UNIVERSITY OF TRENTO - Italy

PhD Dissertation

International Doctorate School in Information and Communication Technologies
DISI - Dipartimento di Ingegneria e Scienza dell'Informazione

Event-centric management of personal photos

Eduardo Javier Paniagua Laconich

Advisor:
Prof. Fausto Giunchiglia

April 2015

Abstract

Since the last decade we have been observing a tremendous growth in the size of personal photo collections. For this reason, and due to the lack of proper automatic classification and annotation in standard album-centric photo software, users find it increasingly difficult to organise and make use of their photos.

Although automatic annotation of media content can work to achieve more sophisticated multimedia classification and retrieval if its used in combination with rich knowledge representations, it still requires the availability of well-annotated training sets to produce the type of higher-level descriptions that would be of interest to casual users. Thus, the applicability of this approach is highly unlikely in the broad domain of personal photography.

Recent developments in the media industry show an interest towards the organisation and structuring of media collections using an event-centric metaphor. This event-centric approach is inspired by strong research in psychology on how our autobiographical memory works to organise, recollect and share our life experiences. While this metaphor is backed by some early user studies, these were led before the large adoption of social media sharing services and there has been little recent research on how users actually use events digitally to organise and share their media.

In this work we first present an updated study on what users are doing with their photos in current online platforms to support the suitability of an event-centric approach. Next, we introduce a simple framework for event-centric personal photo management focused on temporal and spatial aspects and through it we describe our techniques for automatic photo organisation and sharing. Finally, we propose a platform for personal photo management that makes use of these automatic techniques and present an evaluation of a prototypical implementation.

Keywords

media indexing, event-based analysis, context processing, multimedia collection, event detection

Acknowledgements

This research was partially supported by the European Commission under contract FP7-248984 GLOCAL and FP7-287704 CUBRIK.

Special thanks to Prof. Fausto Giunchiglia, Prof. Luca Cernuzzi and my loving family.

List of publications

[P1] Andrews, P., Paniagua, J. and Giunchiglia, F., “Clues of Personal Events in Online Photo Sharing” in *10th International Semantic Web Conference, Germany: DERIVE 2011, Bonn, Germany,* 23-27/10/2011 [1]

Although there is a fair amount of research in user behaviours for personal photography supporting the notion of an event-centric approach that mimics the workings of the autobiographical memory, the bulk of the studies were done in the past, before the use of online photo sharing platforms became mainstream.

In this publication we present a preliminary study of albums shared on Flickr and Picasa. We crawl publicly available albums and process them automatically, inspecting the metadata in the title to detect references to places and dates, classifying our samples into those with references to time, to location, or both. We follow the intuition that the presence of time and location references in the title indicates that the album represents an event.

We observe references to time for 40% of the dataset, to location for more than 25% and to both time and location for more 10% of the dataset. We follow with a manual validation of a small subset to find how well each category predicts event use: 72% for temporal references, 78% for location references and a jump to 98% for combined time-location references.

Additionally, we do a different analysis of the text in the titles searching for event-related vocabulary, using WordNet as a thesaurus, to find that around half of the albums contain at least one word representing an event.

The work related to this publication is presented in Chapter 3.

[P2] Andrews, P., Paniagua, J. and Torsi, S., “Katie’s Swiss Trip: A Study of Personal Event Models for Photo Sharing” in *International Journal on Semantic Web and Information Systems (IJSWIS),* 9(3), 42-56, 2013. doi:10.4018/ijswis.2013070103 [2]

In this publication, we continue the work in [P1] investigating current user behaviours in activities related to personal photography. At first, we analyse and discuss a survey on photo-taking behaviour looking for event-centric thinking when participants capture and organise their photos. We find event-centric thinking for half of the participants when they take new photos, and about two thirds of them think of events when they organise they photos.

With this in mind, we complete the work in [P1] manually covering more of the dataset, also inspecting the actual photos in each album to decide if the album represents an event. We find that time and location are good indicators that the album is in fact representing an event: 77% of albums with time references are events, 83% for locations and 94% for both.

Additionally, to better understand the use of temporal and spatial references in the albums, we classify temporal and spatial information by granularity, finding that 60% of temporal references are years and more than 60% of the locations are towns.

The work related to this publication is presented in Chapter 3.

[P3] Tankoyeu, I., Paniagua, J., Stöttinger, J. and Giunchiglia, F., “Event detection and scene attraction by very simple contextual cues” in *Proceedings of the 2011 joint ACM workshop on Modeling and representing events, J-MRE ’11*, pages 1–6, New York, NY, USA, 2011. ACM. [46]

In this publication we show a simple algorithm to process spatial and temporal information within photos in order to detect boundaries between events in a personal photo collection. We are able to produce hierarchical clustering of events and sub-events if we know the home location of the user. To determine the home location, we process the complete collection analysing spatial and temporal metadata to calculate how frequently the owner was in each location. This lets us classify events into routine (in the home location) and special (outside of the home location) and produce event-sub-event hierarchies for the special events. We do further analysis of the each detected event regarding the density of photos through time. In this way, we are able to estimate how interesting each moment is and use this to recommend the most representative photos.

Its results have been presented as part of a co-author's thesis and I also include certain aspects of it in Chapter 4 (Section 4.2) for automatic detection of personal events.

[P4] Paniagua, J., Tankoyeu, I., Stöttinger, J. and Giunchiglia, F., “Indexing media by personal events” in *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval (ICMR '12),* 2012. ACM, New York, NY, USA, , Article 41 , 8 pages. DOI=10.1145/2324796.2324845 [36]

In this publication we revisit event boundary detection and test it against a much bigger dataset built by crawling publicly available albums from Google Picasa and then inspecting each album manually to only include those that represent events, annotating missing geo-coordinates derived from the inspection. In this way, we end up with 9548 albums from 5 users, representing personal photo collections spanning periods from 4 to 12 years, thus being able to evaluate the performance of our approach for temporal periods of different length to simulate the natural growth of a personal photo collection through the years.

Its results have been presented as part of a co-author's thesis and I also include certain aspects of it in Chapter 4 (Section 4.2) for automatic detection of personal events.

[P5] Paniagua, J., Tankoyeu, I., Stöttinger, J., Giunchiglia, F., “Social events and social ties” in *Proceedings of the 3rd ACM conference on Multimedia Retrieval*, ICMR ’13, pages 143–150, New York, NY, USA, 2013. ACM. [37]

In this publication, we propose an approach to the detection of social events by composing them from similar personal events. Next, we analyse repeated co-participation to see if it can predict the existence of a social between the users.

At first, we analyse personal events from different users looking for a spatial-temporal match that would signal that these users are co-participating on the same social event that happens. This similarity of spatial-temporal features is translated to “being in the same place and at the same time”.

Next, we test if repeated co-participation to the same events by the same pair of users can predict the existence of a social tie. We understand repeated co-participation as a measure of affinity between to users.

Even though [P5] has already been presented as part of a co-author's thesis, it is an integral part of my work and for that reason I revisit it in this document (see Section 4.3) using event co-participation as a justification for automatic photo sharing, implementing and testing this in Chapter 5.

Contents

List of Tables

List of Figures

Chapter 1 Challenges of an event-centric approach to personal photo management

1.1. Introduction

We are what we remember. Be it consciously or unconsciously, all the things that we experience and the knowledge we derive from these experiences define who we are. Moreover, we make ourselves known and get to know others by sharing experiences.

Since its invention in the 19th century, photography has been used as a way to preserve these experiences. People take photos to archive important events and share within their community (as described in [10]).

For a long time since the mass adoption of photography, personal photo collections were relatively small and required no special effort. At first, *albums* or even *shoe boxes* were sufficient to organise personal collections.

With the advent of digital photography and the cost per capture going rapidly towards zero, people started building increasingly larger photo collections. Still, the “album” kept being used as the main metaphor for helping users to organise their personal collections, thus staying close to how physical photo prints were organised previously.

Nowadays and specially due to the rise of the smartphone in the last decade, the typical size of personal photo collections has acquired extreme proportions. In most cases it can no longer be tamed with a simple *folder-as-album* approach.

However, new metaphors of organisation are now emerging to leverage more complex indexing and search in the virtual space. Flickr, for instance, has introduced a very loose organisation system, focusing on tags to group photos. In addition, with the availability of ubiquitous GPS technology, media management services have introduced the possibility to “geotag” media and to browse and search them with location-centric interfaces. Some have also introduced search and navigation services based on who is in the photo and when it was taken, using the metadata provided by the camera and advanced image recognition.

Such uses of newly-available media metadata represent a shift away from the physical photo album metaphor. However, currently available tools are still doing the transition to metaphors that are better suited to making good use of these new features.

Moreover, what is then done with the photos within a collection has drastically changed in the last years with the emergence of popular photo sharing services like Flickr or Picasa and of social networking sites such as Facebook and Instagram.

In this research, we advocate for the use of the *event* metaphor as a way to combine metadata and represent part of the higher-level intent of the users when they organise, retrieve and share their photos.

The event construct, defined according to [9] as *"something that occurs in a certain place during a particular interval of time"*, provides the human mind with a way to segment memories around points of stable spatial and temporal features.

It is this spatial-temporal invariance of events that can be applied to the analogous task of photo management and, similarly, it can be used to aggregate photos in a way that better represents how we humans process, recall and communicate our life experiences.

The focus of this work is in the development of a framework for event-centric photo management, first by shedding definitive light on how people use their personal photos, then by applying an event-centric approach with heavy emphasis on spatial and temporal features to all the stages of the life cycle of personal photos to diminish the limitations stated in the following Section 1.2.

1.2. Problem Statement

A lot of effort has been put into enabling or improving multimedia classification and retrieval. However, the typical approach requires well-annotated media collections in order to work.

Unfortunately, as Smeulders et al. [45] found, we cannot expect high quality manual annotation in personal photo collections, whose domain is on the broad end of the spectrum, for the following reasons:

- Users, who can express themselves in high-level descriptions, typically find this task tedious and avoid it completely (see [15], [48]).
- Even when high-level descriptions are available, labelling is seldom complete and often context-sensitive.
- The unpredictability and variability of the appearance of concepts increases as the domain becomes broader.

Automated annotation algorithms exist (see Lew [26], Fan et al. [11], Li and Wang [28]) that would unburden the user of the annotation task, but their applicability is often limited to very narrow domains. Moreover, the availability of well-annotated collections (in this case ground truth for training and validation) covering broad domains is very unlikely.

Some content analysis algorithms work well if the low-level features (e.g., colour, texture) they work with are highly correlated to the search goals of the particular application (e.g., the work done in [14] and [7] for detection of adult-content photography). Unfortunately, the semantic gap between the user and the system is too big for the more general photo content that can be found in personal photo management tasks, as these algorithms capture metadata that is still too distant from what users find interesting. In these cases, Lew et al. [27] found that multimedia data is usually decomposed into content segments such as shots, speaker segments, image segments, all features at a level too low to be of use to humans who are interested in higher-level content descriptions such as places, objects, people, actions.

There is also a semantic gap issue when sharing media. As described in [6] and [16], standard multimedia search engines suffer from a gap between local content and global concept, due to the diversity of context that exists between the knowledge of the user and the knowledge that the particular search engine can encode.

The work in [20] points out that the typical metadata for an individual photo are still in most cases what can be obtained automatically from the sensors of the capturing device: timestamps and, increasingly, geographical coordinates as this feature becomes more mainstream.

Finally, while event-centric metaphors are backed by some early user studies (see [41], [25], [10], [33]), these were led before the large adoption of social media sharing services and there has been little recent research on how users actually use events digitally to organise and share their media.

In light of the limitations stated this section, we advocate for the use of the event metaphor in a lightweight approach that is founded on the increasing automatic availability of spatial and temporal metadata. The aim of our approach is to resort to clear contextual boundaries that will allow for the design of photo management tasks in terms of spatial and temporal similarities.

1.3. Objectives

The main objective of this thesis is to study and apply an event-centric approach to all the stages of the life cycle of personal photos: producing, organising, retrieving, annotating and sharing. In order to achieve this goal, our work addresses a wide variety of design and technical challenges, which closely relate to the following partial objectives:

- understand how people are currently using their personal photos. Although there is a strong hint in earlier studies that points towards the implicit use of events in photo-related tasks, these were led before the large adoption of social media sharing services and there has been little recent research on how users actually use events digitally to organise and share their media;
- develop a model for event-centric representation of media content that relies mainly on capturing temporal and spatial information increasingly offered in

new mobile photographic devices;

- re-design common media management tasks (i.e., organisation, retrieval and sharing) in terms of the generally available spatial and temporal cues while aiming at diminishing the limitations seen in the previous Section 1.2;
- propose an architecture for personal photo management using the model and re-designed tasks, implementing a prototype that covers a representative subset and then perform a preliminary evaluation that sufficiently demonstrates the value of the proposal.

1.4. Structure Of The Thesis

The reminder of the thesis is structured as follows.

In Chapter 2 we do a survey of past and current studies from four different perspectives that are related to events and photo management:

- Psychology and Neuroscience describe how the autobiographical memory works to build and recall memories of past events,
- Media management tasks in Multimedia Information Retrieval,
- Event modelling,
- Events in current media management tools.

In Chapter 3 we study what users are currently doing with their personal photos to find out that frequently, users are already thinking in terms of events when they engage in photo-related activities.

Encouraged by these findings, in Chapter 4 we present a simple framework that is suited to the needs of our event-centric approach. Using this framework we then propose algorithms that will aid the user to find events within the personal photo collection and then find other users that co-participated in the same event and could potentially be interested in sharing.

Chapter 5 is where we describe a platform for a mobile media management application that exploits events as the providers of context in terms of simple spatial and temporal metadata and facilitate the annotation task. We implement a subset of the design first as a low-fidelity prototype and then as a first iteration of a mobile interface, performing preliminary evaluation to assess the value of the whole event-centric approach.

