Overview of Event-based Collective Knowledge Management in Multimedia Digital Ecosystems

Minale A. Abebe
IT Doctoral Program
Addis Ababa University
3131 Addis Ababa, Ethiopia
minale.ashagrie@aau.edu.et

Joe Tekli
School of Engineering, E.C.E. Dept.
Lebanese American University
36 Byblos, Lebanon
joe.tekli@lau.edu.lb

Fekade Getahun
Computer Science Department
Addis Ababa University
3131 Addis Ababa, Ethiopia
fekade.getahun@aau.edu.et

Richard Chbeir
LIUPPA Laboratory
University Pau & Pays Adour
64000 Anglet, France
richard.chbeir@univ-pau.fr

Gilbert Tekli
Fac. of Technology, Mechatronics Dept.
University of Balamand
100 Tripoli, Lebanon
gtekli@gmail.com

***Abstract*—This paper provides an overview on the problem of event-based collective knowledge management from shared multimedia data. We start by introducing key concepts and constructs related to the problem, including multimedia digital ecosystems, collaborative environments, and collective knowledge management. Then, we utilize a real world motivating scenario to highlight some of the major challenges facing event-based knowledge organization in a multimedia collaborative environment, mainly the need to handle: i) heterogeneous data sources and their unstructured content, ii) large and growing volumes of data published online, iii) non-consistent and ambiguous multimedia data annotations, iv) misleading contents (that are not event related) published by non-experienced users, and vi) multimedia data with missing event-related meta-data. Consequently, we provide a short review of existing methods related to event detection from shared social multimedia data on the Web, contrasting their characteristics with respect to the above challenges, before highlighting potential research directions.**

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

I. INTRODUCTION

Nowadays, emerging technologies such as smart-phones, wireless Internet, 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.). Moreover, they are transforming the Web from a static data publishing platform into a collaborative information sharing environment [4, 34, 61, 81]. Nonetheless, attaining the next stage in Web development and engineering, i.e., the so-called *Intelligent Web*: allowing more sophisticated and meaningful human-machine and machine-machine collaborations, requires yet another breakthrough: allowing the sharing and organization of so-called *collective knowledge* (CK) [74].

In this context, CK can be viewed as the development and aggregation of knowledge assets (i.e., data associated with semantic meaning, e.g., the meaning of *Addis Ababa* is: *capital city of Ethiopia*, as opposed to processing *Addis Ababa* as a piece of textual data made of a bunch of Unicode characters) extracted from a distributed pool of data available on the Web [2]. Organizing individual users' and agents' knowledge assets and information into a meaningful collection of common knowledge (CK) would allow users and especially intelligent terminals (software agents) connected to the Web to easily (automatically) analyze and handle large collections of shared data, along with their links and transactions. Such an organization of CK can improve individual and collaborative Web information management (indexing, storage, exchange, search, and retrieval) [41, 74]. This can be paralleled in human collaboration and work experience: people who share their individual knowledge and know-how with their counterparts (creating some sort of CK between them) are likely to work, collaborate, and execute tasks more effectively and efficiently together (compared with people sharing minimal or no common knowledge) [16, 62].

However, building and handling CK faces many difficulties [74]. According to [14], more than 80% of the data shared on the Web are heterogeneous, streamed, unstructured, massive, multimedia, and are inherently associated to so-called events. An event can be defined as an observable occurrence at a certain time and place that interests a group of people (e.g., soccer match, car accident, heavy storm, etc.) [46]. 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.) [32]. However, annotations of similar multimedia 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 annotations/descriptions to identify meaningful events, which is needed as the building blocks for CK organization, remains an open problem.

In this paper, we aim to shed some light on the problem of event-based CK management from multimedia data. To do so, we start off by describing some of the basic concepts and constructs related to the problem, including multimedia digital ecosystems and collaborative environments in Section II. Then, we use a real world motivating scenario in Section III to highlight some of the major challenges facing event-based CK organization in a multimedia digital ecosystem. These include the identification of event related data and event descriptive features, as well as grasping the semantic meaning and relationships connecting data with events. Subsequently, we provide a concise review of methods related to event detection from multimedia data in Section IV, including cluster-based (unsupervised), classification-based (supervised), and hybrid approaches, highlighting their characteristics with respect to (w.r.t.) the challenges identified in the previous section. We conclude with research directions in SectionV.

II. BACKGROUND

In this section, we provide a brief background description of some of the main concepts related to this study, namely: multimedia digital ecosystems (or MDES, in Section II.A), collaboration in a MDES (Section II.B), and CK management in collaborative environments (in Section II.C).

A. Multimedia Digital Ecosystem

Chang and Boley [8] define a Digital EcoSystem (DES) as *a network of companies, individual contributors, institutions, and customers that interact in line with promoting collaboration among a group of users to achieve common goals and maintain mutual values* (such as maximizing financial profit, minimizing resource consumption, maintaining a stable system, or maintaining user satisfaction). In this context, digital ecosystems are considered as a framework of distributed and heterogeneous collaborative environments that handle data and information effectively to build CK for their community. In other words, a DES can be viewed as a framework to effectively handle data and information in an open, loosely coupled, distributed, and adaptive system, promoting properties such as self-organization, scalability, and sustainability inspired from natural ecosystems, where human users and software agents collaborate, interact, compete, and evolve autonomously to solve complex and dynamic problems [9, 10, 74].

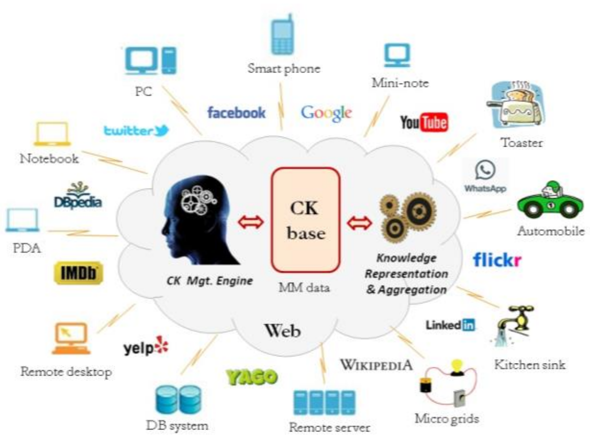

Figure 1. Overview of MDES architecture allowing CK management from social MM data.

