"Architecture of a Video Web - Experience with Annodex"

Dr Silvia Pfeiffer
Vquence, Xiph.org and Annodex.org

Background

Since the year 2000 a project under the name of "**Continuous Media Web**", CMWeb, has explored how to make video (and incidentally audio) a first class citizen on the Web. The project has led to a set of open specifications and open source implementations, which have been included into the Xiph set of open media technologies. In the spirit of the Web, specifications for a Video Web should be based on unencumbered formats, which is why Xiph was chosen.

The project is now "owned" by the **Annodex Association**. One particularly exciting use case of Annodex is the Metavid archive of public domain house and senate footage - presented by Michael Dale in another position paper at this Workshop.

This position paper concentrates on a high-level technical description of Annodex, which incidentally stands for "annotated and indexed media". We share the experiences of the project and encourage the audience to consider it as a technology basis to experiment with further functionality for future Web-delivered media.

Requirements for a Video Web

Annodex follows closely along the tradition of the existing text-based Web in creating a Video-centric Web. It continues to build on the HTTP and URI specifications, which are well suited to delivering video over the Web and addressing video resources. It makes use of CSS, javascript and XML to handle video issues that are identical to text issues.

However, on the current Web, video cannot easily be searched, surfed, recomposed, cached, and addressed. These functionalities need to be created through further specifications.

Search functionality requires a standard description of metadata and annotations. Surfing functionality requires hyperlinks into and out of video. Recompositing (or "mashing up") requires a clean random segmentation and recomposition of video - the same requirements incidentally that caching or proxying have. While addressing requires the definition of standard means to hyperlink into video at time offsets or at named offsets, as well as a means to hyperlink out of video.

After some intensive analysis, we realised that there is a fairly easy approach to extending the Web and making video a first class citizen. All that was required was:

1. a markup language (**CMML**) similar to HTML but for time-continuous media so we could put annotations and metadata alongside video in a similar way that HTML puts these alongside text,

2. a means to hyperlink into time-continous media (**temporal URIs**) by specification of temporal offsets and sections in URIs, and
3. a means to make use of the existing Web delivery and caching infrastructure of HTTP through a recomposable encapsulation format (**Annodex/Ogg Skeleton**).

The HTML-like markup language: CMML

CMML, the "Continuous Media Markup Language", is an XML-based markup language for time-continuous data such as audio and video. It provides a timed text annotation format with the following key functionality:

- it is an **XML file** that describes a video's content, but **can also easily be serialised** into time-continuous frames that can be multiplexed into a audio/video encapsulation format such as Ogg, QuickTime, or MPEG.

- it starts with a header which has annotations and metadata for the video file as a whole.

- it then consists of a set of clips which have a start and end time and can provide annotations, metadata, hyperlinks, a representative keyframe, and captions.

- additionally, CMML typically also contains a stream tag to provide information on the location and format of the media file it describes.

An example CMML file looks like this:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<!DOCTYPE cmml SYSTEM "cmml.dtd">

<cmml lang="en">

<stream basetime="0">
  <import contenttype="video/ogg" src="fish.ogv" start="0"/>
</stream>

<head>
  <title>Types of fish</title>
  <meta name="Producer" content="Joe Ordinary"/>
  <meta name="DC.Author" content="Joe's friend"/>
</head>

<clip id="intro" start="0">
  <a href="http://example.com/fish.html">Read more about fish</a>
  <desc>This is the introduction to the film Joe made about fish.</desc>
  <caption>
    <p id="subtitle1" start="0" end="1" style="text-align: left;">
      This is a left aligned subtitle.
    </p>
    <p id="subtitle2" start="1" end="2" style="text-align: right;">
```

```
      This is a right aligned subtitle.
    </p>
  </caption>
</clip>

<clip id="dolphin" start="npt:3.5" end="npt:0:05:05.9">
  <img src="dolphin.png"/>
  <desc>Here, Joe caught sight of a dolphin in the ocean.</desc>
  <meta name="Subject" content="dolphin"/>
  <caption>
    <p id="subtitle3">
      This is a <span style="fontWeight: bold;">lengthy</span><br/>
      subtitle that is split over two lines.
    </p>
  </caption>
</clip>