Finally, in Chapter 6 we present our concluding remarks and possible future directions.

Chapter 2 State of the art

The purpose of this chapter is to draw inspiration from current research and promising trends in relation to:

- the study of human memory, as it enables us to make useful analogies to the media management task,
- media analysis, to understand why current automatic annotation approaches are generally limited to narrow domains,
- event modelling with the latest attempts at representing event and media metadata in levels closer to what humans typically expect,
- new designs that go beyond the conventional album metaphor and exploit metadata contained in photos.

2.1. Human Studies On Memory, Events And Media

It seems that people take photos to archive important events and share within their close community. Rodden and Wood [41] and Lansdale and Edmonds [25], found clues that people intentionally classify photos according to events in their lives. Chalfen [10] has already observed such a fact before the advent of digital photography and he argues that people do not share pictures per se but use them to tell a story. More recently, the user study conducted by Miller and Edwards [33] similarly concluded that users took photos primarily to archive important events and share within their community; however, at the time of their study, they found that lay users did not share photos actively online and preferred to use prints or email.

This organisation of photos “chronologically by event” eases the search and retrieval of specific photos in personal collections as it aligns with the way human memory is structured. According to Zacks et al. [50], humans identify activity boundaries at points that correspond to a maximum in the number of changing physical features, thus aggregating memories around events. Kurby et al. [24] state that the brain operates in this way to cope with the increased difficulty brought by indexing new information when it is dissimilar from the "current moment" beyond a certain threshold. Some researchers are thus proposing event-centric models to characterise media in terms of the events they are associated with (see [19]; [22]; [16]).

Services such as Last.fm or Upcoming.org already try to link media and event, but do so for public events such as concerts or conferences, and still do not allow users to share their personal events (e.g. weddings, birthdays, holiday vacations). Fialho et al. [12] present a user study to elicit requirements for such services and interaction paradigms that help discover and enrich public events. However, their questionnaire is framed in a way that assumes that users do want to discover and share around public events. In their exploratory approach, they do not try to find out if the event metaphor is one that is useful for the user to organise their media but assume that users are interested in finding future events. The use of media in their model is thus only considered as extra information to describe an event. While this approach is interesting for public events, it does not clarify if users naturally use personal events to organise and share their personal media collections.

Zacks et al. [50] recognise the subject, actors and causal properties as the main components of the human perception of an event, stressing the importance of the *temporal and spatial* aspects to build the event structure. Casati et al. [9] also define events as having a close link to their *spatio-temporal collocation* and to the things that constitute their subject (e.g., a sparrow in the event "a sparrow falls"). Scheffler [42] studies inter-event relations and states that events may be composed of sub-events that are temporally, spatially and causally connected. Jameson and Buschbeck [21] explore some use-case scenarios to show possible ways in which untrained users may organise media in terms of events with complex *spatio-temporal structure,* but again start from the exploratory assumption that events are the central component.

2.2. Media Information Retrieval

In its first years, MIR resorted frequently to approaches from Computer Vision which focused on low-level feature-based similarity search (see QBIC [13] and Virage [3]). Other approaches (see Lew [26] and Fan et al. [11]) focused on extracting features from low-level visual information. SIFT points for images described in [32] are another example of low-level visual features. Although they contribute greatly to any multimedia indexing scheme, in isolation they are not able to capture the conceptual and contextual information that is contained in media.

2.3. Event Modelling

Löffler et al. [30] propose some practical models for events by introducing the IPTC G2 family of news exchange standards. EventML [5] is one of these standards oriented at describing public events in a journalistic fashion, although support for media is limited, and this model is close to Chalfen's [10] idea that media are only used to support a story. A set of requirements for a base model of events is presented in Westermann et al. [49] that categorises all the properties and relations of an event into six aspects: *temporal*, *spatial*, informational, experiential, structural and causal. The F event model (Scherp et al. [43]) specifically addresses most of these requirements, Shaw et al. [44] also address the *temporal*, *spatial* and informational aspects by integrating different ontological models. The Simple Event Model is proposed in (Van Hage et al. [47]) to represent not only who did what, when and where, but also to model the roles of each actor involved, when and for how long this is valid and according to whom.

The crucial role of the entities and their linking relationships has been widely recognised in different research fields such as Semantic Web (see [4] and [8]), Information Extraction (see [29]), Digital Libraries (see [40]). In [4] entities are presented as single units used for reasoning and linking within a conceptual model.

In [29], the tasks of named entity recognition and relation extraction are eased by the use of categorising entities. In [17], linked data is employed in order to support fact-finding and question answering tasks. The work in [8] on globally unique identifiers is used for data integration and the development of entity-centric applications.

In [40] an infrastructure for contextualised search in the Digital library domain is described. It is used to link different types of informational sources among the Internet (e.g., calendar directories, location gazetteer, biographical dictionaries). “Linked Data” principles and practices are described in [18]. More specifically, [18] discusses a mechanism to access data sources from the application layer in a uniform way.

2.4. Events In Photo Management Applications

MediAssist (O'Hare et al. [35]) organises digital photo collections using time and location information combining it with content-based analysis (face-detection and other feature detectors). Rattenbury et al. [39] uses time and latitude/longitude data to analyse tags and unstructured text from photos on Flickr to extract place and event semantics. VisR[1] is a smartphone application that detects events from photos and metadata available on the device. All these studies have in common the predominance of the spatio-temporal aspect of events as it is the one facet that helps users in determining inter-event boundaries, recollect their memories and find their media. Thus, the accepted representation of events is close to Jain's [19] definition: “something that occurs in a certain place during a particular interval of time”; however, there is still confusion between news-worthy, global events (e.g. the football world cup), and personal events (e.g. birthday, wedding). iPhoto[2] and, lately, Photos for Mac, are able to index a personal photo collection in terms of events using temporal boundaries. After the indexing of events, the use is similar to that of albums. Carousel[3] uses temporal and spatial metadata to cluster photos, and lets the user share these clusters through “conversations” with other users. Still, the notion of events is at best implicit, and it lacks any automatic or semi-automatic way of selecting other users that could be potential co-participants to the underlying event.

1 VisR: http://ngs.ics.uci.edu/visr-is-in-android-marketplace/
2 iPhoto and Photos for Mac: https://support.apple.com/kb/PH15130?locale=en_US
3 Carousel: https://carousel.dropbox.com/

Chapter 3 Clues of events in the use of personal photos

While reviewing the State of the Art in the previous chapter, we found out that event-centric metaphors are backed by some early user studies (see [41], [25], [10], [33]). The problem with these studies is that they were led before the large adoption of social media sharing services and there has been little recent research on how users actually use events digitally to organise and share their media.

Consequently, to build a stronger case for events in current personal photo management tasks we see the need to perform further studies to:

- find out what users are currently doing when engaging in photo-taking activities and
- determine if they are thinking in terms of events when they share their media online.

This section features work first published in (Andrews, Paniagua and Giunchiglia, [P1]) and later on extended in (Andrews, Paniagua and Torsi, [P2]).

3.1. A Survey On Photo-taking Behaviour

In order to better understand what users are currently doing with their digital cameras and smartphones, we conducted a study with a total of 40 participants. They were asked to fill in a questionnaire with general demographics (age, gender, work and leisure activities), as well as more specific questions on media-related topics: which photo-taking devices they owned, what kind of subjects they preferred to photograph, how often, and how they proceeded to organise and share these photos.

The questionnaires were distributed during two consecutive Saturdays in March 2012 at the main entrance of the Trento municipality library, located in the city centre. This gave us access to the heterogeneous population borrowing and returning books, CDs and DVDs during the only day of the weekend in which the library is open. About 50 questionnaires were distributed, with a drop out of 20%, for reasons of non-interest, not taking pictures, or without saying the motivation. Interestingly, most dropouts were older users. Therefore 40 fully filled questionnaires were gathered.

The age spans are distributed as following: one participant under 15, seven participants between 15 and 20, twelve participants between 21 to 30, nine people between 31 and 40, eight participants between 41 and 50, two between 51 and 60, and finally one between 61 and 70. The interviewed participants had varied levels of education: ten of them have a master degree, eleven were students at the university, but also nine were aged 21 or older with a high school level while six only had an elementary schooling. The genders were almost evenly distributed, 19 females and 21 males. Most of the participants make use of computers, only nine over 40 of them claimed they never used computers.

We have identified pointers to event-centric thinking by analysing the situations in which the participants take photos, either with a smartphone or with a dedicated digital camera (see Figure 3.1). For smartphones, users mentioned that they took photos in occasions such as “holiday trips”, “nights out with friends” or “birthdays” on 13 times, which correspond to situations that can be categorised as events (casual and planned events combined). For digital cameras, we see an increase, with 31 mentions of event-related situations. In both cases, participants frequently mention that they also take casual photos of kids and pets, outside of specific events. This is also observed in our larger scale data collection of Picasa/Flickr albums (see Section 3.2.2) where we can find occurrences of aggregating albums such as "Sue - 9 Month" or "My Family". Unsurprisingly, almost all users with young kids (four of five with kids up to 13 years old) mention explicitly that they take photos of them.

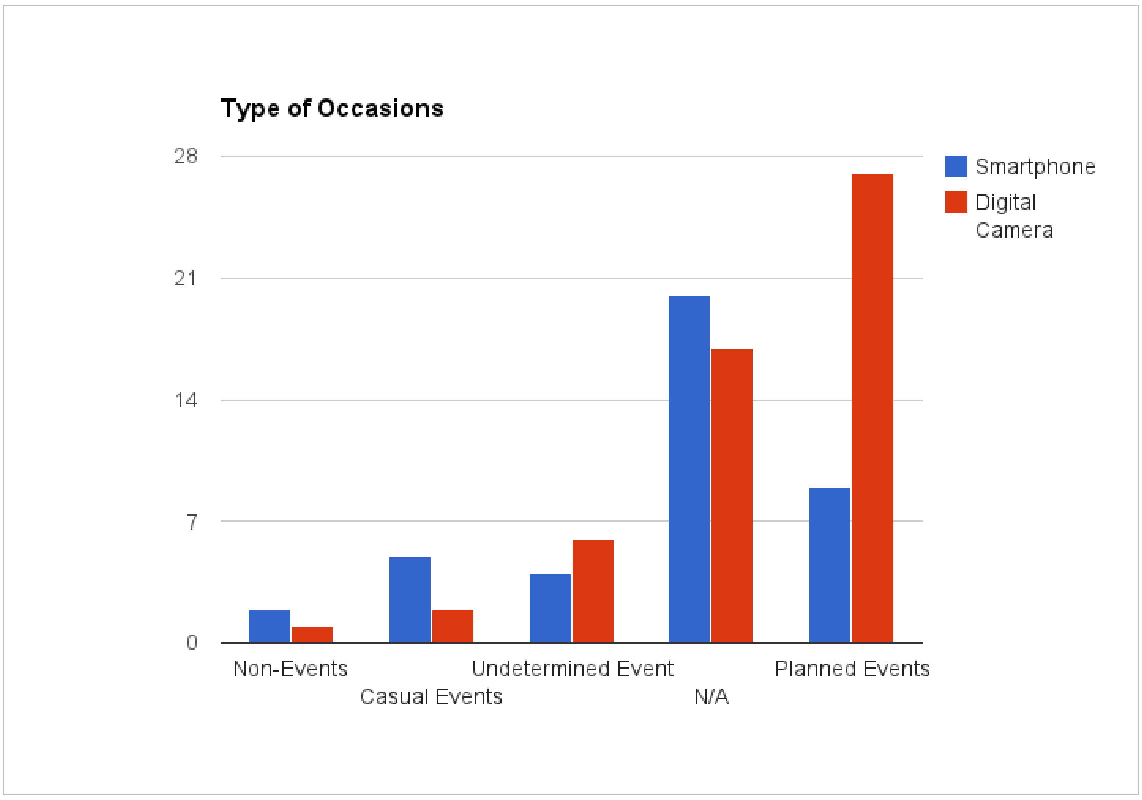

Figure 3.1: Types of occasions for engaging in photo-taking activities (number of mentions in the survey).

Another way to see if users think of their media in terms of events consists in determining if they classify their photos explicitly by events or implicitly by event properties, the most important being time and space. We found that 15 participants answered that they classify their photos in terms of events explicitly; of the ones that do not use events explicitly, eight use dates, two use places and two use both date and place. The rest either chose not to answer (eight), avoid any classification (three) or do it by other methods that not necessarily relate to events (one). We thus see that 67.5% (31 participants) refer to events implicitly or explicitly when organising their photos. In the following section, we show with a large-scale analysis, that we can find a strong connection between photos describing events and the use of dates and places in their organisation by computer users.

3.2. Use Of Personal Events When Sharing Photos

Event-centric services such as Upcoming.org or Last.fm are focused on public events such as concerts or conferences. While datasets (Fialho et al. [12]) based on these websites already provide samples of media organised around event metadata, they do not represent personal events. That is, media of more personal events, such as a birthday or a holiday, are not shared on these websites. However, this kind of media can be found on photo sharing websites such as Flickr and Picasa where users share photos of personal happenings with their family and friends.

These websites do not provide a way to organise photos around events but provide a way to group photos in albums. These albums can only have a very small amount of metadata and are not presented as events to the users. On Picasa, albums can have a title and a description, and optionally a date and a location; on Flickr, sets can only have a title and a description.

We are interested in seeing how users describe albums they share on Picasa and Flickr by using the title and description fields. Our hypothesis is that if they share media related to events, they will provide the event metadata in the fields that are available to them and we will find event references in the titles and descriptions of the albums. We are focusing on these two social sharing sites as they are some of the more popular available at the time of writing; while Facebook is also very popular, it provides very similar features (album based organisation of photos) and does not allow data collection.

