Introduction

In a hypothesized future of pervasive broadband Web-based access to multimedia, where content sources are distributed and heterogeneous, multimedia production and distribution can not rely on manual authoring. A set of requirements for an automated and adaptive multimedia presentation system is presented, which would meet the challenges of the multimedia rich future.

We introduce the idea underlying this thesis (section 1.1) and motivate the need for a new type of multimedia presentation system (section 1.2), named SWeMPs, which forms the core of this research. Placing the proposed contribution in the context of current developments in the Web and multimedia, we outline two typical scenarios (section 1.3) which such a system would be expected to be able to realise. As a result we derive requirements (section 1.4) by which the system can be evaluated. To close this introduction, we outline the structure of this thesis (section 1.5) and place our research within the context of ongoing efforts in the fields of multimedia and Semantic Web (section 1.6).

1.1 The idea underlying SWeMPs

The two key problems of modern society's information overload (as produced by technological advances, most notably the Internet) are the effort involved in finding information AND the problem of making sense of the found information.

The Internet has a lot of information but computers can't do very much with it: this is the reason for the proposed Semantic Web as an evolution of the existing World Wide Web [Berners-Lee,2001]. The principles of the Semantic Web can also be applied to closed, smaller scale data stores such as company intranets. We have access to and a need for a lot of information (thanks to pervasive information-accessing devices such as mobile phones, PDAs, smart devices, TV, radio...) but we don't have time to find and make associations between information every time we want to. Given the need for quick, intuitive access to information, multimedia presentation is an effective form of communication – different media act together as a better means to communication dependant on the context, and synchronization (organisation of media in space and in time) can express non-verbal relationships between the content in a way intuitively understandable by a human user (e.g. linear presentation to represent a timeline of events).

Hence we currently have a situation of inefficiency: we need information and we have access to it, but we still can't easily find and consume it. Even when we find and consume information it may not be communicated to us in an effective manner, e.g. we can't concentrate on pictures while driving our car.

The Web is the natural choice for providing and seeking information as it exists in a ready-to-use network infrastructure (the Internet) and forms the world's largest content repository (with more than 11.5 billion pages at the end of January 2005 [Gulli,2005] and definitely even more than that now, whenever one reads this). However the current situation of Web content discovery and delivery displays

significant disorder and inflexibility: natural language is problematically ambiguous and Web content is only adaptive to a limited extent. The same problems are found in any (closed) information stores.

The Semantic Web is about adding machine-processable knowledge to the Web, or indeed any network. As a result, it is argued that Web or other network clients will be able to deliver the correct information every time by unambiguously understanding the meaning of queries and reasoning on them (through the knowledge available to them on the network) to come to an answer. Likewise, network resources will be described with metadata that supports their selection, adaptation and presentation to the needs of the client. While the Semantic Web is not yet in common use by the average Internet user, it is there and it is growing. The Semantic Web search engine Swoogle[1] has indexed almost 300 million statements of knowledge from over 1.5 million Semantic Web documents online[2]. Furthermore, it is experiencing early adopter take-up in smaller, closed networks such as specialist knowledge access for the health care and life sciences community[3].

Hence, we expect that in the future intelligent agents will be able to handle Web-based tasks, and the Web will shift from being a content repository to being a knowledge repository in which knowledge is often represented by presentable content. In the shorter term, semantics will increasingly be used within dedicated communities to annotate their data and enable better data search and integration. In the longer term, rather than seeing the Web as it is now, with a single client making a request for a particular resource from a certain server and receiving this resource as a response from the server, or when it doesn't know precisely where the desired resource is then using a search engine first to locate that resource, in the future Web a client could formulate a request for information, this request will be processed by intelligent agents that can makes inferences based on the knowledge available to them, and from those inferences content providers will be called to respond with an answer to the clients request. This answer may be in the form of knowledge or of media items which represent that knowledge, and the set of appropriate content must be organized and delivered in an understandable way to the user.

This perceived evolution is illustrated in Figure 1.1: the core principle of the Web was that a user gives an URL and a Web application (i.e. the browser) retrieves and displays the content at that URL. Presently, with the enormous size of the Web, search engines are used as a form of middle agent in this process: the user does not know what he or she wants in terms of an URL but rather in terms of some words that express the information need. However the resolution of resources is made syntactically, i.e. the search engine offers up resources which contain the words provided by the user, and provides all matches in a list which is ordered according to some principle (e.g. Google's PageRank). Finally, in the hypothesized future Web, that need would be expressed semantically – the user's Web application will handle translating the user's expression of his or her need (presumably in natural language) into a formal semantics that a machine 'understands' (i.e. without the ambiguities of natural language). Instead of a search engine, a multimedia presentation system acts

[1] http://swoogle.umbc.edu/
[2] Statistics from 9 June 2006. The latest statistics are at http://swoogle.umbc.edu/index.php?option=com_swoogle_stats&Itemid=8
[3] That this community is an early adopter of Semantic Web technologies is reflected in the formation of a dedicated W3C group to promote activities, see http://www.w3.org/2001/sw/hcls/.

as the middle agent, not only retrieving the appropriate content from the Web but also organizing it into a coherent and meaningful presentation to the user, taking into account the user profile, usage context and so on. This thesis can not realise this scenario alone, but we see it as taking a step in this direction by undertaking research which is necessary to lay the groundwork for such a hypothetical future scenario.

