Proceedings of the Pakistan Academy of Sciences:
A. Physical and Computational Sciences 54 (1): 1–12 (2017)

ISSN: 2518-4245 (print), 2518-4253 (online)

Pakistan Academy of Sciences

Research Article

On Linked Open Data (LOD)-based Semantic Video Annotation Systems

Mumtaz Khan, Shah Khusro*, and Irfan Ullah

Department of Computer Science, University of Peshawar, Peshawar, 25120, Pakistan

Abstract: The advent of Web 2.0 and the rapid growth of video annotation systems have resulted in huge multimedia repositories where multimedia has become among the primary contents that are available on the Web. Annotating videos enable users in easily searching and retrieving multimedia contents on the Web. This practice also enables these systems to share multimedia contents as well. Annotations, if used properly, could be among the key factors in improving search efficiency, interoperability, video indexing and multimedia content analysis. However, user-generated annotations for multimedia content still remain inaccessible to the web of data. The available video-annotation systems provide format-dependent annotations in a proprietary manner. In addition, these annotations are just used within a single system and often cannot be reused, shared, linked, and explored by other communities. This paper aims at video searching problems in different traditional and currently available video-sharing web applications with their annotation tools and their limitations and shortcomings. In addition, this paper focuses on pointing out video searching problem in different ontology-based video-sharing web applications and video annotation systems. We are also investigating the distinguishing features of different LOD-based video-sharing web applications and LOD-based video annotation systems as well as focusing on new research trends to make it an access point for further readings.

Keywords: Multimedia, linked open data (LOD), annotations, video annotation, semantic video annotation systems, ontology

1. INTRODUCTION

The rapid growth of several video annotation systems including YouTube[1], Vimeo[2], Youku[3], Myspace[4], VideoANT[5], SemTube[6], and Nicovideo[7] have generated large amount of multimedia content on the Web that is frequently searched, browsed, and shared. These videos are important because of their multi-purpose usage in e-commerce, advertising, education, linguistic, entertainment, news, product reviews and so on. Similarly, video browsing, serialization, linking, sharing, categorization and filtering is essential in enabling users to locate videos that meet their needs and interests. However, the unstructured nature of videos makes all these aspects of video searching and browsing difficult. In order to handle this dilemma, video annotation systems have been proposed from time to time. However, the lack of a transparent integration with LOD and Semantic Web technologies, their usefulness is buried in the ocean of the huge data of the Web. This is why multimedia is still treated as foreign content to the Web [1]. In addition, although several annotation tools have been proposed, but they are limited in supporting collaborative video annotation in order to create a shared structured knowledge, which can be reused, shared, linked, browsed, queried, and explored by other communities [2, 3].

Most of these systems use high-level semantics in browsing and searching videos such as title, tags and caption etc. In addition, some domain-specific vocabularies have been used in annotations for referring to the defined agreed concepts [4]. However, this activity is highly time-consuming, expensive and needs high-level expertise. Therefore, the solution to all these mentioned problems is only

[1] http://www.youtube.com
[2] https://www.vimeo.com
[3] http://www.youku.com/
[4] https://myspace.com/
[5] http://ant.umn.edu/
[6] http://metasound.dibet.univpm.it:8080/semtube/index.html
[7] http://www.nicovideo.jp/

Received, January 2016; Accepted, March 2017
* Corresponding author: Shah Khusro; Email: khusro@upesh.edu.pk

possible if Semantic Web technologies especially ontologies and LOD are exploited in annotating, searching, browsing, and sharing video content on the Web. This way we may develop more meaningful video annotations to be exploited in organizing, linking, indexing, searching, browsing and sharing video content on the Web in an accurately, precisely, and user-friendly manner. Also, annotating videos or their specific parts such as objects, scenes, and events as well as their summarization based on related objects, scenes, events, and themes.

The available video annotation systems provide support for vocabularies and data schemas, but these annotations are just used within a single system and often cannot be merged, exchanged and explored by other communities. Researchers are trying to find ways for integrating the end-users' knowledge with current collections lying in the form of different data sources on the Web and for annotating web documents and multimedia content in order to improve interoperability and the quality of search. In this regard, various annotation frameworks and research contributions have been made including Annotea [5], LEMO [6], YUMA [7], M3O Ontology [8, 9], Annotation Ontology [10], and Open Annotation Collaboration [11, 12]. For the last few years, several standard vocabularies have been used to expose multimedia content on the Web. These include MPEG-7 [13], MPEG-7 Upper MDS [14], SMIL[8], SVG[9], W3C Media Annotation Working Group[10] (MAWG), COMM [15], VidOnt [16, 17], SWInto[11] [18], Music Ontology [19-21], W3C Media Fragment Working Group[12], IMDB [22-24], Soccer Ontology [25], and Video Movement Ontology [26]. However, these standards are used within single system and allow annotating domain-specific and format-dependent videos. In addition, these standards do not expose, share and interlink data with LOD. Therefore, Linked Data principles should be used in providing uniform access to such data [27]. These principles and rules can also be used in interlinking data among different sources on the Web. This way developers and web applications can easily reuse data in better ways to find new avenues of knowledge and information discovery and access.

This paper investigates the current state-of-the-art in video annotation systems with more emphasis to systems that exploit Semantic Web technologies including ontologies and LOD and identifies the potentials of LOD-based video annotation system in organizing, searching, browsing, and sharing, summarizing and interlinking videos on the Web based on related objects, scenes, events, and themes. Going along this direction, the paper also identifies some prominent issues and challenges, which if mitigated, can result into a global data space for video content on the Web. Rest of the paper is organized as Section 2reviews conventional video-annotation systems along with their shortcomings that led to adopting LOD in video annotation systems, Section 3 presents different annotation models and multimedia ontologies, Section 4 identifies the potential role of LOD in videoannotations, the current trends in LOD-based video annotation systems, and presents the available datasets for multimedia content available on LOD. Section 4 also contributes an evaluation framework by defining some evaluation metrics in order to compare the existing LOD-based video annotation systems and identify their limitations. Finally, Section 5 concludes our discussion and identifies some future directions.

