A General Multimedia Representation Space Model Toward Event-based Collective Knowledge Management

Minale A. Abebe
Addis Ababa University
Addis Ababa, Ethiopia
minale.ashagrie@aau.edu.et

Joe Tekli
Lebanese American University
36 Byblos, Lebanon
joe.tekli@lau.edu.lb

Fekade Getahun
Addis Ababa University
Addis Ababa, Ethiopia
fekade.getahun@aau.edu.et

Gilbert Tekli
Nobatek Research and Development
64600 Anglet, France
gilbert.tekli@nobatek.fr

Richard Chbeir
University of Pau and Adour Countries
64600 Anglet, France
richard.chbeir@univ-pau.fr

***Abstract*—Emergent technologies such as smart phones and wireless Internet have transformed the Web from a static data publishing platform into a collaborative information sharing environment. Yet, attaining the next stage in Web engineering, i.e., the so-called Intelligent Web: allowing meaningful human-machine and machine-machine collaboration, requires another breakthrough: allowing the sharing and organization of collective knowledge (CK), where CK underlines the combination of all known data, information, and meta-data concerning a given concept or event. In this context, various methods have been put forward to perform automatic event extraction and description. Yet, most of them do not capture the semantic meaning embedded in Web-based multimedia data, which are usually highly heterogeneous and unstructured. To address this problem, we introduce in this study a generic Multimedia Representation Space Model called MRSM, designed for multimedia data and multimedia-based event representation, in order to allow event detection and identification based on multimedia CK. We formally define MRSM, its dimensions, their coordinates, and the associated distance (similarity) metrics and properties. We then provide the building blocks for an Event-based Collective Knowledge (CK) Management Framework, built upon MRSM, and geared toward effective CK management. The proposed approach provides a means of extracting, representing, and linking events from heterogeneous multimedia data without any prior knowledge about event-related clues. Preliminary tests confirm the quality and potential of our approach.**

***Keywords*—Multimedia, Metadata, Collective Knowledge, Knowledge Management, Event Detection and Identification.**

I. INTRODUCTION

Nowadays, emerging technologies such as Smart-phones, Wireless Internet, as well as Web and Mobile Services allow users to create, annotate, and share multimedia 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 produces and proactive sensors of information during the 2010s (i.e., producing blogs, publishing and annotating images and videos, commenting on tweets, posting opinions, etc.), transforming the Web from a static data publishing platform into a collaborative information sharing environment [1-4]. Here we distinguish between *data* (containing no meaning, e.g., *"1998"* is considered as a number consisting of 4 digits) and *information* (data having a certain meaning in a certain context, e.g., for a Web expert, *"1998"* is the *year of creation of the XML W3C standard*). Nonetheless, attaining the next stage in Web development and engineering, i.e., the so-called Intelligent Web: allowing meaningful human-machine and machine-machine collaboration, requires yet another breakthrough: allowing the sharing and organization of *collective knowledge* (CK) [5]. Here, *knowledge* represents a higher level of data abstraction: as the combination of all known data, information, and meta-data concerning a given concept, fact, or event, as well as the semantic links between them [6, 7] (e.g., knowing that *"the year of creation of the XML standard"* is *"1998"*, following *Wikipedia* in an article *published in 2000*). In this context, CK can be viewed as the development of knowledge assets or (semantic) information from a distributed pool of contributions [8]. Hence, intelligent terminals (software agents) connected to the Web are expected to automatically analyze and handle large collections of multimedia data with their contents, links and transactions, using the sum of their respective intelligence and knowledge, in order to improve individual and collaborative information management (indexing, storage, exchange, search, and retrieval) [5, 9].

However, realizing the Intelligent Web vision faces many difficulties [5]. According to [10], more than 80% of the data shared in the Web is heterogeneous, streamed, unstructured, massive, multimedia, and are inherently associated to so-called events. An *event* can be defined 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, etc.) [11]. Usually participants of an event capture multimedia data (image, video, audio, etc), annotate, publish and share them to describe the event (e.g., videos from the soccer match, pictures of the storm, opinions about presidential debate, etc.) [12]. However, annotations of similar multimedia data 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 multimedia data descriptions to identify meaningful events, needed as the building blocks for CK organization remains a major challenge.

Addressing the above challenge requires a central stepping stone: producing a generic multimedia data representation model which is machine-readable, openly accessible, and linked to existing knowledge and other datasets, which can be easily used for multimedia object representation and description, event extraction, and intelligent reasoning later on. Here, evaluating the spatial coverage, temporal coverage, and most importantly the semantic meaning of shared multimedia objects become of key importance in performing event detection and identification [13].

In this context, various methods have been put forward to perform automatic event extraction [14-17]. Yet, collecting and organizing shared multimedia data related to a specific event of interest remain a difficult task for existing approaches, due to: i) the complex and heterogeneous nature of multimedia data from different sources [13, 15], ii) using low-level multimedia data descriptors alone, which are prone to noise, rather than handling multimedia semantics (visually similar images might describe totally different things, e.g., an image of a *blue sky* versus an image of a *blue sea*) [14, 16]. In other words, most existing approaches do not capture the semantic meaning embedded in the multimedia objects (e.g., image of the *Mediterranean Sea near the shores of Alexandria* versus image of the *sky above the Atlantic shore of Florida*).

Hence, a new approach that effectively represents multimedia objects, in order to detect and identify meaningful events, considering the heterogeneous and noisy nature of the data is needed. For this purpose, we introduce in our current study a Multimedia Representation Space Model (MRSM), allowing to prepare the stage: from multimedia data representation toward event detection and identification. We formally define MRSM, its dimensions, their coordinates, and associated distance (similarity) metrics and properties. We then provide the building blocks for an Event-based Collective Knowledge (CK) Management Framework, built upon MRSM, and geared toward effective CK management (i.e., recommendation, prediction, versioning, and decision making processes later on). The proposed approach provides a means of extracting, representing, and linking events from heterogeneous multimedia data without any prior knowledge about event-related clues. Preliminary tests confirm the potential of our approach.

The rest of the paper is structured as follows. Section 2 presents a motivating scenario that highlights some of the challenges related to our work. Section 3 briefly reviews related works. Section 4 presents our multimedia representation space model, its properties, and the associated event-based CK management framework. Experimental setup and the results are discussed in Section 5. Finally, we conclude and highlight future research direction in Section 6.

