Multimedia Systems (2016) 22:433–442
DOI 10.1007/s00530-014-0436-3

SPECIAL ISSUE PAPER

Linking socially contributed media with events

Xueliang Liu · Benoit Huet

Published online: 04 November 2014

Abstract Recent years have witnessed the blooming of Web 2.0 content such as Flickr and YouTube, etc. How we can benefit from such rich media resources is still an open and challenging question. In this paper, we present a method combining semantic inferencing and visual analysis for automatically finding 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.

1 Introduction

In recent years, we have witnessed the popularity of social media websites, such as Flickr, YouTube and Facebook. These social media sites provide an interactive sharing platform where huge amounts of unstructured data are uploaded every minute. How we can benefit from such rich media is still an open and challenging problem.

In this framework, some essential questions in information retrieval, such as query expanding with Natural Language Processing, retrieving multi-modal data from different Web services, and depicting result vividly, are investigated comprehensively.

Events are a natural way for referring to any observable happening grouping people at a given time and place with some intent that can be described [21]. 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 websites that provide interfaces to such functionality already exist, e.g., eventful.com, upcoming.org, last.fm, and facebook.com/events to name a few. These services have sometimes explicit connection with media sharing platforms, 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 [9].

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. The contributions of this paper are twofold:

– We investigate the underlying connections between events and social media to allow users to discover meaningful, entertaining or surprising relationships amongst

X. Liu (✉)
School of Computer and Information, Hefei University of Technology, Hefei, China
e-mail: liuxueliang@hfut.edu.cn

B. Huet
EURECOM, Sophia-Antipolis, France
e-mail: benoit.huet@eurecom.fr

them. These connections could be used as means of providing information and illustrations about events that the participants have taken, which provides a way for them to recall their life.

- 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 enriching semantically the descriptions of both events and media.
- The remaining of this paper is structured as follows: at first, we present some related work in Sect. 2. In Sect. 3.1, we present the dataset on which we will evaluate our method. We then detail our approach for associating media with events (Sect. 3). We discuss our results in Sect. 4. Finally, we give our conclusions and outline future work in Sect. 5.

2 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. It is well known that vivid photos attract humans attention more effectively than text description. In [7], to improve the users' attention when reading news articles, a system was proposed to help people reading news by illustrating the news story. The application provides mechanisms to automatically select the best illustration for each scene and to select the set of illustrations to improve the story sequence. In [10], an unsupervised approach was presented to describe stories with automatically collected pictures. In this framework, semantic keywords are extracted from the story, and used to search an annotated image database. Then a novel image ranking scheme automatically choose the most important images. In [23], a Text-to-Picture system was developed that synthesizes a picture from natural language text without limitation. The system firstly identified "picturable" text units by natural language processing, then searched for the most likely image parts conditioned on the text, and finally optimized picture layout conditioned on both the text and image parts. Besides the works that illustrate text work with photos, some studies also have been done to generate video representation from textual content. For example, in [16], a system was presented to create a visual representation for a given, short text. In this system, the authors also used some techniques to query images by the given text string, and the novelty is that the final images are selected in a user-assisted process and automatically used to create a storyboard animation. All of these approaches or systems studied how to illustrate and enrich text content with multimedia data.

In the work presented here, we defined events as the public happening taken on a given location and time [5]. The earlier relevant research focused on the study of news, since the data about news are abundant and easy to collect [4, 17]. With the popularity of social media sites, these repositories are used by users to share their experiences and interests on the Web. These sites also host substantial amounts of user-contributed materials (e.g., photographs, videos, and textual content) for a wide variety of real-world events of different type and scale. How to mine the events information has gathered recent attention. In [20], the authors studied how to employ a wavelet-based techniques to detect events from Twitter stream. A similar method can be found in [6] to detect events from Flickr time series data. In [15], the authors investigate how to filter the tweets to detect seismic activity as it happens. In [2], a system is proposed to detect emerging topics from social streams and illustrate the topics with the corresponding information in multiple modalities. Quack et al. [14] presented methods to mine events and object from community photo collections by clustering approaches. In [3], the authors follow a very similar approach, exploiting the rich "context" associated with social media content and applied clustering algorithms to identify social events. In [12], a demonstration was proposed to categorize photos by events/subevents by visual content analysis. Diakopoulos et al. [8] analyzed Twitter messages corresponding to large scale media events to improve event reasoning, visualization, and analytic. Some other research has been done to find events directly from Twitter post [15, 20]. In [18], a method is introduced to retrieve events-related photos in a collections. In [13], a framework is proposed to automatically mine and select diverse images on historical events of a landmark from Flickr.