This next generation Web, because it represents a paradigm shift in how systems will interact with the Web, requires a similar paradigm shift in system models, architectures, implementations and operations. This shift is particularly relevant to intelligent multimedia presentation systems, as they will be the enabler for the user to be able to benefit from the new means of finding information that the Semantic Web should make possible.

In this thesis we propose an intelligent multimedia presentation system that is designed to meet the requirements of this foreseen next generation Web and that will represent the paradigm shift of the Semantic Web in a knowledge-based approach to multimedia presentation. The system has been titled SWeMPs – a Semantic Web-enabled Multimedia Presentation System[4]. It is a framework for realising individual "intelligent information services" (IIS).

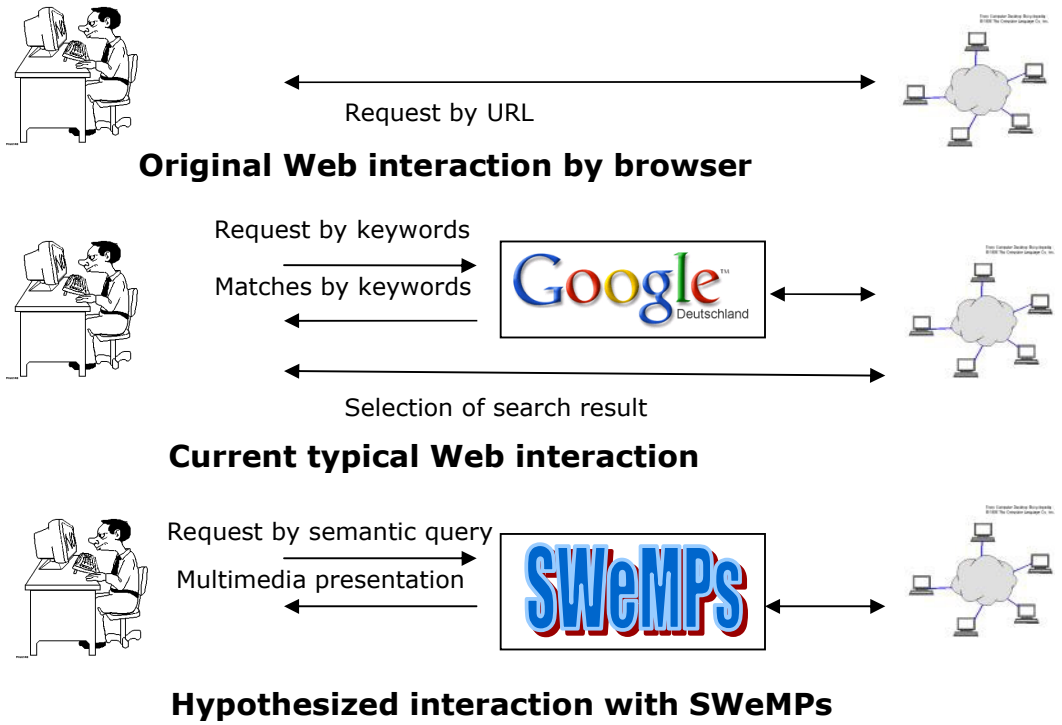

Figure 1.1 The Web evolution, present and future.

1.1.1 Definitions

This thesis deals with aspects of multimedia and multimedia presentation. It is recognised that these terms, in computer science, are often used to mean different

[4] Research website at http://swemps.ag-nbi.de
[6] http://searchwebservices.techtarget.com/sDefinition/0,,sid26_gci212612,00.html

things and so we briefly outline our understanding of this key terminology used in this thesis.

Multimedia – "more than one concurrent presentation medium" (WhatIs.com[6]). While text and images are different presentation mediums, multimedia generally refers to when audio and/or video are also used (other possible media are animated graphics and 3D models). A multimedia item may also be interactive.

Multimedia presentation – often used synonymously with multimedia, a presentation includes a temporal aspect (e.g. a slideshow) and typically offers non-linear interactivity (e.g. jump back or forward, or choose between parallel tracks).

1.2 Motivation

Technological advances have brought a paradigm shift in the accessibility and provision of media. At the turn of the 20th century available media were textual and graphic in nature (print and photography) and their circulation was geographically limited. By the end of the same century, we are experiencing worldwide, almost instantaneous accessibility to and provision of high quality media, including continuous audio and video, in particular through the Internet as well as television and wireless networks.

