

Lecture Notes in Computer Science

1326

Edited by G. Goos, J. Hartmanis and J. van Leeuwen

Advisory Board: W. Brauer D. Gries J. Stoer

Charles Nicholas James Mayfield (Eds.)

Intelligent Hypertext

Advanced Techniques
for the World Wide Web



Springer

Series Editors

Gerhard Goos, Karlsruhe University, Germany

Juris Hartmanis, Cornell University, NY, USA

Jan van Leeuwen, Utrecht University, The Netherlands

Volume Editors

Charles Nicholas

University of Maryland Baltimore County

Department of Computer Science and Electrical Engineering

1000 Hilltop Circle, Baltimore, MD 21250, USA

E-mail: nicholas@csee.umbc.edu

James Mayfield

The Johns Hopkins University, Applied Physics Laboratory

Johns Hopkins Road, Laurel, MD 20723-6099, USA

E-mail: james.mayfield@ihup.edu

Cataloging-in-Publication data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Intelligent hypertext : advanced techniques for the World Wide Web / Charles Nicholas ; James Mayfield (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ; Budapest ; Hong Kong ; London ; Milan ; Paris ; Santa Clara ; Singapore ; Tokyo : Springer, 1997
(Lecture notes in computer science ; Vol. 1326)
ISBN 3-540-63637-4

CR Subject Classification (1991): H.2

ISSN 0302-9743

ISBN 3-540-63637-4 Springer-Verlag Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution under the German Copyright Law.

© Springer-Verlag Berlin Heidelberg 1997

Printed in Germany

Typesetting: Camera-ready by author

SPIN 10545719 06/3142 - 5 4 3 2 1 0 Printed on acid-free paper

Preface

James Mayfield¹ and Charles Nicholas²

¹ Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland, USA

² Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, Maryland, USA *

The papers in this volume grew out of contributions and discussions held at the Workshops on Intelligent Hypertext, which were held in conjunction with the 1993 and 1994 Conferences on Information and Knowledge Management (CIKM'93 and CIKM'94). The theme of the CIKM conference is the interaction between the fields of artificial intelligence, database management, and information retrieval. With their emphasis on intelligent systems for managing information, using the hypertext metaphor, the two Workshops on Intelligent Hypertext were consistent with this theme.

In the years since the workshops took place, the World Wide Web has achieved enormous success. Indeed, for many people, the Web is the only hypertext system they have ever used or even heard of. With this in mind, the editors invited several of the workshop participants to prepare chapter-length papers based on their workshop contributions, with special emphasis on how their results apply to the Web now, or might apply to versions of the Web that appear in the future.

Bieber describes his experience with creating a Web version of the August 1995 issue of *Communications of the ACM*. He points out that the although rich semantics for hypertext links is a familiar idea in the hypertext community, the Web in its current form does little to support this. In particular, no tool to aid authors in adding semantic information to their links is available, making this process laborious and error-prone. Bieber goes on to describe a browser in which semantic information *about* a link, e.g., what sort of relationship exists between the current document and the one being linked to, is presented. One of the important lessons to be learned from this effort is that producing a well-designed hypertext document of non-trivial size is still a significant effort.

Brusilovsky discusses adaptive hypermedia, in which the hypermedia system adapts to its individual human users. This idea goes beyond the idea of "preferences," standard in contemporary browsers. Adaptive hypermedia includes adaptive content, in which different users may see different versions of what is essentially the same hypermedia document, whether that document is an ordinary Web page, for example, or perhaps an annotation or index document. Adaptive hypermedia also includes the notion of adaptive navigation, in which some users see links that others do not. Brusilovsky addresses the

* Supported in part by a Sabbatical Fellowship from the United States Department of Defense.

question of how adaptive hypermedia might be evaluated, and he discusses many example systems.

Gobbetti and Turner describe the Virtual Reality Modeling Language (VRML), and explain how 3D documents may be implemented using VRML and the Web. They present an overview of VRML's historical beginnings, implementation issues, and sample applications. Of particular interest is the "Virtual Sardinia Project," in which geographic information, including satellite images, are used to create a VR model of the island of Sardinia. This model can then be explored with a VRML-enabled Web browser.

One component of intelligent hypertext systems (or indeed of intelligent systems in general) is the degree to which the system adapts (or can be adapted) in order to better serve its users. Kay and Kummerfeld describe their PT system, which builds and maintains a model of each user. For new users, this process begins by PT asking them questions about their background knowledge in computer programming. As the system runs, this user model is updated, and information derived from it is used to generate customized documents, and to customize navigation between documents.