II. MOTIVATION

In this section, we present a motivational scenario that illustrates the need for event-based CK management.

Climate change due to global warming increases the probability of occurrence 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. The flood may cause disruptions of basic utility services such as transportation, electricity, tap water supply, and telephone lines. When such an event occurs in a city nowadays, residents often capture different kinds of multimedia objects (e.g., images, videos, sounds, etc.), annotate, publish, and share them on social media sites like Facebook[1], Flickr[2], and YouTube[3] timelines. They might also post comments on tweeter[4] to share their appreciation and/or criticism regarding the level of preparedness and action taken by the city administration to handle the observed phenomena. Moreover, local media providers may be continually publishing news feeds related to the occurrence of the event.

In order to provide better services to its residents, the city's administration would largely benefit from organizing and processing the CK associated to events that are occurring, in order to make adequate decisions and take reactive/precautionary measures accordingly.

To build such a CK base, the following main challenges need to be addressed:

- Data sources are heterogeneous and their content is unstructured,
- Existence of large volumes of data and content that are published online and grow continuously,
- Multimedia data annotation is not consistent, and depends on annotators' experience,
- Users may publish content that may not be related to events, and thus may be misleading,
- Some of the shared multimedia data may have missing location, date/time, or annotation information.

In this study, we provide an architecture to address the above challenges: introducing a generic MRSM for effective multimedia data description dedicated for event detection and representation.

III. RELATED WORKS

Recently, a number of research works have been conducted targeting knowledge management in open linked data and event detection from social media sources, e.g., [18-23], including a focused European project [24], which we briefly review and discuss in this section.

The open data initiative has availed opportunities for researchers and support innovation. In developing countries such as Kenya, open data initiatives are showing encouraging impact in different areas [25]. Universities in the UK and Australia have also benefited from open research data [19, 26], which are becoming a common trend in the academic arena, allowing to share research-related CK. Here, research data designates not only publications and linked researcher profiles, but also refers to the recorded factual material commonly accepted in the scientific community as necessary to validate research findings. It includes facts, observations, images, computer program results, recordings, measurements, and/or experiences on which an argument, theory, test, or hypothesis, or another research output is based. It can also be data which is collected, observed, or created in a digital form, for the purpose of analyzing and disseminating original research findings. Experience gained from early adopters of open research data shows that open access to such data promotes high quality research (e.g., since researchers are constantly aware of their colleagues' finds and can build on them accordingly), reduce redundant efforts deployed by researchers (e.g., no need to perform the same experiment following the same test protocol redundantly by different researchers), protecting against research fraud, and helping to develop a culture of transparency and sharing of knowledge [18-21].

On the other hand, there is a number of initiatives to detect events from social media data. The authors in [15] used the New York Times news corpus of twenty-two years of coverage, to extract a storyline using a dedicated topic tracking and detection algorithm. They clustered similar text together to enrich the storyline with information extracted from Web knowledge source using the Linked Data platform [22] in order to construct a

[1] www.facebook.com
[2] www.flickr.com
[3] www.youtube.com
[4] www.twitter.com

dedicated predictive model. This central approach shifted the level of analysis from targeting specific entities/topics, to considering broader classes of observations and events. The work made real-time predictions about the likelihoods of future human and natural events of interest. The authors assumed that real-world events were generated following a probabilistic model, and then identified a target predicting event in a certain domain. However, their approach only worked for structured news articles, and did not address unstructured and heterogeneous news contents from distributed sources.

Becker *et al.* [14] 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 and time information are considered for the purpose of detecting events. However, we argue that time and geo-location information might not be enough to effectively detect events, since: i) some social media authoring tools lack location recording components, and ii) the time stamp value of social media contents might be distorted or noisy due to the particular configurations of media capturing tools. Note that the work in [14] focuses on generating events based on predefined preferences stated in advance in existing event aggregation platforms (e.g., anticipated soccer match, or awaited heat wave, which are expected to occur at certain dates/locations, etc.), and does not identify instantaneous/unknown events such as an unexpected flood, earthquake, car accident, or thunderstorm.

Differently from [14], Psallidas *et al*. [13] address the challenge of automatically identifying unknown event contents, considering the high rate of the social streams and the noisy nature of Web data. They propose an online clustering framework that leverages different features associated with each social media document, mainly focusing on Twitter posts (e.g., using *publisher*, publication *time*, and the published *text* features to describe a tweet). Yet, the authors did not discuss how to overcome the often noisy contents of the twitter messages. Also, they solely consider twitter messages in event extraction, without considering any outside knowledge (such as tags or multimedia content from other social media sources for instance).

Another method considering multiple features in identifying events is developed in [23], where the authors attempt to address the problem of structuring social media activities into events. The authors utilize different properties (such as *location*, *time* and user *participation*) from social media sites, based on the assumption that: an event happening in 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 approach in [23] requires a minimal number of photos and videos, above a certain manually defined threshold, in order to perform event detection and identification. In other words, if the number of photos/videos taken by users is less than the threshold, they are disregarded and will not be considered as event representatives.

To summarize, existing event extraction methods in the literature are either: i) domain dependent and consider certain specific kinds of information (e.g., structured news article), e.g., [14], ii) generate events based on predefined clues (and are not able to identify unknown events), e.g., [13], or iii) consider manually defined thresholds which affect event detection coverage (missing certain events) and thus quality, e.g., [23]. In addition, most existing methods, to our knowledge, do not consider the semantic meaning associated to multimedia data and solely focus on time, space, and/or syntactic textual descriptions. Even though the open linked data initiative supports shared knowledge within the research community, e.g., [18-21], yet it generally considers fairly homogeneous data (e.g., publications, books, reports), properly defined and generated by expert (scientific) sources, in contrast with the heterogeneous nature of multimedia data published on the Web, which is often coined with incomplete or noisy descriptions generated by non-experts. Interested readers can refer to [27] for a detailed survey on event detection, identification, and mining techniques.

