The 2012 Social Event Detection Dataset

Symeon Papadopoulos
CERTH-ITI
papadop@iti.gr

Emmanouil Schinas
CERTH-ITI
manosetro@iti.gr

Vasileios Mezaris
CERTH-ITI
bmezaris@iti.gr

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

Ioannis Kompatsiaris
CERTH-ITI
57001 Thermi, Greece
ikom@iti.gr

ABSTRACT

This paper presents the 2012 Social Event Detection dataset (SED2012). The dataset constitutes a challenging benchmark for methods that detect social events in large collections of multimedia items. More specifically, the dataset comprises more than 160 thousands of Flickr photos and their accompanying metadata, as well as a list of 149 manually selected and annotated target events, each of which is defined as a set of relevant photos. This paper discusses the challenges defined as part of SED 2012, the data collection process, the dataset and its basic statistics, the ground truth creation and the suggested evaluation methodology.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

General Terms

Experimentation

Keywords

social event detection, dataset, multimedia, flickr

1. INTRODUCTION

The modeling, detection, and processing of events is an area that has started to receive considerable attention by the multimedia and semantic web communities [1, 4], as testified, for example, by the EiMM workshop series at ACM Multimedia or the DeRiVE workshop series at ISWC. In line with such initiatives, the Social Event Detection (SED) task of MediaEval 2012 [2] requires participants to discover social events and detect related media items. By social events, we mean that the events are planned and attended by people and that the media illustrating the events are captured by people. A lot of multimedia content on the Internet was captured during such events or is otherwise related to events. Automatically establishing these underlying event-media associations represents a big step towards enabling multimedia browsing and search that is more natural to the users. To this end, this paper presents the SED2012 dataset, a large collection of user-generated images and metadata accompanied by task definitions (challenges), ground truth data and evaluation measures. The dataset is publicly available[1]. The dataset described here is the same as the one used in the 2012 edition of the SED task[2]. Compared to [2], this paper provides more extensive descriptions of the dataset (including statistics) and the corresponding collection and ground truth creation methodologies.

MMSys '13, February 26-March 1, 2013, Oslo, Norway.

2. SED2012 OVERVIEW

SED2012 is composed of three challenges, a common test dataset of images with their metadata (time-stamps, tags, geotags for a small subset of them), and ground truth data pertaining to the association of dataset images to a list of target events. Formally, SED2012 is a photo collection denoted by $\mathbb{P} \triangleq \{p\}$, where p is a tuple $(\theta_p, l_p, t_p, u_p, d_p, X_p)$ containing a unique photo identifier θ_p, geotagged with location information l_p (consisting of a pair of latitude-longitude coordinates), captured at time t_p and uploaded by user u_p with the title-description d_p and a set of tags X_p. In addition to the available media content and associated metadata, SED2012 comprises a set of target events $\mathbb{E} \triangleq \{e\}$, each of which is associated with a set $\mathbb{P}_e$ of at least one photo.

Finding the events, according to the task definition, does not mean finding some textual descriptions or metadata of the events. The expected output of a social event detection method is a set of photo clusters $\mathbb{C} \triangleq \{c\}$, each cluster c comprising only photos $\mathbb{P}_c$ associated with a single event (thus, each cluster defining a retrieved event). The "photos associated with a single event" are all photos of the test collection that directly relate (in content, and also in terms of place/time) with the event of interest. The criteria for evaluating social event detection methods in SED2012 take into account the number of correctly detected events (out of all relevant events in the test set) and the number of correct/incorrect media items detected for these events.

As part of SED2012, the following are distributed: (a) a test kit, which includes the definitions of the three SED 2012 challenges and the XML file with the image metadata that can be used for addressing these challenges, (b) the images of the collection (167,332 images) that were captured between the beginning of 2009 and end of 2011 by 4,422 unique Flickr users, and posted to Flickr by their respective owners under

[1] http://mklab.iti.gr/project/sed2012
[2] Except for few minor corrections on the ground truth.

a Creative Commons license), (c) the ground truth annotations for the three defined challenges (task definitions) on the provided dataset, together with a script for evaluating any social event detection results against this ground truth.

