AAL-2009-2-049, ALIAS
D4.6
Cross-media linking of personal events to web content

Due Date of Deliverable	2012-12-31
Actual Submission Date	2013-02-22
Workpackage:	4
Dissemination Level:	Public
Nature:	Report
Approval Status:	Final
Version:	v2.0
Total Number of Pages:	36
Filename:	D4.6-EURECOM-Cross-media-linking.pdf
Keyword list:	

Abstract

History

Version	Date	Reason	RevisedBy
0.1	2012-11-30	1st complete draft	X. Liu
0.2	2012-12-15	2sd complete draft	X. Liu
0.3	2012-12-17	1st final draft	B. Huet
0.4	2013-01-18	Final version addressing comments from first internal review	X. Liu
0.4	2013-02-22	Final version addressing comments from second internal review	B. Huet

Authors

Partner	Name	Phone / Fax / Email
EURECOM	Xueliang Liu	Tel: Fax: Email: Xueliang.liu@eurecom.fr
EURECOM	Raphael Troncy	Tel: +33 4 9300 8242 Fax: +33 4 9300 8200 Email: raphael.troncy@eurecom.fr
EURECOM	Benoit Huet	Tel: +33 4 9300 8179 Fax: +33 4 9300 8200 Email: benoit.huet@eurecom.fr

Table of Contents

1 Summary

The aim of the ALIAS project is to build a system within a mobile robot to assist elderly people in their daily life. In this deliverable, we report our work on the entertainment part of the computer system that will help users enjoy their retired life and feel closer to their family and friends. In details, we propose a framework to support users to choose their social events of interest based on a semantic web dataset. A method for finding media hosted on Flickr that can be associated to an event is presented. A web service, "Event Enricher" is developed to provide users a rich and vivid set media to browse about events . In addition, to remove the noisy media from the inaccurately retrieved during the query process, a novel framework is proposed to model the visual appearance of social events by automatically collected training samples. The visual training samples are collected through the analysis of the spatial and temporal context of media data and events. The resulting event models are effective to filter out irrelevant photos and perform enrichment with a high accuracy as demonstrated on various social events originating for various categories of events. The automated process proposed allows elderly users to transparently browse media related to specific events originating from various web sources straight on the alias platform.

2 Introduction

The aim of the ALIAS project is to build a system within a mobile robot to assist elderly people in their daily life. Besides the expected functionalities of the robot such as intelligent positioning, user interaction, day-to-day cognitive assistance, the entertainment part proposed and described in this documents will help users enjoy their retired life and feel closer to their family and friends. In details, we propose a framework to support users to choose their social events of interest based on the semantic web dataset, EVENTMEDIA[1]. Events are a natural way for referring to any observable occurrence grouping persons, places, times and activities. They are also observable experiences that are often documented by people through different media (e.g. videos and photos). In this documents, a method for finding media hosted on Flickr that can be associated to a public event is presented. It will show the benefits of using linked data technologies for enriching semantically the descriptions of both events and media, so that people can search through content using a familiar event perspective. A web service, "Event Enricher" is developed to help the users browse the event and media content. In addition, to remove the noisy media from the inaccurate query, a novel framework is proposed to model the visual appearance of social events by automatically collected training samples. The visual training samples are collected through the analysis of the spatial and temporal context of media data and events. While collecting positive samples can be achieved easily thanks to dedicated event machine-tags, finding the most representative negative samples from the vast amount of irrelevant multimedia documents is a more challenging task. Here, we argue and demonstrate that the most common negative samples, originating from the same location as the event to be modeled, are best suited for the task. A novel ranking approach is devised to automatically select a set of negative samples. Finally the automatically collected samples are used to learn visual event models using Support Vector Machine (SVM). The resulting event models are effective to filter out irrelevant photos and perform with a high accuracy as demonstrated on various social events originating for various categories of events.

It is worth pointing out that the current Event Enricher engine is only using data from Last.fm/Upcoming/Eventful for identifying an event and Flickr/YouTube for searching for relevant media originating from an event. This deliverable provides with great details the process that takes place when the elderly is asking for information about an event, in order to mine the web for associated media documents. The user (elderly or not) will not access any of those web-sites nor web-services. It is totally transparent to him/her while browsing the EventMedia interface on the Alias robot.

This document describes the proposed approaches to mine the relationship in two aspects, which will support users browsing, querying the past social events, as well making their

[1] http://eventmedia.eurecom.fr

decision on upcoming events.

3 Related Work

The study of events has been addressed in the computer vision community for many years [1]. In computer vision, the objective of event related research concerns essentially the recognition and eventually the localization of special spatial-temporal patterns from a large collection of image sequence or video streams. This is a common yet challenging topic tackled by computer vision/video surveillance scenarios [3] which focus essentially on detecting abnormal or specific behaviors or activities. However, the concept of event addressed in this document is drastically different compared with these works. Here, we define an event as a real life social happening, involving a group of person and occurring at a specific date (or time) and in a specific location. A live concert held in a club on a given night, an international scientific conference or a carnival (lively and animated street celebration) are among the types of events investigated in the work presented in this document.

In the past few years, the study of new methods for organizing, searching and browsing media according to real-life events has drawn lots of attention in the multimedia research community. Much work has been done in very different areas. The methods found in the literature addressing this issue cover many multi-modal processing techniques. Therefore, we address the related work from a number of relevant research directions, including: event illustration by media documents; event detection from social media data; multimedia data tags analysis; as well as content based media analysis.

Illustrating events with media data studies the problem of how to leverage vivid visual content to represent events. In [12], the authors proposed a framework to generate photos collections of news to enhance user's experience while reading news articles. They computed the similarity between news text and image tags and obtained the relevant images using text retrieval techniques. In [25], an approach aimed at creating a vivid visual experience to users browsing public events, such as concerts or live shows, was proposed. They studied the user uploading behaviors on Flickr and YouTube, and matched events with medias based on different modalities, such as text/tags, time, and geo-location. The results is an enriched media set which better illustrates the event. In [19], 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 gave a novel interface to organize the user-contributed content.

The study of "how to detect events?" has also gained a lot of attention in the past years. The objective of event detection is to discover events out by sensing what is occurring at given location and time. To address the problem, Quack *et al.* [30] presented methods to mine events and object from community photo collections. They clustered the photos

with multi-modal features and then classified the results into events and objects. A similar problem was also studied in [15] where Firan *et al.* focused on building a Naive Bayes event models which classify photos as either relevant or irrelevant to given events. In [5, 6], the authors followed a very similar approach, exploiting the rich "context" associated with social media content and applying clustering algorithms to identify social events.

Tagging is popular on media sharing web sites, and such additional information can be extremely valuable for identifying/representing the content the associated media. However, tags can be very diverse in nature. They might describe the visual content of media but can also refer to emotions, or be personalized for a user (or the media owner himself) with the sole aim of triggering his memory or to attract other users' attention. In [37], the authors took tags as a knowledge source and studied the problem of inferring semantic concepts from associated noisy tags of social images. Some other works are done to improve the tag quality. In [22], Liu *et al.* proposed a social image re-tagging approach that aims to assign better content descriptor to the social images and remove noise description. In [2], Arase *et al.* propose a method to detect people's trip based on their research of geo-tagged photos.

Much of the previous approaches aimed at mining the intrinsic connection between events and media are performed by metadata analysis (i.e. time, location, owner, tags, etc...). Only little work has been done on the analysis the visual content of medias in the context of event, and this is precisely the issue we address with this document.

The usage of low-level visual features for improving content-based multimedia retrieval systems has made great progress [11]. To address the problem of web visual data analysis, some large scale datasets have been built using multimedia data crawler from shared portals [10]. Beside those web datasets, a number of learning techniques performed on these datasets have shown acceptable results [42, 18]. Many works [20, 35, 21] have been done to study how to automatically or semi-automatically collect online data for training purposes. In [20], Li *et al.* proposed their work on how to train visual concept model by data collected from Internet automatically. The proposed OPTIMOL model employs a Hierarchical Dirichlet process to learn visual concepts and to make the decision rule on new images. An improved work is reported in [35], where the authors employed text, meta-data and visual information in order to achieve better performance. In [21], the authors tackle the problem of collecting negative training samples to model concepts automatically. This objective is somehow similar to the one addressed in this document. However, their solution exploits the semantics between different visual concepts using related tags. In our work, the objective is to associate media with missing or inaccurate metadata to their corresponding social events. Clearly, the method proposed in [21] cannot be employed to solve our problem, since we cannot define the related and unrelated tags for each event as required by their approach. Our solution leverages the rich contextual information surrounding and defining events to automatically build the collection of online media samples, using an approach inspired from ranking techniques, and training the classifiers individually for each specific event.