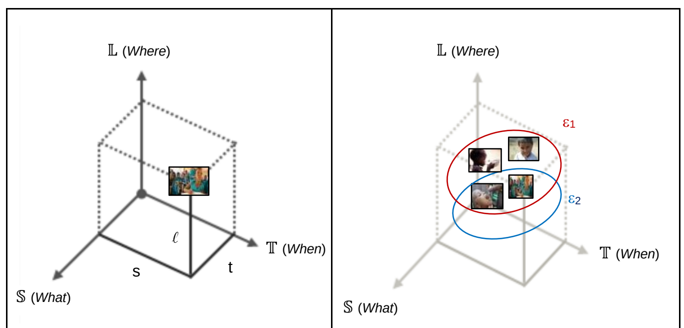
In this paper, we propose a new framework to fill some of the gaps highlighted in existing works, providing a generic model to deal with i) multimedia data heterogeneity and ii) semantic meaning, in performing event extraction.

IV. PROPOSED FRAMEWORK

Our Event-based CK Management Framework consists of two main components: i) our Multimedia Representation Space Model (MRSM), and the ii) Event Extraction process built upon MSRM. In the following we first present each of the aforementioned component in Sections 4.1 and 4.2 respectively, before describing the overall framework and data/control flow in Section 4.3.

A. Multimedia Representation Space Model

Event definitions are theoretically described using the 5W1H model: *When*, *Where*, *What*, *Who*, *Why* and *How* aspects [28-30]. Yet, as described in Section 3, only few of these features are practically covered in existing methods, mainly: *When* (time) and *Where* (location) [13-15]. In our work, we consider an additional feature: the *What* (meaning) of the event, thus covering the temporal, spatial and semantic facets (remaining *Who*, *Why*, and *How* facets will be covered in a subsequent dedicated study). To do so, we define our MSRM as a hyperspace consisting of three composite dimensions: temporal, spatial and semantic, describing each and every multimedia object (as shown in Figure 1.a). Consequently, an event can be represented in the same space, consisting of the collection of multimedia objects describing it (cf. Figure 1.b). In this subsection, we formally describe each dimension, its coverage, and related properties, allowing to describe a multimedia object, and then an event in our MSRM.



a. Describing a multimedia object | **b.** Representing two events e_1 and e_2

Figure 1. Multimedia Representation Space Model (MRSM).

1) Temporal Dimension

Definition 1: [Temporal Dimension ($\mathbb{T}$)]. We define the temporal dimension $\mathbb{T}$ as a finite sequence of discrete and ordered primitive temporal units used to define and interpret a multimedia object's temporal feature values, formally:

$$\mathbb{T} = \{t_0, t_1, t_2, \ldots\} \quad \textbf{(1)}$$

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 $\mathbb{T}$ •

Definition 3: [Temporal Coverage (T)]. It is an ordered collection of time stamps enclosed within a start stamp and an end stamp, describing the temporal coverage of a multimedia object and/or event. We use it to represent the duration or capture of a multimedia object (e.g., a video), or the duration of an event (e.g., duration of a storm). Formally:

$$T = \{t_i \in \mathbb{T} \mid t_i \geq t_s \wedge t_i \leq t_e\} \quad \textbf{(2)}$$

where t_s is the start time stamp of T, and t_e its end time 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:

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

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

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

2) *Spatial Dimension*

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

$$\mathbb{L} = \langle Ø, \lambda, h \rangle \quad \textbf{(4)}$$

where Ø represents the latitude, λ the longitude, and *h* the altitude sub-dimensions (cf. Figure 2.a)•

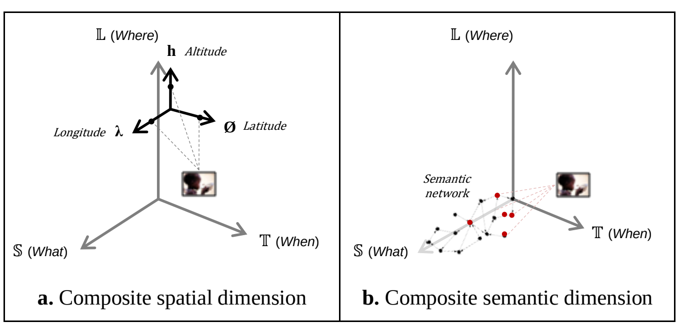


Figure 2. Temporal and semantic dimensions in our MRSM.

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 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 (as in kilometers or miles). In our study, we adopt IS's meter unit (*m*) as the default unit of spatial measure. It can be used trait forwardly with the altitude sub-dimension (h), and is converted to the DMS scale (Degrees, Minutes, and Seconds) or Radians with the latitude (Ø) and longitude (λ), 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 discrete and instantaneous value of the composite spatial dimension $\mathbb{L}$, consisting of a triplet:

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

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

Definition 7: [Spatial Coverage (L)]. It is the set of spatial stamps (surface coverage bounded by local lower and upper spatial stamps) in the composite spatial dimension $\mathbb{L}$, in which a multimedia object was created (e.g., area in which a video stream was recorded) and/or event occurred (e.g., area affected by a storm). Formally, given the composite nature of $\mathbb{L}$, we define L as:

$$L = \int_x \int_y \int_z \phi(x)\, \lambda(y)\, h(z)\, dx\, dy\, dz \quad \textbf{(6)}$$

where Ø(x) is latitude function, λ(y) is longitude function, and h(z) is altitude function •

Definition 8: [Spatial Coverage Representative Point (l_c)**].** It is the midpoint (center of gravity) of a spatial coverage L, formally:

$$\ell_c = \frac{\int_x \int_y \int_z x\, y\, z\, \phi(x)\, \lambda(y)\, h(z)\, dx\, dy\, dz}{\int_x \int_y \int_z \phi(x)\, \lambda(y)\, h(z)\, dx\, dy\, dz} \quad \textbf{(7)}$$

3) *Semantic Dimension*

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

Definition 9: [Semantic Dimension ($\mathbb{S}$)]. It is a set of semantic network units such as *concepts* from ontologies or knowledge bases such as WordNet [31] or Yago [32] linked with semantic relationships (semantic network edges, e.g., *IsA*, *PartOf*, etc.), and ordered following semantic network edge hierarchy (ancestor/child), to define semantic meaning (reflected by concept synonyms, e.g., *car*, *auto*, and *automobile* are all synonyms of concept *car* in WordNet [31]; and/or has the same gloss description, e.g., *a motor vehicle with four wheels*). It can be formalized as a labeled directed graph $\mathbb{S}$=(N, E), where N is the set of concepts (nodes) and E is the set of semantic relationships (edges) •

