Creating Meaningful Multimedia Presentations

Lynda Hardman
Jacco van Ossenbruggen
Centrum voor Wiskunde en Informatica
Kruislaan 413, P.O. Box 94079
1090 GB, Amsterdam, The Netherlands
Email: Firstname.Lastname@cwi.nl

Abstract— Finding relevant information is one step in the chain of understanding information. Presenting material to a user in a suitable way is a further step. Our research focuses on using semantic annotations of multimedia elements to increase the "presentability" of retrieved information. We investigate the use of domain semantics and discourse semantics for improving information presentation. In particular, we present work showing the use of domain semantics for grouping search results, the combined use of domain and discourse semantics for creating a particular genre of multimedia presentation (biography) and finally show the combination of argumentation structures and domain semantics for generating coherent sequences of video material. We provide comments and insights into the viability of these techniques.

I. INTRODUCTION

Generating high-quality hypermedia presentations tailored to different end-users carrying out their own tasks is a goal that is unlikely to be met in the short term. There are, however, a number of steps that can and have been taken which take us closer to this goal. In this paper we describe issues we have come across in different projects and reflect on their implications for the development of the field.

Initial fundamental work carried out is the Standard Reference Model for Intelligent Multimedia Presentation Systems (SRM-IMMPS) [4]. Here the authors abstracted from their experiences with automatically generating multimedia presentations and developed a higher-level model which can be used as a basis for developing future systems. The approach is plan-based, with the underlying assumption that the media assets required in the presentation will be generated. The work reported here, however, considers the pre-existence of a heterogeneous collection of media assets and how these can be selected and combined together into a single coherent presentation.

Personalisation of presentations can be carried out at different stages of the presentation generation process: topic selection within the domain; discourse genre and graphic design style; the environment or context, in particular the user's device capabilities. A theme of our work is the role that explicit discourse information plays in the generation process. This is the difference between a list of retrieval results ordered most relevant first and a presentation that has structure interpretable by the end user, giving the collection of "results" the sense of belonging to the same presentation. We discuss the representation of discourse and the expected benefits in

section II. In section III we explain the process used in two different systems to generate a personalised presentation using the media assets accessible to the system. In section IV we discuss the problem of converting a large semantic graph structure to a human-consumable presentation format and the need for an intermediate processing step.

II. ADDING DISCOURSE TO THE PRESENTATION OF QUERY RESULTS

Broadly stated, our experience has been that the more expressive the discourse the more effort is required to attach meaningful annotations to media fragments. However, a small amount of discourse can go a long way. In essence, different amounts of explicit discourse information can be used to improve the structuring of results sets from queries. More detailed analyses of a number of these systems are given in [6].

No Explicit Discourse: In Noadster [13] no explicit discourse information is included, but rather domain semantics are used to generate clusters of results which are presented to users. This gives the familiarity of hierarchical structuring, while allowing the same result to be classified in all relevant groups. Topia [12] uses similar clustering mechanisms, but is more domain specific. Topia allows more meaningful grouping by pre-selecting concepts in the domain semantics on which to do the clustering. An interface is also provided to allow users to indicate their own indication of importance of a topic. These systems illustrate that grouping, while straightforward to implement, gives a user extra useful information while exploring a repository.

Museo Suomi [9] gives flexible access to cultural repositories in Finland, allowing users to select the topics they want to see. Again, there is no explicit discourse information included, but, by giving users direct access to selection based on the underlying domain semantics information, selection becomes more transparent.

Deriving Aspects of Discourse: Some first steps have been made in incorporating small amounts of discourse semantics into further improving the grouping and structuring of query results. Rutledge et al. [12] investigate how meaning can be given to the clustering process. The clusters determined from the underlying domain semantics have no intrinsic difference to the user (other than the user selecting topics of interest). However, a method was derived that assigns different weights to clusters, allowing only the most significant ones to be

included in the hierarchical structure presented to the user. In addition to selecting what clusters to make and which to present to the user is the question of ordering. In which order should the clusters be presented and, within a cluster, in which order should the items or sub-clusters be presented. This ordering influences the way people perceive the information (for example, consider the numbers of companies who choose a name starting with an "A".) The cluster weights can be used to present the heaviest, hopefully the most significant, clusters first, throughout the hierarchy.