4 Event Enrichment by Social Media data

Organizing media data according to events in the real world is the natural way for human to recall his experience. Exploiting event context to solve the management and retrieval problems raising from the social media draws lots of interest in the multimedia community. In this chapter, we present our work to infer the semantics behind the events and explore social media to illustrate events. With the study of users uploading behavior, we extend the set of illustrating images and videos for a particular event by querying social media with diverse and multi-modality features, and pruning the results with content based visual analysis.

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 section, we present a method for automatically finding medias hosted on Flickr and YouTube that can be associated to public events. We show the benefits of using linked data technologies for enriching semantically the descriptions of both events and media data.

4.1 LODE Ontology 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, user study has been carried out 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 [14]. We argue that linked data technology is suitable for doing this integration at large scale given they naturally based on URIs for identifying objects and a simple triple model (RDF) for representing semantic descriptions. In this section, we revise the LODE event model and describe the techniques to populate this ontology by scraping three large event directories: last.fm, eventful and upcoming.

The LODE ontology[1] is a minimal model that encapsulates the most useful properties for describing events [36]. The goal of this ontology is to enable inter-operable modeling

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

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* were involved. LODE is not yet another "event" ontology *per se*. It has been designed as an *interlingua* model that solves an interpretable problem by providing a set of axioms expressing mappings among existing event ontologies. Hence, the ontology contains numerous OWL axioms stating classes and properties equivalence between models such as the Event Ontology [31], CIDOC-CRM, DOLCE, SEM [40] to name a few.

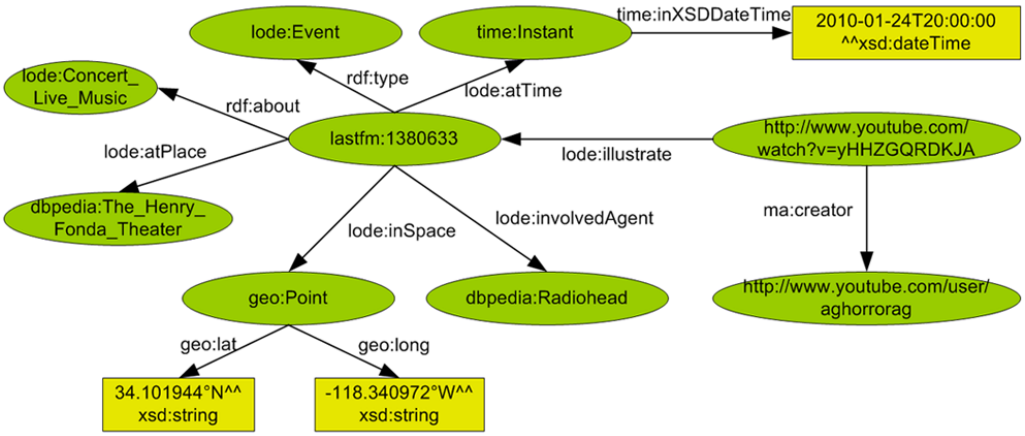


Figure 4.1: The *Radiohead Haiti Relief Concert* described with LODE (*top*) and illustrated with media described by the Media Ontology (*bottom*)

Figure 4.1 depicts the metadata attached to the event identified by id=`1380633` on Last.fm according to the LODE ontology. More precisely, it indicates that an event categorized as a `Concert` has been given on the `24th of January 2010 at 20:00 PM` in the `Henry Fonda Theater` featuring the `Radiohead` rock band. 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. Hence, we see that the video hosted on YouTube has for `ma:creator` the user `aghorrorag`. We use the Last.fm, Eventful and Upcoming APIs to query the online events and then 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 representation of 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 [17]), 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).

4.2 Find Media Illustrating Events

The set of photos and videos available on the web that can be explicitly associated to an event using a machine tag is generally 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 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.

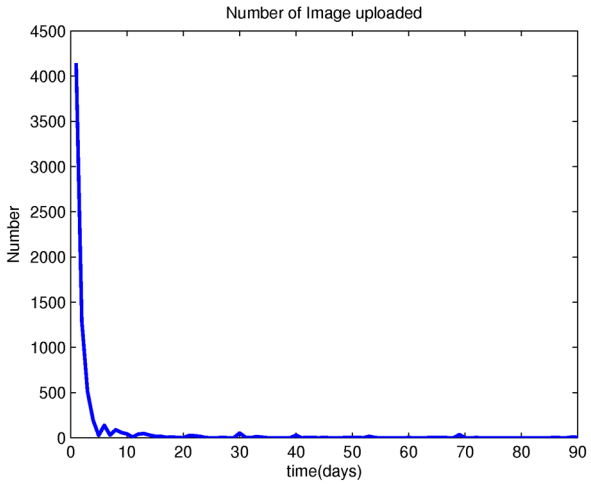


Figure 4.2: Image uploading tendency along time

Starting from an event description, four 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 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 "geo tag" 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 geo tags are hardly found. Hence we use these data to build a visual model to filter the erroneous medias, as described in Chapter 5. The whole framework to enrich event with media data is described in Figure 4.3.

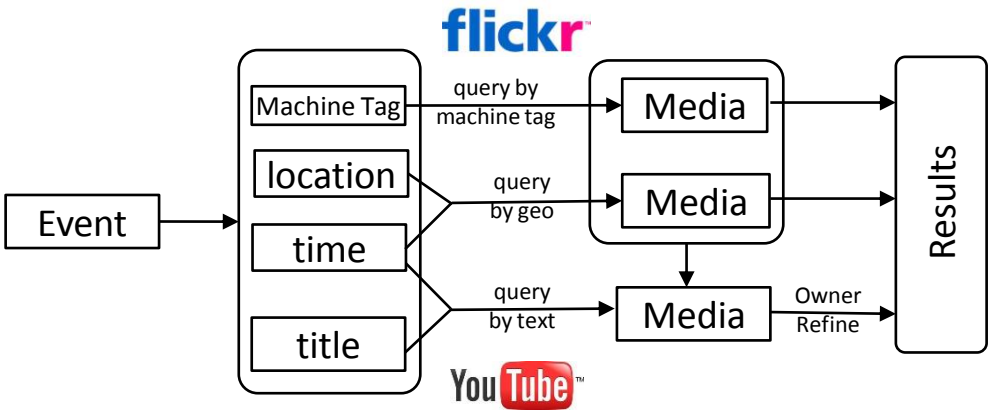


Figure 4.3: The proposed framework to enrich event with photos/videos

4.3 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 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 taken time, media taken 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 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[2]. Figure 4.2 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.

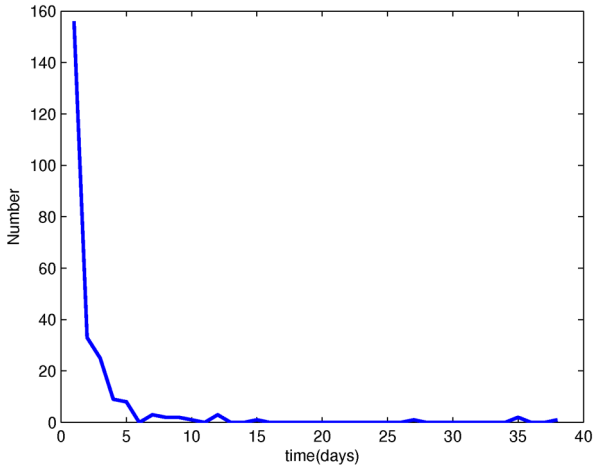


Figure 4.4: Video uploading tendency along time

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" metadata not available for videos in YouTube, we use the"upload time" instead. Figure 4.4 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

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

to search for geo-tagged media on Flickr. Following a bounding 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 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.