The rest of the paper is structured as follows. Section 3 specifies the three challenges defined by SED2012. Section 4 describes the data collection process and basic statistics of the collected dataset. Finally, Section 5 specifies the evaluation methodology for SED2012, comprising the ground truth creation process, basic statistics for the ground truth and the selected evaluation measures.

3. CHALLENGES

SED2012 defines three challenges, each of which focuses on a target class of events. For convenience, the corresponding events are denoted as $\mathbb{E}_1, \mathbb{E}_2, \mathbb{E}_3$.

Challenge 1: *Technical events taking place in Germany.*
Technical events, for the purpose of this task, are public technical events such as exhibitions and fairs. The annual CeBIT exhibition, taking place in Hannover, is a good (but of course, not the only) example of such an event.

Challenge 2: *Soccer events taking place in Hamburg (Germany) and Madrid (Spain).*
This challenge is very similar to the first of the two challenges defined in SED 2011 [3]. Similarly to 2011, soccer events, for the purpose of this task, are soccer games and social events centered around soccer such as the celebration of winning a cup. In contrast, a person playing with a soccer ball out in the street is not a soccer event under the task definition. The reader is referred to [3] for more details on the meaning of a "soccer event".

Challenge 3: *Demonstration and protest events of the Indignados movement occurring in public places in Madrid.*
The Spanish Indignados movement centers around a series of demonstrations and other protests taking place all over Spain, which relate to the recent financial crisis outbreak as well as national politics in general. In contrast to the events that the first two challenges are concerned with, the events that are of interest to this third challenge are not scheduled, well-organized events (e.g., a technical fair or a soccer game, which are typically scheduled several months or days in advance, respectively). Instead, they are to a large extent spontaneous events, with any organization efforts related to them being typically centered around social media channels.

4. DATASET

4.1 Collection process

The data collection process was primarily based on the Flickr API[3]. More specifically, the collection was based on the method **flickr.photo.search** using five different cities (Barcelona, Madrid, Cologne, Hamburg, Hannover) as geographical centres and the interval between the beginning of 2009 and the end of 2011 as temporal constraint. An additional constraint was that the photos should be shared under a Creative Commons license in order to be able to distribute them publicly for research purposes. All collected photos were geotagged (although as will be mentioned later, geolocation information was removed from the majority of photos). In addition to these photos, the collection was enriched with some technical event photos from EventMedia (cf. Section 5.2), also licensed under Creative Commons.

[3] http://www.flickr.com/services/api/

Table 1: Distribution of photos per city

Hannover	Hamburg	Cologne	Madrid	Barcelona
2,936	16,958	16,140	59,043	72,255
1.75%	10.13%	9.65%	35.29%	43.18%

Table 2: Distribution of photos per year

2009	2010	2011
51,019	53,080	63,233

The selection of the five cities listed above was motivated by the following reasons:

- To make the task more challenging by including a large number of non-English textual metadata (cf. Table 3).
- To ensure the existence of numerous events of the three challenges described in Section 3.
- To include distractor images for Challenges 2 and 3. For instance, in the case of Challenge 3, Indignados events taking place in Barcelona could act as distractors to the target Indignados events in Madrid.

4.2 Dataset statistics

The data gathering process described above resulted in a collection of 167.332 photos contributed by 4,422 unique Flickr users. Out of those, 403 photos come from the EventMedia dataset. All photos but the EventMedia ones, were originally geo-tagged. However, before including the photo metadata (including any tags, geotags, time-stamps, etc.) in the XML file that we provide as part of the SED2012 test kit, the geotags were removed for 80% of the photos in the collection (randomly selected). This was done for simulating the frequent lack of geotags in photo collections on the Internet (including the Flickr collection). In addition, we removed all *machine tags* from the dataset. Machine tags are a special type of tag in Flickr that often provide additional information in structured form. Such information can range from location to even direct association to events.

