Distributed Semantic Video Tagging For Peer-to-Peer Authoring System

Giuseppe Rizzo, Biagio Meirone, Pierluigi Di Nunzio, Federico Di Gregorio
Dipartimento di Automatica ed Informatica,
Politecnico di Torino
c.so Duca degli Abruzzi, 24
10129, Torino, Italy
[giuseppe.rizzo|biagio.meirone|pierluigi.dinunzio|federico.digregorio]@polito.it

***Abstract*—In the context of a project aimed at the development of a cooperative and distributed Authoring System to facilitate the production, sharing, and integration of new and existing content from various domains, this work addresses problems related to locating, annotating and reusing audio and video streams distributed over the Internet. The Authoring System allows partial annotations (applied to a part and not to the whole document) and time-based tagging of audio and video streams (where the semantic information is applied only to portions of the stream) and provides tools to extract partial contents and mixes them, creating new resources. These resources are represented with an envelope composed of many references that point to other contents (whole or portions of) within the Internet. This envelope holds only references where multimedia portions, called fragments, are located and they are stored as a meta-data. In such a way, the Authoring System prevents multimedia element duplications. Both original and newly created content are delivered to clients using the state of the art streaming technology that is integrated into the Authoring System to provide a uniform interface both to consume and produce knowledge.**

***Keywords*-P2P authoring system; collaborative authoring system; semantic media editing; semantic media annotation**

I. Introduction

A key focus of the Authoring System is the possibility to share information and to capture common interests among users, merging them in a new user-experience. In such a scenario, the Authoring System can locate, annotate and integrate pieces of information in order to show new information content according to user demand. We propose a generic approach on how information, content and knowledge are managed.

The Authoring System detailed in this paper will not be targeted at a specific problem like helping a student, managing content or driving a work-flow, but instead will provide a solid foundation to easily build such specific applications. First of all it will provide unified management of data and meta-data: everything except raw binary streams (e.g., audio and video, office documents, images) will be stored in a distributed ontology and accessible from any of the many servers composing the authoring network. Then, it will provide tools to import (or connect to) data coming from external sources, including semantic data from different ontologies. Existing data, especially if semantically meaningful, should not be converted to a common format but used by the system as it is, by wrapping it in an appropriate user interface and presented in a uniform way to the user.

A relevant problem in this context is the need to tag and reuse media streams located on different servers to produce new forms of knowledge where considerations about such content converge.When media streams are published in a video repository, they are usually associated to a set of meta-data elements such as the author, the title, a summary description and other semantic information. Some systems allow tagging portions of the stream or place time markers to identify specific features of the stream. But usually this kind of data is available only through a specific user-oriented and stand-alone application running on the original server (or cluster of servers).

We aim to create an efficient and scalable architecture for semantic annotation of multimedia contents (semantic tagging) and to implement this architecture to augment the user-interaction for cataloging, retrieving and visualizing collections of composed multimedia contents using community experience. The key contribution of this paper is that we offer a distributed system which addresses two different issues: real-time multimedia web editing and resource overloading of multimedia content located on different servers by means of multimedia tag sharing. Our idea is to give the user the possibility to create new knowledge, editing new compositions in a web interface, pointing multimedia references stored in different locations. Each composition is represented with an envelope which contains meta-data information and URIs. The Authoring System then aggregates multimedia streams and related semantic information in order to make a new composition. Only footage is stored in the Authoring System in order to avoid uncontrolled resource duplications within the network. Finally, to avoid single point of failure and time-variant response, the System is built upon a peer-to-peer architecture.

The remainder of this paper is organized as follows: a review of the current state of the art in Section 2, key ideas of our approach are introduced in Section 3, the real-time multimedia web editing process addressed in our approach showed in Section 4 and a description of the distributed

Authoring System architecture in Section 5 followed by a general conclusions and future works in Section 6.

II. Related Work

Digital multimedia contents are an integral part of human communication. The ease of creating and capturing them has enabled their proliferation, making the interaction with online information largely visual. Chang et al. [3] were pioneers in this field and they addressed this problem using a repository optimized for querying and retrieval multimedia contents using the Internet. The authors focused on search problems of multimedia elements, which were edited in previous steps by means of dedicated technology (e.g. camera) and then published in the repository.

In previous works, other efforts were focused on building a system able to address the problem of facilitating production and editing of multimedia elements for any end-users. The Authoring System has been a recognized solution within the web user community for managing multimedia contents. Some approaches have developed collaborative tools for addressing the above problem. Sung and Lee [13] focused on an efficient concurrency control, based on user awareness, multiple versions, fine granularity of locking and access permissions of shared objects. Using their system, users in different places can author multimedia presentations in a unified spatio-temporal space, while freely traversing the spatial and temporal domain without changing the context of authoring. This study outlined the collaborative approach in the authoring multimedia system.