The scheme presented in this article aims to connect user contributed images and videos to the social events they depict, by studying users' uploading behaviors on Flickr and matching concert events with photos based on different modalities; such as text/tags, time, and geo-location, and resulting in an enriched photo set which better illustrates events. A similar work was presented in [8], where a strategy was proposed to studies how to extract the valuable information from the overwhelming amount of content on social media on given broadcast news. However, their work focused on filtering noise information and outline the summarization, while illustrating events with media addresses the problem of how to leverage vivid multi-modal content to share experience. In contrast to our work, they do not rely on linked data technologies to realize large scale integration and reconciliation of event directories.

3 Find media illustrating events

3.1 The EventMedia 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. The work of Troncy et al. [19] has explored the overlap in metadata between four popular websites, namely, Flickr as a hosting website for photos and Last.fm, Eventful and Upcoming as a documentation of past and upcoming events. A large dataset called "EventMedia" is presented which is composed of events descriptions together with media descriptions associated with these events and interlinked with the larger Linked Open Data cloud. In this dataset, more than 1.7 million photos are linked by nearly 110.000 events in total (Table 1).

In this section, we consider a subset of this events dataset that corresponds to the intersection of Last.fm, Flickr and YouTube to discover meaningful, surprising or entertaining connections between events, media and people participating in events. 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 tags. Since machine tags are actually not popular in YouTube, the number of YouTube videos that actually contains such a machine tag is unsurprisingly much smaller than the number of Flickr photos. Hence, this intersection yields a dataset of 110 events, 4,790 photos and 263 videos.

The set of photos and videos available on the Web that can be explicitly associated to an event using a machine tag is generally only a tiny subset, lots of media data that are actually relevant for this event are out of the scope. Our goal is to find as many media resources as possible 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 (Fig. 1).

Table 1 Number of event/agent/location and photo/user descriptions in the dataset published in [19]

	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

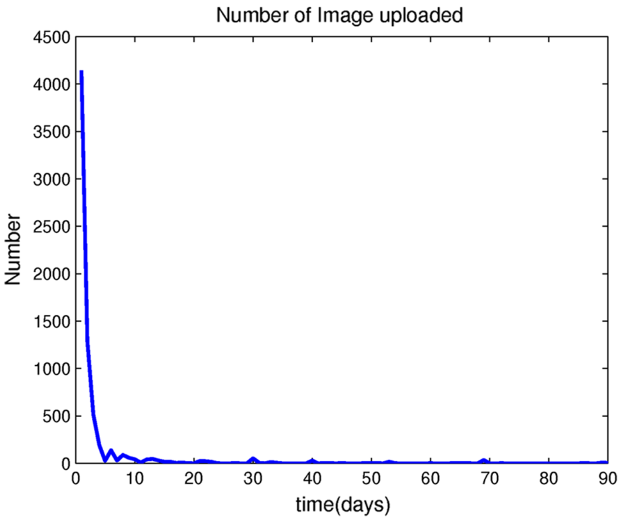


Fig. 1 Image uploading tendency along time

Starting from an event description, four dimensions 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 geo-coordinates attached to a media, the when dimension that is matched with either the taken date or the upload date of a media, and the who dimension that suggests the artists involved in the events. Querying Flickr or YouTube with just one of these dimensions brings 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 "geotag" inaccurate. In addition, we also discard the *who* because of its inconvenience to perform the media query. Actually, there are always too many artists joining an events, and nothing could be found if all of the name unionized as the query parameters. In addition, the artists likely fill the "stage name", other than his/her real name, which are either no meaning at all (for example "Yr Ods", "Yeah Yeah Yeahs"), or with misunderstood meaning(for example "Beach House" "Blue Roses"). So querying with artists names will bring more noisy media another than relevant ones. 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. It should be noticed that the query is not very specific and some irrelevant media data will be retrieved. To prune the noisy media, a visual content analysis technique is developed, which aims at removing the noise images if the visual difference is remarkable enough. Since we know that the media data labeled with machine tag is highly relevant to events and could be obtained easily, they are the best choice as the training samples for filtering noise. However in many events, only few images labeled with machine tags could be queried, and it also be found in these cases, noisy images from the query results with geotags are hardly found. Hence we use these data to build a visual model to filter the erroneous medias, as described in Sect. 3.5. The whole framework to enrich event with media data is described in Fig. 2.

