Finding Media Illustrating Events

Xueliang Liu
EURECOM
Sophia Antipolis, France
xueliang.liu@eurecom.fr

Raphaël Troncy
EURECOM
Sophia Antipolis, France
raphael.troncy@eurecom.fr

Benoit Huet
EURECOM
Sophia Antipolis, France
benoit.huet@eurecom.fr

ABSTRACT

We present a method combining semantic inferencing and visual analysis for finding automatically media (photos and videos) illustrating events. We report on experiments validating our heuristic for mining media sharing platforms and large event directories in order to mutually enrich the descriptions of the content they host. Our overall goal is to design a web-based environment that allows users to explore and select events, to inspect associated media, and to discover meaningful, surprising or entertaining connections between events, media and people participating in events. We present a large dataset composed of semantic descriptions of events, photos and videos interlinked with the larger Linked Open Data cloud and we show the benefits of using semantic web technologies for integrating multimedia metadata.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information System**]: Audio, Video and Hypertext Interactive Systems; I.7.2 [**Document Preparation**]: Languages and systems, Markup languages, Multi/ mixed media, Standards

General Terms

Languages, Hyperlinks, Web, URI, HTTP

Keywords

Events, LODE, media ontology, multimedia semantics

1. INTRODUCTION

Events are a natural way for referring to any observable occurrence grouping persons, places, times and activities that can be described [16]. Events are also observable experiences that are often documented by people through different media (e.g. videos and photos). We explore this intrinsic connection between media and experiences so that people can search and browse through content using a familiar event perspective. We are aware that web sites already exist that provide interfaces to such functionality, e.g. eventful.com, upcoming.org, last.fm/events, and facebook.com/events to name a few. These services have sometimes explicit connection with media sharing platforms, have often overlap in terms of coverage of upcoming events and provide social networks features to support users in sharing and deciding upon attending events. However, the information about the events, the social connections and the representative media are all spread and locked in amongst these services providing limited event coverage and no interoperability of the description [5].

Our goal is to aggregate these heterogeneous sources of information using linked data, so that we can explore the information with the flexibility and depth afforded by semantic web technologies. Furthermore, we investigate the underlying connections between events to allow users to discover meaningful, entertaining or surprising relationships amongst them. We also use these connections as means of providing information and illustrations about future events, thus enhancing decision support. In this paper, we present a method for finding automatically medias hosted on Flickr and YouTube that can be associated to a public event. We show the benefits of using linked data technologies for enriching semantically the descriptions of both events and media.

The remaining of this paper is structured as follow. In Section 2, we briefly describe the LODE event model and how we scrap large event directories. In Section 3, we present the dataset on which we will evaluate our method. We then detail our approach for associating media with events (Section 4). We discuss our results in Section 5 and present some related work in Section 6. Finally, we give our conclusions and outline future work in Section 7.

2. LODE AND EVENT DIRECTORIES

Large numbers of web sites contain information about scheduled events, of which some may display media captured at these events. This information is, however, often incomplete and always locked into the sites. In previous research, we carried out user studies in order to collect end-user experiences, opinions and interests while discovering, attending and sharing events, and user insights about potential web-based technologies that support these activities. The results of this study support the development of an environment that merges event directories, social networks and media sharing platforms [5]. We argue that linked data technologies is suitable for doing this integration at large scale given they naturally based on URIs for identifying objects

ICMR '11, April 17-20, Trento, Italy

and a simple triple model (RDF) for representing semantic descriptions. In this section, we present the LODE event model and how we populate this ontology by scraping three large event directories: last.fm, eventful and upcoming.

2.1 LODE by Example

The LODE ontology[1] is a minimal model that encapsulates the most useful properties for describing events [11]. The goal of this ontology is to enable interoperable modeling of the "factual" aspects of events, where these can be characterized in terms of the *four Ws*: *What* happened, *Where* did it happen, *When* did it happen, and *Who* was involved. LODE is not yet another "event" ontology *per se*. It has been designed as an *interlingua* model that solves an interoperability problem by providing a set of axioms expressing mappings between existing event ontologies. Hence, the ontology contains numerous OWL axioms stating classes and properties equivalence between models such as the Event Ontology [10], CIDOC-CRM, DOLCE, SEM [15] to name a few.

Figure 1 depicts the metadata attached to the event identified by `350591` on last.fm according to the LODE ontology. More precisely, it indicates that an event of type `Concert` has been given on the `13th of July 2007 at 20:30 PM` in the `Nouveau Casino` theater in Paris featuring the Irish singer `Róisín Murphy` known for electronic style.