The unit of the semantic dimension can be a 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 *sport 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 single semantic unit (varying semantic units as groups of concepts to modify the semantic dimension's granularity will be considered in a dedicated study). 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 (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 multimedia object and/or an event. It can be defined as a sub-graph of the semantic dimension $\mathbb{S}$, noted S = ($\mathbb{N}$, $\mathbb{E}$), where $\mathbb{N} \subseteq$ N (set of concepts, i.e., nodes) and $\mathbb{E} \subseteq$ E (semantic relations, i.e., edges, cf. Figure 2.b) •

Definition 12: [Semantic Coverage Representative Point (s_c)**].** It a single concept (semantic stamp) that represents a semantic coverage S. It can be defined as the lowest super ordinate (least common ancestor, a.k.a., most specific common ancestor) of concepts included in the semantic coverage S of a multimedia object and/or event •

After defining our MRSM and its dimensions, we can define the data model for describing a multimedia object and an event.

Definition 13: [Multimedia Object o]. A multimedia object o (e.g., video, image, chart, tweet, or Wiki article) in a given social media environment is defined, following MRSM, as a quadruplet:

$$o = (oid, t_c, l_c, s_c) \quad \textbf{(8)}$$

having a unique object id, *oid*, as well as three representative points: temporal t_c, spatial ℓ_c, and semantic s_c, following each of MRSM's dimensions •

Consequently, and following mainstream event extraction approaches (cf. Section 3), an event can be defined as an aggregation or a group of similar multimedia objects:

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

$$\varepsilon = (eid, T, L, S) \quad \textbf{(9)}$$

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

oid	3452155896
t_c	1238964834
l_c	lat="45.51" long="-73.55" UTM= 18T 612643mE 5041241mN
S	Ian Mosley, Mark Kelly, Pete Trewavas, Steve Hogarth, Steve Rothery, concert, gig, live, weekend, music, progressive, marillion

oid	128796702
t_c	1145040959
l_c	lat="47.43" long="-122.29" UTM= 10T 553236mE 5253825mN
S	Stardance, Norwescon Seattle, DoubleTreeHotel, Nikkor, Washington, conference, cosplay, costume, fantasy

oid	128800481
t_c	1145041372
l_c	lat="47.43" long="-122.29" UTM= 10T 553236mE 5253825mN
S	Stardance, Seattle, Double Tree Hotel, Nikkor, Norwescon, Washington, conference, cosplay, costume, fantasy, scifi

oid	3421558753
t_c	1238956813
l_c	NULL
S	marillion, weekend, montreal

oid	129778685
t_c	1145105698
l_c	lat="47.435" long="-122.294" UTM= 10T 553236mE 5253825mN
S	Norwescon , Seattle, DoubleTreeHotel , Nikkor, Washington, conference, convention, cosplay , costume

oid	3443324510
t_c	1238876657
l_c	lat="45.51" long="- 73.557" UTM= 18T 612695mE 5041397mN
S	Steve Hogarth, concert, gig, live, marillion , weekend, montreal , music, progressive

Figure 3. Sample images extracted from the MediaEvalSED 2013 image dataset [17], described using our MRSM.

Consider for instance the sample 6 images in Figure 3 described following our MRSM. The events extracted based on these images are provided in Figure 4, also described following MRSM.

B. Metric Properties of the MRSM

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

Following our MSRM definition, typical Euclidian distance can be utilized to compare time and location representatives of two multimedia objects/events. As for the semantic dimension, semantic distance can be computed as the inverse of any typical semantic similarity measure comparing two (sets of) concepts in a semantic network [33]. Here, semantic similarity measures can be classified as *edge-based* (estimating similarity as the shortest path between concepts) [34], *node-based* (estimating similarity as the maximum amount of information content concepts share in common) [35], and *gloss-based* (estimating similarity based on word overlap between the concept's gloss descriptions) [36]. In our study, we adopt an aggregate semantic similarity introduced in [37, 38] producing similarity scores $\in[0, 1]$, 0 designating minimal (no) similarity and 1 designating maximum (total) similarity:

$$\begin{aligned} Sim_{Semantic}(c_1, c_2, SN) = \; & w_{Edge} \times Sim_{Edge}(c_1, c_2, SN) + \\ & w_{Node} \times Sim_{Node}(c_1, c_2, SN)) + \\ & w_{Gloss} \times Sim_{Gloss}(c_1, c_2, SN)) \end{aligned} \quad \textbf{(10)}$$

where: c_1 and c_2 are two concepts being compared, SN is the reference semantic network (such as WordNet), $w_{Edge} + w_{Node} + w_{Gloss} = 1$ and $(w_{Edge}, w_{Node}, w_{Gloss}) \geq 0$, Sim_{Edge} is a typical edge-based measure from [34], Sim_{Node} is a typical node-based measure from [34], and Sim_{Gloss} is a typical gloss-based measure from [36], expanded and normalized in [37, 38].

Consequently, similarity between two multimedia objects or two events, represented in our 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*:

$$\begin{aligned} Sim(o_1, o_2) = \; & \alpha \times Sim_{Time}(o_1, o_2) + \\ & \beta \times Sim_{Location}(o_1, o_2) + \\ & \rho \times Sim_{Semantic}(o_1, o_2) \end{aligned} \quad \textbf{(11)}$$

where o_1 and o_2 are two multimedia objects in MRSM, $(Sim_{Time}, Sim_{Location}, Sim_{Semantic}) \in [0, 1]$ designate temporal, spatial, and semantic similarity measures respectively (computed as inverse distances, e.g., $1/(1+Distance(o_1, o_2) \in[0, 1])$, $\alpha + \beta + \rho = 1$ and $(\alpha, \beta, \rho) \geq 0$. The same formula can be applied when computing $Sim(\varepsilon_1, \varepsilon_2)$ where ε_1 and ε_2 are two events represented in MRSM.

