Interlinking Multimedia Annotations

Yunjia Li
School of Electronics and Computer Science
University of Southampton
Southampton, UK
yl2@ecs.soton.ac.uk

Mike Wald
School of Electronics and Computer Science
University of Southampton
Southampton, UK
mw@ecs.soton.ac.uk

Gary Wills
School of Electronics and Computer Science
University of Southampton
Southampton, UK
gbw@ecs.soton.ac.uk

ABSTRACT

With the fast growth of multimedia sharing and annotating applications on the Web, there is an increasing research interests in semantic annotations of multimedia. However, applying linked data principles in multimedia annotations is a relatively new topic, especially when annotations are related to media fragments. This paper, therefore, discusses this problem and further breaks it down into three fundamental sub-questions: 1) choosing media fragment URIs 2) Dereferencing media fragment URIs 3) Ontology alignment related to media fragments and annotations. This paper briefly describes how the interlinking multimedia annotations could be used in the future and concludes with a call for future research to deeply investigate the three research questions. There is a need to develop some working model to address the problems of publishing multimedia resources on the Web as linked data.

Categories and Subject Descriptors

I.7.2 [**Document Preparation**]: Hypertext/hypermedia, Multi/mixed media, Standards; H.5.m [**Information Systems**]: Miscellaneous

General Terms

Languages, Standardization

Keywords

linked data, media fragments, multimedia annotations,semantic Web

1. INTRODUCTION

The Web applications today have been enriched with various multimedia resources and annotations. The consumption of multimedia resources on the Web is tightly coupled with multimedia annotations, which depict the content of multimedia resources. However, to search and share parts of a media resource is still difficult due to the fact that the annotations usually fail to denote the complex nature (temporal, spatial, etc) of media resources. There is an increasing research interest in semantic annotations of multimedia resources and media fragments. Media fragments and semantic annotations have been addressed in various standards and projects, such as MPEG-7 [10], MPEG-21 [6], Core Ontology of MultiMedia (COMM) [2] and temporal URIs [11], but it is still difficult to achieve the interlinking of various multimedia resources on the Web due to the fact that these standards usually are format dependent and there is lack of mutual understandings among applications.

The initiative of Linked data describes a series of methods of publishing structured data using semantic Web technologies and other related standards, such as Resource Description Framework (RDF) [9]. The Linked Data principles [3] offer guidelines of publishing linked data on the Web so that data can be better connected to each other and explored by machines. Generally, there are four rules which must be followed when publishing linked data on the Web [3]: 1) Use URIs as names for things 2) Use HTTP URIs so that people can look up those names 3) When someone looks up a URI, provide useful information 4) Include links to other URIs, so that they can discover more things. The big success of publishing Public Sector Information (PSI) as linked data for the UK government [13] has shown where the Web of data can go as an open and well-connected world.

Linked data should also be able to be applied on the interlinking of more complex multimedia resources on the Web. The W3C Media Fragment Working Group[1] in Video in the Web Activity[2] have collected a wide range of use cases of using media fragments and proposed format independent Media Fragment URI 1.0 draft(MFURI 1.0). Different dimensions for multimedia resources, such as temporal, spatial and track, have been identified. Even though some issues regarding URI dereferencing and multimedia representations using MFURI 1.0 in linked data are also pointed out in [7], little research has been done to apply linked data principles to media fragments and annotations so that multimedia resources can be effectively interlinked to other datasets in the linked data cloud. This paper presents the problems when publishing media fragments and annotations based on the linked data principles and discusses the possible usage of media fragments and annotations in linked data.

[1]http://www.w3.org/2008/WebVideo/Fragments/
[2]http://www.w3.org/2008/WebVideo/

2. DISCUSSION

Linked data principles have been successfully applied in various situations, but compared with other raw data that has been published, such as government data, multimedia resources have more aspects to be considered when applying linked data principles:

1. More complex structure with multi-dimensions (temporal, spatial, tracks, composite elements, etc).
2. Using varieties of codecs and wrapped in a certain format. Different multimedia files need proprietary software or decodecs.
3. Diversity of annotating possibilities. Users can annotate a single time point, or a time interval of whatever length, or the resource as a whole. In a continuous dimension (spatial or temporal), annotation can be made in high granularity.
4. High changing frequencies. Unlike statistic data, which is seldom modified once published, users should be able to frequently create, update or delete media fragments and annotations on a certain multimedia resource.
5. URI schemes used for multimedia delivery on the Web vary. Protocols like Real Time Streaming Protocol[3] (RTSP) and CRID[4] (TV-Anytime Content Reference Identifier) are coexisting with HTTP.