3.2.1 Data collection

We have thus collected a dataset of digital albums shared on Flickr and Picasa. To select users, we use the “explore” pages of each website that feature randomly selected photos; from these photos, we find a set of random users and collect all public albums that are shared by these users. For each album shared on Flickr we retrieve:

- the title of the set,
- the user identification,

- the URL of the set and
- the number of photos and videos within the set.

For each album from Picasa we collect:

- its URL,
- the date specified for this album,
- the number of photos,
- the title,
- the description and
- the user identification

As it is difficult to decide automatically if an album is an event, we have manually annotated a subset of the Picasa dataset, with the help of ten human annotators who have annotated 3,692 albums from Picasa. Each annotator was shown a Picasa album, it's title, description and all of its photos and had to answer the following questions about each album:

- the language,
- the number of date references contained in the text of the title,
- the number of location references contained in the text of the title and
- if the album represented and event. This was not just based on the title, but also on the content of the photo as evaluated by the annotator.

While human annotators were able to identify events, location and dates with accuracy, they cannot easily produce a large set of data, we have thus automatically processed the rest of the Flickr and Picasa datasets to extract the same information as the ones provided by the human annotators.

Because both websites are international, many entries are not written in English. Currently, we are only able to process metadata provided in English and thus want to filter out the other languages. The Perl Lingua::Identify[4] module was used to identify the language of the title and description (when available) in each album entry. The algorithm provided by this module was trained on the EuroParl [23] corpus; based on the manual annotation of the Picasa subset, we have found that the algorithm labels English albums with a precision of 89.0%.

We are interested in seeing if users refer to locations when they describe albums and have thus automatically processed the dataset to find references to geographic locations. The Yahoo! Placemaker[5] service is used to perform this task. This freely available geoparsing service can identify place references in unstructured text. While Yahoo! does not provide information on the accuracy of their algorithm, from our manually annotated sample, we found that Placemaker is able to detect if there is at least one location reference in an English title with 81.2% accuracy.

References to dates are also of interest to us as time is a main attribute of an event. To detect such references, we analysed each title with a custom date parsing algorithm that detects full dates but also partial dates (e.g. “Paris’08”) and date ranges (e.g. “40.5 miler in Sespe Wilderness April 2nd - 5th 2010”). On our manually annotated sample, this algorithm performed with 88.1% accuracy.

We have collected and automatically processed 24,800 sets from Flickr and 62,681 albums from Picasa over the month of July 2011. Randomly crawling the sites provided us with albums posted between 2006 and 2011. As mentioned previously, because of the automated algorithm and of the manual validators main language, in the rest of this article, we are only looking at albums that were annotated as using the English language, resulting in 5,339 (21.5%) sets from Flickr, and 11,355 (18.1%) albums from Picasa. Figure 3.2 shows the main languages detected in both datasets.

4 http://search.cpan.org/~ambs/Lingua-Identify/

5 http://developer.yahoo.com/boss/geo/

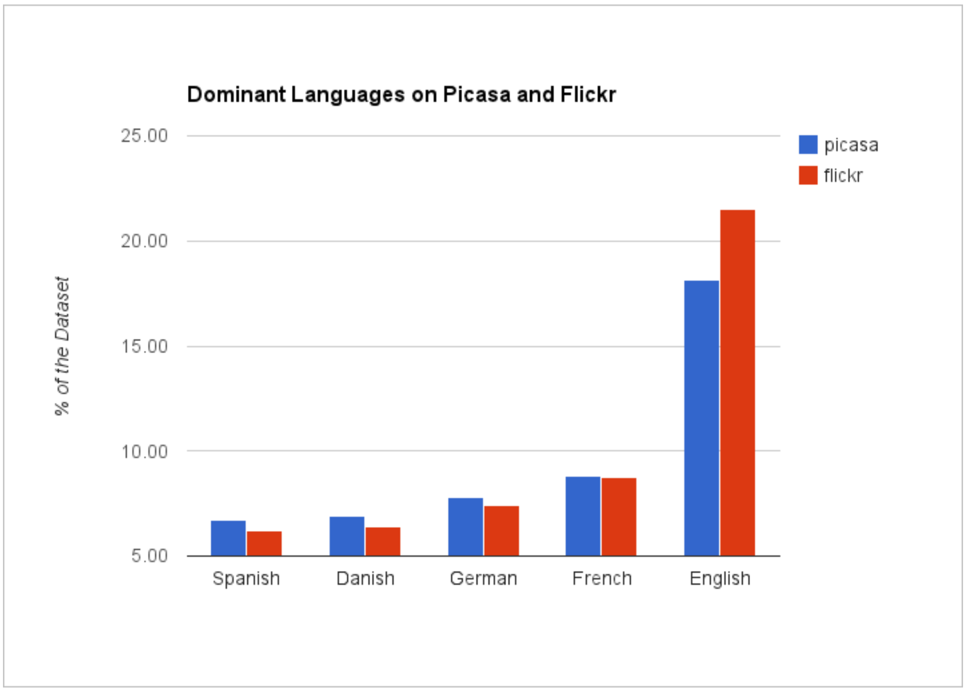

Figure 3.2 Top Languages on Picasa and Flickr

3.2.2 A given place and time

According to the state of the art definition presented in [9], the two main attributes defining an event are its location and when it has happened. Thus, if users are to describe events using albums when sharing their photos, they will probably specify some of these metadata within the available attributes. We found that in the Picasa dataset, only 31% of the albums have a description and thus, in this paper, we focus on the title attribute of the albums, as we do not have enough data to draw conclusions from the descriptions.

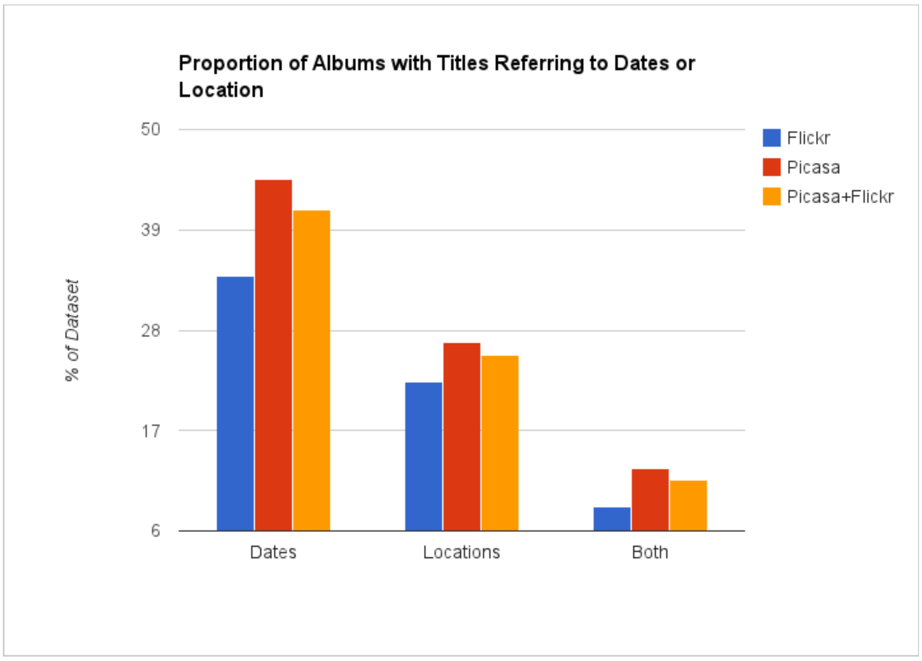

Figure 3.3: Proportion of Albums with Titles Referring to Dates or Location.

Figure 3.3 shows the proportions of albums where date or location references can be found, a test of equal proportion shows that Picasa and Flickr are comparable (p<0.01) and we thus consider that there is no difference in users' behaviour between the two services in the factors we analyse.

The number of albums where an explicit date reference can be found in the title makes for more than a third of the dataset. We can thus see that people do like to share their albums with metadata about the date when the photos were taken. Note that while the users set the title manually, the *date* field on Picasa is filled automatically with the album creation date if the user does not specify any value explicitly. When there is a date in the title on Picasa, it is often not consistent with the album *date* field. It seems that while users are ready and interested to share their photos around dates, they are not motivated to fill in an extra metadata field. The reason behind this might be a limitation in Picasa's interface or it can simply be because the users do not see the gain in filling this extra field.

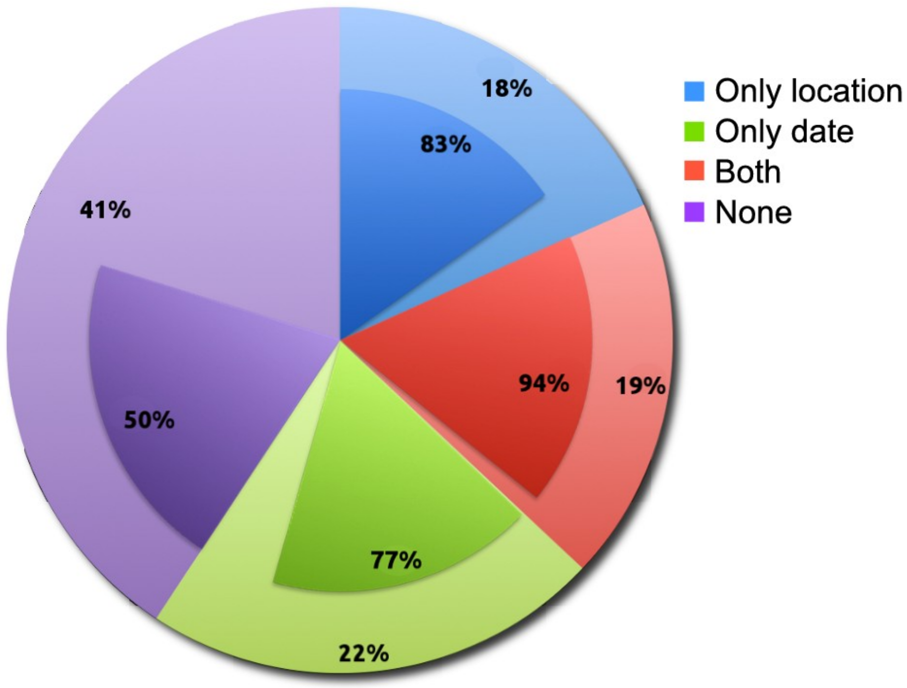

Figure 3.4: Analysis of the validated set for Picasa. The inner slices show the percentage of albums in each category that are about events.

While the date is an important attribute of events, albums with only a date reference are not always events according to our previous definition. In fact, from a manual annotation of the Picasa dataset that can be seen in Figure 3.4, we find that around 23% of the albums with a date reference, but no location reference, are not really events. This is because there are catchall albums for entire years or months, where users put photos of many different events in the album (e.g. "Misc. Apr. 2009"). The album is thus only a way to aggregate photos in a time range and not used to represent a specific event. This happens also when people share photos of their newborn child for milestone periods (e.g. "Jake - 9 months: March").

There are fewer albums with an explicit location reference, but it still makes for a fifth of the dataset. From the manual annotation, we can see that approximately 83% of the albums with only a location reference are actually events. In the same way as with the dates, users use locations for *catch-all* albums where they put photos of a location they visited multiple times but not for any specific event (for instance photos of their home-town).

In these two cases, we can see that the dates and locations are sometimes used only as aggregators for media that could be replaced by automatic metadata based services. However, it seems that the users are not aware of, or willing to use, these services on the studied websites.

94% of the albums with a date and a location together were annotated as being events by the manual validators. While these albums represent a small amount of the dataset, we can already see that when space and time are specified in the title, the users wanted to share an important event.

3.2.3 An event vocabulary

In the previous section, we have looked at how users might use album attributes to describe explicitly an event location or date. However, there are many events represented on Picasa and Flickr that do not include explicit dates or locations. For instance “Janet and Ian’s wedding”, “father’s day”, “Michelle’s shower” or “Christmas Eve” are all titles of albums from our dataset that do represent important personal events with no explicit dates or locations. Thus, there might be more albums in this dataset that represent events than the previous section’s analysis hinted.

In fact, if we look at the most popular words used in the titles (see Figure 3.5), many of them are references to events (e.g. “party, “wedding”, “trip”) or time periods, without having explicit dates. Note that, while not shown in Figure 3.5, the most popular words in the vocabulary are years, in fact on Flickr, 11.0% of the vocabulary are numerals while on Picasa 17.5% of the words used are numbers. Figure 3.5 reports values in per-thousand, while the distribution of the vocabulary follows a very steep long-tail curve, the most popular words still do not cover a large part of the album vocabulary.

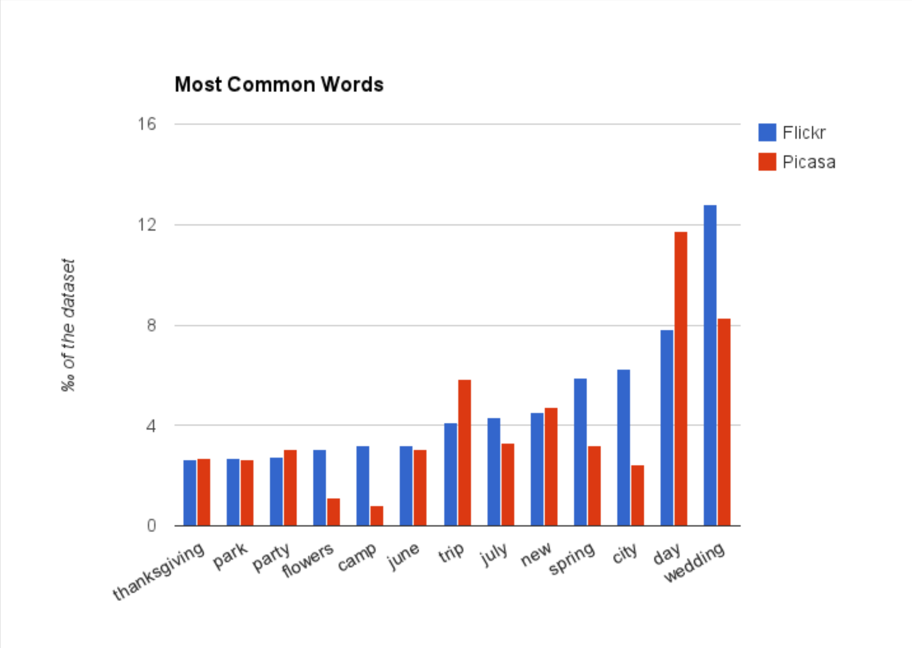

Figure 3.5: Most Frequent Words on Flickr and Picasa (‰ of the datasets).

