



News on Demand

Imagine waking in the morning, turning on your interactive TV,

and being greeted by your intelligent news agent. Your agent has spent the night digitizing, segmenting, indexing, extracting, translating, and summarizing the world's news broadcasts. The agent, knowing your interests from your previous implicit and explicit interactions, selects the most relevant sources and topics and dynamically generates an interactive newscast personalized to your preferred format, length, language, and interests.

Fantasy or reality?

Today you are faced with a barrage of international, national, and local news sources available in a heterogeneous collection of print, radio, television, and Web formats. Recently, News on Demand (NOD) systems have

 **Mark Maybury**
Guest editor

emerged that integrate computing results from the speech, language, and image processing communities to create a new class of multimedia information access systems. Scientifically, these systems are interesting as corpus-based and knowledge-based text, speech, and image processing are converging to support cross-media understanding, resulting in many new research areas such as multi-stream segmentation, topic detection and tracking, and multimedia summarization. Commercially, this technology and related systems enables several new application opportu-

nities in news delivery, distributed education, and intelligence/surveillance. While we have only begun to explore these systems, already several start-ups and innovative corporations are beginning to sell novel systems and services (ISLP.COM, FASTV.COM, NewsEdge, broadcast.com) and national broadcasters such as CNN, MSNBC, and Canadian Broadcasting Corp. (CBC) are beginning to explore personalized news services.

As we begin the new millennium, this special section aims to bring together some of the best work to date in the creation of intelligent NOD systems that automatically process news and provide *personalcasts*, tailored presentations to satisfy individualized user needs. Figure 1 illustrates a generic NOD demand processing flow common to these efforts. In this pipeline, multimedia data streams (text, audio, and video) are captured (digitized and stored), processed (segmented, transcribed,

≈ *This technology and related systems enables several new application opportunities in news delivery, distributed education, and intelligence/surveillance.*

translated, indexed, correlated/clustered, fused, extracted, summarized), and presented to the end user with the possibility of selecting, organizing, coordinating, tailoring, and visualizing the content. Each of these stages of processing presents challenges and new opportunities for research. As is evident in the area of news understanding, NOD efforts are necessarily multidisciplinary, typically requiring cross-organizational and cross-site teams.

We begin the section with an article by Merlino and Boykin, from MITRE, who report how they use hidden Markov models to learn broadcast news structure from annotated broadcast corpora. Speech, language, and image processing are used to extract and then generate tailored news via their Broadcast News Editor (BNE). The news is made searchable via their Broadcast News Navigator (BNN).

The next article in the section describes SRI International's MAESTRO (Multimedia Annotation and Enhancement via a Synergy of Technologies and Reviewing Operators), a system that acts as a conductor for speech, image, and OCR technologies to support synergistic indexing, archiving, and retrieval of multimedia content.

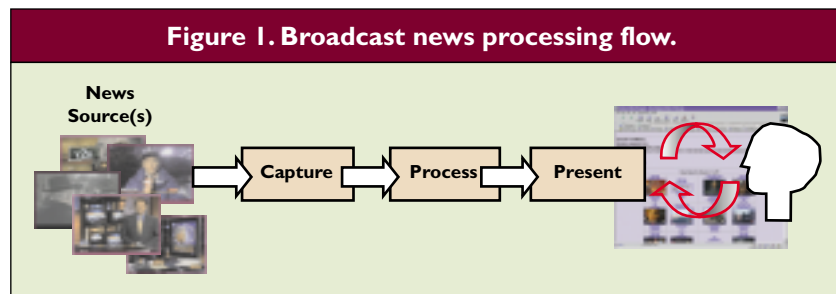
Kubala et al.'s article "Integrated Technologies for Indexing Spoken Language" describes their system Rough'n'Ready, which creates a rough transcription of speech that is ready for browsing. The system extracts features such as the names of people, places, and organizations mentioned in the transcript as well as the identities and locations of the speakers in the recording. It breaks the continuous stream of words into passages that are thematically coherent and automatically summarized with a short list of appropriate topic labels.

The article by Wactlar et al. reports on CMU's Informedia Digital Video Library system, which extracts information from audio and video and supports full content searching over digitized video sources. Two unique features of the Informedia system are highlighted: named face (automatically asso-

ciating a name with a face) and location analysis (displaying and querying news stories from a map).

In "Transcribing Broadcast News for Audio and Video Indexing," authors from LIMSI report on a North American broadcast news transcription system that performs with a 13.6% word error rate and discuss current work in broadcast transcription of German and French broadcasts.

The sidebar by Sadaoki Furui et al. describes joint



research between the Tokyo Institute of Technology and NHK broadcasting in transcription and topic extraction from Japanese broadcast news. They describe efforts to improve processing by modeling filled pauses, performing online incremental speaker adaptation and by using a context dependent language model that models readings of words (which include Chinese characters—Kanji—and two kinds of Japanese characters—Hira-gana and Kata-kana).

The final contribution by Pallett et al. considers how we evaluate such novel systems. The authors report on community evaluation efforts to measure progress in both multimedia information retrieval and Topic Detection and Tracking (TDT) applications via broadcast news corpora made available through the Linguistic Data Consortium.

Enjoy the articles in the section and explore what your news network of tomorrow may be like. ■

MARK MAYBURY (maybury@mitre.org) is Executive Director of the Information Technology Division at the MITRE Corporation in Bedford, MA.

© 2000 ACM 0002-0782/00/0200 \$5.00