

---

**SEMANTIC MODELS FOR  
MULTIMEDIA DATABASE  
SEARCHING AND BROWSING**

---

# The Kluwer International Series on **ADVANCES IN DATABASE SYSTEMS**

Series Editor  
**Ahmed K. Elmagarmid**

*Purdue University  
West Lafayette, IN 47907*

## ***Other books in the Series:***

**INFORMATION BROKERING ACROSS HETEROGENEOUS DIGITAL DATA:**

**A Metadata-based Approach**, *Vipul Kashyap, Amit Sheth*; ISBN: 0-7923-7883-0

**DATA DISSEMINATION IN WIRELESS COMPUTING ENVIRONMENTS**, *Kian-*

*Lee Tan and Beng Chin Ooi*; ISBN: 0-7923-7866-0

**MIDDLEWARE NETWORKS: Concept, Design and Deployment of Internet**

**Infrastructure**, *Michah Lerner, George Vanecek, Nino Vidovic, Dad Vrsalovic*;

ISBN: 0-7923-7840-7

**ADVANCED DATABASE INDEXING**, *Yannis Manolopoulos, Yannis Theodoridis,*

*Vassilis J. Tsotras*; ISBN: 0-7923-7716-8

**MULTILEVEL SECURE TRANSACTION PROCESSING**, *Vijay Atluri, Sushil*

*Jajodia, Binto George* ISBN: 0-7923-7702-8

**FUZZY LOGIC IN DATA MODELING**, *Guoqing Chen* ISBN: 0-7923-8253-6

**INTERCONNECTING HETEROGENEOUS INFORMATION SYSTEMS**, *Athman*

*Bouguettaya, Boualem Benatalhh, Ahmed Elmagarmid* ISBN: 0-7923-8216-1

**FOUNDATIONS OF KNOWLEDGE SYSTEMS: With Applications to Databases**

**and Agents**, *Gerd Wagner* ISBN: 0-7923-8212-9

**DATABASE RECOVERY**, *Vijay Kumar, Sang H. Son* ISBN: 0-7923-8192-0

**PARALLEL, OBJECT-ORIENTED, AND ACTIVE KNOWLEDGE BASE**

**SYSTEMS**, *Ioannis Vlahavas, Nick Bassiliades* ISBN: 0-7923-8117-3

**DATA MANAGEMENT FOR MOBILE COMPUTING**, *Evaggelia Pitoura, George*

*Samaras* ISBN: 0-7923-8053-3

**MINING VERY LARGE DATABASES WITH PARALLEL PROCESSING**, *Alex*

*A. Freitas, Simon H. Luvington* ISBN: 0-7923-8048-7

**INDEXING TECHNIQUES FOR ADVANCED DATABASE SYSTEMS**, *Elisa*

*Bertino, Beng Chin Ooi, Ron Sacks-Davis, Kian-Lee Tan, Justin Zobel, Boris*

*Shidlovsky, Barbara Catania* ISBN: 0-7923-9985-4

**INDEX DATA STRUCTURES IN OBJECT-ORIENTED DATABASES**, *Thomas*

*Mueck, Martin L. Polaschek* ISBN: 0-7923-9971-4

**DATABASE ISSUES IN GEOGRAPHIC INFORMATION SYSTEMS**, *Nabit*

*Adam, Aryya Gangopadhyay* ISBN: 0-7923-9924-2

**VIDEO DATABASE SYSTEMS: Issues, Products, and Applications**, *Ahmed A*

*Elmagarmid, Haitao Jiang, Abdelsalam A. Helal, Anupam Joshi, Magdy Ahmea*

ISBN: 0-7923-9872-6

---

# SEMANTIC MODELS FOR MULTIMEDIA DATABASE SEARCHING AND BROWSING

*by*

**Shu-Ching Chen**

*School of Computer Science  
Florida International University*

**R. L. Kashyap**

*School of Electrical and Computer Engineering  
Purdue University*

**Arif Ghafoor**

*School of Electrical and Computer Engineering  
Purdue University*

**KLUWER ACADEMIC PUBLISHERS**  
New York / Boston / Dordrecht / London / Moscow

eBook ISBN: 0-306-47029-2  
Print ISBN: 0-792-37888-1

©2002 Kluwer Academic Publishers  
New York, Boston, Dordrecht, London, Moscow

All rights reserved

No part of this eBook may be reproduced or transmitted in any form or by any means, electronic, mechanical, recording, or otherwise, without written consent from the Publisher

Created in the United States of America

Visit Kluwer Online at: <http://www.kluweronline.com>  
and Kluwer's eBookstore at: <http://www.ebooks.kluweronline.com>

