFORNIA	CNN Moneyline 08-JUN-2001 BELL BIG BOARD U.S.
CNN Moneyline 08-JUN-2001 CEO DICK GRASSO NEW YORK STOCK EXCHANGE	CNN Moneyline 08-JUN-2001 CARIBBEAN HAVANA LUCIA LOU	CNN Moneyline 08-JUN-2001 CUBAN COMMUNISTS HAVANA LUCIA NEWMAN	CNN Moneyline 08-JUN-2001 U.S. LOU LUCIA
CNN Moneyline 08-JUN-2001 CALIFORNIA HAVANA CUBA LOU	CNN Morning Headline 13-JUN-2001 BRUSSELS BUSH	CNN Morning Headline 13-JUN-2001 BUSH CHINA	CNN Morning Headline 13-JUN-2001 CHINA STATE DEPARTMENT U.S.
CNN Morning Headline 17-MAY-2001	Fox 6PM 18-MAY-2001	CNN World Today	Fox 10PM 18-MAY-2001

Figure 1a. Automated Retrieval of Cuba Stories (*Story Skim*)



7 (455629)	08-JUN-2001	CNN Moneyline	
Length: 00:00:21	TURNING TO BUSINESS NEWS >>> AMERICAN CAPITALISTS AND CUBAN COMMUNISTS GOT TOGETHER TODAY IN AN EFFORT TO BRIDGE THE GULF DIVIDING THE UNITED STATES AND CUBA.		
Real Video	128K		
Similar Stories	BNN Stories		
PERSON	CUBAN COMMUNISTS	LUCIA NEWMAN	
ORGANIZATION	CNN		
LOCATION	CUBA	HAVANA	U.S.
	UNITED STATES		
8 (455630)	08-JUN-2001	CNN Moneyline	
Length: 00:02:27	>> IN FACT THEY SAY DOESN'T MATTER WHAT CUBA IS SAYING TODAY, EVEN IF EMBARGO ISN'T LIFTING -- LIFTED TOTELY THEY BELIEVE IF TRADE REVICIONS OR AT LEAST EASEED SUBSTANTIALLY CUBA WILL AGREE TO PLAY BALL		
Real Video	128K		
Similar Stories	BNN Stories		
PERSON	LOU	LUCIA	AL
	BALL LOU	SAM	
LOCATION	U.S.	CUBA	UNITED STATES
9 (455631)	08-JUN-2001	CNN Moneyline	
Length: 00:00:24	LOU: OKAY, LUCIA, AS ALWAYS THANK YOU, FROM HAVANA CUBA TONIGHT.		
Real Video	128K		
Similar Stories			
PERSON	LOU	LUCIA	
LOCATION	CALIFORNIA	HAVANA CUBA	NEW FRONTIER

Figure 1b. Details of some Cuba Stories (*Story Details*)

Television, like other media such as newsprint, radio and the web, is used both for entertainment and informational/educational purposes which range from public health warnings to educational networks such as the Biography or Discovery channels. Focusing on the latter, BNN supports a range of users, topics, and usage models. Users have included experts (e.g., intelligence analysts, political analysts, systems engineers) and casual users (e.g., scientists, managers, secretaries), although our focus has been on expert users. As exemplified by the evaluation topics presented in the Appendix to this article, topics range from ones that interest users of all ages (such as music, sports, weather, and space) to social/adult issues (such as accidental injuries, crime, gambling and investing) and to expert-specific topics (such as bioterrorism and the Mideast conflict).

The BNN prototype has been used in the MITRE Corporation for research, in the military for daily information monitoring, and for open source intelligence analysis. Current users perform both standing queries and ad-hoc searches. A user can subscribe to be alerted (e.g., via email) when a story containing a keyword or named entity is found.

Corporations are beginning to address user needs for content based video access. For example, Virage's VideoLogger® and Virage Solution Server™ (www.virage.com) incorporates concepts similar to those found in BBN for use by broadcasters, corporate trainers, and analysts. Virage's news on demand system, ViTAP (Video Text and Audio Processing), is used by Government Analysts for retrieval and profile-based alerting of events (Merlino 2002). It incorporates presentation techniques similar to those found in the BNN, augmented with time synchronized playback tracks for machine translation and keyframes. Related research includes meeting and lecture archiving and retrieval (e.g., Hu 2003).

The large broadcasters (e.g., CNN, ABC, BBC, C-NBC) continuously repurpose material. Sometimes local programs from the same broadcaster (e.g., CNN) are repurposed to international (CNN International) or Internet services (CNN Interactive). While current indexing is principally manual, a tool like BNN automates the discovery of specific topics and semi-automates story clip selection.

Video news segmentation performance ranges from 50 to 80% balanced precision and recall. In particular, segmentation algorithms using multimodal cues and trained on a range of broadcast sources such as CNN, MS-NBC or ABC perform with 53% precision and 78% recall (Boykin and Merlino 1999). Broadcast specific models (e.g., ones using visual anchor booth recognition cues specific to a particular program such as ITN) raise the performance to 96% precision and recall. Story segmentation may occur at the source (e.g., by the broadcaster), at an intermediary (e.g., broadband service provider), or by the end user.

As we discuss in the subsequent sections, the BNN system has been extended to support personalization as well as query expansion using Local Context Analysis (LCA). While many user studies with query expansion have been conducted in the past (Attar and Fraenkel 1977; Croft and Harper 1979; Koenemann and Belkin 1996; Koenemann 1996, Xu and Croft 1996, 2000), this research represents the first user study using LCA, the first exploration of LCA in the context of multi-media retrieval on broadcast news (a form of television), and the first user study of personalization and content-based video access.

3. USER MODELING AND TAILORING

The control flow diagram in Figure 2 shows a traditional search session in BNN. The user poses a query and receives a story skim of the kind shown in Figure 1a. The user then selects a story and is provided the details as exemplified in Figure1b. From this story detail, the user can simply review the summary and all named entities, or explicitly choose a media element to display (such as the full video source or the text transcript). User interest profiles can be created from explicit user input and then used to tailor presentations to the user's interests and preferences.

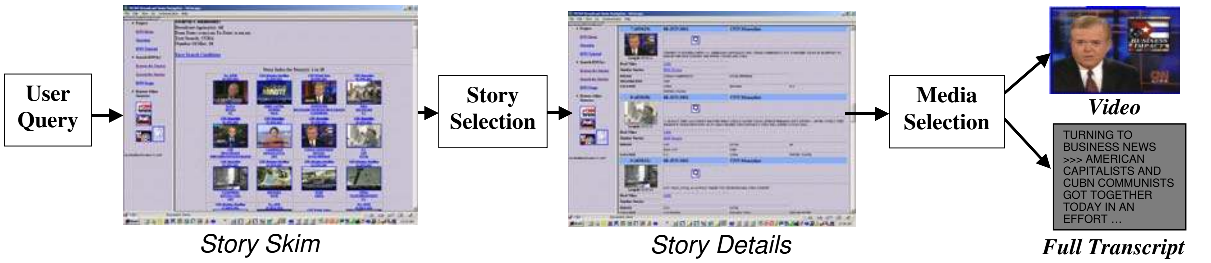
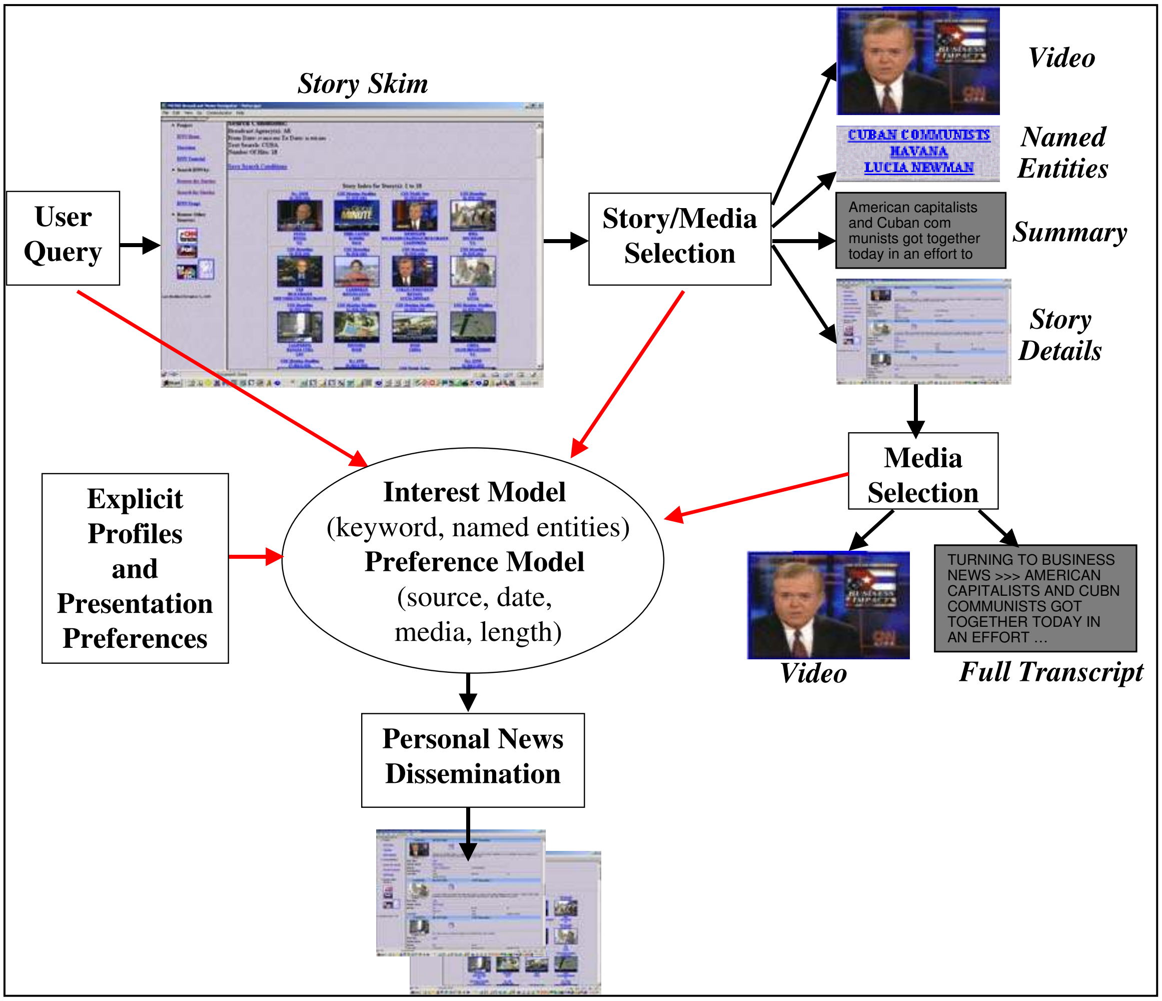

