5 – codecs and standards

Compression and decompression need to be understood by the student of multimedia. Perhaps not in all detail, but in any case to the extent that the consequences of the choice for a particular media type may be evaluated with respect to (network) resource demands.

The MPEG-4 standard is most directly related to the issue of compression and decompression. The other standards, nevertheless, are sufficiently related with MPEG-4 to justify a combined treatment.

Before looking at codecs and standards, we briefly discuss the current state of multimedia on the web.

codecs and standards

- (multimedia) semantic web?
- codecs
- MPEG-4 standard proposal
- standards (MPEG-4, SMIL, RM3D)

The MPEG-4 standard proposal is an adapted version of the original proposal, that is suited for presentation using either one of the slide formats.

0.1 semantic web?

To finish this chapter, let's reflect on where we are now with 'multimedia' on the web. Due to refined compression schemes and standards for authoring and delivery, we seemed to have made great progress in realizing *networked multimedia*. But does this progress match what has been achieved for the dominant media type of the web, that is text or more precisely textual documents with markup?

web content

- *1st generation* – hand-coded HTML pages
- *2nd generation* – templates with content and style
- *3rd generation* – rich markup with metadata (XML)

Commonly, a distinction is made between successive generations of web content, with the first generation being simple hand-coded HTML pages. The second generation may be characterized as HTML pages that are generated on demand, for example by filling in templates with contents retrieved from a database. The third generation is envisaged to make use of rich markup, using XML, that reflects the (semantic) content of the document more directly, possibly augmented with (semantic) meta-data that describe the content in a way that allows machines, for example search engines, to process it. The great vision underlying the third generation of web content is commonly refered to as

the semantic web

which enhances the functionality of the current web by deploying knowledge representation and inference technology from Artificial Intelligence. As phrased in [CWI], the semantic web will bring

structure to the meaningful content of web pages

thus allowing computer programs,such as search engines and intelligent agents, to do their job more effectively. For search engines this means more effective information retrieval, and for agents better opportunities to provide meaningful services.

A great vision indeed. So where are we with multimedia? In [CWI], we read:

multimedia

> *While text-based content on the Web is already rapidly approaching the third generation, multimedia content is still trying to catch up with second generation techniques.*

The reason for this is that processing multimedia is fundamentally different from processing text. As phrased in [CWI]:

processing requirements

> *Multimedia document processing has a number of fundamentally different requirements from text which make it more difficult to incorporate within the document processing chain.*

More specifically it is said that:

presentation abstractions

> *In particular, multimedia transformation uses different document and presentation abstractions, its formatting rules cannot be based on text-flow, it requires feedback from the formatting back-end and is hard to describe in the functional style of current style languages.*

Now this may well be true for specific categories of multimedia on the web. So, for example, rendering presentations written in SMIL is probably not an easy thing to do. But does this really prevent us from incorporating multimedia in the semantic web, or rather create a multimedia semantic web?

As an example, take a *shockwave* or *flash* presentation showing the various musea in Amsterdam. How would you attach meaning to it, so that it might become an element of a semantic structure? Perhaps you wonder what meaning could be attached to it? That should not be too difficult to think of. The (meta) information attached to such a presentation should state (minimally) that the location is Amsterdam, that the sites of interest are musea, and (possibly) that the perspective is touristic. In that way, when you search for touristic information about musea in Amsterdam, your search engine should have no trouble in selecting that presentation. Now, the answer to the question how meaning can be attached to a presentation is already given, namely by specifying meta-information in some format (of which the only requirement is that it is machine-processable). For our *shockwave*or *flash* presentation we cannot dothis in a straightforward manner. But for MPEG-4 encoded material, as well as for SMIL and RM3D content, such

facilities are readily available. You may look at MPEG-7 to get an idea how this might be approached.

Should we then always duplicate our authoring effort by providing (meta) information, on top of the information that is already contained in the presentation? No, in some cases, we can also rely to some extent on content-based search or feature extraction, as will be discussed in the following chapters.

research directions– *agents everywhere*

The web is an incredibly rich resource of information. Or, as phrased in [IR]:

information repository

> *The Web is becoming a universal repository of human knowledge and culture, which has allowed unprecedented sharing of ideas and information in a scale never seen before.*

Now, the problem (as many of you can acknowledge) is to get the information out of it. Of course, part of the problem is that we often do not know what we are looking for. But even if we do know, it is generally not so easy to find our way. Again using the phrasing of [IR]:

browsing & navigation

> *To satisfy his information need, the user might navigate the hyperspace of web links searching for information of interest. However, since the hyperspace is vast and almost unknown, such a navigation task is usually inefficient.*