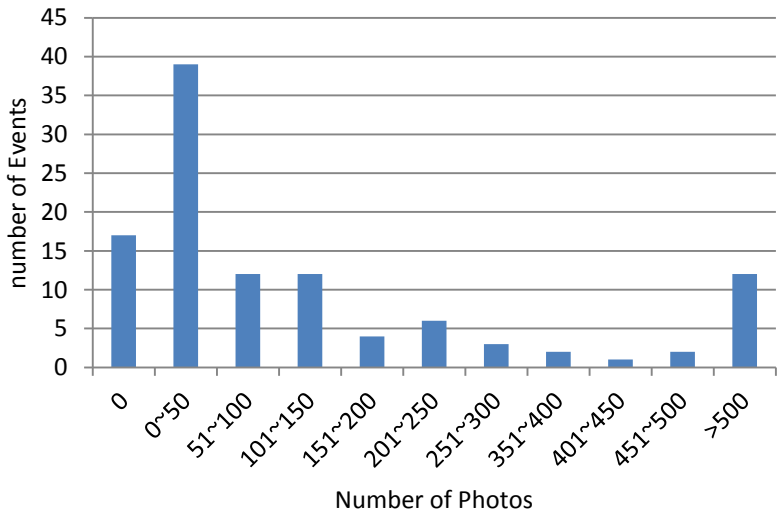
Algorithm 1 Estimate the bounding box for a venue

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($GeoSet$)

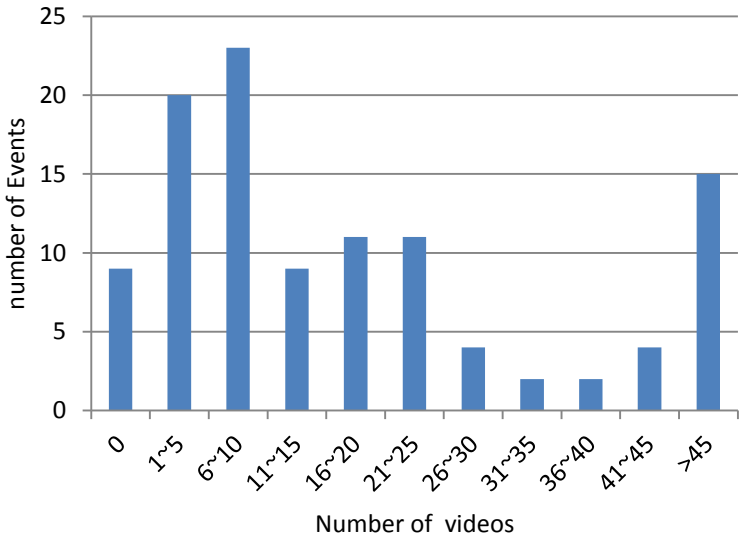
4.4 Query by Geotag

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 [2]. Since we have already known 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 from the LODE descriptions 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. Figures 4.6(a) and 4.6(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.



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



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

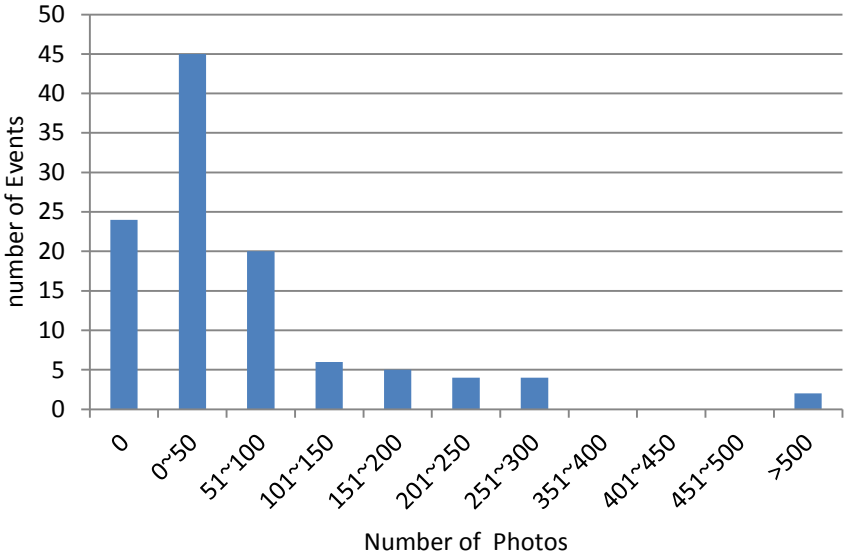
Figure 4.5: Statistics for title based query

4.5 Query by Title

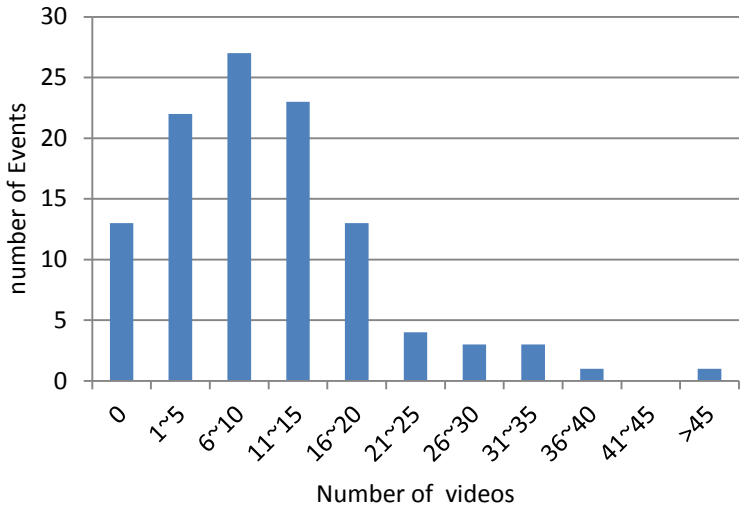
Title is the most describable and readable information for events. Similarly to geo-tagged queries, we perform full text search queries on Flickr and YouTube based on the event titles that are extracted from the LODE description. The retrieved photos and videos are also filtered using a time interval 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 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 Chapter 5 an heuristic for filtering out 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 them are relevant. The distribution of the number of retrieved photos and videos for the 110 events in our dataset is depicted in Figures 4.5a and 4.5b. 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 4.6, we can clearly see that the standard deviation of Figure 4.5 is larger and that again photos are more readily available than videos.

4.6 Event Enricher

Finally, to incorporate the work present in this section, we have developed an online demo to search and browse media illustrating events. The web service is named as EventEnricher and the background service is developed with Python + web.py, while the foreground is developed by HTML + JQuery, as shown in Figure 4.7. On a given event URI defined in EventMedia or an event URL in Last.FM, Upcoming or Eventful, the screenshots show the enriching results in several web pages. In details, with the event URL as the query parameter, the service starts by issuing a query about the event information on EventMedia dataset, and parse the event context such as event title, taken time, taken place, and its original page (as shown in Figure 4.7). Then, the approach presented in this chapter is employed to query and display media data according to machine tag (Figure 4.8), title + time(Figure 4.9), location + time (Figure 4.9) from the Flickr photo sharing platform. After both pruning and refinement processes have operated, the final media data illustrating the event is shown (see Figure 4.11).



