Social Event Detection–A Systematic Approach using Ontology and Linked Open Data with Significance to Semantic Links

Sheba Selvam, Ramadoss Balakrishnan, and Balasundaram Ramakrishnan
Department of Computer Applications, National Institute of Technology Tiruchirappalli, India

Abstract: *With the growing interest in capturing daily activities and sharing it through social media sites, enormous amount of multimedia content such as photographs, videos, texts, audio are made available on the web. Retrieval of multimedia content has now become a trivial task. Generally, people show interest in sharing photographs to a well-known closed community through social media sites like Flickr and Facebook. One solution to retrieve photographs is by identifying them as events. This task is known as Social Event Detection (SED). From the Flickr website, with the use of metadata like photoID, title, tags, description, date, time and geo-location for each photograph, the SED task is performed. As a central piece of the SED task, ontology for events domain is implemented. First half of the work is an explicit knowledge representation by constructing ontology for event detection using Protégé. Then, reasoning is done through HermiT reasoner and later SPARQL query is done to retrieve the media representing each event. The second half of the work involves in linking open description of specific events from different web services like Eventful, Last.fm, Foursquare, Upcoming and GeoNames. SPARQL query is done to measure the retrieval performance of each event after making semantic link using Linked Open Data (LOD). Finally an additional feature, the weather information for events is added, which shows removal of false positives in the SED task.*

Keywords: *Multimedia, social media, social events, photographs, event detection, ontology, linked open data, contextual metadata.*

Received August 15, 2015; accepted December 21, 2016

1. Introduction

The improvement and easy handiness of the digital technology, made usage of digital cameras such as Global Positioning System (GPS) enabled cameras, smart phones by the people to capture the daily activities and then store them as valuable memories. All the users' digital photos are thrown in the social media sites such as Flickr, Facebook, Picasa etc., and so the internet is dumped with huge information. Spending time and effort in organizing the photographs is not done by the users, which leads to large number of photographs being untouched. This creates the need for image indexing and retrieval.

Set of visible activities can be certainly expressed as events. According to [7], 'event is a specific thing happening at a specific time and place'. Day today activities, doings, meeting, etc., of an individual represent events. Events are focused in two different aspects in [7], one as local events and the other as global events. Simple and small personal experiences such as vacation trip, birthday celebration, meeting in an office, get-together in work environment, etc., exemplify local events and the world popular incidents and accidents such as sports match, disaster, public protest, medical achievement etc., represent global events.

The social events according to [16] are events that are planned by people, attended by people and the media illustrating the events are captured by people. A social event of interest can be specified in terms of event-related metadata (e.g., location, time, venue, and performers), example tags or other social information, example media items (images), or a combination of the above. Organizing the social photo collections can be done by making use of the contextual metadata and identifying them as events. This retrieval task is known as Social Event Detection (SED). Ontology based method is given as future work in [23] for better exploitation of textual metadata than the available similarity measures.

The proposed work designates the SED task, by exploiting the different sources of metadata coupled with photos on an image community website Flickr. For the SED task, our proposed work involves semantic analysis of textual metadata by constructing ontology, further with Linked Open Data (LOD) for gathering event related information and finally querying the results.

2. Motivation and Related work

The availability of social websites with huge amount of information dumped in it, created a new research line to identify multimedia as events, which is the SED task. Event types such as entertainment, personal, disastrous detection and landmark identification are usually focused in the SED task.

Topic Detection and Tracking (TDT) in 1998, is the initiative of the topic 'event detection' which is given in [1]. Primarily news event is organized from the web content. Identifying new events from the news feeds or identifying events from a collection started up. Following topic detection, different techniques for event detection rose up such as classification, clustering, algorithmic approach and tool based approach. SED started out with different mining prospect from the available content such as the comments, tags, descriptions accompanied with the photographs. Recently, with increase in the social websites like Flickr, Facebook, etc. many photographs along with their metadata are used for event detection [7, 18]. The proposals of many are on clustering algorithms with the content and time information [8, 12, 26]. The metadata for the photographs, include textual data such as tags, user comments, title and description along with date, time and location. So, simple similarity measures lacks in the accuracy of retrieval of photographs according to events. A simple Jaccard index is used for comparing tags and a better solution addressed by Vizuete *et al*. [23] is to use semantic analysis to handle textual information, which is proved to play a significant role. This twisted a new track to semantic-based approaches for SED.

According to proof given in [7], the Flickr pictures tagged in diverse methods and are not enough to match with the WordNet database, as WordNet is not fully updated with the current activities. This problem is also mentioned in [4] while using WordNet for text categorization.

Similarly Wikipedia devising different classification of events does not cover all descriptions needed for good organizing. Finally a move is made to YAGO [20, 21] ontology which combines knowledge of Wikipedia [24] and semantics from WordNet [25]. For semantic and context based techniques ontological knowledge has become very important. According to the work styled in [16], YAGO ontology does not cover all instances of activities and the availability of context variables such as location, time, and action should be made. So, a solution designed at this stage, is to build an ontology or knowledge base by mining large scale web resources. One more research line given by [10, 23] is ontology based approaches for semantic analysis to yield improvement in the results. In various applications such as classification, searching, context-based reasoning and word sense disambiguation, the importance of constructing ontology has been increased.