Based on the above formula and description, our combined MRSM similarity measure would be consistent with the formal definition of similarity [39, 40], 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*: $Sim(o_1, o_2) = 0 \Rightarrow$ *A* and *B* have no common characteristics,
ii. *Reflexivity*: $Sim(o_1, o_1) = 1$,
iii. *Symmetricity*: $Sim_{XDoc}(o_1, o_2) = Sim_{XDoc}(o_2, o_1)$

In fact, *triangular inequality* is controversially discussed and is usually domain and application-oriented [35, 40]:

iv. *Triangular inequality*: $Sim(o_1, o_2) \geq Sim(o_1, o_3) \times Sim(o_3, o_2)$ (i.e., $Dist(o_1, o_2 \leq Dist(o_1, o_3) + Dist(o_3, o_2))$)

While temporal and spatial similarity measures do satisfy triangular inequality, yet most semantic similarity measures in the literature, e.g., [34-36], fail to satisfy the latter property. An example by Tversky [41], 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 due to fact that semantic similarity is usually evaluated through multiple semantic relations (links) between concepts (nodes), e.g., *geographic proximity* on one hand, and *political affinity* on the other.

A possible solution, allowing to verify *triangular inequality*, would be to consider one kind of semantic relationships (e.g., *geographic proximity* only) when evaluating semantic similarity. In other words, $Sim_{Semantic}$ would be computed as the aggregation of multiple similarities evaluated each w.r.t. the corresponding relationship ($Sim_{Semantic\_GeoProx}$, $Sim_{Semantic\_PoliticalAff}$, etc.), where each measure would (individually, and when aggregated) verify *triangular equality*.

eid	1
T	{1238876657, 1238956813, 1238964834}
L	lat="45.5156" long="-73.5578" lat="45.517" long="-73.5571"
S	Ian Mosley, Mark Kelly, Pete Trewavas, Steve Hogarth, Steve Rothery, concert, gig, live, marillion, weekend, music, progressive, montreal, Steve Hogarth

eid	2
T	{1145040959, 1145041372, 1145105698}
L	lat="47.4357" long="-122.294" lat="47.4357" long="-122.294" lat="47.4357" long="-122.294"
S	Stardance, Norwescon Seattle, DoubleTreeHotel, Nikkor , Washington, conference, cosplay, costume , fantasy, scifi, convention

Figure 4. Events generated based on the sample images from Figure 3, described using our MRSM.

C) Event extraction

Given set of multimedia objects represented in our MRSM, we group them into clusters, based on their time, space, and semantic similarities, where each cluster of similar objects would identify an event. Here, we introduce an adapted graph-based agglomerative group average-link and partitioning clustering method [42], in order to perform event extraction as shown in Algorithm 1 (cf. Figure 5). Given *n* multimedia objects, a fully connected undirected graph *G* with *n* vertices and *n(n-1)/2* weighted edges is created. The vertices represent multimedia objects, and edges between nodes corresponding to the similarities between nodes computed following our aggregate similarity measure (cf. formula 11). Aggregate similarity scores computed for all pairs of input multimedia objects, are stored in an ($n \times n$) matrix of similarity scores (i.e. *SimMat*[][]), provided as input for event extraction as shown in Algorithm1.

The algorithm accepts multimedia objects (using *SimMat*) as input, and then groups objects based on their similarity scores using a hierarchical clustering approach. The algorithm generates clusters by varying the clustering level between l_o and 0, at a constant decrement pace of Dec-value. The group link clusters for a clustering level l_i can be identified by combining those vertices with weights $w \geq l_i$ from the graph *G*. Lines 8 and 9 show clustering at level l_o which group similar objects into *m* partitions calling *Generate_Initial_Clusters* (SimMat) function. Lines 11 to 15 show clustering at level l_i which involves two steps: firstly, computing the similarity distance between the two clusters using UPGMA (Unweighted Pair-Group Averaging Method) [43], as shown in formula 12; and secondly, grouping the clusters if their corresponding weight is greater than or equal to l_i:

$$\mathrm{Avg\_Sim}\left(\mathrm{clus}_1,\ \mathrm{clus}_2\right) = \frac{\sum_{o_i \in clus_1} \sum_{o_j \in clus_2} \mathrm{Sim}(o_i, o_j)}{|\mathrm{clus}_1| \times |\mathrm{clus}_1|} \quad \mathbf{(12)}$$

where o_i and o_j are multimedia objects in clusters $clus_1$ and $clus_2$ respectively, and $|clus_1|$ and $|clus_2|$ are cluster cardinalities (in number of objects).

```
      Input:
1.      SimMat[,]:Decimal  // similarities of pair of MM objects
      Variable:
2.      Dec-value: Decimal// clustering level decrement value (e.g., -0.1)
3.      li: Decimal   // Clustering level
4.      cl: Decimal  // stopping clustering level
5.      lo: Decimal  // initial parameter to have m partitioned clusters
      Output:
6.      Clusters: Collection  // contain the result of clustering
      Begin
7.      For li= lo Down to 0 Step Dec-value
8.        If li = lo Then
9.            Clusters=Generate_Initial_Clusters(SimMat)
10.       Else
11.         For each pair of clusters (ci, cj) in Clusters
                  //Clusters contains group of multimedia objects at level li-1
12.             Average-Similarity = UPGMA(ci,cj))     // eq. 11
13.             If Average-Similarity ≥ li Then
14.                group ci and cj in the same cluster
15.             End If
16.         Next
17.       End if
18.     Next
19.     cl =C-Index(Clusters)   // stopping rule for clustering
20.     Return clusters[cl]
      End
```

Figure 5. Event Extraction Pseudo code.

A stopping rule is necessary to determine the most appropriate clustering level for the link hierarchies. Milligan & Cooper in [44] present 30 such rules, among them, C-index exhibits excellent performance and is thus adopted in our study (line 19). Clusters identified at stopping clustering level are considered and returned as events (as shown in Line 20).

C. Event-based CK Management Framework

Given our space representation model (MRSM) and event extraction algorithm, we design our event-based CK management framework as shown in Figure 6. It consists of five interacting components: i) Identity Manager, ii) Input Data Manager, iii) Event Manager, iv) Profile and Context Manager, and v) CK Miner, which we briefly describe below.