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



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

Figure 4.6: Statistics for geotag based query

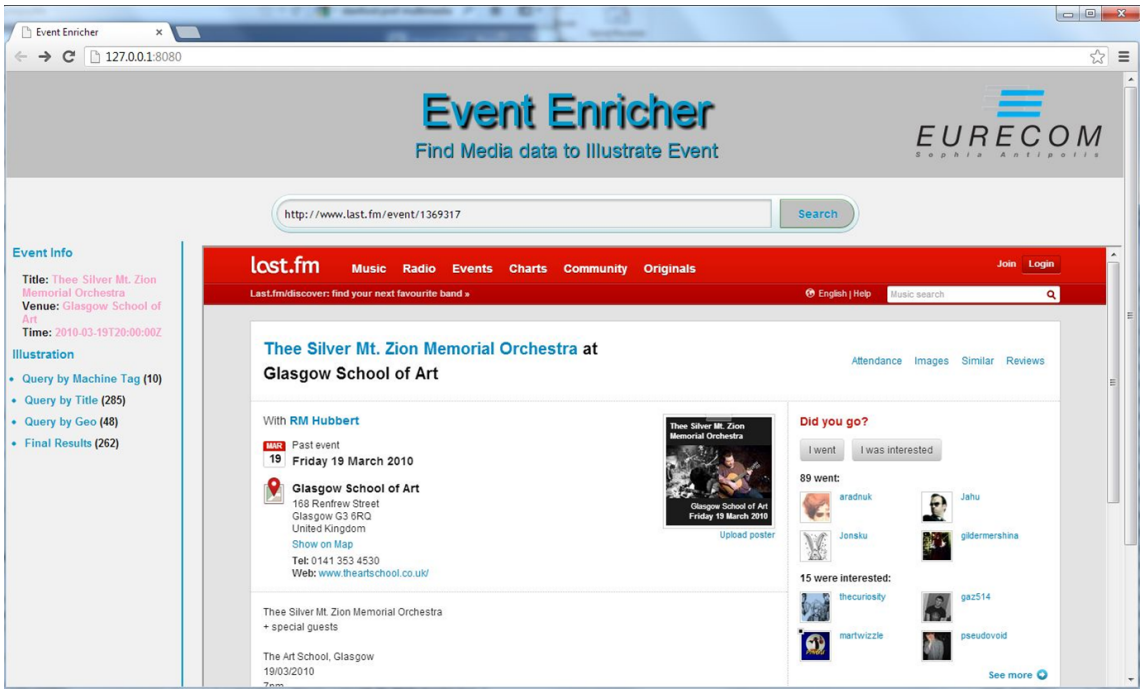


Figure 4.7: The Event Enricher Interface

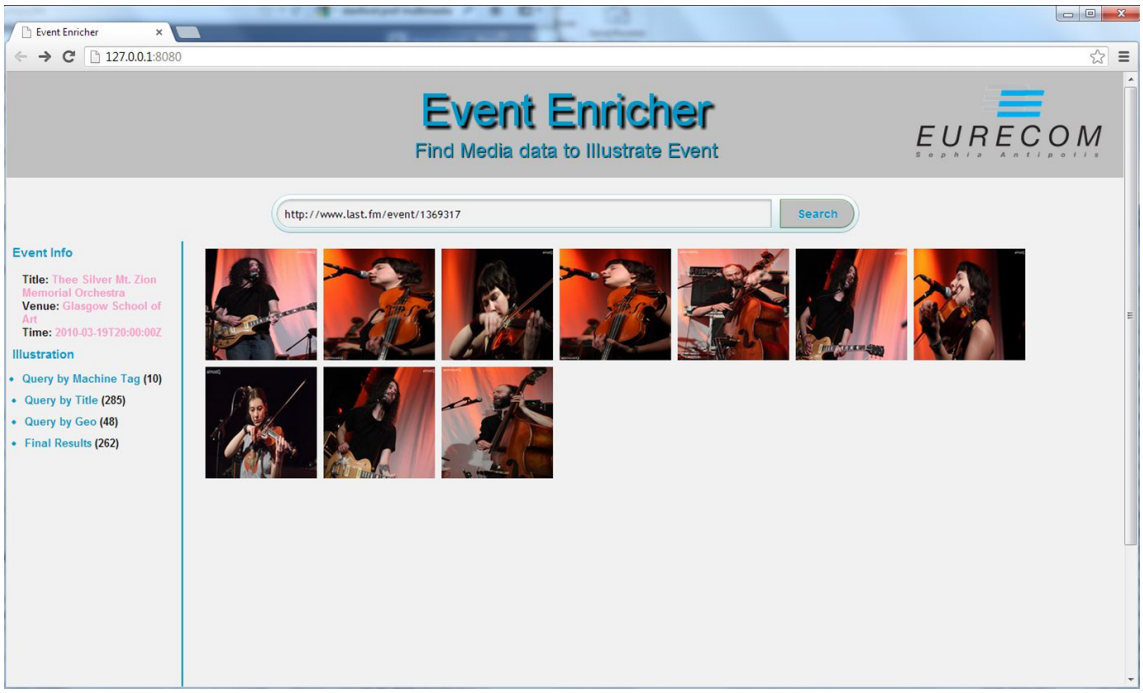


Figure 4.8: Query by Machine Tag in Event Enricher

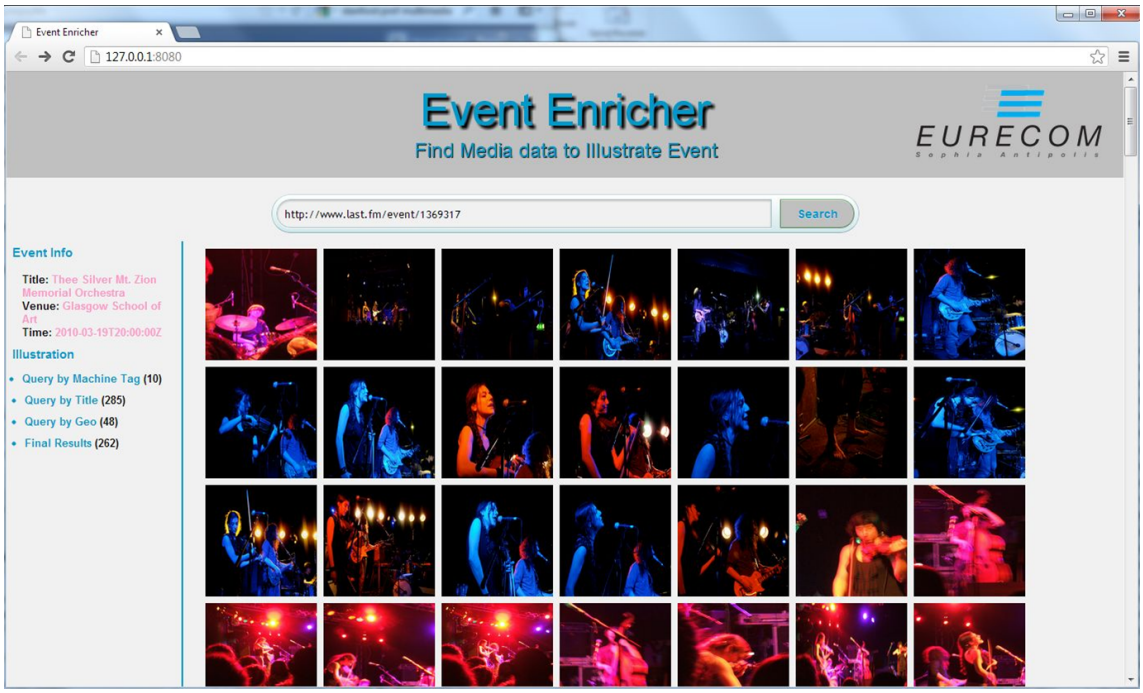


Figure 4.9: Query by Title in Event Enricher

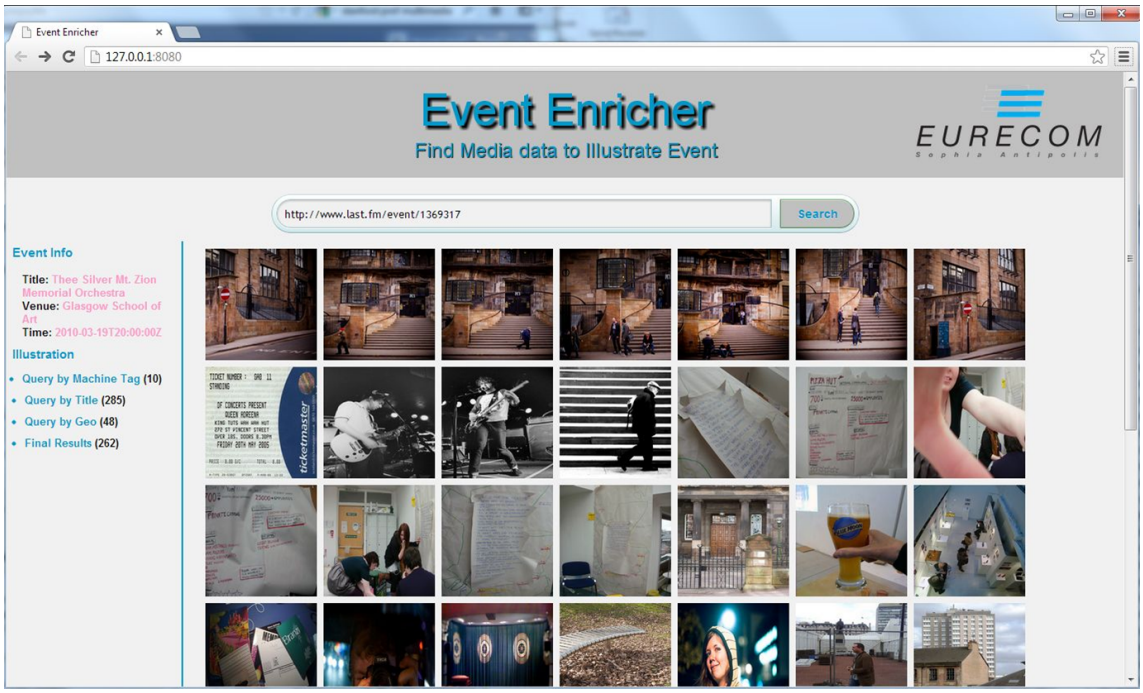


Figure 4.10: Query by Location in Event Enricher

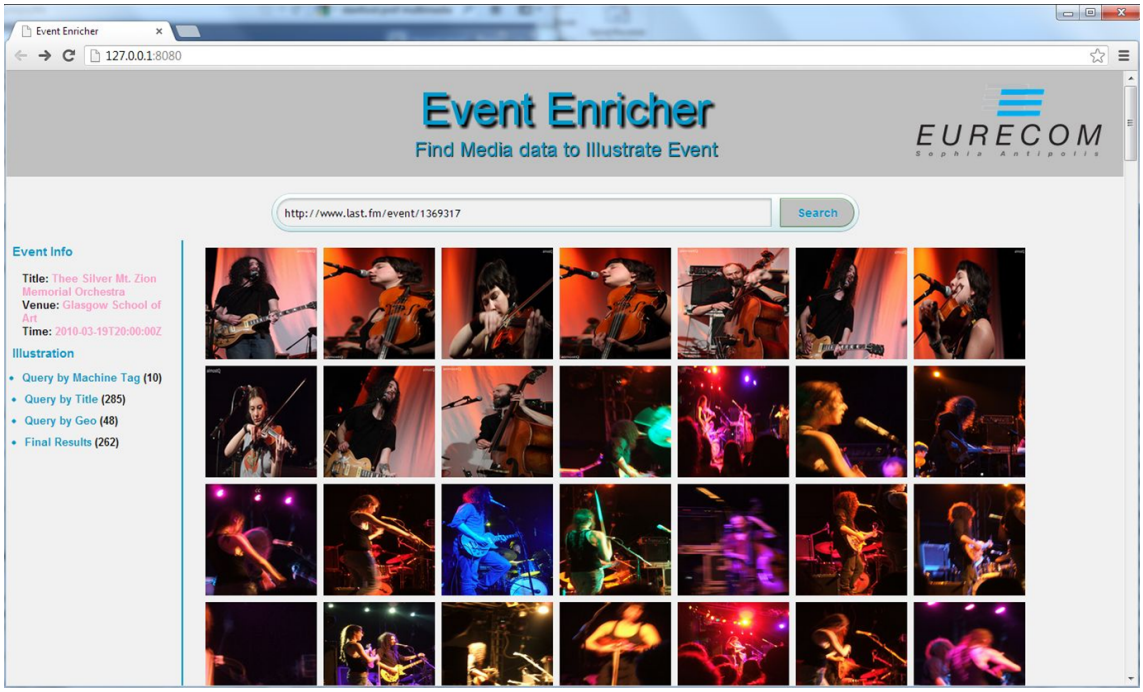


Figure 4.11: Final Result in Event Enricher

5 Modeling Events With Social Media Data

We define a social event as the specific happening that takes place at a given location and time and involve several persons (i.e. concerts, conferences, exhibitions, etc...). This work investigates the feasibility of modeling event visually from automatically collected data. To build a visual event model, one needs a collection of images labeled as positive or negative with respect to the event. Unfortunately, labeling data is a labor intensive and time consuming task. In this document, we propose an original scheme for collecting the training samples for modeling social events visual semantics without any human assistance. Figure 5.1 depicts the automated steps leading to the creation of the dataset to learn event models. The positive samples are collected directly from social media platforms using identification tag based query. The identification tags are the tags that refer to the event content accurately (i.e. event machine tag).

Collecting the representative negative samples is a more challenging task due to the vast amount of irrelevant data available. Here, negative samples are retrieved from online social media data using metadata analysis. We have observed while experimenting that when querying for photos originating from an event, based on its date and location, the negative samples (those photos which do not correspond to this particular event) are photos depicting general concepts for this location. Among such photos one typically finds, buildings, objects and portraits, etc... and some of the tags associated with these media are common for this location. For example, the city name is a popular tag in many situation yet it does not provide much discriminative information to accurately refer an event. In the work presented here, it is reasonable to assume that these photos captured at the same location as events and containing common tags as the most relevant negative samples for this specific event. Common tags, along with their corresponding photos, are identified based on a novel approach inspired from learning to rank [23], which will be detailed in Chapter 5.2.