While we can see in Figure 3.5 that there are concepts representing events amongst the most frequent words in the dataset (e.g. "wedding", "thanksgiving"), it is not a complete view of the dataset as people may refer to events with a very large set of terms. It would be interesting to see how many albums refer to events, however, manually, we cannot exhaustively list the whole vocabulary that could be used to refer to events. We take a semi-automatic approach, using WordNet [34] as a thesaurus, to find all terms that might refer to a concept representing an event. To do so, we have listed all inherited hyponyms of the synset `event#n#1`[6] – which include the words “wedding”, “birthday”, etc. – and of the synset `calendar_day#n#1` – which include the words “Christmas”, “Thanksgiving”, etc. This provides us with a list of 11,092 words and 14,304 “concepts” combined in 15,389 word-concept pairs[7] that we then searched in the titles of the albums in both Flickr and Picasa datasets.

6 In WordNet, synsets represent set of synonyms representing the same concept. The notation event#n#1 is denotes the synset representing the first sense (1) of the noun (n) with the textual form “event”.

7 Note that because of homography, the same word can appear under different concepts.

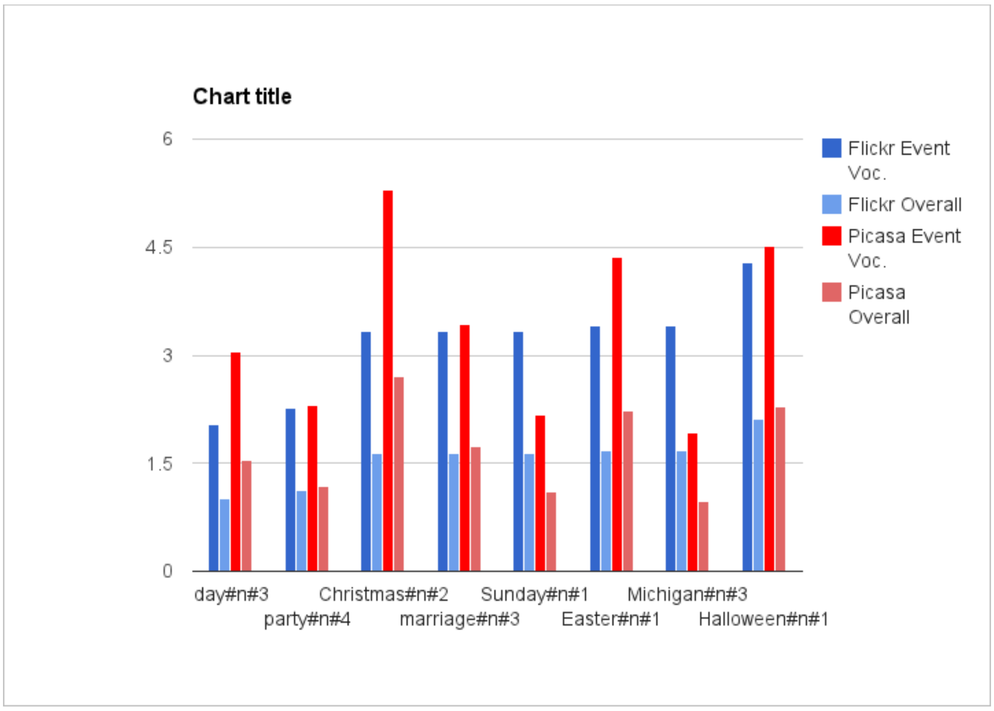

Figure 3.6: Top Leaf Concepts Related to Events in Picasa and Flickr (% of all English albums with an event related word and % of all dataset).

In Figure 3.6 we found that around half of the albums (Flickr: 49.4%; Picasa: 50.9%) have a title with at least one word that represents an event according to WordNet. Of these albums, only 29.6% have a date or a location (or both) in the title. There are indeed many albums that describe events without providing either an explicit date or a location reference (e.g. “Katie’s Swiss trip”, “Field trip - Farm”, “Lily fathers Day”). From a manual analysis of some of these albums’ titles, it seems that many of them either refer to the participants[8] (e.g. “Annete's Prom”, “Father-Daughter Dance”) or to relative dates (e.g. “Father’s Day”, “My Birthday”, “Chloe 1st Day of School”) and locations (e.g. “Trip Home”), which are hard to detect automatically without advanced natural language processing. In fact, we can see in Figure 3.7, that the `day#n#3` and `calendar_day#n#1` synsets are among the most used. This is in line with Jain’s [19] definition of an event: “a significant occurrence or happening, or a social gathering or activity”. However, relative location or participant references are hard to detect automatically and further work is required to check how these are used in the

8 the third important attribute of an event.

album vocabulary.

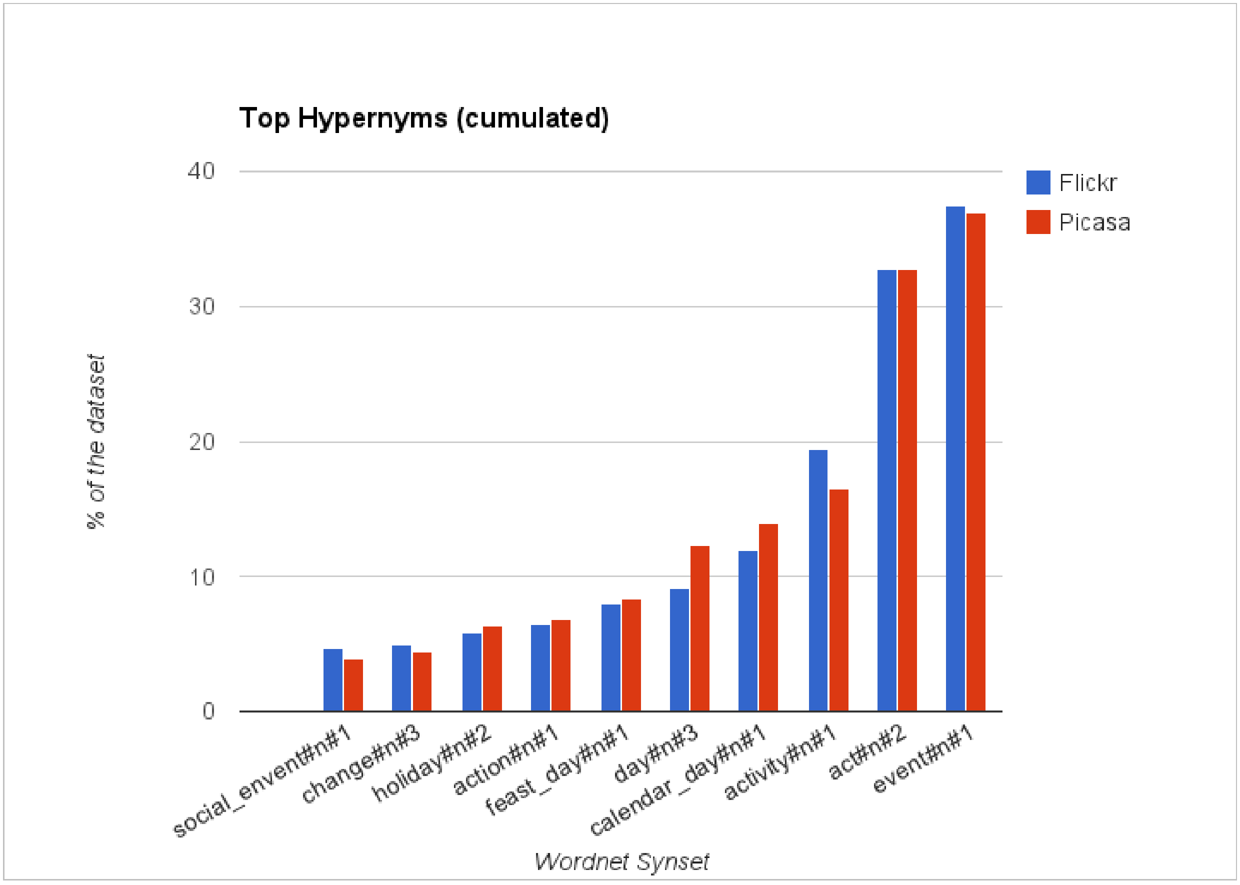

Figure 3.7: Top Concepts Related to Events – cumulating the hyponyms occurrences

WordNet is a very detailed vocabulary and many terms that it declares as relating to the *event* concept might not be used by the users to refer to events. Indeed, there is ambiguity in the vocabulary and we have taken a naive approach where we count the occurrence of all possible words without applying disambiguation (for which we do not have enough context). For instance, `Michigan#n#3` appears as one of the most popular leaf concepts for Flickr; however, this concept represents a card game called “Michigan” but might have been used by users in their album title as the location. The other top concepts however represent less ambiguous event references.

This confirms Chalfen’s [10] conclusions that people like to take photos around personal events that they then share with a community made of close relations. However, as we have discussed earlier, these photos are usually shared without description, and thus Chalfen’s hypothesis that people use photos to tell a story might not be exact on photo sharing websites.

3.2.4 Granularity of the metadata

With digital cameras, photos often have embedded metadata describing with precision when they were taken[9]. In addition, when photos are taken with modern smartphones, they are now also tagged with GPS coordinates and more and more customer “point and shoot” cameras are also integrating this feature. The creation of event related metadata is thus almost fully automated, but as we have discussed earlier, the user rarely remembers of a photo by its exact date and time, but by an event representing their experience at the time of taking the photo.

However, some form of date and time defines an event, and it is interesting to see how users use such dimensions to describe their events. While we do not have direct access to the users for the study we report here, we can extrapolate on their habits from what they do on Flickr and Picasa, and we can see which level of precision is used when referring to personal events.

By definition, an event takes place during a specific duration of time and in a particular location. Clearly, this will not correspond to a timestamp exact to the second or to a GPS point defined within a few meters.

We have thus tried to study the granularity of references to dates and locations used by the users. To do so, we have analysed the information automatically extracted from our dataset to see the precision of the location and date instances that were found in the Flickr and Picasa album titles.

It is relatively straightforward to extract the precision of dates and as we can see in Figure 12, we have three levels of precisions:

- Dates referencing a particular day. For example: "26 March 2011, TUC

9Assuming that the time is set properly on the camera.

March for the Alternative and fringe", "FBU strike, and march against the cuts, Saturday 23 October".

- References to a specific month and year. For instance: "Catalonia and Spain-May 2011".
- References only to a year. For instance: "Quebec City 2006".

In some rare cases (3.3 %), we were unable to decide automatically in which category to classify the date reference. Figure 3.8 shows the proportions of each category in the full Picasa and Flickr datasets. There isn't a significant difference between the two datasets and the dataset of manually validated Picasa events.

We can see that there is a majority (59.5 % combined) of the date references that refer to only a year, users preferring a low level of precision. This, as we have seen in the previous section, happens mostly because of the predominance of recurring -- or once in a year or in a lifetime -- events, such as "Christmas 2012", "Hanshew Wedding 2007" or "Dave's 21st Bday 2008". The full day mentions, with 22.4 % of the date references extracted are just slightly over month only references (18.1 %).

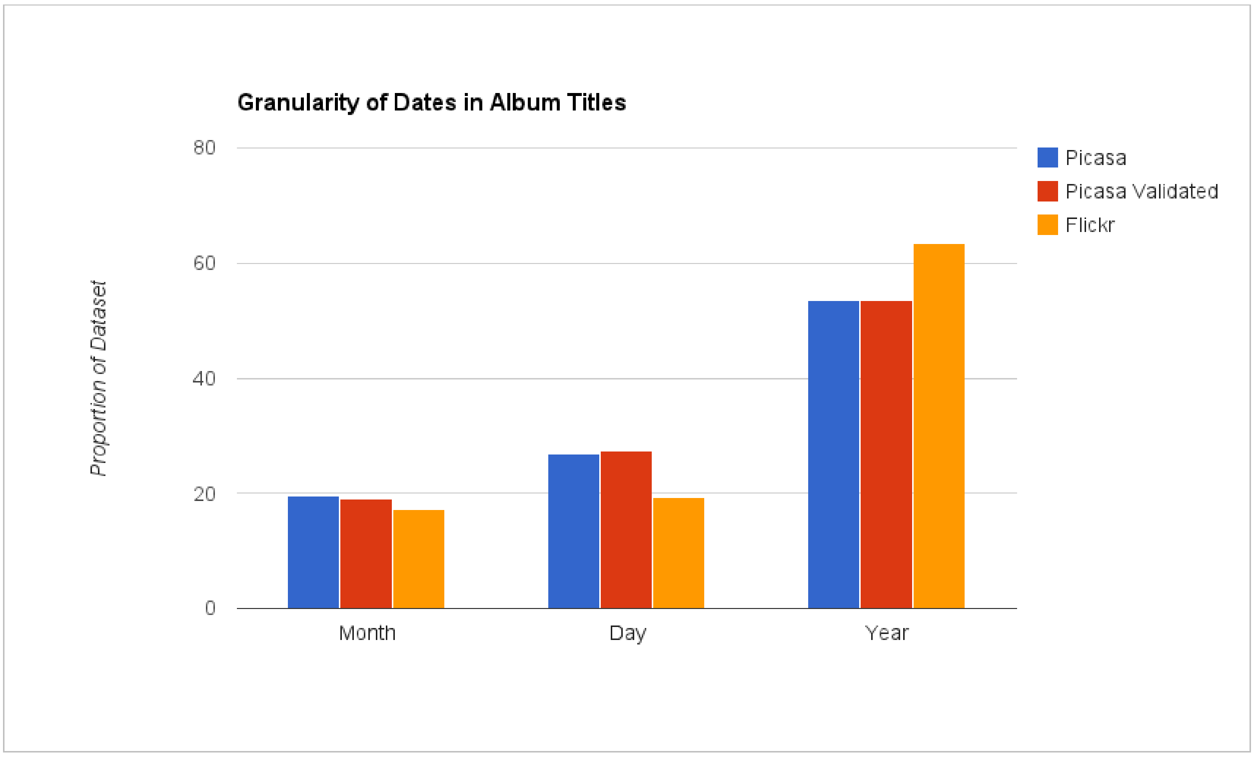

Figure 3.8: Granularity of Dates in Album Titles

In a similar manner to the date granularity, we are interested in seeing with what accuracy users refer to location when describing events. This is not as straightforward to perform in an automated way as there are many different types of location references in the album titles. However, we have already used Yahoo PlaceMaker[10] to extract the references to places in the titles (see Section 3.2.1); this one, in addition to extracting the entity references provides a guess disambiguation to a Where On Earth identifier (Yahoo GeoPlanet, 2013) that can be resolved to location metadata based on the Yahoo GeoPlanet[11] location hierarchy. Yahoo GeoPlanet contains references to six million places around the globe at different levels of granularity and can provide metadata about these places. In particular, we are interested in seeing what types of

10http://developer.yahoo.com/boss/geo/

11http://developer.yahoo.com/geo/geoplanet/guide/concepts.html#placetypes places are used to describe albums in the datasets that we have gathered.