The concept of Multimedia Digital Ecosystem (MDES) is extended from DES in which actors (human users or software agents) with common interests work together by sharing and processing "multimedia" resources: such as text, images, audio files, videos, and animations [37, 74] (cf. Figure 1). As in classical DESs, Actors in a MDES (i.e., users or agents) are responsible for keeping the ecosystem balanced and sustainable by equally contributing to the shared resources. In other words, an MDES is a DES handling multimedia (MM) information, where actors produce shared multimedia contents (e.g., images, videos, etc.) and processes (e.g. image filtering, video rendering, etc.), annotated collaboratively by human users and/or intelligent software agents in order to facilitate multimedia data manipulation, exchange, and accessibility [9, 10, 74].

B. Collaboration in a MDES

To realize the goal of a MDES [9, 10], there is a need for managing CK (cf. Section II.C) with the help of (and in order to produce) collective intelligence. Extracting and handling CK requires intelligent terminals (agents) which are not only capable of understanding and meaningfully processing information, but are also capable of thoroughly collaborating and even "reasoning" together, as a collective, to produce and handle common CK, leading to more sophisticated (intelligent) services, as well as achieving the ultimate goal: *collective intelligence*, where agents are able to automatically sustain themselves and evolve without direct human intervention [74]. In line with the vision of the Semantic Web (SW), a MDES also enables intelligence services (such as information brokers, search, and information filters) allowing to process information more effectively and efficiently [7, 17, 79]. The author of [40] discusses that individual intelligence is coordinated and constantly enhanced through communication, collaboration, and exchange, among different individuals. Similarly, collective intelligence in a MDES effectively mobilizes the CK of a group of actors (users and/or agents) feeding from and simultaneously improving the group's CK [22, 53]. In short, collective intelligence in a MDES aims to realize a "true" collaboration between humans and machines [19, 28, 30].

Collaboration among actors/species of a DES (human and machines) plays a vital role in the sustainability of the ecosystem. To allow an effective collaboration, each member of the ecosystem is required to contribute its part in the form of knowledge (a well-defined representation of information), which can be understood by different actors (human users and software agents) involved in the environment. Knowledge is fundamentally created by an individual [55] but needs to be enriched from a group of users: participating, sharing, adding, and validating the available knowledge for common use. In this context, the knowledge engineering process consists in codifying the knowledge into well formed information, exploring and recommending the existing knowledge, and facilitating collaboration to combine an existing body of knowledge and make it evolve into a new form of (more useful) knowledge [49, 63, 68].

C. Collective Knowledge Management

To better understand the issues and challenges of CK management, we start by explaining the basic concepts of *data*, *information*, *meta-data*, and *knowledge*, which are required to produce CK. These terminologies have been defined inconsistently (and sometimes definitions conflict) in different areas of computer science and other disciplines such as management science, epistemology, and psychology. Sometimes the terms *data* and *information* are used interchangeably (e.g., confusing: the concepts of information processing and data processing, or data management and information management). However, within the scope of this paper, we adopt the most widely accepted definitions in the computer science literature:

Definition 1 – Data (*plural);* **Datum** *(singular)*: is the most basic (raw) representation of facts, concepts, or instructions suitable for communication, interpretation, or processing by human beings and software agents [26, 78]. The datum is usually without context. It basically comes down to numbers, texts, or multimedia objects [85].

For instance, *"2001"* is considered as a number consisting of 4 digits, and highlights no information at all. For this piece of data to be *informative,* it must be interpreted and given a well-defined meaning, such as *"the year of announcement of the Semantic Web"*. More formally:

Definition 2 – Information: is a representation for assigning meanings to the raw *data*, adding context, organization, processing, or visualization, which makes the data more "informational", more understandable, and thus more valuable for human beings and software agents [64, 85] (e.g., drawing a chart highlighting the results of a given computational process based on raw statistics, or visualizing information concerning salient locations on a map based on raw geographic coordinates in a Global Information System [54]. In other words, *information* highlights transcripts of some meanings assigned by human beings [15, 50].

Definition 3 – Meta-data: is a description of data that describes some features of a digital content. It is useful to summarize, search, retrieve, determine access privilege, give usage history, give ownership information, indicate relationship with other recourses, and control the management of digital content [29].

Meta-data can describe who gave the data/information (e.g., *Wikipedia*), when was the data/information given (e.g., *published in 2002*), etc. At a higher level of abstraction, knowledge is viewed as the combination of all known data, information, and meta-data concerning a given concept or fact [85] (like knowing that *"the year of announcement of the Semantic Web"* is *"2001"*, following *Wikipedia* in an article *published in 2002*). More formally:

Definition 4 – Knowledge: is a representation for assigning *semantics* to *information*, i.e., assigning more sophisticated meaning regarding the purpose (goal), the use, and the impact of information w.r.t. a given system, application, or environment. It is a rigorously structured and contextualized representation of information, combining data, information, and meta-data, as well as the semantic links between them, to allow performing more complex tasks (e.g., problem solving, decision making, detecting anomalies, etc.) [38]. Knowledge is usually acquired as the result of computer-simulated cognitive processes, such as perception, learning, association, and reasoning, resulting in transcripts of some semantics acquired by human beings [64, 85].

In this context, knowledge stored in a computer system is referred to as *knowledge representation*. More formally:

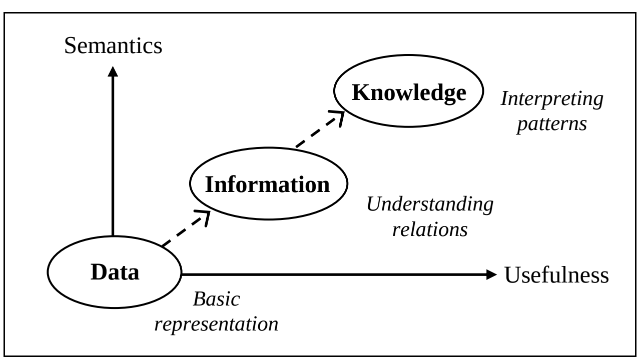

Figure 2. Relationship between data, information, & knowledge.

Definition 5 – Knowledge Representation: It is a structured collection of semantically rich information and a set of associated inference rules, written using dedicated description languages (such as RDF [82] and OWL [48]) that can be used for automated reasoning by software agents. It refers to the application of logic and ontology to the task of constructing computable models for a given domain [60, 65].

Figure 2 below shows the interrelationships between *data*, *information*, and *knowledge* w.r.t. semantic expressiveness and usefulness in practice.

Definition 6 – Collective Knowledge (CK): is defined as a development of knowledge assets or (semantic) information resources from a distributed pool of contributions. It involves combining the knowledge of a crowd of human users and software agents, representing consensus on the commonalities, intersections, and disparities in the knowledge forming the CK, and allowing to infer new information to improve information management, such as indexing, storage, search, manipulation, and retrieval [41].

