

The Benchmarking Initiative for Multimedia Evaluation: MediaEval 2016

The Benchmarking Initiative for Multimedia Evaluation (MediaEval) organizes an annual cycle of scientific evaluation tasks in the area of multimedia access and retrieval. The tasks offer scientific challenges to researchers working in diverse areas of multimedia technology—including text, speech, and music processing, as well as visual analysis. MediaEval tasks focus on the social and human aspects of multimedia. In this way, the benchmark supports efforts of the research community to tackle challenges linked to less widely studied user needs. It also helps researchers investigate the diversity of perspectives that naturally arise when users interact with multimedia content.

Once MediaEval task participants have developed algorithms and written up their results, they come together at the yearly workshop to report on and discuss their outcomes, synthesize their findings, and plan new evaluation tasks and strategies to help advance the state of the art. Here, we present highlights from the 2016 workshop.

A Multimedia Venue

The 2016 MediaEval workshop was held in the Netherlands, right after ACM Multimedia 2016, in Amsterdam. The workshop was hosted by The Netherlands Institute for Sound and Vision, located in Hilversum, about 30 km from Amsterdam (www.beeldengeluid.nl/en). The mission of the institute is to collect and preserve the audiovisual cultural heritage of the Netherlands and to make it open and available for as many users as possible. The institute's archive contains nearly a million hours of television, radio, music, and film, from 1898 through today. The archive is housed in a depot that is five stories deep: an amazing architec-

tural feat for the Netherlands with its proximity to sea level.

Even to the casual visitor, the architecture of the Netherlands Institute for Sound and Vision is striking. The façade is covered with colored glass panels that let through light. These panels are based on nearly 800 different images stored in the archive. They combine relief images with bright colors and were created through a specially designed process (see Figure 1a).

The MediaEval 2016 Workshop was attended by 65 participants from 23 countries (see Figure 1b). The participants submitted algorithms that addressed nine tasks. (See <http://ceur-ws.org/Vol-1739> for the full proceedings.) The workshop consisted of oral and poster presentations. (See <http://multimediaeval.org/video> for videos of the task overview talks.)

A highlight of the workshop was the invited talks. Evangelos Kanoulas, from the University of Amsterdam, Netherlands, talked about information retrieval evaluation and “putting the user back in the loop.” Julián Urbano, from the Universitat Pompeu Fabra, Barcelona, discussed music information retrieval evaluation.¹ Also, Maria Eskevich, from Radboud University Nijmegen, Netherlands, and Robin Aly, from the University of Twente, Netherlands, presented the “TRECVID Video Hyperlinking” task.

MediaEval Shared Tasks

Each year, the MediaEval campaign offers a number of carefully selected and designed evaluation tasks. The tasks are proposed and also coordinated by autonomous groups of task organizers. Each task comprises a definition, a dataset, and an evaluation procedure.

Martha Larson
*Radboud University
Nijmegen*

Mohammad Soleymani
University of Geneva

Guillaume Gravier
IRISA

Bogdan Ionescu
*University
Politehnica of
Bucharest*

Gareth J.F. Jones
Dublin City University



(a)



(b)

Figure 1. The MediaEval 2016 Workshop. (a) Held at the Netherlands Institute for Sound and Vision (Source: Hay Kranen; Creative Commons Attribution 4.0 License Netherlands (CC-BY), <https://commons.wikimedia.org/w/index.php?curid=35124219>), the workshop had (b) 65 participants from 23 countries.

The task-selection process is driven by a large community survey. For a task to be selected, it must match MediaEval's objective of focusing on the human and social side of multimedia, be well designed and feasible, and demonstrate evidence of community interest, as measured by the survey.

Social Multimedia

A first set of tasks asked participants to address challenges raised by social multimedia. The Placing task required participants to estimate the location of multimedia items using user-contributed metadata or content. The task used a subset of the Yahoo Flickr Creative Commons 100 Million (YFCC10M) dataset.²

The Retrieving Diverse Social Images task required participants to refine Flickr query photo results to improve relevance and diversity. The task provided visual, text, and user-tagging-credibility-related features in order to address a broader research community.

Finally, the Verifying Multimedia Use task tackled the problem of images and videos that are shared online (Twitter) but fail to provide a faithful representation of a real-world event.

User Impact

A second set of tasks tackled the challenge of making predictions related to the impact of multimedia on viewers or users. The Emotional Impact of Movies task required participants to predict emotions evoked in viewers. The data consisted of Creative Commons-licensed movies (professional and amateur) together with

valence-arousal ratings collected from annotators.

The Predicting Media Interestingness task was new for 2016. This task required participants to automatically select frames or portions of movies that are the most interesting for a common viewer. To solve the task, participants could use the provided visual, audio, and text content.

Finally, the Context of Multimedia Experience task looked at the suitability of video content for consumption in particular viewing situations. Participants were asked to use multimodal features to predict videos that viewers would rate as suitable for watching on an airplane.

Getting More out of Multimedia

A third set of tasks was devoted to tackling unsolved problems that push beyond the ways in which multimedia is currently used or multimedia collections are accessed. The Person Discovery in Broadcast TV task asked participants to tag TV broadcasts with the names of people who can be seen as well as heard. The task differs from similar tasks in that the list of possible people is not known a priori and their names must be discovered in an unsupervised way.

The Zero-Cost Speech Recognition task reposed the question of how to train an automatic speech recognition system. Instead of assuming that the data is a given, finding data was considered to be part of the challenge. The cost of data and the error rate of the system were minimized together.



(a)



(b)

Figure 2. The Netherlands Institute for Sound and Vision made it possible to (a) demonstrate music queries on a Steinway during the C@MERATA presentation and (b) show videos from past years during discussion and lunch.