# Contents

List of Figures	vii
List of Tables	xi
Preface	xiii
Acknowledgments	xix
1. INTRODUCTION	1
1. Introduction	1
2. Multimedia Information Applications	4
3. Issues and Challenges	5
2. SEMANTIC MODELS FOR MULTIMEDIA INFORMATION SYSTEMS	19
1. Introduction	19
2. Multimedia Semantic Models	23
3. MULTIMEDIA DATABASE SEARCHING	43
1. Introduction	43
2. Image Segmentation	44
3. Video Parsing and Segmentation Approaches	49
4. Motion Detection and Tracking Approaches	52
5. Iconic-Based Grouping and Browsing Approaches	55
6. Object Recognition Approaches	56
7. Knowledge-Based Event Modeling Approaches	57
8. Characteristics of Video Data Modeling	59
9. Content-Based Retrieval	61
4. MULTIMEDIA BROWSING	69
1. Introduction	69
2. Video Browsing	71
3. Key Frame Selections	78

5. CASE STUDY 1 – AUGMENTED TRANSITION NETWORK (ATN) MODEL	81
1. Introduction	81
2. Spatial and Temporal Relations of Semantic Objects	88
3. Multimedia Presentations	90
4. Multimedia Database Searching	92
5. Multimedia Browsing	98
6. User Interactions and Loops	103
6. CASE STUDY 2 – OBJECT COMPOSITION PETRI NET (OCPN) MODEL	111
1. Introduction	111
2. Interval-Based Conceptual Models	113
7. CONCLUSIONS	127
References	129
Index	147

# List of Figures

2.1	Timeline for Multimedia Presentation. $t_1$ to $t_6$ are the time instances. $d_1$ is time duration between $t_1$ and $t_2$ and so on.	24
2.2	A timeline example that includes choice objects	26
2.3	A timeline tree representation of the interactive scenario in Figure 2.2	27
2.4	An OCPN example for Figure 2.1: $D_1$ is the delay for media streams $I_1$ and $A_1$ to display. $D_2$ is the delay for $V_2$ to display.	30
2.5	An example PNBH Petri net.	32
2.6	<i>Transition Network</i> for Multimedia Presentation.	40
3.1	Comparison of video data semantic models.	44
3.2	(a) tessellations with square tiles of equal size, and (b) tessellations with hexagonal tiles of equal size. This kind of tessellation is independent of the image since it does not consider the information content in the image.	45
3.3	(a) an example of a quad-tree tessellation, and (b) tiling with arbitrary shapes like the shapes of a jigsaw puzzle.	46
3.4	A hierarchy of video media stream	50
4.1	The Cataloging Phase Architecture [144]	75
4.2	Augmented Transition Network for video browsing: (a) is the ATN network for a video clip which starts at the state $V_i$ . (b)-(d) are part of the sub-networks of (a). (b) is to model <i>scenes</i> in video clip $V_i$ . (c) is to model <i>shots</i> in scene $S_i$ . Key frames for shot $T_i$ is in (d).	77

- 5.1 Video frame 1. There are four semantic objects: *salesman*, *box*, *file holder*, and *telephone*; *salesman* is the target semantic object. The relative position numbers (as defined in Table 5.1) of the other three semantic objects are in the 10, 15, and 24, respectively. 95
- 5.2 Video frame 52. Semantic object *box* moves from the left to the front of *salesman* (from viewer's point of view). 95
- 5.3 Video frame 70. Semantic object *box* moves from the front to the left of *salesman* (from viewer's point of view). 95
- 5.4 The corresponding subnetwork for multimedia input string in Equation 5.1. 96
- 5.5 A browsing graph of a mini tour of the Purdue University campus: there are seven sites denoted by  $B_i$ ,  $i = 1 \dots 7$  that are connected by arcs. A directed arc denotes a one-way selection and a bi-direction arc allows two-way selections. 98
- 5.6 ATN for the mini tour of a campus: Seven networks represent seven sites which users can browse. Networks  $B_1$  / through  $B_7$  / represent the presentations for sites Purdue Mall, Computer Science Building, Chemical Engineering Building, Potter Library, Union, Electrical Engineering Building and Mechanical Engineering Building respectively. Each network begins a presentation with three media streams: a video, a text, and an audio, and is followed by selections. After a user selects a site, the control will pass to the corresponding network so that the user can watch the presentation for that site continuously. 99
- 5.7 Timelines for presentation  $P_1$  and  $P_2$ : Figures (a), (c), (e), and (f) are the presentation sequence for presentation  $P_1$ . Figures (b), (d), (e), and (f) are the presentation sequence for presentation  $P_2$ . Figures (e) and (f) are two timelines for selections  $B_1$  and  $B_2$ , respectively. 104

