Semantically Enhancing Multimedia Lifelog Events

Peng Wang[1,⋆], Alan Smeaton[2], and Alessandra Mileo[3]

[1] Beijing Institute of System Engineering, Beijing, P.R. China, 100101
pengwangnudt@sina.com
[2] Insight Centre for Data Analytics
Dublin City University, Glasnevin, Dublin 9, Ireland
alan.smeaton@dcu.ie
[3] Insight Centre for Data Analytics,
National University of Ireland, Galway, Ireland
alessandra.mileo@deri.org

Abstract. Lifelogging is the digital recording of our everyday behaviour in order to identify human activities and build applications that support daily life. Lifelogs represent a unique form of personal multimedia content in that they are temporal, synchronised, multi-modal and composed of multiple media. Analysing lifelogs with a view to supporting content-based access, presents many challenges. These include the integration of heterogeneous input streams from different sensors, structuring a lifelog into events, representing events, and interpreting and understanding lifelogs. In this paper we demonstrate the potential of semantic web technologies for analysing lifelogs by automatically augmenting descriptions of lifelog events. We report on experiments and demonstrate how our results yield rich descriptions of multi-modal, multimedia lifelog content, opening up even greater possibilities for managing and using lifelogs.

Keywords: lifelogs, events, semantic web, semantic enhancement.

1 Introduction

Lifelogging is a user-controlled form of gathering personal multimedia information, though not necessarily for sharing with others [1]. Lifelogging is the process of sensing and digitally recording our everyday behaviour, capturing a person's experiences in the form of events, states and relationships. This rich pool of information is collected by wearable sensors by individuals to characterize their own contexts and activities and build applications that can enhance their quality of life.

An important premise for the use of lifelogs is the fact that *events*, discrete and often repetitive activities, are the atomic unit of interaction with lifelogs [2,3]. However, the identification of event boundaries alone is just a first step because we need to know what events actually are and how they relate to each

⋆ Research funded by Science Foundation Ireland under grant SFI/12/RC/2289.

W.T. Ooi et al. (Eds.): PCM 2014, LNCS 8879, pp. 163–172, 2014.

other. In this work we discuss how semantic enrichment of lifelog events and the creation of semantic links between those descriptions and external knowledge can help bridge this data-to-knowledge gap and develop applications that can use lifelog data to determine user contexts and activities more effectively.

In Section 2 we discuss the shortcomings of current event enhancement approaches and the potential of using semantic web for modeling, linking and re-using knowledge about events are reported in Section 3. Experimental results are reported in Section 4, followed by conclusions and future work.

2 Background and Related Work

Lifelogging is concerned with digitally capturing media-rich representations of activities of everyday life and making use of them across a variety of use cases. Arbitrary sets of information including location or geo-tags, textual information, photos, audio and video clips from mobile sensors, etc. can characterize a particular scenario in everyday life.

Among all forms of information, visual information contains more semantics of events which can be used to infer other contexts like "Who", "What", "Where" and "When". Visual lifelogging is the term used to describe both image- and video-based lifelogging from wearables. The SenseCam, shown in Figure 1 and which we use in the work in this paper, is a sensor-augmented wearable camera designed for visual lifelogging by recording a series of images and a synchronised log of sensor data. It captures the view of the wearer from a fisheye lens and pictures are taken automatically at the rate of about one every 50 seconds. On-board sensors are also used to trigger additional capture of pictures when sudden changes are detected in the environment of the wearer. SenseCam has been shown to be effective in supporting recall of memory from the past for individuals [5], as well as having applications in diet monitoring [6], activity detection [7], smoking cessation, sleep monitoring, etc. and these are representative of the kinds of applications to which multimedia lifelogs can be applied [1].