Kent and Neuss argue that World Wide Web search should be carried out by exchange of meta-data, not (as is now practiced) by exchange of the data themselves. They suggest that *formal concept analysis* can provide tools for representing the semantics of hypertext documents, which can serve as meta-data. *Conceptual linkage* in their formulation is the overlay of relationships among the concepts represented in documents onto an existing hypertext link structure. *Conceptual browsing* is browsing along the lines of such an overlay. Conceptual browsing represents a midpoint between traversing the links that are part of a document (*e.g.*, by clicking on a WWW link) and using a global indexing scheme (*e.g.*, by searching the WWW with AltaVista or HotBot). The techniques they report apply both to the structuring of local hypertexts, and to guidance of the user through such a structure.

Mayfield presents an overview of two-level models of hypertext. For a hypertext system to be considered intelligent, it can be argued, requires that the raw data be augmented by some meta-data structure. The purpose of such meta-data is to organize the information so that it can be accessed more easily. Several forms of meta-data are possible, and Mayfield describes a variety of two-level hypertext systems.

In their chapter on the TELLTALE system, Pearce and Miller present their work on the scalability of information retrieval techniques in a hypertext context. TELLTALE is a hypertext system that provides similarity matches between documents (much as traditional information retrieval systems do). But it is integrated with a hypertext interface in a way that greatly improves the speed and ease with which a user can explore the information space. For example, three statistics-based similarity measures are available to the user for controlling how dynamic hypertext links are created. TELLTALE's two main scaling techniques are 1) the use of *n*-grams (sequences of *n* contiguous

characters) as retrieval terms in lieu of more traditional word-based terms; and 2) the distribution of meta-data about the documents in the collection across several machines. The former adds scalability in the types of documents that may be included, while the latter aids scalability in the number of documents indexed.

The notion of meta-data in hypertext emerges several times in this volume. One specific proposal is presented by Wang, Ghaoui, and Rada, who discuss the use of semantic nets as domain models for hypertext systems. Of particular interest is the issue of how to maintain the consistency (of both meaning and style) of a set of hypertext documents created by many people. As Web content becomes more dynamic, the need for techniques to understand (and control) what material exists, and where it resides, will increase. The MUCH and RICH systems described in this chapter are important steps in this direction.

Rus and Subramanian describe an architecture for intelligent software agents, and argue the suitability of such agents to discover and use information distributed over a network. These agents are intelligent in the sense of being able to examine data, make plans in response to the data, and perform actions according to these plans. Agents are endowed with “sensors” and are capable of making sophisticated analysis of documents at different levels of detail. The agents are mobile, and migration from one site to another is one possible action. Rus and Subramanian describe the formal underpinnings of their system, and discuss its operation in detail.

Table of Contents

Preface

James Mayfield and Charles Nicholas	V
---	---

Enhancing Information Comprehension Through Hypertext

Michael Bieber	1
1. Motivation	1
2. Semantic Link Design	3
3. Semantic Labels: Technical Details	7
4. Additional Issues	8
5. The Next Step: Further Enhancing Information Comprehension	9

Efficient Techniques for Adaptive Hypermedia

Peter Brusilovsky	12
1. Introduction	12
2. What Can be Adapted in Adaptive Hypermedia	13
2.1 Adaptive Presentation	13
2.2 Adaptive Navigation Support	14
3. Adaptive Hypermedia Techniques	16
3.1 Adaptive Presentation Techniques	16
3.2 Adaptive Navigation Support Techniques	19
4. Evaluation of Adaptive Hypermedia Techniques	22
4.1 Evaluation of an Adaptive Presentation Technique	23
4.2 Evaluation of an Adaptive Navigation Support Technique: Sorting	24
4.3 Evaluation of Adaptive Navigation Support Techniques: Hiding and Annotation	25
5. Conclusion	27

Exploring Annotated 3D Environments on the World Wide Web

Enrico Gobetti and Russell Turner	31
1. Introduction	31
1.1 Adding 3D Contents to the Web	32

X Table of Contents

1.2	VRML-1.0	33
1.3	VRML-1.0 Applications	34
2.	VRML-1.0 Browsers	35
2.1	Application Overview	36
2.2	User Interaction	37
2.3	World Model	38
2.4	Rendering	38
3.	VRML-1.0 Application Examples	39
3.1	The CERN VENUS Project	39
3.2	The Virtual Sardinia Project	41
4.	Conclusions	43

User Models for Customized Hypertext

Judy Kay and Bob Kummerfeld	47
1. Introduction	47
2. Motivation for generating diverse forms of a text	48
2.1 Customisation based on background knowledge	48
2.2 Customisation based language level and writing style	50
2.3 Customisation of examples and exercises	50
2.4 Customisation by level of abstraction	51
2.5 Summary of motivations for customisation beyond conventional hypertext	51
2.6 Customisation based on varied teaching strategies	51
3. Customisation without duplication	52
4. User models to customise hypertext	54
4.1 Domain representation level	55
4.2 Stereotype level	55
4.3 Individual level	56
4.4 Building the model	57
5. Implementation	57
5.1 Authentication	58
5.2 Concept Inventory	58
5.3 Form of Course Pages	58
5.4 Customisation	61
6. Support for authoring	62
6.1 Customisation tags	62
6.2 Executable pages	64
6.3 Customisation interface	65
6.4 Style sheets	65
7. Related work	65
8. Issues for adaptive hypertext	66
8.1 Document stability	66
8.2 Learner Control	67
9. Conclusion	67