The distribution of photos among the five cities of the dataset appears in Table 1. According to it, Barcelona and Madrid constitute the bulk of the image set, and they are followed by Hamburg and Cologne; Hannover is only sparsely represented in the dataset, with less than 2% of the dataset photos captured there. The yearly distribution of photos is shown in Table 2. There is a gradual increase in the number of photos per year between 2009 and 2011, probably owing to the increased uptake of geotagging practices by Flickr users at that interval. The language distribution of photos appears in Table 3[4]. Moreover, Table 4 indicates the distribution of different CC licenses among the dataset photos.

In total, 51,611 unique tags were used in the dataset, thus creating a diverse vocabulary space. Table 5 lists the top 30 tags in the dataset ranked by usage frequency. As expected,

[4] Computed with: http://code.google.com/p/language-detection/

Table 3: Distribution of top languages, computed using the title (T) or description (D) of photos.

	es	en	de	it	pt	fr	other
T	35,619	27,492	23,893	8,648	7,509	5,442	47,296
D	25,121	27,402	6,003	1,555	1,668	1,468	5,157

Table 4: Distribution of photos per Creative Commons license type. In the URL column, *cc:* stands for *http://creativecommons.org/licenses/*.

License type	URL	Photos	%
BY-NC-SA 2.0	cc:by-nc-sa/2.0/	66,192	39.56
CC BY-NC 2.0	cc:by-nc/2.0/	16,172	9.67
CC BY-NC-ND 2.0	cc:by-nc-nd/2.0/	36,985	22.10
CC BY 2.0	cc:by/2.0/	21,578	12.90
CC BY-SA 2.0	cc:by-sa/2.0/	20,535	12.27
CC BY-ND 2.0	cc:by-nd/2.0/	5,870	3.51

Table 5: Most frequent tags in the dataset

#	Tag	Count	#	Tag	Count
1	barcelona	41,128	2	spain	30,941
3	madrid	30,871	4	españa	16,766
5	catalunya	12,588	6	hamburg	10,495
7	germany	8940	8	catalonia	7,761
9	köln	7,051	10	europe	6644
11	cologne	5,917	12	deutschland	5,676
13	iphoneography	5,420	14	square	5,092
15	square format	4,773	16	gaudi	4,700
17	instagram app	4,699	18	cataluña	4,387
19	music	4,050	20	concert	3,662
21	architecture	3,576	22	skate	3,474
23	espagne	3,405	24	live	3,374
25	plaza	3,027	26	spanishrev...	2,973
27	espanya	2,831	28	concierto	2,665
29	acampadasol	2,583	30	night	2568
102	indignados	1,022	274	football	562
500	conference	332	541	expo	310
911	soccer	196	1990	konferenz	95

there is a prevalence of location specifying tags, e.g. country and city names. Such tags could be helpful in filtering out irrelevant photos given the definition of a challenge (all definitions of Section 3 include a spatial component). The table also contains the frequencies of six tags that are representative of the event classes described in the three above challenges (expo, conference, konferenz, soccer, football, indignados) . The frequencies of the latter reveal that there are considerable numbers of photos related to the event classes of interest, although such statistics can provide only very rough estimates of the actually relevant photos.

Further insights regarding the tag and photo capturing behaviour of users can be gleaned from Figure 1. Figure 1(a) depicts the distribution of the number of tags assigned to each photo. As expected, the large majority (83.9%) of photos are tagged with less than 10 tags. Also, a considerable percentage of photos ($\sim$ 12.5%) does not have any tags associated with them ($X_p = \oslash$). Such kind of skewed tagging behaviour complicates the task of automatic event detection, since tags carry significant amounts of information with respect to the content of an image.

Figure 1(b) illustrates the distribution of tag frequencies in the dataset. The head of the distribution (first 30 points) refers to the tags of Table 5. A look into the tail of the distribution reveals that a large percentage of tags are only scarcely used. More than half of the tags (57.12%) appear just once or twice in the whole dataset. It is obvious that taking into account such tags for event detection may be of limited value due to the sparsity of occurrence.