The solution of the problem of *getting lost in hyperspace* proposed in [IR] is *information retrieval*, in other words *query & search*. However, this may not so easily be accomplished.

data model

> *The main obstacle is the absence of a well-defined data model for the Web, which implies that information definition and structure is frequently of low quality. [IR].*

Now, how would you approach defining a unifying data model for the web? One project in this area that might be worthwhile to look at is the *OntoWeb* project, accessible through

http://www.ontoweb.org

that aims at producing the technology for ontology-based information exchange for both knowledge management and electronic commerce. Such technology allows for adding descriptive information and, equally important, to reason with such information. Moreover, it allows for dealing with information formulated in disparate terminologies by using so-called ontologies, which may be regarded as formalized perspectives or world views.

Standardizing knowledge representation and reasoning about web resources is certainly one (important) step. Another issue, however, is how to support the user in finding the proper resources and provide the user with assistance in accomplishing his task (even if this task is merely finding suitable entertainment).

What we need, in other words, is a unifying model (encompassing both a data model and a model of computation) that allows us to deal effectively with web resources, including multimedia objects. For such a model, we may look at another area of research and development, namely *intelligent agtents*, which provides us not only with a model but also with a suitable metaphor and the technology, based on and extending object-oriented technology, to realize intelligent assistance, [OO].

For convenience, we make a distinction between two kinds of agents, *information agents* and *presentation agents*.

information agent

- gather information
- filter and select

Information agents are used to gather information. In addition, they filter the information and select those items that are relevant for the user. A key problem in developing information agents, however, is to find a proper representation of what the user considers to be relevant.

presentation agent

- access information
- find suitable mode of presentation

Complementary to the information agent is a *presentation agent* (having access to the information gathered) that displays the relevant information in a suitable way. Such a presentation agent can have many forms. To appetize your phantasy, you may look at the vision of *angelic guidance* presented in [Angelic]. More concretely, my advice is to experiment with embodied agents that may present information in rich media 3D. In section **??**, we will present a framework for doing such experiments.

navigating information spaces Having *agents everywhere* might change our perspective on computing. But, it may also become quite annoying to be bothered by an agent each time that you try to interact with with your computer (you know what I mean!). However, as reported by Kristina Höök, even annoyance can be instrumental in keeping your attention to a particular task. In one of her projects, the *PERSONAS* project, which stands for

PERsonal and SOcial NAvigation through information spaceS

the use of agents commenting on people navigating information space(s) is explored. As a note, the plural form of *spaces* is mine, to do justice to the plurality of information spaces.

As explained on the *PERSONAS* web site, which is listed with the acronyms, the *PERSONAS* project aims at:

PERSONAS

> *investigating a new approach to navigation through information spaces, based on a personalised and social navigational paradigm.*

The novel idea pursued in this project is to have agents (*Agneta* and *Frieda*) that are not helpful, but instead just give comments, sometimes with humor, but sometimes ironic or even sarcastic comments on the user's activities, in particular navigating an information space or (plain) web browsing. As can be read on the *PERSONAS* web site:

Agneta & Frieda

> *The AGNETA & FRIDA system seeks to integrate web-browsing and narrative into a joint mode. Below the browser window (on the desktop) are placed two female characters, sitting in their livingroom chairs, watching the browser during the session (more or less like watching television). Agneta and Frida (mother and daughter) physically react, comment, make ironic remarks about and develop stories around the information presented in the browser (primarily to each other), but are also sensitive to what the navigator is doing and possible malfunctions of the browser or server.*

In one of her talks, Kristina Höök observed that some users get really fed up with the comments delivered by *Agneta* and *Frieda*. So, as a compromise, the level of interference can be adjusted by the user, dependent on the task at hand.

Agneta & Frieda

> *In this way they seek to attach emotional, comical or anecdotal connotations to the information and happenings in the browsing session. Through an activity slider, the navigator can decide on how active she wants the characters to be, depending on the purpose of the browsing session (serious information seeking, wayfinding, exploration or entertainment browsing).*

As you may gather, looking at the presentations accompanying this *introduction to multimedia* and [Dialogs], I found the *PERSONAS* approach rather intriguing. Actually, the *PERSONAS* approach is related to the area of *affective computing*, see [Affective], which is an altogether different story.

The *Agneta* and *Frieda* software is available for download at the *PERSONAS* web site.