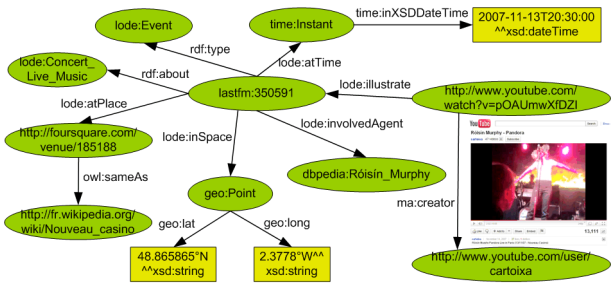


Figure 1: *Róisín Murphy at the Nouveau Casino in Paris* described with LODE

2.2 Scraping Event Directories

We use the Last.fm, Eventful and Upcoming APIs to convert each event description into the LODE ontology. We mint new URIs into our own namespace for events (`http://data.linkedevents.org/event/`), agents (`http://data.linkedevents.org/agent/`) and locations (`http://data.linkedevents.org/location/`). A graph representing an event is composed of the type of the event, a full text description, the agents (e.g. artists) involved, a date (instant or interval represented with OWL Time [7]), a location in terms of both geographical coordinates and a URI denoting the venue and users participation. A graph representing an agent or a location is composed of a label and a description (e.g. the artist's biography).

Event directories have overlap in their coverage. We interlink these events descriptions when they involve the same agents at the same date or when they happen at the same venue at the same date. We invoke additional semantic web lookup services such as dbpedia and freebase, or foursquare and geonames in order to enrich the descriptions of the agents and the locations. Hence, the venue has been converted into a foursquare URI (`http://foursquare.com/venue/185188`) that provides additional information such as the number of different users that have *check in* at this place and the current virual mayor while the wikipedia URI (`http://fr.wikipedia.org/wiki/Nouveau_casino`) provides the history of this venue in French.

The agent URI, which has for label "Róisín Murphy" has also been interlinked with the dbpedia URI (`http://dbpedia.org/resource/RóisínMurphy`) which provides additional information about the solo singer such as its complete discography. This URI is declared to be `owl:sameAs` another identifier from Freebase (`http://rdf.freebase.com/ns/guid.9202a8c04000641\-f80000000004a1685`) which provides information about the 2 bands she has been part of. The linked data journey can be rich and long. One of the challenges we want to address is how to visualize these enriched interconnected datasets while still supporting simple user tasks such as searching and browsing enriched media collections.

3. DATASET

Explicit relationships between scheduled events and photos hosted on Flickr can be looked up using special machine tags such as `lastfm:event=XXX` or `upcoming:event=XXX`. In a previous work, we explored the overlap in metadata between four popular web sites, namely Flickr as a hosting web site for photos and Last.fm, Eventful and Upcoming as a documentation of past and upcoming events [14]. Hence, we have been able to convert the description of more than 1.7 million photos which are indexed by nearly 110.000 events (Table 1).

	Event	Agent	Location	Photos	User
`Last.fm`	57,258	50,151	16,471	1,393,039	18,542
`Upcoming`	13,114		7,330	347,959	4,518
`Eventful`	37,647	6,543	14,576	52	12

Table 1: Number of event/agent/location and photo/user descriptions in the dataset published in [14]

In this paper, we consider a subset of this events dataset that corresponds to the intersection of Last.fm, Flickr and YouTube. In other words, we consider the set of last.fm events for which there is at least one photo and one video shared respectively on Flickr and YouTube that has been tagged with the `lastfm:event=xxx` machine tag. The number of YouTube videos that actually contains such a machine tag is unsurprisingly much smaller. Hence, this intersection yields a dataset of 110 events, 4790 photos and 263 videos.

The Ontology for Media Resource currently developed by W3C is a core vocabulary which covers basic metadata properties to describe media resources[2]. It provides properties for describing the duration of a video, its target audience, copyright, genre, rating or the various renditions of a photos. Media fragments can also be defined in order to have a smaller granularity and attach keywords or formal annotations to parts of a media. The ontology contains a formal set of axioms defining mapping between different metadata formats for multimedia. We use this vocabulary together with properties from SIOC, FOAF and Dublin Core to convert into RDF the Flickr photo and YouYTube video descriptions

[1] `http://linkedevents.org/ontology/`

[2] `http://www.w3.org/TR/mediaont-10/`

(Figure 2). The link between the media and the event is realized through the `lode:illustrate` property, while more information about the `sioc:UserAccount` can be attached to his URI. In Figure 2, we see that both the video hosted on YouTube and the photo hosted on Flickr has the same `ma:creator`: the user `cartoixa`.