Conceptual Analysis of Hypertext

Robert E. Kent and Christian Neuss	70
1. Conceptual Knowledge Systems	70
2. Resource Meta-information	72
3. Conceptual Linkage	77
4. Conceptual Neighborhood	79
5. Conceptualization Processes	83
6. Summary, Implementation, and Future Work	86

Two-Level Models of Hypertext

James Mayfield	90
1. Introduction	90
1.1 Traditional Hypertext	90
1.2 Problems with Hypertext Links	91
1.3 Two-Level Models of Hypertext	92
2. Declarative Two-Level Models	93
2.1 Semantic Net-Based Models	93
2.2 Other Net-Based Models	96
2.3 Clustering Models	99
3. Procedural Two-Level Models	101
3.1 Computed Links	102
3.2 Automated Link Construction	103
4. Conclusions	104

The TELLTALE Dynamic Hypertext Environment:**Approaches to Scalability**

Claudia Pearce and Ethan Miller	109
1. Introduction	109
2. TELLTALE – A Dynamic Hypertext Environment	110
2.1 Dynamic Link Types Using <i>n</i> -grams	111
2.2 Hypertext Interface Overview	116
2.3 Text Highlighting Techniques	117
2.4 Multilingual Text	119
3. Scalability Research Platform	121
3.1 Prototype Goals	122
3.2 Data Structures	123
3.3 Algorithms Used	124
3.4 Scaling Experiments	125
3.5 Future Prototype Experiments	126
4. Summary and Conclusions	128

Domain Model Based Hypertext for Collaborative Authoring

Weigang Wang, Claude Ghaoui, Roy Rada 131

1. Introduction 131
2. Domain Model Based Semantic Nets 132
 - 2.1 Domain Analysis 133
 - 2.2 Semantic Net Representation 133
 - 2.3 Rules and Operators 134
3. Experiences 135
 - 3.1 Creating Multi-hierarchical Hypertext Documents 135
 - 3.2 The Support of Coordination 137
 - 3.3 Construction of HTML Documents 140
4. Discussion 141

**Information Retrieval, Information Structure, and
Information Agents**

Daniela Rus and Devika Subramanian 145

1. Introduction 145
 - 1.1 Related Work 148
2. Organizing Principles for Information Gathering Search Engines ... 150
 - 2.1 Segmenters 151
 - 2.2 Structure Detectors 157
 - 2.3 Assembly of Information Agents 163
3. Example 1: Compiling Reports from Tabular Data 168
4. Example 2: Automatic Information Gathering with Customized
Indexes 170
 - 4.1 Experimental Data 174
5. Discussion 176

List of Contributors

Michael Bieber

Institute of Integrated Systems Research,
New Jersey Institute of Technology,
University Heights, Newark, NJ 07102-1982 USA

Peter Brusilovsky

Human Computer Interaction Institute,
School of Computer Science,
Carnegie Mellon University,
Pittsburgh, PA 15213 USA

Claude Ghaoui

CMS, Liverpool John Moores University,
Liverpool L3 3AF, UK

Enrico Gobbetti

Center for Advanced Studies,
Research and Development in Sardinia,
Via Nazario Sauro 10,
09123 Cagliari, Italy

Judy Kay

Department of Computer Science,
University of Sydney, Australia

Robert Kent

Department of Electrical Engineering and Computer Science,
Washington State University, Pullman, WA 99164-2752, USA

Bob Kummerfeld

Department of Computer Science,
University of Sydney, Australia

James Mayfield

The Johns Hopkins University Applied Physics Laboratory,
Johns Hopkins Road, Laurel, MD 20723-6099, USA

Ethan Miller

Department of Computer Science and Electrical Engineering,
University of Maryland Baltimore County, Baltimore, MD, 21250, USA

Christian Neuss

Technische Hochschule Darmstadt,
Fachgebiet Praktische Informatik,
64289 Darmstadt, Germany

Claudia Pearce

U.S. Department of Defense,
9800 Savage Road, Fort Meade, MD, 20755-6000, USA

Roy Rada

Department of Electrical Engineering and Computer Science,
Washington State University,
Pullman, WA 99164-2752, USA

Daniela Rus

Department of Computer Science,
Dartmouth College, Hanover, NH 03755, USA

XIV List of Contributors

Devika Subramanian

Department of Computer Science,
Rice University, Houston, TX 77005,
USA

Weigang Wang

GMD-IPSI, Dolivostr. 15, D-64293,
Darmstadt, Germany