Alternatively, numeric properties provide a domain-independent means of sorting items sharing a property. The value of a numeric property of each item within a cluster can be used to determine the presentation order. This could be extended to processing which takes into account the minimum, maximum or average of these values.

An item can be included in more than one cluster, and as such it can occur at more than one position in the cluster hierarchy. This information can be used to present the items as part of the hierarchical structure, but using a distinguishing means of display. Thus the number of clusters is not increased but information grouping the items together is not lost but conveyed in a different way.

Fixed Discourse: In the systems discussed above, an attempt is being made to allow discourse to emerge from an analysis of the underlying semantic structures and the associated media assets. An alternative approach to include discourse is to specify it up front.

Media Streams [5] allows a user to query a repository of semantically annotated video fragments. By giving a structured query, the results returned can form a video sequence that can be perceived as part of a continuous whole by the end-user. Here the underlying semantics is used to aid the user in selecting the video fragments and the query structure for ordering them for presentation.

DISC [7] uses an annotated multimedia repository and a domain ontology to create multimedia presentations on demand. A semantic graph is created which consists of a domain ontology of classes, instances and relations between them together with the media material related to the instances. Discourse knowledge is represented explicitly as a set of rules:
- what kind of genre can be applied to a certain main character (a biography for a person);
- what types of narrative units are relevant for a certain genre (a narrative unit describing personal life);
- what types of characters can appear in what narrative unit (a *Personal Life* unit talks about the main character (a person) but can also talk about a spouse of a person as a related character). Characters are mapped to domain classes (Person, Artist);
- what types of domain relationships are relevant for those characters. Inside a *Personal Life* narrative unit isMarried relation leads for example to a Spouse related character.

The goal of Artequakt [10] is to automatically generate biographies of artists from knowledge extracted from the web and maintained in a knowledge base. Each text fragment is annotated with the concepts from a domain ontology.

In order to build a presentation, Artequakt uses human authored templates of discourse structures. Templates have been constructed for the biography genre. A template consists of a number of queries that can either retrieve the desired information from the database or construct sentences dynamically by retrieving specific facts from the ontology. A query is composed using domain classes and relationships between them. For example: ?artist died ?date_of_death in ?place_of_death. The overall structure of a template consists of several sub-structures to define the order in which concepts should be presented: *Sequence*, *Concept* and *Level Of Detail*. The basic *Sequence* structure defines an ordered list of queries that are instantiated from the database. *Concept* represents a set of queries rather than a sequence, meaning that any the queries can be executed at that point of the story building process. A *Level Of Detail* structure allows definition of the ordering preference in the query instantiation meaning that if the highest numbered query cannot be used, the next highest should be taken. A template can also contain additional contextual information which can be used to adjust to different user characteristics (e.g., expert vs. novice).

III. Two example systems

We personalise a presentation by allowing a user to specify what the information is that they are interested in and present it to them in an appropriate way. In this section we describe two example systems in detail.

Personal screen size

While many web sites are generated on the fly from information stored in an underlying database, few allow the structure of the information displayed to change depending on the user. The Cuypers generation system [14] allows a user to select topics available in an underlying database and generates a presentation structure suited to the constraints of the user's device.

The database underlying the web pages of the Rijksmuseum[1] was used to give access to fragments of text and images of paintings in the museum. Cuypers takes as its input a structure expressing a fixed discourse, in this case from Rhetorical Structure Theory [11]. The structure expresses a core concept, selected by the author from the Topic Selection menu, accompanied by examples illustrating the concept. The user can select an artist and the illustrative examples are chosen from the artist's oeuvre. When a selection is made in either menu then the items in the other are updated to ensure that the user is not confronted with an empty set of results. Human-authored text is selected from the database to explain the core concept, while images of paintings are displayed one after the other to illustrate it. Having the text spoken allows a user to listen to the explanation while looking at the different images.