Fig. 1. SenseCam worn by a user

Once lifelogging hardware reached the stage of being useful for real world applications, the focus shifted towards mining deeper meanings from lifelogs i.e. eliciting their semantics. However, the state-of-the-art in enhancing lifelog events is almost all based on low-level feature matching. An everyday event enhancement technique with more comprehensive online resources is introduced in [8], in which SenseCam images are augmented with images from external publicly-available data sources such as Flickr, YouTube, etc. The shortcomings of this are obvious in that the task is performed based on image matching using the scalable colour MPEG-7 image descriptors and geotags. Though a number of approaches have been proposed for multimedia event detection, as shown in TRECVid [13] and MediaEval [14], and egocentric media summarization [15], etc., these approaches still lack capabilities in interpreting events from different contexts, which is more important for lifelogs obtained from sensing devices.

An example of work on enhancing automatically collected lifelogs is reported in [9] which describes the *DigMem* platform which tracks location and combines it with the wearer's heart rate and photos taken by the wearer from a tablet computer. This is turned into RDF and semantically enriched using Linked Data, in a similar way to what is reported later in this paper. What differentiates this work from ours is that our photo capture is true ambient life logging with no user involvement which makes the nature of the photos different with strong similarity among our photos when taken during the same event. Furthermore, our lifelog data is structured into spatio-temporal events which are enhanced semantically through links whereas in previous work there may be only one photo taken during an event thus enhancement is nearly impossible.

3 Semantically Enhanced Lifelogging

The goal of enhancing lifelog events is to inter-link knowledge from different contexts to enhance lifelog applications. We now describe the model and implementation of the EventCube system which does this.

3.1 Contexts and Events: Semantic Representation

Context information can be collected in lifelogging through the deployment of heterogeneous sensors. To model the semantics of events, contexts should be modeled in a way which allows each to be processed separately. That is because high uncertainty is often embedded in context processing due to data loss (e.g. GPS signal dropouts) or detection defects such as Bluetooth signal quality. The following concepts describing multimedia should ideally be included in lifelogs in which events are to be interpreted:

- **Event**: an occurrence as the intersection of time and space.
- **Location**: the geographical context of an event, as a recall cue.
- **Time**: the temporal context as a recall cue for an event.
- **Actor**: the human who carried out the event, normally the user.
- **Attendee**: the human/humans present and possibly involved in the event.
- **Image**: the class abstract for the image.
- **Annotation**: a class abstract for the textual description of events.

For consistency with the definition of an event as "a real-world occurrence at a specific place and time", we explicitly model the event class with spatial and temporal constraints in terms of OWL cardinality restrictions, as shown in Listing 1.1, formalized in the language Turtle [4]. The restriction `owl:cardinality` is used to enforce that one event has exactly one value for the properties of starting time and ending time. In Listing 1.1, the restriction `owl:minCardinality` is stated on the property `:hasLocation` with respect to event class, indicating that any event instance needs to be related to at least one GPS location. In lifelogging, there are cases when more than one GPS coordinate is needed to reflect the spatial characteristics of an event, such as "walking" or "driving".

Besides the raw-content of lifelog events captured directly by wearable sensors, there are three main external types of context, namely spatial, temporal and social for which there are already well-established ontologies to describe them. We investigated existing ontologies which may be reused and integrated into our context enhancing event ontology and chose the OWL-Time and GeoNames ontologies to model spatial and temporal contexts respectively. In our architecture, the people involved in an event including the actor and attendees are modeled by the FOAF (Friend Of A Friend) ontology which describes persons with their properties and relations. The visual information about events which answers the "What" question can be depicted by SenseCam images (addressed by the `FOAF:Image` class), or combined together with semantic annotations obtained from concept detectors.

```
:Event rdf:type owl:Class ;
       rdfs:subClassOf time:TemporalEntity ,
[rdf:type owl:Restriction ;
 owl:onProperty :hasLocation ;
 owl:minCardinality "1"^^xsd:nonNegativeInteger],
[rdf:type owl:Restriction ;
 owl:onProperty :endAt ;
 owl:cardinality "1"^^xsd:nonNegativeInteger],
[rdf:type owl:Restriction ;
 owl:onProperty :beginAt ;
 owl:cardinality "1"^^xsd:nonNegativeInteger].
```

Listing 1.1. Definition of Ontological event classes

3.2 Semantics of Locations

Semantic enhancement of events in lifelogs needs an effective measure to locate the user because location can be a key facet in indexing multimedia lifelogs. Because GPS does not work inside buildings or where satellite signals are weak, we apply a location enhancement algorithm consisting of location clustering, reverse geocoding and an LOD (Linked Open Data) semantic query.

