# GLOCAL: Event-based Retrieval of Networked Media

Pierre Andrews, Francesco De Natale
Università degli Studi di Trento
Via Sommarive 14, 38100 Trento, Italy
{andrews, denatale}@disi.unitn.it

Sven Buschbeck, Anthony Jameson
German Research Center for Artificial Intelligence
Campus D3.2, 66123 Saarbrücken, Germany
{sven.buschbeck, jameson}@dfki.de

Kerstin Bischoff, Claudiu S. Firan, Claudia Niederée
L3S Research Center/Leibniz Universität Hannover,
Appelstr. 9a, 30167 Hannover, Germany
{bischoff, firan, niederee}@L3S.de

Vasileios Mezaris, Spiros Nikolopoulos
CERTH-ITI
57001 Thermi-Thessaloniki, Greece
{bmezaris, nikolopo}@iti.gr

Vanessa Murdock, Adam Rae
Yahoo! Research
Diagonal 177, 08018 Barcelona, Spain
{vmurdock, adamrae}@yahoo-inc.com

## ABSTRACT
The idea of the European project GLOCAL is to use events as the central concept for search, organization and combination of multimedia content from various sources. For this purpose methods for event detection and event matching as well as media analysis are developed. Considered events range from private, over local, to global events.

## Categories and Subject Descriptors
H.3.3 [**Information Systems**]: Information Search and Retrieval;
H.5.2 [**Information Systems**]: User Interfaces

## Keywords
Multimedia, event-based retrieval, media organization, UI design

## 1. INTRODUCTION
The last two decades of research on media retrieval brought enormous progresses on the technologies for description, analysis, and retrieval of media using content and (to some degree) context information. Notwithstanding such advances, current implementations of multimedia search engines are still mainly based on textual queries. This reflects the attitude of human beings to describe the world in terms of their natural language, which represents the most common way of communicating each other and expressing concepts and situations.

On the other hand, our life is a constellation of events which, one after the other, pace our everyday activities and index our memories. A birthday, a wedding, or a car accident are the lens through which we memorize our own personal experiences, the same way as a world sport championships or a natural disasters represent collective experiences that we share within larger-scale communities. Thus, the concept of events and their faceted representations in terms of time, space, entities, and relationships, become a natural and powerful way to attach rich contextual information to media, for the purpose of organization, indexing and retrieval. Although the structuring of media according to reflected events is quite intuitive for a human, the transfer of this idea into an intelligent system for event-centered media organization and access imposes several challenges:

A flexible lightweight **event model** has to be created, which is generic enough to cover the wide variety of possible events (e.g. different levels of structuring, different level of granularity), captures major event properties, enables flexible event media linkage and allows for the representation of the various relationships between events (temporal, causal, sub-events, etc.).

Obviously **linking events to media objects** is in the core of the system. Precise and efficient automatic methods for associating media object with event instances and event classes exploiting all types of available information (visual information, tags, existing classifications) are needed for supporting the user in this task.

For managing and evolving the event space methods are required for event detection, which feed and update the event space as well as for event **matching and merging**, which enable the combination of event spaces and the import of event information.

For making the event space easily digestible for the user, user interfaces that support **effective search and navigation** in the possibly large event and media space are required, exploiting event-intrinsic properties such as time, space and involved entities as well as event structure. **Event-based search** has to be supported by indexing, filtering and ranking technologies, which are tailored to event characteristics.

For regional and global events, media creation and even-centered structuring involves a variety of private as well as professional users. For making full use of these assets approaches for **interlinking and combining** the individual contributions are needed such as methods for sharing and exchanging event-related media objects, for migrating event-related media from local, more private to global contexts and for getting event-based recommendations.


*WWW 2012 Companion*, April 16–20, 2012, Lyon, France.
ACM 978-1-4503-1230-1/12/04.

# 2. PROJECT HIGHLIGHTS

## 2.1 Event Model

The link between media and event has been largely discussed in the state of the art of media management; in particular, [10] and [11] discuss diverse models to link media to events and their metadata. Events are generally understood as something that happens in a particular place at a particular time. In addition, the events that need representation are the ones where the users had an interesting experience. An event is thus described by major metadata such as time and location, but also by other contextual information describing the users' experience and the relationships to their media illustrating this experience. While the models proposed in the state of the art already take many of these aspects in account, we have proposed an extended event model, based on a strong entity centric theory, which can help represent different user experiences of the same event. In preliminary research [13], we have found that users already describe such personal event's experiences and share media around them. As discussed in [12], such a local representation of events and their meronymy relationships can be used to match events between diverse users and aggregate media at a global level. This theoretical model is currently being developed to provide a set of web-services that will manage and share events as part of the GLOCAL project.

