Semi-Automatic Generation of Hypermedia Presentations: a starting point

Lynda Hardman, Brian Bailey, Jacco van Ossenbruggen, Lloyd Rutledge

Introduction

This document describes the work we plan to do in the SION project named above. It is an initial document, stating the results we wish to achieve and the currently proposed means of achieving them. In particular, the Standard Reference Model for Intelligent Multimedia Presentations (SRM) is described from the project's perspective.

The document is structured as follows. First an extract has been made from the original SION proposal describing the aims of the project. Then a detailed description of the SRM is given, including relevant insights for this project. Finally a description is given of the work carried out by Brian Bailey in the Design and Realization Layers during his stay at CWI in the summer of 1999.

The goals of the project

The biggest challenge in creating hypermedia presentations for databases containing multimedia objects is to automate the generation of these presentations as much as possible. This allows presentations to be tailored (or adapted) for the situation and needs of an individual user. The design and generation of such presentations should require only a minimum of human intervention. The research will capitalize on the growing resources of multimedia objects stored and maintained in multimedia databases and use these as a resource for creating hypermedia presentations.

The primary research question is to what extent we are able to automate the traditional hypermedia authoring process in order to generate acceptable quality, cost-effective, user-tailored presentations. Automatic generation of hypermedia presentations involves making design decisions regarding information content, spatial and temporal layout and hyperlink navigation, modalities and media used. Automatic generation of hypermedia presentations is currently only used in text-centric environments as an interface to strongly structured databases. Our research will address presentation generation in media-centric environments to interface semi-structured multimedia databases. It builds upon previous and ongoing research in multimedia presentation (see e.g. Ten Kate et.al. 1998), in timed hypermedia (see e.g. Hardman et.al. 1994), in adaptive hypermedia (see e.g. De Bra et.al., 1999) and in hypermedia presentations of volatile database information (see e.g. Houben et.al., 1998).

The development of an "authoring on demand" model for adaptable hypermedia requires investigation into the most appropriate knowledge representations for selecting and presenting media items in a coherent hypermedia presentation. This development can be taken one step further by adapting the representations and presentations to the individual user automatically, by using adaptive hypermedia techniques (Brusilovsky 96).

The innovation in the work stems from four separate issues:

- firstly, the generation of a document corresponding to a rich hypermedia format, including temporal synchronization of its constituent parts and the generation of links among multiple media (for example, the SMIL format [[ref]]);
- secondly, the re-use of existing media items in the creation of multiple presentations;

- thirdly the creation of presentations for widely varying client-side devices (e.g. palm top, mobile phone, television with set-top box or multimedia PC);
- fourthly, the automatic tailoring to the user's knowledge, preferences and task, in addition to the compute and network environment.

In broad terms, a hypermedia presentation should define what is presented to the user (the content), where it is presented (the spatial layout), when it is presented (temporal layout) and the navigation paths within the presentation (the links). Given a collection of user goals and availability of resources these four aspects leave open an wide variety of possible presentations. An Intelligent Multimedia Presentation System is a reasoning system aimed at generating, if not the optimal one, then at least an adequate one. The Standard Reference Model for Intelligent Multimedia Presentations (SRM, Bordegoni et al. 97) ("the SRM") describes a framework for generating multimedia presentations based on a user input goal. While the SRM seems at first sight also suitable for constructing presentations from existing media items, a number of additions and changes to the original SRM are needed. We describe the SRM and our current thoughts on necessary deviations from the model described in the original paper.

The SRM

The SRM describes a framework for distributing the various tasks necessary for generating a multimedia presentation among a number of modules. The SRM basically separates the procedural parts from the more declarative and knowledge intensive parts of the system. The procedural parts are split into five processing layers, the knowledge intensive parts into four knowledge bases: the "experts". The layers are mainly characterised by their input/output behaviour. They contain little or no expert-specific processing and concentrate more on processing information from multiple experts. An overview of the SRM is given in Fig. 1.

