EURASIP Journal on Applied Signal Processing 2003:2, 91–92

Editorial

Jing Huang
IBM T. J. Watson Research Center, 1101 Kitchawan Road, Yorktown Heights, NY 10598, USA
Email: jghg@us.ibm.com

Mukund Padmanabhan
Renaissance Technologies Corporation, 600 Route 25A, East Setauket, NY 11733, USA
Email: mukund@rentec.com

Savitha Srinivasan
IBM Almaden Research Center, San Jose, CA 95120, USA
Email: savitha@almaden.ibm.com

The recent proliferation of the worldwide web and the low cost of storage have contributed to an explosively growing volume of information. Traditionally, in order to be usable, information needs to be in some form of structured format, such as records in relational databases, XML tagged data types, and so forth. The field of structured-information management deals with techniques to create, store, query, and mine these data types. A fundamental characteristic of accessing such a database is that a data query returns an absolute list of matches in the database.

However, the vast majority of data created and stored today does not exist in structured format. For instance, a recent analytic study reports that only about 20 percent of all corporate content exists in structured formats such as transactional data or product specifications. The rest of the data exists in unstructured, machine-generated formats such as data from medical sensors, security cameras, audio recordings of meetings, broadcasts, traffic video, and so forth. There is often very valuable information buried in such unstructured data (e.g., call-center data may contain information about customer trends); however, the information is not directly accessible, because of its unstructured nature. Although it is possible to convert such data sources to structured forms by manual processing, the high cost associated with this enables only a very small portion of the data to be processed in this fashion. Consequently, there is a great deal of research and commercial value in developing methods both to manage this data and to automatically analyze and extract semantics present in it.

The ease of managing such unstructured data depends on its complexity. One way to characterize complexity is to examine its multimedia properties such as visual, spatial, and temporal components, the ease of data entry, and the existence of well-defined semantic units by which the data can be indexed and searched. Measuring the complexity of unstructured data types along these properties leads to an increasing order of complexity from text and image to audio and video.

For text data types, the basic approach used in information management is to first "extract a sequence of features" from the data; subsequently, the data is "indexed" by the features or the features are compared to templates stored in a library, and the data is "indexed" by a list of templates. A data query of this processed unstructured data would then compute the "similarity" between the query and the indexed data, and return a "ranked list of potential matches" (as opposed to an absolute list of matches as in the case of a query on structured data). Such methods have evolved to some level of maturity in the case of text data types, and in order to capitalize on this, most current methods of dealing with multimedia data first attempt to convert the data into text format and then use text-based techniques to manage it.

We could hence think of an unstructured-information management system as having three phases. In the initial phase of converting multimedia sources into text, research in speech recognition (conversion of speech to text) plays a pivotal role in the processing of unstructured speech data, and research in video processing and content analysis play a pivotal role in the processing of image and video data. As signal processing plays a fundamental role in speech and video processing, we could think of the problem of extracting information from unstructured multimedia sources as an extended application of signal processing. In the second phase of information management, research in feature extraction, indexing, similarity matching, and ranking plays a pivotal role. The third and final phase relates to integrating querying, browsing, and the search paradigm of the complete system. The development of efficient multimedia navigation, summarization, and browsing tools is an important part of this last phase.

This special issue focuses on unstructured-information management across several different unstructured data types. The first paper deals with unstructured text data. In the remaining papers, we transit into other unstructured data types beginning with audio, move on to image, and conclude with video. Each section starts with an overview paper, which attempts to give a high-level picture of the various building blocks used in the solution. This is followed with papers that provide further details about specific building blocks. The section is then concluded with a paper that describes an example of a complete solution or a real application.

The first paper is about a novel feature selection method with applications in managing text data. The next four papers deal with audio as the raw data format (e.g., broadcast news, call-center conversations). The section starts with an overview paper by James Allan that gives a high-level view of the components of a system that starts with audio data as a source and extracts information from it. Subsequently, the papers by Wolfang Macherey et al. and Chiori Hori et al. delve into the theoretical aspects of the system. Finally, the paper by Jean-Luc Gauvain and Lori Lamel describe a system that employs all these methods to successfully process radio-broadcast news. Switching gear from temporal data (audio) to temporal-spatial data (image), the paper by Jing Huang et al. presents a scheme for hierarchical classification of images via supervised learning. The last five papers deal with images and video as the raw data format. The section starts with a paper by Yihong Gong on audio-video summarization that generates a video summary by alignment of the visual summary with the audio summary. The next paper by W. H. Adams et al. that explores semantic indexing of multimedia content building upon well-known techniques for audio, video, and text retrieval and focuses on the use of Bayesian networks for the fusion of different classifiers. The next paper by Thijs Westerveld et al. investigates the effect of language models both in text retrieval and for visual features such as shots and scenes. This is followed by a video classification and retrieval paper that takes advantage of motion patterns. The last paper in this section, by Arnon Amir et al., discusses the practical aspects of a multimedia retrieval system and emphasizes the role of browsing in multimedia retrieval systems.

It is hoped that these papers would give the readers an introduction to the vast field of unstructured-information management and its potential benefits and applications, and also acquaint them with the state-of-the-art in extracting information from various formats of unstructured multimedia data.

Jing Huang
Mukund Padmanabhan
Savitha Srinivasan

Jing Huang is a research staff member at IBM T. J. Watson Research Center. She received the B.S. and M.S. degrees in applied mathematics from Tsinghua University, Beijing, China, and the Ph.D. in computer science from Cornell University. Her Ph.D. work focused on computer vision and content-based image retrieval. After joining IBM T. J. Watson Research Center, she switched to work on automatic speech recognition. Her research interest also includes machine learning and information extraction.

Mukund Padmanabhan received the B.Tech degree in electronics and electrical communication engineering from the Indian Institute of Technology, Kharagpur, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Los Angeles. His interests span a large number of areas, including communications, signal processing, analog integrated circuits, speech recognition, information extraction, and, most recently, statistical financial modeling. He worked in the area of speech recognition at the IBM T. J. Watson Research Center, Yorktown Heights, NY, from 1992 to 2001, where he managed the Telephony Speech Recognition Group. Currently he works for Renaissance Technologies Corp. in the area of financial modeling. He is on the editorial board of the EURASIP Journal on Applied Signal Processing, and also a member of the IEEE SPS Speech Technical Committee. Dr. Padmanabhan was a recipient of the Best Paper Award for a paper in the IEEE Transactions on Speech and Audio Processing in 2001. He is also a coauthor of a book on signal processing and circuits entitled *Feedback-Based Orthogonal Digital Filters: Theory, Applications, and Implementation*.

Savitha Srinivasan manages Multimedia Content Distribution activities at IBM Almaden Research Center. Her group is responsible for multimedia information retrieval and content protection technologies. They are the founding members of copy protection technology currently deployed for DVD audio/video and have been top performers at the recent NIST-sponsored video retrieval task. Her research interests include video segmentation and semantic video retrieval with a focus on the application of speech recognition technologies to multimedia. She has published several papers on speech programming models and multimedia information retrieval. She is on the Scientific Advisory Board of a leading National Science Foundation (NSF) multimedia school and Area Editor of Multimedia in leading journals. She holds three patents related to the use of spelling in speech applications and the combination of speech recognition and audio analysis for information retrieval. Her current expertise extends into pragmatic aspects of multimedia such as digital rights management.