## 2.2 Media and Events

One of the cornerstones of the event-based organization and retrieval of media is the analysis of the media content itself; with the objective to automatically understand the events captured by it (e.g. understand that a given video fragment shows a demonstration). Such an analysis requires performing a wide range of processing steps on the media content, such as the extraction of discriminative features from images, videos and audio files; the decomposition of the content into appropriate segments (e.g., decomposition of the video to shots and scenes); the detection of concepts in images and videos (both concepts corresponding to physical objects, e.g. "person", "airplane", and more abstract ones, e.g. "outdoor", "sunset"); the detection of event classes that express the media content (e.g., "demonstration", "person assembling a shelter"); and the event-based organization of collections of media into meaningful sub-events (e.g. organize a collection of images into sub-events according to what the images show).

Focusing on the more event-oriented steps of such an analysis, one of the key technologies we developed for event detection in images and videos is based on using discriminant concepts. This builds on the results of automatic concept detection, and combines a model vector representation of content segments with a new discriminant analysis (DA) algorithm [1]. The model vector representation is a vector whose elements are visual concept detection scores in the range [0,1], expressing the degree of confidence that a concept is depicted in the given content segment. Such a representation can be derived using a pool of pre-trained concept detectors, which in our approach comprises 346 Support Vector Machine (SVM)-based detectors, taking as input a Bag of Words representation of an image or keyframe. Subsequently, Mixture Subclass Discriminant Analysis (MSDA) [2] is invoked for identifying the semantic concepts that best describe the event, thus, defining a discriminant concept subspace for each event. This method extends Subclass Discriminant Analysis (SDA) so as to further improve recognition accuracy and degree of dimensionality reduction while exploiting the inherent sub-class structure of events. Finally, in the resulting discriminant subspace, the nearest neighbor classifier (NN) is applied to detect an event. The motivation behind the latter choice is that the NN classifier can work with even a very limited number of positive instances of an event, thus alleviating the need for extensive training datasets at the event level. The overall event detection approach has been successfully tested on the GLOCAL dataset as well as in the TRECVID Multimedia Event Detection (MED) tasks of 2010 and 2011 [7], with good results.

A second approach experimented with was the possibility of discovering event-related information from a collection of media in a top-down manner, i.e., jointly exploiting the common characteristics of a set of visual data associated to a given event type. Two techniques were proposed to this purpose. In the first one, a set of descriptors is extracted from a media album, and is used to (i) detect the event class to which the collection refers, and (ii) structure the collection into a set of sub-events [15]. This approach requires a priori knowledge of the possible event classes and of the relevant models, as well as a training phase where a set of classifiers learns the visual models from a set of sample events. The second approach attempts to define a natural "visual signature" of events, by defining a so-called signature image base (SIB) [17]. This technique builds upon the concept that events belonging to the same class share a number of characteristics (e.g., outdoor-indoor, environment, colors, participants, etc.) that directly reflect on the visual characteristics of the relevant media collections, leaving a sort of fingerprint on them. The SIB extracts such fingerprint by jointly capturing the saliency and gist information along the time direction and constructing a simple model that can be used for event and sub-event recognition.

## 2.3 Events and Communities

In the context of social networks the detection of communities has attracted considerable interest and has been subject to various interpretations. The term community is typically used to either refer to groups of users, or groups of media resources. Applying community detection at the level of media facilitates the tasks of new event detection since the resulting groups of resources may potentially embody an event. Thus, given a large set of tagged images with geo-location information our goal within GLOCAL is to identify groups of images representing a new event.

Our method integrates heterogeneous types of information (i.e. visual and tag-based) into a space spanned by a set of latent topics, where events and places can be easily detected via a clustering process that organizes photographic content into groups. The image groups are obtained by means of a graph-based image-clustering algorithm that operates on the fused space of latent topics. This latent space is obtained by the employment of aspect models that are able to express a high dimensional word distribution vector as a low dimensional mixture of latent topics. For representing media in either visual or tag information space we employ the Codebook representation approach where each image is represented as an occurrence count histogram of the representative "words" in its content [6]. Finally, the fused media representation is extracted by employing high order pLSA [5]. High order pLSA is essentially the application of pLSA [3] to more than two observable variables allowing the incorporation of different word types into the analysis process. By treating images, visual content and tags as the three observable variables of an aspect model we manage to extract a set of latent topics that incorporate the semantics of both visual and tag information

space. In this way, we succeed in devising a feature extraction scheme where the co-existence of two "words" that are known from experience to appear together rather frequently is more important in defining the latent topics, than the co-existence of two words that rarely appear together and are likely to be the result of noise. Based in this semantically enhanced feature space GLOCAL facilitates the detection of new (i.e. previously unknown) events from user contributed content.

