Developing Benchmarks: the Importance of the Process and New Paradigms

Roeland J.F. Ordelman
University of Twente & Netherlands Institute for Sound and Vision
The Netherlands
roeland.ordelman@utwente.nl

ABSTRACT

The value and importance of Benchmark Evaluations is widely acknowledged. Benchmarks play a key role in many research projects. It takes time, a well-balanced team of domain specialists preferably with links to the user community and industry, and a strong involvement of the research community itself to establish a sound evaluation framework that includes (annotated) data sets, well-defined tasks that reflect the needs in the 'real world', a proper evaluation methodology, ground-truth, including a strategy for repetitive assessments, and last but not least, funding. Although the benefits of an evaluation framework are typically reviewed from a perspective of 'research output' –e.g., a scientific publication demonstrating an advance of a certain methodology– it is important to be aware of the value of the process of creating a benchmark itself: it increases significantly the understanding of the problem we want to address and as a consequence also the impact of the evaluation outcomes.

In this talk I will overview the history of a series of tasks focusing on audiovisual search emphasizing its 'multimodal' aspects, starting in 2006 with the workshop on 'Searching Spontaneous Conversational Speech' that led to tasks in CLEF and MediaEval ("Search and Hyperlinking"), and recently also TRECVid ("Video Hyperlinking"). The focus of my talk will be on the process rather than on the results of these evaluations themselves, and will address cross-benchmark connections, and new benchmark paradigms, specifically the integration of benchmarking in industrial 'living labs' that are becoming popular in some domains.

Keywords

Benchmarking; Datasets; Audiovisual Access; Video Hyperlinking

Bio

Roeland Ordelman is Innovation Manager at Netherlands Institute for Sound and Vision, one of the largest audiovisual archives in Europe, and principal investigator Speech and Language Technology in the area of Multimedia Retrieval at University of Twente, the Netherlands. He received his PhD in 2003 on the use of speech recognition technology to improve access to audiovisual archives and worked on a wide range of national and international projects in the area of multimedia retrieval. Next to more fundamental research topics such as currently audiovisual (hyper)linking, research topics addressing aspects related to the uptake of access technology in realistic, industrial use cases have his special interest. In that context, he has been involved in organizing benchmark evaluations in the MediaEval Benchmark Evaluation Series and more recently TRECVid. His current work in this area focuses on the integration of (user) evaluation and benchmarking in the context of 'living labs': research infrastructures that involve industry, research and user communities for mutual benefit.

Acknowledgements

His current research is funded by the Dutch National Research Programme COMMIT/ and the CLARIAH project, funded by the Netherlands Organisation for Scientific Research (NWO)

MMCommons'16 October 16-16 2016, Amsterdam, Netherlands

ACM ISBN 978-1-4503-4515-6/16/10.

DOI: http://dx.doi.org/10.1145/2983554.2983562