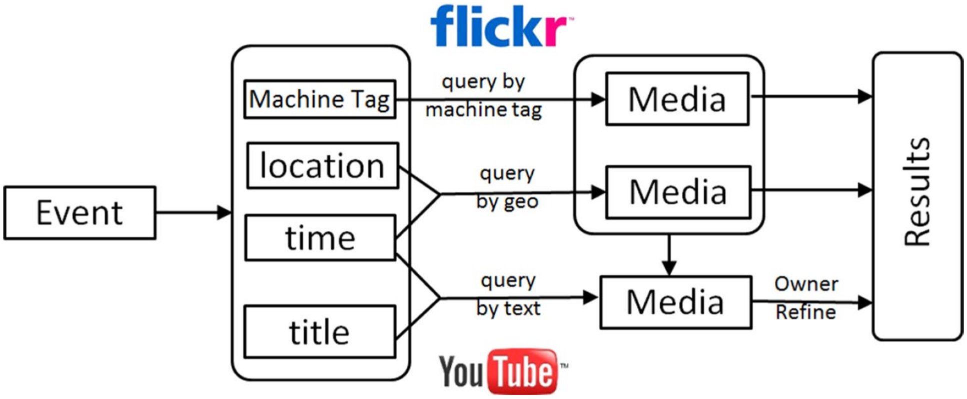


Fig. 2 The proposed framework to enrich event with photos/videos

3.2 Media context analysis

We would like to collect high quality social media data by online query with geographical, temporal, and textual parameters. How to choose the query parameters plays an important role in the process. If loose or imprecise parameters are given, many irrelevant media will be obtained and pollute the results. However, querying with parameters that are too strict will reduce the number of highly relevant media data. To make the tradeoff between quality and number, we should study the time and location trends of the media with machine tags, to infer the proper time and location window corresponding to events.

Since the media documents labeled with machine tags are taken at events, we do temporal–spatial statistics on these data to find out underlying principles. Time is one of the most key components of event, and there are more than one time measurement in events corresponding with media: event date/time, media taken date/time, media post-process time, media uploading time and so on. To find out a reasonable time window to fit our query, 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 4,790 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.[1]. 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.

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 10 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" metadata not available for videos in YouTube, we use the "upload time" instead. Figure 3 shows the results and we observe the same long tail: most the videos are uploaded within the first 5 days following an event.

Following we would like to model the venue location. The Flickr API allows to query photos based on their geographical location. Given region parameters, in the form of center and radius, or rectangle bounding box, the photos taken within a specified location can be retrieved. However, it is not so easy to obtain the geographical area covered for a place, since there are no public data for the size of a venue. We address this issue by leveraging on the event context provided by Last.fm and used by Flickr users. On a given venue (VenueID = V), all of the past events ({*eid*}) which took place there could be retrieved using the Last.fm API. Then the machine tags "lastfm:event=*eid*" is used to search for geotagged media on Flickr. Then, a bounding

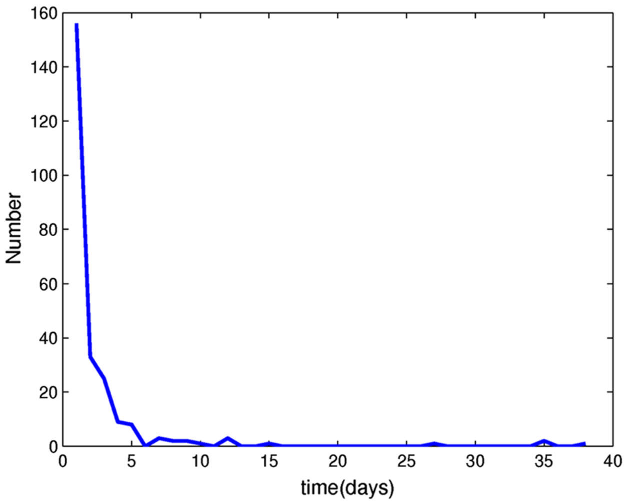


Fig. 3 Video uploading tendency along time

[1] http://en.wikipedia.org/wiki/Exchangeable_image_file_format.

box is computed using the GPS coordinates of the retrieved photos. The basic idea is to compute the bounding box with photos taken near the location, and to filter out the ones which are far from the bounding box. The final bounding box is estimated as the minimized rectangle of the GPS coordinates after removing the outliers [photos which are located further than twice the variance of the set in either direction (longitude or latitude)]. Algorithm 1 details the processing steps leading to the venue's location estimation.

Algorithm 1 Estimate the bounding box for a venue

```
INPUT: $VenueName$
OUTPUT: $BoundingBox$
$PhotoSet = [\,]$
$EventSet$=GetPastEvent($VenueName$)
for each $eventid$ in $EventSet$ do
    $photos$ = GetFlickrPhotos($eventid$, $hasGeo = True$)
    $PhotoSet$.append($photos$)
end for
$GeoSet$ = GetGeoInfo($PhotoSet$)
$GeoSet$.filter()
return MinRect($GeoSet$)
```

Figure 4 shows the result of our bounding box estimation approach for the venue Koko (London, UK). The blue marker is the GPS location of Koko according to Last.fm, and the red markers are the places where photos were taken and labeled by machine tags of past event IDs shared on Flickr. The red rectangle corresponds to the learnt bounding box for the venue Koko. We see (top part of Fig. 4) that some photos taken too far away from the venue have been appropriately discarded.

