Knowledge-Based Systems xxx (xxxx) xxx

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Generic metadata representation framework for social-based event detection, description, and linkage☆

Minale A. Abebe [a], Joe Tekli [b,*], Fekade Getahun [c], Richard Chbeir [d], Gilbert Tekli [e]

[a] College of Business and Economics, Addis Ababa University, 3131 Addis Ababa, Ethiopia
[b] School of Engineering, E.C.E. Department, Lebanese American University, 36 Byblos, Lebanon
[c] Computer Science Department, Addis Ababa University, 1176 Addis Ababa, Ethiopia
[d] LIUPPA Laboratory, University Pau & Pays Adour, 64000 Anglet, France
[e] Fac. of Technology, Mechatronics Department, University of Balamand, 100 Tripoli, Lebanon

HIGHLIGHTS

- Performs semantic-aware event detection, description, and linkage from social media data.
- Represents heterogeneous data in generic model made of temporal, spatial, & semantic dimensions.
- Evaluates data similarity using combined temporal, spatial, and semantic similarity measures.
- Detects events from similar social media objects using adapted unsupervised learning algorithm.
- Describes events in generic model and identifies their directional, metric & topologic relations.

ARTICLE INFO

Article history:
Received 6 January 2019
Received in revised form 7 May 2019
Accepted 25 June 2019
Available online xxxx

Keywords:
Social media
Metadata
Semantics
Similarity evaluation
Event detection
Event relationships
Collective knowledge

ABSTRACT

Various methods have been put forward to perform automatic social-based event detection and description. Yet, most of them do not capture the semantic meaning embedded in online social media data, which are usually highly heterogeneous and unstructured, and do not identify event relationships (e.g., car accident temporally occurs *after* storm, and geographically occurs *near* soccer match). To address this problem, we introduce a generic Social-based Event Detection, Description, and Linkage framework titled SEDDaL, taking as input: a collection of social media objects from heterogeneous sources (e.g., Flickr, YouTube, and Twitter), and producing as output a collection of semantically meaningful events interconnected with spatial, temporal, and semantic relationships. The latter are required as the building blocks for event-based Collective Knowledge (CK) organization, where CK underlines the combination of all known data, information, and metadata concerning a given concept or event. SEDDaL consists of four main modules for: i) describing social media objects in a generic Metadata Representation Space Model (MRSM) consisting of three composite dimensions: temporal, spatial, and semantic, ii) evaluating the similarity between social media objects' descriptions following MRSM, iii) detecting events from similar social media objects using an adapted unsupervised learning algorithm, where events are represented as clusters of objects in MRSM, and iv) identifying directional, metric, and topological relationships between events following MRSM's dimensions. We believe this is the first study to provide a generic model for describing semantic-aware events and their relationships extracted from social metadata on the Web. Experimental results confirm the quality and potential of our approach.

1. Introduction

Nowadays, emerging technologies such as Smart-phones, Wireless Internet, as well as Web and Mobile Services allow users to create, annotate, and share social data on the Web at an unprecedented and increasing pace. These technologies have transformed users from static data consumers during the 1990s (i.e., accessing static Web pages) to intelligent producers and proactive sensors of information during the 2010s (i.e., producing

☆ No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.knosys.2019.06.025.

* Corresponding author.
E-mail addresses: minale.ashagrie@aau.edu.et (M.A. Abebe), joe.tekli@lau.edu.lb (J. Tekli), fekade.getahun@aau.edu.et (F. Getahun), richard.chbeir@univ-pau.fr (R. Chbeir), gilbert.tekli@balamand.edu.lb (G. Tekli).

https://doi.org/10.1016/j.knosys.2019.06.025
0950-7051/

blogs, publishing and annotating images and videos, commenting on tweets, posting opinions, etc.), where most information being shared is multimedia and associated to events [1]. Yet, attaining the next stage in Web engineering, i.e., the so-called Intelligent Web: allowing meaningful human–machine and machine–machine collaboration within a ubiquitous computing environment, requires another breakthrough: allowing the sharing and organization of collective knowledge (CK) [2], where CK underlines the combination of all known data, information, and metadata concerning a given concept or event. In this context, the first step would be to extract and describe the meanings of events and their relationships, to be able to organize their CK later on.

There is no universal definition of an event, but an intuitive notion usually adopted on the Wed and in social media is that of a *social-based event* which can be viewed as a given observable occurrence at a certain time and place that interests a group of people (e.g., soccer match, car accident, heavy storm, presidential debate) [3,4]. Usually participants or observers of an event capture multimedia data (image, video, audio, etc.), annotate, publish, and share them online to describe the event (e.g., videos from the *soccer match*, pictures of the *storm*, opinions about the *presidential debate*, etc.) [5]. However, annotations of similar social media objects (e.g., similar images taken about the same *storm*) might be heterogeneous both in content and format, and would depend on the knowledge and experience of the annotator (e.g., an expert meteorologist would describe a *storm* or a *heat wave* differently from a non-expert observer). Hence, handling diverse and heterogeneous social media descriptions to identify and describe meaningful events remains a major problem.

In this context, various methods have been put forward to perform automatic social-based event detection (cf. literature review in Section 3). Yet, most of them do not capture the semantic meaning (concept definitions) associated with social media data and only focus on their syntactic textual descriptions (e.g., term-frequency weighting), thus missing their semantic relatedness, e.g., [6–8]. Also, most existing methods do not address the issue of identifying meaningful relationships between events (e.g., car accident temporally occurs *after* storm, and geographically occurs *near* soccer match), which we consider as a central requirement toward event-based CK organization. In addition, most methods are domain dependent and consider certain kinds of application specific information (e.g., tweets only, photos only), e.g., [7,9,10], without providing formal definitions of the temporal, spatial, and textual features considered, such as feature dimensionality, points of origin, granularity, coverages, and dedicated similarity measures.

Hence, a new approach is needed to effectively describe social media objects in a generic representation model with formal definitions of all relevant features and their properties, considering the unstructured and noisy nature of the data, in order to detect and describe events and their relationships. For this purpose, we introduce our Social-based Event Detection, Description and Linkage framework titled SEDDaL, taking as input: a collection of social media objects from heterogeneous sources, and then producing as output a collection of semantically meaningful events interconnected with meaningful relationships. SEDDaL consists of four main modules for: (i) describing social media objects in a generic Metadata Representation Space Model (MRSM) consisting of three composite dimensions: temporal, spatial, and semantic, (ii) evaluating the similarity between social media object descriptions following MRSM, (iii) detecting events from similar social media objects using an adapted unsupervised learning algorithm, where events are represented as clusters of objects described in MRSM, and (iv) identifying directional, metric, and topological relationships between events following MRSM's dimensions. We believe this is the first study to formally define a generic model (with its dimensions, properties, and similarity measures), for detecting and describing semantic-aware social media events and identifying their relationships.

The rest of the paper is structured as follows. Section 2 describes some basic concepts, and then presents a motivating scenario highlighting the main requirements toward event-based CK organization. Section 3 briefly reviews methods related to event detection from online social media data. Section 4 describes our SEDDaL framework and its different modules. Experimental results are described in Section 5, before concluding in Section 6 with future research directions.

2. Background, motivation, and requirements

In this section, we first provide a brief background description of some of the main concepts related to event-based CK organization (Section 2.1), and then describe a motivation scenario (Section 2.2) highlighting some of the main needs and requirements (Section 2.3) that we aim to fulfill in our proposal.

2.1. Event-based Collective Knowledge (CK)

To better understand the issues and challenges of event-based CK organization on the Web, we first need to distinguish the concepts of: *data*, *information*, *metadata*, *knowledge*, and *event*. The main difference lies in the level of abstraction of each concept. *Data* is viewed as the lowest abstraction, consisting of the most basic (raw) representation of facts, entities, or concepts, and contains no meaning (e.g., *"2001"* is considered as a number consisting of 4 digits). For the data to be *informative*, it must be interpreted and given a well-defined meaning (such as *"the year of announcement of the Semantic Web"*) and can be therefore qualified as *information* [11]. In this context, *metadata* is viewed as a description about the data and information (such as who gave the data/information – e.g., *Wikipedia*, when was the data/information given – e.g., *published in 2002*, etc.) [11]. At a higher level of abstraction, *knowledge* is viewed as the combination of all known data, information, and metadata concerning a given concept or fact, as well as the semantic links between them [12,13] (like knowing that *"the year of announcement of the Semantic Web"* is *"2001"*, following *Wikipedia* in an article *published in 2002*).

A social-based *event* can be viewed as a special form of knowledge defined following the 5*W*1*H* model [4,14,15]: *When*, *Where*, *What*, *Who*, *Why* and *How* aspects, describing an occurrence of a social or natural phenomenon (*what*, e.g., soccer match, car accident, heavy storm, or presidential debate) of interest to a group of people on the Web (*who*) happening within a certain time (*when*) and location (*where*, e.g., stadium, road, city, or amphitheater), having a certain description (*why*) and identification/traceability (*how*) from the set of social media objects describing it [5,16]. In this context, event-based CK is viewed as a development of knowledge assets or (semantic) information resources from a distributed pool of contributions, produced by human users or software agents, representing consensus on the descriptions and relationships between the events forming the CK [17]. In other words, an event-based CK repository can be viewed as a collection of events, with their descriptions, relationships, and underlying social media data, portrayed following a common representation model that can be used for automated reasoning by software agents [2,18].

Yet, extracting and organizing event-based knowledge from social media data comes with many challenges which we illustrate below using a real world motivational scenario.

Please cite this article as: M.A. Abebe, J. Tekli, F. Getahun et al., Generic metadata representation framework for social-based event detection, description, and linkage, Knowledge-Based Systems (2019), https://doi.org/10.1016/j.knosys.2019.06.025.

2.2. Motivation scenario

Climate change due to global warming increases the probability of some types of unusual weather. One effect of global warming is the occurrence of heavy rainfall. Excessive rain during short periods of time can cause flash floods. A flood may cause disruptions of basic utility services such as transportation, electricity, water, and telecommunication. When such an event occurs in a city, residents often capture different kinds of multimedia data, annotate, publish, and share them on social media sites like Facebook, Flickr, Twitter, or YouTube (cf. Fig. 1). They might also post comments on social media to share their appreciation and/or criticism regarding the level of preparedness and action that should have been taken by the city administration to handle the observed phenomena. Moreover, local media providers may continually publish news feeds related to the event.

In order to provide better services to residents, the city administration would largely benefit from organizing and processing the CK associated with occurring events. As a result, the city administration would be able to make more adequate decisions and take reactive/precautionary measures accordingly. However, user contributed social media contents and metadata on the Web often consist of objects of different types (images, animations, videos, etc.), with different metadata formats (XML, JSON, txt, etc.), coming from different sources (Flickr, YouTube, Twitter, etc.), annotated by different users with different backgrounds (e.g., novice, experts, scientists, etc.) who can sometimes produce inaccurate information or omit relevant information (missing certain event descriptive features following the 5*W*1*H* model), all of which would affect CK organization.

Consider the sample social media objects in Fig. 1, obtained from three different social media sources (Flickr, YouTube, and Twitter), along with their metadata descriptions. Flickr and YouTube use the eXtensible Markup Language (XML) to disclose user contributed contents and metadata, whereas Twitter uses JavaScript Object Notation (JSON). They not only have different data representation models, but also use different tag labels and formats to represent semantically similar (or identical) contents. For instance, Flickr and YouTube use different XML data element names, attribute names, and document structures to represent the date of creation/uploading of social media objects.[1] Moreover, the date, time, and location metadata can be represented in a different format specific to each social media service.[2] Similarly, the location information associated with a social media object might also be represented in different formats.[3] In addition, information published by different social media services can vary in content and structure.[4] Most importantly, different users might publish identical objects (on the same or different social sites) with very different annotations, using free text descriptions or tags which might be syntactically different, yet semantically related, following their own style of writing, vocabulary, and experience in annotation (e.g., an expert meteorologist would describe a *storm* or a *heat wave* differently from a non-expert observer).

2.3. Main requirements

In this context, handling diverse, heterogeneous, and sometimes incomplete social metadata to identify and describe meaningful events highlights various requirements that need to be fulfilled:

1. Converting the source metadata into a uniform data model that is generic enough to model social media objects following a high-level representation[5] suitable/adapted for the purpose of event detection and description,
2. Computing/evaluating the similarity/relatedness between social media objects given their adapted high-level representation, to group related objects together and identify corresponding events (e.g., recognizing and aggregating similar *flood* images published with related metadata, might help identify a *flood* event),
3. Accounting for the relative importance or weight of different event discriminating features (i.e., deciding which dimension of the 5*W*1*H* model is more important) in the event detection process, and adapting them following the user's needs (e.g., the user might be interested in identifying events considering their geographic proximity (*where*), regardless of their temporal (*when*) or semantic (*what*) descriptions),
4. Handling the semantic meaning of the textual descriptions of event discriminating features (e.g., how to understand the semantic relatedness and differences between terms *hailstorm*, *rainstorm*, and *blizzard*, which could be used by different users in describing the same or similar events) remains a central need in performing event detection from social media data,
5. Last but not least, identifying the different *relationships* that can occur between events (e.g., an event occurring *before* or *after* another, *far from* or *near to* another), considering the different available event descriptive features (e.g., temporal (*when*), spatial (*where*), semantic (*what*)), is also required as a building block for event-based CK organization.

[1] With Flickr: <exif tag="DateUploaded" label="Date Uploaded"><raw>2014:07:07</raw></exif><exif tag="TimeUploaded" label="Time Uploaded"> <raw>9:12:10+3:00</raw></exif>, with YouTube:<upload_time>14:48:04 </upload_time>, and with Twitter: "timestamp":1316656366000.

[2] YouTube represents upload date in the form of a complete date along with hours, minutes, and seconds (i.e., YYYY-MM-DD hh:mm:ss) whereas Flickr represents upload date and upload time in the form of a long date (i.e., YYYY:MM:DD) and a separate time representation (in hours, minutes and seconds) following the Coordinated Universal Time (UTC) referential (i.e., hh:mm:ss + UTC). Twitter represents the date/time of a social media object following Unix time, also known as POSIX time or Epoch time stamp, i.e., a single signed integer number that represents the number of seconds elapsed since midnight (00:00:00 UTC) of January 1, 1970 (e.g., 1316656366000 represents the ISO 8601 date format of 2016-7-13 5:6:33 GMT).

[3] YouTube represents geographic coordinates following the degrees, minutes, and seconds format (i.e., <locationlatitude:'' 9° 0' 19.4436'' N'' longitude:'' 38° 45' 48.9996'' E''/>). Yet, Twitter represents location following the decimal degrees format (i.e., "location":{"lng'':38.763611,"lat":9.005401}), whereas Flickr uses a predefined element (*raw*) and predefined attributes (*tag* and *label*) to represent the location information (i.e., <exif tag="City" label="City"> <raw> Addis Ababa </raw></exif><exif tag="Country-PrimaryLocationName" label="Country-Primary Location Name"><raw>Ethiopia</raw></exif>).

[4] For example, YouTube only provides uploaded time stamp, whereas Flickr captures both created and uploaded time stamps. Also, YouTube represents user contributed textual content with different XML elements (i.e., <tags><tag> a heavy rain </tag><tag> thunder shower </tag><tag> downpour </tag><tag> rainfall </tag></tags>, and <description> This is a flood caused by an intense rain for less than an hour. It also created pockets of small businesses. . . </description>). Yet, Twitter represents user contributed textual content as keywords in JSON (i.e., "keywords":[''AddisAbaba", ''Ethiopia'', ''Flood", ''Inundation'', ''Traffic Chaos"]), whereas Flickr uses a predefined element (*raw*) and predefined attributes (*tag* and *label*) to represent user contributed textual content (i.e., <exif tag=''Keywords'' label=''Keywords''><raw> Torrential Rainfall </raw> <raw> Cloudburst </raw><raw> rainstorm </raw></exif>, and <exif tag=''Caption-Abstract'' label=''Caption-Abstract''><raw> It was on July 7, 2014, at around 3:00pm just in the middle of the Meskel Square. . . </raw></exif>).

[5] In contrast with the low level features (such as color histogram) of multimedia objects, the high level features can be user contributed contents and metadata such as title, description, tags, comments, time stamps and location data.

Please cite this article as: M.A. Abebe, J. Tekli, F. Getahun et al., Generic metadata representation framework for social-based event detection, description, and linkage, Knowledge-Based Systems (2019), https://doi.org/10.1016/j.knosys.2019.06.025.

```
<photoid="14646512184"  camera="">
 <exif tag="Keywords" label="Keywords">
 <raw>Torrential Rainfall</raw><raw>Cloudburst</raw>
 <raw>rainstorm</raw></exif>
 <exif tag="DateCreated" label="Date Created">
 <raw>2014:07:07</raw></exif>
 <exif tag="TimeCreated" label="Time Created">
   <raw>8:56:29+3:00</raw></exif>
 <exif tag="DateUploaded" label="Date Uploaded">
   <raw>2014:07:07</raw></exif>
 <exif tag="TimeUploaded" label="Time Uploaded">
   <raw>9:12:10+3:00</raw></exif>
 <exif tag="By-line" label="By-line">
   <raw>Martin Meissner</raw></exif>
 <exif tag="City" label="City">
   <raw>Addis Ababa</raw></exif>
 <exif tag="Country-PrimaryLocationName"
 label="Country-Primary Location Name">
   <raw>Ethiopia</raw></exif>
 <exif tag="Caption-Abstract"
      label="Caption- Abstract">
   <raw>July 7, 2014, at around 3:00pm in the  middle
   of the Meskel Square Following the short but
   torrential rain, the lane that stretches from Bole
   International Airport to Meskel Square, was
   covered by an about 200-metre long stream</raw>
 </exif>
</photo>
```

a. Photo post and its metadata in XML format extracted from Flickr.

```
<video>
 <video_id>544007664 </video_id>
 <yt_id>sJBbE4svRaQ</yt_id>
 <title> Creating of Small business in Addis
Ababa</title>
<description> This is a flood caused by a rainstorm for
less than an hour. Heavy rain also created pockets of
small businesses. This problem created a business
instantly for street boys… </description>
 <username>I'm a Melagoodo.</username>
 <upload_time>2014-07-7 14:48:04</upload_time>
 <duration>62</duration>
 <viewcount>26217</viewcount>
 <locationlatitude :" 9° 0' 19.4436'' N"
         longitude:" 38° 45' 48.9996'' E" />
 <tags>
   <tag>a heavy rain</tag>
   <tag>thundershower</tag>
   <tag>downpour</tag>
   <tag>rainfall</tag>
</tags>
 <video_url>http://www.youtube.com/watch?v=sbE4svRaQ
   </video_url>
 <thumb_url>http://i1.ytimg.com/vi/16uz28/0.jpg
   </thumb_url>
</video>
```

b. Video post and its metadata in XML format extracted from YouTube.

```
{
 "text":"A terrible evening Addis Ababa (around Olompia)
  today, rainstorm cause flood in the new road",
 "cat":[1], "keywords":[“AddisAbaba”, “Ethiopia”,
 “Flood”, “Inundation”, “Traffic Chaos"],
 "location":{"lng“:38.763611,"lat":9.005401},
 "screen_name_lower":“Muradissa","geoflag":false,
 "type":1, "entities":
  {
    "urls":[{"expanded_url":"http://Twitter.com/download/
             android&quot;","url":""}],
    "hashtags":[{"text":“FloodinAddis"}],
    "user_mentions":[{"screen_name":""}]
  },
 "id":489333693576536066,
 "timestamp":1404711029000,
 "source":"web", "trans_text":"", "retweet_count":177,
 "user":
  {
   "location":“Addis Ababa", "verified":false,
    "screen_name":"Muradissa"
  },
  "tweet-lang":9
}
```

c. Tweet and its metadata in JSON format extracted from Twitter.

Fig. 1. Sample social media objects and their metadata obtained from three different social media sources.

The above requirements are partly overlooked by most existing event detection methods as shown in the following section.

3. Related works

Event detection methods from social media data can be categorized as unsupervised (clustering-based), supervised (classification-based), and hybrid approaches (combining clustering and classification processes). We briefly review these approaches in light of the main requirements identified in the previous section. Readers can refer to [19] for a detailed review on event mining.

3.1. Unsupervised approaches

Clustering or unsupervised classification is the process of organizing or grouping a collection of objects into groups (called clusters) based on their similarity values. Similarity is evaluated as the inverse of a distance function in a certain referential space [20,21]. Objects in the same group or cluster are more similar to (less distant from) each other than to those in other groups or clusters. Clustering has been used for various applications (cf. reviews in [22,23]) including event detection from social media data.

The authors in [10] propose an approach for detecting events from photos on Flickr by exploiting the tags supplied by users. The method consists of three steps: (1) identifying whether tags are related to events or not based on their temporal and spatial distributions; (2) detecting event-related tags to classify them into periodic or a-periodic event tags; and (3) retrieving the set of photos for each tag representing an event. A similar data-driven approach in described in [24] where images are first clustered based on their spatio-temporal information (*where* and *when*), where images which do not have spatial information are left out as singleton clusters. The generated and singleton clusters are then compared considering the images' creator (*who*), title, description, tags, and visual information (*what*), to merge similar clusters together. The proposed solutions in [10,24] use the Jaccard (syntactic) similarity measure to compare textual descriptions, and thus do not address their semantic meaning. Also, the methods do not highlight the impact of aggregating different feature similarity measures in the event detection process. Another data-driven approach is developed in [6], where the authors build on an original work from Microsoft Research [25] named PhotoTOC. Clustering is performed using a combination of time-stamps (*when*), spatial information (*where*), textual description labels (*what*), and the photo creator's information (*who*). A training dataset is used to estimate the relevance of each feature type as well as the merging threshold for the combined feature score. Yet, similarly to its predecessors, the solution in [6] does not consider the semantic meaning of the social media objects' textual descriptions, but rather evaluates their syntactic similarities. In [26], the authors attempt to identify social media events based on the assumption that an event happening at a certain place and time, will most probably be coined with a large number of photos and videos taken and shared in different social media sites. Yet, the proposed approach requires a certain number of initial seed photos (i.e., the product of the number of shared images and owners who are posting those images should not be less than a threshold value obtained empirically) to effectively detect events.

In contrast with most of the above studies, few solutions in [27–29] have (partly) considered the semantics of social media objects in the event detection process. The authors in [27] put forward a framework to semantically structure an object collection in social media applications. They use WordNet-based semantic similarity measures [30] where WordNet is utilized as a reference lexical knowledge base [31]. Primarily, only the spatial information (*where*) is used to cluster the object collection. Then the semantic similarities of the objects' descriptive tags (*what*) are utilized to merge the produced clusters. A similar approach

Please cite this article as: M.A. Abebe, J. Tekli, F. Getahun et al., Generic metadata representation framework for social-based event detection, description, and linkage, Knowledge-Based Systems (2019), https://doi.org/10.1016/j.knosys.2019.06.025.

is developed in [29], where initial clusters are identified based on creator (*who*) and temporal (*when*) information, and then the clusters are merged using location distance (*where*), as well as topic[6] and term syntactic similarity. In [28], the authors expand the images' textual descriptions by identifying the synonyms and hypernyms of every term, producing expanded bag-of-words representations which are then compared using the cosine syntactic similarity measure. Yet, the solutions in [27–29] do not evaluate the effect of using aggregated similarity measures (combining different features) on the event detection process. Also, none of the solutions mentioned above addresses the issue of identifying event relationships.