CK is perhaps most relevant w.r.t. Tim Berners-Lee's vision of the Semantic Web (SW), where software agents are capable of automatically analyzing large collections of data with their contents, links, and transactions, in order to improve data accessibility, management and exchange between people and computers [7, 67]. Following the vision of the SW, software agents handle knowledge (cf. Definition 4), i.e., information (cf. Definition 2) given semantic meaning, which allows data and meta-data to be efficiently shared and reused across application, enterprise, and community boundaries. In this perspective, the SW can be referred to as a source of CK about a Web resource or a set of resources, which are created by users participating in the process of building such knowledge voluntarily or purposely.

III. CHALLENGES TOWARD EFFECTIVE CK MANAGEMENT

Performing CK management requires handling event-based resources which are heterogeneous, streamed, unstructured, massive, multimedia, and are usually associated *events* [14]:

Definition 7 – Event: An event is generally defined following the 5W1H model: *When*, *Where*, *What*, *Who*, *Why* and *How* aspects [33, 64, 72], as 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(*who*) happening within a certain time (*when*) and location(*where*, e.g., stadium, road, city, or amphitheater), and having a certain description (*why*) and identification/traceability (*how*) from the set multimedia objects describing it [46].

Usually participants in an event capture data (text, image, video, audio, etc.), understand the content, annotate, publish and share them to describe the event (e.g., videos from a soccer match, pictures from the storm, opinions about presidential debate, etc.) [32]. As a result, agents (human and software) are acting as intelligent and proactive data sensors and actuators and form a collaborative digital environment [4, 34, 61, 81]. Such a collaborative digital environment allows entities having similar profiles or interests to create, annotate, and share digital contents. Whenever an event occurs, an agent may capture data (a photo, audio or video), annotate it and describe it based on its background knowledge and experience. Subsequently, the agent can share or publish the data with its corresponding meta-data (timestamp, location, title, description, tags and etc.,) and knowledge (associated semantic concepts) and make it available to other agents. Yet, extracting and processing events from multimedia data comes with many challenges, such as dealing with different data sources, different kinds of data objects and user provided meta-data, discrepancies in description formats, as well as lack of semantic meaning, among others. In the remainder of this section, we present a real world motivational scenario that illustrates these challenges.

A. *Motivating 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[1], Flickr[2], or YouTube[3] (cf. Figures 3-5). They might also post comments on twitter[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 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.

B. *Multimedia Data Sources on the Web*

Since the past couple of years, social media sites have become the single largest group of contributors of multimedia data on the Web. According to the Pew Research Center 2016 report [21], as of November 2016, 3.419 billion people (from the estimated 7.395 billion world population) have Internet Access. Among these, 67% (i.e., 2.307 billion) are active in social media applications' (such as Flickr, Facebook, YouTube, Instagram, Google+, Twitter, etc.). Social media applications have ushered in one of the largest collaborative environments worldwide in which billions of social media users are daily sharing multimedia contents which interest them. Many of the multimedia objects shared on these sites are related to some events which have important roles in users' lives (e.g., they are part of the event, or they know someone in the event, or they want to take part in the event). For instance, users of YouTube (cf. Figure 4) can share homemade videos and animations associated with their corresponding descriptions or annotations. Once those multimedia objects are shared by the owner (i.e., who captures or creates the video or animation), other users can enhance the shared multimedia contents by adding additional textual information such as a title, description, tags, comments, and ratings, which we refer to as user contributed contents. Moreover, social media services usually capture/associate basic characteristics and meta-data with the published multimedia objects. For instance, YouTube associates with each published video a unique video id, the date/time of creating or uploading the video, the author's name, and the location at which the video was created or uploaded as meta-data. Similarly, the Flickr API serves as a collaborative tool for social media users to share photos/images, graphs, charts and their related tags (cf. Figure 3). Also, the Twitter API allows its users to use up to 140 characters to briefly describe a Multimedia data, and re-tweet shared links, photos, and videos (cf. Figure 5). A major challenge here is how to gather, match, and organize information from all of these difference data sources to make sense out of it and produce a useful CK.

C. *Multimedia Data Description and Representation*

User contributed multimedia contents and meta-data on the Web can be represented in different structures and formats as shown in Figures 3-5. For example, YouTube and Flickr use the eXtensible Markup Language (XML) to disclose user contributed contents and meta-data, 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 example, Flickr and YouTube use different XML data element names, attribute names, and document structures to represent the date of creation/uploading of multimedia objects (e.g., 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> 2014-07-7 14:48:04 </upload_time>`, and with Twitter: `"timestamp":1316656366000`). Moreover, the date, time, and location meta-data can be represented in a different format specific to each social media service. For example, 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)[5] 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 multimedia 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).

Similarly, the location information associated with a multimedia object might also be represented in different formats. For example, 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>`).

---

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

5 The W3C Date and Time Format, available at: https://www.w3c.org/TR/NOTE-datetime


```
<photoid="14646512184"  camera="">
<exif tag="Keywords" label="Keywords">
<raw>Torrential Rainfall</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="By-lineTitle" label="By-line Title">
<raw>STR</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> It was on July 7, 2014, at around 3:00pm just 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,
    near the World Bank Country Office building, covered by an about 200-
    metre long stream.</raw></exif>
</photo>
```


Figure 3. A photo post and its meta-data in XML format extracted from Flickr.

In addition, information published by different social media services can vary in content and structure. For example, YouTube only provides uploaded time stamp, whereas Flickr captures both created and uploaded time stamps. Also, different users might publish identical multimedia objects (on the same or different social sites) with very different annotations, using free text descriptions and tags which might be syntactically different, yet semantically related, following their own style of writing, vocabulary, and experience in annotation. For example, YouTube represents all user contributed textual content with one XML element (i.e., `<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",“Trafic Chaos"]`), whereas Flickr uses a predefined element (*raw*) and predefined attributes (*tag* and *label*) to represent user contributed textual content (i.e., `<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>`).

