

# The Significance of Linking

Paul H. Lewis, Wendy Hall, Leslie A. Carr, and David De Roure

University of Southampton Web: <http://www.soton.ac.uk/>

Multimedia Research Group Web: <http://mmrg.ecs.soton.ac.uk/>  
Highfield Lane, Southampton, SO17 1BJ, UK

Email: [phl@ecs.soton.ac.uk](mailto:phl@ecs.soton.ac.uk), [wh@ecs.soton.ac.uk](mailto:wh@ecs.soton.ac.uk), [lac@ecs.soton.ac.uk](mailto:lac@ecs.soton.ac.uk), and [dder@ecs.soton.ac.uk](mailto:dder@ecs.soton.ac.uk)

Web: <http://www.ecs.soton.ac.uk/~phl/>, <http://www.ecs.soton.ac.uk/~wh/>, <http://www.ecs.soton.ac.uk/~lac/>, and <http://www.ecs.soton.ac.uk/~dder/>

---

**Abstract:** The link has been central to the idea of hypertext since its inception and continues to enjoy widespread popularity. This paper briefly explores the history of links and draws a distinction between navigation and retrieval in information handling. The value of information content in certain navigation and retrieval tasks is examined and the challenges of extending content-based retrieval and navigation to non-text media are identified. Finally, the goal of more versatile content- and concept-based navigation is discussed.

Categories and Subject Descriptors: I.7.2 [Text Processing]: Document Preparation - *hypertext/hypermedia*; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems - *Hypertext navigation*

General terms: content-based navigation, semantics

---

## A Brief History of Links

Ted Nelson defined hypertext as non-sequential reading and writing [Nelson 1981] and more recently Isakowitz et al. [Isakowitz 1995] defined it as the science of relationships and relationship management. Whatever the definition, links have always been at the heart of hypertext and hypermedia, providing the relationships in Isakowitz's definition and allowing the non-sequential modes of access in the definition from Nelson.

Since the early days of Xanadu [Nelson 1981], in which links were designed to be created by the author and permanently fixed by the documents to which they referred, the idea of the link has evolved in a variety of ways. Link architectures, models (most notably the Dexter model [Halasz 1994]) and link typing strategies [Nanard 1991] have been proposed, adopted and often eventually superseded.

Simple, fixed or point to point links have been an enduring feature of hypertext systems, often providing the main linking strategy as, for example, in early systems like Guide [Brown 1987] and HyperCard [Goodman 1987]. Through the ubiquity of the Web they are now the most widely used hypermedia link type. They are embodied in the definition of HTML and are essentially static, point to point connections embedded as pointers in the source document. They have achieved their huge popularity through their universality over the Internet.

---

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Publications Dept, ACM Inc., fax +1 (212) 869-0481, or [permissions@acm.org](mailto:permissions@acm.org)

In contrast to the fixed embedded link, some systems have been designed in which the links are held in link databases, separately from the documents to which they refer. The Intermedia [Haan 1992] system at Brown University was probably the first to introduce this development in hypermedia design and it has become an important feature of many more recent systems. In open hypermedia systems [Davis 1998] links may be manually or system created, are typically stored in a database independently of the documents and can be static or dynamic. The new XLink proposal [Maler 1998] will allow these kinds of links for use with XML [DeRose 2000] data on the Web.

In some hypertext design models [Garzotto 1991] links do not necessarily exist explicitly and are implicit in the structure. In some cases they may be dynamically created by a process which defines how elements of the hypertext structure are related when the process is invoked [Nürnberg 1997a], [Marshall 1997], [Wiil 2000], [Shipman 2000].

## Navigation or Retrieval

Whether implicit or explicit, static or dynamic, links offer a way of defining relationships in information and there are many ways of achieving this goal.

It could be argued that, with the emergence of the Web as the dominant hypertext structure, "hyperlinks" as such are unnecessary. "You can always find any document you need by appropriate use of search engines like Alta Vista." This, however, runs against the serendipity of hypertext and the importance of the relationships or associations that are not born of similarity but through the external knowledge of the link author. Again it might be argued that many of these associations may be achieved by keywording, particularly for associating material of different media types.