5.1 Positive Samples Collection

We collect social events visual positive samples by querying social media platforms with event identification tag. There are different kinds of tags to identify events in social media data. The machine tag is an overlap metadata that is originating from some events repositories (such as LastFM[1], Upcoming[2] or Facebook[3]) and advertised by these web services to their users when they upload media data taken during the event. It is popularly used

[1] http:/www.last.fm

[2] http:/www.upcoming.org

[3] http://www.facebook.com/events/

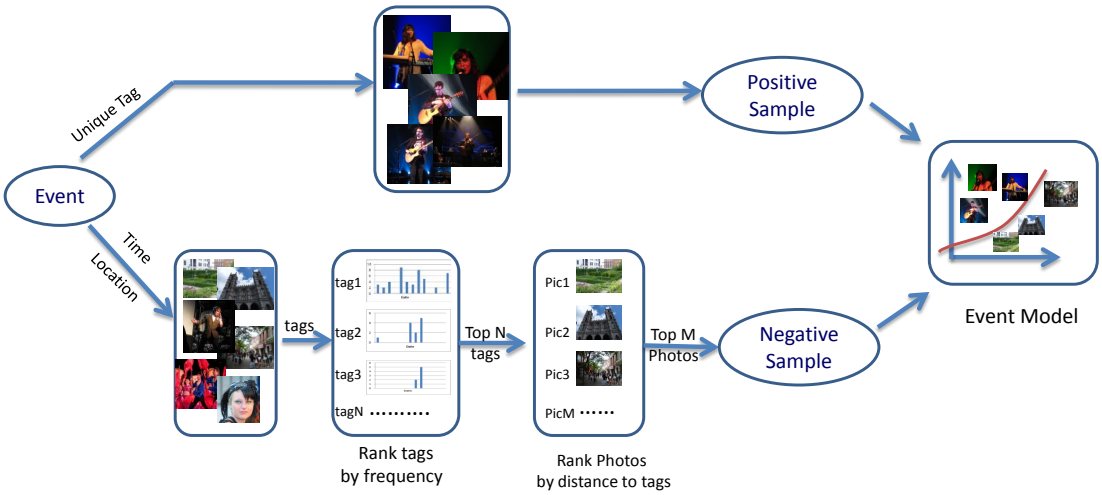


Figure 5.1: Overview of the framework for modeling events semantic. The positive samples are collected based on event machine tags (or specific identifiers), and the negative samples are collected using a learning-to-rank approach, to sort the photos according to the common-ness of its tags with respect to the geographical location.

to connect events and photo/video in media sharing platforms, such as Flickr[4]. In these social event websites, machine tags are formatted as "$DOMAIN:event=$XXX", where "$DOMAIN" is the name of website, and "$XXX" is the unique event id provided by the event sites, for example, "lastfm:event=1842684" is an event registered in Last.FM whose id is 1842684, and "facebook:event=108938242471051" is a facebook event whose id is 108938242471051. When users take photos during the event, they can upload them to media sharing websites with such a tag in order to explicitly associate the photos with the event. The machine tags can be recognized by both kinds of web services and give explicit and accurate links between events and multimedia documents. Hence the media documents containing the appropriate machine tag are taken as positive samples for the corresponding event.

Although machine tags are becoming more frequently used, many events still do not feature such metadata. To overcome this issue, we also use the abbreviated events name to identify certain events. The events abbreviations are well known and popularly used among the attendees. For example,"ACMMM10" is short for the ACM International Conference on Multimedia which took place in 2010, without any ambiguity. All photos with such tag are assumed to be positive samples of this social event in the current work.

[4] http:/www.flickr.com

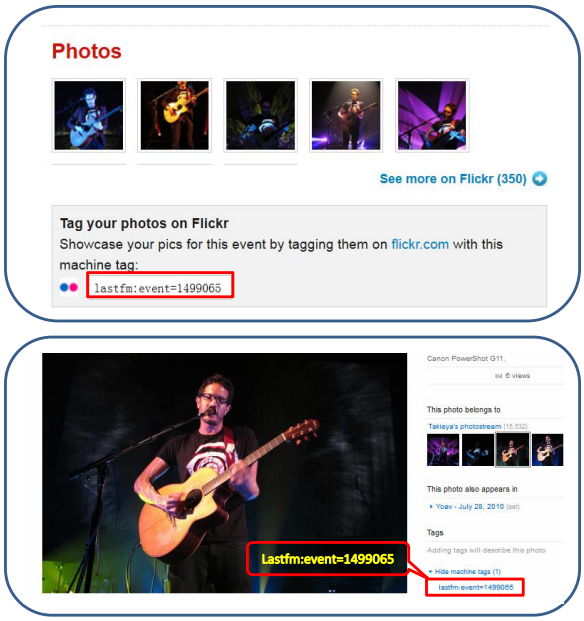


Figure 5.2: Machine Tags Used in Last.fm(Top) and Flickr(Bottom), which provide explicit ground truth on events and media data.