Finally, C@merata (the Querying Musical Scores with English Noun Phrases task) addressed the problem of information retrieval from a collection of musical scores. Effective solutions require combining knowledge of music, natural language processing, and information retrieval (see Figure 2a).

MediaEval is a grassroots initiative that runs without a parent organization. The success of MediaEval is possible due to the hard work of the task organizers and the dedication of the task participants. Hours of intense discussions are needed at the workshop (see Figure 2b) and throughout the year.

In 2017, the MediaEval Workshop will be co-located with Conference and Labs of the Evaluation Forum (CLEF 2017), to be held in September at Trinity College in Dublin, Ireland (<http://clef2017.clef-initiative.eu>). Registration to participate in MediaEval 2017 will run from approximately March to May 2017 (see <http://multimediaeval.org>).

The tasks offered by MediaEval 2017 will likely be a mix of classic and innovative tasks. We anticipate new developments in the areas of music, medical multimedia, and also in the combination of social multimedia and satellite images,³ medical multimedia,⁴ and music.¹ **MM**

Acknowledgments

Support for the MediaEval 2016 workshop was provided by Elias Evaluating Information

Access Systems (<http://elias-network.eu>), friends of ACM SIGIR (<http://sigir.org>), Technicolor (www.technicolor.com), Simula Research Laboratory (<http://www.simula.no>) and the Netherlands Institute for Sound and Vision (www.beeldengeluid.nl). MediaEval connects to the speech processing community by means of the SIG Speech and Language in Multimedia of the International Speech Communication Association (<http://slim-sig.irisa.fr>). We would also like to thank the 2016 task organizers and participants. A special thank you goes to Andrew Demetriou for the effort he invested in making the 2016 workshop a success and to his dedication to supporting multimedia research in crossing disciplinary boundaries.

References

1. B. McFee, E. Humphrey, and J. Urbano, "A Plan for Sustainable MIR Evaluation," *Proc. 17th Int'l Society for Music Information Retrieval Conf. (ISMIR)*, 2016, pp. 285–291.
2. B. Thomee et al., "YFCC100M: The New Data in Multimedia Research," *Comm. ACM*, vol. 59, no. 2, 2016, pp. 64–73.
3. A. Woodley et al., "Introducing the Sky and the Social Eye," *Working Notes Proceedings of the MediaEval 2016 Workshop*, 2016; CEUR-WS.org/Vol-1739/MediaEval_2016_paper_9.pdf.
4. M. Riegler et al., "Multimedia and Medicine: Teammates for Better Disease Detection and Survival," *Proc. 2016 ACM Multimedia Conf. (MM)*, 2016, pp. 968–977.

Martha Larson is a professor at Radboud University Nijmegen in the area of Multimedia Information Technology and also works at Delft University of Technology. Contact her at m.a.larson@tudelft.nl.

Mohammad Soleymani is a Fellow at the Swiss Center for Affective Sciences and Computer Vision and Multimedia Laboratory at the University of Geneva, Switzerland. He is a founding member of MediaEval. Contact him at mohammad.soleymani@unige.ch.

Guillaume Gravier is a senior researcher at IRISA, the Research Institute of Computer Science and Random Systems in France, where he heads the LinkMedia research team. Guillaume served as lead editor of the MediaEval workshop proceedings in 2016. Contact him at guig@irisa.fr.

Bogdan Ionescu is a professor at the University Politehnica of Bucharest, Romania. He serves on the

MediaEval Community Council. Contact him at bionescu@alpha.imag.pub.ro.

Gareth J.F. Jones is a professor at Dublin City University in Ireland. He co-founded VideoCLEF together with Martha Larson in 2008, which became MediaEval in 2010. He is serving as conference chair for CLEF 2017 in Dublin. Contact him at gareth.jones@dcu.ie.

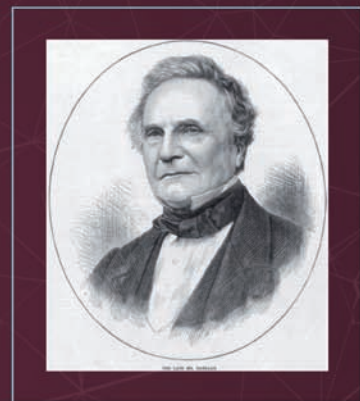


Read your subscriptions through the myCS publications portal at <http://mycs.computer.org>.

IEEE-CS Charles Babbage Award

CALL FOR AWARD NOMINATIONS
Deadline 15 October 2017

- ▶ **ABOUT THE IEEE-CS CHARLES BABBAGE AWARD**
Established in memory of Charles Babbage in recognition of significant contributions in the field of parallel computation. The candidate would have made an outstanding, innovative contribution or contributions to parallel computation. It is hoped, but not required, that the winner will have also contributed to the parallel computation community through teaching, mentoring, or community service.
- ▶ **CRITERIA**
This award covers all aspects of parallel computing including computational aspects, novel applications, parallel algorithms, theory of parallel computation, parallel computing technologies, among others.
- ▶ **AWARD & PRESENTATION**
A certificate and a \$1,000 honorarium presented to a single recipient. The winner will be invited to present a paper and/or presentation at the annual IEEE-CS International Parallel and Distributed Processing Symposium (IPDPS 2017).
- ▶ **NOMINATION SUBMISSION**
Open to all. Nominations are being accepted electronically at www.computer.org/web/awards/charles-babbage. Three endorsements are required. The award shall be presented to a single recipient.



NOMINATION SITE
awards.computer.org

AWARDS HOMEPAGE
www.computer.org/awards

CONTACT US
awards@computer.org