In this context, it is instructive to look at an historical difference between the Information Retrieval and the Hypermedia approaches, although the activities of these two communities are now converging in the search for powerful multimedia information management tools. The Information Retrieval community is particularly concerned with *retrieval*. Retrieval typically answers the request "find me documents containing something like this query" In terms of links or associations, retrieval usually relies on being able to make an association between a query, for example a keyword or phrase, and an information item (document) containing something *similar* to the query. Typically, the association is achieved either through pre-indexing or on-the-fly analysis.

By contrast, the Hypertext community has been particularly concerned with *navigation*. Navigation involves steering across links or associations which do not necessarily require *similarity* between the source (the query) and the destination. The link may represent some meaningful higher level association that is typically, but not necessarily, identified through the mind of the link author.

A synergistic marriage between retrieval and navigation can result in the dynamic provision of links created on the fly using theming technology based on statistical and linguistic techniques to provide links derived from the current document context. This technology has been successfully employed commercially [Multicosm 1999] as an evolution of a dynamic link service [Carr 1995], [Carr 2000].

The importance of associations, perceived by the link author, was a foundation for early work on hypermedia in education. We think and learn through associations, so that if we define, as hypertext links, those associations known to the teacher, then students who subsequently "read" the hypertext will gain a deeper understanding of the material, and learn the associations through navigating the links. This has not yet proven to be as productive in practice but, in a more general form, was the basic premise for Bush's paper "As We May Think" [Bush 1945].

## The Value of Content

Content-based retrieval (CBR) means retrieval using the actual content of the information rather than meta-data, associated keywords or manually constructed indexes. CBR for text is a well established technique, epitomised by

the free text searches and content indexes generated, for example, by Web search engines.

For non-text multimedia information, interest in the research and development of CBR techniques is currently very strong [Computer 1995]. The most active and most advanced area is content-based image retrieval where techniques from image processing and computer vision are beginning to provide useful tools [Eakins 1999] especially in highly constrained application domains. In general, content-based retrieval from raw non-text media is a challenging goal requiring solutions to media processing and analysis problems. The challenge comes from the fact that the media content is not explicit in most raw media formats. For example in pixelated images the only explicit information is the brightness or colour at each pixel position.

The problem is significantly reduced if the media formats are structured in some way to make the content explicit. Examples might be vector maps, CAD drawings and in the audio field, music in MIDI format rather than raw digitised audio. Wide use of formats like the proposed MPEG-7 [MPEG 1999] standard will also enhance content-based techniques but problems of initial production of the "documents" are, in most cases, non trivial.

Content-based navigation (CBN) is less common than CBR but is embodied in the concept of the generic link [Hall 1996]. The source anchors for such links are specified in terms of source content rather than source location. The storage of source content as part of the link structure was facilitated by the adoption of external link databases. Once authored from some source selection, a generic link may be followed from any matching instance of the source content: hence content-based navigation.

The use of CBN gives substantial savings in authoring and link maintenance effort but initially CBN was only possible from text. To provide CBN from non-text media involves a solution to the same problems as in CBR. Approaches to the problem have been made through the MAVIS project at Southampton [Lewis 1996] and the systems at NEC [Hirata 1996].

## Content or Concept?

One of the problems with using content, whether it be for retrieval or for navigation, is that it is not in the content that we are really interested. Words and pictures, videos and speech are representations of objects, ideas and concepts in the real world. In Saussure's terminology [de Saussure 1966], [Smoliar 1996], we are dealing with the difference between the signifier (media representations) and the signified (real objects, concepts, ideas). It is the signified in which we are really interested.

We as humans can make the link from signifier to signified almost automatically, typically drawing on a huge body of prior knowledge. But for our software systems the link (or its absence) is at the root of many of the problems with content-based retrieval and navigation [Grønbaek 1996], [Davis 1999], [Davis 2000]. The same concept can have many different text representations even in the same language (synonyms). In images, instances of the same general concept may be very different (eg arm chairs and office chairs). And even the same object may look vastly different in images of different views.

In an attempt to overcome some of the problems with text, the use of digital thesaurus tools or facilities for statistically based associations have been incorporated into information retrieval systems [van Rijsbergen 1979]. More recently researchers have attempted to introduce layers of associations, above the media based links, which try to capture semantic associations relevant to an application and provide navigation and retrieval based on concepts in addition to content [Tudhope 2000], [Cunliffe 1997], [Beynon-Davies 1994], [Bullock 1998], [Nanard 1991]. The MAVIS-2 project [Dobie 1999] and the COIR project [Hirata 1996] both attempt to associate media based representations directly with a semantic layer in an attempt to provide integrated approaches to content and concept based retrieval and navigation.