5.2 Negative Samples Collection

Since social events are characterized by a grouping of people at a given time and place, the most relevant negative samples are those images taken around the same period and location as the event but which do not originate from the event. Here is an example to motivate our assumption. Given an event held in a city near a famous landmark, it is likely that among the photos taken by attendees some will show the landmark. As a famous landmark, it is expected to be captured frequently by tourist. It is important that such photos are included in the negative samples in order to differentiate between the event and its surrounding. Based on this assumption, we collect negative samples with tags referring to the commonest concepts in that location. We measure the commonness of a tag by its frequency over a given period, and our approach to collect negative samples from localized data is composed of three steps.

The first step consists in gathering the photo candidates. For each event, online services such as Last.FM or Facebook/events are used to identify the location and date. These parameters are then employed to query the Flickr API for a photo set (P). The location is defined by a circle, whose center is determined by the GPS coordinates of the event venue and radius value (R). The time interval is the period of D days before and after the event's date. In order to obtain a large set of candidate photos, appropriate values should be set for both D (days) and R (kms). The influence of those two parameters will be studied in the experiment section 5.4.2.

The second step is to build the text ranking model to identify the "common tags". Here, we define "common tags" as tags that are commonly and frequently associated with a set of photos. In effect, a group of "common tags" represents the most general concepts associated with photos taken in a location. The commonness of a tag can be represented by the fraction of the number of days it appears within a given period. More formally, the commonness of tag t over a time period of D days can be calculated as:

$$Score(t) = \sum_{i=1}^{D} SD(t,i)/D$$

where the value of $SD(t,i)$ is 1 if tag t appears on day i, and 0 if not.

We rank the tags according to their $Score()$ decreasingly. The top N tags (with the largest $Score()$) are kept as the group of common tags $CTags$ for the given period at this location. These tags are prevalently used and highly relevant to the location but do not represent an event due to the fact that they cover a too large time-span. The effect of N, the number of common tags kept to represent the location, is also studied in the experiment section 5.4.2.

The last step is to select of the negative photo samples based on commonness ranking. For each photo p of P, we extract the title and tags as their text description $Text(p)$, and compute the similarity between these terms and the common tags obtained previously. The measure used here is the cosine distance [28].

$$Similarity(CTags, Text(p)) = \frac{CTags \cdot Text(p)}{\|CTags\|\|Text(p)\|}$$

All of the negative candidates are ranked by their textual similarity to the common tags set ($CTags$) and the top M photos are kept as negative samples for training the visual models.

Having collected both positive and negative visual examples of a particular event, machine learning approaches can be employed to learn the visual model. The methodology used to train the Support Vector Machines used in this work is detailed in 5.3.

5.3 Model Training

The collected data is adapted to training the event-specific models with different visual features and classifiers. Since SIFT feature is an effective feature to represent image content [27, 43], we follow this opt for its use for representing the content of the photos. The classifier used in this work is Support Vector Machine, which has been popularly used in different domains [7] and is nowadays prevalently employed for modeling visual content in multimedia indexing and retrieval systems [29]. Each individual event model is obtained as follows; First, 128D Scale Invariant Feature Transform (SIFT) feature is

computed over the local region detected by Difference of Gaussian (DoG) filter, then we cluster all the visual feature with K-means for each event, and the SIFT description is quantized to generate 400-dimensional Bag of Visual Words. The event model is learned by Support Vector Machine with Radial Basis Function kernel. Model parameters are optimized using cross-validation method.

5.4 Experiments

5.4.1 Data Set and Experiment Setting

Our proposed algorithm is evaluated on different types of events, including 10 concerts from LastFM, 3 scientific conferences and 1 popular street carnival. The photo source used here is Flickr, although other media and sources could be easily added to the framework. The details of each event in the dataset is presented in Table 5.1.

Table 5.1: The event dataset used in our experiments includes 10 concerts, 3 international conferences and 1 carnival.

EventID	Title	Date	Latitude	Longtitude
lastfm:804783	Metallica	03/03/2009	54.964053	-1.622136
lastfm:1830095	Hole in the Sky Bergen Metal Festival XII	24/08/2011	60.389585	5.323773
lastfm:1858887	Duran Duran	23/04/2011	41.888098	-87.629431
lastfm:1499065	Osheaga en Ville	28/07/2010	45.509788	-73.563446
lastfm:1787326	The Asylum Tour: The Door	03/03/2011	34.062496	-118.348874
lastfm:1351984	Bospop 2010	10/07/2010	50.788893	5.708738
lastfm:1842684	Buskers Bern	11/08/2011	46.947232	7.452345
lastfm:2020655	Lacuna Coil - Darkness Rising Tour	18/11/2011	50.723090	-1.864967
lastfm:1301748	End Of The Road Festival	10/09/2010	50.951341	-2.082616
lastfm:1370837	Into The Great Wide Open	03/09/2010	52.033333	4.433333
ACMMM10	the ACM conference on Multimedia 2010	25/10/2010	43.777846	11.249613
SIGIR2010	ACM Special Interest Group on Information Retrieval,2010	19/07/2010	46.194713	6.140347
ACMMM07	the ACM conference on Multimedia 2007	24/09/2007	48.334790	10.897200
NICECarnival2011	the Carnival de Nice 2011	05/03/2011	43.701530	7.278240

For our experiments, three photo sets are created. The first set contains all the Flickr photos which match the identification tag (EventID) of the 14 selected events. We randomly split the positive photos originating from each event into two equal parts according to usage: 50% for training, 50% for verifying.

The second set contains the negative candidates. Photos that are taken within a given spatial distance (less than R Kms from) and a given temporal interval (less than D days away) of each selected events are retrieved from Flickr. The process of common tags generation and photos ranking is performed on each event photo set in order to retain only the 200 most common photos (which corresponds to the average number of positive training samples) for each event as negative samples for training the model.

The third set of media is called Real Online data (**RO**) and is used to evaluate our approach in a real life situation. The collection is obtained using Flickr queries combining text, location and time as presented in [25]. This collection process is somehow similar to the one anyone would use to gather photos about an event from any user contributed content platform.The irrelevant photos in this dataset can not be filtered just according their metadata. The ground truth on this collection is provided by manual human labeling.

The number of photos for each event of the three sets can be found in Table 5.2 Since the data is collected based on a realistic scenario, it is diverse in terms of size and content. Clearly the number of photos for each events ranges from very few to several hundreds, while the photos describe different concepts, such as performers, buildings, sky etc...

Table 5.2: The number of media collected for the 14 events. Positive samples are collected with unique tags, negative samples are the photos taken near the event location (pre-ranking and selection) and RO data is collected by the methods proposed in [25], and are manually labeled.

EventID	**Positive Samples**	**Negative Candidate**	**RO**	
			Pos	**Neg**
lastfm:804783	441	1063	466	64
lastfm:1830095	716	748	398	134
lastfm:1858887	408	745	431	266
lastfm:1499065	348	712	16	153
lastfm:1787326	446	913	0	313
lastfm:1351984	307	584	498	19
lastfm:1842684	602	1125	535	78
lastfm:2020655	538	745	750	6
lastfm:1301748	944	541	1157	80
lastfm:1370837	592	1025	592	115
ACMMM07	100	557	178	23
SIGIR2010	30	525	0	201
ACMMM10	118	64	15	44
NICECarnival2011	52	848	60	209
Total	5642	10195	5096	1705

We use half the positive samples and the negative samples to train the SVM model for each event, and optimize the parameters D, R and common vocabulary size N using the remaining part of the positive samples.

In our experiments, the results are measured in terms of accuracy, a criteria commonly used for evaluating classification tasks [28]. Accuracy is defined as the number of true predicted elements divided by the total number of elements in the dataset. To be more precise, four values, True Positive(**TP**), True Negative(**TN**), False Positive(**FP**) and False Negative(**FN**) can be used to measure the performance of a classification or recognition

system. The terms Positive and Negative refer to the results that are predicted by a system, while True or False refer to whether the prediction is correct with respect to the ground truth. The accuracy measure is defined as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

This measure will be used for comparing the performance of various approaches of this document.

