Industry and Standards

Anthony Vetro
Mitsubishi Electric Research Labs

Video in the Web: Technical Challenges and Standardization

Soohong Park
Samsung Electronics & Kyung Hee University

Erik Mannens and Rik Van de Walle
University of Ghent

Joakim Soderberg
Ericsson Research

Glenn Adams
XFSI

Philippe Le Hegaret
World Wide Web Consortium

Choong Seon Hong
Kyung Hee University

Web-based video has improved the richness of the user experience but has led to challenges in content discovery, searching, indexing, and accessibility. Nevertheless, Web video is being used for advertising, enterprise collaboration, entertainment, product reviews, and social applications. As prices drop for consumer electronics devices that can create high-quality video, amateurs and professionals alike are producing increasing numbers of high-quality videos. Meanwhile, social networks are facilitating the proliferation of Web-based video, effectively making video part of the Web instead of a mere extension that doesn't take full advantage of the Web architecture.

This column describes the current status of Web-based video in World Wide Web Consortium (W3C) standardization efforts. The new W3C specifications are designed to facilitate the cross-community integration of media objects, online media captioning, and temporal and spatial media fragments using uniform resource identifiers (URIs) designed to make video a first-class citizen on the Web.

Media annotation

With the increasing amount of online video and audio, it will become more difficult for viewers to find content using current search tools. Deducing certain video or audio metadata information, such as title, author, or creation date, can be a complex activity in light of the number of competing metadata formats. A standardized ontology would provide a common set of terms to define the basic metadata needed for cross-community integration of multimedia objects. Currently, for example, an image could potentially contain Exchangeable Image File Format (EXIF), International Press Telecommunications Council (IPTC), and Extensible Metadata Platform (XMP) information. There are several metadata solutions for video and audio content, including MPEG-7,[1] IPTC, iTunes XML, Yahoo! MediaRSS, Video Sitemaps, CableLabs VOD Metadata Content, TV-Anytime, EBU Core Metadata Set, and XMP. However, these solutions are not easily used across the domains of interest.

A simple ontology that supports cross-community data integration will be the first public deliverable of W3C's Media Annotation Working Group (MAWG). This ontology is designed to help circumvent the current proliferation of video metadata formats by providing full or partial translations and mapping between existing formats. The core vocabulary, which targets media resources mostly on the Web, is based on a common set of XMP properties that cover basic metadata. For example, *creator* is a common property that is supported in several existing metadata formats and is therefore part of the core vocabulary defined by the ontology. In total, 28 metadata schemes used on the Web are analyzed in the current version 1.0 and mapped to our ontology's XMP properties.

The choice of metadata standards was motivated by their popularity in their respective domains. The ontology, with the properties' definitions and mappings, provides the basic information needed by target applications for supporting interoperability among the various kinds of metadata formats related to media resources, especially those that can be found

Editor's Note

Publishing and interacting with video on the Web in a seamless manner with commonly available browsers is not possible with today's Web standards. This article outlines several World Wide Web Consortium initiatives that aim to change that including work on an ontology to support media annotation, addressing of media fragments for improved access to content, and timed text to enable online captioning.

—*Anthony Vetro*

1070-986X/10/$26.00 © 2010 IEEE Published by the IEEE Computer Society

on the Web. In addition, the ontology is accompanied by an API that provides uniform access to all of the elements it defines. This new API framework is designed, as shown in Figure 1, to query the media ontology.

Media fragments

To make video and audio real first-class citizens on the Web, we should be able to link to and from the media in the same way that authors can create hyperlinks to Web pages. Creating this functionality would allow for other uses to be implemented, such as video highlights, search results, mash-ups, or caching. And it would provide a fragment identifier that could be used for identifying and attaching metadata. The ongoing discussion covers spatial and temporal addressing.