While many earlier steps have laid the foundation - such as radio and film, music and video storage, and the development of television and computers - we pick out three trends which came to prominence in the last decade of the 20th century which are setting the scene for this possible future of pervasive multimedia access:

- The World Wide Web (WWW)
- Interactive Television (iTV)
- Wireless and Mobile Networks and Devices

The **World Wide Web** grew out of a proposed new means to manage information through a distributed hypertext system, written by Tim Berners-Lee at the physics laboratory CERN [Berners-Lee, 1989]. The result was characterized as "a wide-area hypermedia information retrieval initiative aiming to give universal access to a large universe of documents."

Through the use of common, open schemes such as HTTP, HTML and URIs, and their use in Web browser software, Internet-based data became accessible and viewable regardless of its storage location, server hardware or local file addressing scheme.

While the Web was originally a *hypertext* system technological progress rapidly evolved it into a *hypermedia* system. Besides the early introduction of graphical support, increased storage capacity and processing power in computers and higher network bandwidths have made Web-based audio and video technically feasible, which could be pre-loaded or (increasingly) streamed. MIME (Multipurpose Internet Mail Extensions) headers are used by the HTTP protocol to inform clients of the format of requested data. Web-based multimedia - content consisting of multiple composite media types – is being supported not only from a plethora of proprietary

media players but also open standards such as SMIL [W3C,1998; W3C,2001], which has been developed by the World Wide Web Consortium as a simple to author language for synchronised Web-based multimedia presentations.

A direct result of the distributed decentralised nature of the Web alongside the simplicity and effectiveness of the URL referencing scheme is that Web content (e.g. a HTML page) could consist of references to different media. Media could be of different types and originate from different sources and yet be presented through the common user interface of a web browser or a media player.

Today's Web is characterized by the significant rise in availability and importance of non-textual media, as well as its use by non-Personal Computer devices. The emerging multimedia Web, as part of a general new trend in Web use termed by some as 'Web 2.0', will be returned to in the conclusion of this thesis. The first widespread introduction of Web-based content outside of the Personal Computer environment took place with the introduction of **Interactive Television** e.g. Microsoft WebTV [Feinleib,1999].

Televisions equipped with a Set Top Box (STB) can not only decode the digital audio-visual content sent from the broadcaster but can include software components to handle other forms of digital content, including from the Web. Hence, Set Top Boxes with an IP connection and Internet software have made it possible to access the Internet through the television set and offer users tools such as e-mail and Web surfing. However, Web usage is limited on a television set due to the lower resolution of the screen and the difference in usage – the user sits further away from the screen, usage is more commonly communal rather than individual and the input device is different (while TV keyboards and mice are available, the remote control is considered a more appropriate input device). HTML-based Web pages were explicitly created for the TV set (with specific layouts and HTML extensions) rather than existing content being re-used. These factors have limited the uptake of Interactive Television content.

Additionally, Set Top Boxes offer software platforms for which interactive applications can be developed, e.g. the DVB Multimedia Home Platform (MHP)[7]. These are mostly Java based, and can integrate with the Web components. This leads to 'convergence' applications which tie Web content with the audio-visual program. The interactive aspects can be tailored to the television environment and utilize the Web as a content source to 'enhance' broadcasts.

Now the Web is becoming even more pervasive as a wide range of **mobile and wireless** devices incorporate Internet connections and software for retrieving and displaying Web-based media content.

The initial impetus was with the mobile phone, which did not remain long a device purely for telephony. Text messaging (SMS) has been a huge success. A less successful approach has been Web access, hindered by the small screen and need for specific authoring of content (WAP, which used a simple mark up called WML). This resulted in a 'closed garden' approach to Web access, where the user was restricted to a service providers' selection of a specific group of content sources. Many phones support Java applications, providing simple interactive services. With

[7] http://www.mhp.org

the emergence of a new generation of phones incorporating digital cameras, there is an uptake of multimedia messaging (MMS), where images or short videos can be sent. Also, Web access has improved through the iMode specification, which uses a mobile-optimised Web browser which supports a simplified HTML and hence can provide less restricted Web surfing.

There are good reasons to believe that interactive applications, multimedia and the Web will increasingly be consumed by users of mobile phones. The upgrading of wireless networks to broadband (such as UMTS and 3G) will enable improved delivery of multimedia and interactive services. Standards exist which allow for sufficient compression of media to enable audio-visual streaming over wireless networks [Stockhammer,2003], and the next wave of mobile content is expected to be video and television [Södergård,2003]. Usage of mobile phones tends to be primarily personal which better supports the use of interactive content. Equally demand for content may be greater than with television, as the latter is used in the home where there is likely to also be a personal computer and other information sources.

None of these trends would be able to progress without the necessary shift in the software and hardware infrastructure that enables the generation, processing and delivery of multimedia content. On the hardware side, device processing power, network bandwidth and the means to generate and store multimedia content is growing steadily. Given that ever more powerful personal computers, set top boxes (evolving now into home media platforms) and mobile/wireless devices are coming to market and prices fall steadily to make the next generation of device affordable to consumers it seems that processing power to handle ever more complex multimedia will not be a severe problem. Likewise network bandwidth is becoming ever less of an issue as broadband Internet access becomes ever more widespread (in developed countries, at least), also in the mobile and wireless world (3G, WiFi). For content producers and suppliers, the means to create and store multimedia is becoming increasingly cheaper; hence economically there are fewer barriers to making rich multimedia content available.