## The Challenge

Building semantic layers and associating media content with the concepts they represent is currently a labour intensive task. The challenge now is to build systems which extract or learn the semantics from the knowledge implicit in the media and make the associations between the media representations and the semantics without a heavy manual input. Retrieving and navigating more directly with concepts, rather than their manifold representations, will then be a reality.

## Bibliography

**[Beynon-Davies 1994]** Paul Beynon-Davies, Douglas Tudhope, Carl Taylor, and Chris B. Jones. "A Semantic Database approach to Knowledge-Based Hypermedia Systems" in *Information and Software Technology* 36, 6, 323-329, 1994.

**[Brown 1987]** Peter J. Brown. "Turning Ideas into Products: The Guide System" in *Proceedings of ACM Hypertext '87*, Chapel Hill, NC, 33-40, 1987.

**[Bullock 1998]** Joseph Bullock and Carole Goble. "TourisT: The Application of a Description Logic based Semantic Hypermedia System for Tourism" in *Proceedings of ACM Hypertext '98*, Pittsburgh PA, 132-141, June 1998.

**[Bush 1945]** Vannevar Bush. "As We May Think" in *The Atlantic Monthly*, 176(1),101-108, [Online: <http://www.isg.sfu.ca/~duchier/misc/vbush/>], July 1945.

**[Carr 1995]** Leslie A. Carr, David C. De Roure, Wendy Hall, and Gary J. Hill. "The Distributed Link Service: A Tool for Publishers, Authors, and Readers" in *Proceedings of the Fourth International World Wide Web Conference*, Boston, MA, 647-656, [Online: [http://www.staff.ecs.soton.ac.uk/lac/dls/link\\_service.html](http://www.staff.ecs.soton.ac.uk/lac/dls/link_service.html)], December 1995.

**[Carr 2000]** Leslie A. Carr, Wendy Hall, and David C. De Roure. "The Evolution of Hypermedia Link Services" in *ACM Computing Surveys*, Symposium on Hypertext and Hypermedia, 2000.

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

**[Cunliffe 1997]** Daniel Cunliffe, Carl Taylor, and Douglas Tudhope. "Query-based Navigation in Semantically Indexed Hypermedia" in *Proceedings of ACM Hypertext 97*, Southampton, UK, 87-95, April 1997.

**[Davis 1998]** Hugh C. Davis. "Referential Integrity of Links in Open Hypermedia Systems" in *Proceedings of ACM Hypertext '98*, Pittsburgh, PA, 207-216, June 1998.

**[Davis 1999]** Hugh C. Davis, David E. Millard, Siegfried Reich, Niels Olof Bouvin, Kaj Grnbæk, Peter J. Nürnberg, Lennert Sloth, Uffe K. Wiil, and Kenneth M. Anderson. "Interoperability between Hypermedia Systems: The Standardisation Work of the OHSWG" in *Proceedings of ACM Hypertext '99*, Darmstadt, Germany, 201-202, February 1999.

**[Davis 2000]** Hugh C. Davis. "Hypertext Link Integrity" in *ACM Computing Surveys*, Symposium on Hypertext and Hypermedia, 2000.

**[de Saussure 1966]** Ferdinand de Saussure. *Course in General Linguistics*. McGraw Hill, New York, 1966.

**[DeRose 2000]** Steven J. DeRose. "XML Linking" in *ACM Computing Surveys*, Symposium on Hypertext and Hypermedia, 2000.

