Chapman & Hall/CRC
Data Mining and Knowledge Discovery Series

Information Discovery on Electronic Health Records

Edited by
Vagelis Hristidis

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Information Discovery on Electronic Health Records

Chapman & Hall/CRC
Data Mining and Knowledge Discovery Series

SERIES EDITOR

Vipin Kumar
University of Minnesota
Department of Computer Science and Engineering
Minneapolis, Minnesota, U.S.A

AIMS AND SCOPE

This series aims to capture new developments and applications in data mining and knowledge discovery, while summarizing the computational tools and techniques useful in data analysis. This series encourages the integration of mathematical, statistical, and computational methods and techniques through the publication of a broad range of textbooks, reference works, and handbooks. The inclusion of concrete examples and applications is highly encouraged. The scope of the series includes, but is not limited to, titles in the areas of data mining and knowledge discovery methods and applications, modeling, algorithms, theory and foundations, data and knowledge visualization, data mining systems and tools, and privacy and security issues.

PUBLISHED TITLES

UNDERSTANDING COMPLEX DATASETS: Data Mining with Matrix Decompositions
David Skillicorn

COMPUTATIONAL METHODS OF FEATURE SELECTION
Huan Liu and Hiroshi Motoda

CONSTRAINED CLUSTERING: Advances in Algorithms, Theory, and Applications
Sugato Basu, Ian Davidson, and Kiri L. Wagstaff

KNOWLEDGE DISCOVERY FOR COUNTERTERRORISM AND LAW ENFORCEMENT
David Skillicorn

MULTIMEDIA DATA MINING: A Systematic Introduction to Concepts and Theory
Zhongfei Zhang and Ruofei Zhang

NEXT GENERATION OF DATA MINING
Hillol Kargupta, Jiawei Han, Philip S. Yu, Rajeev Motwani, and Vipin Kumar

DATA MINING FOR DESIGN AND MARKETING
Yukio Ohsawa and Katsutoshi Yada

THE TOP TEN ALGORITHMS IN DATA MINING
Xindong Wu and Vipin Kumar

GEOGRAPHIC DATA MINING AND KNOWLEDGE DISCOVERY, Second Edition
Harvey J. Miller and Jiawei Han

TEXT MINING: CLASSIFICATION, CLUSTERING, AND APPLICATIONS
Ashok N. Srivastava and Mehran Sahami

BIOLOGICAL DATA MINING
Jake Y. Chen and Stefano Lonardi

INFORMATION DISCOVERY ON ELECTRONIC HEALTH RECORDS
Vagelis Hristidis

Chapman & Hall/CRC
Data Mining and Knowledge Discovery Series

Information Discovery on Electronic Health Records

Edited by
Vagelis Hristidis

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

Chapman & Hall/CRC
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2010 by Taylor and Francis Group, LLC
Chapman & Hall/CRC is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number: 978-1-4200-9038-3 (Hardback)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data

Information discovery on electronic health records / editor, Vagelis Hristidis.
p. cm. -- (Chapman & Hall/CRC data mining and knowledge discovery series)
Includes bibliographical references and index.
ISBN 978-1-4200-9038-3 (alk. paper)
1. Medical records--Data processing. 2. Data mining. I. Hristidis, Vagelis.

R864.I53 2010
610.285--dc22 2009025359

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

To my wife Jelena and my family.

Contents

Preface

Electronic health records (EHRs) are a key component of the information technology revolution occurring in healthcare. EHRs can help improve the quality of healthcare and reduce healthcare costs. Most of the research efforts so far have studied the important and critical problem of standardization of EHRs and interoperability of healthcare information management systems. However, little work has been conducted on the problem of leveraging the rich information found in EHRs, which can improve the quality of medical practice at the point-of-care, or facilitate research. The information stored in EHRs is valuable for practitioners and researchers from the areas of medicine, public health, nursing, law, and health insurance.

In this book, we study the problem of information discovery on EHRs, which involves (a) *searching* the EHR collection given a user query and returning relevant fragments from the EHRs and (b) *mining* the EHR collection to extract interesting patterns, group entities to various classes, or to decide whether an EHR satisfies a given property. An example of searching would be "find the patients related to 'asthma.'" This seemingly simple query turns out to be challenging for many reasons. Should we rank higher a patient with "asthma" in the diagnosis of a past hospitalization, or a patient for whom "asthma" is mentioned in her medical history, or a patient whose EHR refers to "respiratory distress," which is a term related to "asthma"? An example of mining would be "identify patients with high probability of developing asthma." Answering this question involves learning correlations in the EHR collection and using classification algorithms. Most of the book focuses on textual or numeric data of EHRs, where more searching and mining progress has occurred. We also include a chapter on the processing of medical images.