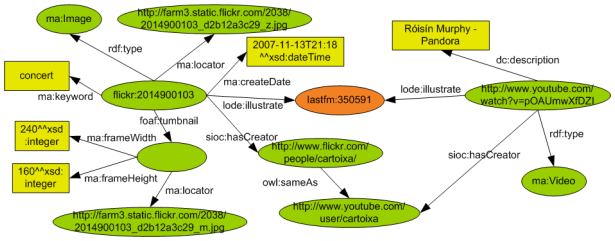


Figure 2: A photo and a video taken by the same user at the *Róisín Murphy Concert* described with the Media Ontology

4. FIND MEDIA ILLUSTRATING EVENTS

The set of photos and videos available on the web that can be explicitly associated to a Last.fm event using a machine tag is generally a tiny subset of all media that are actually relevant for this event. Our goal is to find as much as possible media resources that have **not** been tagged with a `lastfm:event=xxx` machine tag but that should still be associated to an event description. In the following, we investigate several approaches to find those photos and videos to which we can then propagate the rich semantic description of the event improving the recall accuracy of multimedia query for events.

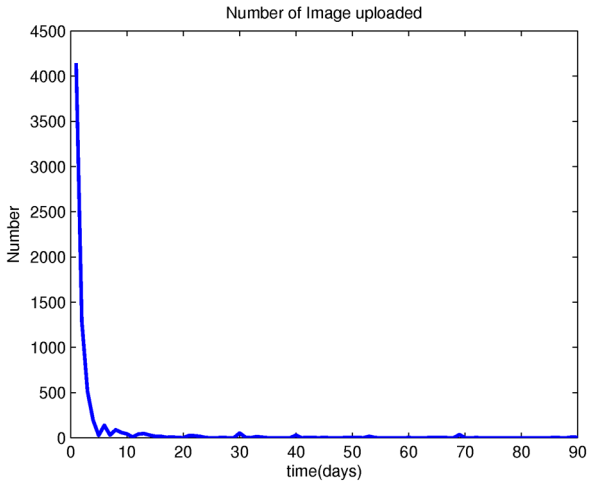


Figure 3: Image uploading tendency along time

Starting from an event description, three dimensions from the LODE model can easily be mapped to metadata available in Flickr and YouTube and be used as search query in these two sharing platforms: the *what* dimension that represents the title, the *where* dimension that gives the geocoordinates attached to a media, and the *when* dimension that is matched with either the taken date or the upload date of a media. Querying Flickr or YouTube with just one of these dimensions bring far too many results: many events took place on the same date or at nearby locations and the title is often ambiguous. Consequently, we will query the media sharing sites using at least two dimensions. We also find that there are recurrent annual events with the same title and held in the same location, which makes the combination of "title" and "geo tag" inaccurate. In the following, we consider the two combinations "title" + "time" and "geotag" + "time" for performing search query and finding media that could be relevant for a given event.

4.1 How Fast Media are Uploaded?

We first investigate the time difference between the start time of an event and the upload time of Flickr photos attached to this event. For the 110 events composing our dataset, we analyze the 4790 photos that are annotated with the Last.fm machine tag in order to compute the time delay between the event start time and the time at which the photos were captured according to the EXIF metadata. Figure 3 shows the result: the y-axis represents the number of photos uploaded on a day to day basis, while the x-axis represents the time (in days) after the event occurred.

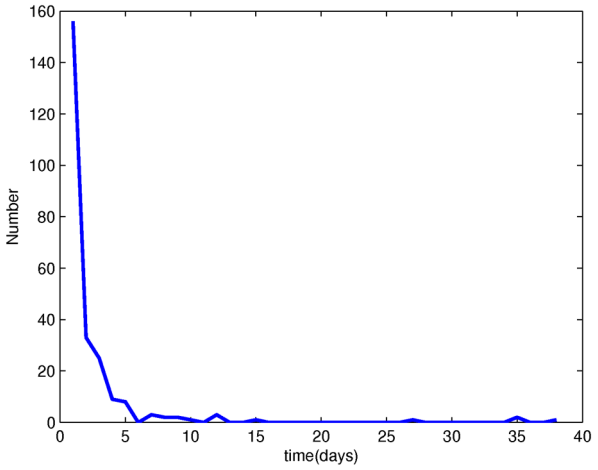


Figure 4: Video uploading tendency along time

The trend is clearly a long-tail curve where most of the photos taken at an event are uploaded during or right after the event took place and within the first 5 days. After ten days, only very few photos from the event are still being uploaded. In the following, we choose a threshold of **5 days** when querying the photos using either the title or the geotag information.

We conduct a similar analysis with the 263 YouTube videos that are annotated with the Last.fm machine tag. The "taken time" being not available for videos from the YouTube API, we use instead the "upload time". Figure 4 shows the results and we observe the same long tail: most the videos are uploaded within the first 5 days following an event.

4.2 Query by Geotag

Geotagging is the process of adding geographical identification metadata to a media and is a form of geospatial metadata. These data usually consist of latitude and longitude coordinates, though they can also include altitude, bearing, distance, accuracy data, and place names. They are extremely valuable for application to structure the data

according to location and for users to find a wide variety of location-specific information [1, 17]. Considering that a place is generally a venue, we assume that at any given place and time there is a single event taking place.