Therefore, detecting events using meta-data and user contributed contents from heterogonous sources highlights various needs. Firstly, there is a need to convert the source meta-data into a uniform data model. The data model should be generic enough to model multimedia objects following a high-level representation[6] suitable/adapted for the purpose of event detection and identification. Secondly, we need to compute/evaluate the similarity/relatedness between multimedia objects given their adapted high-level representation, in order to match and merge similar (or identical) contents, and thus avoid useless duplications. Thirdly, we need to aggregate related multimedia objects together, and process their common features in order to detect and identify relevant events (e.g., recognizing and aggregating similar *flood* images published on Flikr and Facebook, with related meta-data, might help identify a new*flooding* event).


```
<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 an intense rain for less
      than an hour. It 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></tags>
<video_url>http://www.youtube.com/watch?v=sbE4svRaQ</video_url>
<thumb_url>http://i1.ytimg.com/vi/l6ZB2LSuz28/0.jpg</thumb_url>
</video>
```


Figure 4. A video post and its meta-data in XML format extracted from YouTube.


```
{"text":"A terrible evening  Addis Ababa (around Olompia) today, heavy
rain cause flood in the new road", "cat":[1],
"keywords":[“AddisAbaba",“Ethiopia”,”Flood",“Trafic 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":1316656366000,
"source":"web", "trans_text":"", "retweet_count":177,
"user":{"location":“Addis
Ababa","verified":false,"screen_name":"Muradissa"}, "tweet-lang":9}
```


Figure 5. A tweet and its meta-data in JSON format extracted from Twitter.

---

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

D. Organizing Multimedia Objects and Related Events

In line with the basic principle of the Linked Open Descriptions of Events (LODE) ontology [36], a given multimedia object can be classified as event related or not if its corresponding meta-data maps with the (so-called) factual aspects of an event. The factual aspects of an event are information characterized by the 5W1H model: *What? Where? When? Why*? *Who*? and *How?* [33, 64, 72]. Yet, in most existing studies related to event detection (cf. Section IV), location (*Where*) and time (*When*) related information are considered as the minimum sufficient constituents required to determine whether a given shared multimedia object is event related or not.

```xml
<page>
        <title>Lionel Messi wins 2014 World Cup Golden Ball </title>
        <ns>0</ns><id>7049</id>
        <timestamp>2014-07-13T 22:04:04Z</timestamp>
        <contributer>22:04:04
                <username>Jack Sargeant</username>
                <id>4460</id>
        </contributer>
        <model>wikitext</model>
        <format>text/x-wiki</format>
</page>
```

a. XML document extracted from Wikinews.

```xml
<?xml version="1.0" encoding="utf-8" ?>
<rsp stat="ok">
<photo id="14646512184" secret="1fb9e7d2a3" server="3916"
farm="4" camera="">
…… output omitted
<exif tagspace="IPTC" tagspaceid="0"
tag="Country-PrimaryLocationName"
label="Country- Primary Location Name"><raw>BRA</raw></exif>
<exif tagspace="IPTC" tagspaceid="0"
tag="Caption-Abstract" label="Caption- Abstract">
<raw>Germany's goalkeeper Manuel Neuer, left, winner of the Golden
Glove award for best goalkeeper, stands with Argentina's Lionel Messi…
Germany won the match 1-0. (AP Photo/Martin Meissner)</raw>
</exif>
</photo>
</rsp>
```

b. XML document extracted from Flikr.

```xml
<tweet>
     <text>Messi wins World Cup Golden Ball... but did he deserve it?
     http://t.co/FMJnUI5he1 via@MailSport .. Absolutely NOT.</text>
     <created_at>Wed Jul 16 09:00:18 0000 2014</created_at>
     <name>FAYE2117</name><screen_name>faye2117</screen_name>
     <source>&lt;a ref=&quot;http://twitter.com/download/android&quot;
     rel=&quat; nofollow&quot;Twitter for Android&lt;/a&gt;</ source>
     <id>489333693576536066</id>
</tweet>
```

c. XML document extracted from Twitter.

Figure 6. Sample XML meta-data extracts

A major issue here is that some shared multimedia object meta-data on social media sites do not contain minimum event determining feature sets. For instance, meta-data fragments in Figure 6.a and b do not have a location attribute. Also, descriptions in Figure 3 and Figure 6.c only implicitly[7] state the location and time information.

E. Main Challenges

To sum up, sharable multimedia data often consists of objects of different types (images, animations, videos, etc.), formats (bmp, svg, mpeg, x3d, etc.), coming from different sources, annotated by different users with different backgrounds (e.g., novice, experts, scientists, etc.) who can sometimes produce misleading information or omit relevant information (missing certain event discriminating features, following the 5W1H model), all of which would affect how knowledge would be processed based on the latter (e.g., an expert meteorologist would describe a storm or a heat wave differently from a non-expert observer). A first major challenge here is how to determine the relative importance or weight of different event discriminating features (i.e., which dimension of the 5W1H model is more important) in the event extraction process. Secondly, how to classify multimedia objects as event related or not in the absence of (minimum or full) event discriminating features. Thirdly, how to properly handle the semantic meaning, ambiguity, and relatedness of the textual descriptions of event discriminative features (e.g., how to understand the semantic relationships and differences between terms *hailstorm*, *rainstorm*, and *blizzard*, which could be used by different users in describing the same or similar events).

In this context, handling diverse, heterogeneous, and sometimes misleading or incomplete multimedia meta-data to identify meaningful events is needed as a central building block for CK management. Also, handling minimum event discriminative features, namely the spatial coverage (location of multimedia object, e.g., picture or video), the temporal coverage (time frame of multimedia object, e.g., instance of time at which picture was taken, or time interval at which video was taken), and most importantly the semantic meaning of shared multimedia objects (nature and meaning of multimedia data to help identify and resolve misleading or ambiguous information, e.g., identifying that picture of *heavy clouds* is related to video of *ranging tropical storm*, based on the semantic similarities of their textual descriptions) become of key importance in performing event detection and identification from multimedia data, which we review in Section IV.

IV. EVENT DETECTION IN SHARED MULTIMEDIA OBJECT COLLECTIONS

Over the past few years, different approaches have targeted event detection from social media streams and shared multimedia objects. These can be categorized as cluster-based (unsupervised), classification-based (supervised), and hybrid approaches (combining clustering and classification processes). In this section, we briefly review these approaches considering the main challenges identified in the previous section. Readers can refer to [82] for a detailed review on event mining.

A. Cluster-based Approaches

Clustering or unsupervised classification is the process of organizing or grouping a collection of objects into groups (called clusters) based on their similarity value. Similarity is evaluated as the inverse of a distance function in a certain referential space [76, 77]. Objects in the same group or cluster are more similar to(less distant from) each other than to those in other groups or clusters. Cluster analysis has been used for a variety of applications in data mining (cf. reviews in [1, 31, 42]) including event detection. The following paragraphs discuss cluster-based event detection methods from Web-based social and multimedia data.

