

John R. Smith
IBM Research

A Great Challenge

Every machine loves a good challenge. At least that's what we must think given the number of competitions we've created for them. Although the nature of the classic man-machine contests has changed since the seminal challenge of human pile driver versus steam-powered drill in creating the US railroads in the 1800s, speed, strength, and cunning are still what matters most.

In multimedia, computers are essential for almost every task. The huge volumes of multimedia data and relentless audio and video streams require computers to work quickly. The complexity of analytics and other multimedia processing tasks requires deep and capable algorithms. Intelligence or something like it is needed to effectively work at the level of multimedia content recognition and understanding to meet the expectations of humans.

Benchmarks have long been critical in making scientific progress in the multimedia field. The National Institute of Standards and Technology's Text Retrieval Conference Video Retrieval Evaluation (Trecvid) has been important for advancing research in content-based retrieval of digital video since 2001 (see <http://trecvid.nist.gov/>). The initial tasks were basic (recognizing tens of categories) and data sets were small (tens of hours of video).¹ But in 2001, they were a significant challenge for computers. Since then, Trecvid has pressed hard on all of the dimensions, requiring systems to be faster, more capable, and more accurate.² The community has responded in a large way with more than 50 organizations, research labs, and universities typically participating each year.

Similarly, The Cross-Language Evaluation Forum' Cross Language Image Retrieval Track (ImageCLEF) has become important for advancing research on content-based image retrieval (see <http://www.imageclef.org/>). ImageCLEF started in 2003 with tasks based on automatic and interactive image

retrieval of 50 queries over a collection of 30,000 images,¹ which is clearly small by today's standards. Today, a single eager mobile-phone user might take this many photos in a year. But, the tasks have also grown significantly since then, and today they include medical image retrieval, photo annotation, plant identification, and patent image retrieval and classification.

Other challenges have been more fun in nature. The VideOlympics was unique in being the only refereed multimedia evaluation to give out Golden Retriever awards for a number of categories, such as best performer, most impressive interface, public favorite, and most useful (see <http://www.videolympics.org/>). VideOlympics worked as a companion to Trecvid, emphasizing live demonstration of multimedia retrieval systems.³ More recently, ACM Multimedia has been running a series of grand challenges associated with its annual conference. These challenges address significant problems for the multimedia industry over a 2 to 5 year horizon (see <http://www.acmmm10.org/program/competitions/multimedia-grand-challenge/>). The specific challenges include problems such as photo location and orientation detection, video genre classification, multimedia content adaptation, video segmentation, image understanding, and photo set theme identification.⁴ In 2010, 18 systems were submitted for the Multimedia Grand Challenge with prizes going to three winners chosen by industry scientists, engineers, and business luminaries.

Taking on humans is still a favorite for machines. The recent trouncing of the world's top Jeopardy winners by IBM's Watson computer system has reinvigorated the epic man-machine struggle and given the upper hand to computers (see [http://en.wikipedia.org/wiki/Watson_\(artificial_intelligence_software\)](http://en.wikipedia.org/wiki/Watson_(artificial_intelligence_software))). The challenge was a significant one, given

the long-standing difficulties with natural-language understanding and open-domain question-answering. But, it was not rich from a multimedia perspective. Images and video contain a wealth of information and could have been part of the contest given their potential for answering questions and providing insights about the human experience and the real world.⁵

It's this integration of all modalities of multimedia information—images, video, audio, text, and languages—that still needs to come together in a Jeopardy-like challenge. The various benchmarks and evaluations are essential for driving advancement in the multimedia field. Given the good progress that is being made, we now need a great challenge that will mark our progress in the multimedia field.

Everybody loves a great challenge. **MM**

References

1. P. Over et al., "Multimedia Retrieval Benchmarks," *IEEE MultiMedia*, vol. 11, No. 2, 2004.
2. M. Naphade et al., "Large-Scale Concept Ontology for Multimedia," *IEEE MultiMedia*, vol. 13, no. 3, 2006, pp. 86-91.
3. C.G.M. Snoek et al., "VideOlympics: Real-Time Evaluation of Multimedia Retrieval Systems," *IEEE MultiMedia*, vol. 15, no. 1, 2008, pp. 86-91.
4. C.G.M. Snoek and M. Slaney, "Academia Meets Industry at the Multimedia Grand Challenge," *IEEE MultiMedia*, vol. 18, no. 1, 2011.
5. J.R. Smith, "The Not So Newlywed Game," *IEEE MultiMedia*, vol. 17, no. 3, 2010.

Contact John R. Smith at jsmith@us.ibm.com.



Selected CS articles and columns are also available for free at <http://ComputingNow.computer.org>.

Software Experts Summit

The Computer History Museum, Mountain View, CA

Managing the Pace of Innovation

Come to the Summit & learn from the best!



Grady Booch, IBM



Jan Bosch, Intuit



Gary McGraw, Cigital



Grigori Melnik, Microsoft



Presented by
Software
IEEE
computer
society

17 May 2011
Register by 15 April and save!

WWW.COMPUTER.ORG/SES11