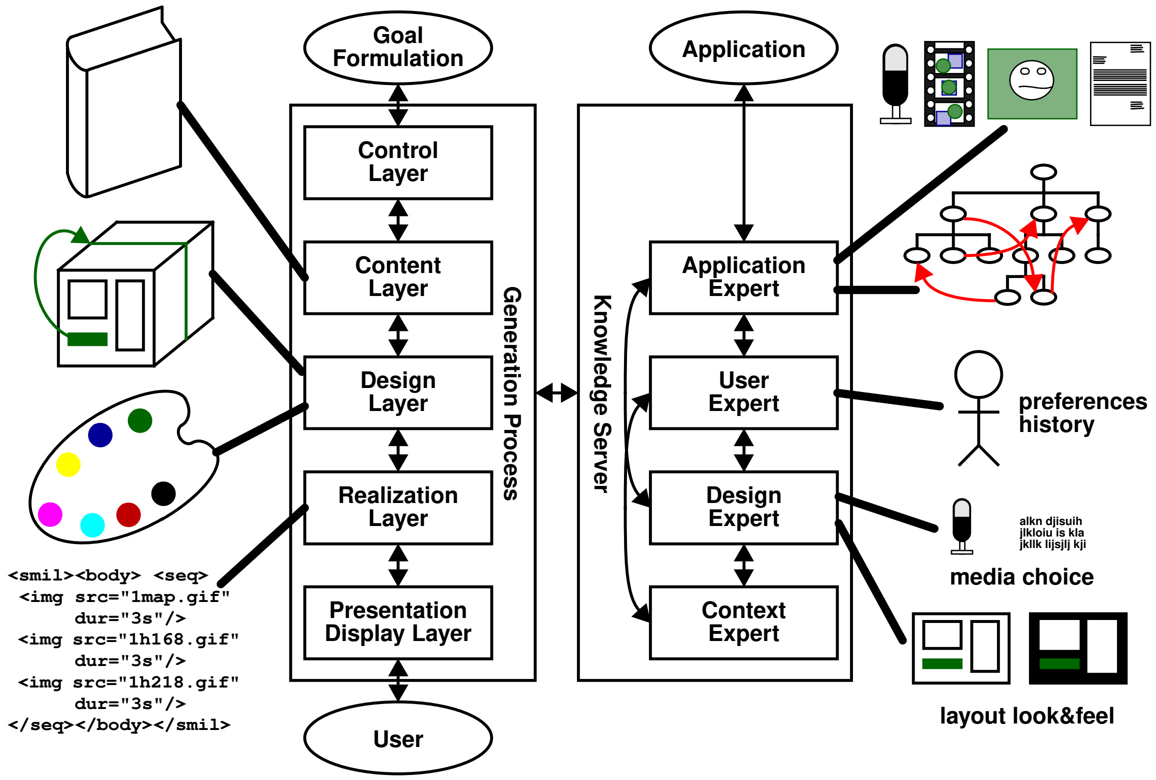

Figure 1. An overview of the SRM

The SRM divides dynamic presentation generation into two areas: generation process and knowledge server. The generation process performs the run-time generation of the presentation based on the user's interaction and information provided by the knowledge server. The knowledge server stores and provides explicit information needed to generate presentations. It also keeps track of the history of the presentation. The generation process is divided into layers, each of which controls one aspect of dynamically generating the presentation. The knowledge server is divided into experts, each of which maintains information about one aspect of what the generation needs to take into account.

We describes the SRM by going through each of the experts, detailing the information stored, and the layers, detailing the information needed and the processing to be carried out. We state where our description differs from the original SRM because of considerations for generating hypermedia presentations from existing media items.

The Experts

Application Expert

In the original SRM, this provides the application, or domain, specific knowledge. Namely:

- interfacing with the application system(s);
- information/data format conversion;
- provision of information/data;
- information/data characterization.

In particular, for this project, we use the application expert to access the following and pass on requested information to the various layers:

- a semantic representation[1] of the domain;
- the data stores containing all the media items. The media items are annotated with (possibly multiple) terms from the semantic representation. (We are assume for the project that this has already been done, e.g. the work being carried out at INA.)

The original SRM was tailored for the generation of media items, in particular text and graphics, rather than the selection of existing ones. The content of the application expert for graphics is, for example, a 3D frame model of, e.g., a modem or an espresso machine. At some point in the generation process a particular view of the object is generated. The semantics are determined initially and then the media item is generated to fit the requested semantics. In our case, the media item already exists, and is annotated with its relevant semantics. We will have to select the "best fit" of the available media items, rather than generating the "correct fit".

User Expert

This maintains the knowledge that forms the user model, comprising information such as:

- goals and plans;
- physical or mental abilities;
- attitudes and preferences;

1. As yet we do not know what form this would take. It may be a thesaurus, an ontology, a semantic net, a description logic. At this stage it is not important what the representation is, just that there is a representation and that it can be consulted and used for annotating the media items. Generalisations and specifications should also be possible. A list of requirements for the knowledge representation would be useful in the short term, where we modify these, on the basis of experience, during the course of the project.