Chen and Roy [43] propose an approach for detecting events from photos on Flickr by exploiting the tags supplied by users. Given a set of Flickr photos, with both user tags and other meta-data, including time and location (latitude and longitude), the proposed solution attempts to discover a set of photo groups where each group corresponds to an event. The method consists of three steps: (1) identifying whether tags are related to events

[7] There are no XML tags named *location*, *time* or *timestamp* as shown in Figures 4 and 5. Rather, the location and time information are described indirectly using the XML tag attribute name such as: **tag="City" label="City"><raw>Addis Ababa</raw>**

or not based on their temporal and spatial distributions; (2) detecting event-related tags to classify them further into periodic or non-periodic (a-periodic) event tags; and (3) retrieving the set of photos for each tag representing an event. The authors also consider tag usage occurrences as an additional feature to detect events. However, the proposed solution does not consider the semantic similarities/differences of user contributed textual features in the event detection task. In addition, the method does not show the impact of aggregating the similarities of different features (e.g., location, time, and high-level multimedia descriptors) in the event detection process.

The authors in [44] attempt to address the problem of structuring social media activities into events, by utilizing different properties (such as location, time, and user participation) from social media sites, 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. Moreover, the authors use visual summaries in order to visually filter/prune multimedia objects related to those detected events. Yet, the approach in [44] requires a certain number of initial seed photos (i.e., the product of shared images and owners who are posting those images should not be less than a threshold value obtained empirically) in order to effectively detect events.

Rafailidis *et al.* [59] present a data-driven approach which consists of three steps: (1) having images clustered based on their spatio-temporal information (*where* and *when*), producing so-called “anchored” clusters, whereas images which do not have spatial information are left as a singleton clusters; (2) calculating pair-wise and aggregated similarity measures between the “anchored” and the singleton clusters, considering various information such as: the multimedia object creator (*who*), title, description, tags, and visual information of images (*what*); (3) merging the clusters based on the calculated inter-correlations of the second step. The proposed solution uses the Jaccard (syntactic) similarity measure to compare textual descriptions, and thus does not address their semantic meaning while detecting events. Moreover, the authors do not highlight the effect of using an aggregated similarity measure (combining location, time, semantic, and authorship features) in the event detection task. A similar data-driven approach is developed in [47], where the authors build on an original work from Microsoft Research [57] named PhotoTOC. Initially, a time-stamp (*when*) based clustering is performed based on spatial information (*where*), textual description labels (*what*), and the photo creator’s information (*who*). Weights for each feature are manually tuned and used for the first solution. A training dataset is used to estimate the relevance of each feature type as well as the merging threshold for the combined feature score. An estimation of the relevance of each feature type is computed. Yet, similarity to its predecessors, the solution in [47] does not consider the semantic meaning of multimedia objects’ textual descriptions, but rather evaluates their syntactic similarities.

In contract with most above studies which do not capture the semantics of multimedia objects, the authors in [24] put forward a framework to semantically structure a multimedia object collection in social media applications. The authors use WordNet based semantic similarity measures [11] for the purpose of event detection, where WordNet is utilized as a reference machine-readable knowledge base [18]. Primarily, spatial information (*where*) is used to cluster the multimedia object collection based on the semantic similarities of their descriptive tags (*what*). When the multimedia collection does not have spatial information, the temporal information (*when*) is used to cluster multimedia objects. A similar approach is developed in [84], introducing a three stage clustering solution: (1) clustering images based on the multimedia object creator information (*who*) and temporal (*when*) feature, (2) merging the obtained clusters based on the location’s distance (*where*); (3) temporally and spatially similar clusters with similar textual descriptions are merged by a combined clustering scheme that takes both topic (semantic) and term syntactic similarity (*what*) into account. Cluster merging and updating is performed iteratively to successively grow clusters. Nonetheless, the solutions in [24, 84] do not evaluate the effect of using aggregated similarity measures (combining different features) on the event detection process.

B. Classification-based Approaches

In addition to clustering-based methods, various classification-based solutions were also developed to perform multimedia event classification in social media. 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 [39]. When a newly unlabeled pattern is encountered, it is compared with existing classes in order to be assigned to the most similar (related) class, or to be considered as a new class of its own if it is significantly different from all existing ones [76]. The following paragraphs discuss classification-based event detection methods from Web-based social and multimedia data.

Liu *et al*. [45] present a method that combines semantic inference and visual analysis for finding multimedia illustrating events. Their goal is to design a web-based platform that allows web users to explore and select events and related multimedia contents. They present a large dataset composed of semantic descriptions of events, photos and videos interlinked with the larger Linked Open Data (LOD) cloud in order to show the benefits of using semantic web technologies for integrating multimedia meta-data. The authors use special machine tags (e.g., *lastfm:event*=XXX, *upcoming:event*=XXX) associated with their multimedia 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/locations, etc.). Yet, the proposed solution does not identify instantaneous/unknown events such as an unexpected flood, earthquake, or thunderstorm. 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 [6] 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 [6] 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 shared multimedia objects’ textual descriptions and the impact, nor do they discuss the impact of an aggregated similarity measure combining different event descriptive features in the process of event detection.

The authors of [83] propose a fusion-based method to detect and identify events. The approach relies on learning a similarity metric between two documents (i.e., multimedia objects). The authors use Factorization Machines[8] to learn the similarity between two documents. Once they get the similarity metric, they use incremental clustering with a quality threshold to detect events. Yet, this work considers multimedia object features holistically, and does not consider the effect of individual object features in the process of event detection.

C. Hybrid Approaches

In addition to clustering and classification methods, a few hybrid solutions (combining both clustering and classification) to perform social multimedia event detection have been proposed.

In [5], the authors propose ensemble and classification-based similarity learning techniques to detect events. The authors discuss the different distinctive representations of social media documents. They define similarity methods for each document representation and explore various techniques for combining them into a single measure of social media objects' similarity evaluation. Both ensemble and classification-based similarity learning techniques are used in conjunction with an incremental clustering algorithm to generate a clustering solution. Each cluster corresponds to an event and includes the social media documents associated with the event. Events are determined based on clustering techniques which are not known beforehand. Also, the proposed solution is focused on music resources, and the authors aim to provide tag recommendations in terms of musical themes, moods, genres, or styles. Yet, the authors do not discuss the effects of spatial and semantic aspects of the shared multimedia objects in the event detection task.