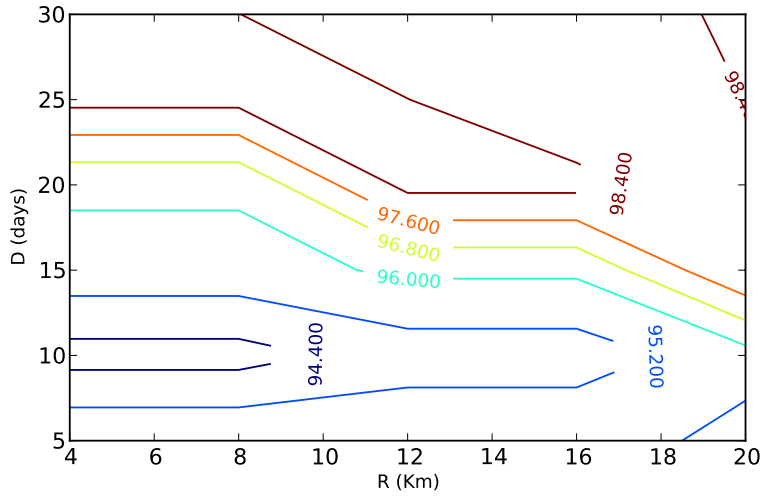
5.4.2 Location Distance, Time Interval and Tags Size

We investigate the impact of parameter R, and D, the location distance and time interval between photo taken and event held, to the final event model. We change the two parameters gradually and test the trained model accuracy on the verification dataset. Specifically, R is chosen from 4 to 20 kms with step of 4 kms, and D is set from 5 to 30 days with step 5 days. Cross-validation on the two parameters is performed in the process. Figure 5.3 shows 3 examples of resulting classification accuracy averaged over the different value of R, and D. Results for all selected events favor the use of rather large parameters for both time interval and location distance. This finding is supported by the fact that the larger the values of D and R, the more photos are retrieved from Flickr and this results in increased diversity within the selected negative samples. Based on the results obtained, the parameters of R and D are set as 20 kms and 30 days respectively.

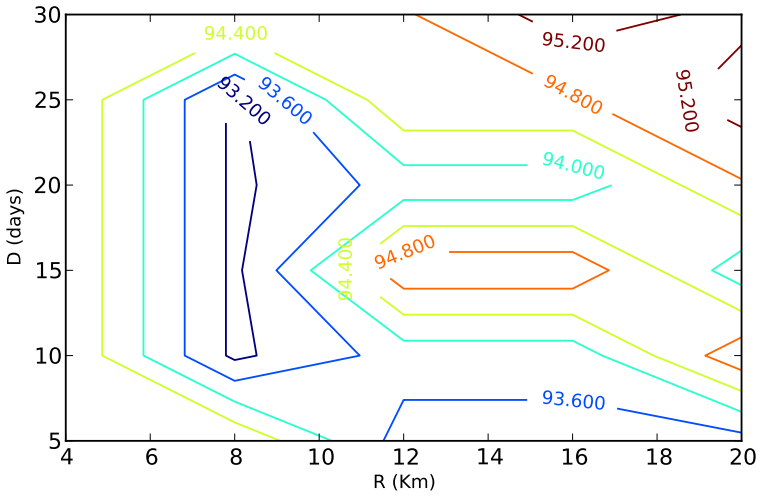
We also evaluate the influence of N, the number of common tags employed, with respect to the resulting event model accuracy. For each combination of parameters R and D, we optimize the model with vocabulary size varying from 5 to 50 tags. The results, presented in Figure 5.4, clearly indicate that the best performance is obtained when the negative vocabulary contains 10 tags.

5.4.3 Performance Evaluation

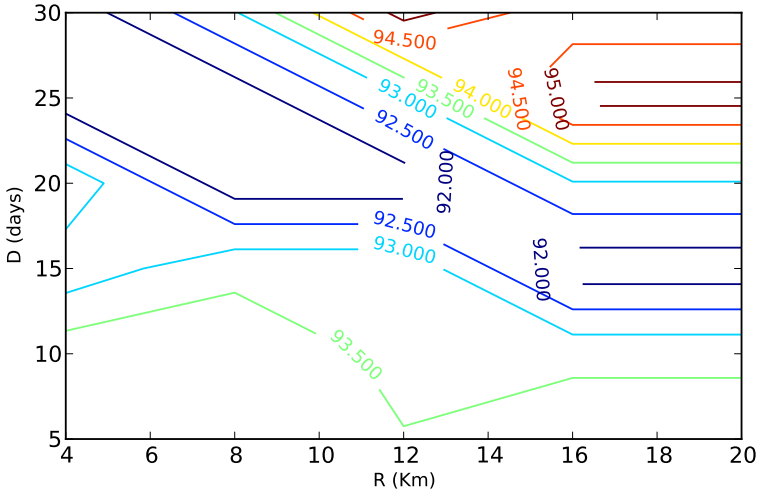
In our experiments, the automatically learned visual event models are compared with four other approaches at the task of mining online media illustrating events and collecting training sample effectively. The first and also the most basic approach, consist in simply running a Flickr query (the one used to create the real online data) and assuming all returned media are positive. In other words, the accuracy value reported in the column **Flickr Query**, indicates the precision in the **RO** test dataset. The second approach reported for comparison is similar to the K-NN visual filter proposed in previous work [24]. In this approach, photos in the test dataset are assigned to the event if and only if their visual similarity with their nearest neighbor is above a high threshold (i.e. 95%). This approach is fast, since it does not require any training nor collection of negative samples. However, the pruning rule is based solely on the analysis of positive samples.



(a) Event 804783



(b) Event 1351984



(c) Event 1351984

Figure 5.3: Cross Validation on R and D for 3 Events (Performance of classification measured by accuracy)

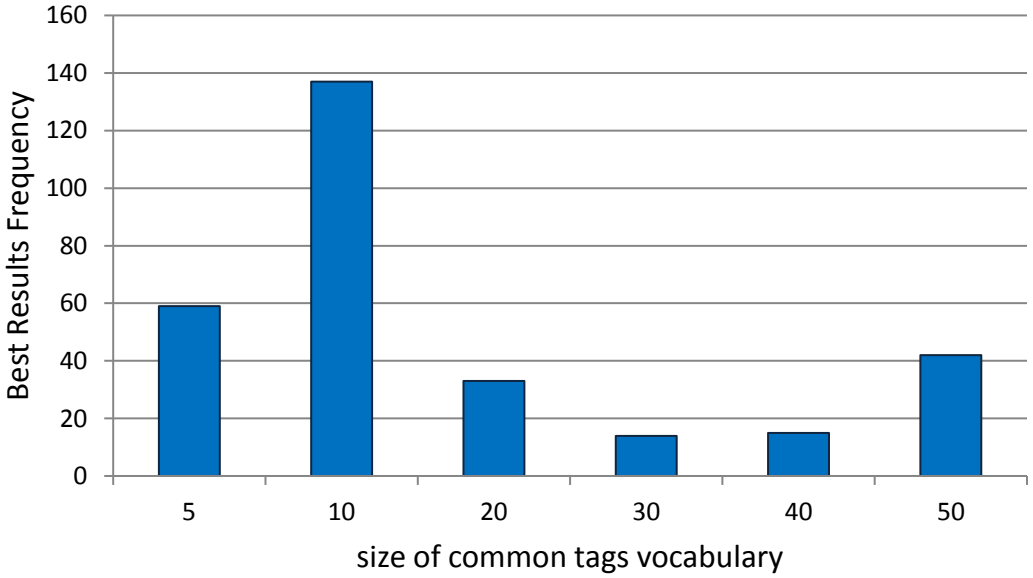


Figure 5.4: Performance vs size of common tag vocabulary. The best results are achieved when the 10 most common tags are employed.

In addition, we compare our approach with two different negative sample collection methods. In the third approach (column **Localization Aware**), rather than ranking photos based on the commonest tags, we use the negative samples randomly selected from the localized negative candidates to train the SVM models. In order to evaluate the influence of "location", a unique set of 200 negative samples is randomly selected from the entire set of (200 photos * 14 events) negative samples and used to train all SVM models. The results corresponding to this approach are reported in column **Localization Un-aware**.

It should be noticed that the values in the column **Flickr Query** shouldn't be compared with the values in the following 4 columns since it measures "Accuracy" in different context. Nonetheless, it is interesting to bear in mind the ratio between the number of positive samples and negative samples in the RO dataset for each event in order to better interpret the results obtained using of the four classification alternatives.

From the results presented in table 5.3, it is interesting to note that the approach proposed in [25] for analyzing visual content using K-NN filtering achieves, on average, almost the same performance as the **Flickr Query**. In other words, such a pruning approach is not very effective at identifying positive and negative illustrations of an event. When compared with the approach in [25], our learned visual model performs significantly and consistently better (83.3% vs 68.6% on average over all 14 events). This result shows the importance of exploiting negative samples to training the events visual content models where the margins between positive and negative samples can be maximized.