G. Stamou and S. Kollias [12] joined the semantic topic and multimedia elements. In their work, the authors investigated the semantic representation of multimedia content, mainly analyzing the community utilization and interaction. A multimedia content may be represented as a resource which transports a piece of information and can be augmented with user descriptions like tags, also called annotations. In this supervised approach, an important role is performed by the domain expert, who becomes the end user. Following this idea, Naphade et al. [8] provided a first approach of taxonomy in the video ontology domain. They optimized tools to facilitate end-user access, covered a large semantic space, made automated extraction feasible and increased observability in different broadcast news video data sets by means of adding context information also called annotations.

Indeed manual tagging of visual content has become widespread on the Internet through what is known as "folksonomy" in which human annotators provide descriptive content tags. Choy and Lui [4] showed how collaborative tagging system augments the user experience. One of the main problems is still the diversity and the inconsistency among these approaches. In the area of human tagging, a common challenge is to generate consistency among annotations in terms of both vocabulary and strategy. To overcome it, a common approach would be to use an ontology to map semantic annotations related to managed concepts. Kiriakov et al. [7] analyzed the usage of a lightweight upper level ontology to represent knowledge in a way which allowed automatic interpretations and inferences. The definition of a multimedia-driven ontology remains an issue, which we will address in the rest of this paper.

In [9], Paolillo and Penumarthy showed an example of "folksonomy" in a real video repository. They investigated the auto-clustering by means of "folksonomy" in a community of users. This approach showed an inconsistency among annotations because of the large number of tags created by users, but pointed out the meaning of semantic feature in a tagging element.

By means of this considerations, we address community oriented mechanism to manage the tag sharing of multimedia contents, which implement the web multimedia editing.

III. Approach

A first step in the definition of the Authoring System is to adopt a specialized ontology that will hold semantic information. In this scenario, [5] and [7] demonstrate that "logically extensive" upper-level ontologies are extremely hard to agree on, to build, maintain, understand and use. This seems to give enough evidence that a lightweight upper level ontology is what semantic annotations need as a basis for representing the tag concept and for describing the community interactions among the users. In the scenario of global definition of semantic for managing Authoring Tagging System in online communities, one of the main work is Semantically-Interlinked Online Communities (SIOC) Core Ontology [2]. Bojars et al. [1] proposed new usage scenarios for online community site data and lets developers build innovative semantic applications on top of the existing Social Web, by means of the SIOC ontology.

Many researchers investigated in ontology creations for representing the semantic of video elements [6], [14]. They obtained different results which addressed one particular topic, e.g. to represent taxonomy of a kind of media or to identify event sequences or concepts. These outlined the difficult to create a standard video description domain. One of our goal is to preserve the semantic meaning of the multimedia content, wrapping it in a reference of the above ontology and augmenting its information by means of user-annotations. For such reason we focus on the representation of a media as simple item, which is described by means of tag concept.

First of all, we map our usage-scenario in the above ontology; in particular we need to represent multimedia contents and user interactions in a common way, by means of ontology entities and inferences. We use the Item concept as the envelop of our main element (multimedia stream), after this we use the community concept to exploit the Friend of a Friend (FOAF) inferences among users. The information

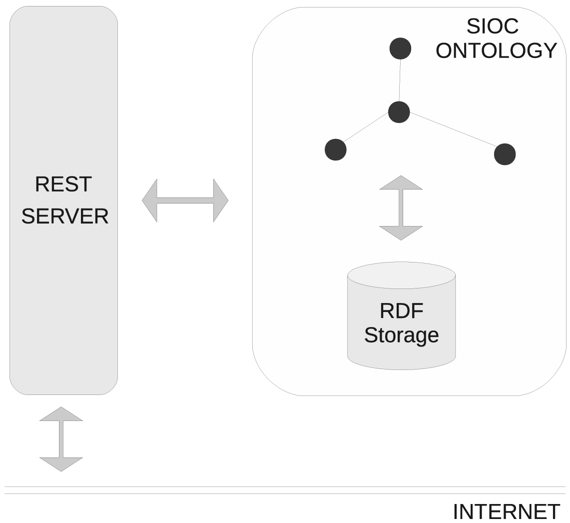

Figure 1. The architecture of our Authoring System aimed to get the user demand and to show the inferred data according to the ontology schema.

needed to represent a tag process is stored in a Tag entity. The inference between the multimedia content (Item) and the Tag is provided by means of the topic inference. To sum up, we use a sub-set of the overall entities contained in SIOC ontology to manage the relation between media elements and their annotations. In particular, the Item represents multimedia content descriptions (e.g. start-time, end-time, duration, streaming protocol, streaming source locator, etc.), the Tag holds category selection list and descriptions. Each user may perform the Post of any new resource (multimedia content) and may infer the Tag information, by means of topic inference.

