Panel Session: Hypertext: "Growing Up?"

Panel Chair: M.E. Frisse, Washington University, St. Louis, USA;
Panelists: Maristella Agosti, Dipartimento di Elettronica e Informatica, Universita' di Padova, Italy; Marie-France Bruandet, Equipe Systems Intelligents de Recherche d'Informations, Grenoble, France; Udo Han, University of Freiburgh,West Germany; Stephen F. Weiss, University of North Carolina at Chapel Hill, USA

Abstract:

This panel will employ two different interpretations of the phrase "growing up" to address areas of common interest between hypertext and information retrieval researchers. First, the panelists will question whether or not hypertext is "growing up" as a scientific discipline; They will discuss characteristics that separate hypertext research from other related disciplines. Second, the panelists will discuss the problems encountered when a hypertext system "grows up" in size and complexity; They will discuss the very real problems expected when representing and integrating large knowledge bases, accommodating multiple users, and distributing single logical hypertexts across multiple physical sites.

The panelists will not lecture, but they will advance a number of themes including "the Myth of Modularity" (Frisse), "New Architectures Employing Hyperconcept Databases" (Agosti), "Hypertext in Software Engineering" (Bruandet), "Automatic Hypertext Generation" (Hahn), and "Large-Scale Hypertexts" (Weiss).

1. Overview

Hypertexts consist of a database of discrete components (text, graphics, video, audio), links connecting the text components, and tools for creating and navigating through the combination of components and links [9,25,28]. Hypertext's flexible structure facilitates creation of multiple unique "paths" through a single corpus of literature. These capabilities promote the use of hypertext in areas as diverse as software engineering, technical documentation, and computer-assisted instruction.

A number of issues seem common to most hypertext discussions. These include assumptions of hypertext component modularity, the need for index structures "external" to the hypertext, the importance of link and component typing, the utility of abstract models for hypertext, and the feasability of automatic generation of hypertexts from media currently distributed in some other form. Addressing the critical aspects of these issues should lead to a number of provocative questions regarding the practical role hypertext can play in the management of information.

2. Components and Links.

One of hypertext's principle features is the ability to support direct machine-supported links between discrete hypertext components [9]. Although hypertext components often play the role of "tiny documents" and serve as the fundamental unit of indexing and retrieval, they differ from traditional documents in their smaller size and explicit relationships defined by the links. Component size may diminish retrieval performance because few concepts are associated with each component. If the components are derived from a larger document, they may violate independence assumptions because a concept formerly expressed completely in one document is now distributed across many components.

Hypertext links serve three functions [10,12,16,29]. First, links denote associations between two highly related hypertext components. Second, links representing hierarchical structures allow for generalization and abstraction. Third, human computer interfaces can use links to facilitate visualization of relationships denoted by the hypertext.

3. *Indexes "external" to the hypertext*

Traditional hypertext link-based browsing is a search process whose performance is profoundly affected by the degree to which locally optimal choices result in a globally optimal hypertext path; performance deteriorates as a hypertext grows in size or becomes less formally structured. Halasz and others argue that in these circumstances browsing must be augmented with "global" search capabilities [11,12,13,21,23]. These capabilities require the addition of an index structure "external" to the components and links constituting the hypertext database. The added index structure allows quicker access to individual components and allows access directly to two or more documents unconnected by links if the components share some string, keyword or attribute. Agosti calls this structure, the "hyperconcept database" [1,2,3,4].

Hyperconcept databases allow one to infer new query terms on the basis of user queries and reader feedback [8,14]. In principle, one could employ a number of hyperconcept databases, each created by different indexing methods and each operating in parallel on the same hypertext database. Users or programs could then select at "run time" the hyperconcept database most appropriate to the circumstance.

4. *Node and Link Types*

The hypertext database and the hyperconcept database each can have different node and link types. Text nodes represent hypertext components in the hypertext database and topic nodes represent indexing concepts in the hyperconcept database. Cross-reference links are used in the structuring of hypertext database components; Semantic links make associations between topic nodes in the hyperconcept database; Connection links connect components in the hypertext database with topic nodes in the hyperconcept database.

5. *Abstract Models*

Hypertext systems have appeal to the software engineering field because they can combine database methods that provide direct data access with interface methods that permit document browsing [17, 24]. Software engineering applications, some believe, will require abstract models more complex than the simple directed graph model representing most hypertexts [7]; Models reported in the literature include those based on first-order logic [18], hypergraphs [30], Petri-nets [29], and object-oriented representations [22]. Abstract models would allow one to search on the basis of hypertext structure [10].