A comparable approach is provided in [70], where the authors introduce a constrained clustering algorithm method, adapted from the spherical K-Means algorithm [69], to detect events from a social multimedia object collection. The K-Means method compares each document with all cluster centroids and forms the clusters iteratively. The number of initial clusters *k* is set in the training phase. All text data such as title, tag, username, and description are combined into a text field and are treated as a short document. Cosine similarity is used to measure the distance between a document and centroids based on the text information, combined with spatial and temporal distances into an aggregate linear similarity measure. Threshold values of temporal, spatial, and textual features are used to decide whether a document is assigned to one of the existing clusters or forms a new cluster. However, the approach does not consider the semantic aspect of textual features: it rather computes syntactic similarity at the text "surface" level using TF-IDF[9] term weights. Moreover, the authors do not discuss the effect of the aggregated similarity measure on the process of event detection.

The authors of [52] propose the watershed-based method with external data sources. They introduce a user-centric data structure, named UT-image (user-time image), to store a multimedia collection's meta-data. The whole meta-data set is turned into a UT- image, so that each row of an image contains all records that belong to one user; and the records are sorted by time. Then after, the merging process is performed based on either the temporal (*when*), spatial (*where*), or textual (tag/title/description, i.e., *what*) feature distance of pairs of clusters, considering the temporal, spatial and textual thresholds set by the user. Moreover, the solution uses the most common user provided keywords in order to make further cluster merging. But, the proposed method does not consider the effect of an aggregated similarity measure combining temporal, spatial, and textual features in the process of cluster merging. In addition to that, the authors themselves state that using the Jaccard (syntactic similarity) measureto compare the textual features distance fails to address the challenge of capturing the semantics of collaborative tags.

The work in [56] uses the Chinese Restaurant Process to cluster a multimedia collection of social media applications. The authors assume that multimedia objects have a unique timestamp and arrive sequentially in a streamed fashion. Thus, when a new multimedia object (photo) comes in, similarity is computed with the already existing multimedia objects. Then, a single pass incremental clustering algorithm is used to make a stochastic decision to either merge the object with the clusters (events) which already exist, or to create a new cluster (event) around it. However, the similarity measure developed in this study is based on a probability model constructed from the training data set in which the authors consider that two data points sharing the same value for a specific attribute, will also belong to the same event. Nonetheless, it is not always true that two multimedia objects having the same temporal value belong to the same event (for example, two multimedia object representing two different events taken at the same time, i.e., a football match between St. Gorge and Ethio-Buna hosted July 5, 2016 at 4:30 p.m. in Addis Ababa Stadium, and a Wedding ceremony celebrated July 5,2016 at 4:30 p.m. in Hawass may not belong to the same event). Similarly, two multimedia objects having the same spatial value but with different temporal time stamps, may not necessarily belong to the same event. Moreover, the approach does not show the impact of using an aggregated similarity measure combining temporal and spatial features in the event detection process.

D. Discussion

To summarize, most 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., [6], ii) generate events based on predefined clues (and are not able to identify unknown events), e.g., [58], or iii) consider manually defined thresholds which affect event detection coverage (missing certain events) and thus quality, e.g., [44]. Also, most existing approaches iv) do not apply weighted aggregated similarity measures to combine different event descriptive features in the event detection process. In contrast, one approach in [83] combines all features holistically, without allowing the user to evaluate the impact of every feature separately. In addition, v) 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. With textual descriptions, most approaches utilize syntactic similarity measures (such as Jaccard or cosine, coined with TF-IDF weighting), thus only capturing the surface level similarity of textual descriptors, and missing their semantic relatedness. Even though the open linked data initiative supports shared knowledge within the research community, e.g., [3, 20, 23, 83], 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.

[8] Factorization Machine (FM) is a new classification model that combines Support Vector Machine (SVM) functionality with factorization models [83].

[9] Term Frequency – Inverse Document Frequency

V. CONCLUSION

A. Recap

In this paper, we gave a brief overview on the problem of event-based CK management from multimedia data. We started by providing a concise description of some key concepts and constructs related to the problem, including MDES, collaborative environments, and CK management. Then, we used a real world motivating scenario to highlight some of the major challenges facing event-based CK organization in a MDES, mainly the need to handle: i) heterogeneous data sources and their unstructured content, ii) non-consistent and ambiguous multimedia data annotations, depending on annotators' background, iii) misleading contents (that are not event related) published by non-experienced users, and iv) multimedia data with missing event-related meta-data such as location (*where*), date/time (*when*), or annotation (*what*) information[10]. Consequently, we provided a concise review of existing methods related to event detection from shared social multimedia data on the Web, contrasting their characteristics w.r.t. the above challenges. In short, applying existing clustering and classification methods to process multimedia objects with event-related features published on social sites has been shown to perform poorly in various event detection and identification tasks, e.g., [6, 58, 80].

B. Research Directions

1) Addressing Major Challenges

To start with, more effort needs to be put in publishing digital content, associating it with proper and sufficient meta-data, to be able to effectively/efficiently use the data/meta-data for later processing. In this context, using data normalization [25, 27], data merging [12, 71], and semantic mediation [35, 51] solutions could help reduce or eliminate the discrepancies among multimedia data formats, structures, and semantics especially when acquired from different users and data sources. In this context, evaluating the semantic meaning of multimedia contents and meta-data becomes of key importance to improving data storage and normalization, as well as detecting events from multimedia collections. This requires extracting semantic concepts (meaningful terms defined w.r.t. a reference knowledge source, such as a digital dictionary or ontology), to structure the annotation, and to link them in a standardized knowledge representation format in order to perform multimedia retrieval and logical reasoning [75]. Performing semantic disambiguation to associate proper meanings with textual terms, which could help minimize the impact of ambiguous or misleading information, would be of key importance here [73]. The resulting CK then needs to be represented in a common form to be shared among members of the digital ecosystem (authoring), exploring and recommending the existing knowledge (finding and reminding), evaluating the produced CK's accuracy in correctly describing events (through collaborative user feedback, data mining, or statistical analysis, etc.), and combining existing and external bodies of knowledge (knowledge reuse) to build a new (more useful/up-to-date) one [44, 58].

2) Toward Intelligent MDES

More interestingly, with the development of the mobile Internet (smart phone-based) and communication technology, the SW vision is being extended toward a new innovative stage: the *Intelligent Web* (IW), also known as *Internet of Things* (IoT), where semantically rich objects: i) either physical objects of the real world, with added digital components (e.g., smart phones, smart cars, robotic systems, etc.) also referred to as intelligent terminals; and ii) software agents (e.g., scripts, applications, APIs, etc.) autonomously interact, sustain themselves and evolve in a MDES, provided with embedded communication capabilities, common (collective) semantics, and addressing schemes. This requires machines to store and manipulate even more knowledge about shared contents in the MDES.