Additionally, GLOCAL also assigns images to already existing events. As presented in [14], this functionality relies on a Naïve Bayes Multinomial classifier, having as features user generated content like tags, description, etc. and as classes the list of events. Users are thus notified of currently existing, appropriate events, when they add a new image to GLOCAL.

Automatically sharing media and events by recommending them to friends connected in the platform is a popular feature. However, deciding when to propagate events and media is not trivial, since not all social connections are just as useful. The importance of not treating all online relationships as equal has just been accounted for in platforms like Google+ and Facebook. While in these platforms users have to manually maintain circles or lists of close friends, family etc, we are developing machine learning methods that identify strong friends automatically based on demographics, social network overlap, taste preferences and interaction data (e.g. comments exchanged). Exploiting such features we also recommend 'new' friends, i.e. people you might know, with over 94% accuracy. For event recommendations, the different kinds of ties are supposed to hold different potentials: Strong ties offer trust and reciprocity as well as shared interests, while weak ties bring the benefit of novelty and diversity.

## 2.4 Event Localization

Social media is a rich source of information about the global and local events in a person's life. Due to the prevalence of smart phones, much of the social media data available to researchers is associated with geographic coordinates. This allows us to build models of locations, to predict where a user is, what is around them and what events they might be participating in as well as to infer the geographic intent of their interaction with a given application. It also allows us to discover new venues and previously unknown places of interest where events might take place.

The basic approach to identifying the location of an event or of a user is to subdivide the globe into grid cells of 1 km sq. We then associate with each cell the textual metadata of the social media that emanated from the cell [8]. For example, we extract the tag sets from the geotagged images uploaded to Flickr, and associate each tag with the cell indicated by the geographic coordinates of the image. We estimate the term distribution according to the number of users who used a given term in that grid cell, smoothing from the term distribution of the entire globe. We can then compute the geographic intent of a user (represented by a query history, or a set of image tags) by the probability that the cell term distribution is similar to the user's query terms (or image tags). Figure 1 shows a screenshot of the demonstrator for this service, which is accessible via a public REST API (see http://glocal.research.yahoo.com).

To predict the location of the user, the same type of setup is used, but rather than associating cells with the social media emanating from the cell, we associate the cell with the collection of social media for a given user, for users whose most frequent location is a given cell [4]. This allows us to predict the location of users according to their term distributions. Annotated social media data can also be used to learn how to connect events with specific venues or points of interest (POIs). In order for this to be useful, knowledge about such POIs needs to be gathered before it can be associated with events. However, existing commercial and academic sources for such data tend to be expensive, incomplete, prone to staleness, biased in their coverage of the world and usually dependant on significant manual labor.

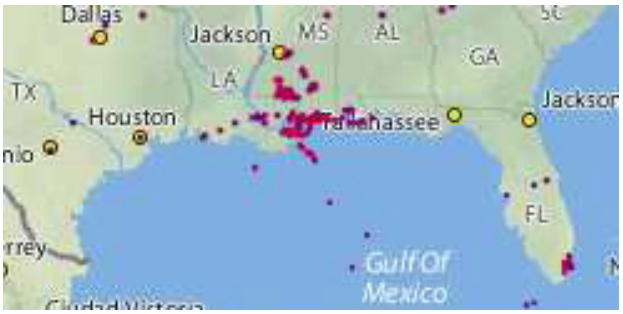


**Figure 1: Locations deemed relevant to the query "Hurricane Katrina". More relevant locations are denoted with a brighter red. Less relevant locations are denoted with a darker blue**

In order to mitigate these problems, we automatically detect and learn about the points of interest (POIs) from text on the web by using a conditional random fields approach to sequence tagging that allows us to annotate text with respect to whether the tokens in the text represent a mention of a POI [9]. No large-scale manually annotated data collections of sufficient quality exist that use the specific label of POI, and so we have developed a technique that uses a number of social media data sources to produce effective training data. We extract check-in status updates from Foursquare and Gowalla via Twitter, as well as page titles from geographic Wikipedia pages (according to the Yago taxonomy) to produce seed queries that we submit to a web search engine. The resulting snippets are then used as example of POIs used in sentences that our tagger can be trained on. The resultant tagger is capable of detecting POI mentions in manually annotated test data sets derived from text found online more effectively than the state-of-the-art Yahoo! Placemaker web service.

# 3. DEMONSTRATION AND THE UI

The services described above have been integrated into the GLOCAL system and equipped with an event-oriented, user-friendly UI, which will be used to demonstrate the GLOCAL functionality (see also http://www.dfki.de/glocal/latest for an annotated UI prototype and [16] for further information) for Soccer and an Arab Spring event scenario. The UI especially supports the navigation in structured events as well as the faceted search in the event and media spaces. Figure 2 illustrates the navigation capabilities for the example of use case Soccer World Championship, which has a strongly hierarchical structure, connecting an event with its sub-events via a sub-event link. Sub-events can be shown and hidden on demand (see Figure 2).