For all events of our dataset, we extract the latitude and longitude information from the LODE descriptions and we perform search query using the Flickr API applying a time filter of 5 days following each event date. We perform the same query using the YouTube API although the number of video that are geotagged is much smaller than for photos. Figures 5(a) and 5(b) show the distribution of the number of retrieved photos and videos for the 110 events in our dataset. We observe that the data is centralized in the left bins which means that for most of the events (n=95), the number of photos (resp. videos) retrieved with geotags is within the 0-100 range (resp. 0-20 range). The largest bin is composed of 45 events that have each between 1 and 50 photos retrieved.

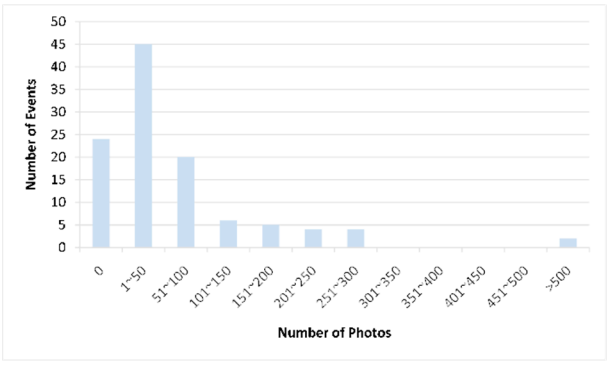
4.3 Query by Title

The title is often the most useful information for describing the events. Similarly to geo-tagged queries, we perform full text search queries on Flickr and YouTube based on the event title that is extracted from the LODE description. The photos and videos retrieved are also filtered using a time interval of five days following the time of the event. When performing search query using the Flickr API query, we use the "text mode" rather than the "tag mode" since the latter is missing in many photos. The number of photos retrieved at this stage is however in an order of magnitude greater than with geo-tagged queries. Due to the well-known polysemy problems of textual-based query, the title-based query brings lots of irrelevant photos. We describe in the Section 4.4 an heuristic for filtering out those irrelevant media.

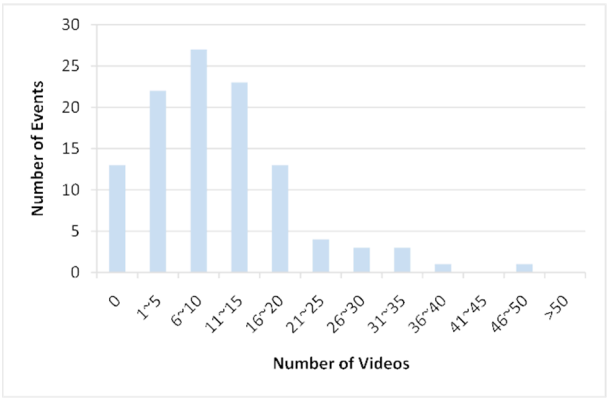
In contrast, we do not observe this noise when querying the YouTube API with only the event title (filtered by the time of the event) using a strict match mode. Hence, the number of videos retrieved per event is rather small and most of the time relevant. The distribution of the number of retrieved photos and videos for the 110 events in our dataset is depicted in Figures 6a and 6b.

Generally, the results of query by title have a similar distribution than the result of query by geotag. For most of the events, a lower number of photos is obtained. Out of the 110 events under investigation, there are 80 events with less than 150 photos, and 83 events with less than 25 videos. However, for some events, a large number of media is retrieved: 12 events (resp. 15) with more than 500 photos (resp. 50 videos). Compared with Figure 5, we can clearly see that the standard deviation of Figure 6 is larger and that again photos are more readily available than videos.

Table 2 shows the overall number of photos and videos retrieved for each strategy for the 110 events that composed our dataset. We first observe that these two strategies allow to retrieve an order of magnitude more media that using solely machine tags. Hence, while 4790 photos are tagged with the `lastfm:event=xxx` machine tag, 6933 photos can be retrieved using the geo-location of the event and 32583 photos can be retrieved using the event title. After removing the duplicated ones, we obtain 36412 photos that are candidate to illustrate an event which is 7,6 times more than the ones labeled by a machine tag. For the videos, the number of candidates is 19,6 times more than the ones with machine tags. Unsurprisingly, most of the media uploaded and shared on the web do not have machine tag.