Hence given the need for CK in MDES, providing effective CK management capabilities in such a large and distributed environment as the IW inherently requires superior processing capabilities. This could be achieved through interconnectivity between intelligent systems (human users, but mostly intelligent software systems). Here, the term *intelligence* refers to the system's ability to cope with new problems and use the power of knowledge analysis, reasoning, and inference efficiently. An intelligent system is expected modify its course of action in light of ongoing events, and to learn and interact with its environment [74]. Such a system would evaluate the alternate courses of action and judge their effectiveness and capacity toward an optimal and proactive behavior [13]. Developing a *collective intelligence*, as an interconnected network of intelligent agents, communicating and collaborating, exchanging data and services, and able to properly create and make use of meaningful multimedia-based CK, in order to autonomously adapt to a new environment or to changes occurring in the current environment, remains a major research direction in the decade to come [74].

[10] We set aside the obvious challenge of handling large volumes of data and content published online and growing continuously, which comes down to the infamous *Big Data* problem.

ACKNOWLEDGMENTS

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

REFERENCES

[1] Algergawy A. *et al.*, *XML Data Clustering: An Overview.* ACM Computing Survey 2011. 43(4):25
[2] Armbrecht R. *et al.*, *Knowledge Management in Research And Development* Research-Technology Management, 2001. 44(4):28-48
[3] Auer S. *et al.*, *Dbpedia: A nucleus for a web of open data.* Inter. Semantic Web Conf. (ISWC'07), 2007. Springer, pp. 722-735
[4] Batty M. *et al.*, *Smart cities of the future.* The European Physical Journal Special Topics 2012. 214(1): 481-518
[5] Becker H. *et al.*, *Learning similarity metrics for event identification in social media.* ACM Inter. Conf. on Web Search and Data Mining, 2010, 291-300
[6] Becker H., e.a., *Identifying content for planned events across social media sites.* ACM Inter. Conf. on Web Search & Data Mining, 2012, pp. 533-542
[7] Berners-Lee T. *et al.*, *The Semantic Web.* Scientific American, 2001. 284(5):1:19
[8] Boley H. and Chang E., *Digital Ecosystems: Principles and Semantics.* Digital EcoSystems and Tech. Conf.(DEST '07), 2007, pp. 398-403
[9] Briscoe G. and De Wilde P., *Computing of Applied Digital Ecosystems* The International ACM MEDES'09 Conf., 2009. pp. 28-35.
[10] Briscoe G., Sadedin S. and De Wilde P., *Digital Ecosystems: Ecosystem-Oriented Architectures.* Natural Computing, 2011. 10(3):1143-1194
[11] Budanitsky A. and Hirst G., *Evaluating WordNet-based Measures of Lexical Semantic Relatedness.* Computational Linguistics, 2006. 32(1): 13-47
[12] Cardoso J. and Bussler C., *Mapping between heterogeneous XML and OWL transaction representations in B2B integration.* Journal of Data & Knowledge Engineering, 2011. 70(12): 1046-1069
[13] Carlos R. *et al.*, *Ambient Intelligence the Next Step for Artificial Intelligence.* IEEE Intelligent Systems, 2008. 23(2): 15-18
[14] 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
[15] Chen M., *et al.*, *Data, Information and Knowledge in Visualization.* IEEE Computer Graphics and Applications, 2009. 29(1):12-19

[16] Collins C. J. and Smith K. G., *Knowledge Exchange and Combination: The Role of Human Resource Practices in the Performance of High-Technology Firms.* The Academy of Management Journal, 2006. 49(3):544-560

[17] Davies J., *et al.*, *Towards the Semantic Web.* Ontology-Driven Knowledge Management, 2003. J. Davies (Ed.), Wiley.

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

[19] Fong T. *et al.*, *Collaboration, Dialogue, and Human-Robot Interaction.* Inter. Symposium of Robotics Research, 2001. Lorne, Victoria, Australia.

[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] Greenwood S. *et al.*, *Social media update 2016: Facebook usage and engagement is on the rise, while adoption of other platforms holds steady.* Washington, DC: Pew Research Center. Retrieved from http://www.pewinternet.org/2016/11/11/social-media-update-2016, 2016

[22] Gruber. T., *Collective knowledge systems: Where the Social Web meets the Semantic Web.* Journal of Web Semantics, 2008. 6(1):4-13

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

[24] Gupta I. *et al.*, *Vit@MediaEval 2013 Social Event Detection task: Semantic structuring of complementary information for clustering events.* Proc. of MediaEval 2013 Workshop, 2013

[25] Gutierrez C. *et al.*, *Foundations of Semantic Web databases.* J. Comput. Syst. Sci. , 2011. 77(3):520-541

[26] Herold K., *An Information Continuum Conjecture.* Minds and Machines (Springer), 2003. 13(4): 553-566

[27] Herrera R. T. *et al.*, *Toward RDF Normalization.* Inter. Conf. on Conceptual Modeling (ER'15), 2015. pp. 261-275, Stockholm, Sweden

[28] Hoc J.M., *From Human–Machine Interaction to Human–Machine Cooperation.* Ergonomics, 2000. 43(7), pp.833-43

[29] Iannella R. and Waugh A., *Metadata: enabling the Internet.* DSTC Pty Limited, 1997, http://www.dstc.edu.au/RDU/reports/CAUSE97

[30] Inagaki T., *Smart Collaboration between Humans and Machines based on Mutual Understanding.* Annual Reviews in Control, 2008. 32(2):253-61

[31] Jain A. K. and Dubes R. C., *Algorithms for Clustering Data.* Prentice-Hall advanced reference series. NJ., 1988. pp. 320

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

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

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

[35] Kerzazi A. *et al.*, *A semantic mediation architecture for RDF data integration.* In Semantic Web Applications & Perspectives Conf., 2008, p. 3

[36] Khrouf H. and Troncy R., *Hybrid event recommendation using linked data and user diversity.* In Proceedings of the 7th ACM conference on Recommender systems. ACM, 2013. pp. 185-192

[37] Kidanu S. *et al.*, *A Multimedia-Oriented Digital Ecosystem: a New Collaborative Environment.* In Proc. of IEEE/ACIS 14th Inter. l Conference on Computer and Information Science (ICIS'15), 2015. pp. 411-416

[38] Kidd A., *The marks are on the knowledge worker.* In Proceedings of the SIGCHI conf. on Human factors in computing systems, 1994. pp. 186-191.