Furthermore, the UI supports media upload, the manual as well as the semi-automatic annotation of media object with event and event class information based on the services described in section 2 and the import of media from other sources such as Flickr. In addition, professional or amateur users can link navigable views


**Figure 2: (A) A filtered and partly collapsed representation of the 2010 soccer World Cup as a hierarchy of events); (B) The user has zoomed in on a single game and clicked on the "media" links for two goals, so as to be able to compare the associated media.**

of the event space with textual narratives using them as illustrations instead of merely commenting on individual events or media. The respective view can be interactively defined via the interface and stored. An Android application providing an additional UI for GLOCAL has also been developed and brings the GLOCAL functionality closer to where media is actually created, introducing the event-centered approach early in the media creation and management process.

## 4. FURTHER WORK

Activities that are planned for the coming year of the project include the development of methods for contextualizing local event knowledge as well as for combining and merging autonomously created event representations. Moreover, media and textual content will be augmented by opinion analysis, providing diverse points of view over entities involved in the events. Other important targets are security concerns and the validation of GLOCAL technologies through large-scale user trials, involving communities of professional as well as generic users. Future directions of GLOCAL research are manifold. First of all, there is a plan to extend the current set of event models to cover a larger variety of situations, as well as to broaden the media technology toolbox to automate as much as possible the event-based organization of media. Finally, we imagine the application of the GLOCAL backend to create a wide set of new facilities, making it possible to exploit event-based information to automatically create storyboards, multimedia summaries, historical reviews, or virtually any desired style of media presentation.

## 5. ACKNOWLEDGMENTS

This work was supported by the GLOCAL project funded by the European Commission under the 7th Framework Programme (Contract No. 248984).

## 6. REFERENCES

[1] Gkalelis, N., Mezaris, V. and Kompatsiaris, I. High-level event detection in video exploiting discriminant concepts. International Workshop on Content-Based Multimedia Indexing, 2011.

[2] Gkalelis, N., Mezaris, V. and Kompatsiaris, I. Mixture subclass discriminant analysis. IEEE Signal Processing Letters, 2011.

[3] Hofmann, T. Probabilistic latent semantic analysis. Uncertainty in Artificial Intelligence, 1999.

[4] Kinsella, S., Murdock, V., O'Hare, N. "I'm Eating a Sandwich in Glasgow: Modeling Locations with Tweets" Workshop on Social Media and User-Generated Content, 2011.

[5] Nikolopoulos, S., Giannakidou, S., Kompatsiaris, I., Patras, I., and Vakali, A., Combining multi-modal features for social media analysis. Social Media Modeling and Computing. Springer 2011.

[6] Sivic, J. and Zisserman, A. 2003. Video google: A text retrieval approach to object matching in videos. Proce. of the 9th International Conference on Computer Vision, 2003.

[7] Moumtzidou, A., Sidiropoulos, P., Vrochidis, S., Gkalelis, N., Nikolopoulos, S., Mezaris, V., Kompatsiaris, I. and Patras, I. 2011. ITI-CERTH participation to TRECVID 2011 Workshop.

[8] O'Hare, N., Murdock, V. Modeling Locations with Social Media (in submission).

[9] Rae, A., Murdock, V., Popescu, A., Bouchard, H. Bootstrapping Points of Interest with Social Media (in submission).

[10] Westermann, U., Jain, R.: Toward a common event model for multimedia applications. IEEE MultiMedia, 2007

[11] Shaw, R., Troncy, R., Hardman, L.: Lode: Linking open descriptions of events. ASWC, 2009

[12] Giunchiglia, F., Andrews, P., Trecarichi, G., Chenu-Abente R. Media Aggregation via Events. Workshop on Recognising and Tracking Events on the Web and in Real Life, 2010

[13] Andrews, P., Paniagua, J., Giunchiglia F. Clues of Personal Events in Online Photo Sharing; DERIVE, 2011

[14] Firan, C.S., Georgescu, M., Nejdl, W., Paiu, R. Bringing Order to Your Photos: Event-Driven Classification of Flickr Images Based on Social Knowledge. CIKM, 2010.

[15] Mattivi, R., Boato, G., De Natale, F.G.B. Event-based media organization and indexing. Infocommunications Journal, 2011.

[16] Buschbeck, S., Jameson, A., Schneeberger, T., New Forms of Interaction With Hierarchically Structured Events; DERIVE, 2011.

[17] Dao, M.-S., Dang-Nguyen, D.-T., De Natale, F.G.B. Signature-Image-Based Event Analysis for Personal Photo Albums. ACM Multimedia 2011