Motivated by this approach, constructing domain ontology for event detection is on the drive. News agencies in the news business produce a large amount of digital contents for their customers. For the contents to be managed and disseminated, ontology for news domain [5] has been implemented making use of the News Engine Web Services (NEWS) project [6].

For many semantic ambient media applications such as smart phones, automobiles, surveillance and others, events play a central role. After constructing ontology, the next move is to incorporate various knowledge sources using LOD as needed by the application. Linking of event descriptions is done with the LOD datasets such as DBpedia, Foursquare, Geonames, Last.fm, Eventful, Upcoming etc. Relevant information for events from these web services is linked to the ontologies to increase the semantic retrieval process. In [19], a model is recommended to publish events as linked data along with a prototype "event directory" web service, which locates stable URIs for events and also provide RDFS+OWL description of the events and links to the other related resources. This model is known as Linking Open Descriptions of Events (LODE). The goal is to model "factual" aspects of events, characterized as four Ws: What happened, where it happened, when it happened and who is involved. These four Ws relate to events, places, times and people, forming LODE ontology. In [11, 13, 14, 19, 22] LODE ontology is used to aggregate heterogeneous information from sources such as eventful.com, upcoming.org, last.fm / events, facebook. com/events, dbpedia.org etc., using linked data, to select events and to discover meaningful connections between them. The problem with this ontology is that, multiple textual descriptions cannot be made to same attribute. For e.g., for a soccer match, textual description may include football, soccer, to name a few may be available. These two textual descriptions represent the same event but it is not clear.

So, the proposed work involved in constructing domain ontology with available contextual metadata and creating semantic links with resources as needed for efficient SED. The focus of this work is on a domain of pictures, in specific on Flickr data. Also, one more feature is added to improve the detection process, the weather information at the time of event. For this, the Weather Underground Web Service, is used and finally, completed with progress in the retrieval result.

3. Problem Definition

The problem addressed in the paper is to identify social events by discovering the media items related to specific social event from the Flickr website. The media items referred are the contextual features (metadata) such as title, description, tags, geo-location, and date and time information for every event. This SED task is not with finding the textual description but aims for a set of photo clusters, where each cluster represents photographs for a single event. For identifying the photo clusters, no visual information was considered, instead, combination of available metadata were processed for finding the events. For e.g., consider photos of event x, where photos of gathering, inside the event venue during/ few moments before/ few moments after the event x, photographs of the gathering moving out at the end of the event x are all related to event x. On the other side, the photographs showing only the event venue without any gathering and relations to event x, are marked irrelevant photographs of event x.

4. Event Detection Ontology Description Scheme

Ontology is proposed for SED task as it is a formal and explicit representation of knowledge. Ontologies are collection of classes or concepts, properties of classes and relations between the classes for describing knowledge of domain and to achieve an effective ontological analysis of the domain. Web Ontology Language (OWL), an XML-RDF based language from the World Wide Web Consortium (W3C) is used for ontology construction.

Resource Description Framework (RDF) is a domain independent data model which is used for the development of web ontologies. Figure 1 shows the proposed domain ontology constructed for SED task, with querying and further enhanced by semantic linking of description of events for good organization of the photographs based on events.