3.2. Supervised approaches

Various supervised or classification-based solutions have also been developed to perform event detection from social media data. We recall that classification or supervised learning is the process of organizing a collection of objects into pre-classified groups or labeled patterns based on their similarities with the training patterns [33]. Classification methods have been used for a variety of applications in data mining (cf. reviews in [33,34]) including event detection from Web and social media data.

In [35], the authors introduce a variety of text-based query building strategies designed to automatically augment user-contributed information for planned events with dynamically generated Twitter content. A planned event is described using time (*when*), location (*where*), and textual metadata (*what*, e.g., title, description, retrieved messaged). Queries include different combinations of features, such as location + title, title + description, location + time + title, etc. Term-frequency analysis is used, treating a predefined event's textual metadata and any retrieved tweets from the previous step as "ground truth" data describing the event. While the authors consider different combinations of features, nonetheless, they do not empirically evaluate their impact on the event detection/augmentation process. Also, the approach does not consider the semantic meaning of textual descriptions and only focuses on term-frequency analysis. The authors in [36] present a method that combines semantic inference and visual analysis for finding events. They present a large dataset composed of semantic descriptions of events, photos, and videos interlinked with the larger Linked Open Data (LOD) cloud. They use special tags (e.g., *lastfm:event* = XXX, *upcoming:event* = XXX) associated with their social media data, in order to detect events, an approach which is only applicable for planned (pre-defined) events posted (in advance) on event aggregating platforms (e.g., anticipated *soccer match*, or awaited *heat wave*, which are expected to occur on certain dates or in certain locations). Yet, the proposed solution does not identify instantaneous/unknown events[7] such as an unexpected *flood* or *car accident*. Also, the authors do not show the effect of aggregating different similarity measures to compare different event descriptive features in the event detection process.

The authors in [37] use event aggregation platforms (such as Last.fm, EventBrite, LinkedIn and Facebook events) to generate planned events. In this work, only social media contents which have location (*where*) and time (*when*) information are considered for the purpose of detecting events. As mentioned before, we argue that time and geo-location information are not enough to effectively detect events, since: (i) some social media authoring tools lack location recording components, and (ii) the timestamp values of social media contents might be distorted or noisy due to the particular configurations of media capturing tools. Note that the work in [37] focuses on generating events based on predefined preferences stated in advance in existing event aggregation platforms. Moreover, the authors do not consider the semantic meaning of social media objects' textual descriptions, nor do they discuss the impact of an aggregated similarity measure combining different event descriptive features in the event detection process. Also, the issue of identifying event relationships is not addressed in the above mentioned solutions.

3.3. Hybrid approaches

Few hybrid solutions, combining supervised and unsupervised techniques to perform social event detection, have been proposed. In [9], the authors utilize ensemble and classification-based similarity learning techniques to detect events. Both ensemble and classification-based similarity learning techniques are used in conjunction with an incremental clustering algorithm to generate a clustering solution. Yet, the authors do not discuss the effects and impact of spatial and semantic features of the shared social media objects in the event detection task. In [38], the authors propose a fusion-based method to detect and identify events. They use Factorization Machines (FMs)[8] to learn the similarity between pairs of social images, considering their creation time (*when*), location (*where*), associated tags and textual descriptions (*what*), as well as authorship (*who*). The latter are then run through an incremental clustering process to identify groups of related images where every group designates an event. This work considers image features holistically, and does not consider the effect of individual features in the event detection process. Moreover, it processes image textual descriptions syntactically and does consider their semantic meaning. The authors in [39] use the Chinese Restaurant Process to cluster a collection of photos and videos from social media applications. They assume that objects arrive sequentially in a streamed fashion, where every new object is compared with the already existing objects based on a probability model constructed from the training data set. Then, a single pass incremental clustering algorithm is used to merge the object with the clusters (events) which already exist, or to create a new cluster (event) around it. Yet, the authors in [39] do not evaluate the impact of aggregating temporal and spatial feature similarity in the event detection process. Moreover, they do not consider textual descriptions or semantic meaning, and only focus on temporal and spatial features.

In [8], the authors introduce a constrained clustering method, adapted from the spherical k-Means algorithm [40], to detect events from a social media object collection. The number of

[6] The authors extract so-called implicit semantic concepts, i.e., latent semantic concepts, inferred from the statistical and algebraic analysis of image textual descriptions (the authors in [29] utilize Latent Dirichlet Allocation), following the basic idea that images that share many textual terms in common are semantically closer others. Implicit concepts are represented numerically as hyper-dimensions in a latent semantic hyperspace, and do not align with any human-interpretable concept [32].

[7] *Unknown* events are events which are unexpected and are not being monitored by users (e.g., an unexpected accident, or an unexpected rainstorm). In contrast, a *known* event is one that is expected or that is being monitored by users (e.g., an expected soccer match or a pre-scheduled music concert). On one hand, detecting unknown events usually requires the use of unsupervised learning techniques (such as the one utilized in our study) where the system identifies events without any previous knowledge about their existence or nature. On the other hand, supervised learning methods are usually utilized to detect *known* events, where users provide the system with some description about the nature of their target events as input (e.g., monitoring *thunderstorms* or *car accidents*, where each event would be described with some metadata), and then the system identifies the corresponding events accordingly (e.g., identifying all occurring thunderstorms, or all occurring car accidents), by matching the incoming social media objects' descriptions with those of the pre-defined events.

[8] A Factorization Machine (FM) is a classification model that combines Support Vector Machine (SVM) functionality with matrix factorization models [38].

Please cite this article as: M.A. Abebe, J. Tekli, F. Getahun et al., Generic metadata representation framework for social-based event detection, description, and linkage, Knowledge-Based Systems (2019), https://doi.org/10.1016/j.knosys.2019.06.025.

initial clusters k is set in the training phase. Cosine similarity is used to measure the distance between an object and the cluster centroids based on the temporal, spatial, and textual features combined into an aggregate linear similarity measure. Yet, the approach does not consider the semantic aspect of textual features and rather computes syntactic similarity using TF–IDF[9] term weights. A similar solution is described in [41], where the authors introduce a user-centric data structure, named UT-image (user-time image), to store a social image collection's metadata. The whole metadata set is turned into a UT-image, so that each row of an image contains all records that belong to one user. Then after, cluster merging is performed considering temporal (*when*), spatial (*where*), or textual (tag/title/description, i.e., *what*) feature similarity thresholds set by the user. The proposed method does not consider the effect of an aggregated similarity measure combining different features together in cluster merging. In addition, the authors themselves state that using the Jaccard (syntactic similarity) measure to compare the textual features fails to address the challenge of capturing the semantics of collaborative tags.

3.4. Event-based knowledge organization

Recently, there has been an increasing interest in automatic knowledge graph construction, where most effort has been dedicated toward the development of statistical models to infer facts about entities in a graph [42]. Some projects have been developed to extract knowledge from semi-structured resources such as Wikipedia (cf. DBpedia [43], Freebase [44] or Google Knowledge Vault [45]), but the extracted information is centered on collecting facts around entities rather than events. Some works have targeted news articles [46], extracting information like persons, organizations, and locations, resulting in a grouping of news stories by topics and entities. Another approach in [47] organizes news articles around stories, which imply events, by computing word and phrase co-occurrence in a sequence of news articles, producing a chain of news articles that form a story. The authors in [48] introduce an approach for automatically extracting named events from news article, while [49,50] discuss the use of so-called semantic roles (i.e., *who*, *what*, *where* of an article) to extract related events using hybrid event extraction approaches. In [51], the authors model the time, dependency, and reference relationships between so-called component events (i.e., episodes) in order to find and understand the "whole picture" of the bigger event. They specifically target the problem of temporal event search and introduce a framework for temporal event relationship analysis, studying the dependency between component events in the evolution of the bigger event that is targeted by the user query. In a subsequent study in [52], the authors introduce a solution to identify the first story of a previously unknown event, combining temporal information, named entity recognition, and topic modeling to associate multiple events with news stories, taking into account the events' evolution over time. Recent methods in [53,54] extract information from multi-lingual news articles, and convert them to a common representation. The authors in [53] use unsupervised clustering to identify related articles in every language separately, using latent semantic indexing for article similarity evaluation. Then, a supervised (Support Vector Machine) classifier is used to merge clusters from different languages describing the same event, based on manual expert training. In [54], the authors use deep natural language processing techniques to extract the different entities in every news article and the events within it. Entities and events are then represented as RDF triples (e.g., sentence "Volkswagen acquires Porsche in 2009" is represented as a set of triples: <event1, *hasActor*, Porsche>, <Porsche, *AquiredBy*, Volkswagen>, <event1, *hasTime*, 2009>) forming an event-centric knowledge graph.

The knowledge graph described in [54] and the time dependency relationships in [51] are seemingly the closest to the notions of event-based collective knowledge and event relationships described in our study, yet with different objectives and coverages. While the authors in [54] target event extraction from text-rich news articles and linguistic-based *entity–event* relationships, as well as time dependencies between *component* events and their description of the (bigger picture) *target* event [51], our present study targets text-poor social medial objects (where the text consists of tags and short comments), and the extraction and representation of temporal, spatial, as well as semantic *entity–entity* relationships, with their directional, metric, and topological variants, which are not addressed in – and would be complementary to – the latter studies.

3.5. Discussion

To summarize, most existing event detection methods in the literature either: (i) are domain dependent and consider certain specific kinds of information (e.g., tweets only, Flickr photos only), e.g., [7,9,10], (ii) generate events based on predefined clues and are not able to identify unknown events (except for unsupervised methods), e.g., [7,35,37], (iii) consider event descriptive features (e.g., time, space, text) separately and do not combine or evaluate their impact on the event detection process (one approach in [38] combines all features holistically, yet without allowing the user to adapt or evaluate the impact of every feature separately), or (iv) do not (or only partly) consider the semantic meaning associated with social media data and focus on syntactic textual descriptions (they use syntactic similarity measures such as Jaccard or cosine, coined with term-frequency weighting, thus only capturing the surface level similarity of textual descriptors, and missing their semantic relatedness, e.g., [6–8,41]). Most importantly, most existing methods to our knowledge (v) do not address the issue of identifying meaningful relationships between events (one approach in [54] identifies linguistic-based entity-event relationships from news articles, which would be complementary to this study), and which we consider as a central requirement toward event-based CK organization.

4. Proposed framework

To address the requirements and limitations identified in the previous sections, we introduce SEDDaL, as an unsupervised and semantic-aware framework for Social Event Detection, Description and Linkage. SEDDaL's overall architecture is depicted in Fig. 2. It consists of four main modules: (i) Metadata Representation Space Model (MRSM) which allows representing the source metadata in a uniform and generic data model to describe social media objects and events (addressing requirement #1 in Section 2.3), (ii) Similarity evaluation module, allowing to compute the similarity/relatedness between social media objects given their uniform representation in MRSM, while considering the relative importance of different features (temporal, spatial, and semantic) in the similarity evaluation process, and adapting the features' weights following the user's needs (answering requirements #2 and #3), (iii) Event detection module built upon MSRM, allowing to group similar/related objects together considering the semantic meaning of their textual descriptions (addressing

[9] Term Frequency–Inverse Document Frequency.

Please cite this article as: M.A. Abebe, J. Tekli, F. Getahun et al., Generic metadata representation framework for social-based event detection, description, and linkage, Knowledge-Based Systems (2019), https://doi.org/10.1016/j.knosys.2019.06.025.

requirement #4), in order to identify corresponding events,[10] and (iv) Event relationships identification module, allowing to identify the different relationships that can occur between events (directional, metric, and topological), considering the different features (temporal, spatial, and semantic) of interest to the user (addressing requirement #5). We describe each of the latter modules in the following sub-sections.

4.1. Metadata representation space model

Event definitions are theoretically described using the 5*W*1*H* model: *When*, *Where*, *What*, *Who*, *Why* and *How* aspects [4,14,15]. Yet, as described in Section 3, only few of these features are practically covered in existing methods, mainly: *When* (time) and *Where* (location) [7,37,55]. In our work, we consider an additional feature: the *What* (meaning) of the event (the remaining *Who*, *Why*, and *How* facets will be covered in a subsequent study). To do so, we define MSRM as a hyperspace consisting of three composite dimensions: temporal, spatial, and semantic, describing every social media object (as shown in Fig. 3a). Consequently, an event can be represented in the same space, consisting of the collection of objects describing it (cf. Fig. 3b). In this subsection, we formally describe each dimension, its coverage, and related properties.

4.1.1. Temporal dimension

One of the three main features used to describe social media objects following MRSM is their temporal coverage. Here, temporal coverage consists of a set of timestamps, where each timestamp is an instance or single occasion related to the object (or event), as shown in Fig. 4a. In the following, we formally define the notions of temporal dimension, temporal stamp, temporal coverage, and temporal coverage representative points.

Definition 1 (*Temporal Dimension ($\mathbb{T}$)*)**.** The temporal dimension $\mathbb{T}$ is defined as a finite sequence of discrete and ordered primitive temporal units used to represent and interpret a social media object's temporal feature values, formally:

$$\mathbb{T} = \{t_0, t_1, t_2, \ldots\} \tag{1}$$

where t_i is the i*th* temporal unit, and t_0 the initial temporal value •

The unit of measurement of the temporal dimension can be chosen by the user (or the system admin) based on the kinds of events to be detected. For instance, detecting a soccer player's maneuvers in a soccer match would require a small time unit (like seconds) whereas detecting thunderstorms and weather-related events can be handled using bigger time units (like hours or days). In our study, we consider the International System (IS)'s second unit (s) as the default time unit, such that the dimension's origin (t_0) is the UNIX time (a.k.a. POSIX or Epoch time, describing instants in time since 00:00:00 UTC, January 1, 1970).

Definition 2 (*Temporal Stamp (t)*)**.** It designates a single discrete value of the temporal dimension $\mathbb{T}$ •

[10] Note that our framework allows describing social media objects and events in the same generic representation space, where the dimensions' properties (stamps, coverages, and coverage representative points) and associated similarity measures are formally defined. The latter allow for a "straightforward" usage of our solution with typical similarity-based machine learning solutions, both supervised and unsupervised: since any such solution would require: (i) a clear description of the features of the data objects (which we provide), as well as (ii) proper methods to compare and evaluate the similarity between objects (which we also provide).

While a still image or a photo object can be described by a single temporal stamp, yet a video object consists of a set of framesets and thus requires a range of temporal tamps to designate its temporal coverage. This is formally stated in Definition 3:

Definition 3 (*Temporal Coverage (T)*)**.** It is an ordered collection of temporal stamps enclosed within a start and an end stamp, describing the temporal coverage of a social media object or event. It is used to represent the duration or capture of an object (e.g., a video), or the duration of an event (e.g., duration of a storm). Formally:

$$T = \{t_i \in \mathbb{T} \mid t_i \geq t_s \wedge t_i \leq t_e\} \tag{2}$$

where t_s is the start temporal stamp of T, and t_e its end temporal stamp •

Note that most current multimedia object authoring tools such as smart-phones or video cameras, as well as most social media services do not capture the temporal stamp of each object's frameset. However, they capture either the start temporal stamp of the object, in the form of a *created time*/*date* attribute and its *duration* (e.g., Facebook Live and Snapchat), or they capture the end temporal stamp of the object in the form of an *upload time*/*date* and its *duration* (e.g., YouTube). When such two conditions occur, the missing value is computed by considering the upload time as an end temporal stamp, such that the start temporal stamp is obtained by subtracting the video duration from end temporal stamp (e.g., in the case of videos uploaded on a social media service once captured). Similarly with social media services providing live streaming, the created time can be considered as a start temporal stamp, and the end temporal stamp can be then computed by adding the object's duration to the start temporal stamp.

Definition 4 (*Temporal Coverage Representative Point (t_c)*)**.** It is the middle time stamp of a temporal coverage T, representing the temporal coverage's center of gravity. Formally:

$$t_c(T) = \frac{t_s + t_e}{2} \tag{3}$$

where t_s is the start temporal stamp of T, and t_e its end temporal stamp •

Temporal coverage representative points are introduced to simplify computations when comparing the temporal coverage of social media objects or events: instead of comparing the whole coverages, we compare their representative points (Section 4.2).

4.1.2. Spatial dimension

In this subsection, we describe the spatial dimension of our MRSM model, as well as the related notions of spatial stamp, spatial coverage, and spatial coverage representative points required for comparing objects (and events later on):

Definition 5 (*Spatial Dimension ($\mathbb{L}$)*)**.** The spatial dimension $\mathbb{L}$ is defined as a composite dimension consisting of three components (sub-dimensions) representing geographical position following Earth's geo-referential system, formally:

$$\mathbb{L} = \langle \emptyset, \lambda, h \rangle \tag{4}$$

where $\emptyset$ represents the latitude, λ the longitude, and h the altitude sub-dimensions (cf. Fig. 4b) •

Similarly to the temporal dimension, the unit of measurement for the spatial (sub) dimension(s) can be chosen by the user (or system admin) based on the kinds of events to be detected. For

Please cite this article as: M.A. Abebe, J. Tekli, F. Getahun et al., Generic metadata representation framework for social-based event detection, description, and linkage, Knowledge-Based Systems (2019), https://doi.org/10.1016/j.knosys.2019.06.025.

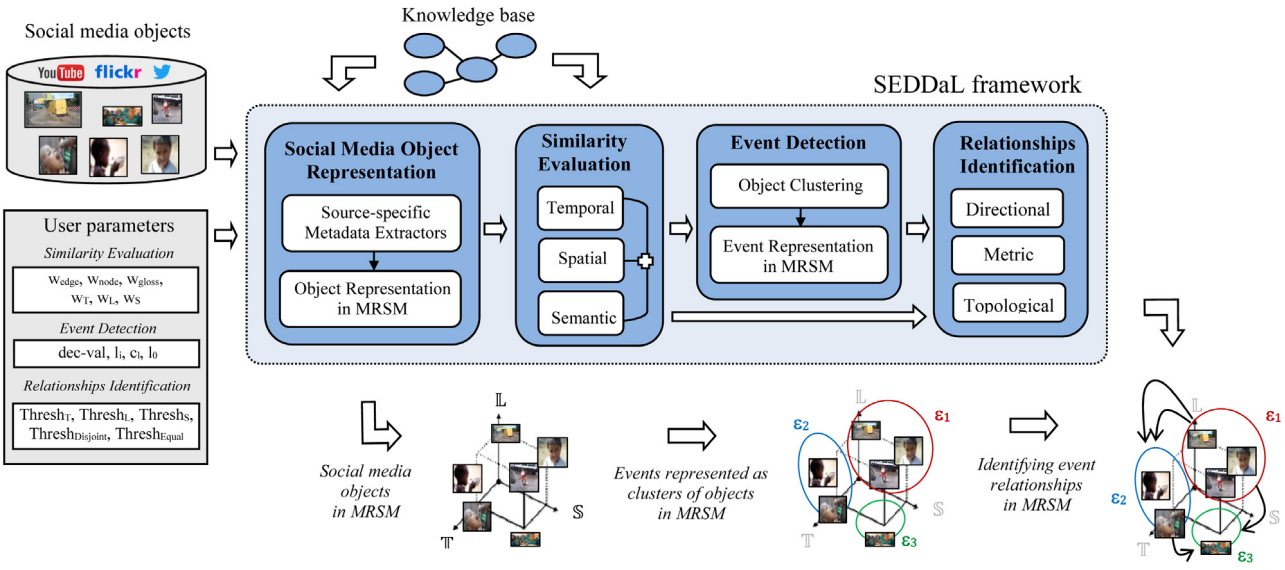


Fig. 2. Overall architecture of our SEDDaL framework.

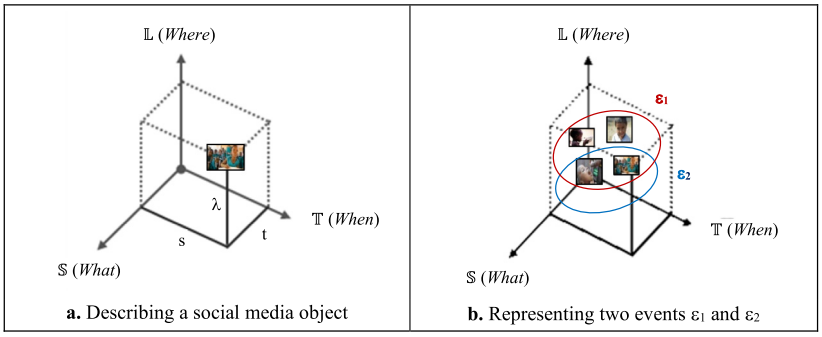


Fig. 3. Metadata Representation Space Model (MRSM).

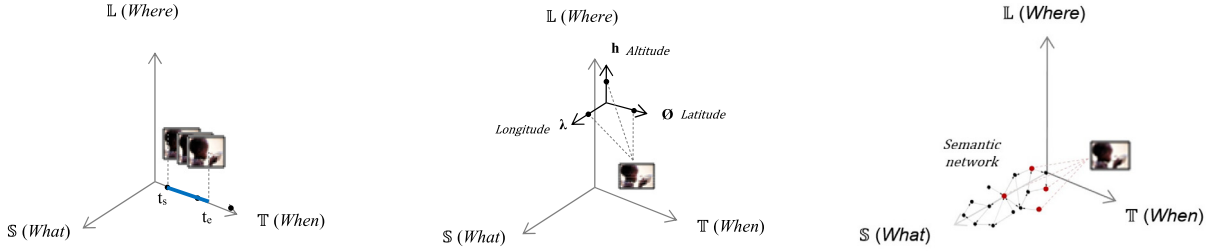


Fig. 4. Temporal, spatial, and semantic dimensions in MRSM.

instance, detecting a soccer player's maneuvers in a soccer match would require a small spatial unit (like meter or foot), whereas detecting thunderstorm or weather-related events would require bigger spatial scales (such as kilometers or miles). In our study, we adopt IS's meter unit (m) as the default unit of spatial measure. It can be converted to the DMS scale (Degrees, Minutes, and Seconds) or Radians with the latitude (Ø) and longitude (λ) sub-dimensions, based on user preferences. We adopt as point of origin for the spatial dimension the geographic center of the surface of the Earth (i.e., the intersection of the Equator and Prime Meridian (0, 0), or Greenwich meridian), even though the point of origin can also be modified/chosen by the user (system admin).