The introducing of media fragments and annotations into linked data must follow the four linked data principles. So the key problems when publishing media fragments and annotations can be summarised as:

Q1 How to choose URI to identify media resources, especially media fragments

Q2 How to return appropriate representations when dereferencing the URIs of media fragments in different contexts

Q3 How to tackle problems of ontology alignment related to media fragments and annotations

To follow the first and second rules of linked data principles, HTTP URIs needs to be used to identify media fragments and they should be applicable to all commonly used media formats on the Web. The third rule of linked data requires that the URIs should be dereferencable. Applications have to be able to either directly return or redirect the request to both original representations (the original file of image, audio and video) and RDF representations. So there must be a mechanism to decide which representation should be returned in different conditions and how the representations can be returned. Q3 in this context is actually a problem of choosing appropriate vocabularies to describe and link media fragments and related annotations. The fourth rule of linked data indicates that domain-specific vocabularies are necessary to link things together. In order to achieve better interlinking, mappings between different vocabularies will be needed. The addressing of media fragments enables the possibility to attach annotations to media fragments, so that the inner content of multimedia resources can be better explored by users.

[3]http://tools.ietf.org/html/rfc2326
[4]http://tools.ietf.org/html/rfc4078

The current situation of multimedia authoring and sharing applications on the Web is that most of them have repositories and multimedia management systems. However, the data, especially data about media fragments, is not published as linked data. So an important guideline when applying linked data principles into multimedia applications is that, on the application level, developers should not totally abandon ”existing data management system and business applications”, but add an ”extra technical layer of glue to connect these into the Web of Data” [8].

2.1 Choosing URIs for Media Fragments

Many standards try to expose the media fragments for annotations, but some of them cannot be applied under linked data principles as they are non-URI based mechanisms, such as MPEG-7, Synchronized Multimedia Integration Language (SMIL)[5] and SVG[6] . In these standards, the descriptions of temporal and spatial dimensions are divided into several attributes, thus the media fragment is not represented by a single URI.

According to the URI definition in RFC3986 [4] and ”Cool URIs for the Semantic Web” [12] ,there are generally three valid ways to add fragment information into URIs: URI query, slash namespaces, hash namespaces. URI query is not widely used to identify a resource in RDF (even though MFURI 1.0 also defines the usage of URI query). The query will return a completely new resource from the server and thus lose the affiliation between parent and children resources. However, many applications do not host the videos and they are fetched directly from the host servers. So it is not realistic to require all the host servers to write extra programmes to handle the query. It is the same case to adopt slash namespaces. Otherwise, a ”404 not available” response will be returned to the client.

Hash namespaces perfectly fit the requirement that multimedia hosting server can be totally unaware of the fragments because the HTTP request will not pass the fragment to the server. Two typical examples of using hash namespaces to denote media fragment are MFURI 1.0 and MPEG-21. MPEG-21 has the limitation that it only applies to MPEG format. MFURI 1.0 in this case is better in that the URI does not restrict the media format it applies. One problem of hash namespaces is that the semantics of URI fragments for most multimedia formats are undefined[7]. Therefore, user agents cannot understand the media fragment just like URI fragments in HTML document. The client side needs extra scripts to handle the fragments. It is quite likely that different mechanisms representing URIs for media fragments will co-exist in the future, so the mappings among these standards might be necessary.

[5]http://www.w3.org/TR/SMIL/
[6]http://www.w3.org/TR/SVG/
[7]http://www.iana.org/assignments/uri-schemes.html

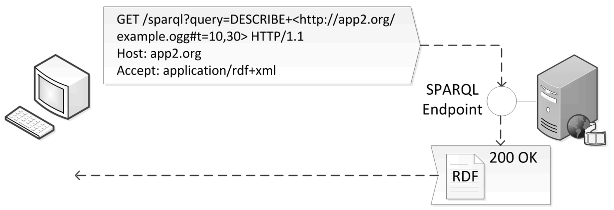

Figure 1: Dereferencing RDF representation of media fragments from SPARQL endpoint

According to the second rule, HTTP should be chosen as the URI schema, but MFURI 1.0 also discusses the possibility of applying RTSP in linked data[8]. RTSP has merits when delivering streamed media and has been widely used on the Web, such as BBC iPlayer[9]. RTSP provides better media control when delivering the original representation of multimedia resources, so sending bit stream is relatively easier compared with mechanisms provided by MFURI 1.0. There is also the possibility that the RDF representation can be included in the response of "DESCRIBE" activity. However, the dereferencing of RTSP URIs is quite different from HTTP URIs. In addition, proprietary servers need to be used to stream the audio/video.