2. CONVENTIONAL VIDEO-ANNOTATION SYSTEMS

The idea of LOD-based video annotations came into existence due to the inherent issues in conventional desktop and Web-based video annotation systems. Therefore, before discussing LOD-based video annotation systems in detail, it is necessary to present a bird-eye-view of these conventional systems along with their limitations and shortcomings. Desktop-based video annotation systems have been designed from the perspective of a particular domain and therefore, these systems address video-related problems of a particular group or community. On the other side, Web-based video annotation systems have wide applicability as they are accessed and used by users from all over the world. Web-based video annotation systems are also superior to their counterpart as they allow multiple users to annotate a video with unique concepts and point of views present in their minds. Conventional video annotation systems vary in annotation mechanism, user interface and annotation complexity from simple to complex. By simple we mean an annotation system that allows users only to upload and annotate video, and by complex we mean an annotation system that enables users to annotate videos or their specific objects, scenes, and events. Here, annotation approach can be either be

[8]http://www.w3.org/TR/smil/
[9]http://www.w3.org/TR/SVG/
[10]http://www.w3.org/2008/WebVideo/Annotations/
[11]http://smartweb.dfki.de/ontology_en.html
[12]http://www.w3.org/2008/WebVideo/Fragments/

manual, automatic, or semi-automatic [28]. In the following paragraphs, we present a brief investigation of conventional desktop and Web-based video annotation systems.

Several desktop-based video annotation systems are available enabling users to annotate videos or their specific objects, scenes, and events. For example, ANVIL [29, 30] annotates MPEG-1, MPEG-2, quick time, AVI, and MOV videos with descriptive, structural and administrative metadata that can be attached with a temporal segment, object or entire resource [29, 30]. Some desktop-based video annotation systems including ELAN[13] [31] also support full-text searching of videos based on their textual annotations. Similarly, Semantic Multimedia Annotation Tool (SMAT) annotates videos using MPEG-7 standard along with object recognition, tracing, arranging multimedia contents, configuring annotation session, visualizing annotations and reporting statistics. A very similar tool to SMAT is Semantic Video Annotation Suite (SVAS)[14] that annotates videos using MPEG-7 standard and adds descriptive and organizational metadata for searching and organizing videos. It facilitates annotations of specific object and scene and can relate similar objects [32].

Several desktop-based video annotation systems use Semantic Web technologies for organizing, indexing, and searching videos based on annotations. For example, Ont oELAN[15] extends ELAN by adding features including opening and displaying ontologies in Web Ontology Language (OWL) and developing language profiles for free-text and ontological annotations [33]. Video Annotation Tool[16] (VAT) and Video and Image Annotation[17] (VIA) annotate MPEG videos during live recording as well as frame by frame. Users are allowed to attach free-text and ontological annotations to specific regions and import OWL ontology files.

Desktop-based video annotation systems suffer from a number of limitations including complex user interfaces, slow and resource-consuming algorithms, limited coverage of video and audio formats, and lack of mechanisms for sharing annotated videos on the Web. Most of these systems cannot properly exploit annotations in organizing, indexing, and searching videos and lack in domain-level ontologies that could solve this issue. To the best of our knowledge, we found no desktop-based system that can annotate a video on specific object, scene, event and theme. The annotations generated by these systems could not be exposed and shared on the Web, which could be beneficial for other similar users, and therefore, adapting Web-based video annotation systems is required.

The Web-based video annotation systems facilitate users not only in accessing videos covering different aspects of life but also allow users to upload, share, annotate, and search videos based on these annotations. For example, YouTube is a well-known and largest video-annotation system that allows users to upload, share, annotate and search videos, where the uploader can annotate specific objects and events. Video fragments are expressed at the level of HTML pages containing videos. Similarly, VideoANT allows users to annotate YouTube videos on temporal basis. For correcting errors, a feedback is automatically generated and sent to the uploader and annotator of the video in order to remove errors, if any [17].

The Web-based video annotation systems also suffer from several issues, e.g., using temporal fragments of YouTube videos, a user cannot point to the specific event and limited to the annotating video at the page level of its web document [34, 35]. Searching specific objects, scenes, events and theme in the video is not supported. Annotations are not properly organized, which makes it difficult for the upload to detect flaws in the object, scene, event, and theme. The annotations cannot be linked to the specific video fragments. Also, objects and themes cannot be annotated with VideoANT. Similarly, Web-based video annotation systems lack in mechanisms for sharing their annotations on the Web and do not make use of LOD for annotation purposes. Moreover, like desktop-based video annotation systems, these systems also exploit different domains separately where data sources are not linked on the Web. Therefore, it is necessary to design and develop LOD-based video annotation systems by taking full advantage of the LOD-based video annotations, which are available on different data

[13]http://tla.mpi.nl/tools/tla-tools/elan
[14]http://www.joanneum.at/digital/produkte-loesungen/semantic-video-annotation.html
[15]http://emeld.org/school/toolroom/software/software-detail.cfm?SOFTWAREID=480
[16]http://mklab.iti.gr/project/vat
[17]http://mklab.iti.gr/via/

sources for organizing, indexing, linking, searching, browsing, and sharing videos based on related objects, scenes, events, and themes.

