Towards the Generation of Hypermedia Structure

Lynda Hardman, Dick C.A. Bulterman

CWI

We present an approach for generating hypermedia presentations from multimedia information items distributed around a network. Our goal is to create a media-independent description of a presentation, from which multiple final presentations can be generated, taking into account the user's information need, the user's task and network and end-user platform resources.

In order to generate the structure of a hypermedia presentation from existing media items we need to define a way of grouping similar items and making links among the groups. This grouping can be based on semantic annotations attached to the media items. Current approaches to video annotation, as a complex example, are analysed. A number of research questions arising from our approach are discussed.

1 Introduction

More and more information, in a diversity of media types, is becoming available online, and is being changed and added to continuously. Users who access this information have differing information requirements, and differing hardware environments ranging from low-end personal computers to high-end graphic workstations. Furthermore, the information is accessed over networks with fluctuating available capacity. Publishers, whether corporate or individual, do not have the time, or money, to create the diversity of presentations needed to meet all these individual demands. Our goal is to (semi-)automatically generate these presentations, thus reducing the effort required by the publisher to cater for such a variety of users.

Our approach is to create a media-independent description of a presentation from which the required final presentations can be generated. A characteristic of material suitable for this generation process is that the domain is fixed, while the information itself is continually changing and/or being added to. Example application domains where this approach would be appropriate are news [12], weather reports, and patient medical records (e.g. X-ray results with voice annotations of diagnoses).

In the following section we describe the stages in the process of generating a hypermedia presentation — our "information pipeline". In this article we concentrate on one stage of the pipeline, that of generating the structure of a hypermedia presentation from previously retrieved media items. We discuss this process of hypermedia structuring in Section 3, along with the need for semantic annotation of the media items. Work on annotating video is used as a starting point for analysing the pros and cons of current annotation approaches. The generation of hypermedia structure is only possible if sufficient information is known about the candidate media items. In Section 4 we list a number of concrete research issues raised by the suggested approach. A final section describes the status of the work.

2 An information pipeline for networked multimedia

We divide the process of creating a hypermedia presentation from multiple sources into a number of stages, as shown in Fig. 1. Media items, (a), are objects of a single (possibly composite) data type which are classified, stored and retrieved as logical units. These are semantically annotated with descriptions of their contents (shown in the figure as small hatched boxes). Some items (b) are selected (for example by a retrieval program) to be part of a presentation. These are assembled into a coherent presentation by selecting (small) groups of items which can be displayed together and creating hyperlinks among the groups. The resulting document structure (c), conforms to, for instance, the hypermedia model described in [8], which defines the information required for specifying hypermedia grouping and linking of static and dynamic media. The document structure is then used to generate a platform-independent presentation specification including layout and synchronization constraints (d). Note that the semantic information (denoted by the small hatched boxes in (c)) used to generate the playback constraints is no longer needed for the selection of media items or structure generation. The final step is to take the machine-independent representation, (d), adapt it according to the end-user's hardware (which can assumed to be static) and (constantly changing) network load and play it back on an end-user's hardware (e).

The part of the pipeline on which we concentrate here is the creation of a hypermedia structure, (c), from the relevant selected fragments, (b).

3 Media-independent annotation for structuring

The problem we investigate here is of generating structure from pre-selected, relevant media items. The resulting structure has to be media independent, since we know only at a later stage in the information pipeline which media are most appropriate and what resources will be available for transporting the media items.

Information required as input to the structuring process is:

- the user's information need (in terms of the domain description);
- a selection of relevant media items annotated with semantic information (derived from the domain description);
- a domain description, which expresses the relationships among the annotations.

The output of the structuring process will be a document description conforming to [8], specifying hypermedia composition, links and contexts, but at this stage not specifying layout and timing constraints.