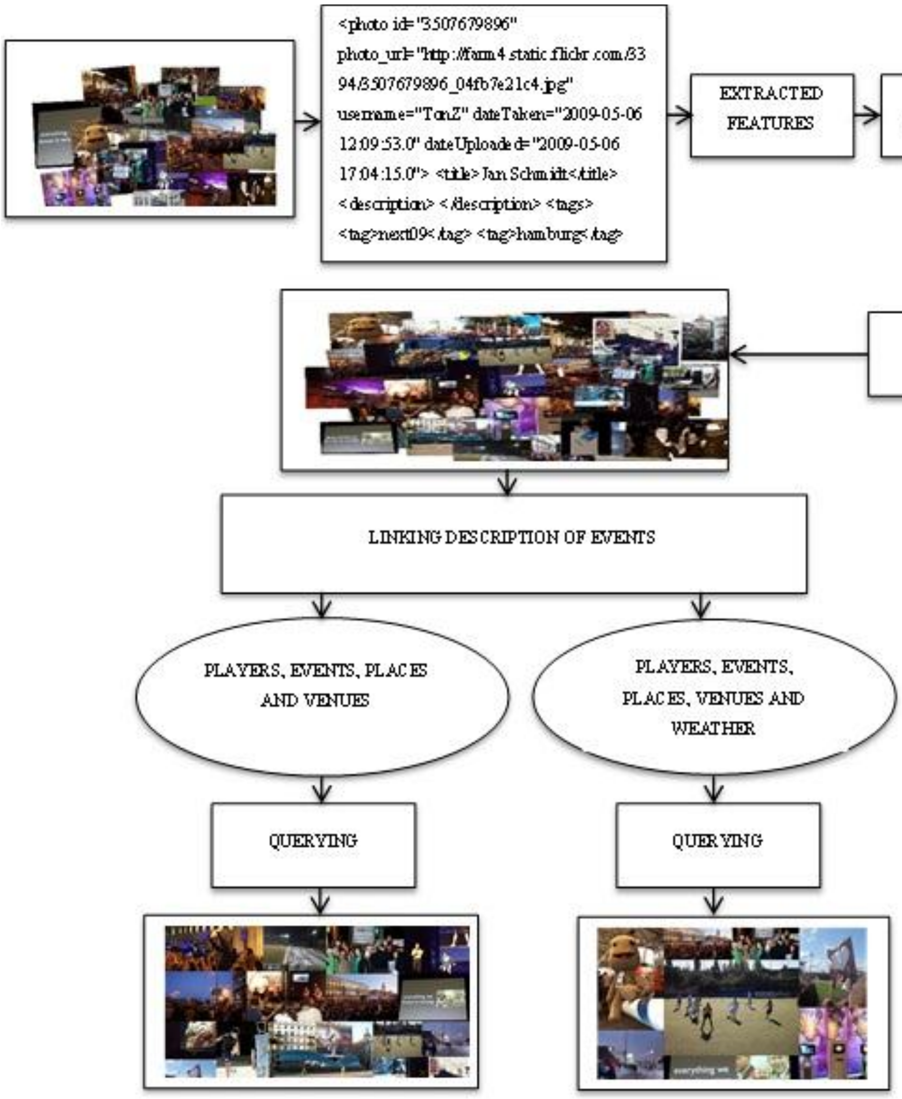

Figure 1. Proposed methodology for SED.

4.1. Key Factors of Event Detection Ontology

As aforementioned, event detection ontology is a formal and explicit representation of knowledge, which consists of three key factors: event concepts, concept properties and relations among the concepts. The ontology is represented as a triplet < C, P, R >, where, C, P, R denotes the event concepts, properties and relations.

Let $E_i = \{E_1, E_2 \ldots E_n\}$ be the events, where i =1, 2… n. C_{ij} be the classes of the events E_i such that j=1, 2 …m and SC_{ijk} be subclasses of the class C_{ij} for the event E_i where k=1, 2 ….q.

Let P_r be the object properties assigned to the classes in the events E_i such that r=1, 2...s and let D_t be the datatype properties assigned to the classes in the events E_i, where t=1, 2 …u.

4.2. Building an Event Detection Ontology

For integrating domain knowledge to the ontology, the concepts are first built followed with properties to describe the concepts and at last relations among the concepts and the properties.

4.2.1. Concepts in Event Detection Ontology

The event concepts act as the basis of the ontology by providing the terms to describe the fact in the event detection domain. In our case, event concepts represent the contextual metadata of events. As presented in Figure 2, the event photograph can be represented by contextual metadata "PhotoID", "EventName", "DateTime", "Country", "City", "Venue", "Latitude" and "Longitude". We define a class, event detection as a superclass of all the dimensions of information. The other concepts describing the event ontology includes Events, EventName, PhotoID, DateTime, EventPlace and EventLocation.

Figure 2. Illustration of an event photograph which is represented by Contextual Metadata: "3854127800", "Soccer", "2009-08-22 12:57:31.0", "Germany", "Hamburg", "FcStPauli", "53.554688" and "9.968639".

Some of the classes have subclasses to define the concepts. In this case, EventPlace has subclasses Country, City and Venue. Similarly, EventLocation has subclasses Latitude and Longitude.

$$\begin{aligned}
&Events && \sqsubseteq && EventDetection \\
&EventName && \sqsubseteq && EventDetection \\
&PhotoID && \sqsubseteq && EventDetection \\
&DateTime && \sqsubseteq && EventDetection \\
&EventPlace && \sqsubseteq && EventDetection \\
&EventLocation && \sqsubseteq && EventDetection \\
&Country && \sqsubseteq && EventPlace \\
&City && \sqsubseteq && EventPlace \\
&Venue && \sqsubseteq && EventPlace \\
&Latitude && \sqsubseteq && EventLocation \\
&Longitude && \sqsubseteq && EventLocation
\end{aligned}$$

4.2.2. Properties in Event Detection Ontology

Event properties are used to characterize the event concepts within the ontology definition and helps in event detection. The properties in this ontology are a part of OWL specification which represents relationship. Two types of properties are used as shown in Figure 3. One of which, the Object Property, describes relationships between two individuals.

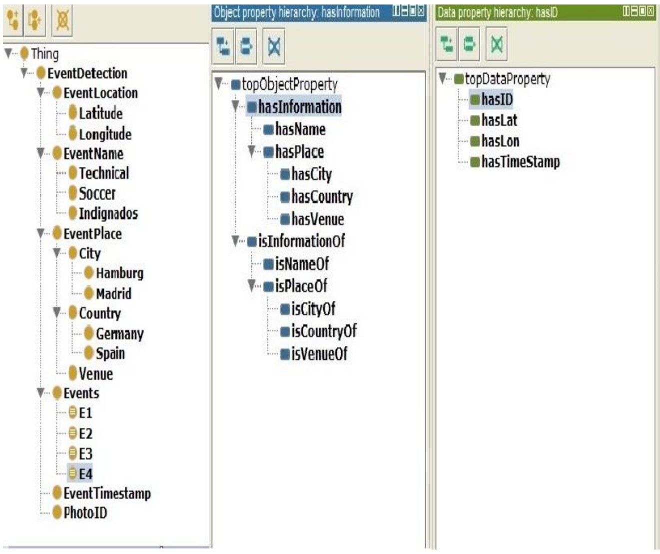

Figure 3. Representing object properties and datatype properties for SED using protégé.