</cmml>
```

The DTD for CMML can be found here:
http://svn.annodex.net/standards/cmml_3_1.dtd

The current draft specification can be found here:
http://svn.annodex.net/standards/draft-pfeiffer-cmml-current.txt

Hyperlinking into and out of video: temporal URIs

There are many different possibilities of hyperlinking into and out of video that should be provided:

- linking to a temporal offset into a video
- linking to a time segment inside a video
- linking to a named segment inside a video (the identifier of a CMML video clip works well for naming segments)
- linking from the inside of a video to some other Web resource (through a hyperlink inside the clip tag of a video)

The final item is provided through inclusion of "a" elements inside CMML clips. In this way it is possible to link out of a video during a certain time period.

Here are some example URIs that use the specification that we came up with in the CMWeb project. More exact details on the temporal URIs are addressed in a separate position paper.

http://example.com/video.axv?t=npt:15.2 ---
video.axv is transferred from 15.2 seconds into video.axv to the end of the file/stream.
http://example.com/video.axv?t=15.2/18.7 ---
video .axv is transferred from 15.2 seconds into video.axv to 18.7 seconds; the default time scheme "npt" is used.

http://example.com/video.axv?t=15.2/18.7,23 ---
video.axv is transferred from 15.2s to 18.7s and from 23s to the end of the file/stream.
http://example.com/video.axv#t=15.2/18.7,17.4/30.1 --
video.axv is transferred from 15.2 seconds into video.axv to 30.1 seconds.

http://example.com/video.axv?id="dolphin" --
video.axv will be transferred from 3.5s which is where it's CMML places it. Specification of a clip always implicitly specifies a segment since a clip always has a start and end time – even though the end time may be the end of the file/stream.

The draft specification of temporal URIs can be found here:
http://annodex.net/TR/draft-pfeiffer-temporal-fragments-03.txt

A recomposable file format: Annodex / Ogg

With CMML and the extension on the URI sepcification, it is basically to create a video web. However, to gain all information about a video file now requires the transfer of a CMML and a video file as two separate documents. Also, the Web client now needs to synchronise the CMML with the video. Another challenge is: what do you do in situations where the CMML file is still in production because the file on the server is actually being created life? And finally: when storing a video file, you don't really want to have to deal with two files all the time – not on the server and not on your local hard drive.

Therefore, the CMWeb project invented the "Annodex" format to encapsulate CMML into the video file. Annodex is really the Ogg container plus a Skeleton and a CMML track apart from the audio and video tracks. The Skeleton track was added to allow dissection of Ogg files and recomposition without a need for decoding and re-encoding.

The draft specification of Ogg Skeleton can be found here:
http://svn.annodex.net/standards/draft-pfeiffer-oggskeleton-current.txt

Outlook

CSIRO's Shane Stephens has recently released a library called liboggplay which implements all of the Annodex specifications and a javascript API as a basic libary to enable codebases such as Mozilla/Firefox or Opera to provide Ogg support according to the new WHATWG video and audio tags. liboggplay also implements Annodex, thus enabling a much richer interaction with the media data.

A demonstration of liboggplay can be provided at the Workshop consisting of a demo of Firefox 3 running liboggplay and the javascript API to control it.

We are fully aware that at this stage in the development of CMML, temporal URIs and Annodex, they are fairly simple and possibly incomplete. However, they fulfil the basic requirements to create a Video Web and can be extended easily to meet further requirements

such as digital rights management, accessibility, and privacy. These should be solvable in a similar manner to how they work with HTML. So, the foundations are laid, should the audience want to continue experimenting with Annodex and add further functionality.

REFERENCES

S. Pfeiffer, C. Parker, and A. Pang, "The Continuous Media Web: A Distributed Multimedia Information Retrieval Architecture Extending the World Wide Web", Multimedia Systems Journal, Special Issue on Systems and Architectures of Multimedia Information Retrieval, 2005; 10(6):544-558.

S. Pfeiffer, C. Parker, and A. Pang, "Continuous Media Web: Hyperlinking, Search and Retrieval of time continuous data on the Web." In: Srinivasan, U. and Nepal, S., eds. Managing Multimedia Semantics. IRM Press; 2005; pp. 160-180. Chap. 7. ISBN: 1591405696;1591405424 (pbk.);1591405432 (e-bk).

C. Parker, A. Pang, and S. Pfeiffer, "Demonstrating a Video and Audio Web", Proceedings of the ACM Multimedia Conference, 2004 ,October 2004, New York, pp. 168 – 171.

A. Pang, C. Parker, and S. Pfeiffer, "Challenges of Networked Media: Integrating the Navigational Features of Browsing Histories and Media Playlists into a Media Browser", Proceedings of the ACM Multimedia Conference, 2004, October 2004, New York, pp. 480 – 483.

S. Pfeiffer, C. Schremmer, and C. Parker, "Annodex: A Simple Architecture to Enable Hyperlinking, Search & Retrieval of Time-Continuous Data on the Web", Proceedings of the ACM Multimedia Conference 2003, 5th Intl. Workshop on Multimedia Information Retrieval (MIR), November 2003, Berkeley.