Multimedia Information Extraction: A Roadmap

Mark Maybury
MITRE
Information Technology Division
maybury@mitre.org

Abstract

Significant advances have been made in information extraction from unstructured multilingual text (see trec.nist.gov). However, increasingly large collections of audio (e.g., iTunes), imagery (e.g., Flickr), and video (e.g., YouTube) together with rapid and widespread growth and innovation in new information services (e.g., blogging, podcasting, media editing, social networking) is driving the need not only for multimedia retrieval but also for information extraction from and across heterogeneous media. A group of researchers and practitioners from the language processing, speech processing, image/video processing and spatial/temporal reasoning communities created the roadmap shown below together with associated grand challenges during the 2008 AAAI Fall Symposium on Multimedia Information Extraction. The roadmap consists of three lanes that predict developments in this field over the next five years together with resultant capabilities in knowledge aiding, presentation design, learning, and performance enhancement. The first lane, data, resources, and tools, shows how sources to be processed are transforming from structured single media (e.g., text, images) to large scale unstructured data (e.g., multiple media, social media). Representations are advancing from named entities and relations to events, emotions, intentions, and social structures requiring an underlying backplane of a multimedia ontology. The second lane, evaluation, illustrates a focus on cross media entity mention/disambiguation and co-reference to multimedia usage scenarios in the near term and perceptual/cognitive load in the midterm to the detection of style and behavior that occurs across multiple media and processing of large scale data (2-3 TB/day). The third lane, components and applications, progresses from media segmentation, feature extraction, and search in the short term to human activity recognition, media integration, and sentiment analysis in the medium term, to bias detection, cross media summarization, and user awareness in the long term. One roadblock to progress is the need for understanding complex events. Collectively these capabilities will support the advancement of multimedia question answering, moving from today's single, factual, isolated questions about single documents to tomorrow's multi-document, scenario-based, contextual question answering from unstructured, multilingual, and multimodal documents.

Biography

Dr. Mark Maybury is Executive Director of MITRE's Information Technology Division (itc.mitre.org). Dr. Maybury is a member of the Board of OMG and the Air Force Scientific Advisory Board. Mark is editor or co-author of nine books including Intelligent Multimedia Interfaces, Intelligent Multimedia Information Retrieval, Question Answering, Intelligent User Interfaces, Advances in Text Summarization, Advances in Knowledge Management, Personalized Digital Television, and Information Storage and Retrieval. Mark has been awarded patents for broadcast news understanding and personalized news navigation. Mark received a B.A. from Holy Cross, an M.B.A. from RPI, and an M. Phil. and Ph.D. in Artificial Intelligence from Cambridge University, England.