

Towards Multimedia Thesaurus Support for Media-based Navigation

Paul H. Lewis, Hugh C. Davis, Mark R. Dobie and Wendy Hall
The Multimedia Research Group,
Department of Electronics and Computer Science
University of Southampton, England, SO171BJ,
tel: +44 1703 593715, E-mail phl@ecs.soton.ac.uk

November 5, 1996

Abstract

This paper describes an open hypermedia system which allows content based navigation and retrieval from both text and non-text media. The system provides a variety of link types including the generic link which, once authored from a particular selection, may be followed from any instance matching that selection. The paper also describes the concept of a multimedia thesaurus and illustrates how it may be used to enhance the navigational capabilities of the system, allowing links to be authored on one representation of an object and followed from a different representation.

1 Introduction

The use of hypermedia links to navigate through multimedia information collections has increased dramatically with the increasing popularity of the World Wide Web. Typically the links in the web are from one point to another and the source anchors of links tend to be text based.

During the last six years the Multimedia Research Group in Southampton University has developed Microcosm, an open architecture for hypermedia systems handling large multimedia information collections[4, 7, 6, 11]. The architecture is three layered: an applications layer in which general and special purpose media viewers, together with third party applications packages, provide the user interface to the multimedia collection, a link service layer in which any number of communicating processes provide the hypermedia functionality and a storage layer where the documents, link information and other databases are maintained.

One of the important features of the architecture is that link information is held in link databases, separately from the documents being linked, so that documents remain in their native format and may be prepared and viewed via third party applications packages. This also means that links may be established with media on read only devices such as CD-ROM.

A second important feature is the generic link. Generic links differ from the more common point to point links in that, once a link has been authored between a source selection and a destination point, it may be followed from every occurrence of the source selection in any document which has access to the linkbase. Thus a generic link from the word *Amsterdam* to a video of the city may be followed from any occurrence of the word *Amsterdam*. The generic link works at the link authoring stage by storing the content of the source selection (ie the word *Amsterdam*) as the source anchor of the link and at the link following stage, by matching the user selection with source anchors in the linkbase.

We have recently extended the Microcosm architecture to create MAVIS (Microcosm Architecture for Video, Image and Sound) which provides a framework for authoring and following generic links from non-text media[11]. Thus, a user may select part of an image by dragging a rectangle over the required region with the mouse, and may specify this selection as the source anchor for a generic link. A user making another sub-image selection in another image may follow the link if the selection is similar to the source anchor on which the link was authored. The generic links in non-text media work by extracting a variety of representations or signatures from the selection and using pattern matching between signatures at the link following stage.

It should be clear that the generic link works by matching representations of the selection *content*. Hence we can refer to navigation with generic links as content based navigation.

In the next section we briefly refer to related work and in section 3 we give more detail of generic link following with MAVIS. In section 4 we discuss how a digital thesaurus could enhance generic link following from text and then extend the approach by describing, with a prototype example, how a multimedia thesaurus could provide enhanced generic link following from text and non-text media. The paper concludes with some final comments and a note on future work.

2 Related Work

There has recently been a rapid development in the use of content for retrieving images from image databases with many papers describing a wide range of techniques [1, 10, 13]. Sometimes these are tuned to specific application domains but some systems provide basic content based image retrieval tools of wider applicability. A good example is the QBIC, Query By Image Content, system [5,

14], developed by IBM, which uses combinations of colour, shape and texture for image retrieval and the Manchester Content Addressable Image Database [12].

Typically the aim is to retrieve an image from the database if it contains *features* specified by the user, either by drawing a sketch or identifying similar features in a displayed image. However, little work has been published on navigating from images to other parts of a multimedia information collection when using image content as the key with which to determine the available routes to be navigated. Part of the reason for this is that most hypermedia systems rely on fixed or hard wired links between parts of the multimedia information collection even when navigating from text. A notable exception is the system under development at NEC [8, 9], which is a hypermedia system in which non-textual information may be retrieved using media-based queries.