The description so far explains personalisation only at the topic level. The system also allows personalisation of the

presentation to the screen size. In the demonstration[2], the user can select different screen sizes to generate to. In a fully functioning system the intention would be to allow the device itself to communicate to the server to state its requirements (e.g., colour depth, screen resolution) and preferences.

Altering a multimedia presentation to the screen size requires an understanding of how the different parts of the presentation relate to each other. One extreme is to display all media assets at once. This would require large amounts of screen estate and potentially conflicting audio tracks. The other extreme would be to display one media asset at a time. This would take a long time and any coherence within the presentation would be lost. An explicit representation of the rhetorical structure is used by the system to make informed decisions about where to separate (groups of) elements for presentation to the user. This is similar to the approach in [1], developed for page-based documents. The system attempts to display as much of the presentation as possible at the same time, but where this overflows the screen space limits, tries another way of partitioning the presentation so that it fits [8]. For example, while each member of a sequence of example images is shown the title and descriptive text should remain on the screen. Only if screen space is extremely limited is the descriptive text shown first, followed by the examples. The rhetorical structures used in Cuypers are encoded explicitly and are not themselves generated. The generation of discourse structures dynamically is addressed in the following section.

Personal opinion

Current television and film culture has developed the documentary genre, where a director carefully selects and orders fragments from large amounts of filmed material for presentation to a wide audience. The majority of the captured footage is not used and many arguments are not presented to the user. Vox Populi [2], [3] is an attempt to give users control of the documentary they see and also to make them aware that there are always many sides to an argument.

Vox Populi[3] supports the generation of argument structure within video sequences, ordering the constituent fragments according to argumentation discourse rules. A database of video fragments has been annotated with concepts from the argument structure (explicit discourse), the domain ontology — the question of initiating a war in Afghanistan, and characteristics of the speaker. A user is able to submit their own query, for example, to allow an argument clash to be generated between two speakers or to generate support between the same two speakers. The terms used in the annotations are related through a thesaurus, which allows the system to find terms which are more generic than, more specific than, synonym of or antonym of a term. Argument structures are created using statements consisting of a `subject`, a `modifier` and a `predicate`. A user can query the system by selecting a question to be discussed, an initial opinion, characteristics of

[2]http://www.cwi.nl/~media/demo/cuypers/
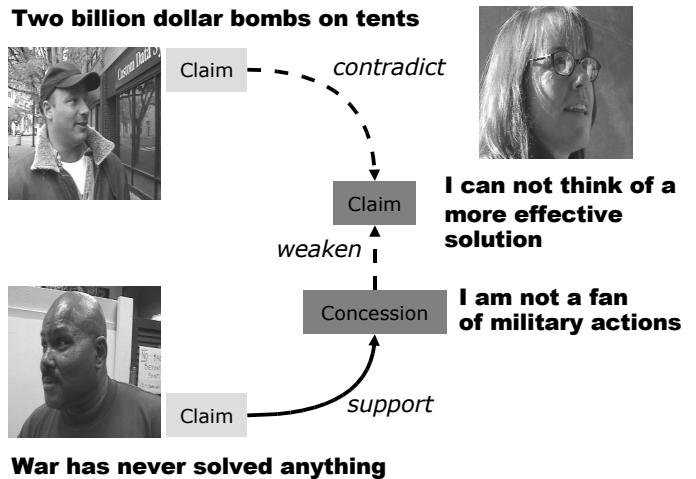[3]http://www.cwi.nl/~media/demo/IWA/

Fig. 1.   Vox Populi: assembling contrasting points of view [3]

two speakers and whether the argument should be supported or contradicted.