(a) Number of photos per event in geotag based query



(b) Number of videos per event in geotag based query

Figure 5: Statistics for geotag based query

4.4 Pruning Irrelevant Media

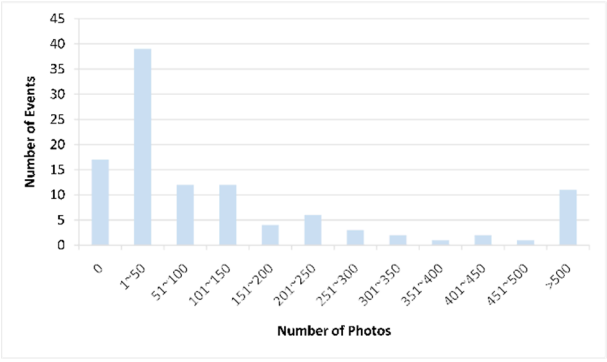
Images and videos with specific machine tags such as `lastfm:event=207358` can be unconditionally associated to events. We consider that media retrieved with geotag queries during a correct time frame should also be relevant for those events. The problem arises with the media retrieved with text-based queries (using the event title) where one can find many irrelevant media. For example, the event identified by 207358 has for title `Malia`. However, a search on Flickr or YouTube with this keyword returns photos about cities, different people (Malawian singer, French swimmer, daughter of the US president Barack Obama) or even hotels with this name.

In order to filter out this noise and to avoid propagating rich event descriptions to those medias, we propose a method for pruning the set of candidates photos using visual analysis. The photos captured at a single event are already very diverse, depicting the artist, the scene, the audience or even the tickets. The diversity of the data makes it difficult to remove all the noisy images that should not be associated with the event considered, while keeping as much as possible the good ones. We address this issue in two steps to ensure high precision and recall ratio.

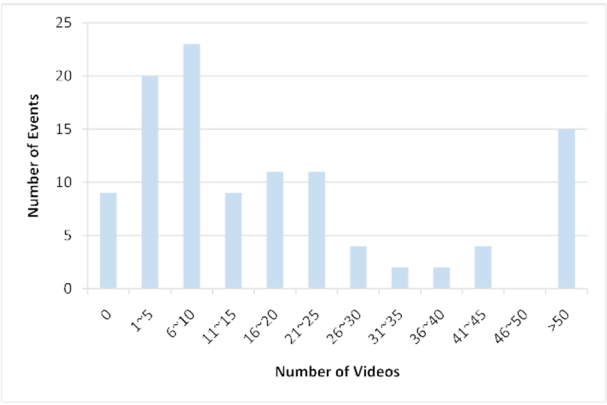
First, we build a training dataset composed of the media containing either the event machine tag or a combination of geo-coordinates and time frame corresponding to the event

Table 2: Number of photos and videos retrieved for 110 events using the event machine tag (ID), the geo-coordinates or the event title

	QueryByID	QueryByTitle	QueryByGeo	ID∩Title	Geo∩Title	Geo∩ID	Geo∩ID∩Title
Photos	4790	32583	6933	2350	494	484	405
Videos	263	4237	1163	103	39	115	29



(a) Number of photos per event in title based query



(b) Number of videos per event in title based query

Figure 6: Statistics for title based query

dimensions. The photos resulting from query by title compose the testing dataset. The visual features employed are 225D color moments in Lab space, 64D Gabor texture, and 73D Edge histogram. For each image in the training data, the nearest neighbors using the ***L1*** distance measure in the training set are found and the smallest distance taken as threshold. Second, images originating from the title query are compared with training images. Images for which the distance to images in the test set is below the threshold are candidates for illustrating the event. The algorithm can be formalized as followed:

Algorithm 1 Prune function

INPUT: $TrainingSet$,$TestingSet$
OUTPUT: $PrunedSet$

for each img *in* $TrainingSet$ **do**
  $D = [\,]$
  for each $imgj$ *in* $TrainingSet$-$\{img\}$ **do**
    D.append(dist_L1(img,$imgj$))
  end for
  $Threshold = \min(D)$
  for each $imgt$ *in* $TestingSet$ **do**
    if dist_L1($imgt$,img)< $Threshold$ **then**
      $PrunedSet$.append($imgt$)
    end if
  end for
end for
return $PrunedSet$

We adopted an adaptive threshold because of the visual diversity within the training dataset. Even for the images belong to the same event, the concept can vary from the musicians, singer to venue, or event ticket. In order to remove noisy images in the testing data, the threshold should adjust respectively. Figure 7 shows the value of threshold used in the experiments which range from 0.01 to 0.346.