3 Generic Links and Non-Text Media

In the Microcosm architecture, a link in the linkbase, which has been authored from text, consists of information about the source anchor and the destination anchor of the link. For a specific or point to point link the source anchor includes the selection content, the file in which it occurs and the location in the file at which it occurs. To be able to follow a specific link from a particular selection in a particular document, all three items in the source anchor must match those for the selection. When authoring a generic link from text only the selection itself is recorded in the linkbase for the source anchor. To follow a generic link from a particular selection, only the selection needs to match the selection in the link. Thus, a generic link may be followed from any instance of the source selection in any document. This gives a substantial reduction in link authoring effort, but until the introduction of MAVIS the generic link was only available from text.

In order to provide generic links from non-text media, the MAVIS architecture was developed. The architecture recognised that it was not simply a matter of recording the media based selection in the link source anchor and then matching selections when link following. Text selections are matched exactly, but with non-text selections, similarity estimation is required. The MAVIS architecture also recognises that in order to provide content based retrieval and navigation from non-text media it is necessary to be able to extract and match a wide range of representations from a range of different media types. This functionality is provided by modules controlled by a media table subsystem. Each module is responsible for handling all the processing associated with one particular representation for one particular medium and must contain algorithms to extract the representation, to index the representation for rapid retrieval and to estimate a similarity value, given two representations to be matched.

The main modules implemented so far include a colour histogram module based on the Tek HVC colour representation, a texture module which uses

statistical geometric texture statistics [3], a shape module which uses rotation, scale and translation invariant moments, a further shape module which uses chord length distribution and a demonstrator sound module which uses the Fourier transform of the selection from a digital sound file as the representation. The media table subsystem maintains an association between the different media and the modules which can process selections from them. The user can control which modules are active via the subsystem and the weighting associated with different representations if retrieval or link following is to be based on more than one representation. The architecture is designed so that additional modules may be added to support other representations as required. More details of the MAVIS architecture and its use for content based retrieval and navigation have already been published [11].

4 Thesaurus Enhanced Generic Links

In the previous sections we saw that generic link following succeeds by matching representations of user selections with the source anchors of links. In the case of text, an exact match or word stem match is appropriate whereas for non-text it is necessary to use similarity matches. However, even in the case of text, the basic generic link has a severe limitation. For the generic link to be successful the user must choose the same word or words as those used when the link was authored. Synonyms and semantically linked concepts are not catered for in the current implementation. If a generic link is authored on the word car, it can not be followed from the word automobile. For non-text media there are equivalent problems. For example, for images, if a link is authored on the shape of an object from one particular view it will not be possible to follow it from an image in which the view is substantially different.

One way to approach a solution to this problem for text is to introduce a digital thesaurus. This should contain all the vocabulary (words or terms) appropriate to the application domain, arranged as a network with at least the following relations: broader term, narrower term, equivalent term and related term. The thesaurus could be invoked either at the link authoring or the link following stage or both. In the first case, if the thesaurus is invoked when a generic link is authored, links could be automatically generated and stored from all terms equivalent to the selected source anchor term. This would increase the linkbase size but have the advantage of increasing the speed of link following. Alternatively, the authoring stage could remain unchanged, ie a single link is generated from the author's selection, but at the link following stage the user's selection is expanded via the thesaurus to a set of equivalent terms and each is compared with the linkbase source anchors to find any available links. This will be less expensive in linkbase storage but will slow down the linkbase search.

More elegantly the thesaurus could contain an indicator of a preferred term for each set of equivalent terms. At the authoring stage, the source anchor

term is switched to its preferred term via the thesaurus and at link following, the user's chosen term is also switched to its preferred term, via the thesaurus, before the linkbase search is commenced.

All transactions via the thesaurus described so far can be invisible to both author and user, providing an appropriate thesaurus is available containing all terms likely to be encountered. Indeed, within the Microcosm architecture, the thesaurus could be introduced as a filter between the viewers and the linkbase, without substantial disturbance to the current architecture. However, an additional level of functionality may be achieved if, at the link following stage, the user is given access to the thesaurus to broaden or narrow down the generic links from a particular selection. For example, a user reading a document on analytical chemistry may wish to try and follow links on *mass spectrometry*. If an option to view the thesaurus at this point is available the user may choose narrower terms such as *high resolution mass spectrometry* or broader terms such as *spectrometry* to act as the selection for link following. Indeed the user could also choose to switch on a particular level of automatic broader or narrower term fan out for selections when generic link following. In this case, whenever a term is chosen, all broader terms or all narrower terms or both could also be compared with the linkbase source anchors to find appropriate links to follow.