The object properties used for this event ontology are hasName, hasCountry, hasCity and hasVenue. These object properties describe relationship between classes. Each object property also has a corresponding inverse object property. A class is denoted by representing an event with object properties hasName, hasCountry, hasCity and hasVenue along with corresponding inverse properties is Name Of, is Country Of, is City Of and is Venue Of.

$$\begin{aligned}
&Is\ NameOf && \equiv && has\ Name^{-} \\
&Is\ CountryOf && \equiv && has\ Country^{-} \\
&Is\ CityOf && \equiv && has\ City^{-} \\
&Is\ VenueOf && \equiv && has\ Venue^{-}
\end{aligned}$$

The second type is the Datatype Property, which links an individual to a datatype value. The datatype properties used for this event ontology are hasID, hasDateTime, hasLatitude and hasLongitude. Similar to object property, datatype property also has the corresponding inverse property. The inverse datatype properties of the constructed ontology include isIDOf, isDateTimeOf, isLatitudeOf and isLongitudeOf.

$$Has\ ID \equiv is\ IDOf^{-}$$

$$\begin{aligned}
&has\ DateTime && \equiv && is\ Date\ Time\ Of^{-} \\
&has\ Latitude && \equiv && is\ Latitude\ Of^{-} \\
&hasLongitude && \equiv && isLongitudeOf^{-}
\end{aligned}$$

4.2.3. Relations in Event Detection Ontology

A single photograph contains a finite number of high level semantic content. The semantic content described with the photographs include the contextual metadata (“Photo ID,” “Event Name,” “Date Time,” “Country,” “City,” “Venue,” “Latitude” and “Longitude”) as shown in Figure 2. Relations help to capture the semantics among the concepts. Assigning properties among the concepts make relations to the ontology.

The relations in the Figure 4 describe the semantic levels of the proposed ontology as onto graph, where, all the elements needed for the EventDetection is used. For an event, the properties such as hasName, hasCountry, hasCity and hasVenue, hasDateTime have universal restrictions.

$E_i \rightarrow \forall P_r$

$E_i \rightarrow \forall D_t$

For e.g., for event E1, the universal restriction is given by,

$E_1 \rightarrow \forall hasName\ Soccer$

$E_1 \rightarrow \forall hasCountry\ Germany$

$E_1 \rightarrow \forall hasCity\ Hamburg$

$E_1 \rightarrow \forall hasVenueFcStPauli$

$E_1 \rightarrow \forall hasDateTime$ *“2009-08-22 12:57:31.0”*

For events E_i, with the classes representing the events C_{ij}, the value of one event is disjoint with the other event. i.e., no individual can be at the same time an instance of both CE_i and CE_j, $i \neq j$. DateTime, EventName and Venue of one event should be disjoint with the other event values. For an event, the EventName and Venue are necessary and the necessary and sufficient conditions include event name, date time and venue.

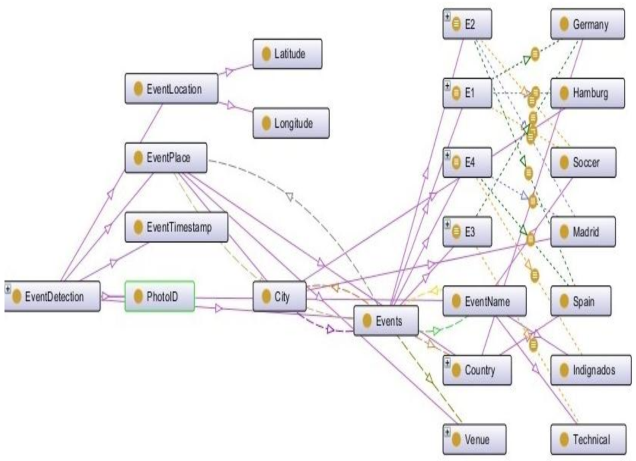

Figure 4. Onto graph representing relations in the ontology.

After construction of the domain ontology, querying is done to retrieve the related media for each event. For this SPARQL query in protégé is used. SPARQL [17] is a standardized query language designed for Resource Description Framework (RDF) databases.

4.3. Linked Open Data

Adding data into the web does not create a semantic web. If a specific term is in requirement, search is essential possibly on that term and finally concludes with lots of outcome. Later the exact word or term is reached on an individual website. The web service helps by disambiguating the search in the start and then all categories of related information is connected which is updated dynamically within the same web space. For this to happen, LOD is needed.

Tim Berners-Lee [2] the inventor of the World Wide Web (WWW) achieves LOD by 4 steps.

1. The data should be open to interrupt.
2. The data is placed on the web and the connecting links are represented in a simplest form. Each of these links is represented using RDF or RDF triple because it has three parts and this triples helps to interconnect.
3. When the data is online, a unique address is needed, Unique Resource Identifier (URI).
4. To put them online a standard web resource protocol, Hypertext Transfer Protocol (HTTP) is required.

Now the LOD can be connected to other LOD, creating a much richer knowledge network. To create high quality domain ontology, with the defined classes and properties, we connect richer knowledge by the available LOD sources needed for the domain ontology.

The work discussed by Hintsa *et al*. [9] is denoted as Model 1, which is an event detection approach using LOD, by linking data sources to get additional information relating to events, artists, venues and places. From the event services available such as Last.fm and Upcoming, the information relevant to events (description, title, venue, time and artist information) were stored in a database. Using GeoNames and Freebase additional information relating to places, including the coordinates for the venues and cities was got. This updated the event profile, to increase more relevant matches between events and the photos. In [9] matching against the database was completed, but lacks in semantically analysing the tags, which can still more determine the place-related tags and remove false positives in the outcomes.