User Query
Story Selection
Media Selection
Video
TURNING TO BUSINESS NEWS >>> AMERICAN CAPITALISTS AND CUBN COMMUNISTS GOT TOGETHER TODAY IN AN EFFORT ...
Story Skim
Story Details
Full Transcript

Figure 2. Traditional Searching using BNN

As shown in Figure 3, in Personalized BNN (P-BNN) users can explicitly define user profiles indicating their interests by specifying simple keywords or named entities such as individuals, locations, or organizations. They can also specify preferred broadcast sources to search (e.g., CNN, ABC News). This profile also captures preferences for controlling the display of various media elements by manipulating media properties such as the source, date, time, length, and preference type for media presentation (e.g., key frame only, story details, full video, text summary). The user's interest profiles can be run periodically and the retrieval results sent to the requester as an alert. This alert can be pointers to a story, a story skim or story details like those shown in Figures 1a and 1b.



American capitalists and Cuban com munists got together today in an effort to

TURNING TO BUSINESS NEWS >>> AMERICAN CAPITALISTS AND CUBN COMMUNISTS GOT TOGETHER TODAY IN AN EFFORT ...

Figure 3. User Modeling and Tailored Presentation in Personalized BNN

Because the original broadcast news source is segmented into its component parts, key elements can be extracted and others summarized. This enables a system not only to select stories based on a user's content interest, but also to assemble them in the manner a user prefers. For example, the user can be presented with only a key frame, with summary sentences, with people or place names, or with the entire source. A natural extension of this work would be the addition of a feedback and collaborative filtering mechanism. An individual user's model could then be modified with each search or within or across sessions, and the user could benefit from searches performed by others in a community.

4. QUERY EXPANSION AND RELEVANCE FEEDBACK: LOCAL CONTEXT ANALYSIS

We use two methods to find information relevant to users' information needs. First, we expand their original query to the most related terms in the corpus. Second, we allow them to provide relevance feedback, so we can provide stories more similar to those they indicated as being relevant. Neither of these methods takes the user's interest model into account that was described in Section 3. For query expansion, we use a technique called Local Context Analysis (LCA) (Xu and Croft 1996, 2000). Figure 4 illustrates the control flow of LCA use. Given a query specified by the user, the system selects those passages containing at least one of the terms and assigns to each of these passages a Retrieval Status Value (RSV) according to the scoring formula employed by the retrieval engine. For all experiments discussed in this article, the Okapi formula (Robertson and Walker 1994) was used. The Okapi formula is commonly considered one of the most robust and effective formulas developed to date for information retrieval. What is treated as a passage is application dependent. Passages can be paragraphs or paragraph-sized fixed windows of text. Sentences can also be treated as passages. At the other extreme, entire documents can be the passages used for LCA. In this article, a passage is considered as co-extensive with a news story. In the future, however, we plan to run experiments where passages are associated with smaller sections of news stories. This may prove to be more appropriate for collections containing a mix of longer and shorter story segments.

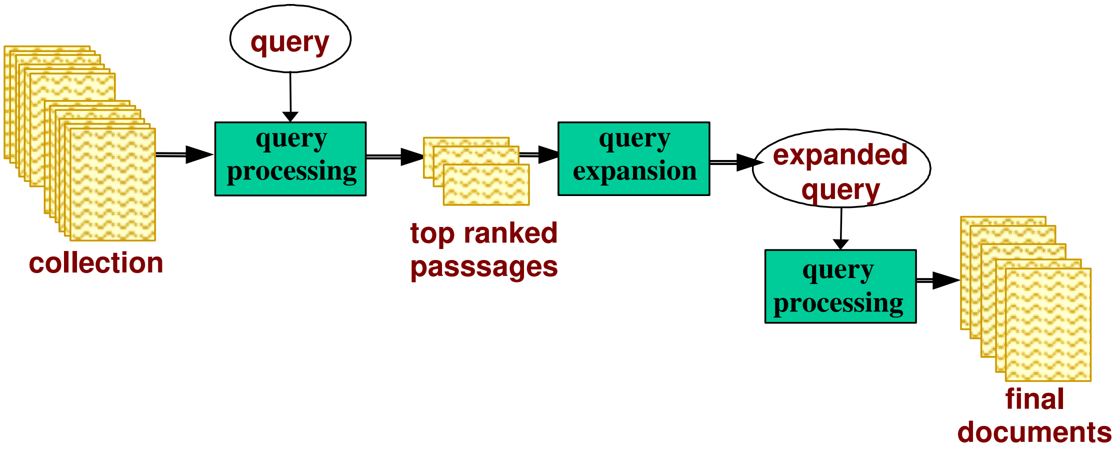


Figure 4. Control Flow for Local Context Analysis

The next step in the process is to mine the top passages for promising concepts that can be used as additional query terms. In the LCA of the BNN version discussed here, concepts are simple words. But concepts can correspond to any lexical, syntactic or semantic marking of documents. An obvious possibility in the context of BNN is to use named entities as concepts. Given an algorithm for the automatic association of named entity tags to snippets of text, named entities can serve as LCA concepts. Alternatively, concepts can be limited to some subset of named entities – persons, for example. An interesting possibility here is the provision of user control over the concept space on a per-query basis. User intuition may deem persons to be critical elements for one information need, where locations are likely to be most helpful for another. Syntactic units, such as noun phrases, can also serve as concepts, assuming the availability of a system component to provide the requisite syntactic analysis.

Once passages have been scored and ranked, LCA selects the top *N* ranked passages and considers all the concepts appearing at least one time in these top *N* passages. Each concept is then scored using the following formula:

$$f(c) = \prod_{i-1}^{M} \left[0.1 + \text{co\_degree}(c, w_i)\right]^{idf(w_i)}$$

$$\text{co\_degree}(c, w_i) = idf(c) \cdot \frac{\log(co(c, w_i) + 1)}{\log(N)}$$

$$co(c, w_i) = \sum_{p \text{ is a passage}} f_{p,w_i} \cdot f_{p,c}$$

$$idf(x) = \min\left(1.0, \frac{\log(N / N_x)}{5.0}\right)$$

The LCA formula for scoring concepts is designed to assign high values to concepts co-occurring with a large number of the query terms in a large number of the top ranked passages. The greater the number of passages, the greater the score. The greater the number of terms it co-occurs with in a given passage, the greater the score is incremented for that passage. The number of times these terms occur, as well as the number of times the concept itself occurs, also affect the degree to which a given passage augments the overall score.

For each of the M query terms, w_i, a value $co(c,w_i)$ is calculated. The *co* function measures how much the concept *c* co-occurs with term w_i. Each passage that contains both the concept and the term contributes a value. This value is equal to the product of the number of times *c* occurs and the number of times w_i occurs in the passage. The log of this measure (1 is added to avoid the possibility of taking the log of 0) is normalized relative to one occurrence of both *c* and w_i in every passage and then multiplied by *idf(c)*, giving the *co_degre*e. The *idf* statistic is a measure of how rare a word is. The *idf* fomula used for LCA, a variant of *idf* weighting used in most modern information retrieval systems, is a function of N_c, the number of passages containing the concept *c*, out of the total set of passages, which is of size *N*. The fewer passages containing the word, the greater the *idf* value. The *co_degree* is a measure of co-occurrence between the concept *c* and query word w_i. A weighted product of the *co_degrees* (weighted by the *idf* values for the query words) yields a measure of how valuable the candidate concept *c* is taken to be relative to the given query.

Once all concepts have been evaluated, a predetermined number of the most highly scoring concepts are chosen. These concepts are then added to the original query terms. If necessary, collection statistics – which may not be pre-computed for concepts as they are for simple query terms – are gathered for the expansion concepts. The enhanced query is then evaluated and the top ranking documents are retrieved. The following are two examples of query expansion terms resulting from LCA on the collection of news stories on which this study was based:

initial query1: **palestinian israeli conflict**

top 10 query1 expansion concepts:	**israel, violence, palestinians, hebron, sources, gaza, gunmen, tanks, city, hamas**
initial query2:	**bush budget**
top 10 query2 expansion concepts:	**surplus, medicare, funding, cut, tax, fall, spending, fought, fund, popular**

Although LCA was developed for automatic query expansion, and we have explained it in that context, the basic approach can be used in a number of different ways. First, its primary use in BNN is not for automatic query expansion (although this capability is provided), but for suggesting possible query expansion terms to the user, leaving to them the assessment of which combination of the suggested terms, if any, will be most beneficial if used as part of the query (see Fig. 5).