6. *Automated hypertext generation*

Many hypertexts are constructed from books or document collections. In these circumstances the text must be decomposed into individual components united by *structural links*. This process is dependent on modularity assumptions used to define component scope and indexing techniques used to represent the documents. When indexing, individual components are characterized and represented deterministically with structured data (e.g., author name, date of publication). Additional structured terms (auxiliary data) are produced by an index process in order to represent further the component's semantic content. Auxiliary data are neither complete nor unique descriptors of component content. The indexing process used to produce auxiliary data is not deterministic; The content can be interpreted and described in different ways depending on the observer's perspective and information needs. Accordingly auxiliary data pointers perform many of the same tasks as hypertext content links.

Automatic indexing remains a topic of active research in the hypertext field [5,11,12,14,15]. Salton and colleagues suggest that their approach towards full-text document retrieval can be used successfully to determine content links between hypertext components [27], but this claim is controversial. Hahn's TOPIC system employs a model of knowledge-based text condensation that transforms text-representation structures into more abstract thematic descriptions of text content. This process attempts to discard irrelevant knowledge structures and to retain only salient concepts. The topical structure of text is then represented in a hierarchical text graph which supports variable degrees of abstraction for text summarization as well as content-oriented retrieval of text knowledge. These text graphs provide a methodology for the automatic generation of hypertexts from full-text files. When applied to hierarchical hypertexts, these principles permit multiple levels of information granularity and abstraction - features useful to design and use of large-scale hypertext systems [19,20,26].

8. Discussion Topics

Hypertext developers often ignore link semantics when creating information retrieval systems for hypertexts and address information retrieval needs by equating hypertexts with a collection of unrelated "tiny documents." If each hypertext component has less information content than a conventional document and if the organizational information expressed through links is not considered, it becomes harder to argue that conventional methods will prove satisfactory. This argument would hold true even if very elegant external "hyperconcept databases" are applied to a hypertext.

Automatic hypertext production is difficult to achieve. Decomposing conventional text into a network of components requires many decisions which may impact greatly on component classification and hypertext use. Although success has been realized when documents are modular in nature and hierarchical in organization, it is not clear how these techniques can be generalized to other literature, nor is it clear how one can integrate multiple smaller hypertexts into a single larger hypertext.

Although abstract models have the potential to enhance retrieval from and understanding of hypertexts, very little work has moved from the speculative stage to the large-scale experimental stage. Among the many factors responsible for this slow movement is the lack of robust hypertexts of sufficient size and complexity to warrant full-scale research projects. The usual claim is that lack of information retrieval techniques impedes the delivery of large-scale hypertexts, but it may be instead that appropriate experiments are not available because large-scale hypertexts have not been made available to the appropriate researcher.

Controversy over the applicability to hypertext of techniques like Salton's will parallel that taking place when addressing full-text document collections [6]. It is possible however that in hypertext the discussions are impeded not so much by a lack of information retrieval techniques as by a lack of experimental methodology suitable to evaluate a retrieval technique's efficacy when applied to hypertext.

As hypertexts become larger, hypertext servers will be necessary. This necessity will require developers to understand how multiple users will share the same hypertext and how hypertexts can be distributed across multiple servers. In contrast to distributed file systems, distribution of hypertext components will be made arbitrarily in order to preserve transparency of the distributed system. This distribution scheme will produce an unbounded amount of interconnection among the hypertext components, and even small changes could place a tremendous burden on the network uniting the servers . Full transparency could bring unpredictable performance.

Citations

1. Agosti, M. Is hypertext a new model of information retrieval? In Proceedings of OnLine Information 88: 12th International Online Information Meeting. London, UK, 57-62,1988.

2. Agosti, M., R. Colotti, G. Gradenigo, P. Mattiello, A. Archi, R. DiGiorgi, B. Inghirami, R. Nannucci and M. Ragona. New prospectives in information retrieval techniques: a hypertext prototype in environmental law. In Proceedings of OnLine Information 89: 13th International Online Information Meeting. London, UK, 483-494,1989.

3. Agosti, M., F. Crestani and G. Gradenigo. Towards data modelling in information retrieval. Journal of Information Science. 25(6): (in press), 1989.

4. Agosti, M., G. Gradenigo and P. Mattiello. The hypertext as an effective information retrieval tool for the final user. In Proceedings of Pre-proceedings of the III International Conference on Logics, Informatics, and Law. Firenze, 1-19,1989.

5. Bernstein, M. An Apprentice That Discovers Hypertext Links. Eastgate Systems, PO Box 1307, Cambridge, MA 02238,1990.

6. Blair, D. C. and M. E. Maron. An Evaluation of Retrieval Effectiveness for a Full-text Document Retrieval System. Communications of the ACM. 28(3): 289-299, 1985.

7. Campbell, B. and J. M. Goodman. HAM: A General Purpose Hypertext Abstract Machine. Communications of the ACM. 31(7): 856-861, 1988.