3. MULTIMEDIA ANNOTATION MODELS AND ONTOLOGIES

Annotations are understood and perceived in different ways. According to Simon et al. [7] annotations are common multidisciplinary practices that enable a scholar to organize, share and exchange knowledge with others about the source material. Annotations are additional information called metadata that are attached to a resource. This metadata can be used for a variety of purposes. Researchers are trying to integrate the knowledge of end-users with the available multimedia content that is available on different data sources on the Web and are trying to dig out how to annotate multimedia contents and improve the quality of video searching and interoperability. Therefore, several semantic annotation models have been developed and still efforts are going on.In this Section, we investigate the available existing annotation models.

Due to the wide applicability and usage of semantic video annotations, a number of annotation models and vocabularies have been developed and used in several multimedia standards and architectures. Annotea is the first W3C Semantic Web adopter for collaborative annotation [5]. Using client-server architecture, it enables users to annotate webpages. These annotationsare stored in RDF database and annotation queries are responded through HTTP front end. It uses Xpointer for locating the annotated documents and Xlink for interlinking documents with each other. RDF is used for describing and interchanging these annotations. However, it suffers from several limitations, e.g., annotations are limited to webpages; limited or weak annotations of multimedia objects; requiring clients to be aware of Annotea-specific protocol; and ignoring the dynamic status of webpages.

Open Annotation Collaboration (OAC) [11, 12] is an open annotation model that annotates audios, videos, images, and webpages and allows sharing annotations across different data sources on the Web. In addition, it supports direct addressing of fragments that allows users to annotate the same fragment. Furthermore, the model provides support for structured annotation bodies and overlaying semantic description related to one or more annotation targets.

LEMO [6] supports all types of annotations and uses linked data principles. It uses MPEG-21 fragment URI for media fragment identification and supports only MPEG media types. However, it has complex and ambiguous media fragment URI syntax when compared to W3C media fragment URIs. In [36], the annotations of LEMO media fragments are linked with LOD cloud. An extension of this model is YUMA [7] that uses Open Annotation Collaboration (OAC) model in combination with LEMO. It is an open annotation model for multimedia objects on the Web and annotates digital objects or specific parts of the digital object and publishes annotations through linked data principles.

Multimedia Metadata Ontology (M3O) is used in combination with several metadata models standards support semantic annotations for complex types of multimedia content [8, 9]. It uses Scalable Vector Graphics (SVG) and Synchronized Multimedia Integration Language (SMIL) for the integrating annotations with rich media representation. M3O uses annotation patterns that formally express annotations, which can be assigned to arbitrary information entities. It also fills the gap of structure metadata models and metadata standards such as XMP, JEITA, MPEG-7 and EXIF.

Annotation Ontology (AO) [10] is an OWL open annotation that enables the online annotations of scientific documents such as web documents, images and their fragments. It is technically similar to Open Annotation Collaboration (OAC) model but differs from OAC in terms of fragment annotations, representation of constraints and constraint targets as first-class resources. It also provides convenient ways for encoding and sharing annotations in FRD format.

Linked Media Framework (LMF) extends the basic Linked Data principles to Linked Media principles and concerns about media resources [37]. LMF is Linked Data server used for annotating videos, storing metadata, indexing, searching and browsing the multimedia content. However, media fragments and their annotations are not supported. In addition, rendering of media annotations have not been properly exploited.

Different multimedia standards including MPEG-7 and Synchronized Multimedia

Integration Language (SMIL) also incorporate features of semantic video annotations. However, these standards use non-URI based mechanisms for fragment identification. Also the descriptions of temporal and spatial media content are divided into multi-dimensions. Therefore, media fragments are not represented by a single URI. MPEG-21 defines normative URIs for fragment identification in MPEG compatible files. However, syntax of MPEG-21 for fragments identification is difficult, ambiguous and format dependent.

Several state-of-the-art ontologies are available for describing multimedia content. For example, Ontology for Media Resources[18] 1.0 (ORM) [38] is a core ontology for describing multimedia content. It is a collection of different descriptions of media resources for supporting a core set of properties of media resources. In addition, it uses hash URI mechanism for fragment identification from temporal, spatial, track and named dimension [38]. However, it suffers from issues including: dereferencing of media fragments; aligning legacy metadata standards and methods to interlink multimedia content using ORM 1.0 [39].

The music ontology [19, 20] annotates audio-related data such as the artist, albums, tracks and characteristics of business-related information about the music. This ontology uses existing ontologies including FRBR final report, eventontology, timeline ontology, ABC ontology from the Harmony project, and the FOAF project. The expressiveness of annotations is defined at three levels including: (i) support for information about tracks, artists and releases; (ii) support for vocabulary about the music creation workflow such as composition, arrangement, and performance recording; and (iii) providing vocabulary about decomposing complex events like e.g., the performance of a particular artist in the event etc. It contains 141 classes, 260 object properties, 131 data type properties and 86 individuals. Similarly, SWInto[19] [18], Soccer Ontology [25], Video Movement Ontology (VMO) [26] are other domain-specific ontologies.

LOD is one of the most important and increasingly adopted ways for publishing, sharing, and interlinking data resources on the Web. The RDF standard format links and integrates the former proprietary data on LOD. In the last few years, a huge collection of multimedia content has been generated that can be seen as a globally linked and distributed data space. In this regard, annotations can play a critical role to efficiently manage, share, reuse, retrieve, and organize multimedia content on LOD. In the Section 4, we present state-of-the-art in LOD-based video annotation systems by investigating their need, current trends and the datasets for multimedia content that are available on LOD.