5.8	Augmented Transition Network: (a) is the ATN network for two multimedia presentations which start at the states $P_1/$ and $P_2/$ , respectively. (b)-(d) are part of the subnetworks of (a). (b) models the semantic objects in video media stream $V_1$ , (c) models the semantic objects in image media stream $I_1$ , and (d) models the keywords in text media stream $T_1$ . In (e), the “Get” procedure is to access an individual media stream. “Display” displays the media streams. “Next_Symbol( $X_i$ )” reads the input symbol $X_i$ . The “Next_State” is a procedure to advance to the next state. “Start_time( $X_i$ )” gives the pre-specified starting time of $X_i$ . User thinking time is accounted for by the <b>Delay</b> variable. $\theta$ is a parameter.	105
6.1	(a) Temporal relations represented graphically by a timeline representation. (b) The corresponding Petri net (OCPN) representation for the temporal relations.	117
6.2	A unified OCPN model.	118
6.3	The $n - ary$ temporal relations.	120
6.4	The forward and reverse relations and intervals.	122
6.5	Partial interval evaluation.	124

# List of Tables

2.1	Classification of Selected Semantic Models	25
2.2	Operators in TCSP	38
3.1	Comparison of Several Video Parsing Approaches	51
3.2	Characteristics of Several Selected Video Data Modeling Approaches	60
5.1	Three dimensional relative positions for semantic objects: The first and the third columns indicate the relative position numbers while the second and the fourth columns are the relative coordinates. $(x_t, y_t, z_t)$ and $(x_s, y_s, z_s)$ represent the X-, Y-, and Z-coordinates of the target and any semantic object, respectively. The " $\approx$ " symbol means the difference between two coordinates is within a threshold value.	89
5.2	Condition and action table: Get procedure is to access an individual media stream. Get-Symbol is a procedure to read the next input symbol of multimedia input string. Next-State is a procedure to advance to the next state in ATN. Display procedure is to display the media streams. $\theta$ , $\tau$ and $\rho$ are the parameters.	93
5.3	The trace of ATN for the specified browsing sequence.	101
5.4	The trace of ATN for presentation $P_1$ .	106
5.5	Continuation of Table 5.4 if $B_1$ is chosen.	107
5.6	Continuation of Table 5.4 if $B_2$ is chosen.	107
6.1	Temporal Parameters of the Unified Model in Figure 6.2 ( $P_\alpha$ TR $P_\beta$ ).	118
6.2	$n$ - ary Temporal Constraints.	121

6.3	Temporal Parameter Conversions ( $P_\alpha$ $TR$ $P_\beta$ to $P_{\alpha-r}$ $TR_r$ $P_{\beta-r}$ ).	123
6.4	$n$ - ary Temporal Parameter Conversions.	123

# Preface

The objective of this book is to provide a survey of different models and to cover state of the art techniques for multimedia presentations, multimedia database searching, and multimedia browsing. Therefore, the readers can have an understanding of the issues and challenges of multimedia information systems. As more information sources become available in multimedia systems, the development of the abstract models for video, audio, text, and image data becomes very important. The pros and cons of the different models for multimedia information designs are discussed in this book. In addition, this book will cover most of the recent works that were published on the prestigious Journals and Conferences such as IEEE Transactions on Knowledge and Data Engineering, ACM Multimedia System Journal, Communications of the ACM, IEEE Computer, IEEE Multimedia, ACM SIGMOD and so on.

This book is aimed at the general readers who are interested in the issues, challenges, and ideas underlying the current practice of multimedia presentation, multimedia database searching, and multimedia browsing in multimedia information systems. It will also be of interest to university researchers, scientists, industry professionals, software engineers, graduate students, and undergraduate students who need to be become acquainted with this new multimedia technology, and to all those who wish to gain a detailed technical understanding of what multimedia information systems involve.

This book is organized in the way that makes the ideas accessible to the readers who are interested in grasping the basics, as well as to those who would like more technical depth.

The first chapter introduces multimedia information applications, the need for the development of the multimedia database management systems (MDBMSs), and the important issues and challenges of multimedia systems. With the increasing complexity of real world multimedia applications, multimedia systems require the management and delivery of extremely large bodies of data at very high rates and may require the delivery with real-time constraints. The applications expected to benefit enormously from multimedia technologies and multimedia information systems include remote collaboration via video teleconferencing, improved simulation methodologies for all disciplines of science and engineering, and better human-computer interfaces.

Also, a multimedia system should be able to accommodate the heterogeneity that may exist among the data. Hence, a new design of an MDBMS is required to handle the temporal and spatial requirements, and the rich semantics of multimedia data such as text, image, audio, and video. The purpose of the design and development of an MDBMS is to efficiently organize, store, manage, and retrieve multimedia information from the underlying multimedia databases. In other words, an MDBMS should have the ability to model the varieties of multimedia data in terms of their structure, behavior and function.