These trends in Web content and delivery demonstrate that there will be a growing need and capability for access to information which potentially benefits from using multimedia in its presentation.

1.3 Use Cases

In this thesis we envisage a multimedia presentation system that will be able to dynamically retrieve media from different providers and can handle the appropriate presentation of that media to the user. What is required from such a multimedia presentation system in this hypothesized future of pervasive multimedia access?

A means to perform requirements analysis for a new computer system or piece of software that is common in Computer Science is to formulate scenarios, or ‘use cases’ [Bittner,2003]. The use case describes how the system will interact with users and other systems (or software components) in order to achieve a specific goal. For the thesis we formulate two scenarios, where given our motivation, it seems reasonable to expect that:

- Both scenarios are Web-based, i.e. they take place over the Web and use data retrieved from the Web.
- A scenario relates to the paradigm of interactive television, i.e. the resulting multimedia presentation shall include an audio-visual stream and is to be presented to a user on a television device (passive interaction style using e.g. a remote control).
- A scenario relates to the paradigm of mobile or wireless Web access, typically on a device with lower bandwidth and processing power than a household Web-accessing device such as a personal computer.

Hence in this thesis, two scenarios are used which exploit Web-based data and which are conceived as being targeted at users on either a (broadband) television-type device or a (narrower band) mobile/wireless device.

The first concerns family trees, and recognizes that a family tree as a multimedia presentation (consisting of images, text and synthetic shapes such as lines) could be expressed in a final format such as SVG Mobile [W3C,2003] and hence presented on mobile devices. Genealogies are not an uncommon example in Semantic Web/logic literature[8], given their obvious relevance to logical reasoning (e.g. being able to know that a grandson is a son of a son, or that daughter is the inverse property of mother). There is a standardized form of expressing genealogical data (GEDCOM[9]) and this has already been ontologized[10]. Furthermore, from the multimedia presentation standpoint a family tree has a graphical representation which follows clear rules and is widely understood.

The second concerns tourism, which is also a common theme in logic/Semantic Web scenarios[11]. Its economic value as a basis for e-commerce (online bookings) as well as the recognized problems regarding obtaining the ideal travel results which can encourage online bookings (e.g. personalization to user preferences, combination of different complementary offers, saving user time and effort by automating the retrieval of travel data) form the motivation for a number of research projects in this area, including – to mention only a selected few:

- *ReiseWissen*, a German national project focusing on providing improved hotel preselection and ranking personalized to users http://reisewissen.ag-nbi.de
- *SATINE*, a EU project which aims to integrate different tourism services to enable a comprehensive travel solution http://www.srdc.metu.edu.tr/webpage/project/satine/
- *E-Tourism working group* at DERI http://e-tourism.deri.at

[8] For example, as an example in the seminal "The Semantic Web" article by Tim Berners-Lee, Ora Lassila and Jim Hendler (2001). Other examples of using genealogies to illustrate logic/the Semantic Web are John Ramsdell's "A Foundation for the Semantic Web"(2001), "TRELLIS: An interactive tool for capturing information analysis and decision making" by Y Gil and V Ratnakar (2002), "Ontology Translation on the Semantic Web" by D Dou, D McDermott and P Qi (2003) and "KAON Server – A Semantic Web Management System" by R Volz, D Oberle, S Staab and B Motik (2003). This is not a comprehensive list but serves to illustrate that genealogy is a common example used by the logic/Semantic Web community.

[9] http://www.familysearch.org/GEDCOM/GedXML60.pdf

[10] http://www.daml.org/2001/01/gedcom/

[11] To give a few examples, "TourisT: the application of a description logic based semantic hypermedia system for tourism" (1998) by C Goble and J Bullock, "Applying Semantic Web technologies for tourism information systems" by A Maedche and S Staab (2001) and "Harmonize: a solution for data interoperability" by M Dell'Erba, O Fodor, F Ricci and H Werthner (2002).

Multimedia has proven to be a valuable asset in the tourism sector, as it can be used to present aspects of a travel request to users in an immersive fashion[12]. Systems designed for the multimedia presentation of tourism information are called "tourist information systems". Images, audio, video and interaction are used to illustrate e.g. a tourism destination, a hotel room or cultural aspects of a location. For our scenario, we consider a tourism-related television program. In the interactive television paradigm this program could be supplemented with additional tourism information accessible to the user through his or her remote control. Having motivated our choice of these two scenarios, we describe them in some more detail:

1.3.1 Photo-based Family tree scenario

In this scenario, we suppose a future situation where a family has a home network in which various consumer electronic devices are linked wirelessly and share data. We can imagine that the family members take their photos digitally and upload those photos into the home network (where the photos can then be viewed over various devices). Furthermore, these photos are organized in meaningful ways. While this could well be done manually by the members placing the photos in particular folders or tagging them with particular words, we hope that in the future semi-automatic annotation will be part of the technology. For example, image-based object recognition can be effective for well defined sets of objects. If we consider that set of objects to be the family members themselves, it seems that photos stored in the home network could be annotated effectively with the family members which appear in them[13].