An example query is the following. The engine selects an interview which is in favour of the chosen subject (the woman on the top right of the figure 1 saying: "I am not a fan of military actions, but in the current situation I cannot think of a more effective solution"). The discourse annotations for this statement decompose it into two parts, the Claim ("I cannot think of a more effective solution") and the Concession ("I am not a fan of military actions"). To contrast her point of view, the engine chooses to support the Concession and contradict the Claim: for the former it selects the man on the lower left saying "War has never solved anything" and for the latter the man on the upper left saying "They are using two billion dollar bombs on ten dollar tents". The Concession is used as a rebuttal of the Claim, i.e. it contradicts the claim even though it is less strong than the claim. Supporting the Concession makes the rebuttal of the claim stronger. Contradicting the Claim constitutes another form of rebuttal. The final video sequence is as follows: woman saying "I am not a fan of military actions"; lower man saying "war has never solved anything"; woman saying "in the current situation I cannot think of a more effective solution"; upper man saying "two billion dollar bombs on tents".

Current work has focussed on the creation of the framework which allows the explicit manipulation of the argument structures described. In addition to the underlying argument structures described, cinematic cues (such as fade-in, fade-out, cross-cut) can be given to make it easier for a viewer to follow the underlying argument. Future work requires further investigation into creating longer term argument structures, with arguments being built up over time and overall conclusions being made.

IV. THE ROLE OF THE STRUCTURED PROGRESSION

Using domain and discourse structure to influence the construction of a presentation is a first step in the creation

of a final-form presentation. For mono-medium cases, such as continuous text or video sequences, all that remains is to select an ordering for the fragments and present them to the user. For presentations that use combinations of media, or even those that impose navigable hierarchical structure on pieces of text (such as Noadster [13]), the step from complex semantic representation to interactive hypermedia presentation is not simple. The semantics of both domain and discourse need to be translated to hierarchical structures that can be expressed through layout, navigational links or temporal information.

To go directly from the complex structure containing all information to a final form presentation would require processing both semantic and presentation information simultaneously. A number of systems, such as DISC [7], demonstrate that an intermediate format is required. We call this intermediate format a structured progression [12]. This encapsulates the flow of a presentation derived from the domain and discourse semantics to a structure which contains only grouping, ordering, importance and cross-hierarchical relationships. The structured progression can then be handed on for processing to a final-form presentation that, e.g., suits the requirements of the end user's platform.

The question that still remains is that while it is apparent that such an intermediate structure is helpful for delimiting processing concerns, it is not yet clear how much discourse information has to be explicitly included in the structured progression. For example, when generating multimedia presentations, one needs to be able to convey semantic and discourse relations directly by means of spatial and temporal layout.

V. Conclusion

This paper provides only a brief overview of the work that has been carried out in this area. While our goal of generating meaningful presentations tailored to the end user remains long term, the building blocks needed are gradually beginning to emerge. Headway has been made with understanding what role explicitly encoded discourse can play in the construction of presentations from query results; experience has been gained in understanding the transformation processes that need to take place and we can demonstrate examples of some of these mechanisms working in practice.

Acknowledgments

References

[1] John Bateman, Jörg Kleinz, Thomas Kamps, and Klaus Reichenberger. Towards Constructive Text, Diagram, and Layout Generation for Information Presentation. *Computational Linguistics*, 27(3):409–449, September 2001. Available from World Wide Web: http://acl.ldc.upenn.edu/J/J01/.

[2] Stefano Bocconi, Frank Nack, and Lynda Hardman. Supporting the Generation of Argument Structure within Video Sequences. In *Proceedings of the sixteenth ACM Conference on Hypertext and Hypermedia 2005*, pages 75–84, September 2005. Available from World Wide Web: http://www.cwi.nl/˜media/publications/ht2005.pdf.

[3] Stefano Bocconi, Frank Nack, and Lynda Hardman. Using Rhetorical Annotations for Generating Video Documentaries. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME) 2005*, July 2005. Available from World Wide Web: http://www.cwi.nl/˜media/publications/icme2005.pdf.

[4] 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*, 18(6-7):477–496, December 1997. Available from World Wide Web: http://www.dfki.uni-sb.de/imedia/lidos/papers/csi97/.