4. STATE-OF-THE-ART IN LOD-BASED VIDEO ANNOTATION SYSTEMS

Multimedia content has become one of the primary content of the Web [1]. Today, users can bookmark, annotate, upload, search, browse and share videos on popular video annotation and video sharing web applications like YouTube etc. However, because of the huge size and unstructured nature, it becomes difficult to properly organize, index, interlink, browse, search, share and summarize video content based on related objects, scenes, events, and themes as in the conventional systems there is no support for videos to be interlinked and share though video annotations in forming a global (LOD-based) data space of videos. Therefore, exploiting Semantic Web technologies especially ontologies and LOD in designing video annotation systems could enable us to not only browse and search videos but also interlink, summarize and share them based on related objects, scenes, events, and themes. In this Section, we are trying to expose this potential role of LOD-based video annotation systems along with identifying current trends, future directions, and datasets available for multimedia content especially videos on the Linked Open Data.

4.1 The Need for LOD-based Video Annotations

The LOD-based video annotation systems have potential applications in a number of domains, which can be understood with the help of some example scenarios. Suppose a politician discusses the issue on human rights in a specific scene or event in a video available on one data source while, on the other side he/she discusses the same issue in another video that is available on some the other data source. Such related scenes or events in these videos can be utilized through LOD in interlinking these videos enabling users to browse

[18]http://www.w3.org/TR/mediaont-10/
[19]http://smartweb.dfki.de/ontology_en.html

or search videos for related objects, scenes, events and themes. One possible application of this scenario is in talk shows where multiple videos on the same issue can be interlinked in order to produce the true picture of the motive of the politician on a particular issue of political interest. Take another example, where a research scholar wants to search and browse video lectures of a particular professor or other researchers in order to get the basic knowledge/opinions of researchers on the topic of interest. In such situations, LOD-based video annotation systems could exploit the available annotations in relating and retrieving relevant videos based on related objects, scenes, events or according to a specific theme. Developing such systems will help the researchers in navigating through similar and cross-domain resources and datasets and will help them in establishing links among concepts, problems and possible solutions by taking full advantage of annotative and social applications' datasets that are linked and available on LOD. According to Hausenblas et al. [39] applying Linked Data principles on media fragments will make these fragments globally identified through URIs which will also facilitate their linkage to global data through LOD-based annotations. Hence, it will allow in better organization, indexing, searching, browsing, and sharing multimedia resources [39].

Li et al. [40] raised three questions using Linked Data principles on media fragments and annotation. The questions are: how to find media fragments through URI? How to display the appropriate representations while dereferencing the media fragments' URIs in different context? And how to mitigate problems related to aligning ontologies related to annotations and the annotated media fragments? [40]. To answer these questions, several LOD-based video annotation systems have been developed in order to interlink the annotations of multimedia resources across different repositories and to achieve better organization, indexing, browsing and searching. These tools are presented in Section 4.2.

4.2 Current Trends in LOD-based Video Annotation Systems

A number of LOD-based video annotation systems have been developed and used including e.g., LUCERO [41, 42], KMI[20], NoTube[21], YUMA [7], (ECMAP)[22], Project Pad[23], Synote[24], SemTube [4, 43], Connect ME [44, 45], MyStoryPlayer [46], SemVidLOD[25], and SemWebVid [47], etc. This Section reviews these video annotation systems in detail.

LUCERO [41, 42] is an LOD-based project developed by Joint Information Systems Committee (JISC) for exposing and connecting educational institutes and research material on the Web. The project has to hand institutional repositories containing educational and research resources and uses set of tools for to extract RDF data from these resources, load this RDF data into a triple store and expose it through the Web for the purpose of interlinking resources through LOD [41, 42].

KMI is an LOD-based annotation tool from Department of Knowledge Media Institute, Open University[26]in order to annotate educational material that is produced by different educational resources from Open University including online teaching facilities like course forums, multi-participant audio environments for language and television programmes on BBC. This tool enables users to annotate video with Linked Data sources consequently navigating them and enriching them with additional materials. In addition, it uses SugarTube[27] browser to search the annotated videos and explores related content through the Linked Data resources [48]. However, it does not annotate theme and specific object in the videos.

NoTube aims to interlink traditional TV environment to the Web contents through LOD for providing an enhanced and more personalized TV experience, e.g., automatic recommendations of different TV programs and personalized advertisements based on the preferences of the individuals by securely linking to the user personal data available on the social Web using Linked Data principles. For establishing links between TV content and the Web, NoTube uses the alignment of existing vocabularies and thesauri, interoperability of content metadata, user profiling, content filtering, and enriching metadata. The project tries to (implicitly) connect the passive TV-related user activities to the

[20]http://annomation.open.ac.uk/annomation/annotate
[21]http://notube.tv/
[22]http://dme.ait.ac.at/annotation/
[23]http://dewey.at.northwestern.edu/ppad2/index.html
[24]http://www.synote.org/synote/
[25]http://vidont.org/semvidlod/
[26]http://www.kmi.open.ac.uk/
[27]http://sugartube.open.ac.uk/

dynamic activities e.g., rating, tagging, sharing, etc., in order to reuse and integrate this Social Web data with user TV experiences through LOD in making personalized recommendations of TV programs [49, 50].

YUMA is led-based open-source and open annotation framework for multimedia objects. It is an extended form of LEMO annotation model. It provides integrated collaborative annotation of multimedia collections of a digital library. The YUMA framework annotates images, maps, audio and video and uses OAC annotation model for interoperability. It also provides semantic enrichment, a method that allows users to easily augment annotations with links that are contextually relevant resources on the Web [7].

EUROPEANA Connect Media Annotation Prototype (ECMAP) [36] annotates videos with bibliographic information on spatial and temporal basis using free-text as well as Geo Names and DBpedia. The annotations are further enriched using semantic tagging and LOD principles[36]. A very similar web application to ECMAP is SemTube video-annotation web application, which aims to develop Model-View-Controller-based configurable video annotation system that can be easily pugged in and integrated with other similar systems/web applications for annotating digital objects with meaningful metadata. This way, Semantic Web technologies are used in enhancing the current state of digital libraries, where the focus is to overcome challenges in searching and browsing videos through the effective linkage of resources and interoperability through the LOD principles. SemTube, provides a collaborative annotation framework using RDF, media fragment URI and Xpointer and pluggable with other ontologies [4, 51].

