Chapter 1

INFORMATION ORGANIZATION AND DATABASES

Katsumi Tanaka
Graduate School of Science & Technology,
Kobe University, Kobe 657-8501, Japan
tanaka@db.cs.kobe-u.ac.jp

Shahram Ghandeharizadeh
Dept. of Computer Science,
University of Southern California, Los Angeles,
CA 90089-0781, USA
shahram@pollux.usc.edu

Yahiko Kambayashi
Dept. of Social Informatics,
Graduate School of Informatics,
Kyoto University, Kyoto 606-8501, Japan
yahiko@isse.kuis.kyoto-u.ac.jp

Abstract In this introduction, we describe needs and current state-of-the-art technologies on information organization, and the overview of the content of the book.

Keywords: information organization, databases, continuous media, semistructured data, spatial data, multimedia, data mining, data warehousing, indexing, query processing, WWW, hypermedia
1. INTRODUCTION

Conventional DBMS technology lacks effective ways to organize information since it assumed formatted data and rigid database schema structures. The major characteristics of data appearing in WWW and multimedia databases are the semistructuredness and the continuity. Both of these two characteristics lead to the difficulties in defining rigid database schema structures.

The present book intends to provide recent developments of the technologies concerned with the information organization technology. The book mainly consists of the following seven parts:

- semistructured data
- multimedia database organization
- multimedia database environments
- data mining and data warehousing
- indexing and query processing
- spatial data access and indexing
- WWW and hypermedia

2. SEMISTRUCTURED DATA

Most of WWW data are called semistructured since it does not have a rigid database schema that controls all the Web pages. Therefore, we need methodologies and tools to store, retrieve and share those semistructured data instead of conventional DBMS technology. The three articles in Part 2 discuss important research issues in semistructured data, such as query languages and integrity constraints. Suciu (Chapter 2) discusses the treatment of semistructured data, especially, the relationship between semistructured data and XML. Tajima (Chapter 3) proposes a new querying facility, called entity-oriented queries, for semistructured data by discovering meaningful subgraphs in semistructured data. Oomoto et al. (Chapter 4) discuss a class of integrity constraints, called path existence constraints, which can be often found in OODBs and semistructured data.

3. MULTIMEDIA DATABASE ORGANIZATION

In the area of multimedia databases, the continuity becomes a crucial point for effective indexing, querying and delivering. The five articles