Location clustering is first explored on GPS coordinate records using k-means clustering [10] in which we use 100m as the clustering radius and a filtering time span of 10 minutes. Reverse geocoding is used to translate latitude/longitude pairs into semantically meaningful address names and returns the closest addressable location. Once the placename is obtained, we query all the semantics (properties and values) of the placename in DBpedia's RDF repository through a SPARQL query and provide links for the user to navigate the returned RDF graphs.

Current reverse geocoding web services label GPS coordinates with semantic tags by returning the nearest place names however due to GPS error margins and the sizes of what we define as places (city vs. streetname), the nearest place is not necessarily the correct answer for the target event. To address this, we provide nearby places as a ranked list for the user and we enhance selected places at the user's request. We rank the place list according to popularity, analyzed from Flickr social tags. The underlying assumption here is that the better-known places will be easier to recall when a user reminisces about the event and the most popular places are usually a benchmark of the region.

3.3 Semantics of Social Profiles

As social contexts, the actor and attendee concepts together reflect the agent aspect of a lifelog event. While these two contexts may answer the "Who" is carrying out the event and "Who" else is involved, enhancing social context aims to enrich the social profiles of these agents. In our implementation the FOAF profile and the lifelogger's personal information in Facebook are combined.

FOAF profiles are datasets in LOD and contain personal information modeled in RDF. While FOAF profiles contain information about millions of persons including persons relevant to the event, lifeloggers' social profiles like Facebook contain more semantics which have been customized and might have higher correlation with lifelog events. When a user re-experiences events, social information can improve the understanding of the "Who" aspect. The combination of FOAF profiles and Facebook involves transforming the XML feeds from Facebook to RDF, the FOAF profiles and Facebook are integrated in the same data model for which the same vocabularies like the FOAF ontology are needed for consistent semantic representation. Finally the RDF statements are populated in the event model thus enhancing the semantic description of the social context for the event.

3.4 EventCube: Conceptual Architecture

In designing an application for enhancing the representation of multimedia lifelog events, we mimic the traditional behavior of users in organizing personal digital photos. In [11], users state that the most important feature of photo organization is to automatically place photographs into albums and as shown in [12], albums are suggested to be more desirable for image organization and retrieval. There is also evidence from memory science that organising our lives into events assists with recall of the past, which is done through these events [2]. Motivated by this, we propose an event enhancement architecture – EventCube – to enhance the descriptions of events. The architecture of EventCube is shown in Figure 2.

For the application of this architecture to lifelogs, we employed SenseCam (shown in Figure 1) and Bluetooth-enabled handsets plus GPS trackers as context-sensing devices. The processing of raw lifelog sensor data into enhanced events is in three steps: first, the user uploads sensor readings to the database where SenseCam images and other sensor streams are segmented into events. A keyframe is automatically selected as a thumbnail for each event. Second, the

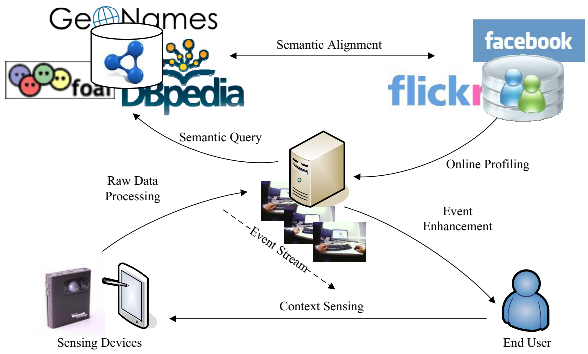

Fig. 2. Event enhancement architecture

recorded GPS coordinates are clustered (as described in Section 3.2) and stored together with Bluetooth proximity records. In this step, sensor readings are synchronized with segmented events. Finally, online knowledge bases and social profiles are accessed and combined to create the relevant semantics for current event contexts. These enhanced contents are provided as links for further use.