ProjectPad is a collaborative video annotation web application developed for research, teaching and distance learning, and for making online notebook of the annotated media segments. The set of tools in Project Pad allow users to organize, browse, and search rich media and collect digital objects in presenting selected parts, descriptions, parts, and annotations on the Web. Similar to ProjectPad, KMI[28], an LOD-based annotation tool, annotates educational resources of course forums and audio/video environments for BBC's language and TV programs. Users are allowed to annotate videos as well as search and browse video-related information through LOD and related technologies [48].

Synote [52, 53], is a Web-based multimedia annotation system for publishing multimedia fragments and user-generated content, i.e., multimedia annotations through the principles of Linked Data. Itallows for synchronized bookmarking, comments, synmarks, and notes attached to audio and video recordings whereas transcripts, images, and slides are exploited in finding and replaying audio/video recordings. It improves online discovery of media fragments, use annotation for indexing so that search engine can easily find out such media fragments. It manually embeds RDF a in Synmarks Note and RDF content editor such as RDFaCE and triples in RDFa are published along with media fragments [53].While watching and listening to the lectures, transcripts and slides are displayed alongside. Browsing and searching for transcripts, synmarks, slide titles, notes and text content are also supported.

ConnectME a nationally funded project in Austria aiming to develop a hypermedia platform based on open Web standards for delivering interactive video experience and web services with support for annotating videos with concepts, providing Web-based linkage among concepts and contents, and on-the-fly augmenting videos content by taking into account the aspects of personalization and contextualization. Under the umbrella of this project, a Web-based hypervideo annotation tool has been developed that annotates videos on spatial and temporal basis using free text as well as DBPedia concepts. It also supports searching for geographic locations in GeoNames [44, 45].

MyStory Player is a video player allows annotating multi-angle videos especially those in education domain. Its user interface allows users in interactively using their provided annotations in analysing actions, gestures, and postures focusing the formal representation of relationships in RDF among depicted elements. It powers the European e-Library website for performing arts, and general metadata such as title, technical metadata such as duration, and timestamp-based data to be used in annotating human dialogues, presentations, and video events [46].

[28]http://annomation.open.ac.uk/annomation/annotate

SemVidLOD uses terms from LOD cloud in semantically enriching video resources, files, and streaming media with high-level descriptions. It uses VidOnt ontology for representing technical, licensing, and administrative metadata with high-level RDF descriptions of the content.

SemWebVid is an AJAX-based Web Application that automatically generates RDF descriptions for YouTube videos by processing the manually added tags and closed captions. It makes use of natural language processing APIs to analyse the descriptors and map the results to LOD concepts using DBPedia, Uberblic, Any23, and rdf: about APIs [47].

By looking into the state-of-the-art literature, the existing LOD-based video annotation systems are limited have some limitations such as lacking support for annotating a specific object, scene, event and theme of the video through LOD as well as the linking related objects, scenes, events, and themes, which are available on different data sources. In addition, searching relevant videos based on related objects, scenes, events, and themes is difficult and challenging task. For example, there is no support of theme-based annotation in ECMAP and searching videos by related themes is not available; Project Pad has no support for searching videos based on specific objects, scenes, events, and themes in a video with no relationships among video annotations. Similarly, KMI, SemTube, ConnectME, Synote, SemWebVid, SemVidLOD, and MyStoryPlayer do not support theme-based video annotations, inter-linking videos as well as browsing and searching related objects, events, scenes, and themes. By carefully, critically and analytically reviewing the state-of-the-art in video annotation systems, we have tried to identify the requirements and features of LOD-based video annotation systems in Table 1, and developed an evaluation framework that evaluates and compares the available LOD-based video annotation systems and tools, shown in Table 2.

4.3 Datasets for Multimedia Contents available on LOD

In this Section, we identify the datasets developed for multimedia content available on Linked Open Data developed for different domains. Each data set has their own classes and properties, which are defined under specific requirements and purpose and explains the facts and knowledge of the corresponding domain.

a. BBC Programs[29]containsinformation about TV and radio programs broadcasted by BBC with 60,000,000 triples, with 12,666 in-links, and 33,237out-links.
b. BBC Music[30] ontology contains information about music such as artists, albums, tracks, performances, and arrangements. Itcontains 20,000,000 triples with 11,009,200 in-links and 23,000out-links.
c. BBC Wildlife Finder[31] contains information about wildlife biota, habits, adaptations, video clips and photos. It contains 23861 triples, 318 in-links and 2373 out-links.
d. DBtune[32] is a collection of music-related data sets, which are exposed as part Linked Open Data. It includes, amongst others, MusicBrainz, AudioScrobbler, and data extracted from the MySpace social web application and uses the music ontology and uses an online agent named Henry[33] for performing signal analysis of media resources found on the Web.
e. EventMedia[34] contains information about media events. It contains 36274454 triples.
f. Linked Movie Database[35] contains information about movies and contains 6148121 triples, 1,883 in-links, and 162,756 out-links. This dataset is published onLOD through D2R Server. It provides contents from the sources including Wikipedia, FreeBase, and GeoNames.
g. EUROPEANA Linked Open Data[36] contains information about photos, video clips gathered by EUROPEANA. It contains 2.4 million triples.

These datasets are freely available on the LOD cloud and every one can easily use its classes and properties. Unfortunately, there is limited number of datasets for multimedia contents on the cloud. Similarly, due to lack of appropriate exploitation of Semantic Web technologies, users are unable to easily use these datasets for annotations in the LOD-based video annotation systems.