The process of step (b) to (c) in Fig. 1 is the following:

- compose related media items into groups suitable for simultaneous display;

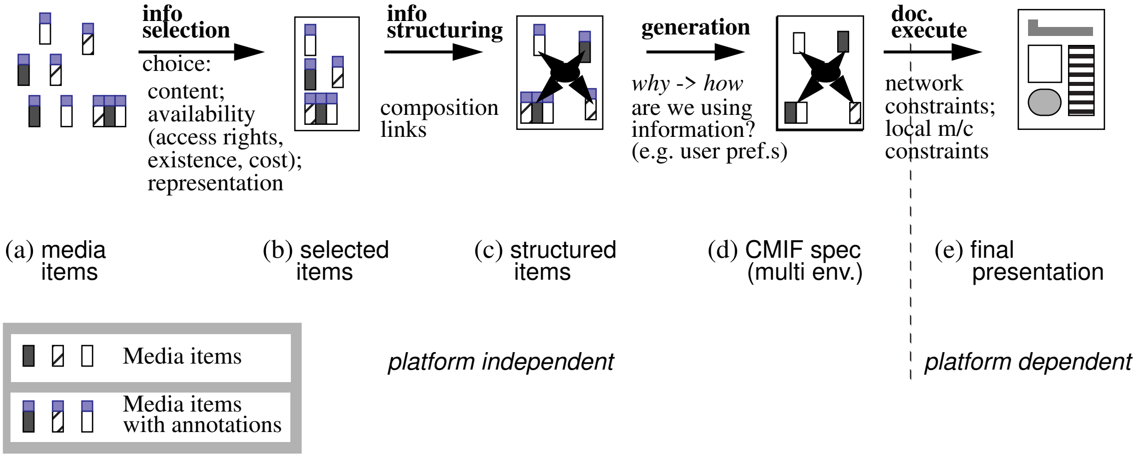

Figure 1. *The Information Pipeline*

- create links among the groups.

The most important requirement for this approach is that semantic annotations are attached to the media items. The annotations should be media independent so that media trade-offs in the presentation can be made further down the pipeline. This media independence is guaranteed through the use of anchors, introduced into hypertext for preserving the media-independence of links. An *anchor* [6] specifies a part of a media item by combining a unique, media-independent identifier, referring to the part, with a data-format dependent description of the part. In our case, where we wish to annotate part of a media item, the annotation is bound to an anchor. This conceals the data-format of the annotated part, and so makes the annotation media-independent.

To obtain a sufficiently powerful annotation scheme, we have investigated video as a complex example—it has two spatial dimensions, as well as a temporal dimension (although it lacks the symbolic "dimension" of text). Burrill et al. [4] attach annotations via fully-fledged video anchors, corresponding to the real-life objects, e.g. the anchor corresponding to a bouncing ball is the collection of locations of the ball in consecutive frames. The anchors are annotated with descriptions, although it is not clear how the descriptions relate to one another, and whether there is a domain description connecting them. Davis [5], on the other hand, has a large, and apparently encompassing, domain description which is used to annotate video clips. The clips are frame extents that can have multiple (and overlapping) descriptions attached. Note that these annotate sequences of video frames, and do not specify parts of the video frame using anchors. Hjelsvold & Midtstraum [9] also use sequences of frames as a base for annotations. Smoliar & Zhang [14] use an approach similar to that of Davis, but the domain representation is less powerful (being a tree structure rather than a directed acyclic graph). Smoliar & Zhang, however, use a partial anchor concept, where areas within a frame are also annotated (a step towards Burrill et al.'s full anchor annotation).

The work of the WIP project, [1], [2], where multimedia presentations are generated, is also relevant to the discussion of annotation. Here, a very detailed domain description has graphical items associated with objects in the domain (e.g. a picture of the lid of an espresso machine). Text can also be generated for the final presentations. The emphasis, however, is very much on the domain description and not on the annotation of a large number of media items. This seems to be similar to the goals of Lester & Porter, [10], who describe their project on generating explanations for biology students, using an annotated store of diagrams, photographs and animations, along with generated text.

