The Informedia Digital Video Library Project is developing new technologies for creating full-content search and retrieval digital video libraries. Working in collaboration with WQED Pittsburgh, the project is creating a test-bed that will enable K–12 students to access, explore, and retrieve science and mathematics materials from the digital video library. The library will initially contain 1,000 hours of video from the archives of project partners: WQED, Fairfax Co. VA Schools’ Electronic Field Trips, and the British Open University’s BBC-produced video courses. (Industrial partners include Digital Equipment Corp., Bell Atlantic, Intel Corp., and Microsoft, Inc.) This library will be installed at Winchester Thurston School, an independent K–12 school in Pittsburgh.

One of the most interesting research aspects of the project is the development of automatic, intelligent mechanisms to populate the library through integrated speech, image, and language understanding. The Informedia digital video library system uses Sphinx-II to transcribe narratives and dialogues automatically. Sphinx-II is a large-vocabulary, speaker-independent, continuous speech recognizer developed at Carnegie Mellon. With recent advances in acoustic and language modeling, it has achieved a 90% success rate on standardized tests for a 20,000-word, general dictation task. By relaxing time constraints and allowing transcripts to

Figure 1. Combining speech, image, and natural language to create a full-context, searchable library
be generated off-line, Sphinx-II will be adapted to handle the video library domain’s larger vocabulary and diverse audio sources without severely degrading recognition rates. In addition to annotating the video library with text transcripts, the videos will be segmented into smaller subsets for faster access and retrieval of relevant information. Some segmentation is possible via the time-based transcript generated from the audio information.

Segmenting video clips via visual content is also being performed based on work at CMU’s Image Understanding Systems Laboratory. Rather than manually reviewing a file frame-by-frame around an index entry point, machine vision methods that interpret image sequences are used to automatically locate beginning and end points for a scene or conversation. This segmentation process can be improved through the use of contextual information supplied by the transcript and language understanding.

Finding desired items in a large information base poses a major challenge. The Informedia Project goes beyond simply searching the transcript text and is, in addition, applying natural-language understanding for knowledge-based search and retrieval. One strategy employs computational linguistic techniques from the Center for Machine Translation for indexing, browsing, and retrieving based on identification of noun phrases in text. Along with improving query capabilities, the Informedia Project is researching better ways to present information from a given video library. Once users identify video objects of interest they will need to be able to manipulate, organize, and effectively reuse the video. To aid the user, the system will use cinematic knowledge to enhance the composition and reuse of materials from the video library.

The Informedia Project’s first version drew on a small (three gigabyte) database of several hundred digital video objects, text, graphics, and audio material drawn from WQEDs Space Age series, distinguished lectures in computer science, and software engineering training lectures. Early user feedback has shown the benefits of automatic indexing and segmentation, illustrating the accurate search and selective retrieval of audio and video materials appropriate to users’ needs and desires. The system demonstrates the practicality of combining speech, language, and image understanding technologies to create entertaining, educational experiences.

© ACM 0002-0782/95/0400

Figure 2. A prototype system display