The proposed work is described as Model 2. In Model 2, semantic analysis is done by constructing ontology followed by linking event related information by LOD and SPARQL query is done to measure the performance. The work in Model 2 is compared with similar event detection approach using LOD that is described as Model 1. The Lack in Model

1 is the semantic analysis of the contextual information which can outperform the database matching. The proposed work, described as Model 2, involves in semantic analysis of tags by constructing SED ontology.

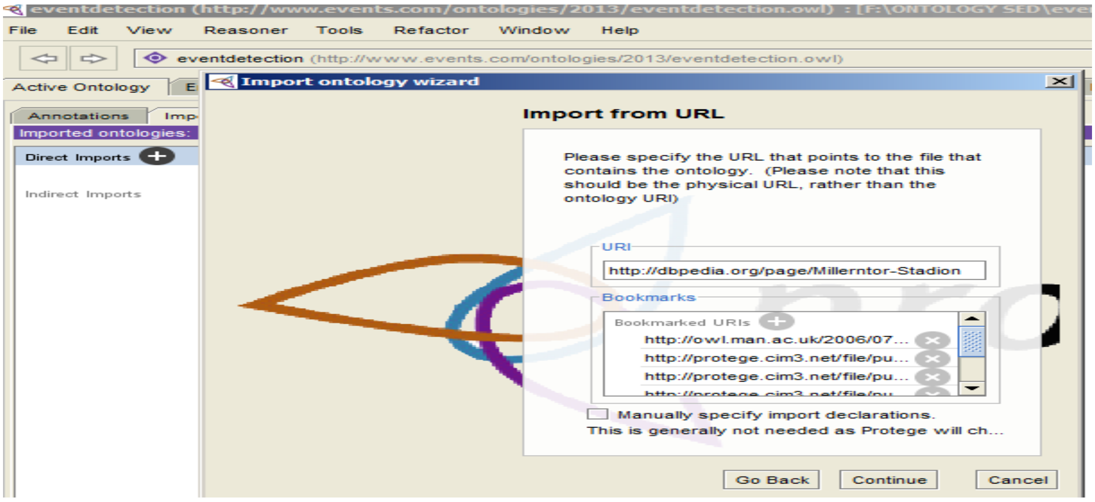

Figure 5. Linking LOD to protégé ontology.

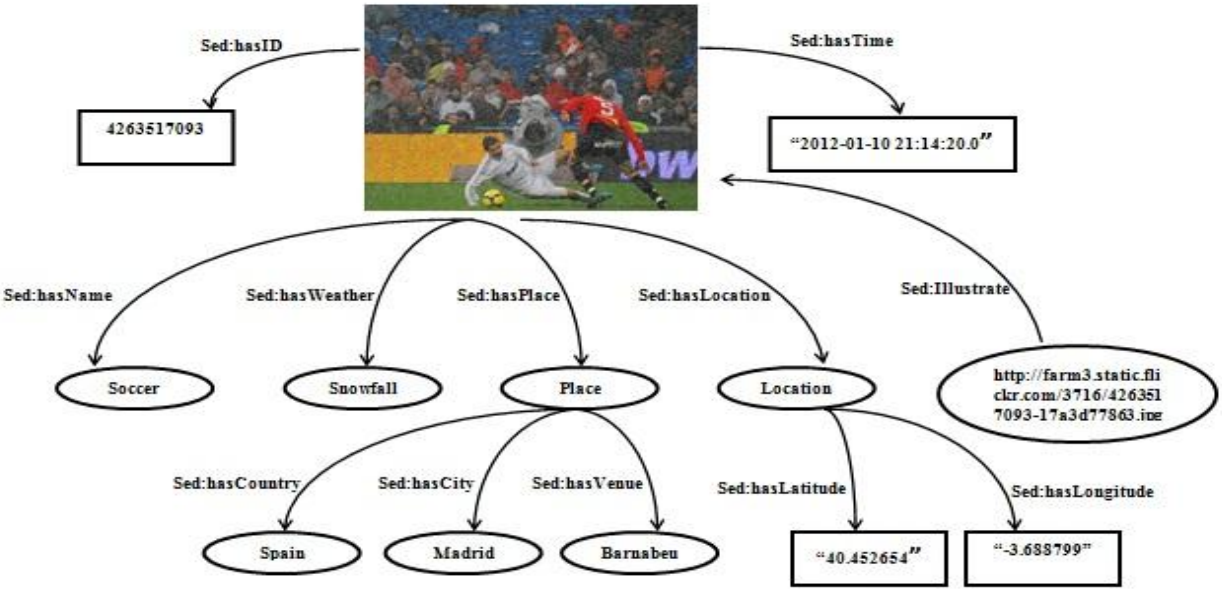

Figure 6. Illustrating Media describing a soccer event at bernabeu.

Further the work is extended by linking the relevant event information from the LOD sources as given by [9] into the event ontology.