Of course, users would like to have innovative ways to use these annotated images. One possibility is to display selected photos as a way to visually illustrate a family tree. Given that the home network can also be accessed over the Web (i.e. the annotated media are available to authenticated users) this would be a means for a family member to visually introduce his or her family to others while away from home (and presumably unable to introduce the family personally!) and possibly on the move (using a wireless, mobile device).

Figure 1.2 illustrates this use case, with the proposed multimedia presentation system as a client application in the home network which handles the organization of the photos and their delivery as a presentation to the user. Any user of the home network may at any time upload content to the network (e.g. the latest photos from their digital camera). Another user in an external location calls from his or her mobile device an "intelligent information service" – the family tree service, which is implemented in the multimedia presentation system – and receives as response the multimedia presentation of his or her family tree. The multimedia presentation system has the knowledge necessary for it to be able to access the photos and their metadata from the home network storage over some well defined API.

[12] Immersion of course could lead one to imagine here some sort of Virtual Reality scenario where a user dons a headset and finds him or herself "inside" the travel destination, Despite the apparent value of such an experience, Virtual Reality technology does already exist and yet has failed to capture the market (arguably due to resistance by users). Here, we mean something less than this 'total' immersion where audio, video and images are combined to communicate to a user the experience of being in some other location over a Web-enabled device such as a PC, television, mobile phone etc.

[13] An interesting example of this is the website Riya (www.riya.com) which uses face recognition software combined with initial manual annotations by users to automatically identify subsequent instances of people in photos uploaded onto the site.

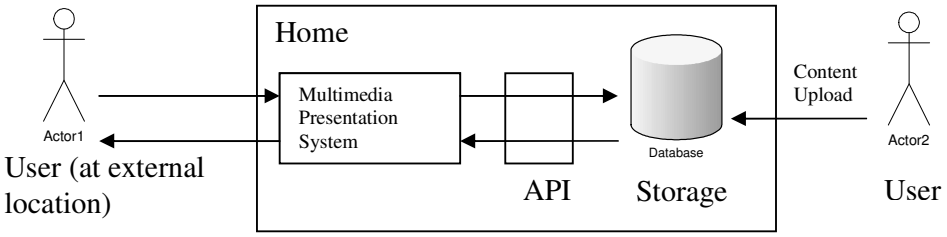

Figure 1.2 Family tree scenario

1.3.2 Interactive tourism television scenario

Digital interactive television is in a position to radically change how developed societies view and use television. It will result in a panoply of channels, programming on demand, and interaction with programmes based on elements inserted into the broadcast stream. There will also be new content services, based often on retrieving and displaying Internet content. Given the potential for rich interactive multimedia through digital television, we find a scenario based on interactive TV programming a suitable example for the proposed multimedia presentation system.

An audiovisual program guide to the sights of a city would be viewed traditionally in a linear and passive way and supplemented by static information sources that must be manually accessed by the user (videotext, SMS, Internet...). This has the disadvantage of limiting content to that which is supplied in advance by the system or incorporating a return channel which requires that the user leaves the programme being viewed or interrupts the viewing. Additional content is not tightly integrated with the programme and its content and is not usually immediately available at the (possibly brief!) moment when the user's interest is awakened. Furthermore, there is a limit in how much the content can be dynamically adapted or personalised to the user.

The scenario proposes to take travel information a step further by introducing a non-linear, interactive approach supplemented by dynamic, real-time, "push" and "pull" information from other sources. The actual provision of and interaction with the complete content bundle could be determined by the changing context, e.g. the user, device and location.

There are three types of sub-scenario foreseen where a client seeks travel information (outside of the normal, passive viewing where no push of interactive multimedia content need take place):

- Thematic interest viewing – segmentation of video into parts of interest with added background information
- Focused touristic viewing – supply of targeted tourist information answering the request of the user
- Immediate viewing – personalisation and dynamic search for information relevant to the user at the present location

We describe each of these sub-scenarios in more detail:

(i) **Thematic interest**

The user can choose a theme and views only the segments of the video which relate to that theme.
Each segment contains additional background material related to the theme.
Accessing the background material would be achieved by selecting an on-screen interactive element.
The background material is intended to add to the users understanding of the theme in the context of viewing the current segment.

(ii) **Focused touristic viewing**

The user can express a particular tourist interest and will receive content focused on that tourist interest. This additional information is designed to be obtrusive, i.e. that it leads away from main programme and the navigation of the user to another set of content which may be text, images or also audio or visual material with their own navigational structure. This content is provided to supply to the user the information about the city that he or she desires to have, that could not have been fulfilled by the passive viewing of the static tourist video.