2.2 Dereferencing URIs for Media Fragments

A media fragment on the Web can have different representations. Dereferencing different representations can be done through content negotiation, where *Accept: application/rdf+xml* or *Accept:text/html* can distinguish if the response is RDF or HTML document. The recipes listed in [5] are applicable to media fragments, but the choosing of URI for media fragment will affect the dereferencing process as it decides whether or not the fragment information be passed to the server. If not, the client side has to use extra programme to include the fragment information in the HTTP request. Then, the server side can process the fragment information and return an appropriate response.

When dereferencing the RDF representation, the ideal result is returning a smaller RDF graph containing only triples related to the requested media fragments. MFURI 1.0 proposes several ways of processing media fragments through HTTP Accept Range header. The server then can compose a RDF file only about that media fragment as a response. The problem about this method is that currently it cannot be done automatically and some client-side programme has to insert the Range header. Another way to tackle this problem is similar to "Redirecting to a SPARQL endpoint" recipe [5] (Figure 1). Client side has to encode the media fragment URI into the query string and server side only needs to expose a SPARQL endpoint to handle the HTTP GET query.

Another problem is the flexibility when dereferencing the URIs for media fragments. For example, when *#t=20s,40s* is requested, the server can choose to return RDF representations for the exact fragment, or within the interval such as *#t=25s,35s*, or slight around the fragment such as *#t=18s,38s*. This is due to the reason that when users connect annotations to media fragments, their choosing of fragment might be slightly different, even though they are trying to annotation the same fragment.

2.3 Ontology Alignment

There are many vocabularies existing to describe multimedia resources, such as MPEG-7, EXIF[10]. MFURI 1.0 is employed by Ontology for Media Resource 1.0[11] (OMR 1.0), which defines the core vocabulary for multimedia resources. It also provides ontology (or vocabulary) mappings to other existing formats, which are very useful when interlinking multimedia resources in different formats from varieties of repositories.

URIs for media fragments offers the opportunity to link annotations to a certain media fragment using linked data principles, but it is a domain specific problem to choose an appropriate vocabulary to describe the relationships of annotations. For example, if an application makes annotations on UK Parliament debates, except for the vocabularies used to describe media fragments, domain vocabularies about debate events, parliament decision making processes and MPs' profiles may also be used to semantically describe the debates. Applications in e-learning environment can also apply media fragments to the multimedia learning objects and use vocabularies of teaching, learning target and modules to describe the annotations.

Many automatic and manual annotation methods can be applied to interlink annotations to other datasets. There are many methods of extracting metadata and other data representing content in the media such as mountain, trees, people, sea, etc. In addition, some fixed relationships on temporal or spatial dimensions can be explored automatically. However, human interference is still needed to point out which media fragments are valued and should be interlinked to other datasets.

3. USAGE OF INTERLINKING MULTIMEDIA ANNOTATIONS

The ultimate goal of introducing media fragments and linked data into multimedia applications is to enable the interlinking among media fragments and annotations. The following scenario explains how this could be possibly done:

Steve is watching a lecture recording made by Bob about Semantic Web. Between the 80th second and 90th second, Bob mentioned the term linked data principles, which Steve could not understand. He asks a question "What are linked data principles". Bob, the lecturer, on seeing Steve's comment, relates this media fragment to a media fragment (between 20th second and 500th second) of Alice's lecture recording, in which she further explains what are linked data principles. It is not necessary that the two recordings are located in the same repository. Bob also links the official W3C page about linked data to this media fragment, so that Steve, as well as other users with similar questions, can do some further reading. More importantly, when searching "linked data

[8]http://www.w3.org/TR/media-frags/#rtsp-media-fragment-processing
[9]http://www.bbc.co.uk/iplayer/
[10]http://www.exif.org/Exif2-2.pdf
[11]http://www.w3.org/TR/mediaont-10/