Information discovery on EHRs has some unique challenges compared to information discovery on other domains such as the Web or a bibliographic database. Some of these challenges are medical privacy concerns, lack of standardization for the representation of EHRs, missing or incorrect values, availability of multiple rich health ontologies, and the often small statistical samples. Addressing these challenges requires interdisciplinary collaboration, which is often difficult to achieve, and this has led to relatively little and narrow public information on this important topic.

In this book, we have assembled an extraordinary interdisciplinary team including scientists from the areas of computer science, medicine, law, math, decision sciences, and biomedical engineering. The book, therefore, covers multiple aspects of information discovery on EHRs, such as ethics/privacy, EHR creation, and EHR processing.

To ensure consistent style and flow across the book, I have, in addition to being the editor, coauthored four chapters, and closely reviewed the rest of the chapters. One of the key goals was to minimize the use of technical jargon in most of the book, so that readers from different disciplines and disparate backgrounds can appreciate the content. In each chapter, we have tried to push the technical material to the second half of the chapter, to allow both experts and nonexperts of the specific chapter's material to satisfy their learning needs. Chapters present state-of-the-art research topics from the perspective of the chapter authors, but also present a survey of the achievements in that area.

The book is organized as follows. Chapter 1 presents an overview of the Extensible Markup Language (XML), which is the data model adopted by most of the recent EHR formatting standards. Chapter 2 presents an overview of EHRs, including what information they include, how they are formatted, and what software systems manage them. Chapter 3 defines the term "information discovery," clarifies related terminology, and presents an overview of the challenges and solutions in different aspects of information discovery on EHRs. Chapter 4 discusses data quality and integration issues in EHRs, including how EHRs are created, which help the reader better understand the processing and discovery challenges of EHRs. Chapter 5 discusses the ethical, legal, and social issues around EHRs, which must be known to everyone who processes or manages EHRs. Chapters 6 to 10 present in detail various aspects of information discovery on EHRs. Chapters 6 and 7 discuss the problems for searching and mining EHRs, respectively. Chapter 8 focuses on how data mining techniques, such as those discussed in Chapter 7, can be adapted in a way that the privacy of the data is preserved; that is, specific data for a specific patient are not revealed. Chapter 9 investigates a different setting, where EHR data are collected or processed by mobile devices. The real-time data analysis needs are also discussed. Finally, Chapter 10 tackles the problem of searching and processing medical images, and in particular the problem of medical image segmentation.

Target Audience

A key goal set before the writing of this book was to make it appropriate for multiple disciplines and a wide audience. This is why we tried to minimize the technical jargon and explain the used terminology where possible. Some chapters are more technical than others. In particular, we believe that Chapters 1 to 5 are appropriate for any audience with basic scientific backgrounds. In Chapters 6 to 10, which are more technical, we tried to contain the more technical material to the second half of the chapter, in order to allow nontechnical readers to absorb the key ideas of the chapters.

The following are some examples of the target audience of this book.

- Medical informaticians, who are interested on how the EHR data can be searched and mined
- Computer science students and researchers, who want to make the jump to healthcare research
- Medical students who want to learn about EHRs and the way they are leveraged to extract useful knowledge
- Medical, statistical, or other types of researchers who study medical trends or patterns
- Medical, computer science, or information technology students taking a course on “mining medical data”

Vagelis Hristidis

Acknowledgments

I would like to thank all the contributors of this book for their effort and dedication and for believing in the success of this book. Obviously, this book would not be possible without their support. I would also like to thank Randi Cohen and Professor Vipin Kumar, who are the executive and series editors for this book series, for their support.

About the Editor

Vagelis Hristidis (also Evangelos Christidis) received his bachelor's degree in electrical and computer engineering at the National Technical University of Athens in 1999. He later moved to San Diego, California, where he finished his master's and doctoral degrees in computer science in 2000 and 2004, respectively, at the University of California, San Diego. Since 2004, he has been an assistant professor at the School of Computing and Information Sciences at Florida International University in Miami, Florida.