4.1 A Multimedia Thesaurus

The thesaurus described so far is similar to thesauruses which have been used in information retrieval systems for many years. However, the integration of a digital thesaurus and generic link following offers a significant enhancement in the potential of generic link following from text. More importantly, we propose a multimedia extension to the basic digital thesaurus concept which will provide enhanced generic link following from non-text media as well as from text.

The multimedia thesaurus(MMT) consists of a network of representations. Many of these will be text, corresponding to a traditional digital thesaurus, but some will be representations of terms extracted from other media. We have seen that in order to provide generic links from non-text, indexes of features such as shape or texture description vectors or sound representations are accumulated. In the multimedia thesaurus these will be associated with equivalent text representations. A new set of relations in the network will, at a minimum, consist of broader representation, narrower representation, equivalent representation and related representation as generalisations of the text term relations mentioned earlier. The MMT will also include preferred representation indicators which will typically be attached to a text representation as this will offer maximum storage efficiency and matching efficiency when link following. An *is_a* relation will also be used to indicate representations of specific instances of objects. For image based representations, thumbnails of the image selection will be associated with the node to facilitate more informed navigation around the MMT.

In a natural extension to the MMT, other relations could be introduced such as *is-part-of*, *is-older-than* etc making the MMT a semantic network of media based representations and allowing more powerful search facilities to users at the link following stage.

It is possible to view the MMT structurally in two parts. The first is a network of concepts and their relationships which is the MMT but with only the preferred terms. This contains all the important conceptual/semantic relationships in the MMT. The second part consists of all the equivalent representations which are linked to their preferred term nodes in the concept network. We have already prototyped a concept network as an aid to enhanced link following[2].

By associating media based representations with equivalent term nodes in the MMT at the generic link authoring stage (if they are not already in the MMT) the following new levels of functionality become possible.

- When a generic link is authored from a representation of an object which is in the MMT, it will be possible to follow the link from any of the media based representations for the object which are also present. Thus, a generic link could be authored from an image of a Greek amphora using a shape representation and followed from a text document via selection of the text term *Greek amphora*.
- In particular application areas, multiple views of objects may be available. For example in a multimedia museum application, multiple views of an exhibit may be provided and representations of these could be associated with each other via the MMT. A generic link from one view could then be followed from any of the other views of the same artefact.
- As in the case of text described earlier, it will be possible to use the MMT at the link following stage to broaden or narrow the specificity of the attempt at generic link following by navigating the broader and narrower representation relations in the MMT.

Although the use of an MMT offers significant enhancements for generic link following, the construction of the MMT is a non-trivial activity and will involve substantial time overheads. However, for some applications the text thesaurus may already exist or the the time investment may be deemed worthwhile.

4.2 The User Interface

When authoring a generic link in an MMT supported, open hypermedia environment, the user may wish to attach the source selection representation to the MMT at the authoring stage if it is not already present. In figure 1 a generic link is being authored from the shape of a particular archaeological artifact (a multiple barbed point) to some text about bone points (the end link). The shape for the start link has been selected by dragging a rectangle around it and

specifying the shape representation module to be used in MAVIS via the media table (not shown).

Before completing the link by selecting *Ok* in the complete link window (bottom right) the author could link the representation into the MMT by selecting an MMT tools menu and choosing an *Add* option. A prototype example is shown in the figure.

The list of top level terms from the MMT is displayed and the appropriate preferred term is selected by scrolling and clicking. Selecting *Add Representation* would then complete the addition of the representation to the MMT. If no appropriate entry points are present, MMT building tools would be used to add appropriate new terms and their relations.

Once the representation is linked in to the thesaurus, the generic link may be followed not only from similar shape selections but also from the text representation *multiple barb points*. No user interaction with the thesaurus would be required here if automatic MMT based generic link following is initiated. A user, following a generic link in this way, may also decide to use the MMT explicitly to explore links from representations nearby. The user will be able to navigate to other representations within the MMT and then choose to follow links from selected representations.