1) Identity Manager

It is responsible for managing system user identities, using a role-based approach, where users are categorized broadly into two main categories: i) expert and ii) non-expert, with dedicated roles being associated to each group. An expert user gets support from the knowledge base system, empowered to add domain knowledge into the framework, and analyze the produced CK. Non-expert users have access and are able to upload event related multimedia objects, and can use data and knowledge search facilities using dedicated interfaces.

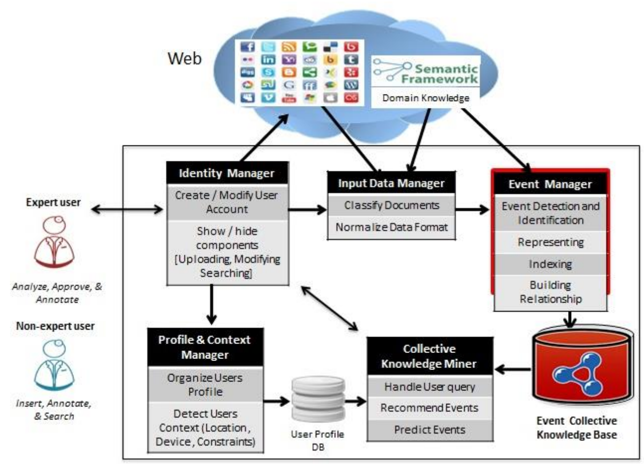


Figure 6. Event-Based CK Management Framework.

2) Input Data Manager

It categorizes input Web data sources into: i) event-related collections and ii) none event collections. It uses dedicated document classification algorithms to filter out none event collections from event collections. The heterogeneity in data representation formats adopted by different data sources is handed using our uniform data representation model. Thus the data from each source is mapped to our MRSM using dedicated mediators.

3) Event Manager

It is responsible for detecting candidate events from the multimedia data collection, identifies relevant events, and linking related events based on their semantic relationships. This component learns event discriminating features and their granularity values from source datasets. Then after, our unsupervised event detection algorithm (cf. Figure 5) is applied to identify potential events from massive data sources.

4) Profile and Context Manager

This component of the framework mainly handles user profile information. It activates when users create their accounts, and updates user context information when the latter connects to the system (context information including: user location, kind of device utilized by the user, as well as any other relevant constraints such as connectivity bandwidth and so on). The user profile and context information is later used to provide near real time notification of events, as well as to recommend the likelihood of related events as an effect of the preceding event for system users.

5) Collective Knowledge (CK) Miner

It uses the evolved event CK base to respond to user search requests. In addition, the component mines and provides recommendations and prediction values to registered system users. The event CK base is in turn responsible for maintaining event knowledge in a machine readable format (such as RDF/XML[5], or N-Triples[6]). The user profile database stores user personal information (such as user name, birth date, and hometown), as well as user interests and preferences.

In other words, we go from raw images with their descriptions (cf. Figure 3), to clusters of similar images where similarity is evaluated following the different dimensions of our MRSM. Clusters represent events which can also be described following our MRSM (cf. Figure 4). Event descriptions provide the seeds for knowledge (after being run through semantic analysis and disambiguation [45]) which, when grouped together with knowledge from other events, produces CK.

Note that in this paper, we mainly focus on the multimedia representation space and its properties, considering the temporal *Where*), location (*When*), and semantic (*What*) dimensions, while disregarding user related information (i.e., *Who*, *Why*, and *How* dimensions) to be covered in a future dedicated study.

V. EXPERIMENTAL SETUP & RESULTS

A. Experimental Dataset and Pre-Processing

We utilized the MediaEvalSED 2013 image dataset [17] to evaluate our event extraction approach. The dataset contains a collection of 131,211 photos and their associated metadata in XML (extensible Markup Language) format, uploaded between January 2006 and December 2012. Moreover, the dataset contains the ground truth event annotations which had been created by human users. The ground truth consists of associating each image to a single label designating an event, such that no image can belong to more than one event. Image metadata consist of an XML document associated to each image, containing *image_id*, *photo_url*, *username*, *dateTaken*, *dateUploaded*, *title*, *description*, *tags*, and *location* (defined in terms of *latitude* and *longitude*) among others. Based on our 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.

We utilized regular expressions to clean out the HTML (Hyper Text Markup Language) tags from the XML documents. In order to simplify similarity computations: i) temporal values were converted into UNIX epoch, ii) Non-English textual metadata were translated using the Google API Translate[7] services, iii) stop words were removed based on the WordNet stop word lists, and iv) terms were stemmed using the Porter stemmer[8].

B. Evaluation Metrics

To evaluate the quality of our event detection process, we used the *Normalized Mutual Information* (NMI) and *F-score* measures [46] which are commonly in the literature. On one hand, *NMI* is an informed probabilistic measure that evaluates the clustering accuracy (purity) of extracted events, computed by comparing the generated events (clusters) and the available ground truth (user defined clusters):

$$NMI(\Omega, C) = \frac{I(\Omega, C)}{[H(\Omega) + H(C)]/2} \quad \textbf{(13)}$$

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 entropies of the sets of generated clusters and predefined clusters respectively.

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) measure widely utilized in information retrieval [47].

$$F-Score = \frac{2 \times PR \times R}{PR + R} \quad \textbf{(14)}$$

C. Experimental Results

We ran a battery of experiments using different parameter values for weight parameters α, β, and ρ highlighting the impact of temporal (α), spatial (β), and semantic (1-α-β) dimensions when performing similarity-based image clustering to extract events. The top 5 results for both NMI and F-score are shown in Table 1. Results clearly highlight three observations: i) all three dimensions seem to be almost equally important in extracting meaningful events, since the best results were obtained with very close weight values for α, β, and ρ; ii) it is also clear that considering 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 ρ designating the impact of semantic similarity evaluation; iii) considering semantic information only (neglecting temporal and spatial dimensions, i.e., α=β=0 and ρ=1), likewise when considering temporal only or spatial only information, produced lower quality results, which points back to our first observation, i.e.: integrating all dimensions seems to be key in improving event extraction quality.

TABLE I. The best five results of our experiment by varying the temporal, spatial and semantic parameter values.