We intend to use a combination of these annotation approaches, in particular the anchor annotation described by Burrill et al. [4] and the domain description such as used by Davis [5].

4 Research issues

The stages along the information pipeline described in Section 2 are not straightforward to implement. At the beginning of each process a number of inputs are required, and the output has to conform to that required by the following stage. In this section we discuss a number of the problems we expect to encounter at different stages in the pipeline.

Existence of annotation

In order to make use of media annotations for grouping items into presentations (Fig. 1 (b) to (c)), the annotations need to exist. Annotations are not yet common place, but are becoming more common as non-text databases of information are being created [9], [12]. The ultimate solution will be when the different media can be processed for automatic annotation, e.g. video [14]. Much of the work on annotations is to enable retrieval, and it is not yet clear whether the same annotations are sufficient, or necessary, for composition.

Part of the contribution of our work is to help define the minimal requirements for these annotations, rather than demanding maximally rich or complex annotations. In particular, investigating what the trade-offs are with different complexities of annotations. For example, annotating the objects visually present in video may be sufficient, rather than trying to attach higher level concepts [4], [5].

Determining semantic distance between annotated items

A possible method for selecting items to be displayed together on the screen (i.e. producing the hypermedia composition) is to choose the items semantically closest to the user's "query". The domain description can be used to calculate such a semantic distance. This method may not deliver the optimal result, since there may be other items further away from the query but whose mutual semantic distances are closer, making them more suitable for grouping. If a measure of semantic distance is used, then it remains unclear how to calculate this distance, since one media item may have multiple annotations (e.g. a video containing images of a cars (transport) and houses (buildings).

Timing relations

Timing relations need to be taken into account when grouping items and creating links to groups, e.g. anchors are not present at all times in temporal media. When linking to an anchor we need to decide where to start displaying the media item containing it—at the start of the anchor or at the beginning of the item (this holds not only for temporal media such as video and sound, but also for text and large pictures). Specifying hypermedia contexts at the beginning and ends of the links, as described in [7], can be used as a first approach to this problem. These allow the specification of exactly how much information is departed from or arrived at when following a link.

Final selection of media items

Having created a hypermedia structure, a final selection as to which media types should be transmitted through the network has to be made—(d) to (e) in Fig. 1. Although the choice of items so far has been made on semantic relevance, it is possible that certain media are inappropriate, for example because the user is carrying out an eyes-on task (e.g. landing an aeroplane), or working in a noisy environment. Appropriate selections of media for cognitive processing also need to be made, e.g. two videos are difficult to follow simultaneously. Even in human-authored hypermedia applications we are still learning about the best ways of using the different media [11]. We do not intend to state the rules for these choices, but to allow the choices to be taken into account in the selection process.

Consistency of history

When a user is navigating through generated hypermedia presentations it would be very disconcerting to go back to a previously generated presentation which has been re-generated (from the same initial conditions) but now contains different media items. To give the user a more stable environment a global history can be kept of all the presentations ever created. The user's local environment is unlikely to be able to store all the data ever down-loaded to it, but a list of the media items used would at least give the ability of recreating previous presentations. (This would be similar to the global history of previously accessed information on WWW browsing systems, such as Mosaic [13] or Netscape.) The current network load would determine whether better or worse quality versions of the media items can be sent, but the *same* items should be sent.

Influences along the information pipeline

Having selected the most appropriate media items to be transmitted to the user, it may be that the network is too busy and the best we can do is to send low quality representations of the chosen items (e.g. send only the first frame of a video accompanied by text subtitles rather than speech). On the other hand, we may want to take the media-type into account earlier in the pipeline, e.g., knowing that the network is full, selecting only low-bandwidth items (e.g. a picture and text) to be composed into the hypermedia presentation—otherwise, computational effort is spent on creating structures using items that were not suitable in the first place. It may be, however, that to generate a potential presentation is not costly, so that a number of possible solutions can be tried and thrown away. The desired solution should take decisions about which particular media items to use at the latest possible moment, but the information in the structures created up until that point should allow the generation of the desired presentation.