8. Clitherow, P., D. Riecken and M. Muller. VISAR: A System for Inference and Navigation of Hypertext. In Proceedings of Hypertext'89. Pittsburgh, PA, 293-304,1989.

9. Conklin, J. Hypertext: an introduction and survey. IEEE Computer. 20(9): 17-41, 1987.

10. Consens, M. P. and A. O. Mendelzon. Expressing structural queries in GraphLog. In Proceedings of Hypertext'89. Pittsburgh, Pa., 269-292,1989.

11. Coombs, J. J. Hypertext, Full Text, and Automatic Linking. In Proceedings of SIGIR 90 (to appear). Brussells, 1990.

12. Croft, W. B. and H. Turtle. A retrieval model incorporating hypertext links. In Proceedings of Hypertext 89. Pittsburgh, PA, 1989.

13. Frisse, M. E. Searching for Information in a Hypertext Medical Handbook. Communications of the ACM. 31(7): 880-886, 1988.

14. Frisse, M. E. and S. B. Cousins. Information Retrieval from Hypertext: Update on the Dynamic Medical Handbook Project. In Proceedings of Hypertext'89. Pittsburgh, PA, 199-212,1989.

15. Frisse, M. E. and S. B. Cousins. Guides for Hypertext: An Overview. To appear in Artificial Intelligence in Medicine (Elsevier). Fall 1990.

16. Furuta, R. and P. D. Stotts. Programmable Browsing Semantics in Trellis. In Proceedings of Hypertext'89. Pittsburgh, Pa., 27-42,1989.

17. Garg, P. G. and W. Scacchi. A software hypertext environment configured software description. In Proceedings of International Workshop on Software Version and Configuration Control. Grassau, FRG, 1988.

18. Garg, P. K. Abstraction Mechanisms in Hypertext. Communications of the ACM. 31(7): 862-870, 1988.

19. Hahn, U. and U. Reimer. Automatic generation of hypertext knowledge bases: Conference on OFfice Information

Systems, Palo Alto. SIGOIS Bulletin. 9(2): 182-188, 1988.

20. Hahn, U. and U. Reimer. Knowledge-based text analysis in office environments: the text condensation system TOPIC. In Proceedings of Office Knowledge: Representation, Management and Utilization. Selected Full Papers based on Contributions to the IFIP TC 8/WG 8.4 International Workshop. 1988. Toronto, Ontario, 1988.

21. Halasz, F. G. Reflections on Notecards: Seven Issues for the Next Generation of Hypermedia Systems. Communications of the ACM. 31(7): 836-855, 1988.

22. Jarwa, S. and M-F.A.Bruandet. A hypertext model for information management in software management in software engineering. In Proceedings of DEXA'90, International Conference on Database and Expert Systems Applications. Austin, 1990.

23. Marchionini, G. and B. Shneiderman. Finding Facts vs. browsing knowledge in hypertext systems. Computer. 21(1): 70-89, 1988.

24. Norman, M. D. and D. S. Mayer. Context- a partitioning concept for hypertext. ACM Transactions of Office Information Systems. 5(2): 168-186, 1987.

25. Rada, R. Writing and Reading Hypertext: An Overview. Journal of the American Society for Information Science. 40(3): 164-171, 1989.

26. Reimer, U. and U. Hahn. Text condensation as knowledge base abstraction. In Proceedings of Fourth Conference on Artificial Intelligence Applications. San Diego, CA, USA, 1988.

27. Salton, G. Personal Communication.

28. Smith, J. B. and S. F. Weiss. An overview of hypertext. Communications of the ACM. 31(7): 816-819, 1988.

29. Stotts, P. D. and R. Furuta. Petri-net-based hypertext: document structure with browsing semantics. ACM Transactions on Information Systems. January: 1989.

30. Tompa, F. W. A data model for flexible hypertext data systems. ACM Transactions on Office Information Systems. 7(1): 85-100, 1989.

Complete Addresses of Panelists:

Maristella Agosti
Dept. Electronics & Informatics, Via Gradenigo, 6/a, 35131 Padova,Italy.
EMAIL: agosti@ipdunivx.unipd.it

Marie-France Bruandet and Sahar Jarwa
Laboratoire Genie Informatique, IMAG BP 53X 38041 Grenoble Cedex France.
EMAIL: bruandet@imag.imag.fr

Mark E. Frisse
Department of Internal Medicine (Box 8121), Washington University School of Medicine, 660 S. Euclid Avenue, St. Louis, MO. 63110 , USA.
EMAIL: frisse@informatics.WUSTL.EDU

Udo Hahn
Department of Math and Computer Science, University of Freiburgh,West Germany.
EMAIL: udo.hahn@informatik.uni-freiburgh.databasep.de

Stephen F. Weiss
Department of Computer Science, University of North Carolina, Chapel Hill, NC 27599
EMAIL: weiss@cs.unc.edu