4 EventCube Experiments and Evaluation

4.1 Experiment Setup

SenseCam was used to collect image sequences while GPS recording and nearby Bluetooth device detection were implemented on a Nokia n810 internet tablet communicating with an external GPS module. GPS data was recorded every 10 seconds. The unique addresses and friendly device names of nearby Bluetooth devices were recorded every 20 seconds. All these sensor readings including SenseCam images were recorded with timestamps and synchronized. One subject wore the above recording devices for our initial event enhancement experiments. We processed logged data on a daily basis as the subject uploaded the SenseCam, GPS and Bluetooth readings after each day's continuous recording. For retrieval of physical information recorded by ambient sensing devices, we segmented each day's SenseCam data streams into individual events indexed by keyframes selected for their visual representation of events [3].

Besides local ambient information sources, the environment also includes the information space constructed by online semantic repositories and users' social profiles. Retrieval from online knowledge bases such as datasets in LOD enhances the "Who" and "Where" aspects of events. We used the web services listed in Table 1 to access the data sources.

4.2 Event-Centric Enhancement Application

We now present an event-centric enhancement application which includes event visualization and allows a user to browse through their own multimedia lifelog.

Table 1. Online data sources employed

Dataset	Web Service Endpoints	Event Aspects
DBpedia	http://dbpedia.org/sparql	Who, Where
GeoNames	http://www.geonames.org/export/	Where
Flickr	http://www.flickr.com/services/api/	Where
DBLP	http://dblp.l3s.de/d2r/	Who
Facebook	http://api.facebook.com/1.0/	Who

This is a browser-based application with a SenseCam event viewer, geospatial map and contextual enhancement embedded, shown in Figure 3.

The event viewer presents event keyframes allowing the user to view events on a day-by-day basis. The calendar on the left corner provides a navigation and a display option. After the user selects a target date, the event viewer will list all events for that day organized in chronological order. Figure 3 illustrates the temporal sequence when the user attended a presentation.

After the user picks an event s/he is interested in, the GPS location is queried and located on the map. To enhance location context, relevant place names are retrieved according to the event's GPS coordinates. The abstract information is shown in the browser as a brief description of the most relevant named place plus links from the web pages or RDF triple repositories. Social context is enhanced in the same manner with brief information and links to external semantics using DBLP and DBpedia. The temporal context is visualized with a timestamp indicating the start time, end time and duration of the selected event.

4.3 Assessing Context Enhancement

We use 25 consecutive days of lifelog data to evaluate our methodology for semantic enhancement of events. The final dataset includes 38,026 images, 327,244 GPS records and 45,898 Bluetooth detections involving 958 unique devices. We applied the location-clustering described in Section 3.2 to find significant places in the lifelog. Since user behavior is usually periodic over a relatively long period, images in the same cluster are more conceptually similar and can be well-represented using keyframes. GPS records were validated in order to filter invalid coordinates such as empty GPS records logged when satellite signals were not visible. Ultimately, 59,164 GPS coordinates were selected for location clustering and each day's locations were clustered with 2,400 coordinates on average.

For those detected significant places we enrich the location context by DBpedia and exemplar results are shown in Table 2, in which the abstracts (defined by predicate `dbpedia-owl:abstract`) and home pages (by `foaf:homepage`) are shown where available. After applying a SPARQL query, semantics about places can be retrieved from DBpedia. Besides abstracts and home pages, there may be dozens of other properties queried from DBpedia for location enhancement. Relevant properties about the target location could include the type of place,