- knowledge and beliefs.

For this project, preferences could include language, media/modality and rhetorical style. We may also consider some notion of "group profile", e.g. "MTV viewers".

Design Expert

The design expert in the SRM is pretty straightforward. It contains information needed, at least by the design layer, for selecting the appropriate media and presentation styles. (In the SRM it is also the expert where "anything else" gets put.) The SRM design expert contains:

- Media/modality model;
- design constraints;
- device model, that is a (partial) model of the user's current computational environment, comprising knowledge about the characteristics and availability of output devices.

For this project, we wish to be more explicit about the design knowledge, so that we can experiment with different types of design knowledge and how these affect the resulting presentation. In particular, we have divided the design expert into a number of "sub-experts":

- rhetorics expert. For the purposes of this project we will hard code the rhetorics we will use, but the idea is that for future work we could select which type of rhetorics to use for different circumstances (e.g. educational vs. entertainment applications);
- dimensionality expert [[better name?]] which determines in which way the chosen rhetorics are converted into the dimensions of space, time and navigation;
- style expert which determines the "look and feel" of the presentation, e.g. window layout, tempo for video/audio/animations/series of graphics, font colour and size for text;
- device model (as above).

[[I'm not sure whether the media/modality model becomes subsumed in the rhetorics expert, combined with the user expert, or whether this remains a separate entity. This can only be answered in the light of further work.]]

Context Expert

In the original SRM, this records the history of the interaction with the complete system. It includes:

- generation context, that is the presentation that has been generated so far;
- presentation context, that is the part of the generation presentation which has already been presented to the user.

In addition to this basic history information, for this project we also include the locus of interest and an intermediate representation format:

- The *locus of interest* (LoI) is the part of the semantic representation the user is currently interested in. It's a subset of the graph[2] annotated with strength values for the user's current interest focus. (This should be seen as very similar to Gene Golovchinsky's work (HT 97) with navigating around an information retrieval system. His locus of interest is translated into a query he passes to the IR engine.) The LoI may change with every user interaction. The history of the LoI is also recorded in the context expert.

2. Assuming our semantic representation is some sort of labelled graph.

- An intermediary format, generated by the content layer (see below), is stored here so that it can be referred to from any of the layers.

The Layers

The layers comprise a chain from a user entering a goal to a presentation being displayed to the user. Roughly speaking, the control layer processes input from the user into a presentation goal, the content layer determines the content (in semantic and media terms) to be displayed to the user, the design layer generates the layout and media item content, the realization layer produces a representation of a complete presentation and the presentation layer does the runtime display of the presentation to the user.

Control Layer

The control layer is responsible for generating a presentation goal on the basis of input from the user. Conceptually, the presentation goal represents the overall message that the system has been requested to convey to the user. As a result, the quality of the presented information is commensurate with the quality of the original goal specification. Once the presentation goal has been defined, the control layer decomposes the goal into smaller subgoals, encodes the subgoals into a machine-readable format, and then passes the encoded sub-goals to the content layer for further processing.

For the purposes of this work we assume that an appropriate sub-goal can be passed to the content layer. Within the scope of this project, we do not intend to implement anything more than an absolute minimal control layer.

Within the context of generating a presentation as the result of an (implicit) query on the existing media item stores, the output of the control layer can be seen in terms of a *locus of interest* of the user. An example of the locus of interest in database terms would be a query, expressed in terms of the database schema. In terms of more knowledge-intensive representations, the goal can be seen as an annotated subset of the information available from within the domain description (stored in the application expert).

The input to the control layer is a goal in terms of an information request (locus of interest).
The control layer splits the goal into a number of sub-goals.
The output from the control layer is a refined information request which can be processed by the content layer.

Content Layer

The content layer is responsible for selecting the appropriate media content that meets the information needs of the specified goal. The selection process is non-trivial and relies on extensive domain knowledge in order for the correct content to be selected. When the content is originally stored, semantic information must be attached that describes how the different content items relate to one another.

The original SRM describes a number of processes that take place within the content layer. We have taken these and adapted them to the case of selecting and combining existing media items, while also attempting to make the rhetorics of the presentation more explicit. At the moment we have identified four distinct steps which have to be carried out. We describe these below, but in outline the first step combines the semantics of the presentation with the desired rhetorics, the second step selects potential media types and modalities, the third step selects potential media items and the fourth step imposes an ordering on the presentation of the items.