Figure 3.9 illustrates the distribution of the geo-references provided in text by the users in the Picasa and Flickr dataset. We are using the place categories provided by Geoplanet, and we can see common types such as *Country* or *Town*. Geoplanet also provides more ambiguous classes, such as *Colloquial* (e.g. "Western US", "Napa Valley"), *Point of Interest* (e.g. "Legoland", "The London Eye") or *Land Feature* (e.g. "Sleeping Giant State Park", "Grand Canyon").

We can see that there is no significant difference between the use of geo-references in the dataset of validated events and the other datasets. *Town* makes for the majority (57.9% overall) of the dataset and *Country* comes second with 16.4%, followed by *State* at 6.1% and other less important types all under 5% of the dataset.

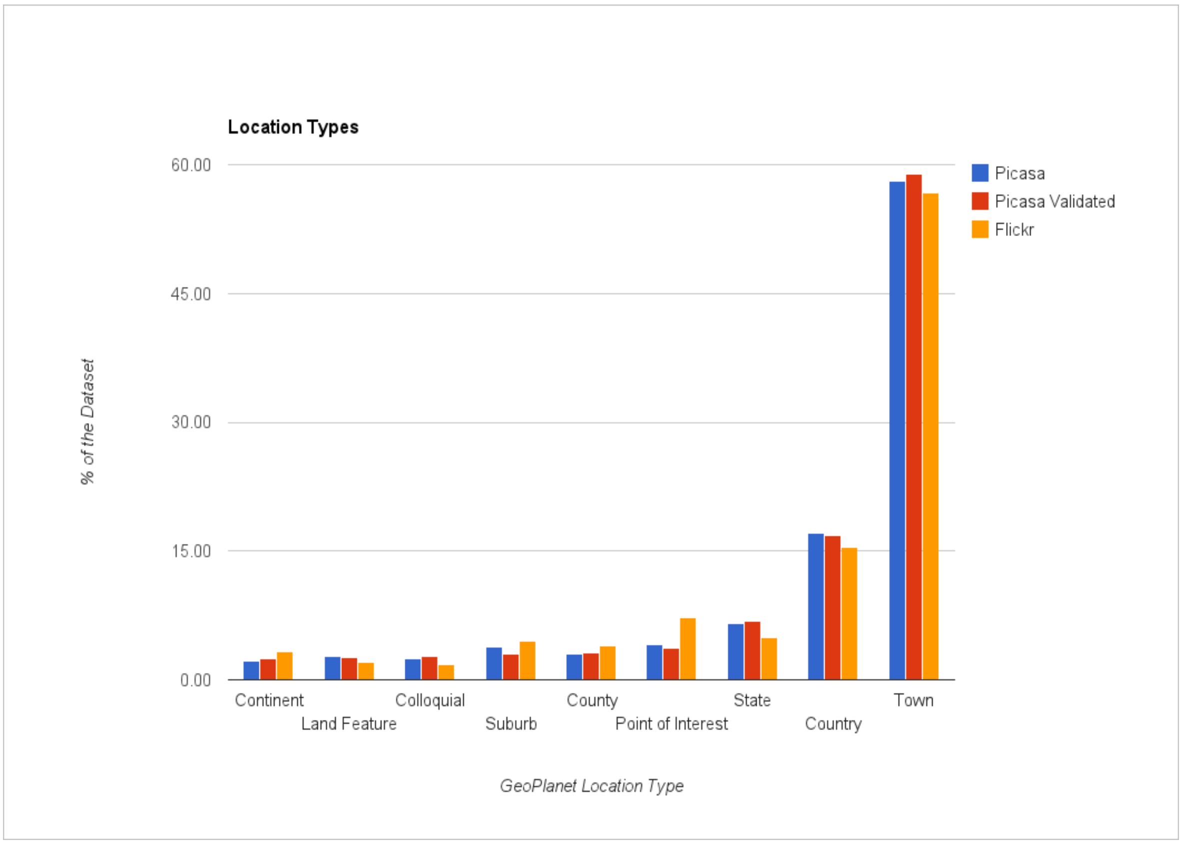

Figure 3.9 Top Types of Locations in the Datasets

3.2.5 Discussion

The results we describe in this article show that people sharing photos online have a tendency to organise their photos around dates and locations. While this is not a guarantee that they are sharing albums about specific personal events, it seems to align with the previous observations of Zacks and Tversky [50] and Kurby et al. [24] who found that users like to segment their memories around time and space.

While most semantic event models discussed in the state of the art (for instance Löffler [30], or Shaw et al. [44]) represent events around dates and locations too, they do not seem to fit perfectly the behaviour of the users that we observed on the sharing sites. In particular, some users seem to aggregate media around date or location without describing events (e.g. the child growth album examples pointed out earlier). While this could be done automatically from the metadata of the photos, there might be a higher semantic to this grouping when sharing. As Chalfen [10] discusses, even if his research was for printed photos, people group photos together to support a story and not always just for the content of the photos per se. That is, the grouping of photos of the “second month” of a baby is not a specific event according to most of the existing metadata models but is still an event of importance for the users that share them online.

In addition, in accordance to Chalfen [10] and to Miller and Edwards [33], people share photos around important personal events. These events (e.g. Christmas, trips, visits) are not always global events and their scope is limited to the close circle of personal relationships. This kind of sharing has probably a different purpose from the one of exploring concert or conference photos (for instance) as described by the LODE ontology (Fialho et al. [12]; Shaw et al. [44]), or from the news outlet use-cases for which the IPTC standards have been developed (Löffler [30]). These semantic models of events also consider that location and date are given with an exact precision, making it difficult to afford the rough mental representation of events' metadata that we have observed (see Section 2.3). We therefore believe that we need custom semantic models and services that would be more adapted to the organisation and sharing requirements of the lay user.

As was pointed out by Miller and Edwards [33], there is also a stronger issue of privacy and access control when dealing with the sharing of personal events. On Picasa and Flickr, we were able to crawl public albums – featured on the website main pages – that were of highly personal nature but are accessible to anyone online. While this is not the scope of this work, we believe that there is a need for better privacy services directly integrated with the event models to deal with the personal media sharing use-cases.

3.3. Conclusion

In this chapter we present results from a survey of camera users and a study of a dataset of albums shared on Flickr and Picasa.

Even though the reasons and the ways people approaches to photos can be very different, we have seen that the surveyed photo takers tend to use an event mental model to organise and share their photos and we believe that allowing event-centric media management would thus be more consistent with the cognitive framing of experience of users. Event-centric models are often embodied in the media organisation state of the art, and in social media applications, in order to support the concept of experience recollection and sharing. We have seen that, following assumptions of the state of the art in event modelling, space, time, people and type of activity are the main different facets converging into a single event.

To confirm the conclusion of our survey, we have conducted a larger scale experiment on Picasa and Flickr, two popular photo-sharing services. We have found that while these two services have different interfaces and features, none of them directly support events as a metaphor to organise and share media. However, on both websites, users tend to divert the organisation system provided to group their media around events. As we have observed this behaviour in large scale on two very different websites, we believe that this is some general intent of the users more than site-specific behaviours biased by their interfaces.

We have found that a significant amount of users share media online illustrating personal events, and use time-location metadata to describe them. In fact, we have found that more than a third of the albums shared reference a date in their title and more than a fourth refer explicitly to a location. Users also seem to group their photos around important personal events (e.g. birthdays, wedding, festivities) without always specifying explicitly a location or date.

Chapter 4 Events for personal photo management

In the previous Chapter 3 we found out that users frequently think in terms of events when they engage in personal photo management tasks. These tasks include typically photo capturing, organisation, annotation and sharing.

Our key intuition is that by leveraging the contextual information represented as spatial-temporal metadata contained in photos, we can describe typical photo management tasks using an event-centric framework to provide automatic or semi-automatic aids.

To that end, we present in the next Section 4.1 an event-centric framework that will be used throughout the rest of this document. In this framework we describe the photo, its contents and context in terms of entities and relations between entities. In this work we only deal with the Event as the preferred way to achieve photo organisation but it could be easily extended to also include the Album to represent photo aggregations that are not events.

In Section 4.2, we test the suitability of our event-centric approach for the organisation of large personal photo collections.

Finally, in section 4.3 we take personal photo collections of different users and analyse how temporal and spatial metadata can be of help in the general task of sharing photos with other users that may be interested.

4.1. A Simple Framework For Event-centric Photo Management

The proposed model for our framework (presented in [16]) uses entities to model tangible or abstract objects. Each entity has a set of attributes, relations to other entities (to represent relations such as *is-a* or *part-of*, among others) and services that can be performed by the entity as sort of abstract class or on a particular instantiation.

In the rest of this section we describe the essence of the model with all the entities (including their attributes, relations and services) that our entity-centric framework comprises.

4.1.1 The generic Entity

This generic entity type is abstract (direct instances are not allowed) and encodes essential information in order to uniquely identify each instance. It also encodes the basic services for CRUD operations that all instances must perform.

Attributes

- *uri*: a unique identifier assigned to the entity at creation time.
- *name*: a rich string (that can contain references to entities within, and as such, aligns conveniently with named entity recognition approaches such as the one in [38]) representing the label given to the entity. Examples of an entity name can be:
  - the simple “Trento” string naming a Location,
  - the rich string “trip to →[Caldonazzo]” in which the “→” indicates a external reference and [Caldonazzo] an instance of the location named “Caldonazzo”.
- *description*: a rich string to describe the entity in greater detail.

4.1.2 Entities for content description within the photograph

Object

The *instantiable* version of an Entity that encodes information about a real world object. It inherits the definition of Entity.

Person

A specialisation of Object to represent people (e.g., event participants) and in particular, to model the user within the photo management task.

Attributes

- *first_name, last_name*: that supersede the more general *name* of Object entities.
- *friend*: a relational attribute (often a set) linking to other Person entities to model friendship-like relations. For the purposes of this thesis we assume all Person-Person relations fall in that category.

Location

Location is a specialisation of Entity that refer to all of those entities for which the spatial dimension is significant. It models spatial objects that occupy regions of space (e.g., cities, natural-bodies, buildings).

Attributes

It inherits all attributes from the generic Entity and includes the following:

- *latitude, longitude, altitude*: all according to the WGS84[12] standard.
- *category*: an identification of the type of location (e.g., a city, a country, a well-known city landmark).

4.1.3 Photograph-related entities

Image

An image is a two-dimensional, static representation of a visual experience. This entity type encodes the low-level representation typically understood by computational machines and correspond to the typical concept of image files captured by digital cameras.

Attributes

- *bitmap*: the quantised two-dimensional representation of light in terms of pixels
- *timestamp*: registered by the photo-taking device at creation time

12 World Geodetic System 1984: http://earth-info.nga.mil/GandG/wgs84/

- *latitude, longitude, altitude*: geo-location information from GPS sensors (in the same format of their counterparts in Location entities)
- other information from sensors (e.g., orientation, tilt)
- *photo*: a relational attribute linking the Image to a corresponding Photo entity (see the next Photo entity description)

It is important to notice that depending on the sophistication of the device, higher-level metadata can also be automatically annotated at the moment of producing a new photograph (e.g., detected faces, camera parameters interesting to photographers). Nevertheless, for the Image entity we are only interested in lower-level metadata.

Photo

The Photo entity complements the Image entity previously described. In photo entities we represent what users are typically interested in when looking at an image: the location depicted in the photograph, faces, objects, the event being depicted.

Attributes

- *image*: a relational attribute linking the Photo to its corresponding lower-level representation
- *region-of-interest*: a structured attribute with
  - a region (e.g., a bounding box) that corresponds to the area in the image bitmap that has something deemed interesting by the user,
  - a semantic string: supporting the reference to any type of entity, typically Person, Object, Location. In the case of a Location, it is frequent to find a correspondence between the geo-location information found in the Image, but this may not be the case if the Location depicted is sufficiently far from the position the user was at when taking the photograph.
- *organised-in*: a relational attribute to link the Photo to an Event.

4.1.4 Entities for photo organisation

Event

A specialisation of the albums, an Event adds well-defined spatial-temporal context to the set of photos within. In this regard it follows closely the definition in [19] and model our Event entity to answer the What, When, Where and Who aspects.