Information relating to events, players, places and venues are linked to the ontology to create relevancy between the events and the photographs. Linking the other sources to the ontology is carried out by importing the URI's as shown in Figure 5. The needed resource alone is linked using linked data, rather than concerning the whole linked data which is unnecessary. The results of Model 1 and Model 2 are tabulated in Table 1.In the social websites such as Flickr, Twitter, Facebook, etc. the environmental conditions and changes during the event are also commented by the public. For e.g., in a photograph of an event, the contextual information available is "4263508963, 2010-01-10 21:31:35.0, Snow in Bernabeu, Party with heavy snowfall Real Madrid 2 - Mallorca 0.The photos do not have grain, is the snow, Real Madrid, bernabeu, Cristiano Ronaldo". The contextual information describes the weather condition "snowfall" during the event. Also in [15], weather information was a feature to organize the photographs. This motivated the need of weather information in the proposed work. The information of how the weather condition, of the event is linked to the photograph is described later in this section. Figure 6 illustrates the event related information linked to the photograph with PhotoID "4263517093", DateTime " 2012-01-10 21:14:20.0", EventName "Soccer", Country " Spain", City "Madrid", Venue "Bernabeu", Latitude "40.452564", Longitude "-3.688799" and Weather "Snowfall".

One of the contextual metadata given in [15], to describe an event was the weather condition. With the help of the Weather Underground Web Service, the weather information was cached and used for photographs that appear between two hours before and after the photo. This was because a single photo can be related with two weather conditions, (rainy and cloudy). So average of temperatures was used to know the weather conditions and temperature. Motivated with the idea given in [15], to the ontology constructed with LOD, (Model 2 Without Weather information), one more feature, the weather information of the place where the event took place is added. With the help of the Weather Underground Web Service, the (zip-code, date) pair or a (city name, date) pair or a (weather station, date) pair is queried on the same date in the same venue, to get the weather information. The problem faced was, even with average of temperatures, two weather conditions were got. For e.g., when the calculated average temperature was 7◦ C, the weather conditions got was (cloudy, overcast). So not only average of temperatures is taken into consideration, even, the average of wind chill, humidity and pressure is considered. The time considered was an hourly basis during the event along with an hour before and after the start and end of the event.

Let T_o represent temperature, W_o the wind chill, H_o indicates the humidity and P_o the pressure for the events E_i. The weather information is derived with the average of all the values, which is given by the equations below.

$$T_o = \frac{1}{n}\sum_{x=1}^{n} T_x \quad (1)$$

$$W_o = \frac{1}{n}\sum_{x=1}^{n} W_x \quad (2)$$

$$H_o = \frac{1}{n}\sum_{x=1}^{n} H_x \quad (3)$$

$$P_o = \frac{1}{n}\sum_{x=1}^{n} P_x \quad (4)$$

By considering the average values of temperature, wind chill, humidity and pressure the weather information is got and added as additional information related to events. By considering the average values of temperature, wind chill, humidity and pressure the weather information is got and added as additional information related to events. Querying result, given in Table 2, after linking the weather information is increased. Also, some of the false positives are removed as shown in the Figure 9. The photographs in Figure 7 represent photographs which do not belong to an event, but contains contextual information similar to the photographs belonging to an event.

Figure 7. Photographs signifying false positives.

For e.g., a photograph captured during a technical event in Germany, has the contextual information, PhotoID “4905176789”, DateTime “2010-08-18 17:22:17.0”, EventName “Gamescom”, Country “Germany”, City “Cologne”, Venue “Koln Messe”. Another photograph which does not belong to the technical event has the contextual information, PhotoID “4923396302”, DateTime “2010-08-19 11:31:33.0”, EventName “Gamescom”, Venue “Messe”. Here, the DateTime is different with some other information missing. In this case, weather information is calculated as described earlier in this section and the false photographs are removed. Even when detection is between multiple events happened in one city, the weather based filtering approach will work as all the events will not happen at same time and same day. Even if multiple events happen the same day in one city, SED is achieved as other contextual features are in combination with the weather information. This proves that by adding weather information, surprising heterogeneous relationship is aggregated to the events.

5. Experimental Results

5.1. Challenges

The three challenges of the SED 2012 task were on a collection of images collected from Flickr. The first challenge we considered was “Find all soccer events taking place in Hamburg (Germany) and Madrid (Spain) in the test collection”. This challenge was considered into two events. Event 1, Soccer events taking place in Hamburg (Germany) and Event 2, Soccer events taking place in Madrid (Spain). The text challenge was “Finding technical events that took place in Germany in the test collection.” Event 3 was to find the technical events in Germany. Technical events, for the purpose of this task, are public technical events such as exhibitions and fairs. The third challenge reads “Finding demonstration and protest events of the Indignados movement occurring in public places in Madrid in the test collection.” Event 4 was to find the Spanish Indignados movement. This task centres on a series of demonstrations of the Indignados movement and other protests taking place all over Spain in 2011-2012. Protest events relate to the financial crisis outbreak as well as national politics in general.

5.2. Dataset

A collection of 70,000 photos were considered where all were captured between the beginning of 2009 and the end of 2011. For all the photos, the XML metadata includes Photo ID, Photo Url, User Name, Date Taken, Date Uploaded, Title, Description, Tags (user-supplied comments), latitude and longitude. A sample event with its contextual metadata, '3854127800, Soccer, 2009-08-22 12:57:31.0, Germany, Hamburg, FcStPauli, 53.554688 and 9.968639', as shown in Figure 2 are the instances that exists for the 'Soccer' event. Initially, to the dataset, pre-processing and feature extraction is done. The contextual information in the XML file is in many different languages, so there is a necessity for translating and cleaning the data. This text processing is needed for the SED task to group the photographs to events, since the processing is done with the metadata. The pre-processing step started with translating the text to English language using Google Translate API. Next punctuations, stop words like you, and, of, the etc. are discarded. The words are split into tokens as the text assigned is not clean and the process of stemming is done. With this pre-processed data, the feature extraction continued with the textual features like title, description, tags along with photo ID and date. The features extracted contains information about event name, country name, city name, venue name, date, time of the event, latitude and longitude values of the event. This SED task is not with finding the textual description but aims for a set of photo clusters, where each cluster represents photographs of a single event. For identifying the photo clusters, no visual information is considered, instead, the available metadata are processed for finding the events.