Parameter values	α (0.5) β (0.5) ρ (0)	α (0.4) β (0.4) ρ (0.2)	α (0.4) β (0.35) ρ (0.15)	α (0.35) β (0.35) ρ (0.3)	α (0.35) β (0.3) ρ (0.35)
NMI	0.9667	0.9759	0.9792	0.9826	**0.9865**
F-score	0.9255	0.9372	0.9388	0.9407	**0.9435**

Moreover, we compare our best experimental results with those of some of the main related works in the literature, namely those approaches who have also utilized the MediaEvalSED 2013 [17] test dataset as benchmark for applying unsupervised clustering approaches in order to extract events.

[5] http://www.w3.org/TR/rdf-syntax-grammar/
[6] http://www.w3.org/2001/sw/RDFCore/ntriples
[7] http://code.google.com/p/google-api-translate-java/

[8] http://tartarus.org/martin/PorterStemmer/

TABLE II. Comparison of experimental result with the existing state-of-the arts.

Method	Features	NMI	F-Score
Nguyen *et al.* [16]	Temporal, Spatial, Textual	0.9849	0.9320
Manchon-Vizuete *et al.* [48]	Temporal, Spatial, Textual	0.9731	0.8833
Sutltano *et al.* [49]	Temporal, Spatial, Textual	0.9540	0.8120
Our Method	Temporal Spatial Semantic	**0.9865**	**0.9435**

Results in Table 2 show that our MRSM-based approach was able to improve the event extraction process, which is mainly due to the fact that our approach considers the semantic descriptions and semantic similarities of user contributed metadata (title, tags and description) in the aggregated similarity evaluation process when performing similarity-based clustering, whereas existing methods focus solely on the temporal/spatial aspects and disregarding multimedia metadata semantics.

VI. CONCLUSION

In this paper, we introduce the core components of an event-based CK management framework, using shared multimedia data from social media sources to identify and describe events and their multimedia object constituents. It is built around a generic Multimedia Representation Space Model called MRSM, designed for multimedia data and multimedia-based event representation from heterogeneous multimedia data without any prior knowledge about event-related clues. Preliminary tests highlight the effectiveness of our multimedia representation space, integrating spatial, temporal, and semantic information for event extraction and representation. We are currently conducting additional tests to evaluate the scalability and adaptability of our solution when dealing with different kinds of multimedia objects with different sizes and properties. As upcoming works, we are investigating crowd-sourcing (using FOAF for instance) [50] as a supplementary metadata source, adding a fourth dimension to our MRSM for user-sensitive event detection and identification. We are also investigating auto-calibration techniques, allowing to choose the proper unit of measurement for each dimension of our MRSM based on the properties of media objects and events described. In the near future, we plan to implement and evaluate recommendation, inference, and query management functionality, using our MRSM, in order to infer related future events based on current ones, and handle related users query. Also, we aim to study the semantic relationships among events (i.e., their identification and formal representation following MRSM), toward creating an open linked event based CK base.

ACKNOWLEGMENTS

This study is partly funded by the National Council for Scientific Research (CNRS-L) Lebanon, project: NCSR_00695_01/09/15, by LAU research fund: SOERC1516R003, and LAU travel grant: SOERC1516T015.

REFERENCES

[1] Schaffers H. et al., *Smart Cities and the Future Internet: Towards Cooperation Frameworks for Open Innovation.* Future Internet Assembly, 2011. pp. 431-446.

[2] Batty M. et al., *Smart cities of the future.* The European Physical Journal Special Topics 2012. 214(1): 481-518.

[3] Vakali A. et al., *Smart Cities Data Streams Integreation: Experimenting with Internet of Things and Social Data Flow.* In 4th Inter. Conf. on Web Intelligence, Mining and Semantics (WIMS'14), 2014. ACM NY, USA.

[4] Kakarontzas G. et al., *Conceptual Enterprise Architecture Framework for Smart Cities: A Survey Based Approach.* In Proceedings of the 11th International Conference on E-Business (ICE-B'14), 2014. pp. 47-54.

[5] Tekli J., Abou Rjeily A., Chbeir R., et al., *Semantic to intelligent web era: building blocks, applications, and current trends.* . International Conference on Managment of Emergent Digital EcoSystems (MEDES), 2013. pp. 159-168.

[6] Spiegler I., *Technology and knowledge: Bridging a "Generating" Gap.* Information & Management, 2003. 40(6), 533-539.

[7] Zins C., *Conceptual Approaches for Defining Data, Information, and Knowledge.* Journal of the American Society for Information Science and Technology, 2007. 58(4), 479-493.

[8] Armbrecht R., Chapas B., Chappelow C., et al., *Knowledge Management in Research And Development* Research-Technology Management, , 2001. Volume 44, Number 4, pp. 28-48(21).

[9] Liang P., Avgeriou P., He K., et al., *From collective knowledge to intelligence: pre-requirements analysis of large and complex systems.* ProceedingProceedings of the 1st Workshop on Web 2.0 for Software Engineering (Web2SE '10), 2010. pp. 26-30, NY.

[10] Chbeir R. and Oria V., *Editorial preface: Special Issue on Multimedia data annotation and retrieval using web 2.0.* Multimedia Tools and Applications, 2013. 64(1): 1-5.

[11] Ma Z. et al., *Knowledge Adaptation for Ad Hoc Multimedia Event Detection with Few Exemplars.* In Proceedings of the 20th ACM Inter. Conf. on Multimedia, 2012. pp. 469-478.

[12] Jain U. and Westermann R., *Toward a Common Event Model for Multimedia Applications.* IEEE Multimedia, 2007. pp. 19-29.

[13] Psallidas F. et al., *Effective event identification in Social media.* IEEE Data Eng. Bull. , 2013. 36(3): 42-50.

[14] Becker H., e.a., *Identifying content for planned events across social media sites.* In Proceedings of the 5th ACM international conference on Web search and data mining, 2012. ACM, pp. 533-542.

[15] Horvitz K. and Radinsky E., *Mining the web to predict future events.* In Proceedings of the 6th ACM International Conference on Web Search and Data Mining, 2013. ACM, pp. 255-264.

[16] Nguyen T.V. et al., *Event Clustering and Classification from Social Media: Watershed-based and Kernel Methods.* In Proceedings of the MediaEval'13 Multimedia Benchmark Workshop, 2013. Barcelona.