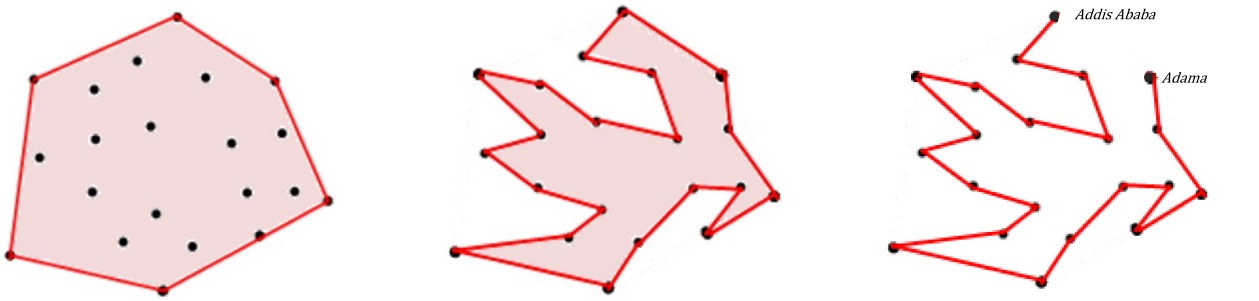
Definition 6 (*Spatial Stamp* (ℓ))**.** It is a discrete and instantaneous value of the spatial dimension $\mathbb{L}$, consisting of a triplet:

$$\ell = \langle Ø, \lambda, h \rangle \quad (5)$$

where Ø, λ, and h designate individual coordinate values defined with respect to (w.r.t.) each of the latitude ($Ø \in Ø$), longitude ($\lambda \in \lambda$), and altitude ($h \in$ h) sub-dimensions of $\mathbb{L}$ •

Definition 7 (*Spatial Coverage (L)*)**.** It is the set of spatial stamps designating the surface coverage in which a social media object is created (e.g., area in which a video stream is recorded) or in which an event occurs (e.g., area affected by a storm). Formally, given the composite spatial dimension $\mathbb{L}$, we define L as:

$L = \{\ell_i \in \mathbb{L} | \ell_i = \langle Ø_i, \lambda_i, h_i \rangle$ is a spatial stamp recorded by



Identified based on the Cartesian coordinates of the spatial stamps[11] using a typical *convex hull* algorithm.

Identified based on the Cartesian coordinates of the spatial stamps using Gauss's area formula:

$$\frac{1}{2}\left|\sum_{i=1}^{n-1} x_i \times y_{i+1} + x_n \times y_1 - \sum_{i=1}^{n-1} x_{i+1} \times y_i - x_1 \times y_n\right|$$

Identified based on spatial stamps' coordinates using the sum of their pair-wise Euclidian distances:

$$\sum_{l_i, l_j \in L} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$$

a. Minimum boundary coverage **b.** Actual surface coverage **c.** Path (trail) coverage

Fig. 5. Different coverages that can be identified from a set of spatial stamps (disregarding the altitude dimension (h) to simplify) (see Refs. [56,57]). [11]

the social media object authoring tool} (6)

where ($\emptyset_i \in \emptyset$) is the latitude, ($\lambda_i \in \lambda$) the longitude, and ($h_i \in h$) the altitude coordinates of every ℓ_i in $\mathbb{L}$ •

For example, a video camera-man can start to capture a video shoot at *Addis Ababa* (8.9806° N, 38.7578° E, 2355 m) and then finish when arriving at *Adama* (8.5263° N, 39.2583° E, 1712 m). Following our model, each frameset of the video has its own spatial stamp, and thus the spatial coverage of the video object is designated by the set of recorded spatial stamps, from *Addis Ababa* till *Adama*. Yet, identifying the actual surface covered by the object's spatial stamps can vary based on the nature of the object created and the metadata provided by the object's authoring tool. For instance, identifying the minimum boundary coverage (Fig. 5a) can be useful in describing the coverage of a video object describing a storm event, whereas identifying the path coverage (Fig. 5c) can be more useful in describing the trail of a video shoot between *Addis Ababa* and *Adama*. Hence, evaluating the similarity/distance between social media objects' spatial coverages (i.e., $Sim(o_1, o_2)$ or $Dist_{\mathbb{L}}(o_1, o_2)$) is not trivial.

Hence, we introduce the notion of spatial coverage representative point to simplify similarity computations: instead of comparing the spatial coverages of two objects (or events), we compare their representative points. We define the representative point as the *geographic midpoint* of a set of spatial stamps, following a well-known procedure from Earth geometry [58]:

Definition 8 (*Spatial Coverage Representative Point* (ℓ_c))**.** It is the geographic midpoint of a spatial coverage $L = \{\ell_i \in \mathbb{L} | \ell_i = \langle \emptyset_i, \lambda_i, h_i \rangle\}$, representing the spatial coverage's center of gravity, formally:

$$\ell_c(L) = \langle \emptyset_c, \lambda_c, h_c \rangle \quad (7)$$

where: $\emptyset_c = atan2(Z, \sqrt[2]{X^2 + Y^2}) \times \frac{180}{\pi}$, $\lambda_c = atan2(Y, X) \times \frac{180}{\pi}$, $X = \frac{\sum_{i=1}^{n} x_i}{n}$, $Y = \frac{\sum_{i=1}^{n} y_i}{n}$, $Z = \frac{\sum_{i=1}^{n} z_i}{n}$, and $\langle x_i, y_i z_i \rangle$ represent the Cartesian coordinates of spatial stamp $\ell_i = \langle \emptyset_i, \lambda_i, h_i \rangle$[9] where $x_i = \cos(\emptyset_i) + \cos(\lambda_i)$, $y_i = \cos(\emptyset_i) + \sin(\lambda_i)$, $z_i = \sin(\emptyset_i)$, and $h_c = avg_{\ell_i \in L}(h_i)$ •

A simplified method to approximate the geographic midpoint is to calculate the mathematical average of the $\langle \emptyset, \lambda, h \rangle$ coordinates of the spatial stamps (without translation into Cartesian space). Yet, the latter would only produce accurate results with distances less than 400 km (250 miles) [58] (i.e., equivalent to finding the midpoint on a flat rectangular projection map).

4.1.3. Semantic dimension

While temporal (*When*) and spatial (*Where*) information have been considered with many existing event detection methods (cf. Section 3), yet the semantic (*What*) facet has been mostly disregarded. Hence, we include a semantic dimension in our MRSM as described hereunder.

Definition 9 (*Semantic Dimension* ($\mathbb{S}$))**.** It is a lexical knowledge based represented as a semantic network made of a set of *concepts* representing groups of words/expressions having identical semantic meanings, and a set of links connecting the concepts representing *semantic relations* (*hypernymy* (*isA*), *holonymy* (*partOf*), *relatedTo*, etc. [30,31]). We represent it as a labeled directed graph $\mathbb{S} = (N, E, R, f)$, where: N is the set of nodes designating concepts; E is the set of edges connecting the nodes, i.e., $E \subseteq C \times C$; R is the set of semantic relations; and f is a function designating the nature of edges in E, i.e., $f: E \rightarrow R$ •

For instance, typical lexical knowledge bases like WordNet [31] or Yago [59] define concepts as so-called synsets: sets of synonyms terms (e.g., *car*, *auto*, and *automobile*) having the same gloss description (e.g., *a motor vehicle with four wheels*), connected with various hierarchical relationships (e.g., *hypernymy*, *holonymy*, etc.) and cross relationships (e.g., *relatedTo*, *derivedFrom*, etc.). The unit of the semantic dimension can be a single concept, or a group of concepts, following the user (system admin)'s perception of semantic meaning. For instance, a user might not care to distinguish between concepts *sports car*, *sedan*, *SUV*, and *muscle car*, and might prefer to refer to all of them as the more general concept *vehicle*. Here, concept *vehicle* would subsume the group of aforementioned concepts, designated as one single semantic unit. In this study, and for the sake of simplicity, we consider each individual concept to be a single semantic unit.[12] The origin of the semantic dimension can be defined as the *root* node of the corresponding semantic network. If the reference semantic network contains multiple root nodes

[11] The Cartesian coordinates (x_i, y_i, z_i) of a spatial stamp ℓ_i can be obtained from its latitude and longitude coordinates ($\emptyset_i$, λ_i), based on the *xy* plane lying within the equatorial plane, with its origin at the center of the earth. Looking down onto the North Pole, the positive *x*-axis passes through the Greenwich meridian (0° E), the positive *y*-axis passes through the 90° E meridian, and the positive *z*-axis extends from the center of the earth through the North Pole [58].

[12] Varying semantic units as groups of concepts to modify the semantic dimension's granularity will be considered in a future study.

(such as in WordNet which has more than 11 root concepts), then we create an artificial root which subsumes all of them.

Definition 10 (*Semantic Stamp (s)*)**.** It is an instance or a single concept of the semantic dimension $\mathbb{S}$ •

Definition 11 (*Semantic Coverage (S)*)**.** It is a set of concepts (semantic stamps), along with their semantic relationships, highlighting the semantic description of a social media object or an event. It can be defined as a sub-graph of the semantic dimension $\mathbb{S}$, noted $S = (N, E)$, where $N \subseteq N$ (set of concepts, i.e., nodes) and $E \subseteq E$ (semantic relations, i.e., edges) •

While various methods for comparing pairs of concepts in a lexical knowledge base have been proposed in the literature, e.g., [30,60], nonetheless, capturing the semantic relatedness between two groups of concepts or concept sub-graphs (e.g., two semantic coverages) has attracted less attention. Two complementary approaches have tackled the issue in [61,62], developed in the context of concept similarity of ontology management systems [62], and concept similarity in geographic information systems [61]. Yet the solutions in [61,62] are computationally expensive and require $O(N!)$ time where N is the number of concepts being compared. Other studies have addressed similar problems in the context of XML sub-tree semantic analysis and disambiguation (comparing groups of XML node labels), e.g., [63,64], schema mapping (matching schema element/attribute definitions) e.g., [65,66], and ontology mapping (matching concept sub-graphs), e.g., [67,68], yet require at least polynomial $O(N^2)$ time. Hence to simplify mathematical computations, we introduce the notion of semantic coverage representative point:

Definition 12 (*Semantic Coverage Representative Point (s_c)*)**.** It is a single concept representing the middle semantic stamp of a semantic coverage S, which we define as the semantic concept that is, on average, most similar to all other concepts in S:

$$s_c = s \in S \,/\, \forall s_i \in S, \ \underset{\text{All } s_j \in S}{\text{Avg}}\,(Sim_S(s, s_i)) \geq \underset{\text{All } s_j \in S}{\text{Avg}}\,(Sim_S(s_j, s_i)) \quad (8)$$

where $\text{Sim}_S(s_i, s_j)$ represents the semantic similarity between (concepts) s_i and s_j (developed in Section 4.2) •

In other words, every concept is compared with all other concepts in semantic coverage S. Consequently, the representative point is identified as the concept having the maximum average similarity w.r.t. all other concepts in S. Fig. 6 provides sample semantic coverages with their coverage representative points.

As a result, instead of comparing the semantic coverages (i.e., the groups of concepts) of two social media objects (or events), we can efficiently compare their representative points.[13]

4.1.4. Data model

After defining MRSM's dimensions, we define its data model for describing a social media object and an event.

Definition 13 (*Social Media Object (o)*)**.** A social media object (e.g., video, image, chart, tweet, or Wiki article) is defined, following MRSM, as a quadruplet:

$$o = (oid, t_c, \ell_c, s_c) \quad (9)$$

having a unique object identifier, *oid*, and three representative points: temporal t_c, spatial ℓ_c, and semantic s_c, following MRSM •

[13] This naturally comes to the expense of reduced semantic expressiveness and thus reduced accuracy in the comparison process, as a consequence of reducing of the whole semantic coverage to one single representative point. The same happens when reducing the temporal and spatial coverages into individual representative points.

We can refer to the above as a *restricted representation* of a social media object in MRSM. Yet, MRSM can also allow an *extended representation* of a social media object using the object's temporal, spatial, and semantic coverages, T, L, and S:

$$o_{\text{Extended}} = (oid, T, L, S) \quad (10)$$

We adopt the *restricted representation* as the *default* representation of a social media object in our study in order to allow for efficient processing: handling the whole coverages of a large number of objects is significantly more computationally complex than handling their coverages' individual representative points (especially when dealing with the spatial and semantic dimensions, as highlighted in the previous sections).

Consequently, an event can be defined as an aggregation or a group of similar social media objects:

Definition 14 (*Event (ε)*)**.** An event ε is an occurrence of a social or natural phenomenon happening at a certain time and location, and can be identified/described by the set of social media objects O describing it, formally:

$$\varepsilon = (eid, T, L, S) \quad (11)$$

where *eid* is a key value used to uniquely identify an individual event ε, $T = \underset{for\ all\ o_i \in O}{\cup}(T_i)$, $L = \underset{for\ all\ o_i \in O}{\cup}(L_i)$ and $S = \underset{for\ all\ o_i \in O}{\cup}(S_i)$ designate respectively: the union of the set of social media objects' temporal coverage representations $U(T_i)$, spatial coverage representations $U(L_i)$, and semantic coverage representations $U(S_i)$, for all objects $o_i \in O$ belonging to event ε •

We can refer to the above as an *extended representation* of an event in MRSM. Yet, MRSM can also allow a *restricted representation* of an event by identifying the event's temporal, spatial, and semantic coverage representative points (similarly to object representative points):

$$\varepsilon_{\text{Restricted}} = (oid, t_c, \ell_c, s_c) \quad (12)$$

Nonetheless, we adopt the *extended representation* as the *default* representation of an event in our study for more expressiveness, and especially since event descriptions are produced after the social media objects have been processed, and thus do not impact the time complexity of our solution.

Consider for instance the 9 sample images shown in Fig. 6 described following MRSM. The events extracted based on these images are provided in Fig. 7. Note that social media objects' textual descriptions generally consist of concatenations of keywords or of short sentences (as shown in Fig. 1). Hence, several linguistic pre-processing steps are required to identify semantically meaningful words, including *stop word removal* (removing prepositions and semantically meaningless words such as: *the*, *a*, *of*, *to*, etc.), *tokenization* (parsing names into tokens based on punctuation and case, to form simple expressions, e.g., *Amb_Temp* → *Ambient Temperature*), and *stemming* (reducing inflected or derived words to their stem, i.e., base or root, e.g., *raining*, *rains* → *rain*) [70,71]. The root words are then matched with the MRSM semantic dimension's concepts (we use a semantic network representation of WordNet 3.0 [31] as the semantic dimension in our study) to identify the corresponding semantic concepts. Concept identification is straightforward when the word has one single meaning, and consists of identifying the concept (synset) that subsumes the word in its definition. In the case of polysemous words (i.e., words with multiple senses), *word sense disambiguation* is utilized to select the semantic concept that most likely describes the meaning of the word among its surrounding keywords or within its containing sentence, e.g. [32,72]. Linguistic pre-processing operations are executed offline to obtain the social media objects' semantic descriptions following MRSM, and do not affect system performance (cf. Section 4.5).

Please cite this article as: M.A. Abebe, J. Tekli, F. Getahun et al., Generic metadata representation framework for social-based event detection, description, and linkage, Knowledge-Based Systems (2019), https://doi.org/10.1016/j.knosys.2019.06.025.

a.

oid 14646512184	
Extended	Restricted (default)
T T_s=1404712589 T_e=1404712589	t_c 1404712589
L NULL	ℓ NULL
S { torrent, rainfall, cloudburst, rainstorm July, Meskel Square, torrent, rain, lane, Bole International Airport, stream, water}	s_c rainfall

b.

oid 544007664	
Extended	Restricted (default)
T t_s=1404744484 t_e=1404744561	t_c 1404744523
L lat=9.005278 long=38.763334 alt=NULL	ℓ_c lat=9.005278 long=38.763334 alt=NULL
S {thundershower, rainfall, rain, downpour, flood, rainstorm, rain, hour, pockets, small businesses, problem, street boys}	s_c rainstorm

c.

oid 14646512184	
E tended	Restricted (default)
T t_s=1404711029 t_e=1404711029	t_c 1404711029
L lat=9.005401 long=38.763611 alt=NULL	ℓ_c lat=9.005401 long=38.763611 alt=NULL
S { Ethiopia , Addis Ababa, flood, inundation, traffic, chaos, terrible, evening, Olompia, today, rainstorm, flood, road}	s_c flood

d.

oid 452155896	
Extended	Restricted (default)
T t_s=t_e=1238964834	t_c 1238964834
L lat=45.51 long=-73.55 alt=NULL	ℓ_c lat=45.51 long=-73.55 alt=NULL
S {Ian Mosley, Mark Kelly, Pete Trewavas, Steve Hogarth, Steve Rothery, concert, gig, live, weekend, *music*, progressive, marillion}	s_c concert

e.

oid 3443324510	
Extended	Restricted (default)
T t_s=t_e=128876657	t_c 128876657
L lat=45.51 long=-73.557 alt=NULL	ℓ_c latt=45.51 long=- 73.557 alt=NULL
S {Steve Ho arth, concert, gig, live, marillion , weekend, montreal, music, progressive}	s_c gig

f.

oid 3421558753	
Extended	Restricted (default)
T t_s=t_e=1238956813	t_c 1238956813
L NULL	ℓ_c NULL
S {marillion, *concert*, rock, weekend, montreal}	s_c concert

g.

oid 18796702	
Extended	Restricted (default)
T t_s=t_e=1145040959	t_c 1145040959
L lat=47.4357 long=-122.294 alt=NULL	ℓ_c lat=47.4357 long=-122.294 alt=NULL
S {Stardance, Norwescon Seattle, Double Tree Hotel, conference, cost me play, antasy}	s_c fantasy

h.

oid 128800481	
Extended	Restricted (default)
T t_s=t_e=1145041372	t_c 1145041372
L lat=47.4357 long=-122.294 alt=NULL	ℓ_c lat=47.4357 long=-122.294 alt=NULL
S {Double Tree Hotel, Nikkor, Seattle, conference, costume play, fantasy, science fiction}	s_c fantasy

i.

oid 129778685	
Extended	Restricted (default)
T t_s=t_e=1145105698	t_c 1145105698
L lat=47.4357 long=-122.294 alt=NULL	ℓ_c lat=47.4357 long=-122.294 alt=NULL
S {Norwescon, Nikkor, Washington, conference, convention, costume play}	s_c convention

Fig. 6. Images from our motivation example (cf. Fig. 1) as well as 6 sample images from the MediaEvalSED 2013 image dataset [69] described following MRSM. Note that image descriptions were obtained using dedicated metadata extractor methods specifically tailored to extract social media object descriptions from the concerned social media sites and MediaEvalSED into MRSM. We provide both the *extended* and *restricted* representations of objects, where the latter is utilized as the default representation in our approach.

4.2. Similarity measures and their metric properties in MRSM

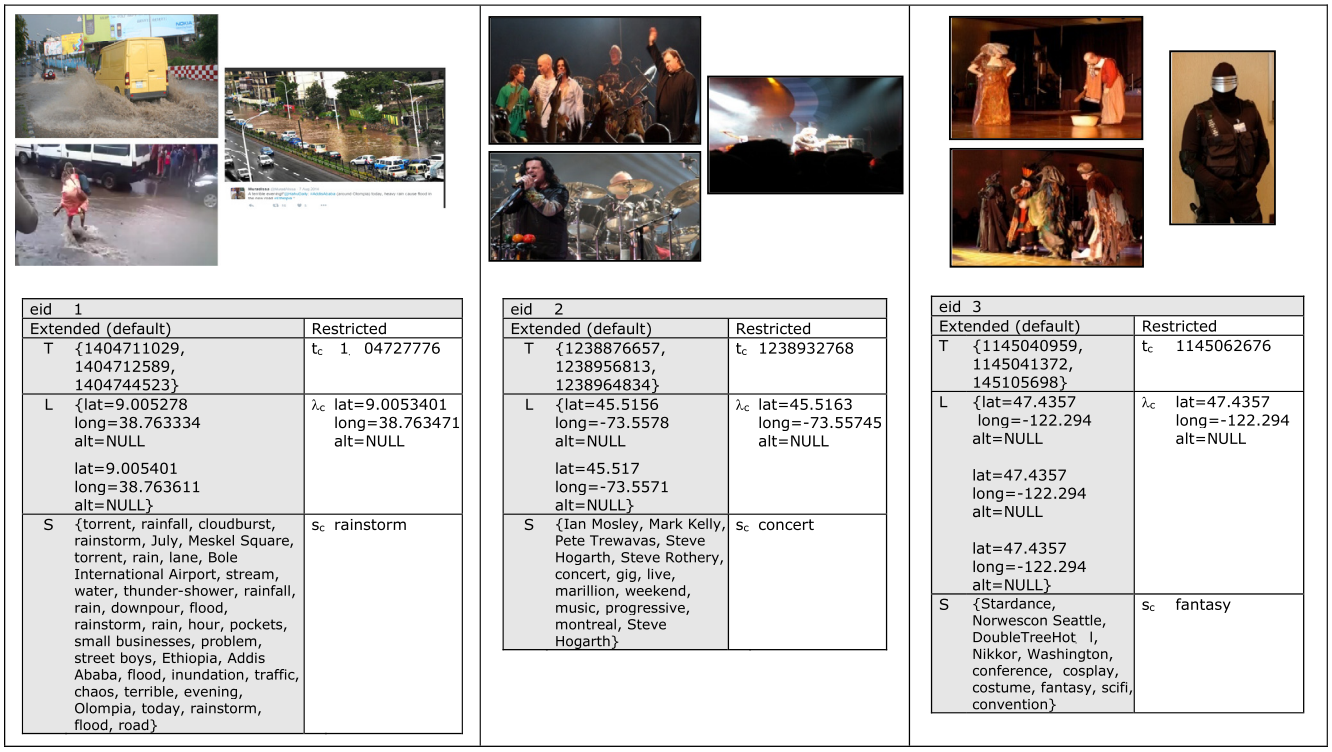
A key issue when defining a space model (such as MRSM) is to define distance (similarity) measures allowing to compare and order entities (i.e., objects or events) represented in the space, and to study their properties which will govern the space model.

4.2.1. Similarity measures used in MRSM

Following our MSRM definition, typical Euclidian distance can be utilized to compare the time coverage representative points of two social media objects or events:

$$\mathrm{Sim}_T(o_1, o_2) = \frac{1}{1 + \mathrm{Dist}_T(o_1, o_2)} \in [0, 1] \qquad \text{where}$$

Please cite this article as: M.A. Abebe, J. Tekli, F. Getahun et al., Generic metadata representation framework for social-based event detection, description, and linkage, Knowledge-Based Systems (2019), https://doi.org/10.1016/j.knosys.2019.06.025.



eid 1			
Extended (default)		Restricted	
T	{1404711029, 1404712589, 1404744523}	t_c	1. 04727776
L	{lat=9.005278 long=38.763334 alt=NULL lat=9.005401 long=38.763611 alt=NULL}	λ_c	lat=9.0053401 long=38.763471 alt=NULL
S	{torrent, rainfall, cloudburst, rainstorm, July, Meskel Square, torrent, rain, lane, Bole International Airport, stream, water, thunder-shower, rainfall, rain, downpour, flood, rainstorm, rain, hour, pockets, small businesses, problem, street boys, Ethiopia, Addis Ababa, flood, inundation, traffic, chaos, terrible, evening, Olompia, today, rainstorm, flood, road}	s_c	rainstorm

eid 2			
Extended (default)		Restricted	
T	{1238876657, 1238956813, 1238964834}	t_c	1238932768
L	{lat=45.5156 long=-73.5578 alt=NULL lat=45.517 long=-73.5571 alt=NULL}	λ_c	lat=45.5163 long=-73.55745 alt=NULL
S	{Ian Mosley, Mark Kelly, Pete Trewavas, Steve Hogarth, Steve Rothery, concert, gig, live, marillion, weekend, music, progressive, montreal, Steve Hogarth}	s_c	concert

eid 3			
Extended (default)		Restricted	
T	{1145040959, 1145041372, 145105698}	t_c	1145062676
L	{lat=47.4357 long=-122.294 alt=NULL lat=47.4357 long=-122.294 alt=NULL lat=47.4357 long=-122.294 alt=NULL}	λ_c	lat=47.4357 long=-122.294 alt=NULL
S	{Stardance, Norwescon Seattle, DoubleTreeHot l, Nikkor, Washington, conference, cosplay, costume, fantasy, scifi, convention}	s_c	fantasy

Fig. 7. Events generated based on the sample images from Fig. 6, described following MRSM. We provide both the *extended* and *restricted* representations of events, where the former is utilized as the default representation in our approach.

$$\text{Dist}_T(o_1, o_2) = |t_{c_1} - t_{c_2}| \tag{13}$$

The Haversine formula, commonly employed in geographic navigation to determine the great-circle distance between two points on a sphere [73], can be utilized to evaluate the geographic distance between two objects o_1 and o_2 in MRSM, based on their spatial stamps' longitude and latitude coordinates[14]:

$$Sim_L(o_1, o_2) = \frac{1}{1 + \text{Dist}_L(o_1, o_2)} \in [0, 1] \text{ where}$$

$$\text{Dist}_L(o_1, o_2) = 2 \times r \times \arcsin\left(\sqrt{\sin^2\left(\frac{\phi_2 - \phi_1}{2}\right) + \cos(\phi_1) \times \cos(\phi_2) \times \sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right) \tag{14}$$

where r is the average radius of the Earth (i.e. 6371 km), $\emptyset_1$ and $\emptyset_2$ are the latitude values of the spatial coverage representative points of objects o_1 and o_2 (in radians), and λ_1 and λ_2 are their longitude values (in radians).