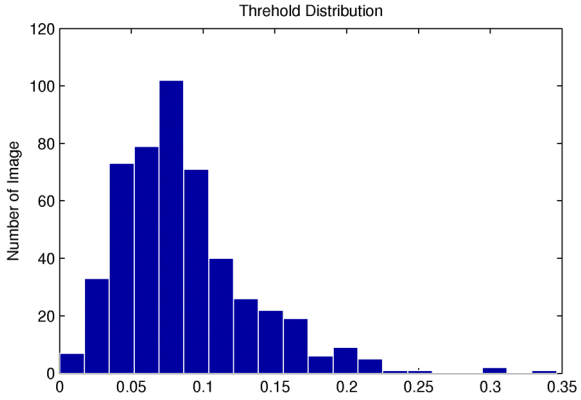


Figure 7: The distribution of threshold

Table 3: Number of photos for 20 events, results of the pruning algorithm and results of the simple heuristic extension

ID	DataSet (nb of photos)			Pruning Result			Extended Heuristic	
	TrainingData	TestingData	GroundTruth	Pruned	Precision	Recall	Extend	NewRecall
346054	2	24	2	1	1	0.500	1	0.500
158744	3	48	48	23	1	0.479	44	0.917
371981	4	16	6	4	1	0.667	4	0.667
341832	7	0	0	0	1	1.000	0	1.000
362195	7	0	0	0	1	1.000	0	1.000
235445	10	1	1	0	1	0.000	0	0.000
42644	13	85	81	13	1	0.160	13	0.160
165697	23	1	1	0	1	0.000	1	1.000
137530	24	9	4	0	1	0.000	1	0.250
517159	24	0	0	0	1	1.000	0	1.000
222241	36	204	180	33	0.97	0.183	72	0.400
234649	45	35	4	1	1	0.250	1	0.250
207358	54	68	4	4	1	1.000	4	1.000
429517	60	171	169	27	1	0.160	41	0.243
437747	65	144	142	8	1	0.056	13	0.092
117886	68	99	97	4	1	0.041	11	0.113
150390	71	16	16	1	1	0.063	1	0.063
350591	79	85	85	6	1	0.071	66	0.776
472733	93	500	478	8	1	0.017	18	0.038
176257	97	260	255	47	1	0.184	147	0.576

4.5 Experiments

For evaluating our pruning algorithm, we take the top 20 events from our 110 events dataset. For these 20 events, there are 785 images in the training set (photos containing either an event machine tag or a geotag) and 1766 photos in the testing set (photos retrieved by event title). We build manually the ground truth for those 1766 photos selecting which ones should be attached to an event and which ones should not (Table 3). The 20 events were all concert events and photos are often depicting artists, venues, stages or audience. Some photos were, however, sometimes hard to judge but the manual assessor used all metadata available around each photo such as the entire list of tags or the albums in which the photos were gathered to decide whether the photo should be discarded or not. In the end, we manually remove 193 irrelevant images by their visual appearance and metadata. The remaining 1593 images are used as ground truth dataset.

The results of the pruning algorithm detailed in the Section 4.4 applied to the 1766 photos are shown in the Table 3. The threshold used is quite strong in order to guarantee a precision of **1** for most of the events. However, this causes about 80% of the candidate images to be excluded, including many relevant photos.

In order to increase the recall ratio, we extend the selected images by our prunning algorithm with all the ones uploaded by the same uploader. The rationale is that if one photo can reliably be attached to an event, we infer that this person indeed attended this event and that all the others photos taken by this person during this time frame are likely to be illustrative media for this event. This simple heuristic allows to significantly improve the recall ratio without sacrifying to the precision.

5. DISCUSSION

Event directories are largely overlapping, providing multiple identifiers for the same venues, artists, and events. We argued that linked data technologies help to integrate at large scale all data sources because of the use of URIs for identifying objects and a simple triple model for representing all metadata yielding a giant graph. Rich semantic descriptions of events can then be propagated to the media to which they are attached. Hence, for the dataset[3] presented in the Section 3, 1,248,021 photos (that is 73 %) have been geotagged for free since Flickr had no geo-tagged information for those photos but only knowledge of an event machine tag that points to a rich description of an event including venues that are geo-localized. Similarly, the propagation of semantic metadata enables to detect inconsistencies between data sources such as the misplacement of a venue.

We have proposed a method for finding media that are relevant for an event based on queries using several dimensions of the event, and pruning the resulting results using visual similarity. However, we observe that there is limited value in pruning video results yet while it is very necessary for photos. Although the total number of video uploads is still exponentially increasing[4], duplicates or videos with absolute no metadata and a very small number of views are important which prevent their discovery.

We also investigate the concept shift as the set of relevant media increases by looking at the tag cloud associated to these media attached to a particular event. Figure 8 depicts tag cloud examples for the event `1097166` that corresponds to the live concert of Alela Diane which took place on Tuesday 14 July 2009 at 7:30pm at the venue `Tivoli De Helling` in Utrecht, The Netherlands. From the three sub figures, we can clearly see the topic shift when new metadata is added

[3]The entire dataset is composed of more than 30 million RDF triples and is available as a dump at `http://www.eurecom.fr/~troncy/ldtc2010/`

[4]YouTube reported in November 2010 that 35 hours of video content are uploaded every minute.