As described previously, the ontology represents an upper-level domain description, which is composed of entities linked using relations. These data are mapped in triples (s,v,o) and stored in a Resource Description Framework (RDF) repository. In particular, our data-base holds semantic information about audio and video streams: streaming sources locators, per-user annotated media fragments and per-user augmented composite streams. Streaming source locators are used to identify and access standard audio and video streams over the Internet. Users can tag, annotate and split into fragments such streams using tools provided by the Authoring System. Also, they can import any semantic information embedded in the stream itself using a standard data format like RDF or link to knowledge available on the originating server using a SPARQL endpoint.

Figure 1 shows the architecture of our Authoring System. The Representation State Transfer (REST) application server manages web resources between the outlined system with customers and it provides an interface for accessing the semantic data. These data are stored according to the SIOC ontology schema in a RDF storage. For each client interaction the Authoring System materializes the RDF data in a client managed way and sends the answer according to the HTTP protocol.

IV. Web multimedia editing

A YouTube like system is developed to manage the web multimedia editing, performed by means of semantic tagging inferences. A web user is able to compound new Items which are made merging the semantic annotations of different Items. An important role is represented by the Item time-line. A user may select a part of the Item, which is called fragment and is characterized by a start-time and an end-time, and may create an annotation. In this way, the Item may be made from many fragments or annotations. Then, a user may divide the Item in a sequence of new elements, which are meaningful for him in a semantic way. Augmented information are expressed by means of triples and sent or gathered from the user through HTTP/SPARQL standard protocol.

Additionally, the gathered information can be assembled by users to define new virtual media streams that, from the consumer point of view, are played as a sequence of audio and video fragments that are joined to form a novel creative composition. The audio and video footage of this new composition is not stored anywhere in the network. It is provided in real time by the Authoring System, which will use streaming source locators to transparently join right media fragments following a user's play request. For such reason, the multimedia content remains the same and it is not copied in any clients, but it is stored only in one single source. Then, our Authoring System limits the network bandwidth consumption (it only exchange the RDF data) and it prevents the data redundancy among the different users.

The informational content, related to original media streams available in the Authoring System, increases implicitly with the number of annotated fragments. Those are created by the user-interaction through the mix of semantic relations. Even though this architecture is strictly dependent from the original source (the source change may require a new index schema for all fragments), on the other hand, it provides a consistency check on data represented by annotations. In addition, this kind of information becomes available for all users and, in a trusted network, the global information grow continuously with community interactions.

Figure 2 shows a mix of new stream, which is composed of many parts of contents created by the community and localized within the Internet. A user client communicates with the Authoring System by means of a HTTP/SPARQL channel. The Authoring System indexes the required resources and sends to the user a list of meta-information according to the demand. This envelop included the URI reference where the user client may stream the multimedia content, with additional information like descriptions and time-line parameters. The user client communicates with

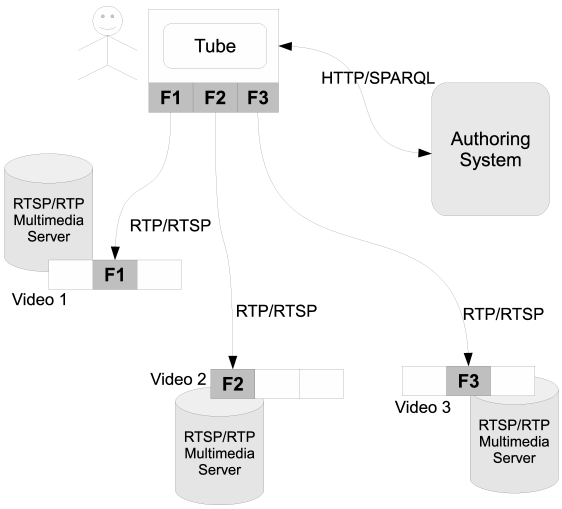

Figure 2. The interaction from the end-user and our Authoring System aimed to discover the meta-data information indexed into our repository.