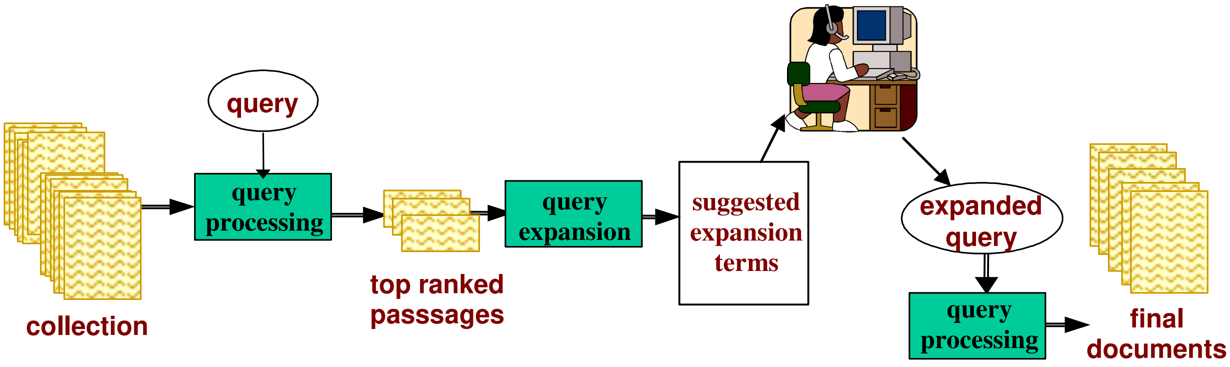


Figure 5. LCA Produces Candidate Query Terms

Second, LCA is considered for use as part of relevance feedback, as in Figure 6. The explanation of LCA given above can be understood as an application of pseudo-feedback, also known as blind feedback, or as it was called when it was originally proposed (Attar and Fraenkel 1977), local feedback. With pseudo-feedback, a query is evaluated and then, for the purposes of query expansion and term re-weighting, the top documents retrieved are treated as if they were known to be relevant. That is, they are treated as if these documents were shown to a user and the user feedback indicated that they were all relevant to the information need that motivated the original query. But given a human at the terminal, we need not depend on pseudo-feedback. Actual relevance feedback can be used instead. The user can be shown the most highly ranked news stories and asked to indicate which of them are indeed relevant to their information needs. Then query expansion can proceed as before, but only the stories marked as relevant by the user will be used for selecting expansion concepts in place of the top *N* stories resulting from the original query.

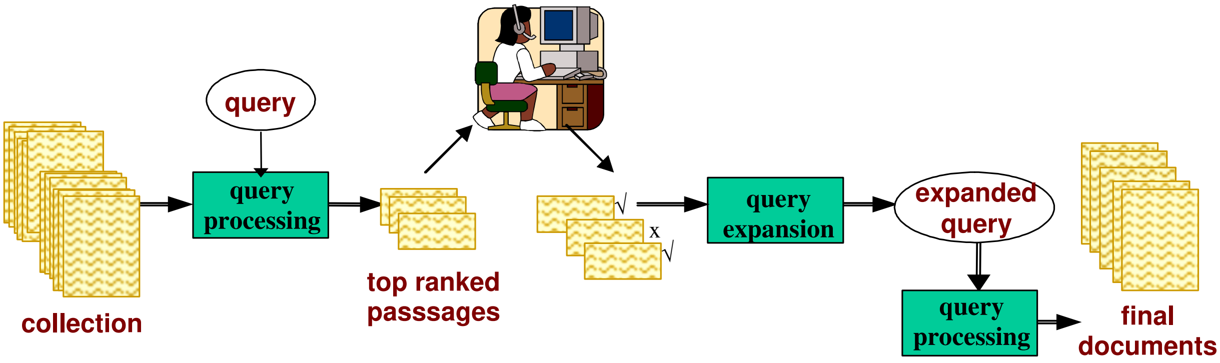


Figure 6. Relevance Feedback

These alternatives can be combined in various ways. BNN can be made to return both the top relevant documents and a list of suggested expansion terms in response to the initial query. The user can then choose to reformulate the query based on the list of suggested terms (and, possibly a quick review of the top ranked documents to get a sense of how the system responded to the initial query), or simply mark the retrieved documents as to relevance. In either case, the system can respond with both a new set of documents and an updated list of potential expansion terms. This cycle can be repeated any number of times. In addition, during any given interaction, the user can request that the system apply automatic query expansion and return the results of a pseudo-feedback cycle in place of the list of candidate expansion terms and the top ranked documents from the unexpanded query.

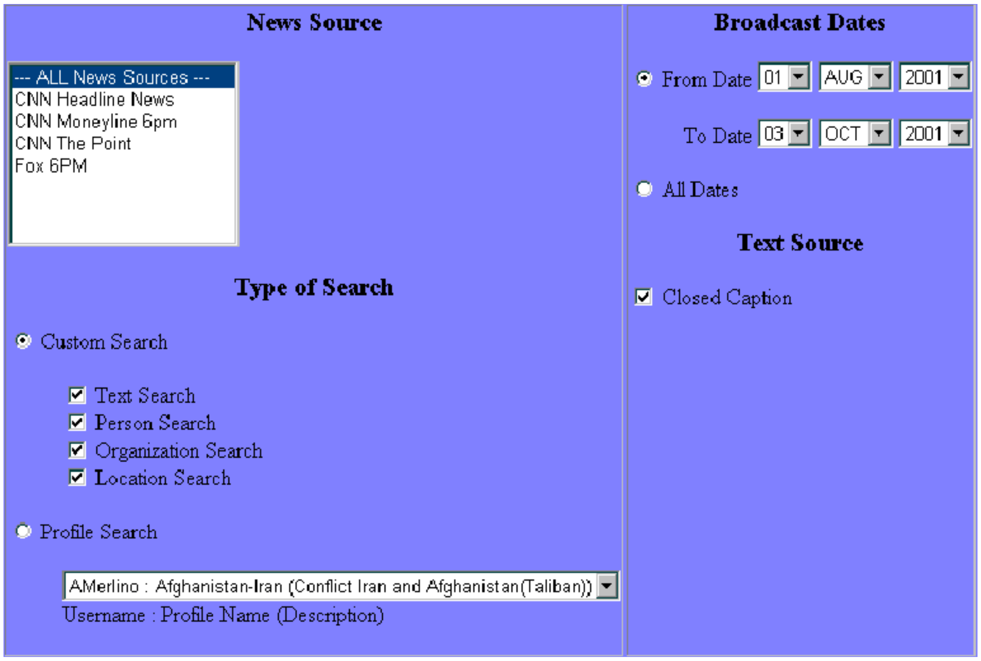
One potential limitation of this approach concerns the size of the broadcast news collection. Clearly, discovery of viable candidates for query expansion is dependent upon the existence of reliable co-occurrence statistics. In order that the statistics upon which LCA calculations are based be robust, they must be extracted from a reasonably sized corpus of related stories.

This can be problematic but need not be an insuperable barrier. A large number of stories with similar content from the same or similar sources can be presumed to be the ideal resource for uncovering quality expansion-term candidates. If this is not available, however, supplementary resources can be used. If the archive of broadcast news stories is not sufficiently large, but a large collection of, say, contemporaneous newspaper articles is available, the corpus of newspaper articles can be used for the mining of additional query terms, in place of, or in addition to, the broadcast news stories.

5. USER INTERESTS AND PREFERENCES

There are a number of methods that can be applied to create and exploit a model of user interests and preferences. Regarding user interests and/or information needs, typically these are captured in the form of user profiles explicitly stated by the user. This can occur in a BNN user profile in which a user can specify their information needs either as a list of keywords and/or a list of named entities; that is, people, organizations, or locations, as illustrated in Figure 7. Figure 7 is the first screen a user sees when they initiate a search in BNN and includes an ability to select program sources, dates, and type of search (e.g., keyword or named entities).

A drop list of stored profiles is displayed in the lower left hand corner of Figure 7. Visible is a stored profile for user "Amerlino" for stories related to conflicts between Afghanistan and Iran. A user can explicitly specify their preferences for particular sources (e.g., CNN, Fox, ABC News) or programs (e.g., CNN Headline News vs. CNN Moneyline), dates (e.g., last three days), time periods, sources (e.g., closed captions, speech transcripts), and type of search (e.g., keyword/text search or named entity search).



News Source

--- ALL News Sources ---
CNN Headline News
CNN Moneyline 6pm
CNN The Point
Fox 6PM

Type of Search

● Custom Search

☑ Text Search
☑ Person Search
☑ Organization Search
☑ Location Search

○ Profile Search

AMerlino : Afghanistan-Iran (Conflict Iran and Afghanistan(Taliban))
Username : Profile Name (Description)

Broadcast Dates

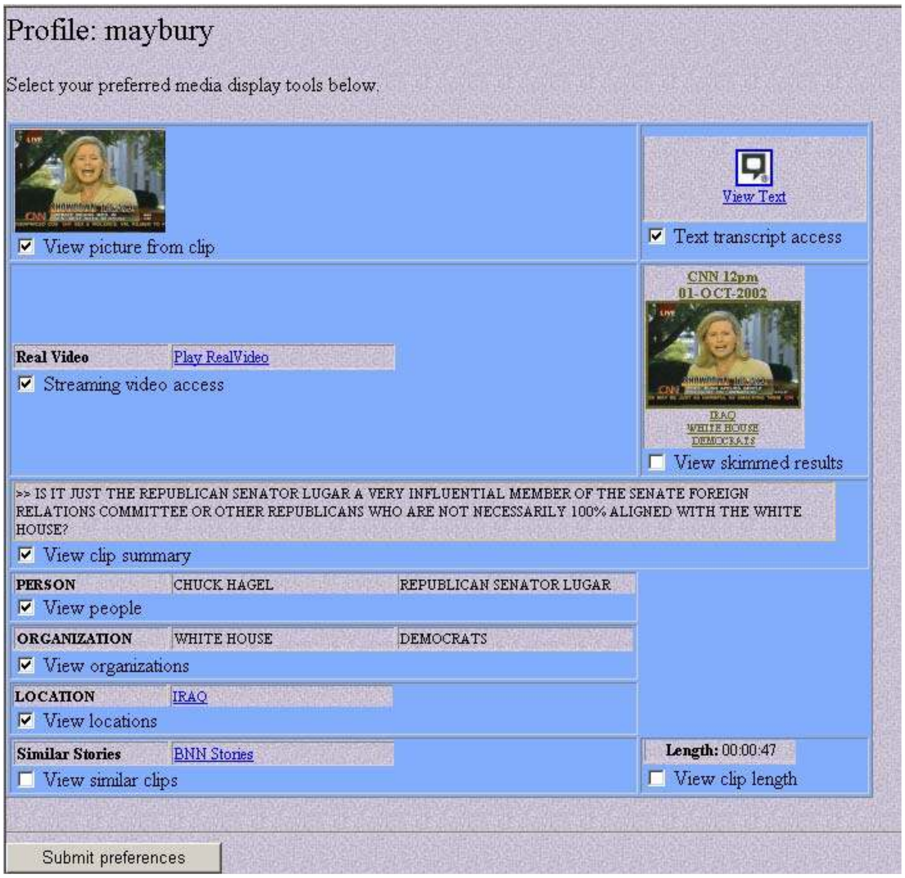
● From Date 01 AUG 2001
To Date 03 OCT 2001

○ All Dates

Text Source

☑ Closed Caption

Figure 7. Search Screen including Explicitly Stated User Profile



Profile: maybury

Select your preferred media display tools below.