[5] Marc Davis. *Media Streams: Representing Video for Retrieval and Repurposing*. PhD thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1995. Available from World Wide Web: http://garage.sims.berkeley.edu/pdfs/1995_Marc_Davis_Dissertation.pdf.

[6] Kateryna Falkovych and Stefano Bocconi. Creating a Semantic-based Discourse Model for Hypermedia Presentations: (Un)discovered Problems. In *Workshop on Narrative, Musical, Cinematic and Gaming Hyperstructure*, Salzburg, Austria, September 2005. Available from World Wide Web: http://www.cwi.nl/˜media/publications/HTW05Falkovych.pdf.

[7] Joost Geurts, Stefano Bocconi, Jacco van Ossenbruggen, and Lynda Hardman. Towards Ontology-driven Discourse: From Semantic Graphs to Multimedia Presentations. In *Second International Semantic Web Conference (ISWC2003)*, pages 597–612, Sanibel Island, Florida, USA, October 20-23, 2003. Available from World Wide Web: http://www.springerlink.com/openurl.asp?genre=article&issn=0302-9743&vo%lume=2870&spage=597.

[8] Joost Geurts, Jacco van Ossenbruggen, and Lynda Hardman. Application-Specific Constraints for Multimedia Presentation Generation. In *Proceedings of the International Conference on Multimedia Modeling 2001 (MMM01)*, pages 247–266, CWI, Amsterdam, The Netherlands, November 5-7, 2001. Available from World Wide Web: http://www.cwi.nl/˜media/publications/mmm01.pdf.

[9] E. Hyvönen, M. Junnila, S. Kettula, E. Mäkelä, S. Saarela, M. Salminen, A. Syreeni, A. Valo, and K. Viljanen. MuseumFinland — Finnish museums on the semantic web. *Journal of Web Semantics*, 3(2-3):224–241, October 2005. Available from World Wide Web: http://dx.doi.org/10.1016/j.websem.2005.05.008.

[10] S. Kim, H. Alani, W. Hall, P.H. Lewis, D.E. Millard, N.R. Shadbolt, and M.J. Weal. Artequakt: Generating Tailored Biographies with Automatically Annotated Fragments from the Web. Presented at the Semantic Authoring, Annotation and Knowledge Markup (SAAKM) 2002 Workshop at the 15th European Conference on Artificial Intelligence (ECAI 2002), Lyon, France.

[11] William C. Mann, Christian M. I. M. Matthiesen, and Sandara A. Thompson. Rhetorical Structure Theory and Text Analysis. Technical Report ISI/RR-89-242, Information Sciences Institute, University of Southern California, November 1989.

[12] Lloyd Rutledge, Martin Alberink, Rogier Brussee, Stanislav Pokraev, William van Dieten, and Mettina Veenstra. Finding the Story — Broader Applicability of Semantics and Discourse for Hypermedia Generation. In *Proceedings of the 14th ACM Conference on Hypertext and Hypermedia*, pages 67–76, Nottingham, UK, August 23-27, 2003. ACM, ACM Press. Available from World Wide Web: http://www.cwi.nl/˜media/publications/ht03.pdf.

[13] Lloyd Rutledge, Jacco van Ossenbruggen, and Lynda Hardman. Making RDF Presentable – Integrated Global and Local Semantic Web Browsing. In *The Fourteenth International World Wide Web Conference*, pages 199–206, Chiba, Japan, May 2005. IW3C2, ACM Press. Available from World Wide Web: http://www2005.org/cdrom/docs/p199.pdf.

[14] Jacco van Ossenbruggen, Joost Geurts, Frank Cornelissen, Lloyd Rutledge, and Lynda Hardman. Towards Second and Third Generation Web-Based Multimedia. In *The Tenth International World Wide Web Conference*, pages 479–488, Hong Kong, May 1-5, 2001. IW3C2, ACM Press. Available from World Wide Web: http://www10.org/cdrom/papers/423/index.html.