(iii) **Immediate viewing**

This scenario filters or orders the previous scenario according to the user's present location. This assumes that the user is now accessing the content over a location-aware device, whether that would be a laptop or television in the hotel room or a mobile device while on the move. From the additional information that is available based on the tourist program and the users own interests, some objects have a given location within the city. Given that the user is also present in the city and making his or her location available, the content is ordered or filtered according to its proximity to the user at the moment of access.

These different sub-scenarios reflect how an interactive audio-visual program can be adapted to different situations. Figure 1.3 illustrates the use case, assuming the audio-visual program to be part of a standard broadcast stream and hence segmentation and non-linearity is supported by the client device through recording the program onto local storage. The enhanced content is generated by the multimedia presentation system broadcaster-side on the basis of the program metadata, available Web content and the user's own preferences and expressed information wish (provided by a IP back channel from the television set top box) and transmitted in a separate IP stream with timestamping to relate it to the program (in this scenario, we do not consider further issues of synchronising IP and broadcast streams as this is generally part of the iTV platform). The set top box of the user has a client application which can communicate with the multimedia presentation system, i.e. provides the user metadata and expressed information wish and receives the enhanced content and synchronizes it with the viewing of the program.

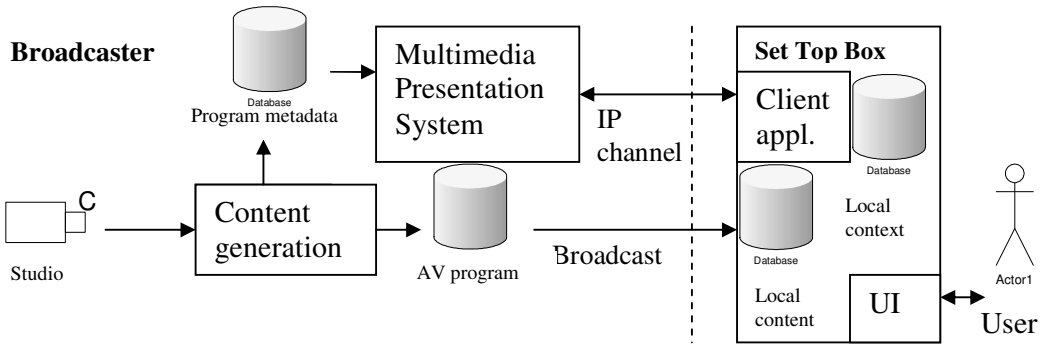

Figure 1.3 iTV tourism scenario

1.4 Requirements

Pervasive access to digital content suggests the end of the 'one size fits all' approach to content delivery and presentation. This is clearly inadequate when the wide variation of environments and situations in which the same content may be accessed is considered. If multimedia content in particular, which may require a higher level of concentration and interaction, is to succeed in communicating its intended message to its consumers, it needs to be relevant, non-intrusive and targeted to the particular user and device.

Fitting multimedia content to an increasing scale of possible contexts implies a non-trivial increase in manual content authoring. Hence content providers have a problem: content is only of value when its delivery and presentation matches the needs of the user and communicates the intended message, yet preparing content for every possible situation involves a prohibitive amount of effort.

Passing some of the task to a computerized system can be considered *automation* while ensuring the content is suited for the situation in which it will be consumed is the task of *adaption*. To support both these aims in combination requires a degree of artificial *intelligence*.

What are the requirements for a multimedia presentation system, which is automated, adaptive and intelligent? We derive the following requirements through consideration of the chosen scenarios.

- **The retrieval of data from different sources**

The Web as content repository has led to a movement away from the use of single centralized media repositories for the delivery of multimedia. Rather, the synchronized coherent multimedia presentation consumed by the user can be a conglomeration of media items from many different sources. A multimedia generation system would certainly need to be able to handle the automated retrieval of content from multiple locations.

For example, the family tree scenario needs to combine images from the family photo database with genealogical knowledge pertaining to the family. In the tourism

scenario the enhanced content added to the audio-visual program may be drawn from different sources: e.g. textual descriptions, images, maps.

- **The processing of data of heterogeneous type and format**

Multimedia is necessarily made up of multiple media objects of different types. Each media type can be represented by a range of (often competing) formats. Composite media formats allow for the representation of multimedia-like presentations, but their encoding fails to make any distinction between individual media objects. As a result, a system automatically retrieving content from multiple sources faces the challenge of being able to support the processing of any of the possible media types and formats it may encounter.

For example, in the family tree scenario the use of a resource-limited mobile device may place additional restraints on the choice of image format used in the presentation. In the tourism scenario, various types of enhanced content – images, animation, video, audio - will be retrieved that must be processed for use in the presentation.