☑ View picture from clip

View Text
☑ Text transcript access

Real Video	Play RealVideo

☑ Streaming video access

CNN 12pm
01-OCT-2002
IRAQ
WHITE HOUSE
DEMOCRATS
□ View skimmed results

>> IS IT JUST THE REPUBLICAN SENATOR LUGAR A VERY INFLUENTIAL MEMBER OF THE SENATE FOREIGN RELATIONS COMMITTEE OR OTHER REPUBLICANS WHO ARE NOT NECESSARILY 100% ALIGNED WITH THE WHITE HOUSE?

☑ View clip summary

PERSON	CHUCK HAGEL	REPUBLICAN SENATOR LUGAR

☑ View people

ORGANIZATION	WHITE HOUSE	DEMOCRATS

☑ View organizations

LOCATION	IRAQ

☑ View locations

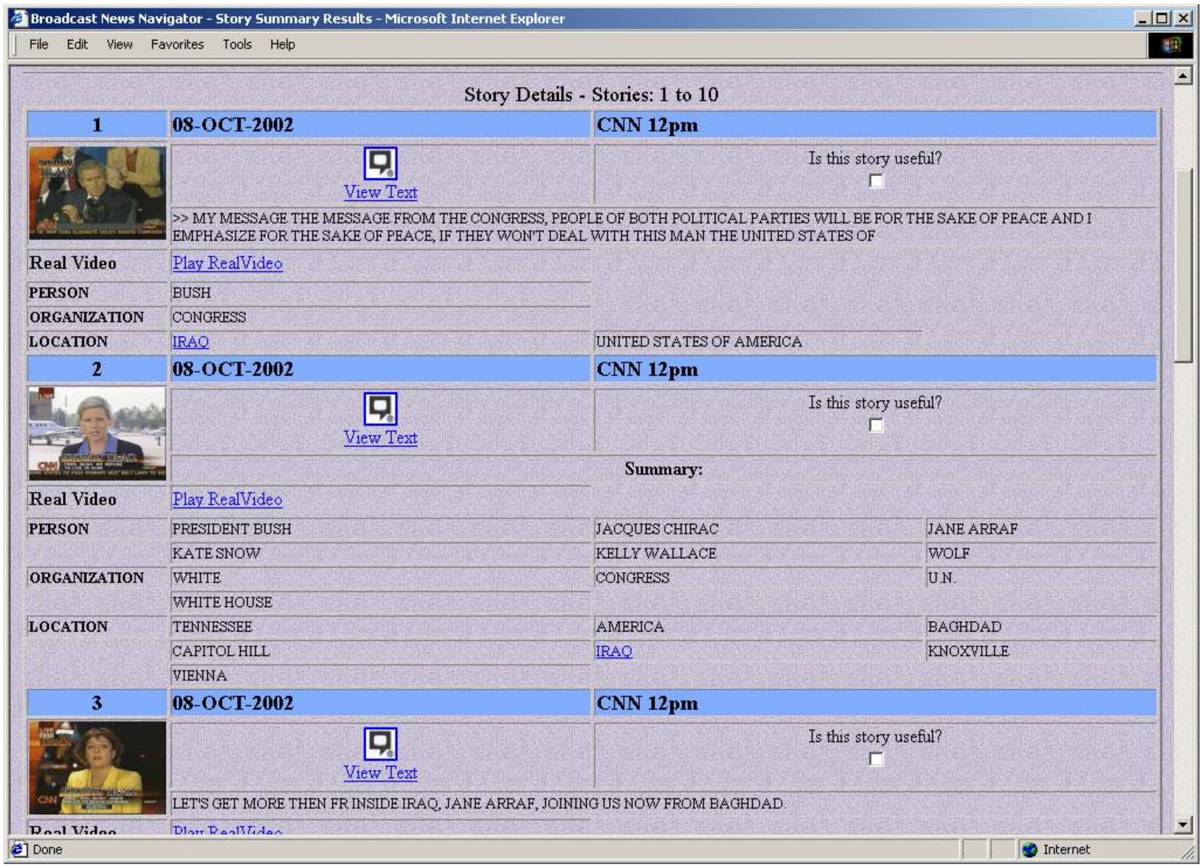
Similar Stories	BNN Stories

□ View similar clips

Length: 00:00:47
□ View clip length

Figure 8. User Media Presentation Profile

Figure 8 shows the user manipulating the system's user model (called a profile) of their media presentation preferences. Note that the user can select what types of media (e.g., key frame picture from a clip, text transcript, video), media properties (e.g., clip length), and/or content (e.g., types of named entities, related stories) to display when viewing story details. In Figure 8, the user has selected all media elements except for similar stories, clip length, and skimmed results. All of these elements are automatically extracted from source stories by BNN.


Broadcast News Navigator - Story Summary Results - Microsoft Internet Explorer
File
Edit
View
Favorites
Tools
Help
Done
Internet

Story Details - Stories: 1 to 10

1	**08-OCT-2002**	**CNN 12pm**	
	View Text	Is this story useful? □	
	>> MY MESSAGE THE MESSAGE FROM THE CONGRESS, PEOPLE OF BOTH POLITICAL PARTIES WILL BE FOR THE SAKE OF PEACE AND I EMPHASIZE FOR THE SAKE OF PEACE, IF THEY WON'T DEAL WITH THIS MAN THE UNITED STATES OF		
Real Video	Play RealVideo		
PERSON	BUSH		
ORGANIZATION	CONGRESS		
LOCATION	IRAQ	UNITED STATES OF AMERICA	
2	**08-OCT-2002**	**CNN 12pm**	
	View Text	Is this story useful? □	
	Summary:		
Real Video	Play RealVideo		
PERSON	PRESIDENT BUSH	JACQUES CHIRAC	JANE ARRAF
	KATE SNOW	KELLY WALLACE	WOLF
ORGANIZATION	WHITE	CONGRESS	U.N.
	WHITE HOUSE		
LOCATION	TENNESSEE	AMERICA	BAGHDAD
	CAPITOL HILL	IRAQ	KNOXVILLE
	VIENNA		
3	**08-OCT-2002**	**CNN 12pm**	
	View Text	Is this story useful? □	
	LET'S GET MORE THEN FR INSIDE IRAQ, JANE ARRAF, JOINING US NOW FROM BAGHDAD.		
Real Video	Play RealVideo		

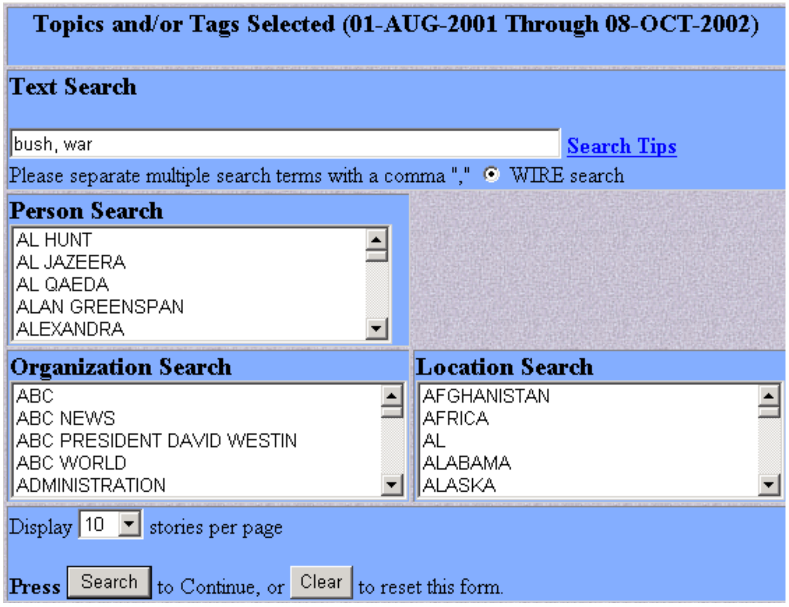
Figure 9. Multiple Media Presentation

Figure 9 illustrates the effect of the media preference profile on the display of stories in the news on October 8, 2002. Based upon preferences selected in Figure 8, only two and a half stories can be displayed on the screen. In contrast, if the user had established a profile stating a preference for only one-line summaries, about twice as many stories could be displayed.

Note however that the automatically generated one-line summary fails completely in story 2. The user would have no description of this story in a summary-only display.

6. DISCOVERING USER INFORMATION NEEDS: QUERY REFINEMENT IN BNN

We can also determine a user's interests not only by what they indicate interest in explicitly (e.g., their search keywords and/or named entities), but also by reasoning about terms and/or entities related to their stated interests. For example, if a user searches for stories about the location "Iraq", we might look into the story set returned by BNN and notice that the person "Saddam Hussein" occurs frequently. Or if the person searches for stories in which the name "Saddam Hussein" and location "Iraq" appear, she might frequently find the terms "weapons of mass destruction" or "UN inspections".