Dr. Hristidis is an expert in database systems and information retrieval (IR). His main research contribution is his work on bridging the gap between databases and IR, by facilitating keyword searching on structured databases. He has successfully applied these techniques to bibliographic, biomedical, and clinical databases, in collaboration with domain experts from the areas of medicine and biology. Dr. Hristidis has also worked in the areas of ranked queries, query results exploration, Web search, storage and parsing of XML data, and spatial databases. Dr. Hristidis's work has resulted in more than 40 publications, which have received more than 1000 bibliographic citations according to Google Scholar. His work has been funded by the National Science Foundation.

Dr. Hristidis has served on numerous program committees of conferences including the Institute of Electrical and Electronics Engineers (IEEE) International Conference on Data Engineering, the International Conference on Extending Database Technology, the IEEE International Conference on Data Mining, the Association for Computing Machinery Special Interest Group on Spatial Information, the International Conference on Advances in Geographic Information Systems, and on the review board of the Proceedings of Very Large Databases Endowment. He has also served as cochair of the International Workshop on Ranking in Databases, and as proceedings, finance, and publicity chair of major database conferences.

Contributors

Filipa Canario Almeida
Department of Biostatistics and Medical Informatics
Universidade do Porto
Porto, Portugal

Daniele Apiletti
Dipartimento di Automatica e Informatica
Politecnico di Torino
Torino, Italy

Elena Baralis
Dipartimento di Automatica e Informatica
Politecnico di Torino
Torino, Italy

Donald J. Berndt
Department of Information Systems
University of South Florida
Tampa, Florida

Giulia Bruno
Dipartimento di Automatica e Informatica
Politecnico di Torino
Torino, Italy

Redmond Burke
Miami Children's Hospital
Miami, Florida

Tania Cerquitelli
Dipartimento di Automatica e Informatica
Politecnico di Torino
Torino, Italy

Rong Chen
Department of Biomedical Engineering
Linköping University
Linköping, Sweden

Altamiro Costa-Pereira
Department of Biostatistics and Medical Informatics
Universidade do Porto
Porto, Portugal

Ricardo João Cruz-Correia
Department of Biostatistics and Medical Informatics
Universidade do Porto
Porto, Portugal

Reid Cushman
Department of Medicine
University of Miami School of Medicine
Miami, Florida

Fernando Farfán
School of Computing and Information Science
Florida International University
Miami, Florida

Alessandro Fiori
Dipartimento di Automatica e Informatica
Politecnico di Torino
Torino, Italy

Alberto Freitas
Department of Biostatistics and Medical Informatics
Universidade do Porto
Porto, Portugal

James Gardner
Department of Mathematics and Computer Science
Emory University
Atlanta, Georgia

Vagelis Hristidis
School of Computing and Information Sciences
Florida International University
Miami, Florida

Xiaolei Huang
Department of Computer Science and Engineering
Lehigh University
Bethlehem, Pennsylvania

Pawel Jurczyk
Department of Mathematics and Computer Science
Emory University
Atlanta, Georgia

James J. Lu
Department of Mathematics and Computer Science
Emory University
Atlanta, Georgia

Stephen L. Luther
Department of Veterans Affairs
Tampa, Florida

Pedro Pereira Rodrigues
Department of Biostatistics and Medical Informatics
Universidade do Porto
Porto, Portugal

Monica Chiarini Tremblay
Department of Decision Sciences and Information Systems
Florida International University
Miami, Florida

Gavriil Tsechpenakis
Department of Computer Science
University of Miami
Miami, Florida

Ramakrishna Varadarajan
Department of Computer Sciences
University of Wisconsin–Madison
Madison, Wisconsin

Li Xiong
Department of Mathematics and Computer Science
Emory University
Atlanta, Georgia

1

Overview of XML

Fernando Farfán and Vagelis Hristidis

CONTENTS

1.1 Introduction

XML, which stands for Extensible Markup Language, is a general purpose language that allows the creation of other new languages to be used in several domains. It is flexible, simple, and designed to meet the challenges of large-scale electronic publishing, facilitating the exchange of data among heterogeneous computer systems (particularly over the Internet), while maintaining the capability of being human-readable.

XML uses a combination of notes and special symbols (called "markup") to express information about the data itself. These markups are basically strings of characters called tags, which are put together to delimit the main portions of data, called elements.

XML is extensible because it lets users to define their own tags, element types, and overall document structure. This extensibility has allowed the development of many application languages for a large number of application domains, ranging from Mathematics (MathML [1]), Graphs and Graphics (GraphML [2]; GML [3]; SVG [4]), Finance (FIXML [5]; FinXML [6]; SwiftML [7]), Internet-related languages (RSS [8]; XHTML [9]), to medicine (CDA [10]).