- **The incorporation of contextual adaptation throughout the process**

Even when content is retrieved and processed to form a multimedia presentation, it is not necessarily in an ideal form for consumption. Pervasive content availability means access to content without restrictions: to any user, on any device, at any time, in any place. If provided content is to remain applicable in any situation, it will require adapting. Simple examples of this would be adding subtitles to a video not in the user's mother tongue, or changing the layout of a presentation to fit in a smaller device screen.

Adaption is made to a particular *context*. This can be the user, the device, the network, or the current location or time, or a given subject or situation. While the range of possible context increases, the importance of delivering appropriate content becomes even more essential. A multimedia generation system will be expected to support factoring all of these possibilities into its multimedia generation process.

For example, in the family tree scenario a mobile device has a small screen and limited resources hence the family tree presentation must be displayed in an effective manner e.g. through minimizing the image size (thumbnails) and providing a full screen image display only when a thumbnail is selected, or replacing images with text labels when network or device constraints make it necessary. In the tourism scenario, user and location can be catered to by selecting enhanced content based on the user's expressed interests and presenting it in terms of the user's location, emphasizing content relating to objects that are closer to the user.

- **The dynamic integration of external knowledge**

All of the requirements presented could expect that the multimedia generation system demonstrates some level of artificial intelligence. After all, the system will need to "know" how to find relevant media content, process it and adapt it to its presentation context. Given the extent of the knowledge that may be necessary to support retrieval from a content source so large as the Web, processing of content of various formats

and adaptation to any possible context in the situation of pervasive access, it seems unlikely and unreasonable to expect that all of this necessary knowledge will be available to the system at execution.

Rather, such a multimedia generation system may reasonably be expected to be able, during run-time, to acquire additional knowledge on the basis of its existing knowledge in order to extend dynamically its capabilities in terms of content retrieval, processing and adaptation.

For example, in the family tree scenario if the system finds that it can not display the images in the final presentation (e.g. the user is using a device that doesn't support them) it will need to determine textual labels to replace them, which it could find by examining the photo annotation metadata. In the tourism scenario, knowledge about the user and his/her location must be acquired, together with background knowledge to the tourist locations in the audio-visual program so that appropriate enhanced content can be found.

- **The presentation of multimedia on the basis of expressed concepts and their relationships**

Any coherent multimedia presentation needs to respect the meaning of the media objects in the context of the presentation as a whole. Media representing key subjects should be emphasized over media representing supplementary subjects, for example, or the ordering of media spatially or temporally should reflect location or chronology of the subjects of that media.

In a system automatically selecting content and making presentation design decisions, it has already been noted that knowledge must be available to the system with respect to how the system will locate, process and adapt content.

However that proves to be insufficient for generating the final presentation. The organization of different content in relation to one another is most meaningfully done on the basis of the relationships between the concepts that the content represents.

For example, in the family tree scenario we have a well defined and understood visual layout based on the genealogical relations between different people. In the tourism scenario, the display of enhanced content is more effective where its relation to the audio-visual program and to each other is taken into account. For example, content is positioned temporally in those segments of the audio-visual program in which the subject of that content appears. Visual ordering can be used e.g. to emphasize the media which relate to objects closer to the user according to his or her present location.

1.4.1 Problem Statement

To summarize then, we present the aim of the thesis as:

"In the light of the emerging media rich Web, to extend the state of the art of Multimedia Presentation Systems in terms of automation and adaptation, so that providing multimedia over the ubiquitous Web can require less manual preparation and exhibit more flexibility"

1.5 Outline of thesis

In this thesis, we apply knowledge representation techniques to the development of a generic multimedia presentation generation system. A conceptual model and framework will be presented and validated by a proof of concept implementation. The structure of this thesis can be understood in terms of the Rational Unified Process (RUP) [Kruchten,2003].

RUP describes how to deploy software effectively using commercially proven techniques. It is a framework or meta-model for software development, in that each software project can select only the needed features that are suited to it. In our case, as the intention is to produce a framework from which we can realise a prototype implementation, we disregard the aspects of RUP which are focused on commercial aspects and slim down the remaining features as this is individual research rather than the work of a large product development team.

Following RUP, we can divide the research carried out into four phases:

1. Inception
2. Elaboration
3. Construction
4. Transition

The initial problem statement, use cases and assessment of the outlined approach given in this chapter represents the inception phase. The following chapters, discussing the relevant fields, the state of the art and defining the proposed model and framework are the elaboration phase. Chapter 5 describes the construction of the prototype, and the evaluation in Chapter 6 plays the role of the transition phase.

Figure 1.4 shows the relationship between RUP and the thesis chapters (chapter 7, the conclusion, is not given here). The main arrow represents the research lifecycle. The thinner arrow which doubles back to inception reflects that RUP foresees now only that each phase is composed of iterations but that also that the development lifecycle can be iterative, i.e. after the evaluation of a developed product, the results may form the basis for a further inception of a new product. Likewise, we note that this lifecycle (which is the subject of this thesis) is in itself a third iteration (after a XML and a Topic Maps based development, respectively). Papers regarding these first two iterations are given in the appendices. Likewise, the conclusion in chapter 7 bears in mind that the results of this work can be the basis for further iterations in the field of semantic multimedia presentation systems.