Finally, Figure 1(c) shows the distribution of user activity, expressed by the number of photos per user. As expected, this distribution is highly skewed, with the 30 most active users contributing approximately 30% of the whole dataset. In contrast, a sizable percentage of users (1046 out of 4422) contributed only a single photo to the dataset, and almost 60% of users contributed less than 10 photos. Thus, analysis methods based on personal user history are applicable only to a limited portion of the dataset.

5. EVALUATION

5.1 Ground truth creation

The ground truth for SED2012 was created with the help of a web-based event annotation tool, namely CrEve [6]. Furthermore, the ground truth for Challenge 1 was enriched by use of the EventMedia dataset [5].

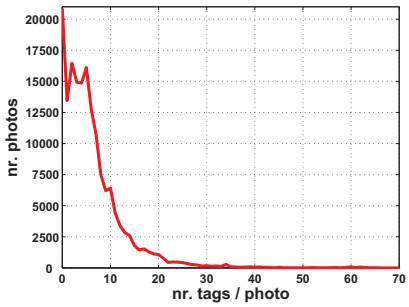
CrEve: The bulk of the annotation was conducted with the help of the CrEve annotation tool [6][5]. The annotation was conducted in two cycles: three annotators, one for each challenge, produced a first round of annotations, and two independent annotators used the results of the first annotation round to refine the ground truth, i.e. by extending or correcting existing events or adding new ones. In each round, the annotation was based on the following sequence:

1. familiarization of annotator with the event domain,
2. usage of a set of relevant keywords as a first source of creating events,
3. usage of location of found events to find additional photos of the same or new events in the same location,
4. usage of usernames who contributed photos to the detected events (to find additional relevant events contributed by the same user),
5. usage of additional keywords that were discovered to be relevant after reviewing the set of collected events.

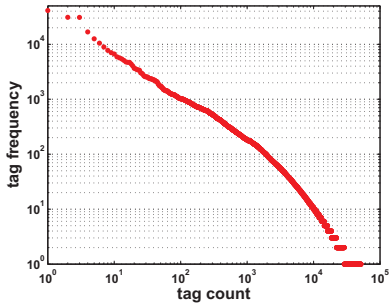
The annotation process was terminated as soon as it was deemed impossible to discover (by means of the search facilities of the tool) new photos related to the events of interest. In total, we estimate that the ground truth construction required approximately 100 hours of manual annotation effort. The amount of expended effort coupled with the two-round annotation scheme and the efficacy of the annotation tool, which, as demonstrated in [6], enabled non-familiar users to produce high-quality and coverage annotations, gives high confidence to the quality and reliability of the produced ground truth. It should be noted that significant amount of the effort was directed to the decision on borderline cases, i.e. photos for which it was not clear whether they should be associated with an event. There were numerous such cases, highlighting the complexity of the task even for a human annotator. However, compared to the number of clearcut cases, the number of borderline cases is small, thus its impact on the reliability of the assessment based on SED2012 should be considered limited.

EventMedia: Part of the events of Challenge 1 were retrieved by use of the EventMedia application [5]. The EventMedia knowledge base is composed of events descriptions together with media descriptions associated with these events and interlinked with the larger Linked Open Data

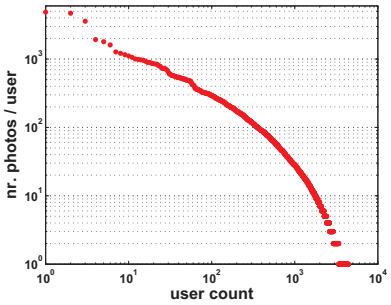
[5] http://clusttour.gr/creve/



(a) Tags per photo.



(b) Tag frequency.



(c) User activity.

Figure 1: Basic statistics of SED2012.

Table 6: Basic event statistics.