5 Conclusions and Future Work

In this paper we have described the Microcosm open hypermedia system and MAVIS, an extension to Microcosm to allow content based navigation from non-text media as well as text. We have discussed the idea of a multimedia thesaurus(MMT) and shown how an MMT could be used to provide more powerful generic link authoring and following.

We are currently in the process of implementing the thesaurus, integrating it into the Microcosm/MAVIS architecture and developing intelligent agents to improve/automate the grouping of representations in the MMT and the matching of selections during link following. There is substantial scope for introducing further enhancements to content based navigation and retrieval facilities using the capabilities of the multimedia thesaurus.

Acknowledgement

This research is being undertaken with the support of the EPSRC, grant number GR/L03446.

References

- [1] Special issue on content based retrieval. *IEEE Computer* 28, 9 (1995).

- [2] BEITNER, N. B., GOBLE, C. A., AND HALL, W. Putting the media into hypermedia. In *SPIE Proceedings, Multimedia Computing and Networking* (1995), pp. 12–23.
- [3] CHEN, Y., NIXON, M., AND THOMAS, D. Texture classification using statistical geometrical features. *Pattern Recognition* 28, 4 (1995), 537–552.
- [4] DAVIS, H., HALL, W., HEATH, I., HILL, G., AND WILKINS, R. Towards an integrated information environment with open hypermedia systems. In *Fourth ACM Conference on Hypertext (ECHT '92)* (Milan, Italy, December 1992), D. Lucarella, J. Nanard, N. M., and P. Paolini, Eds., pp. 181–190.
- [5] FALOUTSOS, C., BARBER, R., FLICKNER, M., HAFNER, J., NIBLACK, W., PETKOVIC, D., AND EQUITZ, W. Efficient and effective querying by image content. *Journal of Intelligent Information Systems* 3 (1994), 231–262.
- [6] HALL, W., DAVIS, H. C., AND HUTCHINGS, G. A. *Rethinking Hypermedia: the Microcosm Approach*. Kluwer Academic Press, Amsterdam, 1996.
- [7] HILL, G. J., WILKINS, R. J., AND HALL, W. Open and reconfigurable hypermedia systems: A filter-based model. *Hypermedia* 5, 2 (1993), 103–118.
- [8] HIRATA, K., HARA, Y., SHIBATA, N., AND HIRABAYASHI, F. Media-based navigation for hypermedia systems. In *Hypertext '93* (Seattle, Washington, USA, November 14-18 1993), ACM Press, pp. 159–173.
- [9] HIRATA, K., HARA, Y., TAKANO, H., AND KAWASAKI, S. Content-oriented integration in hypermedia systems. In *Proceedings of the Seventh ACM Conference on Hypertext* (Washington DC, March 1996), B. Ladd, Ed., pp. 11–21.
- [10] JAIN, R., MURTHY, S., CHEN, P., AND CHATTERJEE, S. Similarity measures for image databases. In *SPIE Proceedings, Storage and Retrieval for Image and Video Databases III* (1995), pp. 58–65.
- [11] LEWIS, P. H., DAVIS, H. C., GRIFFITHS, S. R., HALL, W., AND WILKINS, R. J. Media-based navigation with generic links. In *Proceedings of the Seventh ACM Conference on Hypertext* (March 1996), ACM, pp. 215–223.
- [12] OAKLEY, J., SHANN, R., DAVIS, D., AND HUGUEVILLE, L. A database management system for vision applications. In *5th British Machine Vision Conference* (York, U.K., 13-16 September 1994), BMVA Press, pp. 629–639.

- [13] SMOLIAR, S., AND ZHANG, H. Content-based video indexing and retrieval. *IEEE Multimedia* 1, 2 (1994), 62–72.
- [14] TREAT, H., ORT, E., VO, M., JANG, J.-S., HALL, L., TUNG, F., AND PETKOVIC, D. Searching images using Ultimedia manager. In *SPIE Proceedings, Storage and Retrieval for Image and Video Databases III* (1995), pp. 204–213.

Figure 1: A User Interface for the MMT with MAVIS