5.3. Evaluation and Ground Truth

The ground truth for all the three challenges is given by the EventMedia associations through which the performance is evaluated. Three evaluation measures, used to evaluate the results are precision, recall and f-measure.

- *Precision*: Precision is calculated by identifying correctly detected images w.r.t events by correctly detected images and falsely detected images w.r.t events.
- *Recall*: Recall is calculated using correctly detected images w. r. t events by correctly detected images and missed images w.r.t events.
- *F- measure*: F-measure is the harmonic average of precision and recall.

Table 1, describes the performance of existing work, Model 1 with the proposed work, Model 2. The performance of Model 2 is increase due to the semantic analysis by constructing ontology.

Table 1. Event detection efficiency by linking descriptions of events.

Challenges	**Model 2 (LOD by semantic analysis to ontology)**			**Model 1 (LOD by matching against the DB)**		
Events	Precision	Recall	F-Measure	Precision	Recall	F-Measure
Event 1	0.8517	0.8281	0.8397	0.8318	0.8056	0.8184
Event 2	0.8508	0.8262	0.8383	0.8326	0.8042	0.8181
Event 3	0.8502	0.8254	0.8376	0.8307	0.8035	0.8168
Event 4	0.85	0.8249	0.8372	0.8301	0.8029	0.8162

Table 2. Event detection efficiency with and without weather information-comparison.

Challenges	**Model 2(With Weather Information)**			**Model 2(Without Weather Information)**		
Events	Precision	Recall	F-Measure	Precision	Recall	F-Measure
Event 1	0.8572	0.8272	0.8419	0.8517	0.8281	0.8397
Event 2	0.8530	0.8251	0.8388	0.8508	0.8262	0.8383
Event 3	0.8517	0.8242	08377	0.8502	0.8254	0.8376
Event 4	0.8524	0.8238	0.8378	0.85	0.8249	0.8372

This is because the semantics defined by the ontologies hold richer semantics and also more expressive. Also ontology pulls out the needed information automatically resulting in the recall value to be increased. Table 2, tabulates the proposed work, Model 2 with and without weather information. The weather condition is one of the important clues to describe an event. The climatic condition strongly gets registered in a user mind while participating in an event. So with weather condition added to the contextual metadata, the semantics captured by the relations made among the concepts increased the retrieval performance. Some of the false positives are removed and the value of recall is increased.

6. Conclusions

The huge multimedia information available in the social sites such as Flickr, Facebook, Twitter etc., increased the want of SED task. The existing methods discuss the drawbacks in the SED task, along with the need and motivation of semantic- based approach. The proposed work is ontology based approach with LOD to create semantic links of extra event related information gathered from web sites such as eventful.com, upcoming.org, last.fm/events, and facebook.com/events to name a few. SPARQL querying is done which holds the semantics defined by the ontologies which has richer semantics and more expressive. Then the additional feature, the weather information added, its importance and how the weather condition for an event is gathered are also discussed. The retrieval performance is increased when compared with the existing database matching technique. The results tabulated confirm the significance of the semantic based approach with ontology construction, which holds richer semantics and automatically pulls the needed information, with the false positives to be removed. Our proposed work is useful to the media, location-based applications and also in the field of investigations to a particular activity.

References

[1] Allan J., Papka R., and Lavrenko V., "On-line New Event Detection and Tracking," *in Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Melbourne, pp. 37-45, 1998.

[2] Berners-Lee T., Linked Data-design Issues, http://www.w3.org/DesignIssues/LinkedData, Last Visited, 2011.

[3] Chen L. and Roy A., "Event Detection From Flickr Data Through Wavelet-Based Spatial Analysis," *in Proceedings of the 18th ACM Conference on Information and Knowledge Management*, Hong Kong, pp. 523-532, 2009.

[4] Elberrichi Z., Abdelattif R., and Mohamed B., "Using Wordnet for Text Categorization," *The International Arab Journal of Information Technology*, vol. 5, no. 1, pp. 16-24, 2008.

[5] Fernández N., Fuentes D., Sánchez L., and Fisteus J., "The News Ontology: Design and Applications," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8694-8704, 2010.

[6] Fernández N., Blázquez J., Fisteus J., Sánchez L., Sintek M., Bernardi A., Fuentes M., Marrara A., and Ben-Asher Z., "News: Bringing Semantic Web Technologies into News Agencies," *in Proceedings of The Semantic Web-ISWC*, Athens, pp. 778-791, 2006.

[7] Firan C., Georgescu M., Nejdl W., and Paiu R., "Bringing Order to Your Photos: Event-Driven Classification of Flickr Images based on Social Knowledge," *in Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, Hanover, pp. 189-198, 2010.