[29]http://thedatahub.org/dataset/bbc-programmes
[30]http://thedatahub.org/dataset/bbc-music
[31]http://thedatahub.org/dataset/bbc-wildlife-finder
[32] http://dbtune.org
[33] http://dbtune.org/henry
[34]http://thedatahub.org/dataset/event-media
[35]http://thedatahub.org/dataset/linkedmdb
[36]http://pro.europeana.eu/linked-open-data

Table 1. Features of LOD-based video annotation tools / projects.

Features	Possible Values
Annotation depiction	HTTP-derefrenceble RDF document, Linked Data (LD), Linked Open Data(LOD), Embedded in content representation
Annotation target object type	Web documents, Multimedia objects, Multimedia and web documents
Vocabularies used	RDF/RDFS, Media Fragment URI, Open Annotation Collaborative (OAC), Open Archives Initiative Object reuse and Exchange (OAI-ORE), Schema.org, LEMO, Friend of a Friend (FOAF), Dublin Core (DC), Timeline, Simple Knowledge Organization System (SKOS), W3C Media ontology (W3C MO), Bibliography ontology (Bibbo), Course and AIISo Ontology (CAIISo), Creative Commons Rights (CCR), Expression Vocabulary and Nice Tag Ontology (EVNTO), Sioc Ontology, WP1, WP2, WP3, WP4, WP5, WP6, WP7a, WP7b, WP7c, Basic Geo (WGS84)
Flexibility	Yes, No
Annotation type	Text, Drawing tools, public, private
Definition languages	RDF/RDFS, OWL
Media fragment identification	Xpointer, Media fragment URI 1.0 (MF URI 1.0), MPEG-7 fragment URI, MPEG-21 fragment URI, N/A

Table 2. Feature analysis and summaries of LOD-based semantic video annotation tools/projects.

	Features	Annotation Depiction	Annotation target object	Vocabularies	Flexibility	Annotation Type	Definition Languages	Media Fragment Identification	Browsing, Searching Scene (S), Event (E), Object (O), and Theme (T)	Summarizing related videos based on Scene (S), Event (E), Object (O), and Theme (T)
Projects & Tools	EUROPEANA Connect	LOD	Multimedia and web documents	OAC, LEMO	Yes	Text, Drawing tools, public, private	RDF/RDFS	Xpointer, MPEG-21 fragment URI	Nil	Nil
	SemTube Annotation Tool	LOD	Multimedia objects	RDF/RDFS, OAC	Yes	Text, Drawing tools	RDF/RDFS	Xpointer	Nil	Nil
	YUMA Annotation Framework	LOD	multimedia and web documents	OAC, LEMO	Yes	Text, Drawing tools, public, private	RDF/RDFS	Xpointer, MF URI 1.0	Nil	Nil
	KMI Annomation tool	LOD	Multimedia objects	FOAF, DC, Timeline, SKOS	Yes	Text	RDF/RDFS, OWL	Xpointer, MF URI 1.0	Nil	Nil
	LUCERO	LOD	Multimedia and web documents	FOAF, DC, Timeline, SKOS, W3C MO, Bibbo, CAIISo, CCR, EVNTO, Sioc ontology	Yes	Text, drawing tools, public, private	RDF/RDFS	Xpointer, MF URI 1.0	Nil	Nil
	NoTube	LOD	Multimedia and web documents	WP1, WP2, WP3, WP4, WP5, WP6, WP7a, WP7b, WP7c	Yes	Text, drawing tools, public	RDF/RDFS	Xpointer	Nil	Nil
	Synote	LOD	Multimedia objects	MFURI 1.0,OAC,OAI-ORE,Schema.org	Yes	Text, Private, Public	RDF/RDFa	Xpointer	Nil	Nil
	ConnectME	LOD	Multimedia objects	OAC, RDF, MFURI 1.0	Yes	Text	RDF	MFURI 1.0	Nil	Nil
	MyStoryPlayer	LOD	Multimedia objects	DC, OAC, RDF	Yes	Text	RDF	MFURI 1.0	Nil	Nil
	SemVidLOD	LOD	Multimedia objects	RDF, MFURI 1.0	Yes	Text	RDF	MFURI 1.0	Nil	Nil
	SemWebVid	LOD	Multimedia objects	RDF	Yes	Text	RDF	Xpointer	Nil	Nil

5. CONCLUSIONs AND RECOMMENDATIONS

The conventional video annotation systems have resulted in the production and consumption of huge collection of video content and video annotations that are frequently browsed, searched, and retrieved on the Web. However, these systems do not support sharing and linking the annotated objects, scenes, events in video and linking related videos on thematic basis on Linked Open Data in order to provide a global data space of videos. In addition, the annotation data is just used within their corresponding systems and are not shared and used by other systems. In order to be able to share and use the annotation data, researchers are using Semantic Web technologies in annotating media content and applying LOD concepts on annotated media fragments so that annotations as well as annotated media fragments can be indexed, searched, exposed, and linked to global data sources. However, the state-of-the-art research and development is not mature enough to properly use annotations in searching, reusing and interlinking annotated media fragments, scenes, objects, and themes with global data sources. Moreover, the available LOD-based video annotation systems are limited in several ways because of complex their user interfaces and limited use of Semantic Web technologies as well as the limitations in the available datasets. Similarly, datasets are available for few domains including education, e-Commerce, and news etc., and other domains such as politics etc., are yet to be covered. If such datasets are appropriately developed, organized, and linked, then users could easily annotate and link videos based on related objects, scenes, events, and themes. This way we could be able to provide a more enhanced and user-friendly video searching, browsing, and sharing experience to the users of LOD-based video annotation systems.