Challenge	#events	#photos	#users
$\mathbb{E}_1$ (technical)	18	2,186	55
$\mathbb{E}_2$ (soccer)	79	1,612	29
$\mathbb{E}_3$ (indignados)	52	3,981	116
All	149	7,779	198

cloud. The EventMedia collection is the result of crawling and aggregating data from different event directories, namely Last.fm, Eventful, Upcoming, Lanyrd[6], and its combination with content from Flickr. In total, 403 photos associated with technical events in Germany were retrieved from EventMedia and added to the SED2012 ground truth.

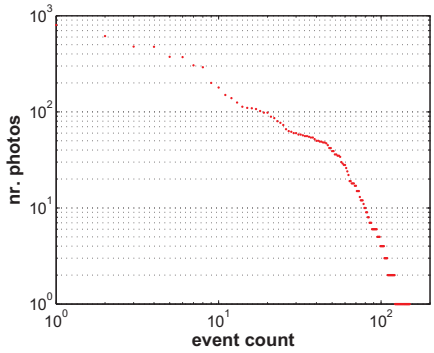
5.2 Ground truth statistics

Table 6 presents some basic statistics of the collected events. In total, the ground truth comprises 149 events, associated with 7779 photos that were captured by 198 distinct users. Thus, about 4.7% of the dataset photos are associated with the events of interest, and 4.5% of the dataset users have contributed photos to them. The largest number of events was collected for the second challenge (soccer). However, the soccer events of the dataset are associated with less photos and users compared to the other two challenges.

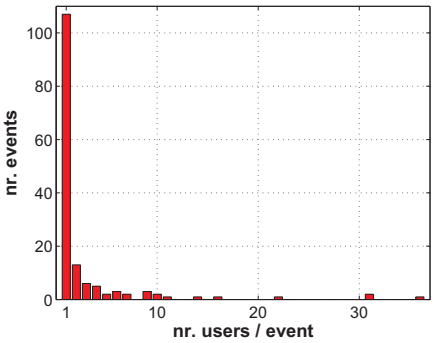
More specifically, each event of Challenge 1 is on average associated with 121 photos and 3 users. Challenge 2 is on average associated with 20 photos and 0.37 users, i.e. each user covers more than one event with his/her photos. Finally, Challenge 3 events are on average associated with 77 photos and 2.23 users. The differentiation of statistics for Challenge 2 can be attributed to the fact that soccer events tend to be covered by committed fans who regularly attend soccer matches (i.e. the vast majority of soccer events).

The average number of photos and users per event reported above do not accurately reflect the event-photo and event-user associations. To this end, Figure 2 presents the distributions of the number of photos and users per event. According to Figure 2(a), there is a small number of events ($\sim$ 10) associated with a large number of photos ($>$ 100), while a significant percentage of the events (27.3%) are associated with only one or two photos. Such events are much harder to detect, especially by methods that make

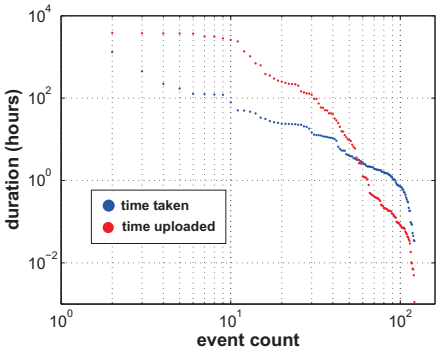
[6] http://www.last.fm, http://www.eventful.com, http://upcoming.yahoo.com,http://www.lanyrd.com



(a) Photos per event.



(b) Users per event.



(c) Event duration.

Figure 2: Basic statistics of SED2012 events.

use of clustering techniques. Similar observations hold for the number of users per event, as illustrated by Figure 2(b). The majority of events (106 out of 149) are captured by single users. Only nine events (one of them from Challenge 1 and eight of them from Challenge 3) are captured by 10 or more users. The largest event, namely the first large gathering of Indignados at Puerta del Sol on 21 May 2011, was captured by 36 Flickr users.