Topics and/or Tags Selected (01-AUG-2001 Through 08-OCT-2002)

Text Search

bush, war

Please separate multiple search terms with a comma "," ● WIRE search

Person Search

AL HUNT
AL JAZEERA
AL QAEDA
ALAN GREENSPAN
ALEXANDRA

Organization Search

ABC
ABC NEWS
ABC PRESIDENT DAVID WESTIN
ABC WORLD
ADMINISTRATION

Location Search

AFGHANISTAN
AFRICA
AL
ALABAMA
ALASKA

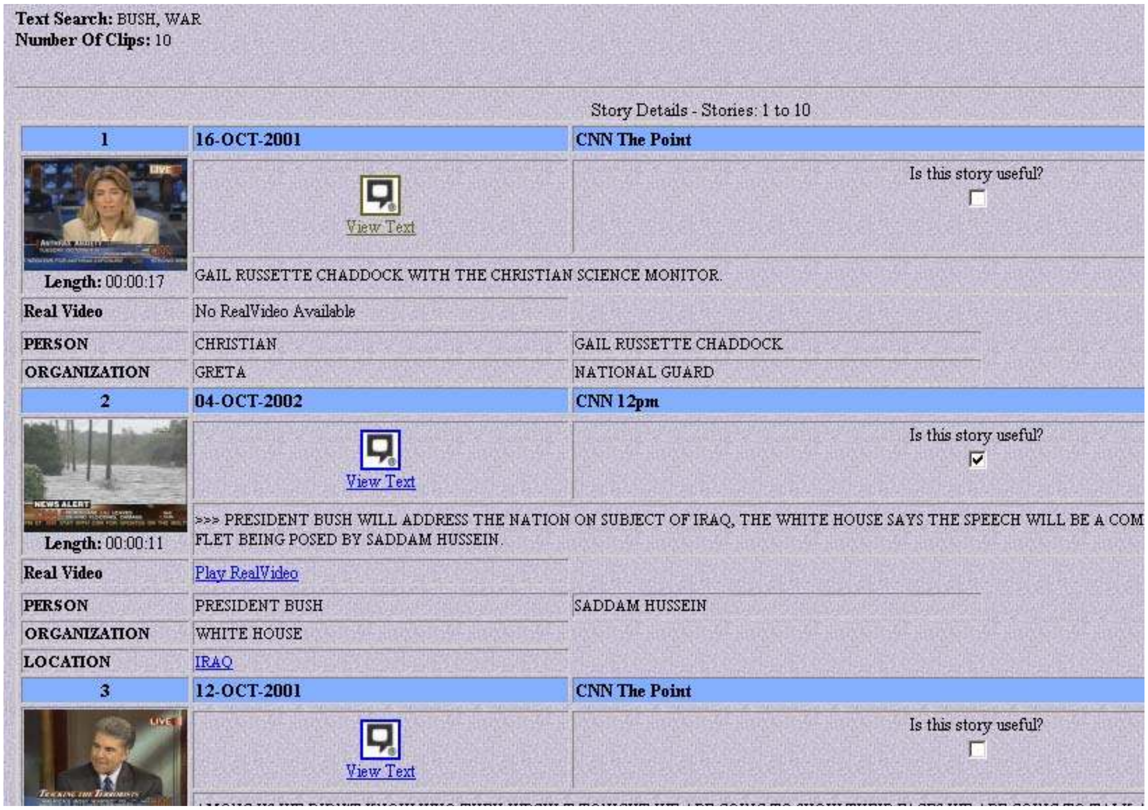
Display 10 stories per page

Press Search to Continue, or Clear to reset this form.

Figure 10a. BNN Search

As in previous versions of BNN, the user initially chooses the set of sources upon which to query, a date range, along with the option to perform either a profile search (saved in a previous session) or a custom search (such as in Figure 7). The choices presented to the user in a custom search include options for searching any named entity category (person, organization, location, etc.) as well as a free-form text search.

Having selected sources, time range, and the type of search, the user is presented the detailed search tool (See Figure 10a). For each category selected in the previous screen, a selectable menu of actual named entities appears with the ability to select one or more elements. Since the text option was selected, a free form text box appears above the selectable menus in Figure 10a. In this example, the user types in the terms "bush" and "war". As show in Figure 10b, the retrieval engine returns a set of stories that are about "bush" and "war".



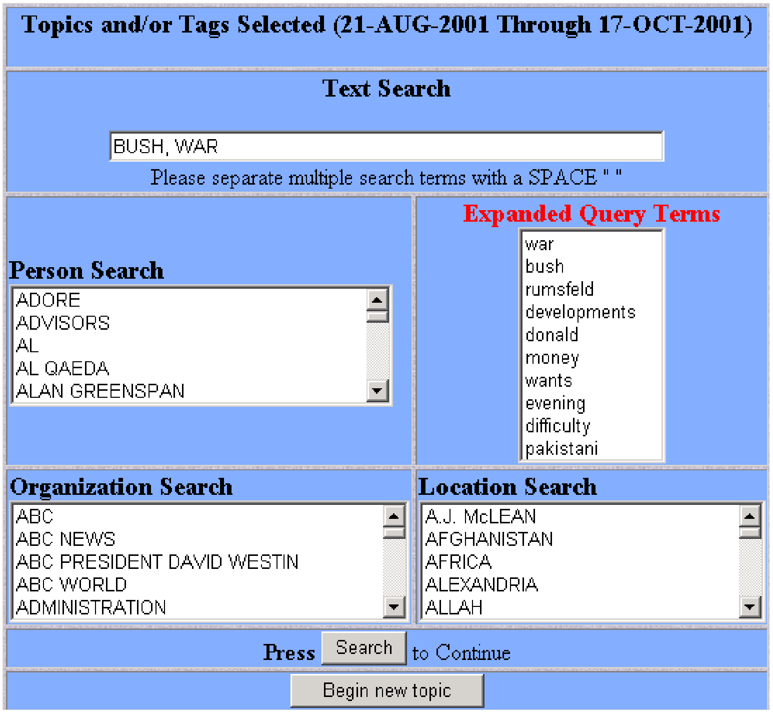
Text Search: BUSH, WAR
Number Of Clips: 10

Story Details - Stories: 1 to 10

1	16-OCT-2001	CNN The Point
	View Text	Is this story useful? □
Length: 00:00:17	GAIL RUSSETTE CHADDOCK WITH THE CHRISTIAN SCIENCE MONITOR.	
Real Video	No RealVideo Available	
PERSON	CHRISTIAN	GAIL RUSSETTE CHADDOCK
ORGANIZATION	GRETA	NATIONAL GUARD
2	04-OCT-2002	CNN 12pm
	View Text	Is this story useful? ☑
Length: 00:00:11	>>> PRESIDENT BUSH WILL ADDRESS THE NATION ON SUBJECT OF IRAQ, THE WHITE HOUSE SAYS THE SPEECH WILL BE A COM FLET BEING POSED BY SADDAM HUSSEIN.	
Real Video	Play RealVideo	
PERSON	PRESIDENT BUSH	SADDAM HUSSEIN
ORGANIZATION	WHITE HOUSE	
LOCATION	IRAQ	
3	12-OCT-2001	CNN The Point
	View Text	Is this story useful? □

Figure 10b. Relevance Selections

The user then selects stories that she finds most relevant to her information need, in this case the second story which is about the threat posed by Saddam Hussein in Iraq. The system uses LCA to expand the user's query terms. In particular, the terms "bush" and "war" are expanded into a list which is displayed in Figure 11a. It includes person names such as “donald” and “rumsfeld”, terms such as “developments” and “money”, adjectival locations such as “Pakistani”, and so on, as described above.



Topics and/or Tags Selected (21-AUG-2001 Through 17-OCT-2001)

Text Search

BUSH, WAR

Please separate multiple search terms with a SPACE " "

Expanded Query Terms

war
bush
rumsfeld
developments
donald
money
wants
evening
difficulty
pakistani

Person Search

ADORE
ADVISORS
AL
AL QAEDA
ALAN GREENSPAN

Organization Search

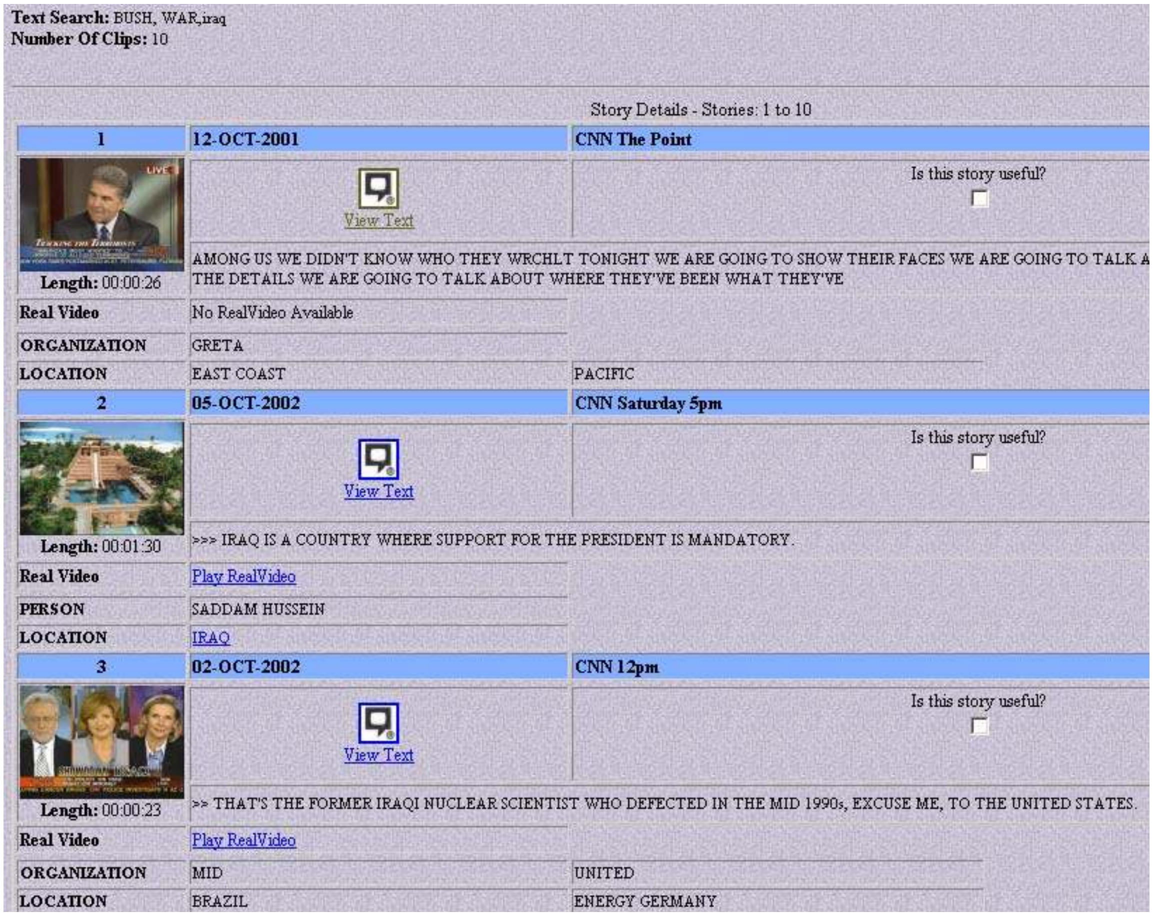
ABC
ABC NEWS
ABC PRESIDENT DAVID WESTIN
ABC WORLD
ADMINISTRATION

Location Search

A.J. McLEAN
AFGHANISTAN
AFRICA
ALEXANDRIA
ALLAH

Figure 11a. Results using Expanded Query

In the example in Figure 11a, the user selects the term “iraq” from the location menu and then reruns the query, expanded now to include the terms “bush”, “war”, and “iraq”. Figure 11b shows the resulting stories retrieved by the expanded query. Notice in Figures 10b and 11b that at this stage the user can select via a check box those stories from the returned story list that she deems most relevant to her information needs. The most frequently occurring terms in these selected stories will be added to further refine the user’s query. At this point the user can further refine their search or simply execute it. The user need not select query expansion terms nor provide relevance feedback, or she can do either or both.