Temporal addressing enables the referencing of a time or a segment of time in video and audio content, including a normal play time (or time offset), a frame-based time, or an absolute time. It allows the media player to jump to the specified time or frame or to only play a specific segment of the file. RFC 2326 (for the Real-Time Streaming Protocol) defines the notion of normal play time, which is the stream's absolute position relative to the beginning of the video. On top of that, the Society of Motion Picture and Television Engineers' time codes define the notion of frame-level accuracy.

Insufficient temporal addressing approaches are more or less available in Synchronized Multimedia Integration Language (SMIL), MPEG-7, MPEG-21,[2] and temporal URI. Both SMIL and MPEG-7 require an indirection, though, in that we need to possess the XML description containing the temporal information before the video content can be fetched. On the other hand, the MPEG-21 and temporal URI approaches rely on defining the URI syntax and thus do not rely on an additional XML description.

These approaches also have limitations. They lack the ability to represent complex fragments, especially when combining temporal and spatial addressing. And when using the fragment identifier component syntax of a URI, the method is dependent on the media type of the retrieved representation. For example, the MPEG-21 URI syntax is tied to the MPEG container. Thus, it would be difficult to apply one generic fragment identifier to the existing video or audio codecs.

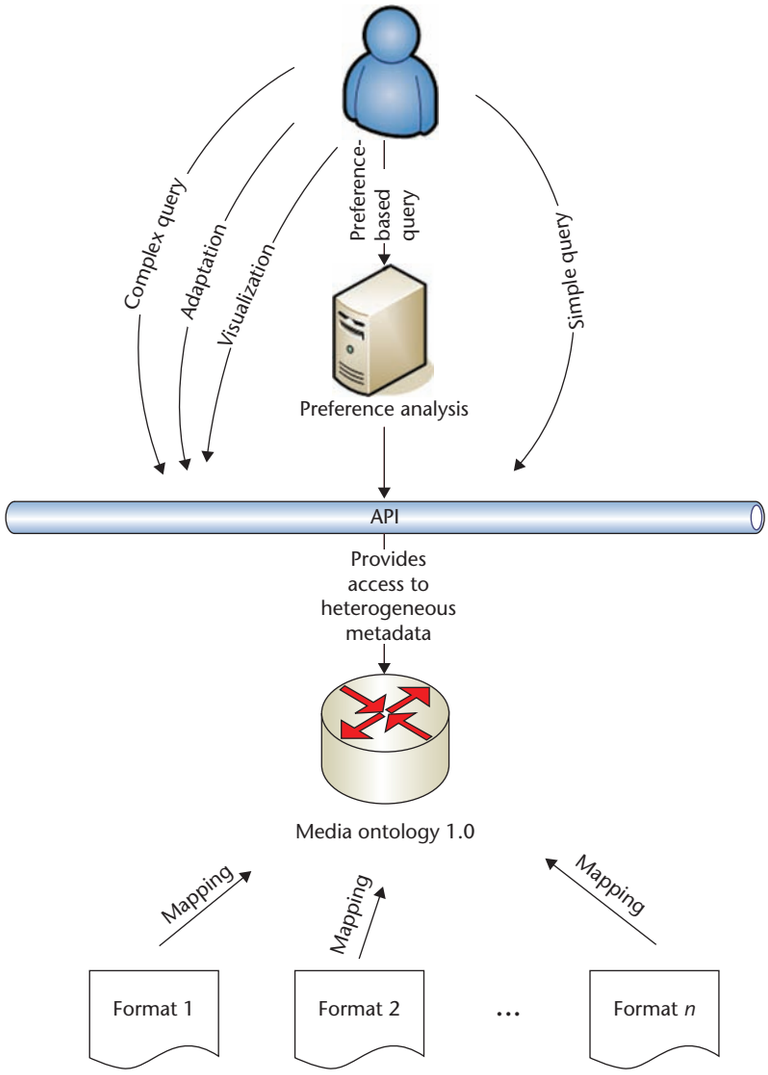

Figure 1. Design of W3C's media ontology and API in the Media Annotation Working Group.