As for the semantic dimension, semantic distance can be computed as the inverse of any typical semantic similarity measure comparing two concepts in a semantic network [30]. Here, semantic similarity measures can be classified as *edge-based* (estimating similarity as the shortest path between concepts) [74], *node-based* (estimating similarity as the maximum amount of information content concepts share in common) [75], and *gloss-based* (estimating similarity based on word overlap between the concept's gloss descriptions) [76]. In our study, we adopt an aggregate semantic similarity measure introduced in [63,77] producing similarity scores $\in$ [0,1], 0 designating minimal (*null*) similarity and 1 designating maximum (total) similarity:

$$\begin{aligned}\text{Sim}_S(o_1, o_2) = {} & w_{\text{Edge}} \times \text{Sim}_{\text{Edge}}(s_{c_1}, s_{c_2}, \text{KB}) + w_{\text{Node}} \\ & \times \text{Sim}_{\text{Node}}(s_{c_1}, s_{c_2}, \text{KB}) + w_{\text{Gloss}} \\ & \times \text{Sim}_{\text{Gloss}}(s_{c_1}, s_{c_2}, \text{KB}) \quad \in [0, 1]\end{aligned} \tag{15}$$

where: s_{c_1} and s_{c_2} designate the two concepts representing the semantic coverage representative points of o_1 and o_2 respectively, KB is the reference lexical knowledge base (we adopt WordNet 3.0[15]), $w_{\text{Edge}} + w_{\text{Node}} + w_{\text{Gloss}} = 1$ and $(w_{\text{Edge}}, w_{\text{Node}}, w_{\text{Gloss}}) \geq 0$, Sim_{Edge} is a typical edge-based measure from [74], Sim_{Node} is a typical node-based measure from [75], and $\text{Sim}_{\text{Gloss}}$ is a typical gloss-based measure from [76], expanded and normalized in [63,77].

Consequently, the similarity between two objects represented in MSRM can be computed as the aggregation of individual dimensional similarity measures, using any convenient aggregation function such as *maximum*, *minimum*, *average*, or *weighted sum*. We adopt the latter in our study to allow more user flexibility in fine-tuning the weights:

$$\begin{aligned}\text{Sim}_{\text{MRSM}}(o_1, o_2) = {} & w_T \times \text{Sim}_T(o_1, o_2) + w_L \times \text{Sim}_L(o_1, o_2) \\ & + w_S \times \text{Sim}_S(o_1, o_2) \quad \in [0, 1]\end{aligned} \tag{16}$$

where o_1 and o_2 are two social media objects in MRSM, $(\text{Sim}_T, \text{Sim}_L, \text{Sim}_S) \in$ [0,1] designate temporal, spatial, and semantic similarity measures respectively, w_T, w_L, w_S designate the similarity measures' coefficients (weight values) respectively where $w_T + w_L + w_S = 1$ and $(w_T, w_L, w_S) \geq 0$.[16] Similarity weight values can be set by the user or obtained empirically.

[14] To our knowledge, there is no geographic or spatial distance measure that considers the *altitude* value in its computations. Yet, we include *altitude* as part of the spatial stamps' description in order to utilize it later on when an appropriate distance/similarity measure becomes available.

[15] Available at: https://wordnet.princeton.edu/wordnet/download/standoff/.

[16] The same formula can be applied when computing $\text{Sim}_{\text{MRSM}}(\varepsilon_1, \varepsilon_2)$ where ε_1 and ε_2 are two events represented in their restricted form in MRSM.

4.2.2. Metric properties of MRSM

Based on the above formula and description, our comb MRSM similarity measure is consistent with the formal definition of similarity [20,78], and comes down to a *generalized metric*, i.e., a similarity (distance) function satisfying *minimality*, *reflexivity* and *symmetricity* properties, but not *triangular inequality*:

i. Minimality: $Sim_{MRSM}(o_1, o_2) = 0 \Rightarrow o_1$ and o_1 have no common characteristics,
ii. Self-similarity or Reflexivity: $Sim_{MRSM}(o_1, o_1) = 1$,
iii. Symmetricity: $Sim_{MRSM}(o_1, o_2) = Sim_{MRSM}(o_2, o_1)$
iv. Triangular inequality: $Sim_{MRSM}(o_1, o_2) \geq Sim_{MRSM}(o_1, o_3) \times Sim_{MRSM}(o_3, o_2)$ (i.e., $Dist_{MRSM}(o_1, o_2) \leq Dist_{MRSM}(o_1, o_3) + Dist_{MRSM}(o_3, o_2)$ where $Dist_{MRSM}$ is the inverse distance function of Sim_{MRSM})

Triangular inequality is usually domain and application-oriented [20,75]. While our temporal and spatial similarity measures do satisfy triangular inequality (following Euclidian and Haversine distances), yet most semantic similarity measures, e.g., [74–76], fail to satisfy the latter property. An example by Tversky [79] illustrates the *impropriety* of triangular inequality with an example about the similarity between countries: *"Jamaica is similar to Cuba (geographical proximity); Cuba is similar to Russia (political affinity); but Jamaica and Russia are not similar at all"*. That is because semantic similarity is usually evaluated through multiple semantic relations between concepts, e.g., *geographic proximity* on one hand, and *political affinity* on the other. A solution would be to consider one kind of semantic relations (e.g., *political affinity* only) when evaluating semantic similarity. Sim_S would be computed as the aggregation of multiple similarities evaluated each w.r.t. the corresponding semantic relation ($Sim_{S\_GeoProx}$, $Sim_{S\_PoliticalAff}$, etc.), where each measure would (individually, and when aggregated) verify *triangular equality*.

4.3. Event detection and description

Given a set of social media objects represented in MRSM, we group them into clusters, based on their time, space, and semantic similarities, where each cluster of objects identifies an event (cf. Definition 14). Here, we introduce an adapted graph-based agglomerative average-link clustering method (refer to [22] for a survey on clustering algorithms) as an unsupervised approach to perform event detection since we do not assume any knowledge about the events prior to the event detection process.[17]

The algorithm's pseudo-code is shown in Fig. 8. Given n input objects, the algorithm starts by computing the similarity between every pair of objects using our aggregate similarity measure (Sim_{MRSM}, cf. Eq. (16)). Aggregate similarity scores computed for all $n \times (n\text{-}1)/2$ pairs of objects are stored in an $(n \times n)$ matrix (i.e. SimMat[,], cf. lines 9–11). Clusters are then generated by varying the clustering level between l_o and 0, at a constant decrement pace of dec-value (line 12). The group link clusters for a clustering level l_i are identified by grouping together objects with similarity scores $\geq l_i$. Clustering at level l_o groups similar objects into an initial set of clusters by calling function *Generate_Initial_Clusters* (lines 13–14). Clustering at level l_i involves two steps (lines 15–21): (i) computing the similarity between the two clusters using UPGMA (Unweighted Pair-Group Averaging Method) [80], as shown in Eq. (17), and (ii) merging the clusters if their average pair-wise similarity is greater than or equal to l_i:

$$\text{Avg\_Sim}(\text{clust}_1, \text{clust}_2) = \frac{\sum_{o_i \in \text{clust}_1} \sum_{o_j \in \text{clust}_2} \text{Sim}_{\text{MRSM}}(o_i, o_j)}{|\text{clust}_1| \times |\text{clust}_2|} \tag{17}$$

where o_i and o_j are objects in clusters $clust_1$ and $clust_2$ respectively, and $|clust_1|$ and $|clust_2|$ are cluster cardinalities (in number of objects per cluster). A stopping rule is necessary to determine the most appropriate clustering level for the link hierarchies. Milligan & Cooper in [81] present 30 such rules, among them, *C-index* exhibits good performance and is thus adopted in our study (line 23). The clusters identified at the stopping level are then described as events following MRSM (line 24), by producing corresponding temporal, spatial, and semantic coverages obtained from their object descriptions (cf. Definition 14). For instance, given the objects in Figs. 6, 7 shows the events produced by our algorithm (considering equal weights for every MRSM dimension, and default parameter values for the event detection algorithm), along with their MRSM descriptions.

4.4. Identifying event relationships

Identifying the relationships between two (objects or) events can be performed by comparing their descriptions, i.e., their temporal, spatial, and semantic coverages and representative points in MRSM. We distinguish between three categories of relationships: (i) directional (e.g., *before*, *after*), (ii) metric (e.g., *far*, *near*), and (iii) topological (e.g., *include*, *intersect*). The following subsections describe each category of event relationships and how to identify them w.r.t. every dimension in MRSM.

4.4.1. Directional relationships

Directional relationships are identified for MRSM's temporal and spatial dimensions, and do not apply to the semantic dimension.

Definition 15 (*Temporal Directional Relationship* ($r_T^{\text{directional}}(\varepsilon_1, \varepsilon_2)$)**.** It refers to the exclusive directional relationship that can exist between two events ε_1 and ε_2 following their temporal coverages in MRSM, specifically: *before* ($\overset{before}{\rightarrow}$) and *after* ($\overset{before}{\rightarrow}$). Formally, considering events ε_1 and ε_2:

$$r_T^{\text{directional}}(\varepsilon_1, \varepsilon_2) = \begin{cases} \varepsilon_1 \overset{before}{\rightarrow} \varepsilon_2 & \text{if } t_e(\varepsilon_1) < t_s(\varepsilon_2) \\ \varepsilon_1 \overset{after}{\rightarrow} \varepsilon_2 & \text{if } t_s(\varepsilon_1) > t_e(\varepsilon_2) \end{cases} \tag{18}$$

where $T(\varepsilon_i) = \{t_s(\varepsilon_i), t_e(\varepsilon_i)\}$ represents the temporal coverage of ε_i, consisting of its start time and end time respectively •

In other words, an event ε_1 occurs *before* event ε_2 if ε_1 ends before ε_1 begins. Similarly, ε_1 occurs *after* ε_1 if ε_1 starts after ε_2 ends.

Definition 16 (*Spatial Directional Relationship* ($r_L^{\text{directional}}(\varepsilon_1, \varepsilon_2)$)**.** It refers to the directional relationship that can exist between two events ε_1 and ε_2 following their spatial coverages in MRSM, specifically: *north* ($\overset{north}{\rightarrow}$), *south* ($\overset{south}{\rightarrow}$), *east* ($\overset{east}{\rightarrow}$), *west* ($\overset{west}{\rightarrow}$), *above* ($\overset{above}{\rightarrow}$), and *below* ($\overset{below}{\rightarrow}$). Formally, considering events ε_1 and ε_2:

$$r_L^{\text{directtional}}(\varepsilon_1, \varepsilon_2) = \begin{cases} \varepsilon_1 \overset{north}{\rightarrow} \varepsilon_2 & \text{if } \phi_1 > \phi_2 \wedge \lambda_1 \simeq \lambda_2 \\ \varepsilon_1 \overset{south}{\rightarrow} \varepsilon_2 & \text{if } \phi_1 < \phi_2 \wedge \lambda_1 \simeq \lambda_2 \\ \varepsilon_1 \overset{east}{\rightarrow} \varepsilon_2 & \text{if } \phi_1 \simeq \phi_2 \wedge \lambda_1 > \lambda_2 \\ \varepsilon_1 \overset{west}{\rightarrow} \varepsilon_2 & \text{if } \phi_1 \simeq \phi_2 \wedge \lambda_1 < \lambda_2 \\ \varepsilon_1 \overset{above}{\rightarrow} \varepsilon_2 & \text{if } h_1 > h_2 \\ \varepsilon_1 \overset{below}{\rightarrow} \varepsilon_2 & \text{if } h_1 < h_2 \end{cases} \tag{19}$$

[17] Note that any general purpose clustering algorithm could have been used here. Yet, we adapt a graph-based agglomerative group average-link approach due to its well know effectiveness and acceptable efficiency (average $O(N^2)$ time) in various application scenarios [22,23,64].

Please cite this article as: M.A. Abebe, J. Tekli, F. Getahun et al., Generic metadata representation framework for social-based event detection, description, and linkage, Knowledge-Based Systems (2019), https://doi.org/10.1016/j.knosys.2019.06.025.

```
Algorithm: Event_Detection
Input:
1.    Objects: Collection          // collection of social media objects represented in MRSM
Variables:
2.    SimMat[ , ]: Decimal         // similarities of pairs of MM objects
3.    dec-value: Decimal           // clustering level decrement value (= -0.1 by default)
4.    Clusters: Collection         // clusters of objects
5.    li: Decimal                  // Clustering level
6.    cl: Decimal                  // stopping clustering level
7.    lo: Decimal                  // initial parameter to have m partitioned clusters (= 0.9 by default)
Output:
8.    Events: Collection           // contains the events detected
Begin
9.    For every oi in Objects
10.       For every oj in Objects
11.           SimMat[i, j] = SimMRSM(oi, oj)      // Computing pair-wise similarities
12.   For li= lo Down to 0 Step dec-value
13.      If li = lo Then
14.           Clusters = Generate_Initial_Clusters(SimMat)
15.      Else
16.           For each pair of clusters (clusti, clustj) in Clusters
              // Clusters contain the groups of objects at level li-1
17.           If Avg_Sim(clusti, clustj) ≥ li Then   // using UPGMA in Formula 11
18.                merge clusti and clustj in the same cluster
19.           End If
20.       Next
21.      End if
22.   Next
23.   cl =C-Index(Clusters)                  // stopping rule for clustering
24.   Events = MRSM(Clusters at cl)          // clusters obtained when stopping rule is reached
25.   Return Events                          // collection of events described in MRSM
End
```

Fig. 8. Pseudo code of our event detection algorithm.

where $\ell_c(\varepsilon_i) = \langle \emptyset_i, \lambda_i, h_i \rangle$ represents the spatial coverage representative point (center of gravity) of ε_i, consisting of its latitude, longitude, and altitude coordinates respectively[18] •

Note that while temporal directional relationships are exclusive (i.e., no two events can share both *before* and *after* relationships simultaneously), yet spatial directional relationships are inclusive and can occur simultaneously (e.g., $\varepsilon_1 \overset{north}{\rightarrow} \varepsilon_2$, $\varepsilon_1 \overset{west}{\rightarrow} \varepsilon_2$, and $\varepsilon_1 \overset{below}{\rightarrow} \varepsilon_2$ mean event ε_1 occurs to the *north west* of ε_2 and is *below* ε_2 in altitude).

4.4.2. Metric relationships

We consider two main metric relationships: *near* and *far*, that can be applied to all three dimensions of MRSM. We make use of MRSM's dimension-specific similarity measures (cf. Section 4.2) to define them.

Definition 17 (*Temporal Metric Relationship* ($r_T^{\text{metric}}(\varepsilon_1, \varepsilon_2)$)**.** It refers to the exclusive metric relationship that can exist between two events ε_1 and ε_2 following their temporal coverages in MRSM: *near* ($\overset{T_{near}}{\rightarrow}$) and *far* ($\overset{T_{far}}{\rightarrow}$). Considering events ε_1 and ε_2:

$$r_T^{\text{metric}}(\varepsilon_1, \varepsilon_2) = \begin{cases} \varepsilon_1 \overset{T_{near}}{\rightarrow} \varepsilon_2 & \text{if } \text{Sim}_T\,(t_c(\varepsilon_1), t_c(\varepsilon_2)) \leq \text{Thresh}_T \\ \varepsilon_1 \overset{T_{far}}{\rightarrow} \varepsilon_2 & \text{otherwise} \end{cases} \tag{20}$$

where $\text{Sim}_T(t_c(\varepsilon_1), t_c(\varepsilon_2))$ computes the temporal similarity (following Eq. (13)) between the temporal coverage representative points of ε_1 and ε_2, and Thresh_T is a (user defined or system computed) temporal closeness threshold •

In other words, an event ε_1 is considered to be temporally near another event ε_2 if ε_1's temporal midpoint is close to that of ε_2. Otherwise, the events are considered to be temporally far from each other.

Definition 18 (*Spatial Metric Relationship* ($r_L^{\text{metric}}(\varepsilon_1, \varepsilon_2)$)**.** It refers to the exclusive metric relationship that can exist between two events ε_1 and ε_2 following their spatial coverages in MRSM: *near* ($\overset{L_{near}}{\rightarrow}$) and *far* ($\overset{L_{far}}{\rightarrow}$). Considering events ε_1 and ε_2:

$$r_L^{\text{metric}}(\varepsilon_1, \varepsilon_2) = \begin{cases} \varepsilon_1 \overset{L_{near}}{\rightarrow} \varepsilon_2 & \text{if } \text{Sim}_L\,(\ell_c(\varepsilon_1), \ell_c(\varepsilon_2)) \leq \text{Thresh}_L \\ \varepsilon_1 \overset{L_{far}}{\rightarrow} \varepsilon_2 & \text{otherwise} \end{cases} \tag{21}$$

where $\text{Sim}_L(\ell_c(\varepsilon_1), \ell_c(\varepsilon_2))$ computes the spatial similarity (following Eq. (14)) between the spatial coverage representative points of ε_1 and ε_2, and Thresh_L is a (user defined or system computed) spatial closeness threshold •

In other words, an event ε_1 is considered to be spatially near another event ε_2 if ε_1's spatial midpoint (center of gravity) is close to that of ε_2. Otherwise, the events are considered to be spatially far from each other.

Definition 19 (*Semantic Metric Relationship* ($r_S^{\text{metric}}(\varepsilon_1, \varepsilon_2)$)**.** It refers to the exclusive metric relationship that can exist between two events ε_1 and ε_2 following their semantic coverages in MRSM: *near* ($\overset{S_{near}}{\rightarrow}$) and *far* ($\overset{S_{far}}{\rightarrow}$). Considering events ε_1 and ε_2:

$$r_S^{\text{metric}}(\varepsilon_1, \varepsilon_2) = \begin{cases} \varepsilon_1 \overset{S_{near}}{\rightarrow} \varepsilon_2 & \text{if } \text{Sim}_S\,(s_c(\varepsilon_1), s_c(\varepsilon_2)) \leq \text{Thresh}_S \\ \varepsilon_1 \overset{L_{far}}{\rightarrow} \varepsilon_2 & \text{otherwise} \end{cases} \tag{22}$$

where $\text{Sim}_S(s_c(\varepsilon_1), s_c(\varepsilon_2))$ computes the semantic similarity (following Eq. (15)) between the spatial coverage representative points of ε_1 and ε_2, and Thresh_S is a (user defined or system computed) semantic closeness threshold •

[18] Note that $\simeq$ identifies whether a pair of latitude/longitude coordinates are almost (approximately) equal, compared with *exact equality* (=). *Approximate equality* is evaluated using dedicated (user/system defined) latitude/longitude similarity thresholds, where $\emptyset_1 \simeq \emptyset_2$ comes down to verifying whether $|\emptyset_1 - \emptyset_2| < \text{Thresh}_{\emptyset}$ (likewise, $\lambda_1 \simeq \lambda_2$ comes down to verifying whether $|\lambda_1 - \lambda_2| < \text{Thresh}_{\lambda}$). We adopt approximate equality here to allow more flexibility in identifying spatial directional relationships.

An event ε_1 is considered to be semantically near another event ε_2 if ε_1's semantic midpoint (concept most similar to all others in ε_1's semantic coverage) is close to that of ε_2. Otherwise, the events are considered to be semantically far from each other.

4.4.3. Topological relationships

We consider four topological relationships: *equal*, *include*, *intersect*, and *disjoint*, applied to all three dimensions of MRSM.