[39] Kotsiantis S. B., *Supervised Machine Learning: A Review of Classification Techniques.* Informatica, 2007. 31:249-268

[40] Levy P., *Collective Intelligence: Mankind's Emerging World in Cyberspace.* Basic Books, 1999. pp. 312

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

[42] Lim P.C. *et al.*, *Towards Semantic Clustering – A Brief Overview.* International Journal of Image Processing (IJIP), 2011. 4(6):557-566

[43] Ling C. and Abhishek R., *Event detection from flickr data through wavelet-based spatial analysis.* In Proceedings of the 18th ACM conference on Information and knowledge management, 2009. pp. 523-532

[44] Liu X. *et al.*, *Using social media to identify events.* In Proc. of the 3rd ACM SIGMM International Workshop on Social Media, 2011. pp. 3-8

[45] Liu X. *et al.*, *Finding media illustrating events.* In Proceedings of the 1st ACM International Conference on Multimedia Retrieval, 2011. p 58

[46] Ma Z. *et al.*, *Knowledge Adaptation for Ad-hoc Multimedia Event Detection with Few Exemplars.* ACM Inter. Conf. on Multimedia, 2012. pp. 469-478.

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

[48] McGuinness D.L. and Van Harmelen F., *OWL 2 Web - Ontology Language Document.* W3C Recomm., 2012. http://www.w3.org/TR/owl2-overview/

[49] McInerney C., *Journal of the American society for Information Science and Technology (Wiley Online Library).* Knowledge management and the dynamic nature of knowledge, 53(12): 1009-1018

[50] Mingers J. C., *An evaluation of theories of information with regard to the semantic and pragmatic aspects of information systems.* Systemic Practice and Action Research (Springer), 1996. 9(3): 187-209

[51] Mrissa M. *et al.*, *Context-based Semantic Mediation in Web Service Communities.* Weaving Services and People on the WWW, 2008. pp. 49-66

[52] Nguyen T.V. *et al.*, *Event Clustering and Classification from Social Media: Watershed-based and Kernel Methods.* MediaEval'13 Workshop, 2013

[53] Noubel J.F., *The Collective Intelligence, the Invisible Revolution.* The Transitioner, 2004

[54] Old L.J., *Information Cartography: Using GIS for visualizing non-spatial data.* In Proc. of ESRI Inter. Users' Conf.. San Diego, CA., 2002.

[55] Oren E. *et al.*, *Semantic Wikis for Personal Knowledge Management.* In Inter. Conf. on Database & Expert Systems Applications, 2006, pp. 509-518

[56] Papaoikonomou A. *et al.*, *A similarity-based chinese restaurant process for social event detection.* In Proc. of MediaEval 2013 Workshop, 2013

[57] Platt J. *et al.*, *PhotoTOC: Automatic clustering for browsing personal photographs.* Joint Conf. of the Fourth International Conference on Information, Communications and Signal Processing, 2003. pp. 6-10.

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

[59] Rafailidis D. *et al.*, *A Data-Driven Approach for Social Event Detection.* Working Notes of the MediaEval 2013 Workshop, 2013

[60] Randall D., Shrobe H. and Szolovits P., *What is a knowledge representation?* AI magazine, 1993. 14(1): 17

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

[62] Schank R. and Abelson R., *Scripts, Plans, Goals, and Understanding: An inquiry into Human Knowledge Structure.* Psychology Press, 1977. p. 256

[63] Schreiber G., *Knowledge Engineering and Management: the Common KADS Methodology.* MIT press, 2000

[64] Smriti S. *et al.*, *News Event Extraction Using 5W1H Approach & Its Analysis.* Inter. J. of Research in Eng. and Tech. (IJRET), 2013. (4) 5

[65] Sowa J. F. *et al.*, *Knowledge representation: logical, philosophical, and computational foundations.* MIT Press, 2000

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

[67] Stroka S., *Knowledge Representation Technologies in the Semantic Web.* Info. Tech. & Systems Mngmt., 2005. Salzburg Univ. of Applied Sciences

[68] Studer R. *et al.*, *Knowledge engineering: principles and methods.* Data and Knowledge Engineering (DKE), 1998. 25:161–197.

[69] Su Z., Kogan J. and Nicholas C., *Constrained Clustering with k-means Type Algorithms.* Text Mining: Applications and Theory, 2010. pp. 81-103

[70] Sutanto T. and Nayak R., *Admrg@ MediaEval 2013 social event detection.* In Proceedings of the MediaEval 2013 Workshop, 2013

[71] Taddesse F.G. *et al.*, *Semantic-based Merging of RSS Items.* World Wide Web Journal: Internet and Web Information Systems Journal, 2010. Vol. 12 (No. 11280), Springer Netherlands

[72] Takahiro I. *et al.*, *Information classification and navigation based on 5W1H of the target information.* Inter. conf. on Comput.l Ling., 1998, pp. 571-577

[73] 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 (IEEE TKDE), 2016. 28(6): 1383-1407

[74] Tekli J., *et al.*, *Semantic to intelligent web era: building blocks, applications, and current trends.* . Inter. MEDES'13 Conf., 2013. pp. 159-168

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

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

[77] Tekli J. *et al.*, *An Overview of XML Similarity: Background, Current Trends and Future Directions.* Computer Science Review, 2009. 3(3):151-173

[78] Teo C. and Gay R., *A Knowledge-driven Model to Personalize e-Learning.* Journal on Educational Resources in Computing (JERIC), 2006. 6(1): 3

[79] Terzi E. *et al.*, *Knowledge Representation, Ontologies, and the Semantic Web.* Asia-Pacific Web Conference (APWeb 2003), 2003. pp. 382-387

[80] Tong W. et al., *E-lamp: integration of innovative ideas for multimedia event detection.* Machine vision and applications (Springer), 2014. 25(1): 5-15

[81] Vakali A. *et al.*, *Smart Cities Data Streams Integration: 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

[82] W3C, *RDF 1.1 XML Syntax.* W3C Recommendation 2014. http://www.w3.org/TR/rdf-syntax-grammar/

[83] Wistuba M. and Lars S. T., *Supervised Clustering of Social Media Streams.* Working Notes of the MediaEval 2013 Workshop, 2013

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

[85] Zaharieva M. *et al.*, *Clustering and Retrieval of Social Events in Flickr.* Working Notes of the MediaEval 2014 Workshop, 2014

[86] 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