Spatial addressing, on the other hand, allows for referencing a region in a video frame. Information can then be attached to the particular region or, when combined with timed-based addressing, objects can be tracked in the video. Three solutions are currently tangible in the W3C's Scalable Vector Graphics (SVG), SMIL and MPEG-7 standards. However, all of these solutions require an indirection for accessing a specific region in a video.

W3C's Media Fragments Working Group (MFWG) provides URI-based mechanisms for uniquely identifying temporal and spatial fragments of media objects in the Web, such as video, audio, and images. The working group is chartered with the task of investigating several possibilities in URI syntax, using URI fragment identifiers or query parameters to

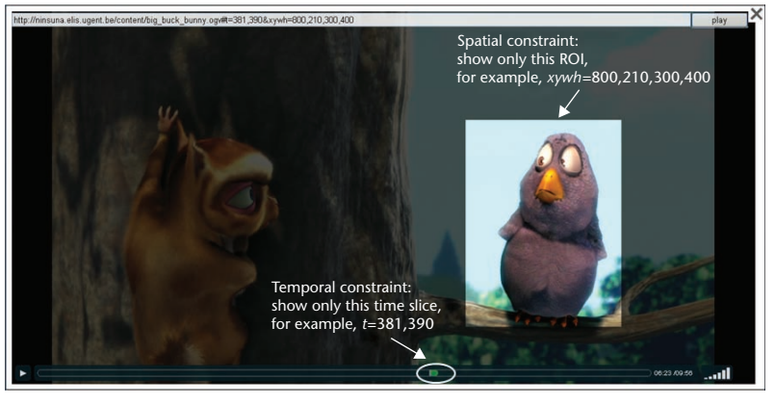

Figure 2. Sharing only parts of a video using both temporal and spatial addressing schemes within an HTML 5 player.

address a region of a media object and to calculate its impact at the application level. To make video a first-class citizen on the Web, we should be able to identify the temporal and spatial regions within the videos.

Having global identifiers for arbitrary media segments or fragments would provide substantial benefits, including linking, bookmarking, caching, and indexing. For example, it would be ideal to be able to point to the twentieth second of a video news report, bringing the user directly to a specific news item within the report. Another example would be the ability to identify individuals or objects that are featured in the video, as can be seen in Figure 2. None of the existing solutions, such as MPEG-21 or temporal URI, are fully satisfactory in this respect or provide a general unique resource identifier for video clips independent of the format in use.

The existing solutions and approaches, such as MPEG-21, SVG, SMIL, or temporal URI, were used as initial background knowledge. The MFWG focuses on developing a mechanism to uniquely identify a temporal fragment within an audio or video object that is independent of the underlying audio or video codec in use. Also, the MFWG is investigating the delivery of the requested resource to allow for full or partial media retrieval using, at least, the Hypertext Transfer Protocol. Furthermore, the MFWG is working on providing a partial mapping between URI syntax and the various ways, in XML or URI, of defining a temporal or a spatial region in W3C recommendations, such as SVG and SMIL.

Timed text

W3C's Timed Text Working Group (TTWG) is chartered with the task of creating a standard XML representation of textual content to which timing and stylistic semantics can be applied. The goal of this research is to produce a standardized content format that will serve as a means for authoring and interchanging caption, subtitling, and other similar types of content, such as marquee and karaoke content. As an authoring format, it will permit the creation of standardized common authoring platforms. As an interchange format, it will permit native distribution and, through transformation processing, conversion to and from existing legacy formats.

The initial work of the TTWG focused on the development of the Timed Text Authoring Format and Distribution Format Exchange Profile (DFXP) specifications developed to a W3C recommendation candidate in late 2009. Although DFXP was explicitly designed to support transformation to and from several existing legacy formats, such as the analog and digital television closed captioning formats used in North America, as well as industry formats such as 3GPP Timed Text, DFXP also supports direct, native distribution to presentation devices or players that include a compliant DFXP presentation processor.