Definition 20 (*Temporal Topological Relationship* ($r_T^{\text{topological}}(\varepsilon_1, \varepsilon_2)$)**.** It refers to the exclusive topological relationship that can exist between two events following their temporal coverages in MRSM: *equal* ($\overset{T_{equal}}{\rightarrow}$), *include* ($\overset{T_{include}}{\rightarrow}$), *intersect* ($\overset{T_{intersect}}{\rightarrow}$), and *disjoint* ($\overset{T_{disjoint}}{\rightarrow}$). Considering two events ε_1 and ε_2:

$$r_T^{\text{topological}}(\varepsilon_1, \varepsilon_2) = \begin{cases} \varepsilon_1 \overset{T_{equal}}{\rightarrow} \varepsilon_2 & \text{if } t_s(\varepsilon_1) \simeq t_s(\varepsilon_2) \wedge t_e(\varepsilon_1) \simeq t_e(\varepsilon_2) \\ \varepsilon_1 \overset{T_{include}}{\rightarrow} \varepsilon_2 & \text{if } t_s(\varepsilon_1) \leq t_s(\varepsilon_2) \wedge t_e(\varepsilon_1) \geq t_e(\varepsilon_2) \\ \varepsilon_1 \overset{T_{intersect}}{\rightarrow} \varepsilon_2 & \text{if } (t_s(\varepsilon_1) \leq t_s(\varepsilon_2) \wedge t_e(\varepsilon_1) \leq t_e(\varepsilon_2)) \\ & \quad \vee (t_s(\varepsilon_1) \geq t_s(\varepsilon_2) \wedge t_e(\varepsilon_1) \geq t_e(\varepsilon_2)) \\ \varepsilon_1 \overset{T_{disjoint}}{\rightarrow} \varepsilon_2 & \text{otherwise} \end{cases} \quad (23)$$

where $T(\varepsilon_i) = \{t_s(\varepsilon_i), t_e(\varepsilon_i)\}$ represents the temporal coverage of ε_i, consisting of its start time and end time respectively •

Definition 21 (*Spatial Topological Relationship* ($\mathbf{r}_{\mathbf{L}}^{\mathbf{topological}}(\varepsilon_1, \varepsilon_2)$)**.** It refers to the exclusive topological relationship that can exist between two events following their spatial coverages in MRSM: *equal* ($\overset{L_{equal}}{\rightarrow}$), *include* ($\overset{L_{include}}{\rightarrow}$), *intersect* ($\overset{L_{intersect}}{\rightarrow}$), and *disjoint* ($\overset{L_{disjoint}}{\rightarrow}$). Considering two events ε_1 and ε_2:

$$\mathbf{r}_{\mathbf{L}}^{\mathbf{topological}}(\varepsilon_1, \varepsilon_2) = \begin{cases} \varepsilon_1 \overset{L_{equal}}{\rightarrow} \varepsilon_2 & \text{if } equal_{\mathbf{L}}(\varepsilon_1, \varepsilon_2) \\ \varepsilon_1 \overset{L_{include}}{\rightarrow} \varepsilon_2 & \text{if } include_{\mathbf{L}}(\varepsilon_1, \varepsilon_2) \\ \varepsilon_1 \overset{L_{intersect}}{\rightarrow} \varepsilon_2 & \text{if } intersect_{\mathbf{L}}(\varepsilon_1, \varepsilon_2) \\ \varepsilon_1 \overset{L_{disjoint}}{\rightarrow} \varepsilon_2 & \text{otherwise} \end{cases} \quad (24)$$

where $equal_L()$, $include_L()$, and $intersect_L()$ are functions specific to the nature of the spatial coverage considered (e.g., *minimum boundary*, *actual surface*, or *path/trail coverage*, cf. Fig. 5) •

For instance, considering the *minimum boundary* or the *actual surface* spatial coverages, $equal_L()$, $include_L()$, and $intersect_L()$ functions come down to evaluating geometric equality, inclusion, and intersection between two surface areas in a Euclidian geometric space. However, considering the *path/trail* spatial coverage:

- Function $equal_L(\varepsilon_1, \varepsilon_2)$ is evaluated by checking whether all spatial stamps in both events' spatial coverages are pairwise identical, i.e., $\forall \ell_i \in L(\varepsilon_1), \forall \ell_j \in L(\varepsilon_2), \exists$ one-to-one mapping between every ℓ_i and $\ell_j / \ell_i = \ell_j$,
- Function $include_L(\varepsilon_1, \varepsilon_2)$ is evaluated by testing whether the path consisting of ε_1's spatial coverage is a sub-path of ε_2's coverage, i.e., $L(\varepsilon_1) = \ell_i, \ell_k, \ldots \ell_m \subset L(\varepsilon_2) = \ell_j, \ldots, \ell_i, \ell_k, \ldots \ell_m, \ldots, \ell_n$
- Function $intersect_L(\varepsilon_1, \varepsilon_2)$ is evaluated by testing whether ε_1 and ε_2's path coverages cross, i.e., $\exists$ path $(\ell_i, \ell_j) \in L(\varepsilon_1) \wedge \exists$ path $(\ell_m, \ell_n) \in L(\varepsilon_2)/(\ell_i, \ell_j)$ and (ℓ_m, ℓ_n) cross.[19]

[19] We adopt the *path/trail* spatial coverage and its $equal_L()$, $include_L()$, and $intersect_L()$ functions since they are processed in linear time (cf. Section 4.1.2).

While temporal and spatial topological relationships can be accurately identified given the events' temporal and spatial coverages, the same cannot be done with semantic topological relationships. The latter are fuzzy by nature due to the linguistic and non-Euclidian nature of semantic coverages, made of sets of concepts referencing a lexical knowledge base. To solve this issue, we adopt the *semantic relatedness* approach from [82,83] originally developed to identify the semantic topological relationships between two RSS feeds. The same approach can be utilized to the semantic coverages of two events in MRSM.

Definition 22 (*Semantic Topological Relationship* ($\mathbf{r}_{\mathbf{S}}^{\mathbf{topological}}(\varepsilon_1, \varepsilon_2)$)**.** It refers to the exclusive topological relationship that can exist between two events following their semantic coverages in MRSM: *equal* ($\overset{S_{equal}}{\rightarrow}$), *include* ($\overset{S_{include}}{\rightarrow}$), *intersect* ($\overset{S_{intersect}}{\rightarrow}$), and *disjoint* ($\overset{S_{disjoint}}{\rightarrow}$). Formally, considering two events ε_1 and ε_2:

$$\mathbf{r}_{\mathbf{S}}^{\mathbf{topological}}(\varepsilon_1, \varepsilon_2) = \begin{cases} \varepsilon_1 \overset{S_{equal}}{\rightarrow} \varepsilon_2 & \text{if } equal_{\mathbf{S}}(\varepsilon_1, \varepsilon_2) \\ \varepsilon_1 \overset{S_{include}}{\rightarrow} \varepsilon_2 & \text{if } include_{\mathbf{S}}(\varepsilon_1, \varepsilon_2) \\ \varepsilon_1 \overset{S_{intersect}}{\rightarrow} \varepsilon_2 & \text{if } intersect_{\mathbf{S}}(\varepsilon_1, \varepsilon_2) \\ \varepsilon_1 \overset{S_{disjoint}}{\rightarrow} \varepsilon_2 & \text{otherwise} \end{cases} \quad (25)$$

where $equal_S()$, $include_S()$, and $intersect_S()$ are defined following the *semantic relatedness* approach in [82,83] •

Following [82,83], the *semantic relatedness* between two semantic coverages $S(\varepsilon_1)$ and $S(\varepsilon_2)$, noted $SemRel(S(\varepsilon_1), S(\varepsilon_2))$, is evaluated as the cosine similarity of the *semantic enclosure* vectors of $S(\varepsilon_1)$ and $S(\varepsilon_2)$. The *semantic enclosure* of one concept c_1 within another concept c_2 designates how much of c_1's *semantic neighborhood* is included in c_2's semantic neighborhood, where the semantic neighborhood of a concept c_i is a set of concepts surrounding c_i in the reference lexical knowledge base (e.g., WordNet). Consequently, *semantic coverage* vectors describing $S(\varepsilon_1)$ and $S(\varepsilon_2)$ are produced, where the vector space dimensions represent each a distinct concept $c_m \in S(\varepsilon_1) \cup S(\varepsilon_2)$, such that the weight of a concept c_m in $S(\varepsilon_i)$ is computed as the maximum *semantic enclosure* of c_m within any of the other concepts $c_j \in S(\varepsilon_i)$. In other words, $SemRel(S(\varepsilon_1), S(\varepsilon_2))$ returns a value $\in [0,1]$ estimating how much of the semantic neighborhoods of $S(\varepsilon_1)$ and $S(\varepsilon_2)$'s concepts – i.e., how much of $S(\varepsilon_1)$ and $S(\varepsilon_2)$'s semantic meanings – are close to each other. As a result:

- Function $equal_S(\varepsilon_1, \varepsilon_2)$ is evaluated by checking whether the semantic relatedness between $S(\varepsilon_1)$ and $S(\varepsilon_2)$ is higher than a (user-defined or system computed) equality threshold, i.e., $SemRel(S(\varepsilon_1), S(\varepsilon_2)) > \text{Thresh}_{Equal}$,
- Function $include_S(\varepsilon_1, \varepsilon_2)$ is evaluated as the product of the semantic enclosure vectors of $S(\varepsilon_1)$ and $S(\varepsilon_2)$, designating whether $S(\varepsilon_1)$'s meaning is semantically included in $S(\varepsilon_2)$ or not,
- Function $intersect_S(\varepsilon_1, \varepsilon_2)$ is evaluated by checking whether semantic relatedness between $S(\varepsilon_1)$ and $S(\varepsilon_2)$ is comprised between two (user/system defined) thresholds for equality and disjointness: $\text{Thresh}_{Disjoint} \leq SemRel(S(\varepsilon_1), S(\varepsilon_2)) \leq \text{Thresh}_{Equal}$.

While semantic *equality*, *intersection*, and *disjointness* relationships are defined as fuzzy (approximate) relationships w.r.t. (predefined or pre-computed) semantic relatedness thresholds (cf. Fig. 9), nonetheless, this is not the case for the *inclusion* relationship which can be accurately identified by evaluating the product of semantic coverages' enclosure vectors (a more detailed description of the *semantic relatedness* approach from [82,83] is provided in the Appendix).

Please cite this article as: M.A. Abebe, J. Tekli, F. Getahun et al., Generic metadata representation framework for social-based event detection, description, and linkage, Knowledge-Based Systems (2019), https://doi.org/10.1016/j.knosys.2019.06.025.

Fig. 9. Basic semantic topological relationships and corresponding thresholds following [82,83].

4.4.4. Relationships identification algorithm

Our event relationships identification algorithm is shown in Fig. 10. It identifies the relationships between a pair of input events following the above definitions, and is iteratively applied on all pairs of events extracted by our event detection algorithm.

For instance, considering sample events ε_1, ε_2, and ε_3 in Fig. 7, the algorithm identifies the following relationships:

Temporal: directional: $\varepsilon_1 \xrightarrow{after} \varepsilon_2$, $\varepsilon_1 \xrightarrow{after} \varepsilon_3$, $\varepsilon_2 \xrightarrow{after} \varepsilon_3$, metric: $\varepsilon_1 \xrightarrow{T_{far}} \varepsilon_2$, $\varepsilon_1 \xrightarrow{T_{far}} \varepsilon_2$, $\varepsilon_2 \xrightarrow{T_{near}} \varepsilon_3$, topological: $\varepsilon_1 \xrightarrow{T_{disjoint}} \varepsilon_2$, $\varepsilon_1 \xrightarrow{T_{disjoint}} \varepsilon_2$, $\varepsilon_2 \xrightarrow{T_{disjoint}} \varepsilon_3$

Spatial: directional: $\varepsilon_1 \xrightarrow{east} \varepsilon_2$, $\varepsilon_1 \xrightarrow{east} \varepsilon_3$, $\varepsilon_2 \xrightarrow{north} \varepsilon_3$, metric: $\varepsilon_1 \xrightarrow{T_{far}} \varepsilon_2$, $\varepsilon_1 \xrightarrow{T_{far}} \varepsilon_2$, $\varepsilon_2 \xrightarrow{T_{near}} \varepsilon_3$, topological: $\varepsilon_1 \xrightarrow{L_{disjoint}} \varepsilon_2$, $\varepsilon_1 \xrightarrow{L_{disjoint}} \varepsilon_2$, $\varepsilon_2 \xrightarrow{L_{disjoint}} \varepsilon_3$

Semantic: metric: $\varepsilon_1 \xrightarrow{S_{far}} \varepsilon_2$, $\varepsilon_1 \xrightarrow{S_{far}} \varepsilon_3$, $\varepsilon_2 \xrightarrow{S_{far}} \varepsilon_3$, topological $\varepsilon_1 \xrightarrow{S_{disjoint}} \varepsilon_2$, $\varepsilon_1 \xrightarrow{S_{disjoint}} \varepsilon_2$, $\varepsilon_2 \xrightarrow{S_{disjoint}} \varepsilon_3$

Fig. 11 shows a simplified representation of the above events following MRSM, along with their temporal, spatial, and semantic relationships. Our approach starts with raw social media objects with their metadata, and then generates semantic-aware events with their relationships, producing an event-based knowledge graph which forms the seed for event-based CK. We mainly focus on the temporal (*Where*), location (*When*), and semantic (*What*) dimensions in this paper. Yet, producing full-fledged CK requires expanding our current event-based knowledge graph to include user related information (i.e., *Who*, *Why*, and *How* dimensions). Also, dedicated inference rules can be developed (e.g., having $\varepsilon_1 \xrightarrow{after} \varepsilon_2$ and $\varepsilon_2 \xrightarrow{after} \varepsilon_3$ means we can transitively infer $\varepsilon_1 \xrightarrow{after} \varepsilon_3$) to enhance the knowledge graph organization and expressiveness, which we aim to investigate in a future study.

4.5. Computational complexity

The time complexity of our social event detection, description, and linkage solution simplifies to $O(|N|^2 \times |\text{KB}| \times depth(\text{KB}))$ where $|N|$ designates the number of social media objects, $|\text{KB}|$ the number of concepts in the reference knowledge base, and $depth(\text{KB})$ its maximum depth. It is evaluated as the sum of the complexities of the four main modules of the *SEDDaL* framework:

i. *Social media object representation within MRSM*: simplifies to $O(|N| \times |S|^2 \times |\text{KB}| \times depth(\text{KB}))$ time, evaluated as the sum of the time complexities of: (i) identifying the temporal coverage representative points of all objects in the data collection (computed as the average of the start and end temporal stamps of an object, cf. Eq. (3)) which requires $O(|N|)$ time, (ii) identifying the spatial coverage representative points of all objects (computed as the geographic midpoints of the objects' spatial stamps, cf. Eq. (7)) which requires $O(|N| \times |L|)$ time where $|L|$ is the number of spatial stamps for a given object (cf. Definition 8), and (iii) identifying the semantic coverage representative points for all objects (computed as the concept that is most similar to all others within a given object's semantic coverage, cf. Eq. (8)) which requires $O(|N| \times |S|^2 \times |\text{KB}| \times depth(\text{KB}))$[20] time where $|S|$ is the number of semantic stamps/concepts for a given object.

ii. *Social media objects' similarity evaluation*: simplifies to $O(|\text{KB}| \times depth(\text{KB}))$ time, and is computed as the sum of the complexities of temporal, spatial, and semantic similarity evaluation measures: (i) temporal similarity evaluation (based the Euclidian distance between two temporal coverages' representative points, cf. Eq. (13), requires constant $O(1)$ time, (ii) spatial similarity evaluation (based on the Haversine's distance between two spatial coverages' representative points, cf. Eq. (14)) requires $O(1)$ time, and (iii) semantic similarity evaluation (combining edge-based, node-based, and gloss-based similarity between two semantic coverages' representative points/concepts, cf. Eq. (15)) requires $O(|\text{KB}| \times depth(\text{KB}))$ time.

iii. *Event detection process* (cf. Fig. 8): simplifies to $O(|N|^2 \times |\text{KB}| \times depth(\text{KB}))$ time, and comes down to the clustering algorithm's complexity: $O(|N|^2)$, combined with the complexity of the similarity evaluation process: $O(|\text{KB}| \times depth(\text{KB}))$.

iv. *Event relationships identification* (cf. Fig. 10): simplifies to $O(|E|^2 \times |S|^2 \times |\text{KB}|)$ time, where $|E|$ is the number of extracted events, and $|S|$ the number of semantic concepts (cardinality of the semantic coverage) for a given event, and is computed as the sum of the complexities of: (i) identifying the metric, directional, and topological relationships between two events, following both temporal and spatial dimensions, requires constant $O(1)$ time, (ii) identifying the semantic metric relationships requires $O(|\text{KB}| \times depth(\text{KB}))$ (to compute semantic similarity between two event's semantic representative points), and (iii) identifying semantic topological relationships requires $O(|S|^2 \times depth(\text{KB}))$ time (to evaluate the semantic relatedness between two event's semantic enclosures and compute their enclosure vectors' products, cf. Appendix).

5. Experimental evaluation

We have implemented SEDDaL to test and evaluate its performance, and compare it with alternative solutions in the literature. Written in Java, our implementation comprises of SEDDaL's four main modules: (i) social media object representation, (ii) similarity evaluation, (iii) event detection and description, and (iv) event relationships identification, and four metadata extractor methods, designed to extract social media objects' temporal, spatial, and textual descriptions obtained from YouTube, Flickr, Twitter, and the MediaEvalSED 2013 and 2014 image datasets [69,84]. It also includes a linguistic pre-processing component (performing *stop word removal*,[21] *tokenization*,[22] *stemming*,[23] and *word sense disambiguation*[24]) allowing to transform the objects' textual descriptions into semantic coverages made of sets of semantic concepts. WordNet 3.0 is utilized as the reference knowledge base in SEDDaL's current implementation,[25] where concepts represent sets of synonymous terms (or synsets).

[20] $O(|\text{KB}| \times depth(\text{KB}))$ underlines the complexity of the combined semantic similarity measure [63] adopted in our study, utilized to identify the concept that is most similar to all others in a multimedia object's semantic coverage. $depth(\text{KB})$ represents the maximum number of edges (semantic relationships) between KB's root node and its farthest leaf node.

[21] Using WordNet's stop word list: http://www.d.umn.edu/~tpederse/Group01/WordNet/wordnet-stoplist.html.

[22] Using the Stanford Tokenizer: https://nlp.stanford.edu/software/tokenizer.shtml.

[23] Using the Porter stemmer: http://tartarus.org/martin/PorterStemmer/.

[24] Using an implementation of the simplified LESK disambiguation algorithm: http://sigappfr.acm.org/Projects/XSDF/.

[25] The more comprehensive Yago knowledge based [59] can be used in the future.

```
Algorithm: Relationships_Identification
Input:
1.    Events: Collection                                   // collection of events represented in MRSM
2.    Thresh_T, Thresh_L, Thresh_S, Thresh_Disjoint, Thresh_Equal   // threshold values ∈ [0, 1]
Variables:
3.    ε_i, ε_j: events
Output:
4.    Rel{} = ∅      // set of relationships between all pairs of events
Begin
5.    For every pair (ε_i, ε_j) in Events
6.      Rel = Rel ∪ T_Directional (ε_i, ε_j)        // r_T^directional (ε_i, ε_j) following Definition 15
7.      Rel = Rel ∪ T_Metric (ε_i, ε_j, Thresh_T)   // r_T^metric (ε_i, ε_j) following Definition 17
8.      Rel = Rel ∪ T_Topological (ε_i, ε_j)        // r_T^topological (ε_i, ε_j) following Definition 20
9.      Rel = Rel ∪ L_Directional (ε_i, ε_j)        // r_L^directional (ε_i, ε_j) following Definition 16
10.     Rel = Rel ∪ L_Metric (ε_i, ε_j, Thresh_L)   // r_L^metric (ε_i, ε_j, Thresh_L) following Definition 18
11.     Rel = Rel ∪ L_Topological (ε_i, ε_j)        // r_L^topological (ε_i, ε_j) following Definition 21
12.     Rel = Rel ∪ S_Metric r_S^metric (ε_i, ε_j, Thresh_S)   // r_S^metric (ε_i, ε_j) following Definition 19
13.     Rel = Rel ∪ S_Topological (ε_i, ε_j, , Thresh_Disjoint, Thresh_Equal)   // r_S^topological (ε_i, ε_j), Definition 22
14.   Next
15.   Return Rel
End
```

Fig. 10. Pseudo code of our event relationships identification algorithm.

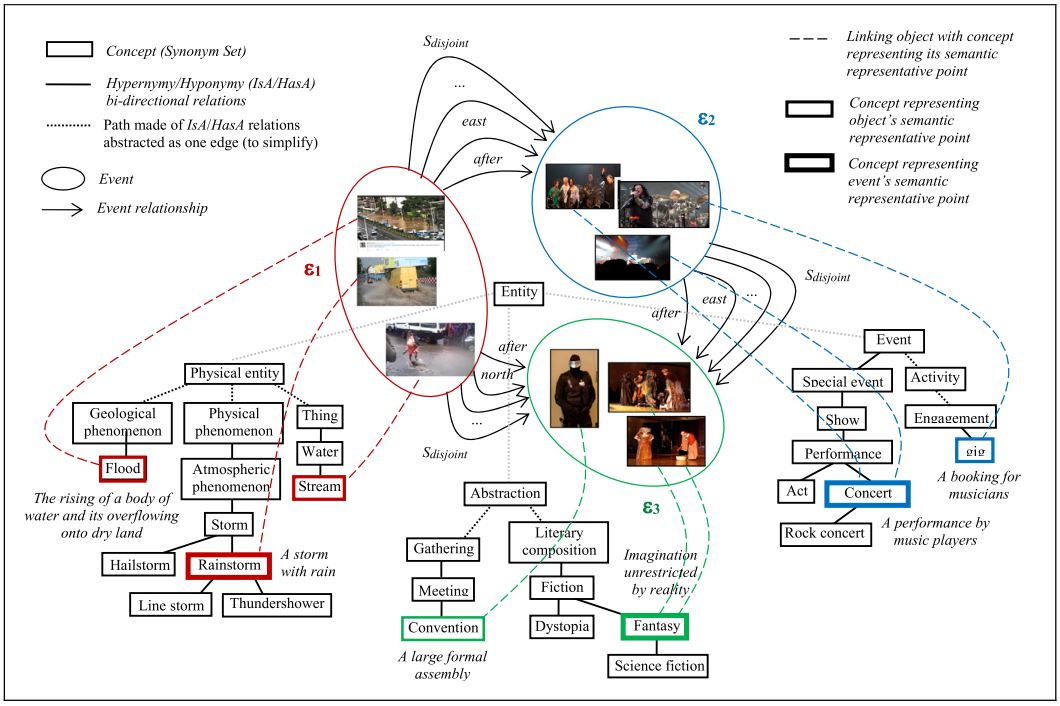


Fig. 11. Simplified event-based knowledge graph, describing the events from Fig. 7.

5.1. Experimental dataset and pre-processing

We utilized the MediaEvalSED 2013 and 2014 image datasets [69,84] to evaluate our event extraction approach. The 2013 dataset contains a collection of 131,211 photos and their associated metadata in XML (eXtensible Markup Language) format, and the larger 2014 dataset contains 362,578 photos with their metadata provided in JSON (Java Script Object Notation) format. Both datasets contain the ground truth event annotations created by human users. The ground truth consists of associating each image with a single label designating an event, such that no image can belong to more than one event. Image metadata contain *image_id*, *photo_url*, *username*, *dateTaken*, *dateUploaded*, *title*, *description*, *tags*, and *location* (defined in terms of *latitude* and *longitude*) among others, associated with every image. Based on MRSM, we only extract and process image metadata associated with temporal features (i.e., *dateTaken* and *dateUploaded*), spatial features (i.e., *latitude* and *longitude*), and semantic features (i.e., *title*, *tags*, and *description*). Note that almost all of the photos have temporal information, but only 46.1% of them have spatial information, 95.6% of them have tags, 97.9% have titles, and 37.9% have description information. The datasets were

pre-processed using our MediaEvalSED metadata extractor to: (i) convert temporal values into UNIX epoch,[26] (i) clean out the HTML tags (e.g., < br>, <i>, etc.) and remove the special characters embedded in the image's textual descriptions (e.g., " & < etc.), (ii) translate non-English textual metadata using the Google API Translate service, and (iii) replace hyphens by spaces or blank characters according to the existence of the word in WordNet. The images' textual descriptions, originally expressed in three elements in the source datasets (i.e., *title*, *descriptions*, and *tags*), were merged into one element (labeled *content*), and then processed through our linguistic-preprocessing component to produce the corresponding semantic coverages.