Attributes

- category: the category to which the event belongs (e.g., Birthday, Graduation, Football Match) to complement the Event title (What?).
- participant: a relational attribute (often a set) that links the Event to its Person participants (Who?).
- temporal_period: to encode the temporal aspect of the event (When?)
- location: a relational attribute linking to a Location entity to model the spatial aspect (Where?) of an Event.
- sub-event: a relational attribute (often a set) linking an Event to another, smaller (sub-) Event, reflecting the compositional aspect.
- parent-event: the inverse of the relational attribute above.

4.2. From Low-level Spatial-temporal Metadata To Hierarchical Events

From the definitions and the study of photo albums in Section 3.2 we know that the most representative aspects of events are *time* and *place*. This is why, when doing manual organisation into events, users are indirectly aggregating photos around points that represent “a certain place, a particular interval of time” (see [9]); that is, around points in which spatial and temporal features remain stable.

In this section we explore different approaches to bridge the gap between low-level temporal and spatial features contained in images and the higher-level encoded in events.

Section 4.2.1 presents a discussion of a well-known approach from Loui and Savakis [31] in which we briefly describe it to later state the successive steps to move from lower-level timestamps to higher-level events. Next, in Section 4.2.2 we take [31] and build on it to achieve hierarchical organisation. This hierarchical organisation is possible because we process the photo collection to categorise each detected event as being either routine or special. For special events we do a temporal finer-grained sub-classification. Finally, in Section 4.2.3 we validate our approach against a big dataset of 6 fully annotated photo collections, taking [31] as a baseline.

The work in this section is related to publications [P3] and [P4].

4.2.1 Temporal Clustering using a Human-centric perception of time (TC-H)

A typical approach (see Figure 4.1) to photo organisation would have the user browse through his personal photos ordered time-wise as a Photo Stream. The user then would look for temporal gaps within the collection and mark some of them as boundaries between events. Time and, in particular, time gaps seem to be a useful feature to look for in order to try and bridge the semantic gap between timestamps and events.

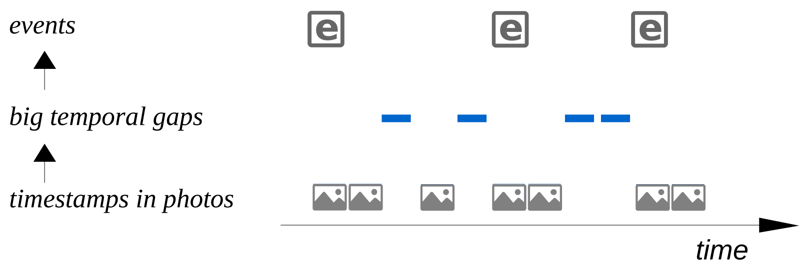

Figure 4.1: From timestamps to events

Loui and Savakis [31] propose a technique that mimics what we just described to achieve automatic organisation. To learn which time gaps are big and small, their method uses the distance between timestamps in consecutive photos where, for each timestamp n_i in the set, $n_{i-1} \leq n_i \leq n_{i+1}$ holds (that is, the set is ordered by these timestamps). Next, distances Δn between each pair of consecutive timestamps n are calculated iteratively for the whole set and clustered using k-means with $k = 2$.

This produces two clusters that group:

- small gaps that will not be considered as event boundaries and
- big gaps that signal inter-event boundaries in the set of photos.

Still, for large photo collections this is not enough to get satisfying results since time gaps often present wide variability. A scaling function is provided to amplify the impact of smaller gaps, based on interpreting time gaps with a meaning higher than just number of seconds between one photo an the next.

Consequently, time gaps are first classified according to the following units of time:

- less than a quarter of a day,
- between a quarter of a day and a day,
- between a day and a week,
- between a week and a month and
- more than a month.

The scaling function then is prepared to follow a human-centric perception of time which is typically non-linear. Under this interpretation an hour and a day are less similar than, say, a week and a month.

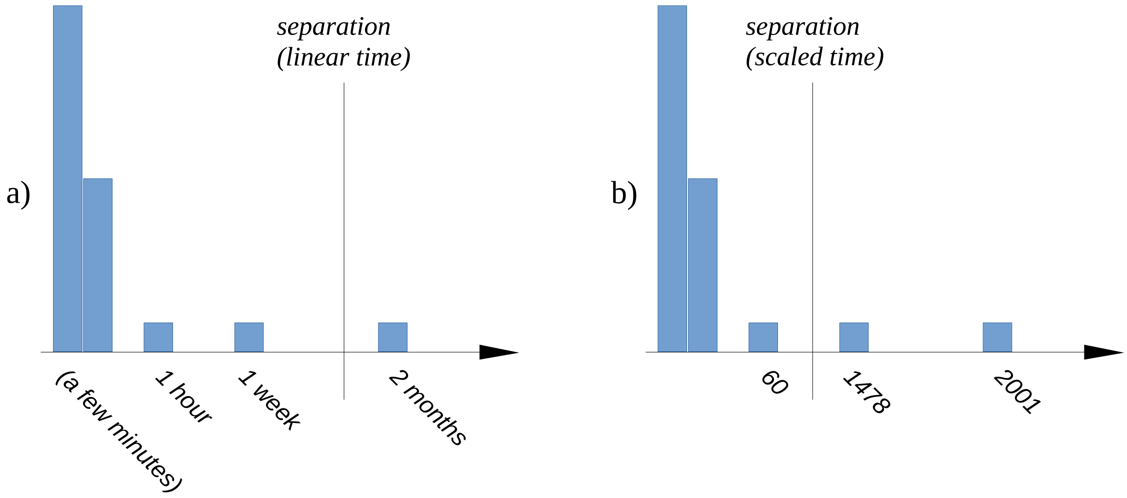

Figure 4.2: Difference histogram showing where the separation is done for a) linear time, and b) scaled time.

An example is illustrated in Figure 4.2. When only linear time is used (see Figure 4.2a) the 2-Means algorithm is applied on the histogram of linear time differences and the resulting separation may not be ideal (the selected granularity is higher than 1 week). In Figure 4.2b the scaling function is applied and the resulting clusters are more appropriate as the granularity now corresponds to a separation of 1 day.

In a nutshell, this technique extracts events from timestamps by analysing the whole photo collection to estimate a temporal gap and use it as separator. To further enhance the estimation, the authors encode a human-centric understanding of time that in their experiments (seen in [31]) produce better results than the purely linear approach.

4.2.2 Hierarchical Clustering through the detection of Round Trips (HC-RT)

The TC-H technique of the previous Section 4.2.1 is achieved by taking a straightforward temporal clustering technique and refining it with a rough modelling of how humans perceive different temporal gaps. Drawing inspiration from it, in this section we aim at further refining the TC-H approach with the incorporation of knowledge about the type of event being analysed.

For our approach, we attempt at understanding if the cluster detected using TC-H represents either a *routine* (*at-home*) event or a special (*away-from-home*) event. It is then necessary, as a previous step, to know which is the home location of the user. Although this may seem an important shortcoming, in [P3] we show that this information can often be inferred from the personal photo collection itself. In any case, for the personal photo management task this is not an issue since we can get input from the user.

Figure 4.3 shows an overview flowchart of the proposed approach in 4 stages:

(a) All photos are processed with TC-H to do a first shallow classification.

(b) Each cluster is categorised either as Routine or Special. Routine clusters correspond to events in the home location and no further processing is done on them.

(c) Re-join: consecutive special clusters (between routine events) are re-joined to represent one big special event.

(d) Each special event is processed with TC-H to find sub-events.

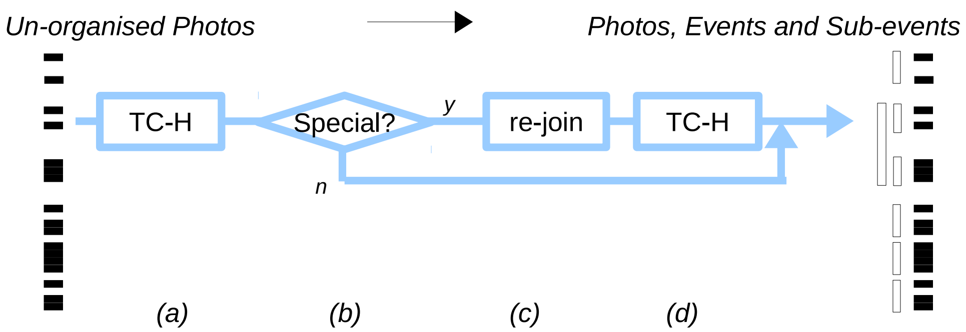

Figure 4.3: Schematic overview of the proposed algorithm.

In a nutshell, our technique builds on the TC-H technique of the previous Section 4.2.1 by analysing spatial information in each cluster found with TC-H to infer if the cluster represents a *routine* event at the home location of the user or instead it is a special *away-from-home* event. We try to overcome the gap between photos with temporal and spatial information and a big event composed of smaller sub-events by detecting round trips.

4.2.3 Experimental validation

In this section we evaluate the performance of our HC-RT approach. First, we describe the dataset used and then we establish the parameters of our evaluation. Afterwards we run our test and calculate the chosen parameters and discuss our findings.

Description of the dataset

For this experiment we use a bigger version of the dataset described in Section 3.2, with 9548 albums that validators selected as probably referring to events. Of these albums, we only include those with timestamps and GPS information, ending up with 5 users and 647 event-related albums, with 64.8 images per album on average. An overview of the data-set is given in Table 4.2.1.

	U1	**U2**	**U3**	**U4**	**U5**
Home location	San Francisco, CA	Miramar, FL	San Jose, CA	Tel Aviv, Israel	Newport, GB
# pics	2186	5803	16041	1366	16527
# events	36	219	180	22	190
# pics/events	61	26	89	62	87
sub-events	no	65 (1509 pics)	no	no	51 (4522 pics)
# locations	34	47	71	9	58
time period	3 years	6 years	11.6 years	4.4 years	12.4 years
movement	193142 km	157799 km	260327 km	9767 km	68823 km

Table 4.2.1: Overview of the data-set used in the experiments crawled from Picasaweb.

The collection of users is very diverse. U1 is a *nomad* user. The number of events almost corresponds to the number of locations. He shares images when he is travelling, and does not return to previous locations. U2 and U5 try to build the event/sub-event hierarchy by naming their albums following an ad-hoc systems like “Wedding_Ceremony”, “Wedding_Reception”, ... U4 is characterised by sharing a comparably small number of distinguished images over the years. All the users like travelling, it being the most common type of personal event. The travelling distance is diverse over the continents. U5, being the only European user, travels a lot, but within a smaller range.

Experimental set-up

The given data-set exemplifies five personal photo collections. As ground-truth, it provides a manual and subjective hierarchy of events done by the very user. All users are unaware of the experiments. The ground-truth is only justified by their personal experience. We regard the data-set as a temporally serial stream of photos, where we detect event borders. Therefore, we evaluate a 2-class classification problem. For each experiment we compute standard information retrieval measures and refer to them in percent.

$$Precision = \frac{100\,tp}{tp+fp} \quad , \quad Recall = \frac{100\,tp}{tp+fn} \quad , \quad F1-measure = \frac{2\,Precision\,Recall}{Precision+Recall} \quad ,$$

where *tp* denotes a true positive detection of event boundaries, *tn* a correctly non-detection of an event boundary, *fp* an event being wrongly segmented and *fn* an actual event boundary not being detected. The hierarchical interpretation of the data-set is seen as a 3-class classification problem: event borders, sub-event borders and no borders.

Results

We consider TC-H (presented in Section 4.2.1) as a baseline. It is based on only on temporal information. With an F1-measure of about 60, it already shows promising results. This clearly illustrates that the temporal aspect is the most important for the detection of event boundaries. Detailed numerical results can be seen on the row corresponding to TC-H in Table 4.2.2.

		U1	U2	U3	U4	U5	Ø
TC-H	Prec	27.87	71.01	32.67	73.68	52.11	51.47
	Recall	94.44	77.17	89.67	71.79	71.96	81.01
	F1	43.04	**73.96**	47.90	72.73	60.44	59.61
HC-lite	Prec	72.34	95.76	66.11	80.00	87.60	80.36
	Recall	94.44	51.60	92.40	75.68	58.89	74.60
	F1	81.93	67.06	77.07	77.78	70.43	74.85
HC-RT	Prec	75.00	84.76	94.71	76.74	78.75	81.99
	Recall	94.29	65.29	80.50	89.19	70.00	79.85
	F1	**83.54**	73.74	**87.03**	**82.50**	**74.12**	**80.19**

Table 4.2.2: Results for event boundary detection within the user's album organisation.

Our HC-RT approach (Section 4.2.2) is evaluated twice. At first, we disregard the hierarchical organisation within special events and only take the root events. The results are seen in the row labelled “HC-lite” with a mean F1-Measure of 75. The second time we compare the full approach (row labelled “HC-RT”), achieving a mean F1-Measure of 80.

In both cases, the richer semantics of the hierarchical approach outperform the TC-H baseline, showing the value of our proposal.

4.3. From Simple Spatial-temporal Cues To Event Sharing

In the previous Section 4.2 we explored how temporal and spatial metadata in photos can be used to find the boundaries between one personal event and the next and in this way organise a personal photo collection.

In this section we take personal photo collections of different users and analyse how temporal and spatial metadata can be of help in the general task of sharing photos with other users that may be interested.

Our key intuition is that we can detect event co-participation (i.e., users attending the same event) by analysing whether 2 users have personal events that are spatially and temporally similar. We describe our approach to event similarity in Section 4.3.1 in which we state that if two personal events are similar enough beyond a certain threshold, they can be considered as different personal accounts of the same underlying event. Consequently, we can take this as a probable signal of event co-participation and use this technique to recommend the sharing of event and photos entities.

The work in publication [P5] is related to this section.