DFXP includes support for block and inline level content, inline and referential styling with inheritance, and inline timing. It draws from the syntax and semantics of existing W3C specifications, including Extensible Stylesheet Language Formatting Objects, SVG, and SMIL, to avoid having to reinvent the wheel. In its initial format, DFXP was designed for delivery as either a complete XML document or as a sequence of XML document fragments; however, DFXP does not define a streaming format. Nevertheless, it's possible to make use of other mechanisms for streaming XML fragments, such as the Binary Format for MPEG-7, to obtain DFXP streaming capabilities. Needless to say, it will be able to use MFWG and MAWG results and vice versa.

Conclusions

The issue of which standard video codec to embrace on the Internet has been debated for several years. HTML 5, which includes a proposal for embedding and controlling audio and video content, has attracted a lot of attention in terms of the choice of video codecs. It doesn't recommend a baseline video codec, but instead elaborates several requirements, such as compatibility with the open source development model

and no patent risks. Nowadays, Adobe, Apple, and Microsoft are already supporting H.264 Advanced Video Coding (AVC)[3] in their product lines. The mobile industry has adopted this format too. However, H.264/AVC currently has licensing liabilities, making it incompatible with the open source community and the goals of the W3C patent policy.

Fortunately, video codecs continue to evolve, and today's codecs will be different from tomorrow's codecs. While a single baseline codec that is royalty-free is critical to some, large video content producers place significant emphasis on the end-user experience. This end-user experience includes widely available players, codecs whose functionalities provide a good viewing experience, ease of use, and standardized means for annotating and exchanging media. Most users take a Flash video for granted or deploy their own video player, using a codec that fits the desired content quality (such as the VP7 On2 codec). As such, Google might have the key to open the codec lock because the company recently acquired On2 Technologies. If Google would make both VP7 and VP8 open source, standardized solutions would have some serious competitors.

Currently, individual content producers still suffer from a lack of a universally supported set of codecs and annotation tools, which makes it challenging to publish video on the Web and to ensure that almost anyone can view, annotate, or exchange that video using commonly available browsers. The ultimate goal is to publish a video the same way one publishes an image today. The sentiment shared among several participants is that W3C should gather the relevant parties and continue to explore this issue.

W3C recently opened up the time dimension on the Internet by specifying a means to address partial media resources (called media fragments) using URI-based mechanisms to uniquely identify temporal and spatial fragments of Web-based media objects. Furthermore, W3C provided a means to annotate these omnipresent media resources in a standardized semantic way using a simple ontology to support the cross-community data integration of information related to media objects on the Web, as well as an API to access this information. In addition, W3C produced a standardized content format that could serve as a means for authoring and interchanging caption, subtitling, and other similar types of metadata frequently used in multimedia resources. All of these initiatives are designed to make video a first-class citizen on the Web. **MM**

Acknowledgments

We thank all of the members of the Video in the Web Activity of W3C and their valuable contributions to it. The research activities described here were partially funded by W3C, Ghent University, Interdisciplinary Institute for Broadband Technology, the Institute for the Promotion of Innovation by Science and Technology in Flanders, the Fund for Scientific Research-Flanders, the European Union, and a grant from the Kyung Hee University in 2010 (KHU-2010372).

References

1. B.S. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG-7: Multimedia Content Description Standard,* Wiley, 2001.
2. I. Burnett et al., ''MPEG-21: Goals and Achievement,'' *IEEE MultiMedia,* vol. 10, no. 6, 2003, pp. 60-70.
3. T. Wiegand et al., ''Overview of the H.264/AVC Video Coding Standard,'' *IEEE Trans. Circuits System and Video Technology,* vol. 13, no. 7, 2003, pp. 560-576.

Contact author Choong Seon Hong at cshong@khu.ac.kr.

Contact editor Anthony Vetro at avetro@merl.com.