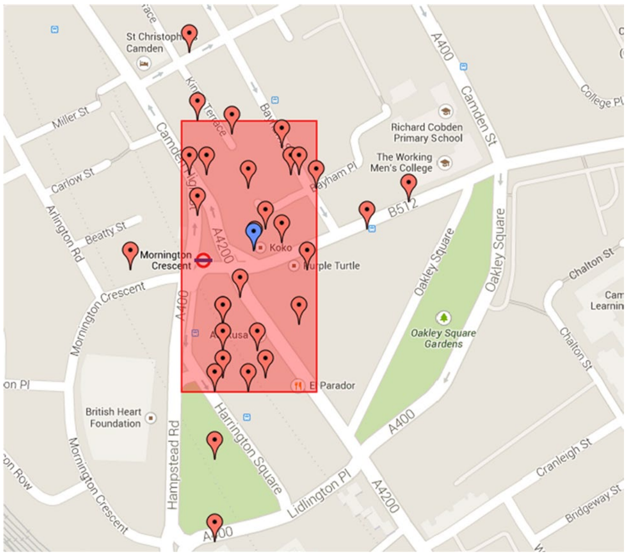


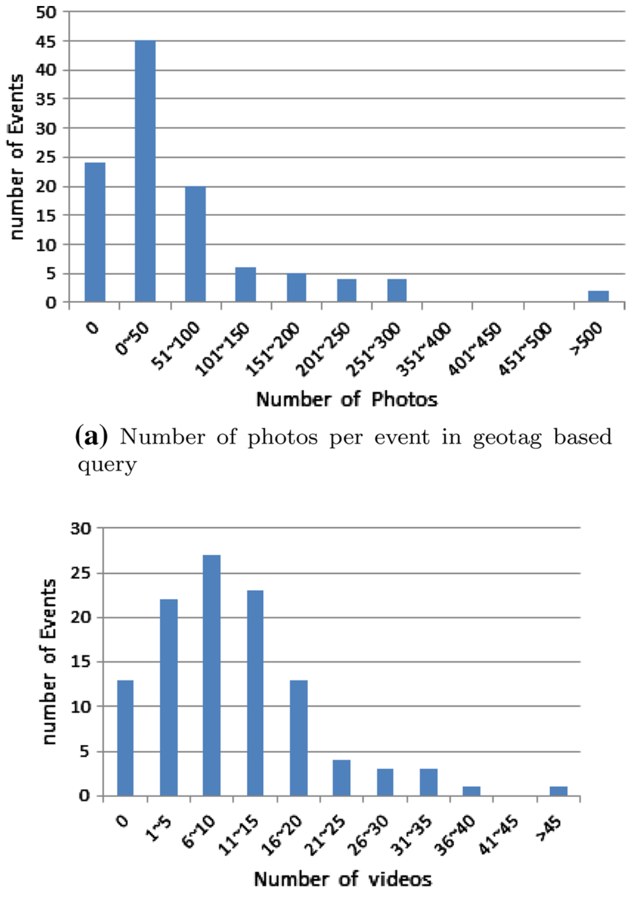
Fig. 4 *Bounding box* for the venue Koko in London (UK)

3.3 Query by geotag

Nowadays, geographical metadata is a common and key component in social media data. It could be labeled by an automatically extracting process if the media is captured by GPS-equipment devices, or be labeled manually when users sharing their media online. The metadata, named as geotags, usually are described in different format. For example, it is always composed as latitude and longitude coordinates, though it can also include altitude, bearing, distance, accuracy data, or place names. Geotags provide information to retrieve and manage media data. They are extremely valuable for application to structure the data according to location and it is also helpful for users to find a wide variety of location-specific information [1, 22]. Since we have already identified that many photos/videos are captured during events, and some of them likely are labeled with geotags indicating event taken places, these media data could be retrieved if querying with geotags parameters. 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 in our dataset, we extract the latitude and longitude information and then perform geographical based 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 videos that are geotagged is much smaller than photos. Figure 5a, b show respectively 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 bins are composed of 45 and 27 events, with about 50 photos and 11 videos retrieved respectively.