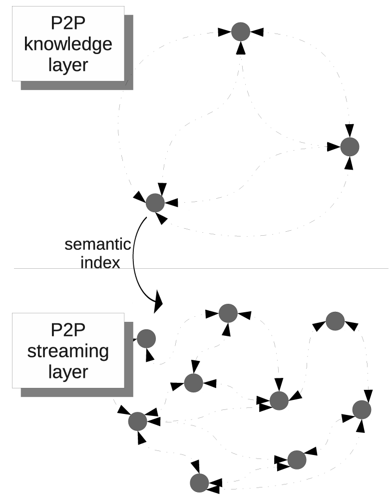

Figure 3. The distributed Authoring System architecture, composed of two layers: the reliable RDF repository layer, built up on the P2P network and BitTorrent streaming layer

multimedia content servers by means of RTSP/RTP (or similar streaming protocol).

V. Distributed Authoring System

As the number of users of the Authoring System increases, we can expect the storage space required by the ontology and the bandwidth needed by the Authoring System to deliver virtual streams to both grow. To overcome this problem, we use a peer-to-peer network to accommodate for the new storage space and bandwidth requirements. Each node is at the same time an autonomous Authoring System and part of two different P2P networks: a distributed RDF storage and a Peer-to-Peer file-sharing cloud.

In our previous work [11], we proposed an architecture for a load balanced and reliable RDF storage system which manipulated semantic information within a distributed peer-to-peer network. Peers are self organized in a ring topology, based on a Distributed Hash Table (DHT), where each node is assigned a segment of the key space that can dynamically change in order to maintain a uniform distribution of the data among the participating peers. Data redundancy is then used to replicate each triple in multiple locations so that, in case of peer failures, neighbor nodes can act on their behalf and return consistent results. Additionally, each node provides an entry point able to resolve atomic, disjunctive and conjunctive SPARQL queries on the network semantic knowledge. We use this architecture to implement a reliable storage layer, distributed among a trusted network of peers.

For what concerns the streaming layer, we adopt the BitTorrent protocol. Pouwelse et al. [10] showed the large adaption of this protocol by the community and the significant performance in the P2P architecture. Additionally, the index mechanism used by the BitTorrent peers can augment our approach based upon the pointing of resource by means of URI references.

Using the distributed RDF storage, each Authoring System has direct access to full semantic information generated by all nodes within the network. Moreover, it can act as a streaming server and when it receives a demand can decide, considering its bandwidth load estimations and network files distribution, to assemble and play the media stream directly, to demand part of the work to its peers or redirect full request to the P2P overlay. In Figure 3, we can analyze the proposed distributed Authoring System approach, composed of two layers: in the top the P2P knowledge layer, which hosts the RDF data of the whole ontology and the P2P streaming architecture in the bottom level.

VI. Conclusions and future works

In this paper we have presented a distributed semantic peer-to-peer Authoring System, joined with a semantic video sharing context and used to obtain web multimedia editing.

A novel approach is used to develop the Authoring System based on two P2P layers that manage the ontology editing real-time and video streaming. The proposed architecture shows an information sharing model which permits to reference in a transparent way portion of media elements, which are published in a network and stored in external video repositories. Only using the web-based application, a user may edit multiple streams and join them in such a way to perform a complete playlist. This process prevents the copy of overall multimedia contents on a client and, by means of tagging information, compounds a new element only pointing parts of contents. For this reason, our system firstly reuses already existing source or parts of and, then, refurbishes new contents. In the post-production step, the new content is shared with the overall community. Additionally,

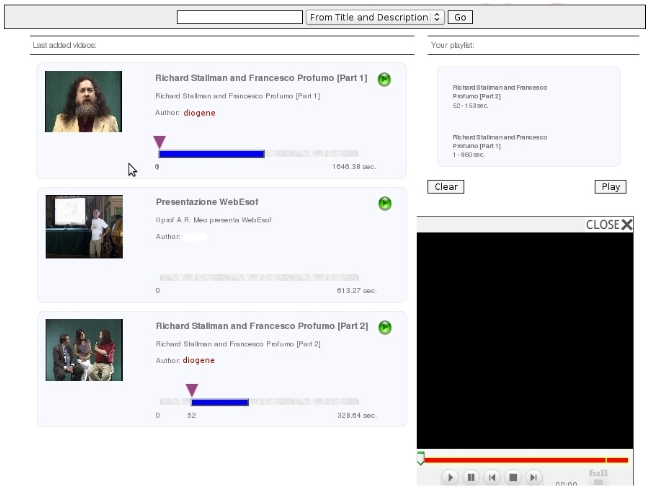

Figure 4. The working prototype of the Authoring System: semantic video annotation and creation of a new composite virtual stream.

the use of a standard ontology, SIOC, makes the communication with other architecture easy and, furthermore, many user interactions improve the knowledge of the Authoring System. To avoid bandwidth bottle-necks and single point of failure issues, we propose a P2P RDF storage network associated to a P2P file sharing network. The system can improve the process of building ontologies by allowing non technical users to easily express their knowledge about audio and video footage and implicitly share it in a semantic way.