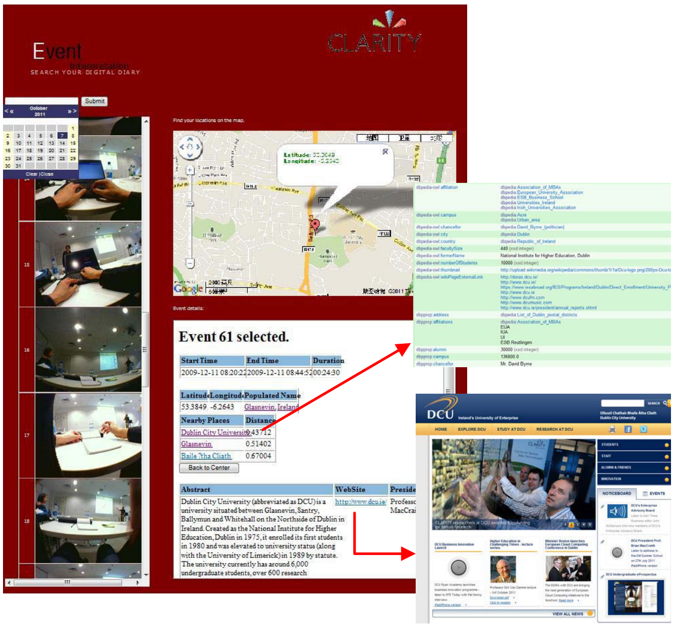

Fig. 3. Event enhancement interface (event viewer and map-enhancement browser)

the exact geospatial location information, affiliation, image, etc., which are all provided as links. As reflected in Table 2, we did not apply placename disambiguation before applying the enhancement. "The Spire"[1] is enhanced as a novel in Table 2, which is not the true interpretation of its meaning as a tourist attraction. However, the `dbpedia-owl:wikiPageDisambiguates` property allows users to navigate various options of resources with the same name as "The Spire" and to choose the correct one, which is described as "the Monument of Light . . . on O'Connell Street in Dublin, Ireland".

For social context while most benchmark locations can be queried from DBpedia, not many persons involved in the event can be enhanced in this way so we enhance the social context by combining resources from DBpedia, DBLP and Facebook. Sophisticated approaches to identifying users are beyond the scope of this paper and in our application we allow the user to map their real friends' names to Bluetooth friendly names. The purpose of this is to focus on event enhancement issues rather than user identification. Social context is then enhanced by querying relevant information from the aforementioned data sources by interlinking the friends' names to those data sets.

Table 3 shows some samples of enhanced social descriptions for lifelog events. For simplicity, we only illustrate the person abstracts obtained from DBpedia. The column for DBLP shows the number of records in DBLP datasets reflected

[1] The Spire is a large, spire-shaped public monument in the centre of Dublin city.

Table 2. Enhanced samples for places

Place Name	Abstract	Home Page
Dublin City Univ.	a University situated between Glasnevin, Santry, Ballymun and . . .	www.dcu.ie
Trinity College	formally known as the College of the Holy and Undivided Trinity of . . .	www.tcd.ie
Glasnevin	a largely residential neighborhood of Dublin, Ireland . . .	–
Baile Átha Cliath	capital and largest city of Ireland . . .	www.dublincity.ie
Croke Park	the principal stadium and headquarters of the Gaelic Athletic Association . . .	www.crokepark.ie
Merrion Square	a Georgian square on the southside of Dublin city centre . . .	–
The Spire	a 1964 novel by . . .	–

by the number of `dc:creator/foaf:maker` properties queried from DBLP. For the Facebook column we illustrate only the hometown defined in the aligned RDF models. Since these profiles are accessed from either the publicly-available LOD repository or the lifelogger's own Facebook account, there are no ethical issues when interlinking and using them in the enhancement application.