4.3.1 Detecting event co-partipation using spatial-temporal event similarity

We detect event co-participation by defining an event similarity function and setting a threshold beyond which two events are similar enough that they are considered to be two different accounts of the same event.

Consider this representation of an event E that is in line with our framework (see Section 4.1):

$$E = \langle t, P \rangle$$

where t is the temporal period of E and P is the set of geographic coordinates (latitude, longitude) in all photos of E.

The algorithm first determines if there is a time overlap between events. If two events E_A and E_B are defined as:

$$E_A = \langle t_A, A \rangle \quad \text{and} \quad E_B = \langle t_B, B \rangle$$

then time similarity is computed by taking into account how much of t_A overlaps with t_B.

We take the inverse of the time overlap function, called *tsim*, to assess temporal similarity between two event instances, defining it as follows:

$$tsim(t_A, t_B) = 1 - \frac{t_A \cap t_B}{t_A \cup t_B} \text{ , if } \quad t_A \cup t_B \neq \emptyset \qquad tsim(t_A, t_B) = 1 \text{ , otherwise}$$

Events that are completely temporally similar will have $tsim = 0$, while events that are not similar at all will have $tsim = 1$.

For spatial analysis we use the haversine distance (denoted $d(p, q)$ for two geographic points p and q) to build another function that selects the minimum distance from a spatial point p to a set B as follows:

$$dmin(p,B)=\underset{i=1}{\overset{n}{MIN}}\, d(p,b_i)$$

where $B = \{ b_1, b_2, \ldots , b_n \}$.

The spatial similarity *ssim* from another set A of geographic points to B is then given by:

$$ssim(A,B)=\frac{\sum_{i=1}^{m} dmin(a_i,B)}{m}$$

where A = $\{ a_1, a_2, \ldots , a_m \}$.

Both *tsim* and *ssim* are directed, that is, $tsim\ (t_A , t_B)$ is not the same as $ssim\ (t_A , t_B)$. We take an optimistic approach towards similarity to produce non-directed versions:

$$ntsim(t_A,t_B)=MIN\{tsim(t_A,t_B),tsim(t_B,t_A)\}$$

$$nssim(A,B)=MIN\{ssim(A,B),ssim(B,A)\}$$

Events E_A and E_B are similar when:

$$ntsim(t_A,t_B)<c_T \quad \text{and} \quad nssim(A,B)<c_S$$

where c_T and c_S are thresholds learned experimentally in the following Section 4.3.2.

4.3.2 Experimental validation

In this section we evaluate the performance of the event similarity approach. First, we describe the dataset used and then we establish the parameters of our evaluation. Afterwards we run our test and calculate the chosen parameters and discuss our findings.

Description of the dataset

To build our dataset we collected events from Upcoming.org, photo albums associated to them from Flickr and their corresponding contact lists.

Our ground truth provides validation for social event detection based on metadata from photos that were annotated with machine tags 8 of the form ‘upcoming:event=*evid*’ where *evid* refers to an event instance in the Upcoming service. In this way an event from Upcoming can be linked to photos in Flickr that constitute the depiction of such event. In a first instance, photos of the same user that are grouped by the same machine tag are ground truth for that user’s personal events. Moreover, personal events from different users that have the same machine tag indicate that these personal events are different accounts of an underlying social event. Social ties are validated using information about Flickr contacts for each user. Two users that took photos and tagged them using the same machine tag provide evidence that they were co-participants in the same event identified by that tag. We exploit this fact to build participant lists from machine tags contained in photos from different users. To build our ground truth we extracted Flickr contact lists for all the users involved. By testing if two users have each other in their contact lists we can establish whether users that co-participate in a particular event are socially connected or not. The dataset consists of more than 11 180 events with photos contributed by more than 4100 different users. Of these events, 1291 have photos owned by more than 1 user, a condition necessary if we are to analyse co-participation and its relation to acquaintance. The dataset includes metadata of photos uploaded in the years 2007 - 2012. For this study we only required the Flickr user ID of the owner of each photo, tags for Upcoming events and the list of Flickr contacts for each user.

Experimental setup

To derive social events from personal photo streams we proceeded in two phases. First, we found personal events within the user’s photo stream using the approach described in previous section. After all individual photo streams were processed and personal event instances detected, we analysed each personal event against all others, using the spatial-temporal similarity function described in Section 4.3.1.

This similarity function works with two thresholds c_T and c_S above which we do not consider that two personal events are accounts of the same social event. To set these two thresholds we divided our dataset in two groups, using 50% of the detected personal event instances to tune these thresholds and the remaining 50% to test. Pairs of personal events for which the similarity function produced results below the thresholds tested were considered to be part of a bigger social event. In this way we were able to combine several personal events into social event candidates.

To validate a detected social event, each one of them was automatically scanned looking for upcoming tags. If the detected social event had only one upcoming tag throughout its whole photo set then we considered this case as a true positive. It was not possible to do automatic evaluation on detected social events without Upcoming tags or with multiple Upcoming tags. Manual inspection was required in these cases, in order to see if photos coming from different users were still about the same event.

c_T	c_S *(km)*	U-Joint	True Pos.	O-Joint	TP%
0.75	0.50	28	265	32	81.54
0.75	1.00	26	265	32	82.05
0.75	5.00	23	264	34	82.24
0.75	10.00	23	263	34	82.19
0.50	0.50	28	267	31	81.90
0.50	**1.00**	**26**	**267**	**31**	**82.40**
0.50	5.00	23	264	34	82.24
0.50	10.00	23	263	34	82.19
0.25	0.50	45	254	14	81.15
0.25	1.00	45	254	14	81.15
0.25	5.00	43	254	15	81.41
0.25	10.00	43	253	15	81.35

Table 4.3.1: Results for the parameter learning phase, comparing true positives against under-joint (U-Joint) and over-joint (O-joint) events. Best results where achieved using c_S = 0.50 and c_S = 1.00.

Results

Table 4.3.1 shows the parameter learning process. We have selected $c_T = 0.50$ and $c_S =$ 1.00 where the percentage (column *"TP%"*) of true positives (column *"True Pos."*) is higher. Incorrect detection happens due to under-joining (column *"U-joint"*) and over-joining (column *“O-joint”*) of social events. We can also see that there are small variations for temporal threshold c_T of 0.75 and 0.50, with c_T = 0.50 resulting in less over-clustering, while a value of 0.25 results in an increased number of partial social events. Based on the set of experiments run, c_T = 0.50 and c_T = 1.00 produce the best results.

After submitting the testing subset to evaluation we found that the event similarity function produced correct results (*“True Pos.”*) in 78.76% of the cases, partial social events (*“U-joint”*) in 9.56% of the cases and social events with multiple tags (*“O-joint”*) in 11.68% of the cases.

Chapter 5 An event-centric platform for Personal Photo Management

In Chapter 3 we found out that even when presented with conventional album-centric design, users are already organising their photos, in most cases, in terms of events. Following assumptions of the state of the art in event modelling, we find that space, time, people and type of activity are the main different facets converging into a single event.

Based on this findings, we presented our event-centric approach in Chapter 2 and then evaluated its utility by describing typical tasks of organisation and sharing around the spatial-temporal context contained in events.

After having shown potential exploitation of this spatial-temporal context within events, in this chapter we present a platform that cover the typical life cycle photos, designing its related tasks around events.

In Section 5.1 we present the general architecture and describe each sub-component in terms of its purpose and its capabilities. Next, in Section 5.2 we describe typical scenarios and their related use cases on a step-by-step basis. Section 5.3 presents our implementation of a subset of these scenarios in the context of a smartphone application. Finally, in Section 5.4 we discuss the evaluation done first on a low-fidelity paper prototype and then on a first release in the context of a visit to a museum.

5.1. System Architecture

This section presents the general architecture and some details of the most important subsystems of the proposed photo management platform.

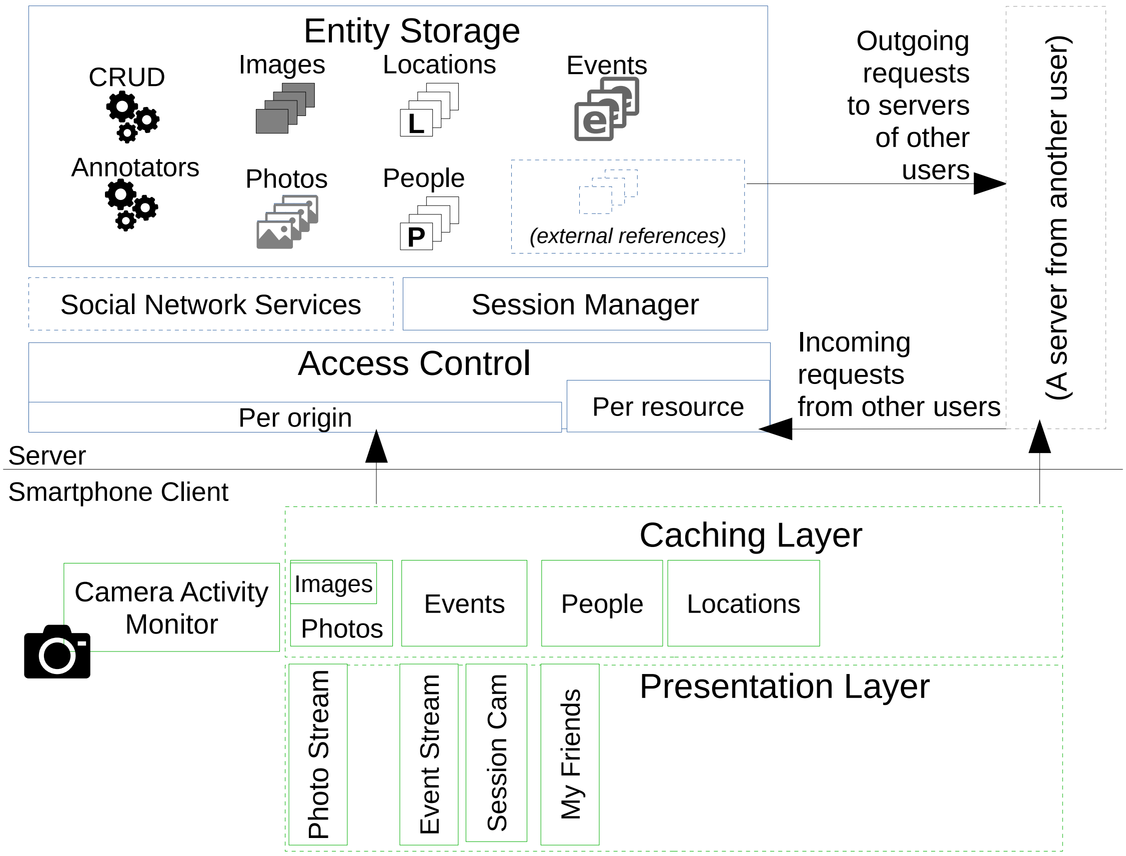

Figure 5.1: General architecture of the Photo Management Platform.

Figure 5.1 shows the main components in our platform, both on the Server and on the Client. We follow a resource based approach for access and manipulation of entities, avoiding the storage of entities belonging to other users. This means that each server is owned but just one user, and is in charge of storing entities owned by that user only. When the user receives access to a photo taken by a friend, all manipulations are requested to the external server that holds the instance of the photo shared. To keep track of known external entities, the server stores references to known external entities. On the client side, the same user can have multiple client applications communicating with his/her server to allow for the use of several photo-taking devices simultaneously (e.g., a user using his smartphone and his DSLR WiFi-enabled camera).

5.1.1 Server

In our vision of the system, the server is in charge of:

- storing the entities of the user, performing various manipulations on the entities on request and keeping track of entities shared by other users, querying their corresponding servers when needed,
- authorising or denying requests using two modalities: full access for an authenticated owner and per-entity-per-user for requests by other users,
- keeping track of realtime collaboration sessions, giving external users access to shared events, performing event-matching against received events.
- keeping track (either locally or by referencing some social network platform) of the friendship relations of the user. In the context of the platform, being friends with another user means knowing the point of access to his or her platform server.

Each of these concerns is represented in Figure 5.1 as a component and described within the next subsections.

Entity storage

The repository of entities indexes all local instances of Objects, People, Locations, Images, Photos and Events (as described in 4.1). Additionally, it also indexes external entities known to the repository due to sharing.

It provides CRUD and more complex services to other components (e.g., event boundary detection, event matching).

Access Control

This component manages external accesses to entities in the repository of the user in two modalities:

- full access: if the authenticated origin and the owner of the server are the same.
- per-resource per-origin: an external query on a specific entity will only be allowed if the tuple <*origin, uri, q*> is present, where origin represents the external agent performing the query, uri refers to the entity to be queried and q to the query itself (e.g., an HTTP GET).

Session Manager

This component is in charge of managing realtime sessions (collaborative and individual). In a session, all uploaded photos are automatically associated with a “live” event.

Users can invite other users and when the invitations are accepted, these other users become contributors and from there on, also their newly captured photos are linked with the live event.

Collaboration ceases when the user leaves the session, but the event is kept on being updated by other users that are still in. When the last user leaves, the session is closed and there are no more live additions.

Social network services

This component is used to query the social network of the user. It is used to determine if the user is “friends” with another user.

In the context of this platform, “being friends” means knowing the entry point to query the server owned by a user that is a “friend” in order to be able to access shared photos.

5.1.2 Client

For the client side of our platform, we have the following components:

Camera activity monitor

To automatically upload all newly captured photos, this component is implemented as a background service that initiates the uploading right after a new photo is taken.

A caching layer

The purpose of this component is to retrieve and operate on a subset of all the entities present in the server and allow for offline operation and transparent synchronisation.

A presentation layer

This layer mediates user input and output. The interfaces designed in this layer allow for manipulation of photos in two levels:

- as un-organised photos in a timeline and
- as events, after automatic organisation.

Additionally, it allows for starting, joining and closing of live event building sessions.

5.2. Scenarios

To test our design we selected three representative scenarios that will be described here.