5.2. Evaluation metrics

To evaluate the quality of our event detection process, we use the *Normalized Mutual Information* (NMI) [85] and *f-score* measures [86] commonly utilized in the literature. NMI is an informed probabilistic measure that evaluates the clustering accuracy (purity) of extracted events, by comparing the generated clusters (events) with the user defined ones (ground truth):

$$\mathrm{NMI}(\Omega, C) = \frac{I(\Omega, C)}{[H(\Omega) + H(C)]/2} \in [0, 1] \tag{26}$$

where: $\Omega = \{w_1, w_2, \ldots, w_k\}$ is the set of generated clusters, $C = \{c_1, c_2, \ldots, c_j\}$ is the set of predefined clusters (ground truth), $I(\Omega, C)$ is the mutual information between the generated clusters and the predefined clusters, and $H(\Omega)$ and $H(C)$ are the entropies of the sets of generated clusters and predefined clusters respectively:

$$I(\Omega, C) = \sum_{c_i \in C} \sum_{w_j \in \Omega} p(c_i, w_j) \times \log_2 \frac{p(c_i, w_j)}{\left(p(c_i) + p(w_j)\right)} \text{ and}$$
$$H(X) = -\sum_{x_i \in X} p(x_i) \log_2 p(x_i) \tag{27}$$

where $p(c_i)$ underlines the probability of an object being in the predefined cluster c_i, and is computed as the number of objects in c_i that truly belong to c_i over the total number of objects in c_i (similarly for $p(w_j)$), and $p(c_i, w_j)$ underlines the probability of an object being in both c_i and w_j, and is computed as the cardinality (number of objects in) the intersection of c_i and w_j, i.e., $|c_i \cap w_j|$ over the cardinality of the union of c_i and w_j, i.e., $|c_i \cup w_j|$. NMI's score varies $\in$ [0,1], where a higher NMI value indicates a better agreement with the ground truth results (NMI = 1 indicates total agreement between generated clusters and predefined ones), whereas a lower NMI value (closer to 0) indicates lesser agreement with the ground truth [85].[27]

On the other hand, f-score measures the goodness of extracted events (clusters of objects), computed as the harmonic mean of *precision* (PR) and *recall* (R) measures widely utilized in information retrieval [87,88]:

$$\text{f-score} = \frac{2 \times PR \times R}{PR + R} \in [0, 1] \text{ where}$$
$$PR = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i} \in [0, 1] \text{ and}$$
$$R = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} a_i + \sum_{i=1}^{n} c_i} \in [0, 1] \tag{28}$$

For an extracted cluster C_i that corresponds to a given user identified event ε_i:

- a_i is the number of objects in C_i that indeed correspond to ε_i (correctly clustered objects).
- b_i is the number of objects in C_i that do not correspond to ε_i (miss-clustered).
- c_i is the number of objects not in C_i, although they correspond to ε_i (objects that should have been clustered in C_i).

High precision denotes that the clustering task achieved high accuracy, grouping together objects that actually correspond to the events mapped to the clusters. High recall means that very few objects are not in the appropriate cluster where they should have been (i.e., few objects are not associated to the proper event). Hence, high precision and recall, and thus high f-score (indicating in our case excellent clustering quality) characterize a good event detection method.

We also utilize typical precision and recall measures to evaluate the quality of our event relationships identification process.

5.3. Event detection quality

We conducted two sets of experiments to evaluate the event detection quality of our approach: (i) considering the impacts of MRSM's temporal, spatial, and semantic feature dimensions, and (ii) comparing our method with existing solutions.

5.3.1. Impact of feature dimensions

We ran four experiments using different parameter values for weight parameters w_T, w_L, and w_S highlighting the impact of temporal, spatial, and semantic feature dimensions.

In **Experiment #1**, we set the value of w_T to 0.0 and apply a stepwise increment of 0.1 until reaching its 1.0 upper bound. The values of w_L and w_S are set to be the same following w_T's variation, i.e., $w_L = w_S = (1\text{-}w_T)/2$. Experimental results in Table 1a and Fig. 12a show that both NMI and f-score values increase: from 0.9637-to-0.9845 and from 0.9863-to-0.9943 respectively, when w_T increases from 0.0-to-0.3. However, both evaluation metrics' values decrease when w_T increases from 0.4-to-1.0. The best NMI (i.e., 0.9943) and f-score (i.e., 0.9845) are obtained when $w_T = 0.3$ with $w_L = w_S = 0.35$. This concurs with the intuition behind the theoretical design of MRSM: highlighting that all three dimensions of the model have a significant impact on event detection. Note that when the value of $w_T = 1.0$, both the NMI (0.8992) and f-score (0.7427) evaluation metrics record their worst results, which shows that relying on the social media

[26] The temporal features in the MediaEvalSED 2013 and 2014 datasets, i.e., *dateTaken* and *dateUploaded*, are represented following the Internet date/time format, RFC 3339 (i.e., YYYY-MM-DD hh:mm:ss.f). But, the RFC 3339 date/time representation lacks time zone information. For example, the timestamp "2007-10-25 20:32:23.0" in Addis Ababa, Ethiopia (which is GMT + 3) and Washington, DC, USA (which is GMT-5) should represent different instants of time (given their time zone differences). To address this issue, we transform the Internet date/time format into the UNIX timestamp (cf. Section 4.1). The value of *dateTaken* is utilized as the object's time stamp.

[27] In cluster evaluation literature, *NMI* and *f-score* are commonly used metrics to evaluate cluster quality. Other metrics include *purity* ("predecessor" to *NMI*) and *Rand index* ("predecessor" to *f-score*) [87]. While the original *purity* measure counts the number of objects correctly assigned to their proper clusters, yet, its main downside is that it tends to increase with the increase in number of clusters, since the clusters become smaller and thus the number of objects put in the wrong clusters tends to decrease accordingly, reaching purity = 1 (maximum) when individual clusters are formed (where every object is put in its own "correct" cluster). NMI was introduced to handle the tradeoff between (i) number of correctly clustered objects and (ii) number of generated clusters, using an information-theoretic approach: evaluating the probability of an object being in the proper cluster. And to handle (penalize) obtaining a larger number of generated clusters (since probabilities would otherwise increase accordingly, which brings us back to the same problem of purity), NMI normalizes the probabilities by dividing them with the sum of the entropies of both the generated clusters and the reference (ground truth) clusters.

Table 1
NMI and f-score values obtained when varying the temporal, spatial, and semantic feature weight values.

a. Experiment 1: Impact of varying w_T				b. Experiment 2: Impact of varying w_L				c. Experiment 3: Impact of varying w_S			
w_T	$w_L = w_S$	NMI	F-score	w_T	$w_L = w_S$	NMI	F-score	w_T	$w_L = w_S$	NMI	F-score
0.0	0.50	0.9863	0.9637	0	0.50	0.9727	0.9285	0	0.50	0.9872	0.9626
0.1	0.45	0.9911	0.9756	0.1	0.45	0.9915	0.9772	0.1	0.45	0.9932	0.9810
0.2	0.40	0.9932	0.9825	0.2	0.40	0.9920	0.9785	0.2	0.40	0.9942	0.9843
0.3	**0.35**	**0.9943**	**0.9845**	**0.3**	**0.35**	**0.9942**	**0.9843**	0.3	0.35	0.9941	0.9841
0.4	0.30	0.9926	0.9792	0.4	0.30	0.9936	0.9830	**0.4**	**0.30**	**0.9943**	**0.9845**
0.5	0.25	0.9832	0.9511	0.5	0.25	0.9935	0.9830	0.5	0.25	0.9882	0.9713
0.6	0.20	0.9493	0.8568	0.6	0.20	0.9901	0.9725	0.6	0.20	0.9780	0.9499
0.7	0.15	0.9464	0.8492	0.7	0.15	0.9896	0.9714	0.7	0.15	0.9576	0.9005
0.8	0.10	0.9409	0.8326	0.8	0.10	0.9890	0.9693	0.8	0.10	0.9336	0.8540
0.9	0.05	0.9337	0.8144	0.9	0.05	0.9885	0.9678	0.9	0.05	0.9227	0.9329
1.0	0.0	0.8992	0.7427	1	0	0.9872	0.9631	1	0	0.9093	0.8095

object's temporal description only (while totally disregarding its spatial and semantic dimensions) does not help in detecting events.

In **Experiment #2**, we vary the value of w_L from 0.0 to 1.0, while applying a stepwise increment of 0.1. The values of w_T and w_S are set to be the same following w_L's variation, i.e., $w_T = w_S =$ (1-w_L)/2. Similarly to the previous experiment's results, Table 1b and Fig. 12b show that both NMI and f-score values increase: from 0.9727-to-0.9942 and from 0.9285-to-0.9843 respectively, when w_L increases from 0.0-to-0.3. Both evaluation metrics decrease when w_L increases from 0.4-to-1.0 (i.e., when w_T and w_S start to decrease from 0.3-to-0). The best NMI value (i.e., 0.9942) and f-score value (i.e., 0.9843) are obtained when $w_L = 0.3$ with $w_T = w_S = 0.35$. Results of Experiment 2 also concur with intuition behind our MRSM design: that all three temporal, spatial, and semantic dimensions have an important impact on event detection. Note that both NMI (0.9727) and f-Score (0.9285) record their worst results when $w_L = 0$, i.e., when totally disregarding the spatial dimension.

In **Experiment #3**, we vary the value of w_S 0.0 to 1.0 with a stepwise increment of 0.1. The values of w_T and w_L are set to be the same following w_S's variation, i.e., $w_T = w_S =$ (1-w_L)/2. Experimental results in Table 1c and Fig. 12c show that both NMI and f-score values increase: from 0.9872-to-0.9943 and from 0.9626-to-0.9845 respectively, when w_S increases from 0.0-to-0.4. Yet, both evaluation metrics decrease when w_S increases from 0.5-to-1.0 (i.e., when w_T and w_L start to significantly decrease from 0.25-to-0). The best NMI (i.e., 0.9943) and f-score (i.e., 0.9845) values are obtained when $w_S = 0.4$ and $w_T = w_L = 0.3$. This concurs with our intuition and the results of the previous experiments, where all three dimensions have a major impact on event detection. When the value of $w_S = 1.0$, both NMI (0.9093) and f-score (0.8095) record their worst results, which shows that using the objects' semantic description only (while disregarding its time and space descriptions) does not help in detecting events.

Experiment #4 set out to empirically identify an estimation of the parametric configuration of w_T, w_L, and w_S producing the best event detection quality. Here, we vary the weight values independently between [0.25, 0.45], where the latter designates the range of values for which each of the parameters produced its best results in the previous experiments. For each parameter, the weight values are incremented by 0.05 from the lower boundary (0.25) until reaching the upper boundary (0.45). This produces 64 different parametric configurations, a subset of which (including the top 10 configurations) is shown in Table 2. Results show that the best NMI (0.9943) and f-score (0.9845) values are obtained with $w_T = 0.25$, $w_L = 0.35$, and $w_S = 0.4$, which concurs with the previous experiments: where all three social metadata features seem important in extracting meaningful events. One could even suggest that the semantic feature dimension ($w_S = 0.4$) has a slightly better impact on event detection, compared with its temporal ($w_T = 0.25$) and spatial ($w_L = 0.35$) counterparts,

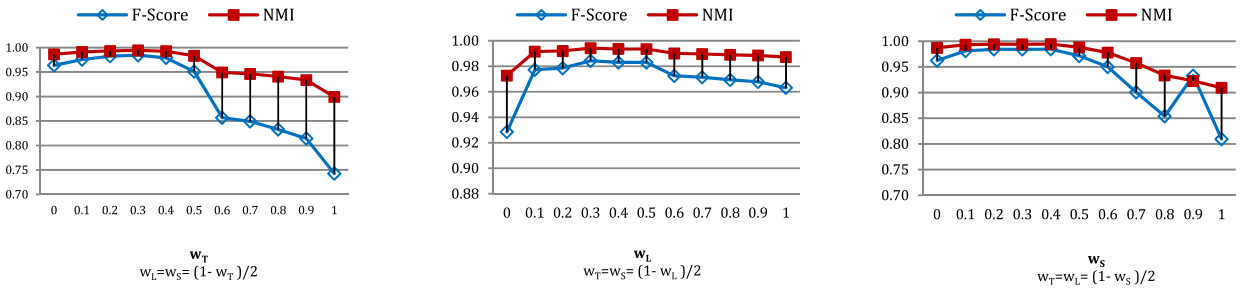
Table 2
The top ten results obtained in Experiment 4, by varying the temporal, spatial, and semantic parameter values.

w_T	w_L	w_S	NMI	F-score
0.25	**0.35**	**0.4**	**0.9943**	**0.9845**
0.35	0.4	0.25	0.9942	0.9844
0.3334	0.3334	0.3334	0.9941	0.9841
0.4	0.35	0.25	0.994	0.9835
0.25	0.4	0.35	0.9937	0.9834
0.3	0.45	0.25	0.9936	0.9831
0.25	0.45	0.3	0.9934	0.9826
0.25	0.3	0.45	0.9933	0.9826
0.35	0.25	0.4	0.9924	0.9793
0.4	0.25	0.35	0.9924	0.9789
0.45	0.3	0.25	0.9924	0.9793
0.3	0.25	0.45	0.9916	0.9778
0.45	0.25	0.3	0.9916	0.9769

and that the impact of the temporal feature seems relatively less important than the other two. Yet by visualizing the variations of all three parameters w.r.t. NMI in Fig. 13, one can also realize that the latter observation does not seem to hold or generalize given the fluctuations of the weights in the experimental process.

Discussion: Results in Experiments 1-to-4 highlight three main observations: First, all three dimensions seem to be almost equally important in extracting meaningful events, since the best results were obtained with close weight values for w_T, w_L, and w_S. Second, considering the semantic descriptions of images and their semantic similarities is beneficial for event extraction since both NMI and f-score regularly increase with the increase of parameter w_S (as long as the weights of the temporal and spatial dimensions are also significant). Third, considering semantic information only (neglecting temporal and spatial dimensions, i.e., $w_T = w_L = 0$ and $w_S = 1$), or considering temporal only or spatial only information, produces lower quality results, which points back to our first observation: integrating all dimensions seems to be key in improving event extraction quality.

Note that identifying and fine-tuning the parametric weight values of social media object features (and other system parameters) can be handled automatically as an optimization problem, such that parameters should be chosen to maximize event detection quality (through some cost function such as NMI or f-score). This can be solved using a number of known techniques that apply linear programming and/or machine learning to identify the best weights for a given problem class, e.g., [67,89,90]. The main idea is to assign a higher (lower) weight to every (combination of) parameter(s), acting like contrast filters in image processing by increasing the contrast on input matrixes. Providing such a capability, in addition to manual tuning, would enable the user to start from a sensible choice of values (e.g., identical weight parameters to consider all features equally, i.e., $w_T = w_L = w_S =$ 0.3334) and then optimize and adapt the event identification



a. Experiment 1: Impact of varying w_T **b.** Experiment 2: Impact of varying w_L **c.** Experiment 3: Impact of varying w_S

Fig. 12. Visualizing NMI and f-score results highlighting the impact of temporal, spatial, and semantic features on the event detection task.

process following the scenario and the nature of the data at hand, giving more emphasis to the temporal, spatial, or (inclusive) semantic features of the data objects being compared. We do not further discuss parameter weight optimization here since it is out of the scope of this paper (to be addressed in a future study).

5.3.2. Comparative study

Table 3 summarizes the main differences between our method and existing social event detection methods. In short, our approach: (i) provides a generic representation model that can describe any kind of social metadata, (ii) does not require any predefined clues to identify events, (iii) considers the semantic meaning associated with metadata using a reference lexical knowledge base, (iv) combines three main event descriptive features: time, space, and semantics, allowing the user to fine-tune their impact in the event detection process, (v) describes the extracted events following the same generic representation model used to describe social media objects, and most importantly: (v) identifies different kinds of relationships (directional, metric, and topological) that can exist between events, which are not addressed in most existing methods.

We experimentally compare our method, considering the best results obtained via our optimal parametric configuration ($w_T = 0.25$, $w_L = 0.35$, and $w_S = 0.4$), with alternative solutions, namely approaches that have also adopted the MediaEvalSED 2013 and 2014 datasets [69,84] as benchmarks for cluster-based event detection. Results in Table 4 show that our approach is able to improve the event extraction process. This is mainly due to the fact that our solution considers the semantic descriptions and semantic similarities of user contributed metadata in the aggregated similarity evaluation process when performing similarity-based clustering, whereas existing methods focus mainly on the temporal/spatial aspects. Most methods, e.g., [6,8,41,91], consider the images' textual descriptions by performing syntactic processing (using term frequency or n-gram vector comparisons) but disregard the semantic meaning of the text. The approach in [27] considers the images' spatial features only to initially cluster the collection of images (temporal features are used only if spatial features are not available), and then only uses semantic similarity to refine/merge the produced clusters (rather than integrating semantics in the initial clustering process), while the approach in [28] expands the images' textual descriptions by identifying the synonyms and hypernyms of every term, producing expanded bag-of-words representations which are then compared using a typical syntactic similarity measure (i.e., cosine). The authors in [91,92] consider, in addition to the temporal, spatial, and textual features, some of the images visual properties using adaptations of the bag-of-visual-words (BoVW) model defined on the images' color and texture features. Yet, results in Table 4b show that considering visual features in both [91,92] in did not improve event detection quality.[28]

Nonetheless, results in Table 4 show that our approach's improvement in event detection effectiveness seems relatively small compared with existing solutions, considering both MediaEvalSED 2013 and 2014 datasets. This is due to two reasons: (i) the nature of social-based events which can be detected fairly accurately using temporal and spatial data only (which is done with most existing solutions), and (ii) certain existing methods consider some form of syntactic textual similarity evaluation or partly consider semantic meaning (e.g., counting the number of common synonyms and hypernyms) which also improves their performance. Here, our contribution is two-fold: (i) our results show that including knowledge-based semantics and full-fledged semantic similarity evaluation further improves quality, even though by a relatively reduced margin (since we are competing at the upper tier of the performance scale), and most importantly (ii) our approach goes farther than event detection, to represent events and extract their different relationships (metric, topological, and directional, following all three temporal, spatial, and semantic dimensions) in a generic representation model. This is central to allow event-based CK representation later on, and requires additional (semantic) processing which is not performed by most existing methods.

Note that further improvement to the quality of the event extraction process could be obtained by utilizing more accurate word sense disambiguation and semantic analysis techniques. While we adopt the commonly used simplified LESK algorithm [94] in our current implementation, yet, exploring more recent and advanced algorithms, e.g., SSI [95] and UKB [96], could help identify more accurate semantic representations of social media objects based on their textual metadata. In addition, while we utilize legacy node-based [74], edge-based [75], and gloss-based methods [76] to evaluate the semantic similarity between pairs of individual concepts (describing objects or events), yet exploring more recent approaches, e.g., for evaluating the semantic similarity between pairs of text sequences [97], or between digital item descriptions [98], could help further improve both the quality and performance of the event extraction process. A dedicated empirical study comparing and evaluating the impact of the latter techniques within the context of our framework is reported to a future extension of this work.

[28] Note that varying the training set size does not affect performance levels in our case since we adopt an unsupervised approach in our study. Yet, varying training set size in a supervised context would be essential to evaluating the effectiveness of the proposed solution.

Please cite this article as: M.A. Abebe, J. Tekli, F. Getahun et al., Generic metadata representation framework for social-based event detection, description, and linkage, Knowledge-Based Systems (2019), https://doi.org/10.1016/j.knosys.2019.06.025.

Table 3
Characteristics of main alternative methods for event detection from shared social media data on the Web.

Approach	Method	Data type	Data source	Temporal feature	Spatial feature	Semantic feature	Other features	Weighted features	External resource	Event relationships
Psallidas et al. [7]	Unsupervised (clustering)	Tweet	Twitter posts	✓	✓	×[a]	Publisher (*who*)	×	URL	×
Ling and Abhishek [10]		Photo	Flickr photos	✓	✓	×	×	×	×	×
Liu et al. [26]		Photo	EventMedia	✓	✓	×	×	×	×	×
Rafailidis et al. [24]		Photo & Video	MediaEval	✓	✓	×	×	×	×	×
Zaharieva et al. [29,93]		Photo	MediaEval	✓	✓	×[b]	×	×	×	×
Gupta et al. [27]		Photo & Video	MediaEval	✓	✓	✓	×	×	WordNet	×
Manchon-Vizuete and Giro-i-Nieto [6]		Photo & Video	MediaEval	✓	✓	×[a]	Author (*who*)	×	×	×
Manchon-Vizuete et al. [28]		Photo & Video	MediaEval	✓	✓	×[a]	Author (*who*) Visual (BoVW)	×	×	×
Becker et al. [35]	Supervised (classification)	Tweet	Twitter posts	✓	✓	×[a]	×	×	URL, and Upcoming Yahoo API	×
Liu X., 2011 [36]		Text, Photo & Video	Flickr photos and YouTube videos	✓	✓	✓	×	×	LOD[c], and WordNet	×
Becker et al. [37]		Photo & Video	Twitter, YouTube, and Flickr	✓	✓	×	×	×	×	×
Becker et al. [9]	Hybrid	Photo	Flickr photos	✓	✓	×	×	×	×	×
Wistuba and Lars [38]		Photo & Video	MediaEval	✓	✓	×	×	×	×	×
Papaoikonomou et al. [39]		Photo & Video	MediaEval	✓	✓	×	×	×	×	×
Sutanto and Nayak [8]		Photo & Video	MediaEval	✓	✓	×[a]	×	×	×	×
Sutanto and Nayak [91]		Photo	MediaEval	✓	✓	×[a]	Visual (BoVW)	×	×	×
Nguyen et al. [41]		Photo & Video	MediaEval	✓	✓	×[a]	Author (*who*)	×	×	×
Guo et al. [92]		Photo	MediaEval	✓	✓	×[a]	Visual (BoVW)	×	×	×
Gregor L. et al. [53]		News articles	News Feed	✓	×[d]	×[a]	Agent (*who*)	×	GeoNames, and DMoz	×
Rospocher M. et al. [54]		News articles	News Feed	✓	✓	✓	Agent (*who*)	×	DBPedia, and NAF[e]	×[f]
Our Approach (SEDDaL)	Unsupervised (clustering)	Photo & Video	Flickr, YouTube, Twitter, & MediaEval	✓	✓	✓	×	✓	WordNet	✓

[a]Processing textual descriptions syntactically, without considering their semantic meaning.
[b]Extracting latent semantics from the statistical analysis of textual descriptions, i.e., implicit semantic concepts which do not align with human-interpretable concepts [32].
[c]Linked Open Data.
[d]Location information is extracted after the events have been identified.
[e]NLP Annotation Framework, available at: http://wordpress.let.vupr.nl/naf/.
[f]Identifies linguistic-based *entity-event* relationships (e.g., <Porsche, *AquiredBy*, Volkswagen>), rather than directional, metric, and topological *event–event* relationships following the temporal, spatial, and semantic event feature dimensions targeted in our study.

5.4. Event relationships identification

We have also evaluated our approach's effectiveness in identifying the different directional, metric, and topological relationships between events. For this purpose, we generated 100 synthetic event representations (consisting of event feature coverages and their representative points) following MRSM, and then varied the event descriptions to highlight different relationship distributions. As a result, we underline the following observations.

First, our approach accurately identifies all ***directional relationships***, following both temporal and spatial dimensions, producing f-score = 1 at all times, since the latter are identified based on crisp and exact rules.[29]

Second, our approach identifies ***metric relationships:*** *far* and *near* w.r.t. all three MRSM dimensions, yet with different accuracy levels depending on the dimension specific similarity thresholds

[29] F-score graphs are omitted for directional relationships since they only show one single and constant value: f-score = 1.