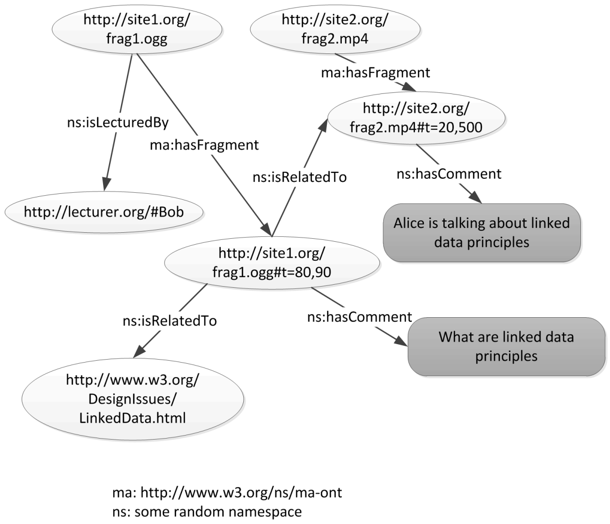

Figure 2: An example of interlinking media fragments

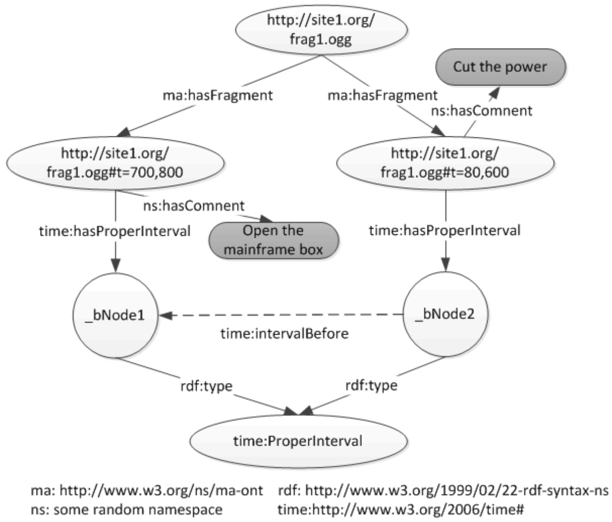

Figure 3: An example of reasoning on temporal relationships

principles", both Bob and Alice's lectures will be listed in the search result and the result can point to the exact fragment where the term "linked data principles" are mentioned, even though the whole lecture may be mainly about some other topics.

Figure 2 presents the RDF graph representing the relationships described in this scenario. Most of the interlinkings in this example should be manually created as it is still difficult to automatically detect the similarity between two media fragments currently unless users are involved to point out the relationships. Some applications enable users to make and share annotations on media fragments, but without media fragment URIs, these annotations are locked inside each application and could not be linked together on media fragments level to benefit search.

When annotations are linked to media fragments, reasoning related to temporal and spatial dimensions of media fragment can be further explored. For example, J.F. Allen summarises the relationships between time intervals as: before, equal, meets, overlaps, during, starts and finishes [1]. So by analysing the URIs of media fragments, it is possible that the temporal interlinking between two fragments can be setup automatically. Similar relationships can also be found in spatial dimension as "within or around an area of something". Figure 3 is an example indicating that "cut the power" happens before "open the mainframe box". This reasoning is based on the fact that the annotation "cut the power" is linked to the media fragment which is earlier than the media fragment "open the mainframe box" annotates.

4. CONCLUSIONS

There is no detailed solution yet to publish multimedia annotations, especially media fragments as linked data. This paper, therefore, discusses three key problems regarding applying linked data principles to media fragments and annotations. Some possible usages of interlinked multimedia annotations have been revealed in this paper. Deeper investigation has to be done to in this area in order to enable the better interlinking of multimedia resources on the Web. Some general working models of publishing and consuming linked media fragments and annotations should be developed and evaluated in real application contexts.

5. REFERENCES

[1] J. F. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, 1983.

[2] R. Arndt, R. Troncy, S. Staab, L. Hardman, and M. Vacura. COMM: Designing a Well-Founded Multimedia Ontology for the Web. In *International Semantic Web Conference*, volume 4825 of *Lecture Notes in Computer Science*, pages 30–43. Springer Berlin Heidelberg, 2007.

[3] T. Berners-Lee. Linked Data. 2006.

[4] T. Berners-Lee, R. T. Fielding, and L. Masinter. RFC 3986 - Uniform Resource Identifier (URI): Generic Syntax, 2005.

[5] D. Berrueta and J. Phipps. Best Practice Recipes for Publishing RDF Vocabularies, 2008.

[6] J. Bormans and K. Hill. MPEG-21 Overview v.5, 2002.

[7] M. Hausenblas, R. Troncy, T. Bürger, and Y. Raimond. Interlinking Multimedia: How to Apply Linked Data Principles to Multimedia Fragments. *Linked Data on the Web Workshop*, 2009.

[8] C. Heath Tom Bizer. Linked Data: Evolving the Web into a Global Data Space, 2011.

[9] F. Manola and E. Miller. RDF Primer, 2004.

[10] J. M. Martinez. MPEG-7 Overview, 2004.

[11] S. Pfeiffer, C. D. Parker, and A. T. Pang. Specifying Time Intervals in URI Queries and Fragments of Time-Based Web Resources, 2005.

[12] L. Sauermann and R. Cyganiak. Cool URIs for the Semantic Web. *W3C Note*, 49(681):1–15, 2008.

[13] N. Shadbolt. data.gov.uk - The Linked Data Revolution. 2010.