5 Current Direction of Work

Our next step in this work is to begin some experimental implementations, to investigate how much domain representation is needed to generate adequate hypermedia presentations, and how satisfactory these generated presentations are. Our goal is to produce acceptable presentations in a timely manner for minimum pre-annotation effort and minimum computational effort. If these first results prove encouraging, we will be able to investigate the gains of investing more effort in better annotations and better grouping/linking heuristics.

This topic is related to other work going on in the Multimedia Kernel Systems group at CWI, [3], which is concentrates further towards the end of the information pipeline, (d) to (e) in Fig. 1.

References

[6] André, Elisabeth, & Thomas Rist, "The Design of Illustrated Documents as a Planning Task" in "Intelligent Multimedia Interfaces", ed. Mark T Maybury, AAAI Press/MIT Press, 1993 ISBN 0-262-63150-4, pp 94 - 116.

[7] André, Elisabeth, & Thomas Rist, "Multimedia Presentations: The Support of Passive and Active Viewing". Deutsches Forschungszentrum fur Kunstliche Intelligenz, Saarbrucken, Germany, Research Report RR-94-01.

[8] Bulterman, D.C.A., "Specification and Support of Adaptable Network Multimedia", ACM/Springer *Multimedia Systems*, 1(2) September 1993.

[9] Burrill, V.A., T. Kirste & J.M. Weiss, "Time-varying sensitive regions in dynamic multimedia objects: a pragmatic approach to content-based retrieval from video", *Information and Software Technology* Journal Special Issue on Multimedia 36(4), Butterworth-Heinemann, April 1994, pp 213-224

[10] Davis, Marc, "Media Streams: An Iconic Language for Video Annotation", Telektronikk 5.93: Cyberspace Volume 89 (4) (1993) 49 - 71, Norwegian Telecom Research, ISSN 0085-7130 http://www.nta.no/telektronikk/5.93.dir/Davis_M.html

[11] Halasz, F. & Schwartz, M., "The Dexter Hypertext Reference Model", *Communications of the ACM* 37(2), Feb 94, pp 30-39

[12] Hardman, L., Bulterman, D.C.A. & Van Rossum, G., "Links in Hypermedia: the Requirement for Context, *Proceedings of Hypertext 93*, Seattle, Nov. 93, pp 183-191

[13] Hardman, L., Bulterman, D.C.A. & Van Rossum, G., "The Amsterdam Hypermedia Model: Adding Time and Context to the Dexter Model", *Communications of the ACM* 37 (2), Feb 94, pp 50-62

[14] Hjelsvold, Rune & Roger Midtstraum, "Modelling and Querying Video Data", in Proceedings of the 20th VLDB, Santiago, Chile, 1994.

[15] Lester, James C., Bruce W. Porter, "Designing Multi-Media Knowledge Delivery Systems: The 'Strong Representations' Paradigm", AAAI symposium on Intelligent Multi-Media Multi-Modal Systems, March 1994, Stanford University, USA, pp 64-72.

[16] Liestøl, Gunnar , "Aesthetic and Rhetorical Aspects of Linking Video in Hypermedia", in proceedings of (6th ACM conference on hypertext) ECHT '94, pp 217 - 223

[17] Miller, Gene, Greg Baber & Mark Gilliland, "News On-Demand for Multimedia Networks", 1st ACM conference on Multimedia '93, 383 - 392

[18] Schatz, B. R., & Hardin, J. B. (1994). NCSA Mosaic and the World Wide Web: Global Hypermedia Protocols for the Internet. *Science 265*, 12 August 1994, 895 - 901.

[19] Smoliar, Stephen W & HongJiang Zhang, "Content-Based Video Indexing and Retrieval", IEEE *Multimedia*, Summer 1994, 62 - 72.