Figure 1.1 shows an example of an XML document. An XML document consists of the following:

> *XML elements*. XML elements are the basic building blocks of XML markup. Lines 5 to 8, for example, correspond to a *name* XML element. The elements may be seen as containers. Each element may have attributes, and may contain other elements, character data, or other types of information. This containment specifies the structure

```
1.  <? xml version="1.0" ?>
2.  <ClinicalDocument>
3.    <id extension="49912" root="2.16.840.1.113883.3.933"/>
4.    <patient>
5.      <name>
6.        <given>Peter</given>
7.        <family>Patient</family>
8.      </name>
9.      <genderCode code="M" codeSystem="2.16.840.1.5.1"/>
10.     <birthTime value="20020924"/>
11.   </patient>
12.   <component>
13.     <StructuredBody>
14.       <component>
15.         <section>
16.           <code code="10160-0" codeSystem="2.16.840.1.113883.6.1"
                codeSystemName="LOINC" />
17.           <title>Medications</title>
18.           <entry>
19.             <Observation>
20.               <code code="84100007" codeSystem="2.16.840.1.113883.6.96"
                    codeSystemName="SNOMED CT" displayName=" medication history"/>
21.               <value xsi:type="CD" code="195967001" codeSystem="2.16.840.1.113883.6.96"
                    codeSystemName="SNOMED CT" displayName="Asthma">
22.                 <originalText>
23.                   <reference value="m1"/>
24.                 </originalText>
25.               </value>
26.             </Observation>
27.           </entry>
28.           <entry>
29.             <Observation>
30.               <code code="84100007" codeSystem="2.16.840.1.113883.6.96"
                    codeSystemName="SNOMED CT" displayName="medication history"/>
31.               <value xsi:type="CD" code="32398004" codeSystem="2.16.840.1.113883.6.96"
                    codeSystemName="SNOMED CT" displayName="Bronchitis">
32.                 <value xsi:type="CD" code="91143003" codeSystem="2.16.840.1.113883.6.96"
                      codeSystemName="SNOMED CT" displayName="Albuterol" />
33.               </value>
34.             </Observation>
35.           </entry>
36.           <entry>
37.             <SubstanceAdministration>
38.               <text>
39.                 <content ID="m1">Theophylline</content>20 mg every other day, alternating
                     with 18 mg every other day. Stop if temperature is above 1103F.
40.               </text>
41.               <consumable>
42.                 <manufacturedProduct>
43.                   <manufacturedLabeledDrug>
44.                       <code code="66493003" codeSystem="2.16.840.1.113883.6.96"
                            codeSystemName="SNOMED CT" displayName="Theophylline"/>
45.                   </manufacturedLabeledDrug>
46.                 </manufacturedProduct>
47.               </consumable>
48.             </SubstanceAdministration>
49.           </entry>
50.         </section>
51.       </component>
52.     </StructuredBody>
53.   </component>
54. </ClinicalDocument>
```

FIGURE 1.1
Sample XML document.

and hierarchy to the document. The *ClinicalDocument* element that starts in line 2, for example, contains all the other XML elements in the document; the elements in lines 6 and 7 contain text data. The element in line 9 includes two attributes, *code* and *codeSystem*, but does not contain any further information; it is called an empty element.

Tags. Each element is delimited with a *start-tag* and an *end-tag*. Line 5 corresponds to the start-tag of the element name, whereas line 8 corresponds to the end-tag of the same element. We can see how the start-tag "opens" the container that is later closed by the end-tag. In the case of empty elements, a pair of *start-tag/end-tag* can be used, or it could be represented by an *empty-element tag* abbreviation, as it is the case in line 9. The attributes are always included in the start-tag, as seen in the element in line 3.

Attributes. Element attributes describe the properties of an element. Each attribute is comprised of a name-value pair. For example, the start-tag in line 9 has two attributes: code="M" and codeSystem="2.16.840.1.5.1". *code* is an attribute name and "M" is its attribute value. Attribute values must be character strings. Note that it is often a design decision whether a piece of information is represented as an attribute or as a subelement.

It is important to understand that XML is not a programming language; hence XML does not do anything by itself. XML is a data representation format.

The syntax (format) of XML is standardized and formally defined by the World Wide Web Consortium (W3C) [11], which is supported by large software vendors as well as the academic community. This is a key reason for the success of XML.