5.2.1 Creating events from selected photos

1. The user selects several photos covering a period of 1 week and invokes the creation of an event.
2. The application runs event boundary detection on the selected set, and creates one or more new event instances.
3. The application shows the event stream, focused on the new events just created.

5.2.2 Reviewing similar events

1. The user selects an event from the event stream to be displayed.
2. A detailed view of the event is presented, including title, list of participants, temporal period, location and associated photos.
3. Since there is a similar event from another user, it is also shown.

4. The user selects this similar event and reviews it.

5.2.3 Building an event collaboratively

1. User A starts a new event manually.
2. User A makes his event visible to all friends nearby.
3. B starts a new event manually.
4. B makes his event visible to all friends nearby.
5. A and B take lots of photos.
6. A and B can see each others photos in realtime.
7. A and B close their respective events.

5.3. Prototype Implementation

We implemented a subset of client and server functionalities in order to test our selected scenarios.

The server is implemented using Ruby on Rails to do CRUD on entities for events, photos, images, locations and people. Additionally, it performs event boundary detection (as seen in Section 4.2.2) and checks for event matching (using the approach in Section 4.3.1) whenever a new instance is detected.

The client is an Android application that communicates with its server using HTTP to exchange JSON objects that represent the entities described in Section 4.1.

5.4. Evaluation And Preliminary Observations

We performed our evaluation with 4 users {U1, U2, U3, U4}, aged 31, 36, 37 and 37.

Evaluation of scenarios A (“Creating events from selected photos”) and B (“Reviewing similar events”), were performed with each user trying the prototype separately. Scenario C (“Building an event collaboratively”) by its very nature had to be performed with all 4 users simultaneously.

At first, we presented Scenario A as a bootstrapping task. We let the user ask for automatic event organisation on his whole collection, pre-loading photos taken from the smartphone of each user, covering a period of three months. It produced a reasonable classification except for U3 for whom the classification produced 4 correct events and 8 non-events (as confirmed by U3) with only 1 photo.

In the second iteration, we discarded our all-or-nothing approach, letting the users select some consecutive photos, create new events and see the results. This was a definite enhancement because users intuitively focused on parts of the photo stream where they knew the classification should be more interesting to them. For example, U1 selected a time period for his summer vacation, and not a week in which he mostly took casual photos of his pet.

For Scenario B (“Reviewing similar events”), we took the same month (June, 2013) for all users and asked them to manually organise and then review these manually produced events. Since all 4 users are friends in real life, the system was able to match their personal events in 1 occasion, for a day they spent together in Lake Levico. It also produced matches in one additional occasion for U2 and U3. In all cases, this scenario was well received (U2 praised the “lazy” approach to photo sharing). Moreover, we visually identified two more matches for U1 and U2 but the corresponding photos in U1's collection lacked the necessary GPS coordinates. This calls for a modification of Scenario B to additionally let users explicitly ask for possible similar events. If an explicit request is made by the user then the application can re-try event matching with looser parameters.

Scenario C was tested by making a group visit to a museum (Trento's Muse), which was selected for its lack of restrictions on taking photos. All 4 users immediately created their respective events and set the visibility to “all friends”. We observed an interesting behaviour in U2 and U3: they took turns at capturing photos of one another. When we asked about this behaviour, we found out this was intentional. U3 stated: “this is so I can take photos and also be in the photos”. This observation might point to appropriation practices in a “ping-pong” pattern of collaboration.

5.4.Evaluation and Preliminary observations

The results from our evaluation can be summarised as follows:

- An all-or-nothing use of HC-RT (see Section 4.2.2) performs poorly in this application. Nevertheless, after a simple modification to give more control to the user, it proves helpful.
- The use of event similarity for automatically finding shared events is useful. It will be made better if we let the user explicitly request for matching when it was not automatically achieved but is still expected, by retrying with less stringent threshold.
- Live event building is well understood and we might have possibly observed how users appropriate this feature.

Chapter 6 General conclusions

Traditional approaches to photo management are no longer producing satisfactory results for the typical user. The increasing availability of contextual information provided by sensors in newer photo capturing devices calls for a transition to more sophisticated paradigms.

Towards this objective, we investigate the use of an event-centric metaphor for personal photo management, inspired in its apparent resemblance to how the autobiographical memory operates, as it too focuses on the event as the unit of organisation, retrieval and sharing of experiences.

Although there is a lot of supporting evidence from past behaviours that this resemblance can be exploited, we felt the need to support it with more recent research. We looked for implicit hints of event use on current online sharing tools, and found that even though these tools do not directly support the use of events, users are still finding ways to encode event information within their classifications.

With this newer supporting evidence, we engaged in the task of describing typical photo management tasks in terms of events and the spatial-temporal context they provide as organisational units. What we found is that often, we can extract higher meaning from simple temporal and spatial cues and achieve automatisation if we represent or at least approximate some knowledge of how time and space are understood by humans. Moreover, even when we only achieve partial automatisation, we can re-frame a task and decrease the effort that the use needs to put in order to achieve its goal.

Finally, we recognise there is a whole untapped well of event-centric features other than space and time, and as such, present interesting challenges for further studies.

Bibliography

[1] Andrews, P., Paniagua, J., Giunchiglia, F., "Clues of Personal Events in Online Photo Sharing" in Proceedings of the 10th International Semantic Web Conference, DeRIVE '11 , pages:. (2011)

[2] Andrews, P., Paniagua, J., Torsi, S., "Katie's Swiss Trip: A study of personal event models for photo sharing" in International Journal on Semantic Web and Information Systems (IJSWIS) 9, pages:42-56. IGI Global (2013)

[3] Bach, J., Fuller, C., Gupta, A., Hampapur, A., Horowitz, B., Humphrey, R., Jain, R., Shu, C., "Virage image search engine: an open framework for image management" in Proceedings of the SPIE Storage and Retrieval for Still Image and Video Databases , pages:76-87. (1996)

[4] Bazzanella, B., Palpanas, T., Stoermer, H., "Towards ageneral entity representation model" in Information Reuse & Integration IRI'09 , pages:431-432. (2009)

[5] Bickford, M., Constable, R., Eaton, R., Guaspari, D., Rahli, V., Introduction to EventML, 2011

[6] Boato, G., Fontanari, C., Giunchiglia, F., De Natale, F., "GLOCAL Multimedia Retrieval" in DISI Technical Report , pages:. (2008)

[7] Bosson, A., Cawley, G., Chan, Y., Harvey, R., "Non-retrieval: Blocking Pornographic Images" in Proceedings ofthe 1st International Conference on Image and Video Retrieval , pages:50-60. Springer-Verlag (2002)

[8] Bouquet, P., Stoermer, H., Niederee, C., Maa, A., "Entity name system: The back-bone of an open and scalable web of data" in Semantic Computing, 2008 IEEE International Conference , pages:554-561. (2008)

[9] Casati, R., Varzi, A., "Events" in The Stanford Encyclopedia of Philosophy , pages:. The Metaphysics Research Lab at Stanford, CA (2010)

[10] Chalfen, R., Snapshot Versions of Life, 1987

[11] Fan, J., Gao, Y., Luo, H., "Multi-level annotation of natural scenes using dominant image components and semantic concepts" in Proceedings of the ACM International Conference on Multimedia , pages:540-547. ACM (2004)

[12] Fialho, A., Troncy, R., Hardman, L., Saathoff, C., Scherp, A., "What's on this evening?" in EVENTS 2010 , pages:. (2010)

[13] Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Qian Huang Dom, D., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., Anker, P, "Query by image and video content: The QBIC system" in Computer , pages:23-32. ()

[14] Forsyth, D., Fleck, M., "Automatic detection of human nudes" in Computer Vision 32, pages:63-77. (1999)

[15] Frohlich D., Kuchinsky A., Pering C., Don A., Ariss S., "Requirements for photoware" in Proceedings of the ACM conference on computer supported cooperative work , pages:166-175. (2002)

[16] Giunchiglia, F., Andrews, P., Trecarichi, G., Chenu-Abente, R., "Media Aggregation via Events" in EVENTS 2010 - Recognising and Tracking Events on the

Web and in Real Life , pages:. (2010)

[17] Hardman, L., Van Ossenbruggen, J. Troncy, R., Amin, A., Hildebrand, M., Interactive information access on the web of data, 2009

[18] Heath, T., Bizer, C., "Linked data: Evolving the web into a global data space" in Synthesis Lectures on the Semantic Web: Theory and Technology 1, pages:1-136. (2011)

[19] Jain, R., "EventWeb: Developing a Human-Centered Computing System" in Computer 41, pages:42-50. (2008)

[20] Jameson, A., "Scenarios and Challenges for the Sharing of Event-Indexed Media" in Workshop on Visual Interfaces to the Social and Semantic Web (VISSW2010), IUI , pages:. (2010)

[21] Jameson, A., Buschbeck, S., "Interaction design for the exchange of media organized in terms of complex events" in EVENTS 2010 , pages:. (2010)

[22] Kim, P., "Event-based Multimedia Chronicling Systems" in Computer Engineering , pages:1-12. (2005)

[23] Koehn, P., "Europarl: A parallel corpus for statistical machine translation" in Machine Translation 5, pages:. (2005)

[24] Kurby, C., Zacks, J., "Segmentation in the perception and memory of events" in Trends in Cognitive Sciences 12, pages:72-79. (2008)

[25] Lansdale, M., Edmonds, E., "Using memory for events in the design of personal filing systems" in International Journal of Man-Machine Studies 36, pages:97-126. (1992)

[26] Lew, M., "Next-generation Web searches for visual content" in Computer 33, pages:46-53. IEEE (2000)

[27] Lew, M., Sebe, N., Djeraba, C., Jain, R., "Content-Based Multimedia Information Retrieval: State of the Art and Challenges" in ACM Transactions on Multimedia Computing, Communications and Applications 2, pages:1-19. ACM (2006)

[28] Li, J., Wang, J., "Automatic linguistic indexing of pictures by a statistical modeling approach" in Transaction on Pattern Analysis and Machine Intelligence 25, pages:1075-1088. IEEE (2003)

[29] Linguistic Data Consortium, Ace (automatic content extraction) en-glish annotation guidelines for entities,, 2004

[30] Löffler, H., Baranger, W., Steidl, M., IPTC Standards for Photo Metadata, 2007

[31] Loui, A., Savakis, A., "Automated Event Clustering and Quality Screening of Consumer Pictures for Digital Albuming" in IEEE Transactions on Multimedia 5, pages:390-402. IEEE (2003)

[32] Lowe, D., "Object recognition from local scale-invariant features" in Proceedings of the Seventh IEEE International Conference on Computer Vision 2, pages:. (1999)

[33] Miller, A., Edwards, W., "Give and Take: A Study of Consumer Photo-Sharing Culture and Practice" in CHI'07 , pages:347–356. (2007)

[34] Miller, G., "WordNet: a lexical database for English" in Communications of the ACM 38, pages:39-41. ACM Press (1995)

[35] O'Hare, N., Lee, H., Cooray, S., MediAssist: using content-based analysis and context to manage personal photo collections, 2006

[36] Paniagua, J., Tankoyeu, I., Stöttinger, J., Giunchiglia, F., "Indexing Media by Personal Events" in Proceedings of the 2nd ACM International Conference on Multimedia Retrieval ICMR '12 , pages:. (2012)

[37] Paniagua, J., Tankoyeu, I., Stöttinger, J., Giunchiglia, F., "Social events and social ties" in Proceedings of the 3rd ACM Conference on Multimedia Retrieval, ICMR ’13 , pages:143-150. (2013)

[38] Pianta, E., Girardi, C. and Zanoli, R., "The TextPro tool suite" in Proceedings of Sixth International Language Resources and Evaluation (ELRA) , pages:. (2008)

[39] Rattenbury, T., Good, N., Naaman, M., "Towards automatic extraction of event and place semantics from flickr tags" in SIGIR '07 , pages:103. (2007)

[40] Richardson, L., Ruby, S., RESTful web services, 2007

[41] Rodden, K., Wood, K., "How do people manage their digital photographs?" in Proceedings of the CHI '03 , pages:409. (2003)

[42] Scheffler, U., "Events as shadowy entities" in Philosophy 2, pages:35-53. (1994)

[43] Scherp, A., Jain, R., "Introducing an ecosystem for semantics" in IEEE Multimedia 16, pages:. (2009)

[44] Shaw, R., Troncy, R., Hardman, L., "LODE: Linking Open Descriptions of Events" in The Semantic Web , pages:153-167. (2009)

[45] Smeulders, S., Worring, M., Santini, S., Gupta, A. and Jain R., "Content-based image retrieval at the end of the early years" in IEEE Transactions on PatternAnalysis and Machine Intelligence 22, pages:1349-1380. IEEE Journals & Magazines (2000)

[46] Tankoyeu, I., Paniagua, J., Stöttinger, J., Giunchiglia, F., "Event detection and scene attraction by very simple contextual cues" in Proceedings of the 2011 joint ACM workshop on Modeling and representing events, J-MRE ’11 , pages:1-6. (2011)

[47] Van Hage, W., Malaisé, V., Segers, R., Hollink, L., Schreiber, G., "Design and use of the Simple Event Model (SEM)" in Web Semantics: Science, Services and Agents on the World Wide Web 9, pages:128-136. Elsevier (2011)

[48] Wenyin L., Dumais S., Sun Y., Zhang H., Czerwinski M., Field B., "Semi-automatic image annotation" in Human-Computer Interaction - Interact '01 , pages:. (2001)

[49] Westermann, G., Mareschal, D., Johnson, M., Sirois, S., Spratling, M., Thomas, M., "Neuroconstructivism" in Developmental Science 10, pages:75-83. Blackwell Publishing Ltd (2007)

[50] Zacks, J., Tversky, B., "Event structure in perception and conception" in Psychological Bulletin 127, pages:3-21. (2001)