[17] Reuter T. et al., *Social Event detection at MediaEval'13:Challenges, Datasets, and Evaluations.* In Proceedings of the MediaEval'13 Multimedia Benchmark Workshop, 2013. Barcelona, Spain.

[18] Yannoukakou A. and Araka. I., *Access to Government Information : Right to Information and Open Government Data Synergy.* In Proceedings of the 3rd Inter. Conf. on Integrated Information (IC-ININF), 2014. pp. 332-340.

[19] Gruen J. and Houghton N., *Open Research Data.* Australian National Data Service (ANDS),, 2014.

[20] Gray J. and Daribishire H., *Beyond Access: Open Government Data & the Right to (Re)Use Public Information.* Access Info Europe and Open Knowledge, 2011.

[21] Auer S. et al., *Dbpedia: A nucleus for a web of open data.* Proceedings of the 6th international The semantic web and 2nd Asian conference on Asian semantic web conference (ISWC'07/ASWC'07), 2007. Springer, pp. 722-735.

[22] Bizer C., Heath T. and Berners-Lee T., *Linked Data - The Story So Far.* Int. J. Semantic Web Inf. Syst., 2009. 5(3): 1-22.

[23] Liu X., Troncy R. and Huet B., *Using social media to identify events.* In Proceedings of 3rd ACM SIGMM International Workshop on Social Media, 2011. pp. 3-8.

[24] Bouchard H. et al., *GLOCAL: Event-based Retrieval of Networked Media.* European Project Grant Agreement No. 248984, 2012. www.glocal--project.eu.

[25] Mutuku L. and Mahihu C.M., *Understanding the Impacts of Kenya Open Data Application and Services.* iHub Research, 2012. Nairobi.

[26] Kelsey T., *Unleashing The Potential - The UK open data white paper.* Open Government Partnership, 2012.

[27] Xie L., Sundaram H. and Campbell M., *Event Mining in Multimedia Streams.* in Proceedings of the IEEE, 2008. 96(4): 623-647.

[28] Jeong-Dong K *et al.*, *Onto: Ontological context-aware model based on 5w1h.* Inter. Journal of Distributed Sensor Networks, 2012. p. 11.

[29] Takahiro I. *et al.*, *Information classification and navigation based on 5W1H of the target information.* In 17th Inter. Conference on Computational Linguistics, 1998. Association for Computational Linguistics, pp. 571-577.

[30] Smriti S. *et al.*, *News Event Extraction Using 5W1H Approach & Its Analysis.* Inter. Journal of Research in Engineering and Technology (IJRET), 2013. (4) 5.

[31] Fellbaum C., *Wordnet: An Electronic Lexical Database.* MIT Press, Cambridge, MA, 1998. 422 p.

[32] Hoffart J., *et al.*, *YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia.* Artif. Intell., 2013. 194: 28-61.

[33] Budanitsky A. and Hirst G., *Evaluating WordNet-based Measures of Lexical Semantic Relatedness.* Computational Linguistics, 2006. 32(1): 13-47.

[34] Wu Z. and Palmer M., *Verb Semantics and Lexical Selection.* Proc. of the 32nd Annual Meeting of the Associations of Comput. Linguistics, 1994. pp. 133-138.

[35] Lin D., *An Information-Theoretic Definition of Similarity.* Intern. Conf. on Machine Learning (ICML), 1998. pp. 296-304. Morgan Kaufmann Pub. Inc.

[36] Banerjee S. and Pedersen T., *Extended Gloss Overlaps as a Measure of Semantic Relatedness.* In Proceedings of Inter. Joint Conf. on AI (IJCAI'03), 2003. pp. 805-810.

[37] Charbel N. *et al.*, *Resolving XML Semantic Ambiguity.* Inter. Conf. on Extending Database Technology (EDBT'15), 2015. Brussels, Belgium, pp 277-288.

[38] Tekli J. *et al*, *Building Semantic Trees from XML Documents.* Elsevier Journal of Web Semantics (JWS), 2016. 37–38:1–24.

[39] Ehrig M. and Sure Y., *Ontology Mapping - an Integrated Approach.* European Semantic Web Conference (ESWC), 2004. pp. 76-91. Heraklion, Greece.

[40] Tekli J. *et al.*, *Approximate XML Structure Validation based on Document-Grammar Tree Similarity.* Elsevier Information Sciences journal, 2015. 295:258-302.

[41] Tversky A., *Features of Similarity.* Psychological Review, 1977. 84(4):327-352.

[42] Jain A.K. *at al.*, *Data Clustering: A Review.* ACM Computing Surveys, 1999. 31(3):264-323.

[43] Sneath P. et al., *Numerical taxonomy.* The Principles and Practice of Numerical Classification, 1973. San Francisco.

[44] Milligan G. and Cooper M., *An Examination of Procedures for Determining the Number of Clusters in a Dataset.* Psychometrika, 1985. 50(52):159-179.

[45] Tekli J., *An Overview on XML Semantic Disambiguation from Unstructured Text to Semi-Structured Data: Background, Applications, and Ongoing Challenges.* IEEE Trans. on Knowledge and Data Engineering (TKDE), 2016. 28(6): 1383-1407.

[46] Ghosh A. and Strehl J., *Cluster ensembles—a knowledge reuse framework or combining multiple partitions.* J. of ML Research, 2003. 3:583-617.

[47] Salton G. and Mcgill M.J., *Introduction to Modern Information Retrieval,.* 1983. McGraw-Hill, Tokio.

[48] Manchon-Vizuete D. and Giro-i-Nieto X., *UPC at MediaEval 2013 Social Event Detection Task.* . In Proc. of the MediaEval'13 Multimedia Benchmark Workshop, 2013. Barcelona, Spain.

[49] Sutanto T. and Nayak R., *Admrg@ MediaEval 2013 social event detection.* In Proc. of the MediaEval 2013 Multimedia Benchmark Workshop, 2013. Barcelona, Spain.

[50] Aleman-Meza B., Nagarajan M., Ding L., et al., *Scalable Semantic Analytics on Social Networks for Addressing the Problem of Conflict of Interest Detection.* ACM Transaction on the Web (TWeb), 2008. 2(1):7.