

# Event-based MultiMedia Search and Retrieval for Question Answering

Benoit Huet  
EURECOM  
06904 Sophia Antipolis (France)  
Benoit.Huet@eurecom.fr

## ABSTRACT

User generated content, available in massive amounts on the Internet, comes in many "flavors" (i.e. micro message, text document, images and video) and is receiving increased attention due to its many potential applications. One important application is the automatic generation of multimedia enrichments concerning users topic of interests and in particular the creation of event summaries using multimedia data [1]. In this talk, an event-based cross media question answering system, which retrieves and summarizes events on a given user generated query topic is proposed. A framework for leveraging social media data to extract and illustrate social events automatically on any given query will be presented. The system operates in three stages. First, the input query is parsed semantically to identify the topic, location, and time information related to the event of interest (News in this scenario presented here). Then, we use the parsed information to mine the latest and hottest related News from social news web services. Third, to identify a unique event, we model the News content by latent Dirichlet Allocation and cluster the News using the DBSCAN algorithm. In the end, for each event, we retrieve both textual and visual content of News that refer the same event [2,3]. The resulting documents are shown within a vivid interface featuring both event description, tag cloud and photo collage [4].

Popular question answering systems (i.e. YahooAnswers) and search engines retrieve documents on the basis of text information. The integration the visual information within the text-based search for video and image retrieval is still a hot research topic. In the second part of this talk, we propose to use visual information to enrich the classic text-based search for video retrieval [5]. With the proposed framework, we endeavor to show experimentally, on a set of real world scenarios, that visual cues can effectively contribute to significant quality improvement of image/video retrieval. Experimental results show that mapping text-based queries to visual concepts improves the performance of the search system. Moreover, when appropriately selecting the relevant visual concepts for a query, a very substantial improvement of the system's performance is achieved [6].

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the owner/author(s).

MARMI'16, June 06-06 2016, New York, NY, USA  
ACM 978-1-4503-4362-6/16/06.

<http://dx.doi.org/10.1145/2927006.2927015>

Based on the various results presented in this talk we argue that question answering (among other application) can greatly leverage from cross media analysis to the benefit of users.

## Keywords

Events, social media, cross media, question answering, visual concepts, multimedia semantic.

## SHORT BIO

Dr. Benoit Huet is Assistant Professor in the Data Science Department of EURECOM (France). He received his BSc degree in computer science and engineering from the École Supérieure de Technologie Électrique (Groupe ESIEE, France) in 1992. In 1993, he was awarded the MSc degree in Artificial Intelligence from the University of Westminster (UK) with distinction, where he then spent two years working as a research and teaching assistant. He received his DPhil degree in Computer Science from the University of York (UK) for his research on the topic of object recognition from large databases. He was awarded the HDR (Habilitation to Direct Research) from the University of Nice Sophia Antipolis, France, in October 2012 on the topic of Multimedia Content Understanding: Bringing Context to Content. He is associate editor for IEEE Multimedia Magazine, IEEE Transaction on Multimedia, Multimedia Tools and Application (Springer) and Multimedia Systems (Springer) and has been guest editor for a number of special issues (EURASIP Journal on Image and Video Processing, IEEE Multimedia, etc...). He regularly serves on the technical program committee of the top conference of the field (ACM MM/ICMR, IEEE ICME/ICIP). He chaired the following international conference: MMM2019, PCM2014 and multiple workshops. He is Technical Program co-Chair of the upcoming ACM Multimedia conference. He was chairing the IEEE MMTC Interest Group on Visual Analysis, Interaction and Content Management (VAIG) from 2010-2014. He has co-authored over 150 papers in Books, Journals and International conferences. His current research interests include Large Scale Multimedia Content Analysis, Mining and Indexing - Multimodal Fusion - Socially-Aware Multimedia.



## ACKNOWLEDGEMENT

The work presented in this keynote was performed with Xueliang Liu, Mathilde Sahuget and Bahjat Safadi. The research leading to this paper was partially supported by the project AAL-2009-2-049 "Adaptable Ambient Living Assistant" (ALIAS) co-funded by the European Commission and the French Research Agency (ANR) in the Ambient Assisted Living (AAL) programme, by the European Commission under contracts FP7-287911 LinkedTV and FP7-318101 MediaMixer and by Bpifrance within the NexGen-TV Project, under grant number F1504054U.

## REFERENCES

- [1] Xueliang Liu and Benoit Huet. 2016. *Event-based cross media question answering*. *Multimedia Tools and Applications*. 75, 3 (February 2016), 1495-1508. DOI=<http://dx.doi.org/10.1007/s11042-014-2085-0>
- [2] Xueliang Liu and Benoit Huet. 2013. *Event Representation and Visualization from Social Media*. In *Proceedings of the 14th Pacific-Rim Conference on Advances in Multimedia Information Processing — PCM 2013 - Volume 8294*, Benoit Huet, Chong-Wah Ngo, Jinhui Tang, Zhi-Hua Zhou, Alexander Hauptmann, and Shuicheng Yan (Eds.), Vol. 8294. Springer-Verlag New York, Inc., New York, NY, USA, 740-749. DOI=[http://dx.doi.org/10.1007/978-3-319-03731-8\\_69](http://dx.doi.org/10.1007/978-3-319-03731-8_69)
- [3] Xueliang Liu and Benoit Huet. 2013. *Heterogeneous features and model selection for event-based media classification*. In *Proceedings of the 3rd ACM conference on International conference on multimedia retrieval (ICMR '13)*. ACM, New York, NY, USA, 151-158. DOI=<http://dx.doi.org/10.1145/2461466.2461493>
- [4] Xueliang Liu and Benoit Huet. 2013. *EventEnricher: a novel way to collect media illustrating events*. In *Proceedings of the 3rd ACM conference on International conference on multimedia retrieval (ICMR '13)*. ACM, New York, NY, USA, 303-304. DOI=<http://dx.doi.org/10.1145/2461466.2461517>
- [5] Bahjat Safadi, Mathilde Sahuget, and Benoit Huet. 2014. *When textual and visual information join forces for multimedia retrieval*. In *Proceedings of International Conference on Multimedia Retrieval (ICMR '14)*. ACM, New York, NY, USA, , Pages 265 , 8 pages. DOI=<http://dx.doi.org/10.1145/2578726.2578760>
- [6] Bahjat Safadi, Mathilde Sahuget and Benoit Huet. Linking text and visual concepts semantically for cross modal multimedia search. IEEE International conference on Image Processing (IEEE ICIP), Paris, France, Oct 2014