Table 5.3: Performance (Accuracy) of alternative classification approaches for associating Media with their corresponding Event

EventID	Flickr Query	Our Algorithm	Pruning in [25]	Localization Aware	Localization Un-aware
lastfm:804783	87.92	88.68	46.98	50.00	75.85
lastfm:1830095	74.81	78.38	80.26	96.62	84.96
lastfm:1858887	61.84	63.41	63.56	76.47	73.89
lastfm:1499065	9.47	90.53	89.94	92.90	89.35
lastfm:1787326	0.00	98.40	92.65	97.12	42.49
lastfm:1351984	96.32	96.32	55.32	86.65	93.81
lastfm:1842684	87.28	87.93	67.86	79.28	87.11
lastfm:2020655	99.21	91.80	71.69	75.00	94.58
lastfm:1301748	93.53	93.53	73.73	64.83	93.21
lastfm:1370837	83.73	85.15	73.83	60.25	80.62
SIGIR2010	0.00	60.19	42.28	16.41	22.38
ACMMM07	25.01	57.62	46.61	28.81	27.18
ACMMM10	85.83	91.04	87.56	86.57	89.05
NICECarnival2011	22.30	76.58	59.10	55.39	56.51
Average	69.41	83.31	68.64	70.07	73.42

Out of the three modeling approaches, our method obtains the best performance with an overall accuracy of 83.31%. Compared with our proposed approach, the models trained using random negative samples expose degraded accuracy (from 83.3% to 70.1%), which shows the importance of carefully selecting the negative samples when building the training collection. The idea of employing the commonest tags to identify nonevent related media proved to be effective. Moreover, the performance of models trained with the uniform negative dataset is better than models trained with random negative event sample, but not as accurate as our approaches. Those results confirm our hypothesis, "location" information plays an important role in negative samples collection and our approach is effective in collecting such negative samples.

In addition, we detail the final statistical results from the four approaches in Table 5.4. In this table, the results are measured in terms of True Positive (**TP**), True Negative (**TN**), False Positive (**FP**) and False Negative (**FN**) ratio. Clearly, although the Location Unaware method obtains the best True Positive ratio, however, it performs worst of all four approaches when dealing with negative sample (**TN**=17.82, **FN**=23.44). Both the K-NN pruning method from [25] and the Location-Aware method fail to correctly classify many positive samples(**FP** are 24.09 and 22.29 respectively) . While not achieving the best result in terms of **TP** alone, our proposed approach handles better than any others methods the negative samples, leading to the best performance overall.

Overall, the experiments have clearly shown the value of using visual analysis to model social events content. Furthermore, we have demonstrated that the construction of the event model can be automated without compromising the resulting performance.

Table 5.4: The detailed classification performance of the four approaches, averaged over all 14 events, measured in terms of True Positive (TP)), True Negative (TN), False Positive (FP) and False Negative (FN) ratio.

	TP	TN	FP	FN
Our Algorithm	49.47	33.85	9.84	6.83
Pruning in [25]	35.56	33.07	24.09	7.28
Localization Aware	37.03	33.05	22.29	7.63
Localization Un-aware	55.63	17.82	3.10	23.44

6 Conclusion and Future Work

This document reports and details the techniques that are currently being developed within the ALIAS project for enabling an easy access to events through media. In particular, it identified the objective and implementation details in the entertainment part of the ALIAS project, and an event based media retrieval/browsing framework is proposed to query the media originating from social events. The work will be helpful to assist the elderly, or any user, to identify new potentially interesting social events that they may wish to attend (in the case of future event) or relive (in the case of a past event). In this deliverable, we mainly focus on the study of public events, such as concerts, conferences, etc...

How to handle private events, such as birthday party, wedding is a promising direction for future work. However, research in this direction is made difficult due to the lack of media available since media originating from private events are as the event.. private.

Bibliography

[1] J. K. Aggarwal and Q. Cai. Human motion analysis: a review. In *IEEE Nonrigid and Articulated Motion Workshop*, pages 90–102, 1997.

[2] 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.

[3] L. Ballan, M. Bertini, A. Bimbo, L. Seidenari, and G. Serra. Event detection and recognition for semantic annotation of video. *Multimedia Tools and Applications*, 51(1):279–302, Nov. 2010.

[4] H. Becker, D. Iter, M. Naaman, and L. Gravano. Identifying content for planned events across social media sites. In *ACM conference on WSDM*, 2012.

[5] 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.

[6] 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.

[7] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[8] J. M. Carroll. *Making use: Scenario-based design of human-computer interactions*. MIT Press, 2000.

[9] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.

[10] T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y.-T. Zheng. NUS-WIDE: A Real-World Web Image Database from National University of Singapore. In *Proc. of ACM Conf. on Image and Video Retrieval*, Santorini, Greece, 2009.

[11] 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.

[12] D. Delgado, J. Magalhaes, and N. Correia. Automated Illustration of News Stories. In *2010 IEEE Fourth International Conference on Semantic Computing*, pages 73–78. IEEE, Sept. 2010.

[13] J. Fan, Y. Shen, N. Zhou, and Y. Gao. Harvesting Large-Scale Weakly-Tagged Image Databases from the Web. In *IEEE Conference on CVPR*, 2010.

[14] 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.

[15] C. S. Firan, M. Georgescu, W. Nejdl, and R. Paiu. Bringing order to your photos: event-driven classification of flickr images based on social. In *Proceedings of the 19th ACM international conference on Information and knowledge management - CIKM '10*, page 189, New York, New York, USA, Oct. 2010. ACM Press.

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

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

[18] R. Hong, G. Li, L. Nie, J. Tang, and T.-S. Chua. Explore Large Scale Data for Multimedia QA. In *ACM conference on CIVR*, Xi'an, China, 2010.

[19] 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.

[20] L.-J. Li and G. Wang. OPTIMOL: automatic Online Picture collecTion via Incremental MOdel Learning. *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 88(2):1–8, 2007.

[21] X. Li, C. G. Snoek, M. Worring, and A. W. Smeulders. Social negative bootstrapping for visual categorization. In *Proceedings of the ACM International Conference on Multimedia Retrieval*, 2011.

[22] 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.

[23] T.-Y. Liu. *Learning to Rank for Information Retrieval*. springer, 2011.

[24] X. Liu, R. Troncy, and B. Huet. Module for cross-media linking of personal events to web content, v1. *Alias project Deliverable*, D4.2.

[25] X. Liu, R. Troncy, and B. Huet. Finding Media Illustrating Events. In *ACM Conference on ICMR*, 2011.

[26] D. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE Conference on ICCV*, pages 1150–1157, 1999.

[27] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, Nov. 2004.

[28] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 1 edition, July 2008.

[29] P. Over, G. Awad, M. Michel, J. Fiscus, W. Kraaij, and A. F. Smeaton. Trecvid 2011 – an overview of the goals, tasks, data, evaluation mechanisms and metrics. In *Proceedings of TRECVID*. NIST, USA, 2011.

[30] T. Quack, B. Leibe, and L. Van Gool. World-scale mining of objects and events from community photo collections. In *ACM conference on CIVR*, page 47, New York, USA, July 2008.

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

[32] C. Saathoff and A. Scherp. Unlocking the Semantics of Multimedia Presentations in the Web with the Multimedia Metadata Ontology. In *19^{th} World Wide Web Conference (WWW'10)*, Raleigh, USA, 2010.

[33] G. M. Sacco and Y. Tzitzikas. *Dynamic Taxonomies and Faceted Search: Theory, Practice, and Experience*, volume 25 of *The Information Retrieval Series*. Springer, 2009.

[34] A. Scherp, T. Franz, C. Saathoff, and S. Staab. F—A Model of Events based on the Foundational Ontology DOLCE+ Ultra Light. In *5^{th} International Conference on Knowledge Capture (K-CAP'09)*, Redondo Beach, California, USA, 2009.

[35] F. Schroff, A. Criminisi, and A. Zisserman. Harvesting Image Databases from the Web. In *IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE, 2007.

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

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

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

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

[40] 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.

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

[42] Z.-J. Zha, T. Mei, J. Wang, Z. Wang, and X.-S. Hua. GRAPH-BASED SEMI-SUPERVISED LEARNING WITH MULTI-LABEL. *ACM Trans. Program. Lang. Syst.*, 20(5):97–103, 2009.

[43] L. Zhang, F. Lin, and B. Zhang. Support vector machine learning for image retrieval. *International Conference on Image processing*, 2(x):721–724, 2001.

[44] 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.

[45] S. Zhu, G. Wang, C. Ngo, and Y. Jiang. On the sampling of web images for learning visual concept classifiers. *Proceedings of the ACM CIVR*, 2010.