6. REFERENCES

1. Van Deursen, D., W.V. Lancker, E. Mannens, & R. Van de Walle. Experiencing standardized media fragment annotations within HTML5. *Multimedia Tools and Applications* 70: 827-846 (2014).
2. Dasiopoulou, S., E. Giannakidou, G. Litos, P. Malasioti, & Y. Kompatsiaris. *A Survey of Semantic Image and Video Annotation Tools*. Springer, p. 196-239 (2011).
3. Khan, M., S. Khusro, & I. Ullah. On Annotation of Video Content for Multimedia Retrieval and Sharing. *International Journal of Computer Science and Information Security* 14:198-218 (2016).
4. Grassi, M., C. Morbidoni, & M. Nucci. *Semantic Web Techniques Application for Video Fragment Annotation and Management*. Springer, p. 95-103 (2011).
5. Kahan, J., M.-R. Koivunen, E. Prud'Hommeaux, & R.R. Swick. Annotea: an open RDF infrastructure for shared Web annotations. *Computer Networks* 39: 589-608 (2002).
6. Haslhofer, B., W. Jochum, R. King, C. Sadilek, & K. Schellner. The LEMO annotation framework: weaving multimedia annotations with the web.*International Journal on Digital Libraries* 10:15-32 (2009).
7. Simon, R., J. Jung, & B. Haslhofer. The YUMA media annotation framework. Springer, p. 434-437(2011).
8. Saathoff, C., & A. Scherp. M3O: the multimedia metadata ontology. In: *Proceedings of the Workshop on Semantic Multimedia Database Technologies (SeMuDaTe 2009)*, Graz, Austria (2009).
9. Saathoff, C., & A. Scherp. Unlocking the semantics of multimedia presentations in the web with the multimedia metadata ontology.In *Proceedings of the 19th International Conference on World Wide Web*., p. 831-840 (2010).
10. Ciccarese, P., M. Ocana, L. J. Garcia Castro, S. Das, & T. Clark. An open annotation ontology for science on web 3.0. *Journal of Biomedical Semantics* 2: S4 (2011).
11. Bargeron, D., A. Gupta, & A.B. Brush. A common annotation framework.*Rapport technique MSR-TR-2001-108 from Microsoft Research, Redmond, US* 3:3 (2001).
12. Haslhofer, B., R. Simon, R. Sanderson, & H. Van de Sompel. The open annotation collaboration (OAC) model. In *2011 Workshop on Multimedia on the Web (MMWeb)*. p. 5-9 (2011).
13. Martinez, J.M. *MPEG-7 Overview (version 10)*. Technical Report. *ISO, IEC JTC1/SC29/ WG112004* (2004).
14. Hunter, J. Adding multimedia to the semantic web: Building and applying an MPEG-7 ontology. In: *Multimedia Content and the Semantic Web: Methods, Standards and Tools*. Wiley Online Library (2005).
15. Arndt, R., R. Troncy, S. Staab, & L. Hardman. COMM: A Core Ontology for Multimedia Annotation. Springer, p. 403-421 (2009).
16. Sikosa, L.F. *Multimedia Ontologies.* http://www.lesliesikos.com/multimedia-ontologies/
17. Hosack, B. VideoANT: Extending online video annotation beyond content delivery. *Tech Trends*. 54: 45-49 (2010).
18. Oberle, D., A. Ankolekar, P. Hitzler, P. Cimiano, M. Sintek, M. Kiesel, B. Mougouie, S. Baumann,