Text Search: BUSH, WAR,iraq
Number Of Clips: 10

Story Details - Stories: 1 to 10

1	**12-OCT-2001**	**CNN The Point**
Length: 00:00:26	View Text	Is this story useful? □
	AMONG US WE DIDN'T KNOW WHO THEY WRCHLT TONIGHT WE ARE GOING TO SHOW THEIR FACES WE ARE GOING TO TALK A THE DETAILS WE ARE GOING TO TALK ABOUT WHERE THEY'VE BEEN WHAT THEY'VE	
Real Video	No RealVideo Available	
ORGANIZATION	GRETA	
LOCATION	EAST COAST	PACIFIC
2	**05-OCT-2002**	**CNN Saturday 5pm**
Length: 00:01:30	View Text	Is this story useful? □
	>>> IRAQ IS A COUNTRY WHERE SUPPORT FOR THE PRESIDENT IS MANDATORY.	
Real Video	Play RealVideo	
PERSON	SADDAM HUSSEIN	
LOCATION	IRAQ	
3	**02-OCT-2002**	**CNN 12pm**
Length: 00:00:23	View Text	Is this story useful? □
	>> THAT'S THE FORMER IRAQI NUCLEAR SCIENTIST WHO DEFECTED IN THE MID 1990s, EXCUSE ME, TO THE UNITED STATES.	
Real Video	Play RealVideo	
ORGANIZATION	MID	UNITED
LOCATION	BRAZIL	ENERGY GERMANY

Figure 11b. Selected Stories

7. PRELIMINARY EVALUATION

Evaluation of user adaptive interfaces is more challenging than typical human-computer interface evaluation, for several reasons. First, the user can influence interface behavior because models of the user change system output and/or behavior. Second, the system's model of the user can influence the behavior of the user (e.g., if it is poor or uncooperative, users can become frustrated; if it is critical or challenging it can inspire new user inferences). Third, there often is diversity in the task being performed, its complexity, and/or the overall environment. This high degree of variability raises the uncertainty and complexity in the operation of the systems and in their adaptation. This, in turn, makes evaluation challenging.

Because of this complexity, we have evaluated the performance of BNN both in terms of content (type and amount) and the form of delivery. We have found that presenting less information to the users (e.g., story skims or summaries versus full text or video) enables more rapid relevance assessment and story comprehension (Light and Maybury 2002). For example, using 20 users performing relevance assessments and information extraction tasks, we demonstrated that users exhibit over 90% precision and recall using displays such as those in Figure 11b in less than half the time required to search digital video sequentially.

	PERFORMANCE (Document Precision)		
	General	**Specific**	
TOPIC	*Query Precision (query term in parenthesis)*	*Query plus Document Relevance Feedback (# selected docs)*	*Query plus Query Expansion Feedback (selected terms in parenthesis)*
1: Iraqi foreign minister	20/20 = 100% (“iraq”)	1/20 = 5% (1)[2]	100% (“tariq”) 100% (“aziz”)
2: Weapons of mass destruction	18/20 = 90% (“weapons”)	19/20 = 95% (18)	100% (“nuclear”) 100% (“inspections”)
3: Chief weapons inspector	8/17 = 47% (“inspector”)	10/20 = 50% (8)	100% (“blix”)
4: Israeli Palestinian conflict	5/19 = 31.5% (“israeli”)	8/20 = 40% (8)	95% (“gaza”) 95% (“hamas”)
5: Washington D.C. sniper	16/20 = 80% (“sniper”)	20/20 = 100% (17)	90% (“shooting”)
AVERAGE	***69.7%***	***58% (10.4)***	***97%***

Table 1. Preliminary Performance Evaluation

Because personalization increases the refinement and focus of a user query, this should translate directly into task performance enhancements. To test this hypothesis, we ran a series of evaluations. We initially tested P-BNN on a collection of 600 news stories (culled out of tens of thousands of stories from several years) primarily in October 2002 from multiple program sources such as CNN Headline News, CNN NewsNight with Aaron Brown,

[2] This is the performance of relevance feedback on only 1 document (#9), so it is very low.

and CNN Moneyline. Based on user queries and user feedback, we returned up to twenty relevant news stories (which we interchangeably call documents) which we then had the user assess for relevance. Table 1 contains illustrative performance evaluations from queries on this collection.

The first column lists the topics investigated in the corpus. The second column reports the precision of documents returned by a single-term user query representing a general information need. The third and fourth columns represent the precision of a more specific or focused search which is enhanced by document relevance feedback and automated query expansion.

Query precision is the percentage of documents out of the top twenty returned (or fewer, if less than 20 documents are returned) that the user finds relevant to their information needs. For example, for the first topic where the user is searching for documents about Iraq's foreign minister, whose name they have forgotten, they first search broadly for "Iraq" and find that all top 20 documents returned are about "Iraq". Thus, the precision on the general term "Iraq" is 20/20 or 100%. However, their information need is unsatisfied as only one story out of 20 (#9) is about "Tariq Aziz", Iraq's foreign minister. Accordingly, the user continues and selects story #9 as relevant and provides this feedback to the search engine. As was detailed in Section 4, LCA extracts a weighted set of the most frequent terms from this document (in this case the terms "iraq", "matter", "telling", "reiterated", and so on) which P-BNN then uses to invoke another search against the entire story collection, and returns another set of documents that match these weighted terms. Since the user provides only a single document for relevance feedback and the words "Tariq" and "Aziz" appear near the end of a twenty term expansion list, the precision performance of this feedback is only 5% (third column of Table 1). That is, only 1/20 or 5% of the documents returned after this feedback are about Iraq's foreign minister. Note that depending upon which documents the user selects and the terms contained therein, document relevance feedback can either refine or broaden the search. In all five queries shown document relevance feedback improves precision.

After indicating relevant documents, the user can also ask the system to suggest, based on real-time analysis of these documents, specific terms to expand their query. As shown in the fourth column of Table 1, when the user selects either the terms “Tariq” or “Aziz” from the term expansion list, the system returns exactly five documents that pertain to the user’s original information need, thus achieving a precision of 5/5 or 100%. At this point the LCA models the user’s information needs more precisely than a set of weighted terms.

The second query in Table 1 concerns weapons of mass destruction. The user types the simple query “weapons”. This retrieves 18/20 = 90% relevant documents about weapons of mass destruction. When ten documents are noted as relevant by the user, a real-time analysis of the most frequent keywords in these documents is performed and is used to retrieve documents, 95% of which are relevant to the user’s information needs.

In the third query example, the user is searching for the lead U.N. weapons inspector. The user starts with a broad search (“inspector”). However, this yields a low 47% relevant documents. Providing user relevance feedback raises the performance slightly, to at least 50%. In fact, relevancy assessment was clear on half of these documents but ambiguous on others. Therefore, we assumed the most conservative interpretation and only counted half as relevant. However, when the user runs LCA query expansion and reviews the list which includes the rank ordered terms “powell”, “hans”, “secretary”, “weapons”, “the”, “resolution”, “inspectors”, and “blix”, the user notes that Hans Blix is the chief weapons inspector and selects the term “Blix” to find 20 documents which are 100% relevant to their needs.

In the fourth query example, the user is searching for stories about the Israeli-Palestinian conflict. When they type in a general query such as “israeli”, they obtain a low yield of relevant stories. Providing feedback about relevant documents raises the performance by adding such expansion terms as “palestinian”, “gaza”, “civilians”, “hamas”, “factions”, “militants”, “sharon”, and “raid”. When the user selects specific concrete terms such “gaza” or “hamas”, precision rises to 95%.

In the fifth topic area, the user is interested in stories about the sniper attacks in Washington, D.C. Using the term "sniper", 16 of 20 documents retrieved were about the D.C. sniper (two stories were irrelevant and two others were errors in story segmentation). When the user selects those documents and requests similar ones, the precision rises to 100%. When the user asks for term suggestions based on their relevance assessments, the system indicates that the terms "sniper, the, police, shooting, maryland, …" and so on are the most typical of the document set. If they select the term "shooting", the precision of the returned document set is 18/20 or 90% (two irrelevant documents are returned about a shooting of a marine and a U.N. protester shooting).

As can be seen from the examples in column two in Table 1, the relation of a keyword and the document collection can dramatically influence performance. A specific term like "iraq" that has many stories in the collection can yield high precision, although users often need to discover these in the search process. Providing document level relevance feedback (shown in column three) improves precision in four out of five cases. However, if only one relevant document is provided (as in query #1), this method performs poorly because of limited evidence to infer the user's information needs. Selection of specific expansion terms by the user (column four) yields a more specific model of their information needs and results in higher precision in all five queries which allows the system to retrieve a more relevant set of stories to their interests.

8. DETAILED USER EVALUATION

Motivated by the promise of query refinement for capturing a more accurate specification of user information need, we performed a detailed study to analyze the following variables:

- *Recall*. We were not only interested in the precision of retrieval (i.e., the ability to only retrieve relevant documents), but also the ability to retrieve all of the relevant documents.
- *Scale.* We need to ensure that the promising performance results that we obtained will be sustained in larger collections, in particular in thousands of stories from several months to several years worth of news.