3.4 Query by title

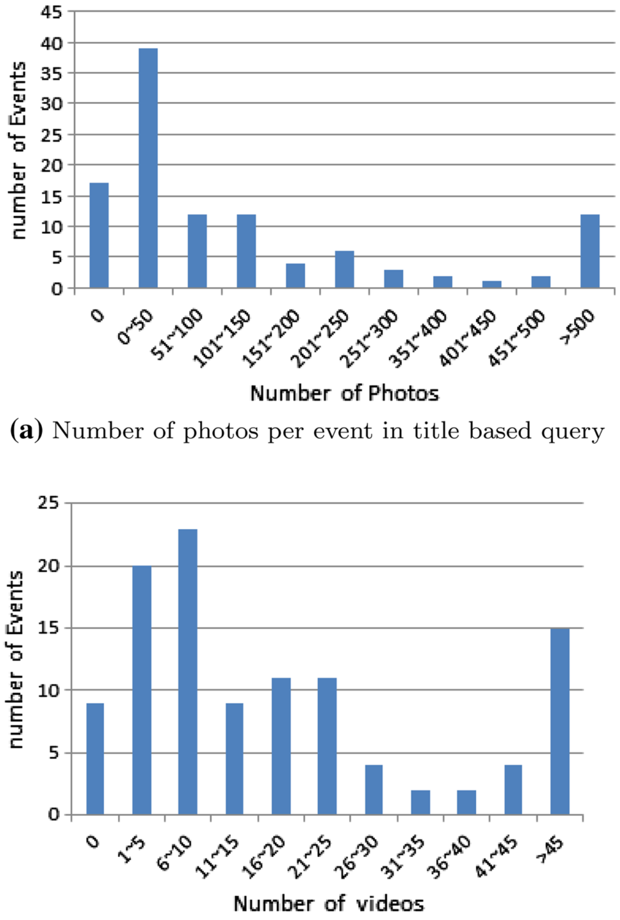
The title is the most descriptive and readable information for events. Similarly to geotagged queries, we perform full text search queries on Flickr and YouTube based on the event titles that are extracted. The retrieved photos and videos are also filtered using a time interval of five days following the event taken time. When performing search query using the Flickr API query, we use the "text mode" rather than the "tag mode" since the latter is more strict and many photos will miss. The number of photos retrieved at this stage is however in an order of magnitude greater than with geotagged queries. Due to the well-known polysemy problems of textual-based query, the title-based query brings lots of irrelevant photos. We describe in Sect. 3.5 an heuristic for filtering out irrelevant media. In contrast, we do not observe this noise when querying the YouTube API



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

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

Fig. 5 Statistics for geotag-based query



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

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

Fig. 6 Statistics for title based query

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 them are relevant. The distribution of the number of retrieved photos and videos for the 110 events in our dataset is depicted in Fig. 6a, b. 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 Fig. 5, we can clearly see that the standard deviation of Fig. 6 is larger and that again photos are more readily available than videos.

3.5 Pruning irrelevant media

Images and videos with specific machine tags such as `lastfm:event=207358` can be unconditionally associated with events. We consider that media retrieved with geotag queries during a correct time frame should also be relevant for those events. In other words we consider that both GPS and time information are accurate. 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 noisy data and to avoid incorrectly propagating rich event descriptions to these media documents, we propose a method for pruning the set of candidates photos using visual content based analysis. The photos captured at a single event are 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 relevant ones. We address this issue in two steps to ensure high precision and recall ratio.

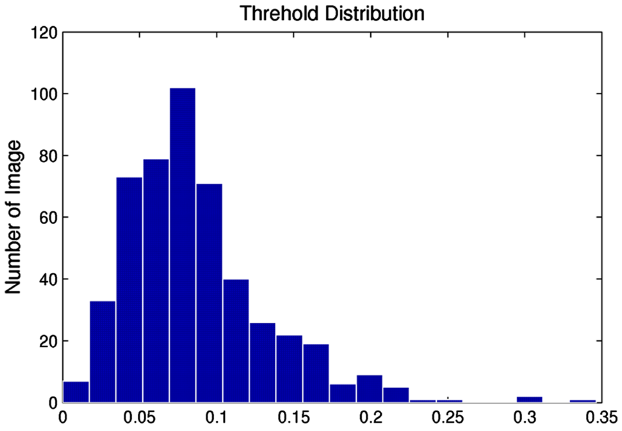


Fig. 7 The distribution of threshold

The main idea of our algorithm is to measure the visual similarity of media documents, that is, to learn a threshold from training set and used to filter noisy data in testing data. First, we build a training dataset composed of the media containing either the event machine tag or a combination of geo-coordinates and time frames corresponding to the event dimensions. The photos resulting from query by title compose the testing dataset. The visual features used in our approach are 225D color moments in Lab space, 64D Gabor texture, and 73D Edge histogram. For each image pairs in the training data, the nearest neighbors algorithm using the *L1* distance measure in the training set is performed and the smallest distance is taken as threshold. Second, images originating from the title query are composed of training images. Images for which the distance to images in the test set is below the threshold are candidates for illustrating the event. Mathematically, let E as the training photos set, and F as the testing photos set. The objective is to select the photos from F which are similar to the photos in E, to additionally enrich the set E illustrating an event. The visual similarity between two images is computed as follows:

$$L_1(F_j, E_i) = \sum_k |F_j(k) - E_i(k)| \tag{1}$$

where $F_j(k)$ and $E_i(k)$ are normalized concatenating low level feature vector of the images. F_j is added to the set of media illustrating the event when

$$\exists E_i \in E : L_1(F_j, E_i) < \text{THD}_i$$

where THD_i is the threshold which is also learned from the E data. As shown in Eq. 2, we use a strict strategy to decide the threshold, which is chosen as the minimal value of similarity of images pairs in training set. And the threshold is also adaptive to different events because of the visual diversity within the training dataset. In order to remove noisy images in the testing data, the threshold should be adjusted respectively. Figure 7 shows the value of threshold used in the experiments which range from 0.01 to 0.346.

$$\text{THD}_i = \min_{\{j\}\setminus i} \sum_k |E_j(k) - E_i(k)| \tag{2}$$

The algorithm can be formalized as follows in Algorithm 2:

Algorithm 2 Pruning 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($imgt$,$img$) $\le$ $Threshold$ then
      $PrunedSet$.append($imgt$)
    end if
  end for
end for
return $PrunedSet$
```

In visual pruning, in order to filter out most of the irrelevant photos, a strict threshold strategy is applied, and some relevant ones are also be discarded, which leads to a lower recall ratio. In order to recover these photos and improve the recall ratio, we exploit the "owner" property in social media and proposed a refinement method. We assume that a person cannot attend more than one event simultaneously. Therefore, all the photos that have been taken by the same owner during the event duration should be assigned to the event. So if the owner has shared more media, capture during this period, they are automatically added as illustrative media for the event. Using this heuristic, it is possible to retrieve photos which do not have any textual/geographical description. As far as we know, "owner refinement" is the only effective approach to match events and media data when not enough metadata (such as textual, graphical metadata) is available.

4 Results

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 are effective to retrieve an order of magnitude more media than using solely machine tags. Hence, while 4,790 photos are tagged with the `lastfm:event=xxx` machine tag, 6,933 photos can be retrieved using the geo-location of the event and 32,583 photos can be retrieved using 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	4,790	32,583	6,933	2,350	494	484	405
Videos	263	4,237	1,163	103	39	115	29

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	Precision
346054	2	24	2	1	1	0.5	1	0.5	1
158744	3	48	48	23	1	0.479	44	0.917	0.977
371981	4	16	6	4	1	0.667	4	0.667	1
341832	7	0	0	0	1	1	0	1	1
362195	7	0	0	0	1	1	0	1	1
235445	10	1	1	0	1	0	0	0	1
42644	13	85	81	13	1	0.16	13	0.16	1
165697	23	1	1	0	1	0	1	1	1
137530	24	9	4	0	1	0	1	0.25	1
517159	24	0	0	0	1	1	0	1	1
222241	36	204	180	33	0.97	0.183	72	0.4	0.944
234649	45	35	4	1	1	0.25	1	0.25	1
207358	54	68	4	4	1	1	4	1	1
429517	60	171	169	27	1	0.16	41	0.243	0.929
437747	65	144	142	8	1	0.056	13	0.092	0.952
117886	68	99	97	4	1	0.041	11	0.113	1
150390	71	16	16	1	1	0.063	1	0.063	1
350591	79	85	85	6	1	0.071	66	0.776	0.91
472733	93	500	478	8	1	0.017	18	0.038	1
176257	97	260	255	47	1	0.184	147	0.576	0.952
Summary	785	1,766	1,573	180	0.998	0.114	438	0.278	0.950

title. After removing the duplicated ones, we obtain 36,412 photos that are candidates 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 tags.

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 1,766 photos in the testing set (photos retrieved by event title). We build manually the ground truth for those 1,766 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 1,573 images are used as ground truth dataset.

The results of the pruning algorithm detailed in Sect. 3.5 applied to the 1,766 photos shown in 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, besides many relevant photos. In order to increase the recall ratio, we extend the resulting images by our pruning algorithm with all the ones uploaded by the same uploader. The reason is that if one photo can reliably be attached to an event, we infer that this participant indeed attended the event and that all the others photos taken by this person during this time frame