We developed a real front-end web-based application (Figure 4) which implements the approach described above. Future works will consist of a large testing step in a real scenario to obtain relevant data about sharing contents and spreading semantic concepts. Topic scenarios may be based on educational or company environments with a large amount of resources to manage and visualize. Also, based on previous lightweight ontology consideration, we will extend it adding MPEG-7 standard related to the multimedia content in order to obtain a more accurate element description. Finally, we will inherit Word Net references to preserve the inconsistency due to different words chosen to represent same semantic meanings.

Besides, we will augment the user-interaction of a end-user with our Authoring System, exploiting semantic inferences among the user community and implementing data mining algorithms for linking main tag topics in the overall (or a part of) set of tags and user descriptions.

ACKNOWLEDGMENT

We wish to thank anonymous reviewers, who helped us in preparing the final version of the paper. Additionally, a special thank to our advisor and professor, A.R. Meo, for his continuous suggestions and encouragements.

REFERENCES

[1] U. Bojars, J. G. Breslin, V. Peristeras, G. Tummarello and Stefan Decker, *Interlinking the Social Web with Semantics*, IEEE Intelligent Systems, Volume 23, pp. 29 - 40, 2008.

[2] J. G. Breslin, A. Harth, U. Bojars, and S. Decker, *Towards Semantically-Interlinked Online Communities*, The Semantic Web: Research and Applications, Volume 3532, pp.500 - 514, 2005.

[3] S. Chang, J. R. Smith, M. Beigi and A. Benitez, *Visual information retrieval from large distributed online repositories*, ACM Communications, Volume 40, Issue 12, pp. 63 - 71, 1997.

[4] S. Choy and K. Andrew, *Web Information Retrieval in Collaborative Tagging Systems*, In Proc. of the IEEE/WIC/ACM International Conference on Web Intelligence, pp. 352 - 355, 2006.

[5] J. Davies, M. Lytras, and A. P. Sheth. "Semantic-Web-Based Knowledge Management", IEEE Internet Computing, Volume 11, no. 5, pp. 14-16, 2007.

[6] Alexandre R.J. Fran?ois, Ram Nevatia, Jerry Hobbs, Robert C. Bolles, *VERL: An Ontology Framework for Representing and Annotating Video Events*, IEEE MultiMedia, Volume 12, Issue 4, pp. 76 - 86, 2005

[7] A. Kiryakov, Borislav Popov, Ivan Terziev, Dimitar Manov and Damyan Ognyanoff, *Semantic annotation, indexing, and retrieval*, Journal of Web Semantics: Science, Services and Agents on the World Wide Web, Volume 2, Issue 1, pp. 49 - 79, 2004.

[8] M. Naphade, J.R. Smith, J. Tesic, Shih-Fu Chang, W. Hsu, L. Kennedy, A. Hauptmann, J. Curtis, *Large-scale concept ontology for multimedia*, IEEE Multimedia, Volume 13, Issue 3, pp. 86 - 91, 2006.

[9] J. C. Paolillo, S. Penumarthy, *The Social Structure of Tagging Internet Video on del.icio.us*, In Proc. of the 40th Annual Hawaii International Conference on System Sciences, Volume 0, pp. 85-95, 2007.

[10] J. Pouwelse1, P. Garbacki1, D. Epema1 and H. Sips1, *The Bittorrent P2P File-Sharing System: Measurements and Analysis*, Peer-to-Peer Systems IV, Volume 3640, pp. 205-216, 2005.

[11] G. Rizzo, F. Di Gregorio, P. Di Nunzio, A. Servetti and J. C. De Martin, *A peer-to-peer architecture for distributed and reliable RDF storage*, In Proc. of the First International Conference on Networked Digital Technologies, pp. 94-99, 2009.

[12] G. Stamou and S. Kollias, *Multimedia Content and the Semantic Web: Standards, Methods and Tools*, John Wiley & Sons, 2005.

[13] M. Y. Sung and Do H. Lee, *A collaborative authoring system for multimedia presentation*, In Proc. of the IEEE International Conference on Communications Volume 3, pp. 1396 - 1400, June 2004.

[14] Z. Zha, T. Mei, Z. Wang, X. Hua, *Building a comprehensive ontology to refine video concept detection*, In Proc. of the International Workshop on multimedia information retrieval, pp. 227 - 236, 2007.