- *Quality*. We need to understand the effects of combining relevance feedback and term selection, to allow users to combine forms of query and document feedback to more accurately specify their needs.
- *Query Characterization and Display*. An effective means for characterizing the effects of various relevance or refinement selections on the weighted term model of the user's information needs is necessary, so the user has a clear characterization of what they are asking for.
- *Speed*. Searches together with refinements take much less than a minute to perform. Nevertheless, we are currently designing user studies to establish the tradeoff between the time necessary to perform query refinement and document relevance feedback, and increases in precision and recall as a result of finer models of user information needs which reduce time required in post retrieval analysis.
- *User Satisfaction*. We are interested whether users believe the system to be more enjoyable to use, and whether they perceive it to improve their performance with respect to accuracy, timeliness, and comprehensiveness.
- *Cognitive Load*. While difficult to measure, we are interested in whether query expansion eases or increases the load on the user's attention and reasoning resources. Indirect measurements of these might include time for manual term generation versus term selection from expansion menus, the number of iterations to converge on a query, and so on.

8.1 Evaluation Corpus and Topic Development

We created an evaluation corpus consisting of the closed-captioned text of nine news broadcasts airing between August 21, 2001 and October 17, 2001. The news broadcasts were automatically segmented into stories by BNN, resulting in 502 stories. It is important to note that while to our knowledge this is the highest performing story segmentation system, it remains inaccurate (Boykin and Merlino 1999). A baseline version of the system over a range of broadcast sources (e.g., CNN, MS-NBC, and ABC) performed segmentation on average with 38% precision and 42% recall across all multimodal cues (i.e., textual, audio, and visual cues). In contrast, performance for the best combination of multimodal cues rose to 53% precision and 78% recall. When visual anchor booth recognition cues are specialized to a specific source (e.g., ITN broadcasts that have more regular visual story change indicators), the performance rises to 96% precision and recall. In the current system we were dealing with accuracy in the 50% to 80% range.

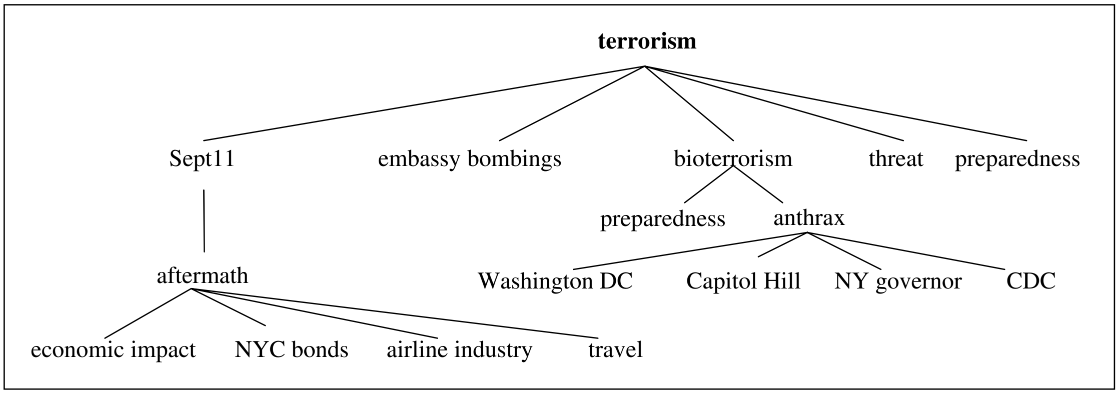


Figure 12. Sample Topic and Subtopic Categories in Evaluation Corpus

Each story can contain zero or more topics – zero for those stories that contained no text, or too little text to decipher. To annotate the corpus for topics, an initial pass was made by one annotator who indicated what each story was about, using no pre-defined topic typology. A senior annotator then reviewed the set of topic labels that emerged, developed a clean typology of 26 topics with subtopics where needed, and made a second pass on the corpus to apply the modified topic labels and to also provide final judgment on the story topics themselves. Figure 12 illustrates one of the resulting top-level topics, Terrorism, and some of the sub-topics in this particular corpus.

The second annotator also evaluated each story in isolation and flagged every topic within a story where automatic story segmentation created a section too brief or too removed from context to reasonably understand what the story was about. This resulted in 121 topics being labeled as “fragments.” In scoring against the stories marked by the subjects in the user study, these fragments were all considered non-relevant.

For the user evaluation we developed 10 topic areas: bioterrorism, U.S. space program, accidental injuries, gambling, investing, Mideast conflict, music, weather, violent crime, and sports. Each experiment topic was manually mapped to the topic annotations in the evaluation corpus to create a gold standard for measuring user performance. For example, the "investing" experiment topic mapped to the topic-subtopic annotations "economy, stock market" and "economy, federal reserve rate," as well as the topic annotation "investing." Each topic area was presented in one or more sentences to give the subjects an idea of the stories that were relevant to the topic, for example[3]:

Bioterrorism: We are interested in any story or story fragment related to bioterrorist events, preparation, threats, or prevention. To be relevant, the biological threat must be initially spread by terrorists, and not by natural processes.

8.2 Experiment Design

We created a fully instrumented version of BNN to allow detailed comparison of time stamped logs of events in both a baseline system that does not contain query refinement (Configuration A) and one augmented with LCA for query refinement (Configuration B). Since we had previously empirically demonstrated the value of personalizing broadcast news layout (Merlino and Maybury 1999), our intent was to more extensively and deeply explore the bounds of performance of personalcasting content via query refinement.

At the beginning of the study, four subjects were given an overview of the purpose and design of the experiment, and a demonstration of the experimental task using both system configurations. Subjects were then given personal computers and an opportunity to use the system themselves with several practice topics. The subjects were allowed to ask questions during this training period, but not during the experiment proper. The total training time was approximately one hour. For each of the ten experimental topics, subjects were then asked to find as many relevant stories as possible. For each topic they were given five minutes. They were instructed to work at a normal pace and to try as many different queries as they wished; they were not required to continue searching for the full five minutes if they had no more

[3] A list of all 10 topics can be found in the Appendix.

query ideas. Two randomly chosen subjects used Configuration A for the first five topics, while the other two used Configuration B, and then all switched. While the 10 topics were randomly ordered, all subjects processed them in the same order. In this way, the conditions under which a given topic was processed were kept as constant as possible. After completing the 10 experiment topics, users were asked to fill out a user satisfaction questionnaire, discussed below.

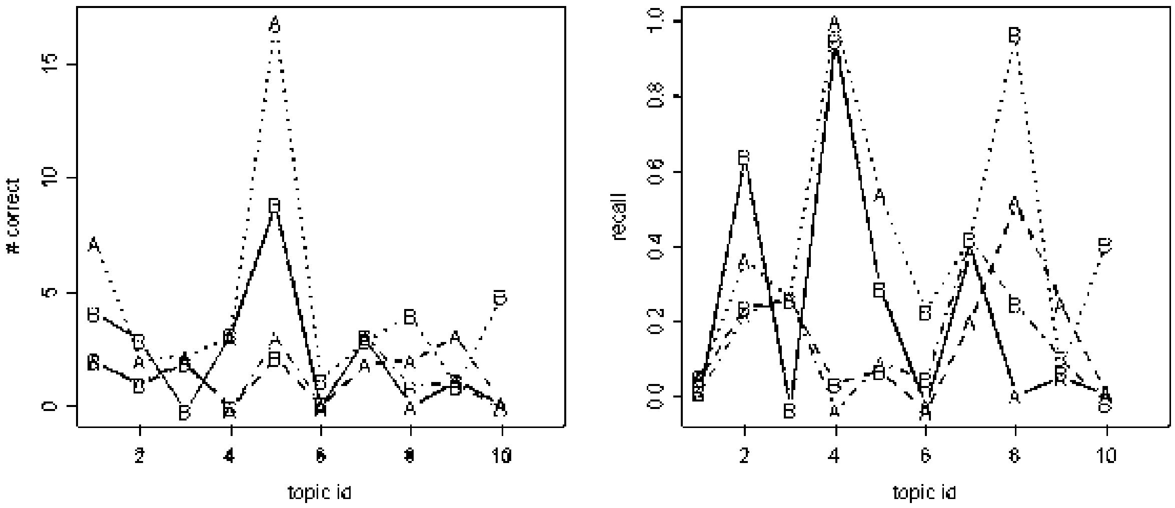


Figure 13. Performance Across Topics and Subjects According to: #-correct & recall

8. 3 Results: Comparative Performance

We based our comparison of the effectiveness of the two system configurations on two metrics: #-correct and recall. The #-correct measure is simply a count of the number of relevant stories found by a given user for a given topic. It does not take into account the number of stories that were considered relevant to that topic. The recall measure, in contrast, is the fraction of relevant stories that the subject was able to find; that is, #-correct divided by the total number of stories in the collection relevant to the topic. We chose not to measure precision, the fraction of stories found by the user that were relevant to the topic. Precision would measure the agreement of the user's assessment of relevance with the judgment as given by the gold standard, which would measure characteristics of the subjects rather than characteristics of the configurations used.

Figure 13 shows how each of the subjects performed on each of the topics. The graph on the left shows performance as measured by #-correct, and the graph on the right, performance as measured by recall. The topics are presented in the order of presentation. The scores for a given user are connected, with a different line style for each user. Each point is labeled with an A or B, indicative of the configuration that was used by the associated subject for processing the corresponding topic. For the purpose of visualization, a small amount of random noise has been added to each score in order to make it easier to distinguish overlapping points and lines.

Inspection of the graphs suggests that there may be a difference in ability among subjects. It also suggests that there may be an intrinsic difference in difficulty of some topics as compared with others, although which topics might be considered more difficult, and which considered relatively easier to resolve, differs according to the metric used. Overall, there is substantial variance in the scores under both metrics, and there does not seem to be any indication that one system configuration dominates the other with respect to either of the two measures studied. Indeed, a statistical analysis of the two configurations showed no significant difference between systems.

However, an analysis of variance showed that, for both metrics, there is a clear effect due to differences in users. Also, for both metrics, the variance attributable to both the user and the topic is far greater than the variance that can be attributed to the different configurations.