At this stage in the project, we assume the process is driven top-down—a presentation goal is passed to the content layer, along with information about the desired rhetorics, and steps are taken towards generating a presentation. This makes sense in the original SRM where the media items in the presentation could be generated as necessary. In this case we wish to reuse existing items, so that the desired semantics may not be available in the desired modality/medium. We expect, therefore, that some bottom-up processing (i.e. starting with the available, relevant media items) will also have to be carried out. We ignore this issue in the following list of steps, but should bear it in mind when we get round to building implementations.

Note that while we go into great detail in this layer, it is not the intention to implement the layer in the short term. We found, however, that we needed a much clearer picture of what was happening in the content layer in order to think about the design layer sensibly. In particular, which information is needed, where it comes from and what it is used for.

Step 1 Narrative generation

The input to the content layer is an information request (the locus of interest).

The content combines the information request with rhetorical information.

The output from step 1 is a set of communicative acts and the relationships among them.

The user has a Locus of Interest (stored in the context expert) which is in terms of the semantic representation. For example, it may be an annotated subset of the semantic representation—where the annotations give an indication of how interested the user is in the various (hopefully related) concepts. Examples of such annotations are: core interest, related interest, marginal interest. Whatever means we choose of representing the LoI, it should express the complete semantics of the user's information interests (for this generation cycle).

Semantic information only is insufficient for generating a presentation (it can be compared with passing an SQL query to a database). In addition, there needs to be some notion of narrative for presenting the semantics to the user. This is where the rhetorical style comes in. The original SRM seems to implicitly assume a particular a rhetorical style. We want to make it more explicit, and ideally have the rhetorics as a potentially exchangeable component in the system. This was the reason we introduced the rhetorics expert within the design expert.

We are not yet clear on where the rhetorical information comes from. It may be passed to the content layer from the control layer, or it may be inferred from the user, design and context experts. In either case, the rhetorical information has to be combined with the semantic information. For example, the user has requested information on movies by a particular director. She is more interested in background information than dates and times of showing the movies. Among the rhetorical styles available are "tutorial", "informative", "entertaining". The previous information presented to the user was using an "informative" style, and there was no evidence to change this, so the semantics will be presented according to the "informative" rhetorics. This entails associating parts of the semantics with rhetorical information.

This results in the content layer producing a set of *communicative acts* along with relationships among the acts. A communicative act is a notion used in the original SRM, where it represents some piece of information to be communicated to the user. It stems from natural language work on speech acts and is intended for communications using multimedia. A communicative act is independent of media and modality but it does model constraints, preconditions and effects of the information to be communicated.

Step 2 Multimedia narrative generation

The input to step 2 is a set of communicative acts and the relationships among them, plus information from the design and user experts.

Preferred modalities and media types are associated with the communicative acts.

The output from step 2 is a set of media communicative acts (MCAs) and the relationships among them.

In step 2, each of the communicative acts from step 1 is allocated a number of media/modalities which can be used to express the communication. Each communicative act can potentially be expressed by a number of media. Some way of recording the preference for each is needed. At this stage we need to work with multiple potential media types, since we have yet to interrogate the store of media items for availability. Information from the user expert, on user media preferences, and the design expert (including platform dependencies), on the appropriateness of different media, is also required for this step. Context information (from the context expert) may also be needed to ensure consistency among several generation passes.

This process results in a set of media communicative acts (MCA's) along with relationships among them. A media communicative act is a communicative act with its associated modalities/media. The relationships have not changed during the process.

Step 3 Media item selection

The input to step 3 is a set of media communicative acts (MCA's) and the relationships among them.

Media items are selected from the store(s) of media items conforming to the desired semantics and media/modalities.

The output from step 3 is a set of media object communicative acts (MOCAs) and the relationships among them.

The output from step 2 is used as a means of querying the media item store(s) for media items suitable for expressing the MCAs derived in step 2.

At some point the MCAs need to be matched against the available media items in the data store(s). If there are too few or too many then some strategy has to be determined to increase or decrease the number passed through to step 4. On the other hand, this decision may be postponed to the design layer.

The result of this layer is the same collection of input MCAs and relationships but enriched with concrete media objects, which become a collection of MOCAs (media object communicative acts).

Step 4 Narrative ordering

The input to step 4 is a set of media object communicative acts (MOCAs) and the relationships among them.

A (partial) narrative ordering is imposed on the MOCAs.

The output from step 4 is a (partially) ordered list of MOCAs.

The ordering should not be seen as a literal order, but as a rhetorical order.