[8] Fung G., Yu J., Yu P., and Lu H., "Parameter Free Bursty Events Detection in Text Streams," *in Proceedings of the 31st International Conference on Very Large Data Bases*, Trondheim, pp. 181-192, 2005.

[9] Hintsa T., Vainikainen S., and Melin M., "Leveraging Linked Data in Social Event Detection," *in Working Notes Proceedings of the Media Eval Workshop*, Pisa, 2011.

[10] Jung Y., Ryu J., Kim K., and Myaeng S., "Automatic Construction of a Large-scale Situation Ontology by Mining How-to Instructions from the Web," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 8, no. 2, pp. 110-124, 2010.

[11] Khrouf H. and Troncy R., "EventMedia: A LOD Dataset of Events Illustrated with Media," *Semantic Web*, vol. 7, no. 2, pp. 193-199, 2016.

[12] Li Z., Wang B., Li M., and Ma W., "A Probabilistic Model for Retrospective News Event Detection," *in Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Salvador, pp. 106-113, 2005.

[13] Liu X., Troncy R., and Huet B., "Finding Media Illustrating Events," *in Proceedings of the 1st ACM International Conference on Multimedia Retrieval*, Trento, pp. 58, 2011.

[14] Liu X., Troncy R., and Huet B., "Using Social Media to Identify Events," *in Proceedings of the 3rd ACM SIGMM International Workshop on Social Media*, Scottsdale, pp. 3-8, 2011.

[15] Naaman M., Harada S., Wang Q., Garcia-Molina H., and Paepcke A., "Context Data in Geo-Referenced Digital Photo Collections," *in Proceedings of the 12th Annual ACM International Conference on Multimedia*, New York, pp. 196-203, 2004.

[16] Papadopoulos S., Schinas E., Mezaris V., Troncy R., and Kompatsiaris I., "The 2012 social Event Detection Dataset," *in Proceedings of the 4th ACM Multimedia Systems Conference*, Oslo, pp. 102-107, 2013.

[17] Prud E. and Seaborne A., Sparql query language for rdf, http://www.w3.org/TR/rdf-sparqlquery, Last Visited, 2006.

[18] Rattenbury T., Good N., and Naaman M., "Towards Automatic Extraction of Event and Place Semantics from Flickr Tags," *in Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Amsterdam, pp. 103-110, 2007.

[19] Shaw R., Troncy R., and Hardman L., "Lode: Linking Open Descriptions of Events," *in Proceedings of the Semantic Web*, Shanghai, pp. 153-167, 2009.

[20] Suchanek F., Kasneci G., and Weikum G., "Yago: A Large Ontology from Wikipedia and Wordnet," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 6, no. 3, pp. 203-217, 2008.

[21] Suchanek F., Kasneci G., and Weikum G., "Yago: A Core of Semantic Knowledge," *in Proceedings of the 16th International Conference on World Wide Web*, Banff, pp. 697-706, 2007.

[22] Troncy R., Malocha B., and Fialho A., "Linking Events with Media," *in Proceedings of the 6th International Conference on Semantic Systems*, Graz, pp. 42, 2010.

[23] Vizuete D., Gris-Sarabia I., and Giro-i-Nieto X., "Photo Clustering of Social Events by Extending Photo TOC to a Rich Context," *in Proceedings of the 1st International Workshopon Social Events in Web Multimedia in Conjunction with the ACM Conference on Multimedia Retrieval*, Glasgow, pp. 11-18, 2014.

[24] Wikipedia, A Free, Web-based Collaborative, Multilingual Encyclopedia, http://en.wikipedia.org, Last Visited, 2017.

[25] WordNet, A Large Lexical Database, http://wordnet.princeton.edu/, Last Visited, 2017.

[26] Yang Y., Pierce T., and Carbonell J., "A Study of Retrospective and on-line Event Detection," *in Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Melbourne, pp. 28-36, 1998.

Sheba Selvam completed her BTech (Hons.) in Information Technology and ME in Computer Science and Engineering. She is currently a Research Scholar in the Department of Computer Applications, National Institute of Technology, Tiruchirappalli (NITT). She works in social media retrieval and her areas of interest include multimedia and information retrieval.

Ramadoss Balakrishnan earned doctorate in Applied Mathematics from Indian Institute of Technology, Bombay India, in 1983 and is awarded M.Tech. in Computer Science & Engineering from Indian Institute of Technology, Delhi in 1995.Currently working as Professor in Computer Applications at National Institute of Technology, Tiruchirappalli, Tamil Nadu, India India(Deemed University, funded by Government of India). Recipient of Medal for "Distinction in Engineering & Technology" and cash prize by the Central Board of Irrigation & Power, New Delhi in 1983 for a research paper published in the Irrigation & Power Journal. Also recipient of "Best Teacher Award" in Computer Application for the year 2006-2007 from National Institute of Technology, Tiruchirappalli.. Having teaching & research experience of more than 25 years with more than 40 research publications in refereed journals & reputed International conferences. Research areas include Data Mining; Networks; Software Testing.

Balasundaram Ramakrishnan completed his MCA and ME in Computer Science and Engineering. He obtained his PhD in the area of e-learning. He is working in the Department of Computer Applications, National Institute of Technology (NIT) as an Associate Professor. He joined NIT (formerly known as Regional Engineering College) during 1987. His areas of interests include object oriented analysis and design, internet and web technologies, HCI and mobile data management. He is involved in the development of e-learning course materials.