- [Dobie 1999]** Mark Ralph Dobie, Robert Tansley, Dan Joyce, Mark Weal, Paul H. Lewis, and Wendy Hall. "A Flexible Architecture for Content and Concept-based Multimedia Information Exploration" in Proceedings of the Second UK Conference on Image Retrieval, BCS Electronic Workshops in Computing, 1999.
- [Eakins 1999]** John Eakins and Margaret Graham. "Content-based Image Retrieval" in A Report to the JISC Technology Applications Programme, JTAP report No. 39, 1999.
- [Garzotto 1991]** Franco Garzotto, Paulo Paolini. "HDM - a Model for the Design of Hypertext Applications" in Proceedings of the Third ACM Conference on Hypertext, 313-328, 1991.
- [Goodman 1987]** Danny Goodman. The Complete Hypercard Handbook. Bantam Books, New York, 1987.
- [Grønbaek 1996]** Kaj Grønbaek and Randall H. Trigg. "Toward a Dexter-Based Model for Open Hypermedia: Unifying Embedded References and Link Objects" in Proceedings of ACM Hypertext 96, Washington DC, 149-160, [Online: <http://acm.org/pubs/citations/proceedings/hypertext/234828/p149-gronbaek/>] March 1996.
- [Haan 1992]** Bernard J. Haan, Paul Kahn, Victor A. Riley, James H. Coombs, and Norman K. Meyrowitz. "IRIS Hypermedia Services" in Communications of the ACM (CACM), 35(1), 36-51, January 1992.
- [Halasz 1994]** Frank G. Halasz and Mayer D. Schwartz. "The Dexter Hypertext Reference Model" in Communications of the ACM (CACM), 37(2), 30-39, February 1994.
- [Hall 1996]** Wendy Hall, Hugh C. Davis, and Gerard Hutchings. Rethinking Hypermedia, The Microcosm Approach. Kluwer Academic, Dordrecht, The Netherlands, 1996.
- [Hirata 1996]** Kyoji Hirata, Yoshinori Hara, Hajime Takano, and Shigehito Kawasaki. "Content-Oriented Integration in Hypermedia Systems" in Proceedings of ACM Hypertext '96, Washington, DC, 11-21, March 1996.
- [Isakowitz 1995]** Tom-s Isakowitz, Edward Stohr, and P. Balasubramanian. "RMM: A Methodology for Structuring Hypermedia Design" in Communications of the ACM (CACM), 38(8), 34-44, August 1995.
- [Lewis 1996]** Paul H. Lewis, Hugh C. Davis, Steven R. Griffiths, Wendy Hall, and Rob J. Wilkins. "Media-based Navigation with Generic Links" in Proceedings of the Seventh ACM Conference on Hypertext, 215-223, 1996.
- [Maler 1998]** Eve Maler and Steven DeRose (editors). XML Linking Language (XLink), World Wide Web Consortium Working Draft3 [Online: <http://www.w3.org/TR/WD-xlink>], March 1998.
- [Marshall 1997]** Catherine C. Marshall, and Frank M. Shipman. "Spatial Hypertext and the Practice of Triage" in Proceedings of ACM Hypertext '97, Southampton, UK, 124-133, April 1997.
- [MPEG 1999]** MPEG. Home Page, 1999. <http://drogo.cselt.stet.it/mpeg/>
- [Multicosm 1999]** Multicosm Ltd. In Chilworth Science Park, Southampton, 1999. <http://www.multicosm.com/>
- [Nanard 1991]** Jocelyne Nanard and Mark Nanard. "Using structured types to incorporate knowledge in hypertext" in Proceedings of ACM Hypertext '91, San Antonio, TX, 329-344, December 1991.
- [Nelson 1981]** Theodor Helm Nelson. Literary Machines. Published by the author. 1981.
- [Nürnberg 1997a]** Peter J. Nürnberg, John Leggett, and Erich R. Schneider. "As We Should Have Thought" in Proceedings of ACM Hypertext '97, Southampton, UK, 96-101, April 1997.
- [Shipman 2000]** Frank M. Shipman and Catherine C. Marshall "Spatial Hypertext: An Alternative to Navigational and Semantic Lnks" in ACM Computing Surveys, Symposium on Hypertext and Hypermedia, 2000.

**[Smoliar 1996]** Stephen W. Smoliar, James D. Baker, Takehiro Nakayama, and Lynn Wilcox "Multimedia Search: An Authoring Perspective" in Proceedings of the ISIS First International Workshop on Image Databases and Multimedia Search Amsterdam, Netherlands, 1-8, 1996.

**[Tudhope 2000]** Douglas Tudhope and Daniel Cunliffe. "Semantically-Indexed Hypermedia: Linking Information Disciplines" in ACM Computing Surveys, Symposium on Hypertext and Hypermedia, 2000.

**[van Rijsbergen 1979]** C. J. "Keith" van Rijsbergen. Information Retrieval. Butterworth, 1979.

**[Wiil 2000]** Uffe K. Wiil, Peter J. Nürnberg, and John J. Leggett. "Hypermedia Research Directions: An Infrastructure Perspective" in ACM Computing Surveys, Symposium on Hypertext and Hypermedia, 2000.