Table 4
Comparison with alternative event detection methods.

a. Results obtained using the MediaEvalSED 2013 dataset				b. Results obtained using the MediaEvalSED 2014 dataset			
Method	Features	NMI	F-score	Method	Features	NMI	F-score
Gupta et al. [27]	Temporal (partly), Spatial, Semantic (partly)	0.1802	0.1426	Guo et al. [92]	Temporal, Spatial, Textual, Visual (BoVW)	0.9018	0.7525
Sutltano and Nayak [8]	Temporal, Spatial, Textual	0.9540	0.8120	Sultano et Nayak [91]	Temporal, Spatial, Textual, Visual (BoVW)	0.9024	0.7533
Manchon-Vizuete and Giro-i-Nieto [6]	Temporal, Spatial, Textual	0.9731	0.8833	Manchon-Vizuete et al. [28]	Temporal, Spatial, User ID, Semantic (partly)	0.9820	0.9240
Nguyen et al. [41]	Temporal, Spatial, Textual	0.9849	0.9320	Zaharieva et al. [93]	Temporal, Spatial, Textual	0.9866	0.9386
Our method (SEDDaL)	**Temporal, Spatial, Semantic**	**0.9865**	**0.9435**	**Our method (SEDDaL)**	**Temporal, Spatial, Semantic**	**0.9880**	**0.9430**

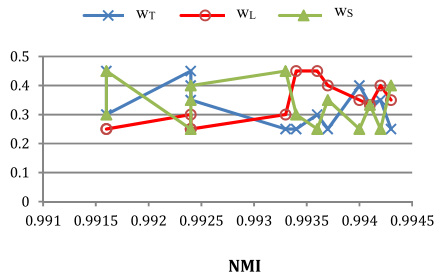


Fig. 13. Visualizing parameter weight variations w.r.t. NMI (a similar graph can be obtained w.r.t. f-score).

($Thresh_T$, $Thresh_L$, and $Thresh_S$) utilized to distinguish between the two relationships. Fig. 14 shows the results obtained with the semantic dimension,[30] on a distribution consisting of: 50 *S_far* and 50 *S_near* relationships, centered on $Thresh_S = 0.3$, following a normal distribution from 0 to 1. Results show that all *S_near* relationships are correctly identified at $Thresh_S = 0.3$ (f-score = 1), such that: (i) the number of false positives increases when $Thresh_S$ increases from 0.3-to-1, highlighting a decrease in precision from 1-to-0.5 (minimum precision = 50 is obtained when all 50 *S_far* relationships are considered as false *S_near* relationships, cf. Fig. 15a), and (ii) the number of false negatives (i.e., the number of *S_near* relationships that are missed) increases when $Thresh_S$ decreases from 0.3-to-0, highlighting a decrease in recall from 1-to-0 (minimum recall = 0 is reached when all 50 *S_near* relationships are disregarded at $Thresh_S = 0$). The behavior of our solution in detecting the *S_far* relationships, reflected in the results in Fig. 14, is inversely proportional to that of detecting *S_near* relationships, which conforms with their definition (if the metric relationship is not *near*, then it is *far*, and vice-versa, cf. Section 4.4.2, hence detecting more *near* relationships means detecting less *far* ones, and vice-versa).

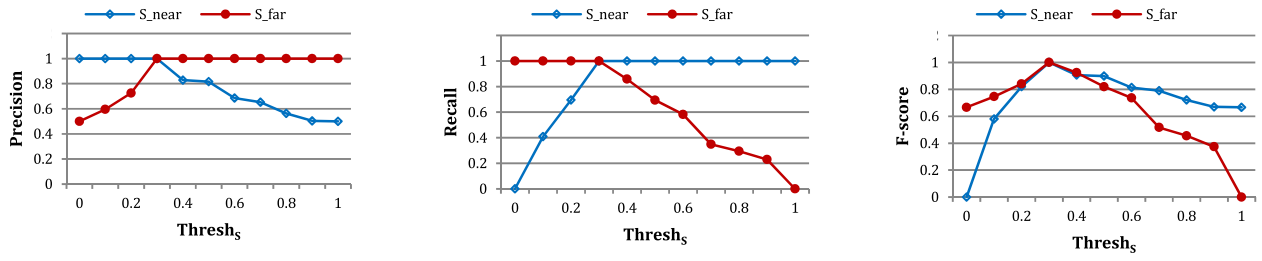
Third, our approach identifies all ***topological relationships***, following both temporal and spatial dimensions, producing f-score = 1 at all times, since the latter are identified based on crisp and exact rules (similarly to directional ones). The same goes for the *S_inlude* topological relationship following the semantic dimension (which can be exactly identified based on the product of the semantic enclosure vectors of the concerned events). As for the other semantic topological relationships: *S_equal*, *S_intersect*, and *S_disjoint*, their accurate identification depends on the similarity thresholds ($Thresh_{Equal}$ and $Thresh_{Disjoint}$) utilized to distinguish between the three relationships (similarly to metric ones).

Figs. 15 and 16 show the precision, recall, and f-score results obtained with a distribution consisting of: 25 *equal*, 25 *include*, 25 *intersect*, and 25 *disjoint* relationships, centered on $Thresh_{Equal} = 0.65$ and $Thresh_{Disjoint} = 0.35$, following normal distributions from 0-to-1. Fig. 15 shows the results obtained when varying $Thresh_{Equal}$ to identify *S_equal* and *S_intersect*, considering a fixed $Thresh_{Disjoint} = 0.35$ (its optimal value in this experiment). Fig. 16 shows the results obtained when varying $Thresh_{Disjoint}$ to identify *S_intersect* and *S_disjoint*, considering a fixed $Thresh_{Equal} = 0.65$ (its optimal value in this experiment).

Results in Fig. 15 show that all *S_equal* relationships are correctly identified at $Thresh_S = 0.65$ (f-score = 1), such that: (i) the number of false positives increases when $Thresh_{Equal}$ decreases from 0.65-to-0.35, highlighting a decrease in precision from 1-to-0.5 (minimum precision = 0.5 is obtained when all 25 *S_intersect* relationships are considered as false *S_equal* relationships, cf. Fig. 15a), and (ii) the number of false negatives (i.e., the number of *S_equal* relationships that are missed) increases when $Thresh_{Equal}$ increases from 0.65-to-1, highlighting a decrease in recall from 1-to-0 (minimum recall = 0 is reached when all 25 *S_equal* relationships are disregarded at $Thresh_{Equal} = 1$, Fig. 15b). Results for detecting *S_intersect* relationships are inversely proportional to those of detecting *S_equal* ones, which is expected following their definition (the topological relationship that can occur on either side of $Thresh_{Equal}$ is *S_equal* or *S_intersect*: if it is not *S_equal*, then it is *S_intersect*, and vice-versa).

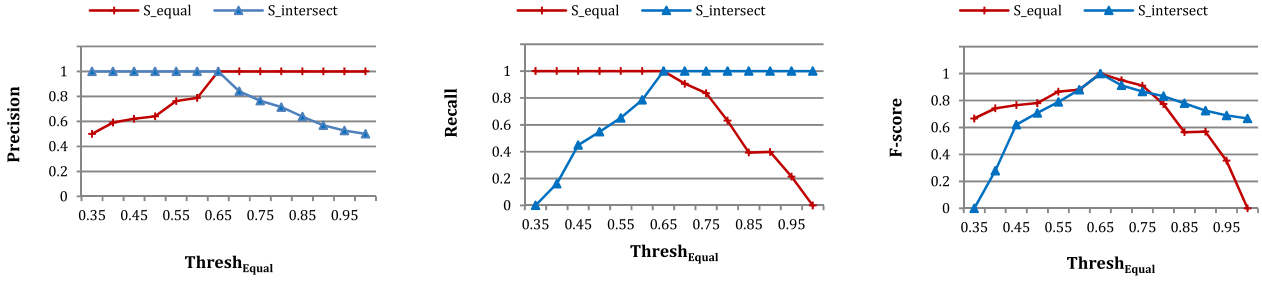
Similar results are obtained in Fig. 16, where all *S_disjoint* relationships are correctly identified at $Thresh_{Disjoint} = 0.35$ (f-score = 1), such that: (i) the number of false positives increases when $Thresh_{Equal}$ increases from 0.35-to-0.65, highlighting a decrease in precision from 1-to-0.5 (minimum precision = 0.5 is obtained when all 25 *S_disjoint* relationships are considered as false *S_intersect* relationships, cf. Fig. 16a), and (ii) the number of false negatives (i.e., the number of *S_equal* relationships that are missed) increases when $Thresh_{Disjoint}$ decreases from 0.35-to-0, highlighting a decrease in recall from 1-to-0 (minimum recall = 0 is reached when all 25 *S_disjoint* relationships are disregarded at $Thresh_{Disjoint} = 0$, Fig. 16b). Results for detecting *S_intersect* relationships are inversely proportional to those of detecting *S_disjoint* ones, which conforms with their definition. Note that

[30] Similar results are obtained with the temporal and spatial dimensions, and thus are omitted here for ease of presentation.



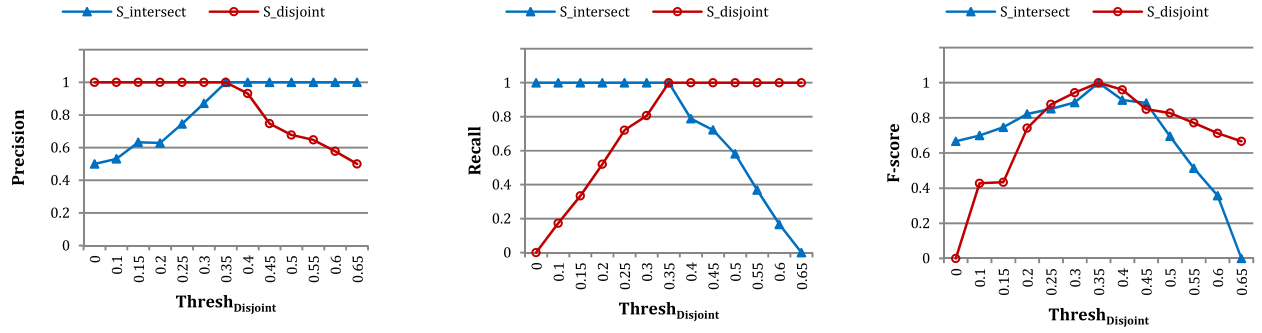
a. Precision **b.** Recall **c.** F-score

Fig. 14. Effectiveness in identifying semantic metric relationships.



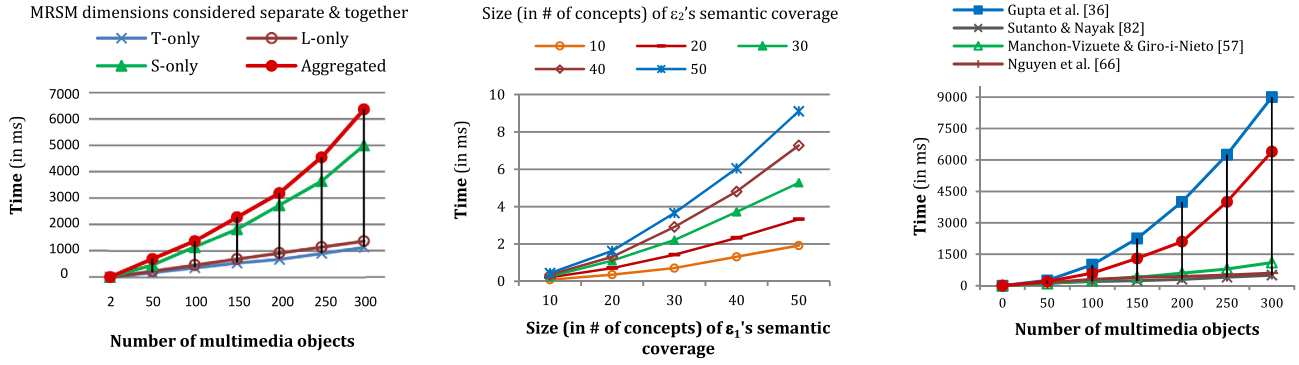
a. Precision **b.** Recall **c.** F-score

Fig. 15. Effectiveness in identifying semantic topological relationships: *S_equal* and *S_intersect*, when varying Thresh_{Equal}.



a. Precision **b.** Recall **c.** F-score

Fig. 16. Effectiveness in identifying semantic topological relationships: *S_intersect* and *S_disjoint*, when varying $\text{Thresh}_{Disjoint}$.



a. Time performance of our event detection process w.r.t. dataset size and MRSM similarity measures

b. Time performance when identifying all semantic relationships between two events.

c. Time performance compared with existing approaches.

Fig. 17. Time performance of our event detection process, w.r.t. dataset size and similarity evaluation measure.

a correlation can be identified between the threshold values and the distribution of event relationships, as well as the interplay between Thresh_{Equal} and $\text{Thresh}_{Disjoint}$. This can be inferred using

learning or regression based optimization techniques as described in Section 4.2, which is outside the scope of this study.

5.5. Time performance

Time experiments were carried out on an HP ProLiant ML350 Generation 5 (G5) Dual-Core Intel® Xeon™ 5000 processor with 2.66 GHz processing speed and 16 GB of RAM. Images from the MediaEvalSED 2013 dataset were utilized as benchmark. As shown in Section 4.5, our SEDDaL framework solution is of $O(|N|^2 \times |\text{KB}| \times depth(\text{KB}))$ where $|N|$ designates the number of social media objects being processed, $|\text{KB}|$ the number of concepts in the reference knowledge base (we utilize WordNet 3.0), and $depth(\text{KB})$ its maximum depth. It mainly comes down to the complexity of our event detection (clustering based) process which we evaluate in Fig. 17a, considering each of MRSM's dimensions separately (temporal only: $w_T = 1$, $w_L = w_S = 0$, spatial only: $w_T = 1$, $w_L = w_S = 0$; and semantic only, $w_T = 1$, $w_L = w_S = 0$) as well all three dimensions put together ($w_T \neq 0$, $w_L \neq 0$, $w_S \neq 0$). Results in Fig. 17a show that time grows in a polynomial fashion with the dataset size, where a clearly recognizable overhead is added when considering the semantic dimension. This concurs with our theoretical complexity analysis where performing semantic similarity evaluation requires an extra $O(|\text{KB}| \times depth(\text{KB}))$ time for every pair of social media objects being compared.

We also evaluated the time required to identify the semantic relationships between events, which is of $O(|E|^2 \times |S|^2 \times |\text{KB}|)$ where $|E|$ is the number of extracted events and $|S|$ the semantic coverage size (in number of semantic concepts) per event. We evaluate the time required to identify the semantic relationships between two individual events, which complexity simplifies to $O(|S|^2)$ when KB is fixed (i.e., size of WordNet). Fig. 17b shows the quadratic dependency on the combined events' semantic coverage sizes, which equally underlines a linear dependency on each event's semantic coverage size. Similar results (omitted here) highlight quadratic time dependency on the number of events. Note that the time performance of our event relationships identification process is negligible (in the order of seconds, Fig. 17b) compared with the time performance of the event detection process (in the order of thousands of seconds, cf. Fig. 17a) since the former depends on number of produced events $|E|$, which is always negligible compared with the number of objects $|N|$ provided as input to the event detection process. In other words, empirical results confirm our complexity analysis and show that the overall performance of our approach is chiefly governed by the performance of the event detection process (cf. Section 4.5).

To sum up, Fig. 17c compares our solution's time performance with existing approaches. Results show that our solution (considering all three temporal, spatial, and semantic dimensions) is more time consuming compared with [6,8,41]. Referring to Fig. 17a, one can realize that the added overhead is due to evaluating the semantic meaning of the textual descriptions (whereas existing solutions in [6,8,41] perform syntactic-only processing). Yet, results also show that our approach is less expensive that Gupta et al.'s semantic-aware solution in [27], since the latter performs semantic processing on all user contributed tags describing every object (amounting to multiple concepts per object), whereas our solution only considers the object' semantic coverage representative point (amounting to one single concept per object) in the semantic similarity evaluation process.

Fig. 18. Basic semantic topological relationships and corresponding thresholds following [82,83] (reported from Fig. 9).

6. Conclusion

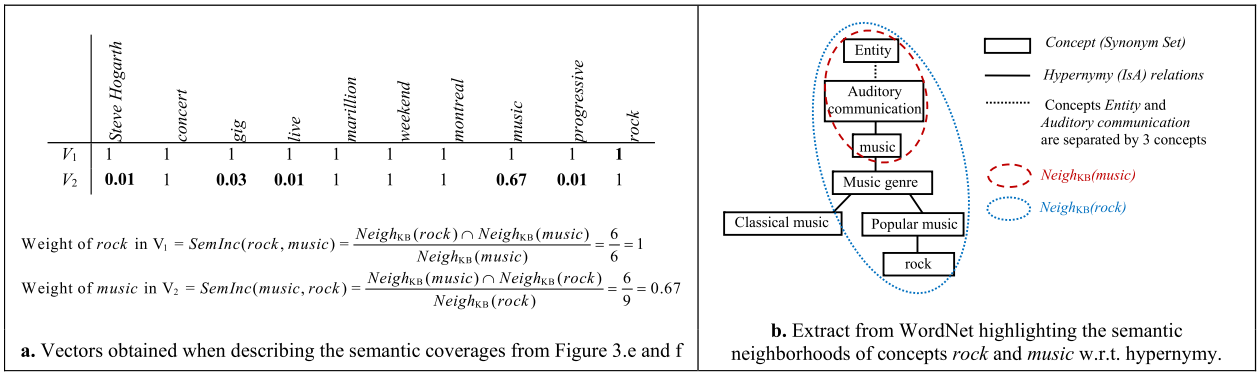
This paper introduces SEDDaL, a framework for Social Event Detection, Description and Linkage from different social media sources. It takes as input: a collection of social media objects from heterogeneous sources, and then produces as output a knowledge graph consisting of a collection of semantically meaningful events interconnected with meaningful relationships, forming the seed of so-called event-based collective knowledge (CK). SEDDaL consists of four modules for: (i) describing social media objects in a generic Metadata Representation Space Model (MRSM) consisting of three composite dimensions: temporal (*When*), spatial (*Where*), and semantic (*What*), (ii) evaluating the similarity between social media object descriptions following MRSM, (iii) detecting events from similar objects using an adapted unsupervised learning algorithm, where events are represented as clusters of objects described in MRSM, and (iv) identifying directional, metric, and topological relationships between events following MRSM's dimensions. This is the first study to provide a generic model for detecting and describing semantic-aware social events and identifying their different relationships. Experimental results highlight the quality and potential of our solution.

We are currently conducting additional tests to evaluate the scalability and adaptability of our solution when dealing with different kinds of objects (e.g., vector graphics, animations, music annotations, and videos) with different sizes and properties. We are also investigating auto-calibration and optimization techniques, e.g., [67,89,90], allowing to choose the proper unit of measurement and proper parameter values for each dimension of MRSM, considering the properties of the media objects being described, in order to adapt the outcome of the event detection process. Other challenges toward producing full-fledged event-based CK within a ubiquitous computing environment include: (i) investigating prominent spatiotemporal indexing structures, like HHCode, QuadTree, Octree, and GeoHash [99–101], to speed up data representation and access in MRSM and improve overall time performance, (ii) expanding the current MRSM model to include user related information and additional semantics (i.e., *Who*, *Why*, and *How* dimensions), (iii) investigating crowd-sourcing (using Wikipedia, or FOAF [102] for instance) as supplementary metadata sources, (iv) developing dedicated event relationship inference rules (e.g., having $\varepsilon_1 \xrightarrow{after} \varepsilon_2$ and $\varepsilon_2 \xrightarrow{after} \varepsilon_3$ means we can transitively infer $\varepsilon_1 \xrightarrow{after} \varepsilon_3$) to enhance the produced knowledge graph's organization and expressiveness, and (v) using formal description languages (such as RDF [103] and OWL [104]) to represent the event-based knowledge graph for querying and automated reasoning by (human users and) software agents, allowing event-based trust management (to distinguish and overcome "fake" events [105]), object recommendation (based on related events, e.g., recommend *items to buy* since they occur in related *sales* events [106]), and event prediction functionality (infer future events based on current ones [55]).

Acknowledgments

This study is partly funded by the National Council for Scientific Research-Lebanon (CNRS-L), and by the Lebanese American University (LAU).

Please cite this article as: M.A. Abebe, J. Tekli, F. Getahun et al., Generic metadata representation framework for social-based event detection, description, and linkage, Knowledge-Based Systems (2019), https://doi.org/10.1016/j.knosys.2019.06.025.



a. Vectors obtained when describing the semantic coverages from Figure 3.e and f

b. Extract from WordNet highlighting the semantic neighborhoods of concepts *rock* and *music* w.r.t. hypernymy.

Fig. 19. Sample semantic enclosure vectors (a) and semantic neighborhoods (b) following [82,83].

Appendix

We briefly describe the concepts of *semantic neighborhood*, *semantic enclosure*, and *semantic relatedness* from [82,83] originally developed to identify the semantic topological relationships between two RSS feeds. The same concepts can be utilized to identify the topological relationships between the semantic coverages of two events in MRSM:

- Similarly to processing two RSS feeds, the two semantic coverages $S(\varepsilon_1)$ and $S(\varepsilon_2)$ being processed are represented as vectors of concepts, V_1 and V_2, where the vector space dimensions represent each a distinct concept $c_m \in S(\varepsilon_1) \cup S(\varepsilon_2)$,
- The weight of a concept c_m in vector V_i, noted w_m^i is $\in [0,1]$. It is maximum, i.e., $= 1$, when the concept $c_j \in S(\varepsilon_i)$. Otherwise, it is computed as the maximum *semantic enclosure* of c_m within any of the concepts $c_j \in S(\varepsilon_i)$:

$$w_m^i = \begin{cases} 1 & \text{if } c_m \in S(\varepsilon_i) \\ max(SemInc(c_m, S(\varepsilon_i))) & \text{otherwise} \end{cases} \quad (29)$$

- The *semantic enclosure* of c_m within another concept c_j is computed as the asymmetric Jaccard similarity measure between the semantic neighborhoods of c_m and c_j:

$$SemInc(c_m, c_j) = \frac{Neigh_{KB}(c_m) \cap Neigh_{KB}(c_j)}{Neigh_{KB}(c_j)} \quad (30)$$

- The *semantic neighborhood* of concept c_m, $Neigh_{KB}(c_m)$, consists of the set of all concepts related directly or transitively with c_m via the hierarchical *hypernymy* (*IsA*) relationship in the reference knowledge base KB (e.g. WordNet). Sample concept neighborhoods are shown in Fig. 19b.
- The *semantic relatedness* between two semantic coverages $S(\varepsilon_1)$ and $S(\varepsilon_2)$ is evaluated as the cosine similarity of their vector representations V_1 and V_2:

$$SemRel(S(\varepsilon_1), S(\varepsilon_2)) = cosine(V_1, V_2) = \frac{V_1.V_2}{|V_1| \times |V_2|} \quad (31)$$

As a result, the following rules are utilized to identify the semantic *equality*, *intersection*, and *disjointness* relationships between two semantic coverages $S(\varepsilon_1)$ and $S(\varepsilon_2)$:

$$\mathbf{r}_S^{\mathbf{topological}}(\varepsilon_1, \varepsilon_2) = \begin{cases} \varepsilon_1 \xrightarrow{S_{equal}} \varepsilon_2 & \text{if } equal_S(e_1, e_2) \Leftrightarrow SemRel(S(\varepsilon_1), S(\varepsilon_2)) \geq \text{Thresh}_{\text{Equal}} \\ \varepsilon_1 \xrightarrow{S_{include}} \varepsilon_2 & \text{if } include_{\mathbf{S}}(\varepsilon_1, \varepsilon_2) \Leftrightarrow \prod_{c_m \in V_1} w_m^1 = 1 \\ \varepsilon_1 \xrightarrow{S_{intersect}} \varepsilon_2 & \text{if } intersect_{\mathbf{S}}(\varepsilon_1, \varepsilon_2) \Leftrightarrow \text{Thresh}_{\text{Disjoint}} \leq SemRel(S(\varepsilon_1), S(\varepsilon_2)) < \text{Thresh}_{\text{Equal}} \\ \varepsilon_1 \xrightarrow{S_{disjoint}} \varepsilon_2 & \text{otherwise} \end{cases} \quad (32)$$

While the equality, *intersection*, and *disjointness* relationships can be defined using *semantic relatedness* thresholds (cf. Fig. 18), this is not the case for *inclusion* relation, which is evaluated as the product of the semantic enclosure vectors of $S(\varepsilon_1)$ and $S(\varepsilon_2)$, designating whether $S(\varepsilon_1)$'s meaning is semantically included in $S(\varepsilon_2)$ or not.