Figure 8: Tag Clouds of photos associated to the event 1097166: *Alela Diane at Tivoli De Helling (Utrecht) on 14 Jul 2009*

from a another source. Figure 8(a) is built from all the tags of all photos retrieved when using the event machine tag: the most frequent keyword in this cloud is, as expected, `lastfm:event`. However, in Figure 8(b), when the tags set are enlarged by the metadata from the photos retrieved by query by title (and pruned with our algorithm), the topic shifts to `aleladiane` who is the artist performing during this event. A similar observation can be made after looking at Figure 8(c), where some metadata from query by geotag are added, and the most significant keyword changes to the location of the event `utrecht`.

Finally, we briefly present initial interfaces that we have started to develop for searching and browsing media through an event perspective [13]. Users wish to discover events either through invitations and recommendations, or by filtering available events according to their interests and constraints. Therefore, the interface allows constraining different event properties (e.g. time, place, category). Mechanisms for providing this desired support include restricting a time period through a timeline slider control input (Figure 9). Categories and location can be filtered using hierarchical faceted metadata [6], allowing users to browse through different dimensions of the collection. The hierarchical facets are presented according to a taxonomy of predefined event categories, and through an event's geo-location information. These properties allow the combination of different event types and locations while visually guiding the user through an interactive query refinement process. Faceted browsing also avoids empty results by restricting the available filtering options to display only non-empty results. Since users are likely to revisit information they have viewed in the past [6], we will also support simple history mechanisms, by saving a list of recently viewed events. To aid search, input boxes with dynamic term suggestions (auto-completion) is used to provide user feedback by suggesting a list of matching terms while typing.

After an event is selected, all associated information is displayed. Media are presented to convey the event experience, along with social information to provide better decision support. According to user interests, social proximity should be emphasized while displaying event attendance (e.g. friends attending). Other information that should be presented includes: performers, topic, genre, price. While scraping the data, some events such as popular music festivals were associated with more than 2,000 photos and videos. In order to deal with this large number, pagination is used while ordering media according to different contexts (e.g. by popularity, time, or social proximity). Alternatively, pictures are clustered according to context or visual similarity, and representative images are shown through Treemaps to present a varied sample of associated media.

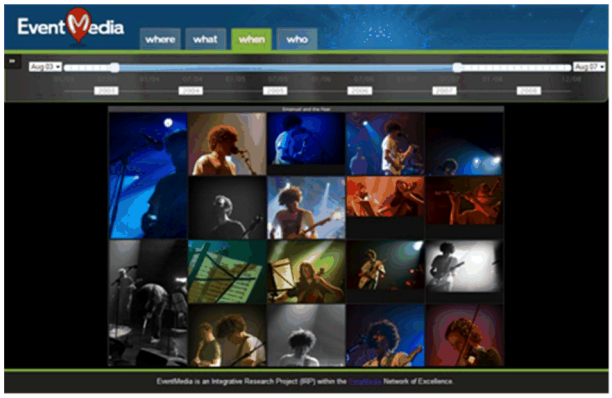


Figure 9: Interface illustrating a set of media associated to an event for a period of time

6. RELATED WORK

In recent years, research on how to better support the end-user experience when searching and browsing multimedia content has drawn lots of attention in the research community. A tremendous amount of work has been done in very different areas. Among the possible directions, the usage of low-level visual features for improving content-based multimedia retrieval systems has made great progress in the past ten years [4]. The drawback of content-based retrieval systems is often the lack of manually labeled data for training systems. Our approach propagates the rich semantic description of events to the media, thus contributes to semi-automatically build reliable large training dataset. In [2, 3], the authors follow a very similar approach, exploiting the rich "context" associated with social media content and applying clustering algorithms to identify social events. In contrast to our work, they do not rely on linked data technologies to realize large scale integration and reconciliation of event directories.

Tagging is popular on media sharing web sites, but tags are also as diverse as there are users. Tags might describe the visual content of media but could also simply refer to emotions of be completely personalized to a user with the

sole aim of triggering the user's memory. In [12], the authors take tags as a knowledge source and they studied the problem of inferring semantic concepts from associated noisy tags of social images. Some other work are done to improve the tag quality. In [9], Liu proposed a social image retagging approach that aims to assign better content descriptor to the social images and remove noise description. In [1], Arase et al. propose a method to detect people's trip based on their research of geo-tagged photos.

A natural extension of our work would benefit from [8]. In this paper, the authors proposed a system to present the media content from live music events, assuming a series of concerts by the same artist such as a world tour. By synchronizing the music clips with audio fingerprint and other metadata, the system gives a novel interface to organize the user-contributed content. We did not yet consider audio fingerprint for tracking down series of events but rely only on semantic metadata so far.