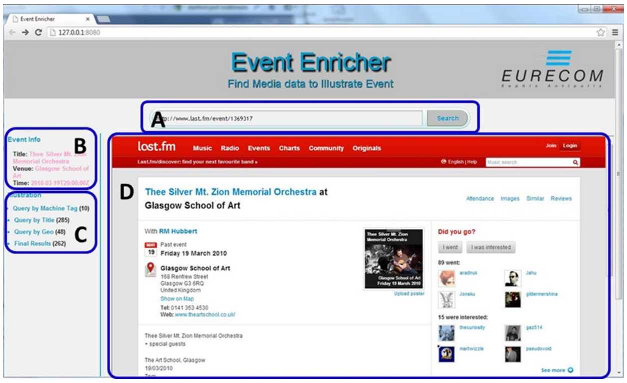

Event Enricher
127.0.0.1:8080
Event Enricher
Find Media data to Illustrate Event
EURECOM
A
tp://www.last.fm/event/1369317
Search
Event Info
B
Title: Thee Silver Mt. Zion Memorial Orchestra
Venue: Glasgow School of Art
Time: 2010-03-19T20:00:00Z
Illustration
Query by Machine Tag (10)
Query by Title (285)
Query by Geo (48)
Final Results (262)
C
D
last.fm
Music
Radio
Events
Charts
Community
Originals
Thee Silver Mt. Zion Memorial Orchestra at Glasgow School of Art
With RM Hubbert
Friday 19 March 2010
Glasgow School of Art
Did you go?

Fig. 8 EventEnricher interface: (*A*) Input URL; (*B*) Event Abstract; (*C*) Navigation of the Results; (*D*) MainView: to show the event homepage and photos in the results

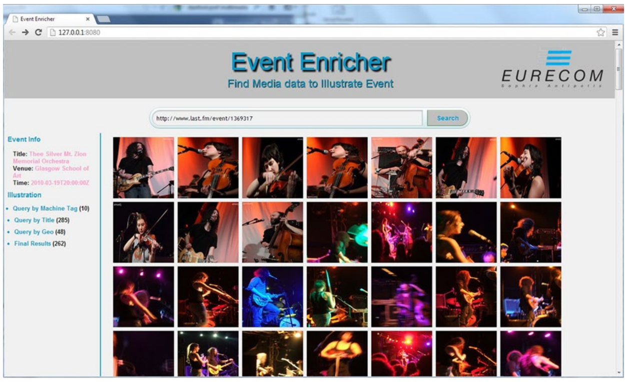


Fig. 9 Results on Lastfm event:1369317

are likely to be illustrative media for this event. This simple heuristic allows to significantly improve the recall ratio (from 0.114 to 0.278) without sacrificing to the precision.

4.1 Demonstration

Based on our proposal, a Web service is built to help user browse media data from events.[2]. Flickr is used as the basic media sharing platform, although EventEnricher can easily be extended to cater for other source such as YouTube, Google Picasa, etc. As shown in Fig. 8, the Web service named as EventEnricher and developed with Python + web.py. On a given event URI defined in EventMedia, the demo showcases the enriching results with several tables. In details, with the event URI as the query parameter, the service firstly query the event information on EventMedia dataset, and parse the event context such as event title, taken time, taken place. Then the approaches presented in this section is employed to query media data in Flickr. After the pruning and refinement process, the final data is presented in four tables named "query by machine tag", "query by title", "query by geo", and "final results" in a new Web page, as shown in Fig. 8(part C).

The users can interact with the system through four parts, as shown in Fig. 8. Part (A) is the input parameter, while an event URL in last.fm, DailyMotion, Upcoming, or URI in EventMedia dataset [19] could be the input to query the event. When event information is retrieved, the abstract is presented in part (B), and the home page of the event is depicted in part (D). Then, the event's machine tag, title, geo location, time metadata are extracted and used to query the photos in Flickr, as described in [11]. The results from the query, as well as from the visual pruning and owner refinement process, can be accessed by the list in part (C). With a mouse click, the photos are presented in part (D), as shown in Fig. 9.

Figure 9 also shows the effectiveness of our system on collecting event relevant photos. For last.fm event (ID = 1369317), only 10 photos are labeled with machine tag, and 285 and 48 photos are retrieved by the location and title based query. After the visual pruning and owner refinement process, a set of 262 photos illustrates the event.

4.2 Discussion

Event directories are largely overlapping, providing multiple identifiers for the same venues, artists, and events. We argued that linked data technology helps 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 to 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 Sect. 3.1, 1,248,021 photos (that is 73 %) have been geotagged for free since Flickr had no geotagged 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.

5 Conclusion

In this paper, we have shown how linked data technologies can be used for integrating information contained in event

[2] The code is available at: https://github.com/MediaAnalysis/EventEnricher.

[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/.

and media directories. 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 scarify the precision. Ultimately, we provide an event-based interface to explore shared media.

Acknowledgments This work was supported by the National Basic Research Program of China (No. 2013CB329604), the European Commission under contracts FP7-287911 LinkedTV and FP7-318101 MediaMixer.

References