8. 4 Results: Comparative Satisfaction

An anonymous survey was administered to the subjects asking for their assessments on a Likert scale of enjoyment, ease of retrieval, trust in the results, completeness (ability to find all relevant stories), utility, and speed. On average, using System B subjects reported they enjoyed the system 12.5% more, trusted the results 7.7% more, and believed 8.3% more strongly that the results they found were more complete, on average a 9.5% perceived improvement. When asked to explicitly compare System A to System B in terms of ease of use, reliability, and speed, users indicated either no difference or a preference for System B. However, these results are only suggestive as experiments with larger sample sizes are needed to obtain statistically significant findings.

9. FUTURE RESEARCH

Many outstanding research problems must be solved to realize automatically created user-tailored news. Important problem areas include:

1. *Automatic logging and inference of user interests.* With users increasingly learning, working and playing in digital environments, monitoring user interactions (e.g., Linton et al. 1999) is feasible and has shown value. In information seeking sessions, detecting selections and rejections of information provides an opportunity to induce individual and group profiles that can assist in content selection and presentation generation. For example, each of the user actions shown in Figure 3 (e.g., query, story selection, media selection) affords an opportunity for modeling user interest in the first two actions and/or preference in the last. In addition to explicit user interest collection, an implicit method could build an interest model by watching the user session to track the user's query, selection of particular stories, and choice of media. The system could then automatically construct a content interest and media preference model.
2. *Tailoring.* More sophisticated mechanisms are required to tailor content to specific topics or users. In addition to content selection, material must be ordered and customized to individual user interests. This will require methods of presentation generation that integrate extracted or canned text with generated text.
3. *Information Extraction.* Over the longer term we are working to create techniques to automatically summarize, fuse and tailor selected events and stories. This requires deeper understanding of the source news material beyond extracting named entities, key frames, or key sentences.
4. *Multilingual content.* Because news is global in production and dissemination, it is important to support access to and integration of foreign language content. This poses not

only multilingual processing challenges, but also requires dealing with different country/cultural structures and formats.

5. *Cross story fusion.* An important problem is not only the summarization of individual stories, but also summarizing across many stories, possibly from difference sources or languages. This is particularly challenging when the sources are possibly inconsistent in content or form. This ultimately requires cross story multimodal presentation generation.
6. *Persistence/transience of interest profiles.* Users' information needs tend to change over time, with profiles rapidly becoming out of date. Monitoring user queries and story selections over time is one method that can address this problem. Generalizing from their specific interests can yield an even richer user model.
7. *Evaluation.* Community defined multimedia evaluations will be essential for progress. Key to this progress will be a shared infrastructure of benchmark tasks with training and test sets to support cross-site performance comparisons.

10. CONCLUSION

We have designed, implemented, demonstrated and evaluated the Personalized Broadcast News Navigator (P-BNN) that provides tailored content and presentation of broadcast video news. We combine automated video understanding and extraction together with user modeling to provide individualized personalcasts at the story level from weeks of network news. Our system supports explicit user content and media preference profiles, it implicitly reasons about terms co-occurring with user query terms, and it accepts and modifies its model of the user's information need based on user feedback on the relevance of provided content. Accordingly, the system overcomes the fixed organization of news programs produced for stereotypical audiences by segmenting, selecting, and reordering content based on user preferences and feedback. Moreover, it represents an advance beyond program-level electronic program guides that are beginning to find their way into the commercial marketplace by not relying upon any externally provided program metadata and by providing more fine-grained content tailoring at the story rather than program level. Accordingly, we believe this kind of interactive, fine-grained, content-based personalization will be fundamental to television and news understanding systems of the future.

ACKNOWLEDGEMENTS

We would like to thank Alfred Kobsa and the anonymous reviewers for their detailed and helpful feedback on earlier versions of this article.

REFERENCES

Ardissono, L., Portis, F. and Torasso, P. 2001. Architecture of a System for the Generation of Personalization Electronic Programming Guides. Eighth International Conference on User Modeling: Workshop on Personalization in Future TV, Sonthofen, Germany. www.di.unito.it/~liliana/UM01/ardissono-etal.pdf.

Attar, R. and Fraenkel, A. 1977. Local Feedback in Full-Text Retrieval Systems. *Journal of the Association of Computation Machinery*, 24(3): 397-417.

Bove, V. M. 1983. Personalcasting: Interactive Local Augmentation of Television Programming. Master's thesis, MIT, 1983.

Boykin, S. and Merlino, A. 1999. Improving Broadcast News Segmentation Processing. IEEE International Conference on Multimedia and Computing Systems. Florence, Italy. 7-11 June 1999.

Boykin, S. and Merlino, A. 2000. Machine Learning of Event Segmentation for News on Demand. *Communications of the ACM*, 43(2): 35-41.

Boyle, C. and Encarnacion, A. O. 1994. An Adaptive Hypertext Reading System. *User Modeling and User-Adapted Interaction*, 4(1): 1-19.

Brusilovsky, P. 1996. Methods and Techniques of Adaptive Hypermedia. *User Modeling and User-Adapted Interaction,* 6(2-3): 87-129.

Brusilovsky, P. 2001. Adaptive Hypermedia. *User Modeling and User-Adapted Interaction,* 11: 87-110.

Croft, W.B. and Harper, D.J. 1979. Using Probabilistic Models of Document Retrieval Without Relevance Information. *Journal of Documentation*, 35(4): 285-295.

Hu, Q. 2003. Audio Hot Spotting. MITRE Sponsored Research Project. http://www.mitre.org/news/events/tech03/briefings/intelligent_information/hu.pdf

Kaplan, C., Fenwick, J. and Chen. J. 1993. Adaptive Hypertext Navigation based on User Goals and Context. *User Modeling and User Adapted Interaction,* 3(3): 193-220.

Koenemann, J. 1996. Supporting Interactive Information Retrieval Through Relevance Feedback. CHI 96 Doctoral Consortium. http://www.acm.org/sigchi/chi96/proceedings/doctoral/Koenemann/Jk2_txt1.htm.

Koenemann, J. and Belkin, N. 1996. A Case For Interaction: A Study of Interactive Information Retrieval Behavior and Effectiveness. In *Proceedings of the SIGCHI Conference on Human Factors and Computing Systems*. Vancouver, British Columbia, Canada. ACM Press: NY. Pages: 205 – 212. http://www.acm.org/sigchi/chi96/proceedings/papers/Koenemann/jk1_txt.htm

Light, M. and Maybury, M. 2002. Personalized Multimedia Information Access: Ask Questions, Get Personalized Answers. *Communications of the ACM* 45(5): 54-59. (www.acm.org/cacm/0502/0502toc.html). In Brusilovsky, P. and Maybury, M. (eds). Special Section on The Adaptive Web.

Linton, F., Joy, D., and Schaefer, H-P. 1999. Building User and Expert Models by Long-Term Observation of Application Usage. In J. Kay (Ed.), UM99: User Modeling: Proceedings of the Seventh International Conference (pp. 129-138). New York: Springer Verlag. [Selected data are accessible from an archive on http://zeus.gmd.de/ml4um/]

Maybury, M. Feb. 2000. News on Demand: Introduction. *Communications of the ACM*, 43(2): 32-34.

Maybury, M., Merlino, A., and Morey, D. 1997. Broadcast News Navigation using Story Segments, ACM International Multimedia Conference, Seattle, WA, November 8-14, 381-391.

Merlino, A. and Maybury, M. 1999. An Empirical Study of the Optimal Presentation of Multimedia Summaries of Broadcast News. Mani, I. and Maybury, M. (eds.) *Automated Text Summarization,* MIT Press.

Merlino, A. 2002. ViTAP News on Demand. Human Language and Technology Conference, San Diego, CA, March 25, 2002.

Robertson, S.E. and Walker, S. 1994. Some Simple Effective Approximations to the 2-Poisson Model for Probabilistic Weighted Retrieval. *Proceedings of the 17th Annual ACM-SIGIR Conference on Research and Development in Information Retrieval,* 232-241. Reprinted in: K. Sparck Jones and P. Willett (eds) 1997, *Readings in Information Retrieval*. Morgan Kaufmann, 345-354.

Salton, G. and Buckley, C. 1990. Improving Retrieval Performance by Relevance Feedback. *Journal of the American Society for Information Science (JASIS),* 41(4): 288-297.

Xu, J. and Croft, W.B. 1996. Query Expansion Using Local and Global Document Analysis. *Proceedings of the 19th Annual ACM-SIGIR Conference on Research and Development in Information Retrieval*, 4-11.

Xu, J. and Croft, W.B. 2000. Improving the Effectiveness of Information Retrieval with Local Context Analysis. *ACM Transactions on Information Systems,* 18(1): 79–112.

APPENDIX: USER EVALUATION TOPICS

1. Bioterrorism: We are interested in any story or story fragment related to bioterrorist events, preparation, threats, or prevention. To be relevant, the biological threat must be initially spread by terrorists, and not by natural processes.

2. U.S. Space Program: We are interested in any story or story fragment related to events and activities associated with U.S. space programs.

3. Accidental Injuries: We are interested in any reports of injuries to people as a result of accidents. Injuries as a result of intentional harmful acts such as crime and terrorism are *not* relevant.

4. Gambling: We are interested in any story or story fragment that reports on gambling, i.e., betting on an uncertain outcome or playing a game for financial gain. Both legal and illegal gambling are relevant.

5. Investing: We are interested in any story or story fragment related to financial investing, such as stock and interest rate reports. Advertisements about financial investing are _not_ relevant.

6. Mideast Conflict: We are interested in any story or story fragment that relates to the conflict in the Middle East and efforts to resolve it. To be relevant, the story must center around issues between Middle Eastern countries and/or territories, as opposed to U.S.-Mideast relations.

7. Music: We are interested in any story or story fragment about music, including musical compositions, musicians, bands, and concert events.

8. Weather: We are interested in any story or story fragment that reports on or forecasts weather events and phenomena.

9. Violent Crime: We are interested in any story or story fragment about violent criminals and/or criminal actions. Stories about terrorists and terrorism are _not_ relevant.

10. Sports: We are interested in any story or story fragment that reports on sporting events or athletes.