Figure 2(c) illustrates the distribution of the events' duration. For each event, we compute its duration as the difference between the timestamps of the last and the first photo associated with it. Obviously, this is possible to compute only for events with more than one photo associated with them. Since in Flickr there are two timestamps for each photo, namely the capture time (or time taken) and the upload time, Figure 2(c) depicts the distribution of durations for both, expecting that the duration based on the capture time (in blue) to be more representative of the true event duration. Obviously, this duration estimation entails a gross approximation since the respective event photos might only cover a small aspect of the actual event. Nevertheless, the diagram can be used to draw some basic conclusions.

By taking a more careful look into the depicted event durations, we could spot two cases of erroneous timestamp information in the respective event photos. For instance, few of the photos associated with the event "Indignados protest against the Europact measures" (19 June 2011) were dated as captured on 19 June 2010 (impossible since no such event had taken place at that time). This gave the impression that this event lasts over a year, while in fact it lasted less than three days. A similar case was observed with respect to the soccer match between Real Madrid and Liverpool. Those two cases correspond to the first two points in Figure 2(c). The rest of the events are associated with "reasonable" durations of seven days or shorter intervals. Technical events tend to last multiple days, while, as expected, soccer events last a few hours. Erroneous timestamp information severely affects the performance of event detection methods, many of which rely on the analysis of temporal media distribution.

The spatial distribution of the ground truth events constitutes another interesting aspect of the dataset. Figure 3 depicts the distribution of those SED2012 events that are located in Madrid and Hamburg. Since Madrid events are scattered over a wide area, they are illustrated in two zoom levels. By inspecting the locations of the events, it becomes clear that technical and soccer events are mostly associated with a limited number of locations (e.g. conference centres, stadiums), while the Indignados events take place in less predictable locations (typically in squares).

Figure 4 illustrates several example photos associated with some of the SED2012 events. The examples demonstrate that the visual content of photos is to some extent distinctive of the event category, but is of limited use for distinguishing between different events of the same category.

5.3 Evaluation metrics

The recommended evaluation of event detection methods is performed with the use of the ground truth described in Section 5.2. In the official benchmark of SED2012, two evaluation measures are used to compare the ground truth sets of events $\mathbb{E}_1, \mathbb{E}_2, \mathbb{E}_3$ with the candidate sets of events $\mathbb{C}_1, \mathbb{C}_2, \mathbb{C}_3$ automatically produced by the method under test:

- Harmonic mean (F-score) of Precision and Recall for the retrieved images. We use the macro version of F-score. This measures only the goodness of the retrieved photos but not the number of retrieved events, nor how accurate the correspondence between retrieved images and events is.
- Normalized Mutual Information (NMI). This compares two sets of photo clusters (where each cluster comprises the images of a single event), jointly considering the goodness of the retrieved photos and their assignment to different events.

Both evaluation measures receive values in the range [0, 1] with higher values indicating a better agreement with the ground truth results. These evaluation measures can be calculated both per challenge and on aggregate (for those teams that submit runs to all challenges). Evaluating a new method with these measures and by use of the evaluation script made available together with the dataset enables its comparison with the set of methods that were submitted for evaluation to SED2012 [2].

6. CONCLUSIONS

The SED2012 dataset gives the opportunity to comparatively evaluate different approaches to the problem of social event detection in multimedia collections. The collection constitutes a large sample of photos emulating well the characteristics of photos shared online, while the accompanying ground truth consists of target events that are representative of a wide variety of real-world social events. The dataset could also be helpful in other multimedia research problems, e.g. image geotagging and collection summarization.

Acknowledgments

The work presented in this paper was supported by the European Commission under contracts FP7-287911 LinkedTV, FP7-248984 GLOCAL, FP7-287975 SocialSensor and FP7-249008 CHORUS+.

7. REFERENCES

[1] V. Mezaris, A. Scherp, R. Jain, and et.al. Modeling and Representing Events in Multimedia. In *19th ACM Int. Conf. on Multimedia*, Scottsdale, AZ, USA, 2011.