Table 3. Enhanced samples for social context

Bluetooth Name	DBpedia	DBLP	Facebook home town
NeilOHare-MacBook	–	13	Drogheda, Ireland
Alan Smeaton's MacBook Pro	Alan Smeaton is an author and academic at Dublin City University . . .	227	Dublin, Ireland
cdvpmini-AlansOffice	Alan Smeaton is an author and academic at Dublin City University . . .	227	Dublin, Ireland
cdvpminiColum	–	12	–
Pete	a British multimedia artist living in Newfoundland, Canada . . .	23	–

A similar problem to that caused by a lack of name disambiguation is mis-enhancement for commonly-used names. For example, the recorded person 'Pete' (Peter as real name), who was a colleague of our subject, is incorrectly enhanced when querying DBpedia. The characteristics of Bluetooth also cause another artifact for social descriptor enhancement. Bluetooth has a range of about 10 meters and in some cases it can penetrate walls. In our enhancement experiment, we rank the Bluetooth records in terms of their frequency of occurrence during the time span of the selected event. In this way, accidentally logged device proximities can be ranked lower and have less chance to be enhanced.

Using Semantic Web techniques, our enhancement leverages information retrieved from public resources by applying SPARQL queries and aligning semantics to a standardized RDF model. The populated profiles provide a comprehensive tool to realize detailed aspects about lifelog events.

5 Conclusion and Future Work

This paper introduced a method for semantic enhancement of multimedia lifelog events to improve their interpretation by leveraging external knowledge. The approach we have taken has only recently become feasible because a critical mass of semantic descriptions of people, places, and activities has now been modeled and published as Linked Data. The effectiveness of detecting significant places and accessing billions of triples in the LOD knowledge bases for enhancing location and social contexts has been demonstrated. In future work we will explore how we can improve our understanding of lifelog events by integrating more semantic information from richer contextual views, and investigate how to use such richer semantic structure to enhance the characterization of events.

References

1. Gurrin, G., Smeaton, A.F., Doherty, A.R.: Lifelogging: Personal Big Data. Foundations and Trends in Information Retrieval 8(1), 1–125 (2014)
2. Doherty, A.R., Pauly-Takacs, K., Caprani, N., et al.: Experiences of Aiding Autobiographical Memory Using the SenseCam. HCI 27(1-2), 151–174 (2012)
3. Doherty, A.R., Smeaton, A.F.: Automatically segmenting lifelog data into events. In: WIAMIS 2008, pp. 20–23. IEEE Computer Society, Washington, DC (2008)
4. Beckett, D.: Turtle-Terse RDF triple language. W3C Technical Report (2007)
5. Silva, A.R., Pinho, S., Macedo, L.M., Moulin, C.J.: Benefits of SenseCam Review on Neuropsychological Test Performance. AJPM 44(3), 302–307 (2013)
6. O'Loughlin, G., Cullen, S.J., McGoldrick, A., et al.: Using a wearable camera to increase the accuracy of dietary analysis. AJPM 44(3), 297–301 (2013)
7. Wang, P., Smeaton, A.F.: Using visual lifelogs to automatically characterise everyday activities. Information Sciences 230, 147–161 (2013)
8. Doherty, A., Smeaton, A.F.: Automatically augmenting lifelog events using pervasively generated content from millions of people. Sensors 10(3), 1423–1446 (2010)
9. Dobbins, C., Merabti, M., Fergus, P., et al.: Exploiting linked data to create rich human digital memories. Computer Communications (2013)
10. Ashbrook, D., Starner, T.: Using GPS to learn significant locations and predict movement across multiple users. Personal UbiCom 7(5), 275–286 (2003)
11. Rodden, K.: How do people organise their photographs? In: Proceedings of the BCS IRSG Colloquium (1999)
12. Platt, J.C.: AutoAlbum: Clustering digital photographs using probabilistic model merging. In: CBAIVL 2000, pp. 96–100. IEEE Computer Society (2000)
13. Little, S., Jargalsaikhan, I., Clawson, K., et al.: Interactive Surveillance Event Detection at TRECVid2012. In: ICMR 2013, pp. 301–302 (2013)
14. Reuter, T., Papadopoulos, S., Petkos, G., et al.: Social Event Detection at MediaEval 2013: Challenges, Datasets, and Evaluation. In: MediaEval Workshop (2013)
15. Lu, Z., Grauman, K.: Story-Driven Summarization for Egocentric Video. In: CVPR, Portland, OR, USA, pp. 2714–2721 (2013)