1. Arase, Y., Xie, X., Hara, T., Nishio, S.: Mining people's trips from large scale geo-tagged photos. In: 18th ACM International Conference on Multimedia, pp. 133–142. Firenze, Italy (2010)
2. Bao, B.-K., Min, W., Sang, J., Xu, C.: Multimedia news digger on emerging topics from social streams. In: Proceedings of the 20th ACM International Conference on Multimedia, MM '12, pp. 1357–1358 (2012)
3. Becker, H., Naaman, M., Gravano, L.: Event identification in social media. In: 12th International Workshop on the Web and Databases. Providence, USA (2009)
4. Billsus, D., Pazzani, M.J.: A hybrid user model for news story classification. In: Proceedings of the Seventh International Conference on User Modeling, pp. 99–108 (1999)
5. Chen, K.-Y., Luesukprasert, L., Chou, S.-C.T.: Hot topic extraction based on timeline analysis and multidimensional sentence modeling. IEEE Trans. Knowl. Data Eng. **19**(8), 1016–1025 (2007)
6. Chen, L., Roy, A.: Event detection from flickr data through wavelet-based spatial analysis. In: ACM Conference on CIKM (2009)
7. Delgado, D., Magalhães, J.A., Correia, N.: Assisted news reading with automated illustrations. In: ACM Conference on Multimedia, pp. 1647–1650 (2010)
8. Diakopoulos, N., Naaman, M., Kivran-Swaine, F.: Diamonds in the rough: social media visual analytics for journalistic inquiry. In: 2010 IEEE Symposium on Visual Analytics Science and Technology, pp. 115–122 (2010)
9. Fialho, A., Troncy, R., Hardman, L., Saathoff, C., Scherp, A.: 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, pp. 40–54. Athens, Greece (2010)
10. Joshi, D., Wang, J.Z., Li, J.: The story picturing engine—a system for automatic text illustration. ACM Trans. Multimed. Comput. Commun. Appl. **2**(1), 68–89 (2006)
11. Liu, X., Troncy, R., Huet, B.: Finding media illustrating events. In: 1st ACM International Conference on Multimedia Retrieval. Trento, Italy (2011)
12. Mattivi, R., Uijlings, J., De Natale, F., Sebe, N.: Categorization of a collection of pictures into structured events. In: Proceedings of the 2nd ACM International Conference on Multimedia Retrieval, p. 1. New York, USA (2012)
13. Min, W., Bao, B.-K., Xu, C.: What happened near big ben: event-driven landmark mining from flickr. In: Proceedings of the 13th Pacific-Rim Conference on Advances in Multimedia Information Processing, PCM'12, pp. 769–778. Springer, Berlin (2012)
14. Quack, T., Leibe, B., Van Gool, L.: World-scale mining of objects and events from community photo collections. In: Proceedings of the 2008 International Conference on Content-Based Image and Video Retrieval, p. 47. New York, USA (2008)
15. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes Twitter users: real-time event detection by social sensors. In: International Conference on WWW, pp. 851–860 (2010)
16. Schwarz, K., Rojtberg, P., Caspar, J., Gurevych, I., Goesele, M., Lensch, H.P.A.: Text-to-Video: story illustration from online photo collections. Knowl. Based Intell. Inf. Eng. Syst. **6279**, 402–409 (2010)
17. Toda, H., Kataoka, R.: A clustering method for news articles retrieval system. In: Special Interest Tracks and Posters of the 14th International Conference on World Wide Web, p. 988. New York, USA (2005)
18. Trad, M.R., Joly, A., Boujemaa, N.: Large scale visual-based event matching. In: Proceedings of the 1st ACM International Conference on Multimedia Retrieval, pp. 1–7. New York, USA (2011)
19. Troncy, R., Malocha, B., Fialho, A.: Linking events with media. In: 6th International Conference on Semantic Systems. Graz, Austria (2010)
20. Weng, J., Lee, B.-S.: Event eetection in Twitter. In: Fifth International AAAI Conference on Weblogs and Social Media, pp. 401–408. HP Laboratories (2011)
21. Westermann, U., Jain, R.: Toward a common event model for multimedia applications. IEEE MultiMed. **14**(1), 19–29 (2007)
22. Zheng, Y.-T., Zhao, M., Song, Y., Adam, H., Buddemeier, U., Bissacco, A., Brucher, F., Chua, T.-S., Neven, H.: Tour the world: building a web-scale landmark recognition engine. In: 22nd International Conference on Computer Vision and Pattern Recognition. Miami, Florida, USA (2009)
23. Zhu, X., Goldberg, A.B., Eldawy, M., Dyer, C.R., Strock, B.: A text-to-picture synthesis system for augmenting communication. In: Proceedings of the 22nd National Conference on Artificial Intelligence, vol. 2, pp. 1590–1595 (2007)