7. CONCLUSION AND FUTURE WORK

In this paper, we have shown how linked data technologies can be used for integrating information contained in event and media directories. We used the LODE and Media Ontology respectively for expressing linked data description of events and photos. We described a method for finding as much as possible photos and videos relevant for a given event: we start from the media that contain specific machine tags and that can be used to train classifiers that will prune results from general queries. We evaluated our approach against a manually built gold standard and we show that we are able to increase significantly the recall with a very conservative approach that does not sacrify the precision. Ultimately, we aim at providing an event-based environment for users to explore, annotate and share media and we present an initial user interface (available at `http://eventmedia.cwi.nl/demo`) that we continue to develop.

We are currently consolidating and cleaning our dataset with more sources and more linkage. We intend to provide soon user participation at events from public Foursquare check-in and live Tweets. Our priority is also to express the right licensing and attribution information to the data that has been rdf-ized. We truly believe that multimedia will then be finally added back to the Semantic Web.

Acknowledgments

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, and by the projects FP7-216444 "Peer-to-peer Tagged Media" (Petamedia).

8. REFERENCES

[1] Y. Arase, X. Xie, T. Hara, and S. Nishio. Mining People's Trips from Large Scale Geo-tagged Photos. In *18th ACM International Conference on Multimedia (ACM MM'10)*, pages 133–142, Firenze, Italy, 2010.

[2] H. Becker, M. Naaman, and L. Gravano. Event Identification in Social Media. In *12th International Workshop on the Web and Databases (WebDB'09)*, Providence, USA, 2009.

[3] H. Becker, M. Naaman, and L. Gravano. Learning Similarity Metrics for Event Identification in Social Media. In *3rd ACM International Conference on Web Search and Data Mining (WSDM'10)*, pages 291–300, New York, USA, 2010.

[4] R. Datta, D. Joshi, J. Li, James, and Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys (CSUR)*, 40, 2008.

[5] A. Fialho, R. Troncy, L. Hardman, C. Saathoff, and A. Scherp. What's on this evening? Designing User Support for Event-based Annotation and Exploration of Media. In *1st International Workshop on EVENTS - Recognising and tracking events on the Web and in real life*, pages 40–54, Athens, Greece, 2010.

[6] M. Hearst. *Search User Interfaces*. Cambridge University Press, 2009.

[7] J. Hobbs and F. Pan. Time Ontology in OWL. W3C Working Draft, 2006. `http://www.w3.org/TR/owl-time`.

[8] L. Kennedy and M. Naaman. Less talk, more rock: automated organization of community-contributed collections of concert videos. In *18th ACM International Conference on World Wide Web (WWW'09)*, pages 311–320, Madrid, Spain, 2009.

[9] D. Liu, X.-S. Hua, M. Wang, and H.-J. Zhang. Image retagging. In *18th ACM International Conference on Multimedia (ACM MM'10)*, pages 491–500, Firenze, Italy, 2010.

[10] Y. Raimond, S. Abdallah, M. Sandler, and F. Giasson. The Music Ontology. In *8th International Conference on Music Information Retrieval (ISMIR'07)*, Vienna, Austria, 2007.

[11] R. Shaw, R. Troncy, and L. Hardman. LODE: Linking Open Descriptions Of Events. In *4th Asian Semantic Web Conference (ASWC'09)*, 2009.

[12] J. Tang, S. Yan, R. Hong, G.-J. Qi, and T.-S. Chua. Inferring semantic concepts from community-contributed images and noisy tags. In *17th ACM International Conference on Multimedia (ACM MM'09)*, pages 223–232, Beijing, China, 2009.

[13] R. Troncy, A. Fialho, L. Hardman, and C. Saathoff. Experiencing Events through User-Generated Media. In *1st International Workshop on Consuming Linked Data (COLD'10)*, Shanghai, China, 2010.

[14] R. Troncy, B. Malocha, and A. Fialho. Linking Events with Media. In *6th International Conference on Semantic Systems (I-SEMANTICS'10)*, Graz, Austria, 2010.

[15] W. van Hage, V. Malaisé, G. de Vries, G. Schreiber, and M. van Someren. Combining Ship Trajectories and Semantics with the Simple Event Model (SEM). In *1st ACM International Workshop on Events in Multimedia (EiMM'09)*, Beijing, China, 2009.

[16] U. Westermann and R. Jain. Toward a Common Event Model for Multimedia Applications. *IEEE MultiMedia*, 14(1):19–29, 2007.

[17] Y.-T. Zheng, M. Zhao, Y. Song, H. Adam, U. Buddemeier, A. Bissacco, F. Brucher, T.-S. Chua, and H. Neven. Tour the World: building a web-scale landmark recognition engine. In *22nd International Conference on Computer Vision and Pattern Recognition (CVPR'09)*, Miami, Florida, USA, 2009.