S. Vembu, & M. Romanelli. DOLCE ergo SUMO: On foundational and domain models in the SmartWeb Integrated Ontology (SWIntO). *Web Semantics: Science, Services and Agents on the World Wide Web* 5: 156-174 (2007).
19. Raimond, Y., F Giasson, K. Jacobson, G. Fazekas, T. Gängler, & S. Reinhardt. music ontology specification. *Specification Document.* Published on November 28, 2010. http://purl.org/ontology/mo/
20. Raimond, Y., S. Abdallah, M. Sandler, & F. Giasson. The music ontology. In: *Proceedings of the International Conference on Music Information Retrieval*, 2007. p. 417-422 (2007).
21. Kanzaki, M. *Music Vocabulary*. Published on July 04, 2003. http://www.kanzaki.com/ns/music
22. Bouza, A. MO - the Movie ontology. Published on January, 2010. http://www.movieontology.org/documentation/
23. Mitra, R.M.S., S. Basak, & S. Gupta. *Movie Ontology*. http://intinno.iitkgp.ernet.in/courses/1170/wfiles/243055
24. Avancha, S., S. Kallurkar, & T. Kamdar. Design of ontology for the internet movie database (IMDb). *Semester Project CMSC* 771 (2001).
25. Da Costa Ferreira, L.A.V., & L.C. Martini. Soccer ontology: A tool for semantic annotation of videos applying the ontology concept. In: *Proceedings of IADIS International Conference on e-Society,* Lisbon, Portugal, July 3-6, 2007. http://www.iadisportal.org/e-society-2007-proceedings
26. De Beul, D., S. Mahmoudi, & P. Manneback. An ontology for video human movement representation based on Benesh notation. In: *2012 International Conference on Multimedia Computing and Systems (ICMCS).* p. 77-82 (2012).
27. Bizer, C., T. Heath, & T. Berners-Lee. Linked data-the story so far.*International Journal on Semantic Web and Information Systems (IJSWIS)* 5:1-22 (2009).
28. Potnurwar, A.V. Survey on Different Techniques on Video Annotations. *International Journal of Computer Applications* (2012).
29. Kipp, M. ANVIL - A generic annotation tool for multimodal Dialogue. In: *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech)* (2001).
30. Kipp, M. Spatiotemporal coding in anvil. In:*Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC-08* (2008).
31. Sloetjes, H., A. Russel, & A. Klassmann. ELAN: A Free and Open-Source Multimedia Annotation Tool. In 2007 *Prococeedings of 8th International Speech Communication Association (INTERSPEECH 2007)*, p. 4015-4016 (2007).
32. Schallauer, P., S. Ober, & H. Neuschmied. Efficient semantic video annotation by object and shot re-detection. In: *2nd International Conference on Semantic and Digital Media Technologies (SAMT) Posters and Demos Session, Koblenz, Germany* (2008).
33. Chebotko, A., Y. Deng, S. Lu, F. Fotouhi, A. Aristar, H. Brugman, A. Klassmann, H. Sloetjes, A. Russel, & P. Wittenburg. OntoELAN: An ontology-based linguistic multimedia annotator. In:*Proceedings of IEEE Sixth International Symposium on Multimedia Software Engineering, 2004,* p. 329-336 (2004).
34. Van, W., Lancker, D. Van Deursen, E. Mannens, & R. Van de Walle. Implementation strategies for efficient media fragment retrieval. *Multimedia Tools and Applications.* 57: 243-267 (2012).
35. YouTube Video Annotations. http://www.youtube.com/t/ annotations_about
36. Haslhofer, B., E. Momeni, M. Gay, & R. Simon. Augmenting Europeana Content with Linked Data Resources. In:*Proceedings of the 6th International Conference on Semantic Systems*, 2010, p. 40 (2010).
37. Kurz, T., S. Schaffert, & T. Burger. LMF: A framework for linked media. In:*2011 Workshop on Multimedia on the Web (MMWeb),* p. 16-20 (2011).
38. Lee, W., W. Bailer, T. Bürger, P. Champin, V. Malaisé, T. Michel, F. Sasaki, J. Söderberg, F. Stegmaier, & J. Strassner. Ontology for media resources 1.0. W3C recommendation. *World Wide Web Consortium* (2012).
39. Hausenblas, M. Interlinking multimedia: how to apply linked data principles to multimedia fragments. Linked Data on the Web. In: *Workshop (LDOW 09) in conjunction with 18th International World Wide Web Conference (WWW 09)* (2009).
40. Li, Y., M. Wald, & G. Wills. Interlinking multimedia annotations. In: *Web Science 2011, Koblenz, Germany, 14 - 18 Jun 2011* (2011).
41. d'Aquin, M., F. Zablith, E. Motta, O. Stephens, S. Brown,S. Elahi, &R. Nurse. *The LUCERO Project.Linking Universrity Content for Education and Research Online*. http://lucero-project.info/lb/about
42. Fernandez, M., M. d'Aquin, & E. Motta. Linking data across universities: an integrated video lectures dataset. Springer, p. 49-64 (2011).
43. Morbidoni, C., M. Grassi, M. Nucci, S. Fonda, & G. Ledda. Introducing the Semlib Project: Semantic Web Tools for Digital Libraries. In:*Proceedings of the 1st international workshopon on Semantic Digital Archives(SDA)*, 2011. p. 97-108 (2011).
44. Nixon, L.J., M. Bauer, & C. Bara. Connected Media Experiences: interactive video using Linked Data on the Web. In:*proceeding of the WWW2013 Workshop on Linked Data On The Web*(*LDOW)*, 2013 (2013).

45. Nixon, L.J., M. Bauer, C. Bara, T. Kurz, & J. Pereira. ConnectME: Semantic Tools for Enriching Online Video with Web Content. In: *2012 I-SEMANTICS (Posters & Demos)*. 932: 55-62 (2012).
46. Bellini, P., P. Nesi, & M. Serena. MyStoryPlayer: experiencing multiple audiovisual content for education and training. *Multimedia Tools and Applications* 74: 8219-8259 (2015).
47. Steiner, T. SemWebVid-making video a first class semantic web citizen and a first class web Bourgeois. In: *Proceedings of the 2010 International Conference on Posters & Demonstrations Track* 658: 97-100 (2010).
48. Lambert, D., & H. Q. Yu. Linked Data based video annotation and browsing for distance learning. In: *SemHE 2010, Southampton, UK* (2010).
49. Schopman, B., D. Brickly, L. Aroyo, C. van Aart, V. Buser, R. Siebes, L. Nixon, L. Miller, V. Malaise, & M. Minno. NoTube: making the Web part of personalised TV. In: *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line, April 26-27th, 2010, Raleigh, NC, USA* (2010).
50. Aroyo, L., L. Nixon, & L. Miller. NoTube: the television experience enhanced by online social and semantic data. In: *2011 IEEE International Conference on Consumer Electronics-Berlin (ICCE-Berlin),* 2011. p. 269-273 (2011).
51. Morbidoni, C., M. Grassi, M. Nucci, S. Fonda, & G. Ledda. Introducing SemLib project: Semantic web tools for digital libraries. In: *International Workshop on Semantic Digital Archives-Sustainable Long-term Curation Perspectives of Cultural Heritage Held as Part of the 15th International Conference on Theory and Practice of Digital Libraries (TPDL), Berlin*, 2011 (2011).
52. Li, Y., M. Wald, G. Wills, S. Khoja, D. Millard, J. Kajaba, P. Singh, & L. Gilbert. Synote: Development of a web-based tool for synchronized annotations. *New Review of Hypermedia and Multimedia* 17: 295-312 (2011).
53. Li, Y., M. Wald, T. Omitola, N. Shadbolt, & G. Wills. Synote: Weaving media fragments and linked data. In: *Linked Data on the Web (LDOW2012), Lyon, FR*, 16 Apr 2012, p. 10 (2012).