Consider for instance two semantic coverages extracted from Fig. 6e and f, which we designate as $S(\varepsilon_1) = \{$*Steve Hogarth*, *concert*, *gig*, *live*, *marillion*, *weekend*, *montreal*, *music*, *progressive*$\}$ and $S(\varepsilon_1) = \{$*marillion*, *concert*, *rock*, *weekend*, *montreal*$\}$ respectively, where ε_1 and ε_2 represent two hypothetical events.[31] Corresponding vector representations V_1 and V_2 are shown in Fig. 19a. Here, one can realize that $\varepsilon_1 \xrightarrow{S_{include}} \varepsilon_2$ since the product of all weights in vector V_1 considering all dimensions from $S(\varepsilon_1)$ and $S(\varepsilon_2)$, i.e., $\prod_{c_m \in V_1} w_m^1 = 1 = 1$. In other words, the semantic meaning of ε_2 is included in (or subsumed by) ε_1.

The reader can refer to [82,83] for a more detailed description of the approach.

References

[1] R. Chbeir, V. Oria, Editorial preface: Special issue on multimedia data annotation and retrieval using web 2.0, Multimedia Tools Appl. 64 (1) (2013) 1–5.
[2] J. Tekli, A. Abou Rjeily, R. Chbeir, G. Tekli, P. Houngue, K. Yetongnon, M. Ashagrie Abebe, Semantic to intelligent web era: building blocks, applications, and current trends, in: Inter. Conf. on Managment of Emergent Digital EcoSystems (MEDES), 2013, pp. 159–168.
[3] Z. Ma, et al., Knowledge adaptation for ad hoc multimedia event detection with few exemplars, in: ACM Inter. Conf. on Multimedia, 2012, pp. 469–478.
[4] S. Smriti, K. Rajesh, B. Pawan, G. Sumita, News event extraction using 5W1H approach & its analysis, Int. J. Res. Eng. Technol. (IJRET) (4) (2013) 5.

[31] We consider hypothetical events here to simplify the computation example.

[5] U. Jain, R. Westermann, Toward a common event model for multimedia applications, IEEE Multimedia (2007) 19–29.
[6] D. Manchon-Vizuete, X. Giro-i Nieto, UPC At MediaEval 2013 social event detection task, in: MediaEval'13 Multimedia Benchmark Workshop, 2013, p. 1045.
[7] F. Psallidas, et al., Effective event identification in social media, IEEE Data Eng. Bull. 36 (3) (2013) 42–50.
[8] T. Sutanto, R. Nayak, Admrg@ MediaEval 2013 social event detection, in: MediaEval 2013 Multimedia Benchmark Workshop, 2013, p. 1043.
[9] H. Becker, M. Naaman, L. Gravano, Learning similarity metrics for event identification in social media, in: Inter. Conf. on Web Search & Data Mining (WSDM), 2010, pp. 291–300.
[10] C. Ling, R. Abhishek, Event detection from flickr data through wavelet-based spatial analysis, in: ACM Conf. on Info. & Knowledge Management (CIKM'09), 2009, pp. 523–532.
[11] M. Chen, D. Ebert, H. Hagen, R.S. Laramee, R. Van Liere, K.-L. Ma, W. Ribarsky, G. Scheuer, Data, information and knowledge in visualization, IEEE Comput. Graph. Appl. 29 (1) (2009) 12–19.
[12] M. Jäger, S. Nadschläger, T.N. Phan, J. Küng, Data, information & knowledge sources in the agricultural domain, in: International Conference on Big Data Analytics and Knowledge Discovery (DEXA'15) Workshops, 2015, pp. 115–119.
[13] C. Zins, Conceptual approaches for defining data, information, and knowledge, J. Amer. Soc. Inf. Sci. Tech. 58 (4) (2007) 479–493.
[14] K. Jeong-Dong, S. Jiseong, B. Doo-Kwon, Onto: Ontological context-aware model based on 5w1h, Int. J. Distrib. Sensor Netw. (2012) 11.
[15] C.C. Karaman, S. Yaliman, S.A. Oto, Event detection from social media: 5W1H analysis on big data, in: Signal Proc. & Comm. Apps. (SIU'17) Conf, 2017, pp. 1–4.
[16] Z. Ma, Y. Yang, Y. Cai, N. Sebe, A.G. Hauptmann, Knowledge adaptation for ad-hoc multimedia event detection with few exemplars, in: ACM Inter. Conf. on Multimedia, 2012, pp. 469–478.
[17] V.D. Nguyen, N.T. Nguyen, H.B. Truong, A preliminary analysis of the influence of the inconsistency degree on the quality of collective knowledge, Cybernet. Syst. 47 (1–2) (2016) 69–87.
[18] M. Cochez, S. Decker, E. Prud'hommeaux, Knowledge representation on the web revisited: The Case for prototypes, Int. Semant. Web Conf. (ICSW) (1) (2016) 151–166.
[19] L. Xie, H. Sundaram, M. Campbell, Event mining in multimedia streams, Proc. IEEE 96 (4) (2008) 623–647.
[20] J. Tekli, R. Chbeir, A. Traina, C. Traina, F. Renato, Approximate XML structure validation based on document-grammar tree similarity, Inf. Sci. 295 (2015) 258–302.
[21] J. Tekli, R. Chbeir, K. Yétongnon, An overview of XML similarity: Background, current trends and future directions, Comput. Sci. Rev. 3 (3) (2009) 151–173.
[22] C.C. Aggarwal, C.K. Reddy, Data Clustering: Algorithms and Applications, CRC Press, ISBN: 978-1-46-655821-2, 2014, p. 49.
[23] A. Algergawy, M. Mesiti, R. Nayak, Gunter Saake, XML Data clustering: An overview, ACM Comput. Surv. 43 (4) (2011) 25.
[24] D. Rafailidis, et al., A data-driven approach for social event detection, in: MediaEval'13 Multimedia Benchmark Workshop, 2013, p. 1043.
[25] J. Platt, M. Czerwinski, B.A. Field, PhotoTOC: Automatic clustering for browsing personal photographs, in: Proceedings of the 2003 Joint Conference of the Fourth International Conference on Information, Communications and Signal Processing, 2003, pp. 6–10.
[26] X. Liu, R. Troncy, B. Huet, Using social media to identify events, in: ACM SIGMM Inter, Workshop on Social Media, 2011, pp. 3–8.
[27] I. Gupta, K. Gautam, K. Chandramouli, Vit@MediaEval 2013 social event detection task: Semantic structuring of complementary information for clustering events, in: Working Notes of the MediaEval 2013 Worksho, Barcelona, Spain, 2013.
[28] D. Manchon-Vizuete, I. Gris-Sarabia, X. Giró i Nieto, UPC At mediaeval 2014 social event detection task, in: MediaEval'14 Multimedia Benchmark Workshop, 2014, p. 1263.
[29] M. Zaharieva, D. Schopfhauser, M. del Fabro, M. Zeppelzauer, Clustering & retrieval of social events in flickr, in: MediaEval'14 Multimedia Benchmark Workshop, 2014, p. 1263.
[30] A. Budanitsky, G. Hirst, Evaluating WordNet-based measures of lexical semantic relatedness, Comput. Linguist. 32 (1) (2006) 13–47.
[31] G.A. Miller, C. Fellbaum, Wordnet then and now, Lang. Resour. Eval. 41 (2) (2007) 209–214.
[32] J. Tekli, An overview on XML semantic disambiguation from unstructured text to semi-structured data: Background, applications, and ongoing challenges, IEEE Trans. Knowl. Data Eng. (IEEE TKDE) 28 (6) (2016) 1383–1407.
[33] S.B. Kotsiantis, Supervised machine learning: A review of classification techniques, Informatica 31 (2007) 249–268.
[34] S. Kotsiantis, I. Zaharakis, P. Pintelas, Machine learning: A review of classification & combining techniques, Artif. Intell. Rev. 26 (2007) 159–190.
[35] H. Becker, F. Chen, D. Iter, M. Naaman, L. Gravano, Automatic identification and presentation of twitter content for planned events, in: Inter. AAAI Conf. on Weblogs & Social Media (ICWSM'11), 2011, p. 3.
[36] X. Liu, et al., Finding media illustrating events, in: ACM Intern. Conf. on Multimedia Retrieval, 2011.
[37] H. Becker, D. Iter, M. Naaman, L. Gravano, Identifying content for planned events across social media sites, in: Inter. Conf. on Web Search & Data Mining (WSDM), ACM, 2012, pp. 533–542.
[38] M. Wistuba, S.T. Lars, Supervised clustering of social media streams, in: Working Notes of the MediaEval 2013 Workshop, 2013.
[39] A. Papaoikonomou, et al., A similarity-based Chinese restaurant process for social event detection, in: Working Notes of the MediaEval 2013 Workshop, 2013, p. 1043.
[40] Z. Su, J. Kogan, C. Nicholas, Constrained clustering with k-means type algorithms, in: Text Mining: Applications and Theory (Wiley Online Library), 2010, pp. 81–103.
[41] T.V. Nguyen, et al., Event clustering and classification from social media: Watershed-based and kernel methods, in: MediaEval'13 Multimedia Benchmark Workshop, 2013.
[42] A. Bordes, E. Gabrilovich, Constructing and mining web-scale knowledge graphs, in: 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'14), ACM, New York, NY, USA, 2014, p. 1967.
[43] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, S. Hellmann, DBpedia – A crystallization point for the web of data, Elsevier J. Web Semant. (JWS) 7 (2009) 154–165.
[44] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, J. Taylor, Freebase: a collaboratively created graph database for structuring human knowledge, in: Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data, ACM, 2008, pp. 1247–1250.
[45] X. Dong, et al., Knowledge vault: a web-scale approach to probabilistic knowledge fusion, in: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'14), 2014, pp. 601–610.
[46] C. Ehrlich, Ontos News Portal. http://www.ontosearch.com/2008/01/identification#, 2010, accessed in 2018.
[47] D. Shahaf, C. Guestrin, Connecting the dots between news articles, in: ACM Inter. Conf. on Knowledge Discovery & Data Mining (KDD'10), 2010, pp. 623–632.
[48] E. Kuzey, J. W.G. Vreeken, A fresh look on knowledge bases: Distilling named events from news, in: ACM Inter. Conf. on Information and Knowledge Management, (CIKM'14), ACM, Shanghai, China, 2014, pp. 1689–1698.
[49] H. Llorens, E. Saquete, B. Navarro, Tipsem (English and Spanish): Evaluating crfs and semantic roles in Tempeval-2, in: Inter. Workshop on Semantic Evaluation (SemEval'10), ACL, Stroudsburg, PA, USA, 2010.
[50] J. Pustejovsky, et al., Timeml: Robust specification of event and temporal expressions in text, in: Inter. Workshop on Computational Semantics (IWCS'03), 2003, pp. 1–11.
[51] Y. Cai, Q. Li, H. Xie, T. Wang, Event relationship analysis for temporal event search, in: Inter. Conf. on Database Systems for Advanced Apps. (DASFAA), 2013, pp. 179–193.
[52] Y. Rao, Q. Li, Q. Wu, H. Xie, F.L. Wang, T. Wang, A multi-relational term scheme for first story detection, Neurocomputing 254 (2017) 42–52.
[53] L. Gregor, B. Fortuna, J. Brank, M. Grobelnik, Event registry: Learning about world events from news, in: 23rd International WWW Conference, 2014, pp. 107–110.
[54] M. Rospocher, et al., Building event-centric knowledge graphs from news, J. Web Semant. (ISSN: 1570-8268) 37–38 (2016) 132–151.
[55] K. Horvitz, E. Radinsky, Mining the web to predict future events, in: ACM Inter. Conf. on Web Search and Data Mining (WSDM'13), ACM, 2013, pp. 255–264.
[56] B. Braden, The surveyor's area formula, College Math. J. 17 (4) (1986) 326–337.
[57] T.M. Chan, Optimal output-sensitive convex hull algorithms in two and three dimensions, Discrete Comput. Geom. 16 (1996) 361–368.
[58] R. Rhoad, G. Milauskas, R. Whipple, Geometry for Enjoyment and Challenge, new ed., McDougal Littell, ISBN: 0-86609-965-4, 1991, p. 784.
[59] J. Hoffart, F.M. Suchanek, K. Berberich, G. Weikum, YAGO2: A spatially and temporally enhanced knowledge base from wikipedia, Artificial Intelligence 194 (2013) 28–61.
[60] Saruladha K., G. Aghila, S. Raj, A survey of semantic similarity methods for ontology based information retrieval, in: Inter. Conf. on ML & Comput. (ICMLC), 2010, pp. 297–301.
[61] D'Ulizia A., Ferri F., P. Formica A. Grifoni, M. Rafanelli, Structural similarity in geographical queries to improve query answering, in: Proceedings of the 2007 ACM Symposium on Applied Computing (SAC'07), 2007, pp. 19–23.
[62] A. Formica, Ontology-based concept similarity in Formal Concept Analysis, Inform. Sci. 176 (18) (2006) 2624–2641.

Please cite this article as: M.A. Abebe, J. Tekli, F. Getahun et al., Generic metadata representation framework for social-based event detection, description, and linkage, Knowledge-Based Systems (2019), https://doi.org/10.1016/j.knosys.2019.06.025.

[63] J. Tekli, N. Charbel, R. Chbeir, Building semantic trees from XML documents, Elsevier J. Web Semant. (JWS) 37–38 (2016) 1–24.
[64] J. Tekli, R. Chbeir, K. Yetongnon, A novel XML structure comparison framework based on sub-tree commonalities and label semantics, Elsevier J. Web Semant. (JWS): Sci. Serv. Agents World Wide Web 11 (2012) 14–40.
[65] A. Algergawy, R. Nayak, G. Saake, Element similarity measures in XML schema matching, Elsevier Inf. Sci. 180 (24) (2010) 4975–4998.
[66] J. Tekli, R. Chbeir, K. Yétongnon, Minimizing user effort in XML grammar matching, Inf. Sci. J. 210 (2012) 1–40.
[67] M. Ming, P. Yefei, S. Michael, A harmony based adaptive ontology mapping approach, in: Inter. Conf. on Semantic Web & Web Services (SWWS), 2008, pp. 336–342.
[68] P. Shvaiko, J. Euzenat, Ontology matching: State of the art and future challenges, IEEE Trans. Knowl. Data Eng. 25 (1) (2013) 158–176.
[69] T. Reuter, et al., Social event detection at MediaEval'13: Challenges, datasets, and evaluations, in: MediaEval'13 Multimedia Benchmark Workshop, 2013, p. 1043.
[70] P.A. Bernstein, J. Madhavan, E. Rahm, Generic schema matching, ten years later, Publ.Very Large Database Endowment (PVLDB) 4 (11) (2011) 695–701.
[71] S. Maßmann, S. Raunich, D. Aumüller, P. Arnold, E. Rahm, Evolution of the COMA match system, in: Inter. Conf. on Ontology Matching (OM'11), Vol. 814, 2011, pp. 49–60.
[72] R. Navigli, Word sense disambiguation: a survey, ACM Comput. Surv. 41 (2) (2009) 1–69.
[73] N.R. Chopde, M. Nichat, Landmark-based shortest path detection by using A* and haversine formula, Int. J. Innov. Res. Comput. Comm. Eng. 1 (2) (2013) 298–302.
[74] Z. Wu, M. Palmer, Verb semantics and lexical selection, in: 32nd Annual Meeting of the Associations of Computational Linguistics (ACL), 1994, pp. 133–138.
[75] D. Lin, An information-theoretic definition of similarity, in: Inter. Conf. on Machine Learning (ICML), Morgan Kaufmann Pub. Inc, 1998, pp. 296–304.
[76] S. Banerjee, T. Pedersen, Extended gloss overlaps as a measure of semantic relatedness, in: Inter. Joint Conf. on Artificial Intelligence (IJCAI'03), 2003, pp. 805–810.
[77] N. Charbel, J. Tekli, R. Chbeir, G. Tekli, Resolving XML semantic ambiguity, in: Inter. Conf. on Extending Database Technology (EDBT'15), 2015, pp. 277–288.
[78] M. Ehrig, Ontology alignment: Bridging the semantic gap, in: Semantic Web and beyond: Computing for Human Experience, Springer, ISBN: 978-0-387-36501-5, 2007.
[79] A. Tversky, Features of similarity, Psychol. Rev. 84 (4) (1977) 327–352.
[80] P. Sneath, R. Sokal, Numerical taxonomy: The principles and practice of numerical classification, Q. Rev. Biol. 50 (4) (1975) 525–526.
[81] G. Milligan, M. Cooper, An examination of procedures for determining the number of clusters in a dataset, Psychometrika 50 (52) (1985) 159–179.
[82] F.G. Taddesse, J. Tekli, R. Chbeir, M. Viviani, K. Yetongnon, Semantic-based merging of RSS items, World Wide Web J. 12 (11280) (2010).
[83] F.G. Taddesse, J. Tekli, R. Chbeir, M. Viviani, K. Yétongnon, Relating RSS News/Items, in: Inter. Conf. on Web Engineering (ICWE'09), 2009, pp. 44–452.
[84] P. Georgios, S. Papadopoulos, V. Mezaris, Y. Kompatsiaris, Social event detection at MediaEval 2014: Challenges, datasets, and evaluation, in: Proceedings of the MediaEval'14 Multimedia Benchmark Workshop, 2014, Available at: http://www.multimediaeval.org/mediaeval2014/sed2014/.
[85] S. Zhong, J. Ghosh, Generative model-based document clustering: a comparative study, Knowl. Inf. Syst. 8 (3) (2005) 374–384.
[86] A. Ghosh, J. Strehl, Cluster ensembles—a knowledge reuse framework for combining multiple partitions, J. Mach. Learn. Res. 3 (2003) 583–617.
[87] C.D. Manning, P. Raghavan, H. Schütze, Introduction to Information Retrieval, Cambridge University Press, 2008, Ch. 1 Boolean Retrieval - A First Take at Building an Inverted Index, https://nlp.stanford.edu/IR-book/.
[88] R. Baeza-Yates, B. Ribeiro-Neto, Modern Information Retrieval: The Concepts and Technology behind Search, second ed., ACM Press Books, 2011, p. 944.
[89] A. Gal, H. Roitman, T. Sagi, From diversity-based prediction to better ontology & schema matching, in: Inter. WWW Conference, 2016, pp. 1145–1155.
[90] J. Hopfield, D. Tank, Neural computation of decisions in optimization problems, Biol. Cybernet. 52 (3) (1985) 52–141.
[91] T. Sutanto, R. Nayak, Fine-grained document clustering via ranking and its application to social media analytics, Soc. Net. Anal. Min. 8 (1) (2018) 29:1-29:19.
[92] X. Guo, K. Nguyen, S. Denman, C. Fookes, S. Sridharan, Single image depth prediction using super-column super-pixel features, in: Inter. ICIP Conf, 2017, pp. 2657–2661.
[93] M. Zaharieva, M. del Fabro, M. Zeppelzauer, Cross-platform social event detection, IEEE MultiMedia 22 (3) (2015) 14–25.
[94] F. Vasilescu, P. Langlais, G. Lapalme, Evaluating variants of the lesk approach for disambiguating words, in: Language Resources & Eval. (LREC'04), 2004, pp. 633–636.
[95] R. Navigli, P. Velardi, Structural semantic interconnections: A knowledge-based approach to word sense disambiguation, IEEE Trans. Pattern Anal. Mach. Intell. 27 (7) (2005) 1075–1086.
[96] E. Agirre, S. Aitor, Personalizing pagerank for word sense disambiguation, in: 12th Conf. of the European Chapter of the ACL, 2009, pp. 33–41.
[97] C. Agirre E. Baneab, C. Cardiec, D. Cerd, M. Diabe, A. Gonzalez-Agirrea, W. Guof, R. Mihalceab, G. Rigaua, J. Wiebe, Semeval-2014 task 10: Multilingual semantic textual similarity, in: 8th International Workshop on Semantic Evaluation (SemEval 2014), Vol. 8, 2014, pp. 1–99.
[98] A. Gonzalez-Agirre, N. Aletras, G. Rigau, M. Stevenson, A. E, Why are these similar? Investigating item similarity types in a large Digital Library, J. Assoc. Inf. Sci. Technol. (JASIST) (ISSN: 2330-1643) 67 (2015) 1624–1638.
[99] S. Aluru, Quadtrees and octrees, in: D. Mehta, S. Sahni (Eds.), HandBook of Data Structures and Applications, Vol. 19, Chapman and Hall/CRC, 2004, pp. 1–26.
[100] A. Fox, E.C, J. Hughes, S. Lyon, Spatio-temporal indexing in non-relational distributed databases, IEEE Int. Conf. Big Data (2013) 1–9.
[101] S. Har-Peled, Quadtrees - hierarchical grids, geometric approximation algorithms, Math. Surveys Monogr. (2011) 173.
[102] B. Aleman-Meza, M. Nagarajan, L. Ding, A.P. Sheth, I.B. Arpinar, A. Joshi, T.W. Finin, Scalable semantic analytics on social networks for addressing the problem of conflict of interest detection, ACM Trans. Web (TWeb) 2 (1) (2008) 7.
[103] W3C, RDF 1.1 XML Syntax. W3C Recommendation 2014. http://www.w3.org/TR/rdf-syntax-grammar/.
[104] D.L. McGuinness, F. Van Harmelen, OWL 2 Web - Ontology Language Document Overview. W3C Proposed Edited Recomm. 2012. http://www.w3.org/TR/owl2-overview/.
[105] N. Nguyen, V. Nguyen, D. Hwang, An influence analysis of the number of members on the quality of knowledge in a collective, Intell. Fuzzy Syst. 32 (2) (2017) 1217–1228.
[106] D.T. Hoang, V.C. Tran, V.D. Nguyen, V.T. Nguyen, D. Hwang, Improving academic event recommendation using research similarity and interaction strength between authors, Cybernet. Syst. 48 (3) (2017) 210–230.

Please cite this article as: M.A. Abebe, J. Tekli, F. Getahun et al., Generic metadata representation framework for social-based event detection, description, and linkage, Knowledge-Based Systems (2019), https://doi.org/10.1016/j.knosys.2019.06.025.