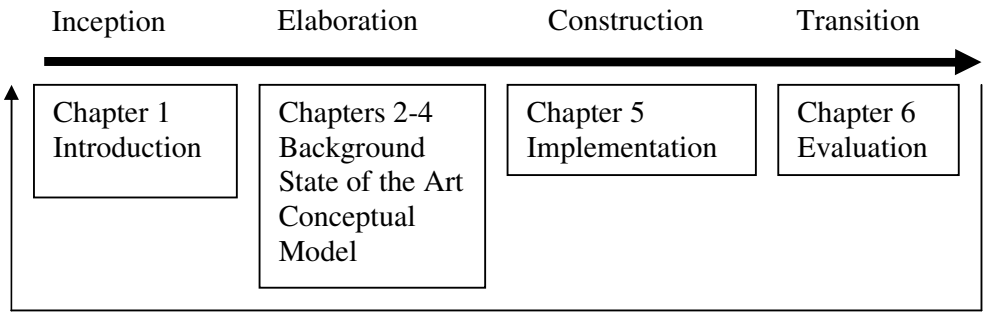

Figure 1.4 Visualisation of thesis structure in terms of RUP

The hypothesis of the work is that the use of a conceptual model to represent the generation process can improve upon applications for the automation and adaption of multimedia acquisition and provision. By introducing an automated and adaptive approach to multimedia generation, it may be possible to better support multimedia content acquisition and provision to the expected pervasive, digital, multimedia content consuming devices of the future. Now we introduce each of the subsequent chapters in a little more depth.

Chapter 2 introduces the groundwork for this and other contemporary research. Noting the current evolution in the Web and multimedia research communities to knowledge-based solutions, the identified problems in the field of multimedia generation, prior to the use of knowledge-based components, are reviewed. Knowledge representation techniques common to the Web and to multimedia are introduced, and the generic architecture for an IMMPS ("Intelligent Multimedia Presentation System") is outlined.

Contemporary knowledge-based approaches are reviewed in Chapter 3, in the context of the set of requirements laid out in this introduction. As a result of this review, it is possible to note that no single approach meets all of these requirements while the aims in general are in the scope of the work in Intelligent Multimedia Presentation Systems (IMMPS).

As a result of these findings we propose in Chapter 4 the framework for a Semantic Web-enabled Multimedia Presentation System (SWeMPs). This consists of a definition of the system in terms of components and process, a formal model of the domain ('ontology') of multimedia presentation generation and a generic rulebase for realising the multimedia generation process.

In Chapter 5, an implementation of the proposed approach is described. It has been created as part of this research to act as a 'proof of concept'. We describe the concrete implementation decisions that were made and outline the final versions of the rule-set and ontology which were used in the implementation.

An evaluation of this implementation, given in Chapter 6, focuses on the benefits of this approach in terms of supporting authoring multimedia presentation generation at the real-world conceptual level rather than at the digital media representation level. It analyses the value of this approach through the prototypical realisation of the two scenarios that were chosen to illustrate future requirements of such an intelligent multimedia presentation system.

Finally we conclude in Chapter 7 with an overview of the contributions of this research and proposals for future outlook and research directions for the field of semantically enabled multimedia presentation generation.

1.6 Scope of the work

While there is much research in the application of the Semantic Web to content retrieval and handling, its use with multimedia has not been a major part of the field. This is a result of Semantic Web practitioners not having multimedia backgrounds and likewise, multimedia practitioners not having Semantic Web backgrounds. As will be shown in Chapter 2, the use of semantic approaches for mixed media content retrieval has been researched but the issues of multimedia adaptation and presentation with semantic approaches have been less explored. The emergence of the awareness that Semantic Web and multimedia practitioners can learn much from one another can be seen in the launching of workshops to explore this overlap (European workshop on the integration of knowledge, semantics and digital media technologies, London, November 2004 and 2005; ESWC 2005 Workshop 'Multimedia and the Semantic Web'; WWW 2006 Workshop 'Semantic Web for Multimedia Annotation'). This research is positioned at this overlap between Semantic Web and multimedia and hence has been presented at three of these events and we participated in the remaining event.

The review of contemporary knowledge-based multimedia generation systems (Chapter 3) also shows that, in the context of research in multimedia presentation generation systems specifically, the use of semantic approaches is both a very recent development and yet has not made any significant progress in the past few years despite trends in the Web – broadband access, device ubiquity, more media content – making it ever more relevant, hence we consider this work to be part of a necessary emerging field for research and development (which we have christened 'Semantic Enabled Multimedia Presentation Systems') which is in need of fresh approaches. This thesis intends to motivate further research in this area and presents our contribution known as SWeMPs, both in the form of a conceptual framework and model as well as a prototypical implementation as a proof of concept.