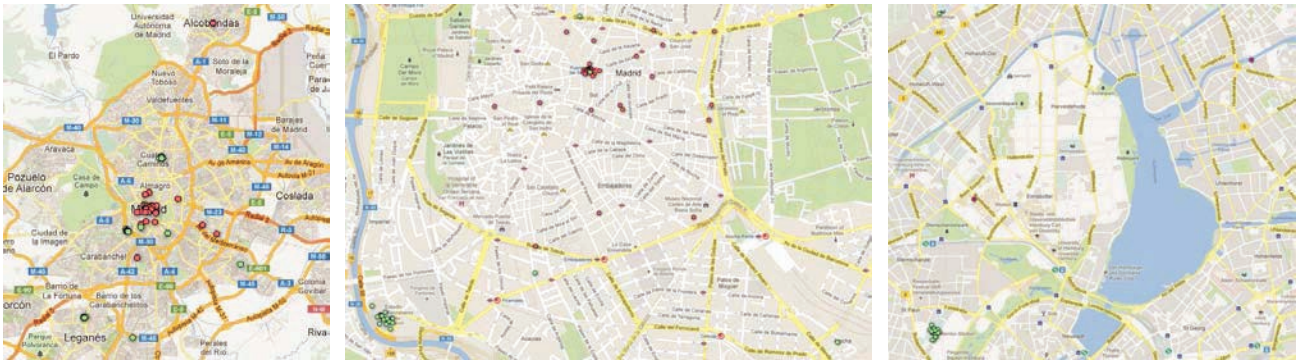
[2] S. Papadopoulos, E. Schinas, V. Mezaris, R. Troncy, and I. Kompatsiaris. Social Event Detection at MediaEval 2012: Challenges, Dataset and Evaluation. In *Proc. MediaEval 2012 Workshop*, Pisa, Italy, October 2012.

[3] S. Papadopoulos, R. Troncy, V. Mezaris, B. Huet, and I. Kompatsiaris. Social Event Detection at MediaEval 2011: Challenges, Dataset and Evaluation. In *Proc. MediaEval 2011 Workshop*, Pisa, Italy, September 2011.

[4] T. Rattenbury, N. Good, and M. Naaman. Towards automatic extraction of event and place semantics from flickr tags. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '07, pages 103–110, New York, NY, USA, 2007. ACM.

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

[6] C. Zigkolis, S. Papadopoulos, G. Filippou, Y. Kompatsiaris, and A. Vakali. Collaborative Event Annotation in Tagged Photo Collections. *Multimedia Tools and Applications*, 2012.



(a) Madrid events (low zoom) (b) Madrid events (high zoom) (c) Hamburg events (high zoom)

Figure 3: Spatial distribution of events in Madrid and Hamburg. The markers are color coded to denote the respective challenge. More specifically, magenta denotes technical events, light green denotes soccer events, and red denotes Indignados events.

(a) PHP Unconf. 2010 (b) Gamescom 2009 (c) CeBIT 2010 (d) Convention Camp 2011

(e) Real Madrid - Milan (2010) (f) World Cup 2010 (g) St. Pauli - HSV (2010) (h) Spain - Colombia 2011

(i) Inaugural march, 15 May (j) Large gathering, 20 May (k) Gathering, 15 Oct (l) Demonstration, 17 Nov

Figure 4: Example images from SED2012 events. Each row corresponds to events of the same challenge. The photos are licensed under Creative Commons by the following flickr users (user ids in parentheses): (a) Rob_L. (robertlippert), (b) gnislew (gnislew), (c) Randolf Jorberg (3gstore), (d) traukainehm (trau_kainehm), (e) Jan S0L0 (jansolo09), (f) Daniel Dionne (mrzeon), (g) jmm-hamburg (jmm-hamburg), (h) elchicogris (elchicogris), (i) ACido (acido), (j) Adolfo Chaves (63101856@N07), (k) Daniel Dionne (mrzeon), (l) @Popicinio (popicinio).