Note that in the original SRM, ordering was geared towards text-oriented documents, and a separate ordering step may oversimplify the problem for hypermedia. Instead, the rhetorics can be used to generate the inter-dimensional constraints directly in the design layer. On the other hand, order could be sufficiently important for the narrative, that a separate step is justified.

Prioritising is not explicitly modelled by the SRM, but for a hypermedia presentation it may very well be as important as ordering. Priorities can be used by the design layer (see below) to determine which items are part of the main presentation, and which items are only accessible via links, or come later in a temporal ordering. We think prioritizing should be done before the design layer.

Design Layer

At this stage in the processing we no longer do any content selection or media allocation. If the results passed to the design layer from the content layer are unworkable, then the design layer has to communicate this back to the content layer in some way.

In the design layer we concentrate on inter-media layout/style and intra-media layout/style. The former is concerned with the appropriate assignment of spatial, temporal and linking structures to the partially ordered MOCAs. The latter is more concerned with the characteristics of the individual media items.

Step 5 Inter-media layout

The input to step 5 is a (partially) ordered list of MOCAs.
This step determines the spatial, temporal and linking relations among individual media items.
The output of step 5 is a spatial/temporal/link "layout" for the presentation.

The decision as to whether items are laid out spatially, temporally or with links is made here. In addition to the output from step 4, consultation with the design expert is required in order to determine which space/temporal/link layout is the most appropriate to convey the rhetorics contained in the MOCAs.

The result is a skeletal hypermedia presentation including media items, spatial constraints, temporal constraints, navigational linking (constraints?).

Step 6 Inter-media style

This step determines the style for the overall presentation.

At this stage decisions on background colour for visual media, correct relative volumes for audio, size/colour or borders are made. This is in overleg met the design layer.

Step 7 Intra-media layout and style

This step determines the style for the individual media items in the presentation.
The output of step 7 is a complete, platform independent, description of the hypermedia presentation.

At this stage decisions on the layout within media items, in particular text, and media item styles are determined. In addition, highlighting of parts of media items most relevant to the user is carried out, e.g. underlining in text.

Realization Layer

The main task of the realization layer is to maintain format independence in the other layers of the processing model. The realization layer needs to know how to generate the different required formats, e.g. SMIL, MPEG4, dynamic HTML etc. This information is particular to the realization layer, and as such is stored as private knowledge in the layer itself.

Presentation Layer

The presentation layer takes the document format from the realization layer and presents it to the user. The software components of this layer will need to be integrated with the user interface components of the control layer. For the initial stages of this project we will use standard browsers.

Further Remarks

Prototype code has already been created for the realization and design layers by Brian Bailey during his 3 month visit in summer 1999. The results of his work are in the piece entitled "Automatic Generation of Electronic Program Guides”. Please note that this is an internal project document.

One of the things we have considered is the collecting together of a number of characteristics (from the design layer) into profiles. For example, the MTV profile would use a particular rhetorical style, particular user characteristics, preferred style rules and even some influence on the selection of content—the “BBC” profile would select longer pieces of text compared with the “MTV” profile. Creating and experimenting with these profiles would be a valuable part of the project.

References

M. Bordegoni, G. Faconti, M.T. Maybury, T. Rist, S. Ruggieri, P. Trahanias and M. Wilson, "A Standard Reference Model for Intelligent Multimedia Presentation Systems" , Computer Standards & Interfaces, vol 18 (6-7) 477-496.

Paul De Bra, Geert-Jan Houben and Hongjing Wu, "AHAM: A Dexter-based Reference Model for Adaptive Hypermedia", Proceedings of the ACM Hypertext 99 Conference, February 1999.

P. Brusilovsky (1996). Methods and Techniques of Adaptive Hypermedia. Journal on User Modeling and User-Adapted Interaction, Vol 6, 87-129, (Kluwer Academic Publishers).

Lynda Hardman, Dick C. A. Bulterman and Guido van Rossum, "The Amsterdam Hypermedia Model: Adding Time and Context to the Dexter Model", Communications of the ACM vol 37 (2), pp 50-62, 1994.

Geert-Jan Houben and Paul De Bra, "Generating hypermedia applications for volatile database output", Proc. of the AACE WebNet’98 Conference, pp. 220-225, November 1998.

Ten Kate, W., Bulterman, D.C.A., Deunhouwer, P., Hardman, L. and Rutledge, L. "Presenting Multmedia on the Web and in TV Broadcast", Proceedings of the 3rd European Conference on Multimedia Applications, Services and Techniques (ECMAST 98), May 1998