The issues and challenges discussed in this chapter include:

- formal semantic modeling techniques
- indexing and searching methods
- synchronization and integration modeling
- formal query languages
- data placement schemas
- architecture and operating system support
- distributed database management
- multimedia query support, retrieval, and browsing

The second chapter discusses the temporal relations, the spatial relations, the spatio-temporal relations, and several semantic models for multimedia information systems. As more information sources become available in multimedia systems, the development of abstract semantic models for multimedia information becomes very important. An abstract semantic model should be rich enough to provide a friendly

interface of multimedia presentation synchronization schedules to the users and should be a good programming data structure for implementation to control multimedia playback. In other words, the models must be devised able to support the specification of temporal constraints for multimedia data and the satisfaction of these constraints must be at run-time. The use of a model that can represent the temporal constraints on multimedia data makes it easier for the satisfaction of these constraints at presentation time.

The semantic models can be classified into the following categories:

- timeline models
- time-interval based models
- graphic models
- petri-net models
- object-oriented models
- language-based models
- augmented transition network (ATN) models

Several existing models in each category are introduced in this chapter. Some models are primarily aimed at synchronization aspects of the multimedia data while others are more concerned with the browsing aspects of the objects. The former models can easily render themselves to an ultimate specification of the database schema. Some models such as based on graphs and Petri-nets have the additional advantage of pictorially illustrating synchronization semantics and are suitable for visual orchestration of multimedia presentations.

The third chapter introduces the issues for multimedia database searching. Multimedia database searching requires semantic modeling and knowledge representation of the multimedia data. Two criteria are considered to classify the existing approaches of modeling multimedia data, especially video data. These two criteria are level of abstraction and granularity of data processing. Based on these two criteria, several classes of approaches employed in modeling video data are compared in this chapter.

Some important issues are discussed here. They include:

- image segmentation and image segmentation techniques

- video segmentation and video parsing
- motion detection and tracking approaches
- iconic-based grouping and browsing approaches
- object recognition approaches
- knowledge-based event modeling approaches
- content-based retrieval

The fourth chapter discusses the issues for multimedia browsing and introduces several existing multimedia browsing systems. Cataloging and indexing of video is a critical step to enable intelligent navigation, search, browsing, and viewing of digital video. While the importance of a seamless integration of querying, searching, browsing, and exploration of data in a digital library collection is recognized, this chapter focuses on the challenges associated with video browsing. An increasing number of digital library systems allow users to access not only textual or pictorial documents, but also video data. Digital library applications based on huge amount of digital video data must be able to satisfy complex semantic information needs and require efficient browsing and searching mechanisms to extract relevant information. In most cases, users have to browse through parts of the video collection to get the information they want, which address the contents and the meaning of the video documents. Hence, a browsing system has to provide the support for this kind of information-intensive work.

The browsing systems introduces in this chapter include:

- a client/server architecture-based browsing system
- the VideoQ browsing system
- the CueVideo browsing system
- the Informedia browsing system
- the augmented transition network (ATN) browsing system

In addition, the importance of key frame selection algorithms is also included, and some of the key frame selection algorithms are discussed in this chapter.

Two case studies are given in the fourth and the fifth chapters. The fourth chapter introduces the augmented transition network (ATN) model

and the fifth chapter introduces the object composition Petri net (OCPN) model. These two models are proposed by the authors of this book. The ATN model has the capabilities to model multimedia presentations, multimedia database searching, and multimedia browsing. Also, the temporal, spatial, or spatio-temporal relations of various media streams and semantic objects can be captured by the proposed ATN model. On the other hand, the OCPN model is based on the logic of temporal intervals and Timed Petri Nets to store, retrieve, and communicate between multimedia objects. These two models are discussed in details in these two chapters.

SHU-CHING CHEN, R. L. KASHYAP, AND ARIF GHAFOR

# Acknowledgments

Much of the information and insight presented in this book was obtained from recent works that were published on the prestigious Journals and Conferences such as IEEE Transactions on Knowledge and Data Engineering, ACM Multimedia System Journal, Communications of the ACM, IEEE Computer, IEEE Multimedia, ACM SIGMOD and so on.

The authors would like to thank Prof. Mei-Ling Shyu at Department of Electrical and Computer Engineering, University of Miami for her pertinent and constructive comments that helped us to improve this book significantly. We would like to thank Dr. Srinivas Sista for his valuable suggestions and comments on the image/video segmentation.

**Acknowledgement:** The work presented in this book has been supported, in part, by the NSF grants IRI-9619812 and IIS-9974255.