According to the W3C [12], the key characteristics of XML are

- XML is a markup language much like Hypertext Markup Language (HTML).
- XML was designed to carry data, not to display data.
- XML tags are not predefined. You must define your own tags.
- XML is designed to be self-descriptive.
- XML is a W3C Recommendation.

1.1.1 Does XML Have Semantics?

XML has a strict and formally defined syntax, which specifies when a document qualifies to be an XML document. Furthermore, an XML element has a tag that generally specifies the type of the element and some value. For instance, in Figure 1.1, we can tell that "Peter" is the "name" of a "patient." However, all of these factors do not mean that XML has semantics. This is a common misunderstanding. Intuitively, the reason is that a computer does not know what a "patient" is. Furthermore, two persons may use different tag names to denote the same real-life entity, for example, "patient" versus "client" for a hospital database. To add semantics to XML data, we need to define the semantic meaning of the XML tags. One popular means of doing this is by using ontologies (which will be discussed in Chapter 2).

1.1.2 Related Work and Further Readings

The W3C [11] is an international organization devoted to the definition of Web standards. This consortium, which was formed by industry giants, academia, and the general public, creates standards for the World Wide Web. Within these standards and recommendations, W3C has defined markup languages such as Standard Generalized Markup Language (SGML) [13] and XML [14], as well as technologies and query languages around XML, such as the Document Object Model [15] for document parsing, XML Path Language [16] and XML Query Language [17]. Also, application and domain languages based on XML have been defined by the W3C, such as Extensible Hypertext Markup Language [9], Scalable Vector Graphics [4], and the Resource Descriptor Framework (RDF; [18]).

The storage of XML documents has received attention from academia and industry, with several directions being followed. Many independent works have studied new native storage solutions for XML [19], or created native XML databases and storage systems, such as Lore [20], TIMBER [21], Natix [22, 23], and eXist [24]. Another direction exploits the maturity of relational systems to store XML [25]. Some of these works include STORED [26] and those carried out by Florescu and Kossmann [27] and Tatarinov et al. [28]. Moreover, major commercial Relational Database Management Systems (RDBMSs), such as Microsoft SQL Server [29], Oracle [30], and IBM DB2 [31], provide support to store and query XML data.

In addition, XML schema has been considered as an adequate means to close the gap between relational databases and XML. Some works exploit XML schema to create mappings from XML to RDBMSs [32], or to represent relational data as XML [33, 34].

Several query languages for XML have been developed by W3C, such as XPath [16] and XQuery [17]. A large amount of scholarly work has been devoted to optimizing the processing of XPath and XQuery queries. Works on optimizing XPath query processing include BLAS [35], the Natix project [23, 36], and the work done by Barton et al. [37]. Similarly, XQuery process optimization has been addressed by May et al. [38] (Natix), Zhang et al. [39] (Rainbow), and Liu et al. [40].

Another popular topic in XML research is the study and optimization of XML parsing, which especially considers tree-based representations of XML documents. Nicola and John [41] have identified the XML parsing process as a bottleneck to enterprise applications. Their study compares XML parsing in several application domains to similar applications that use relational databases as their backend. Operations such as shredding XML documents into relational entities, XPath expression evaluation, and XSL Transformations [42, 43] processing are often determined by the performance of the underlying XML parser [41], limiting the massive embracement of native XML databases into large-scale enterprise applications.

Noga et al. [44] presented the idea of *lazy parsing*. The virtual document tree can potentially be stored on disk to avoid the preparsing stage; however, the virtual document tree has to still be read from disk. Schott and Noga [45] applied these ideas to XSL Transformations. Kenji and Hiroyuki [46] have also proposed a lazy XML parsing technique applied to XSL Transformation stylesheets, constructing a pruned XML tree by statically identifying the nodes that will be referred to during the transformation process. We extended these ideas and developed a double-lazy parser [47, 48], which treats both phases of the DOM processing (*preprocessing* and *progressive parsing*) in a lazy fashion.

Lu et al. [49] presented a parallel approach to XML parsing, which initially preparses the document to extract the structure of the XML tree, and then perform a parallel full parse. This parallel parsing is achieved by assigning the parsing of each segment of the document to a different thread that can exploit the multicore capabilities of contemporary CPUs. Their preparsing phase is more relaxed than the one proposed by Noga et al. [44] and that we use throughout our work; this relaxed preparsing only extracts the tree shape without additional information, and is used to decide where to partition the tree to assign the parsing subtasks to the threads. This partitioning scheme differs from ours since it is performed after the preparsing phase is executed, whereas ours is performed a priori, with the objective of optimizing such preparsing stage.

There have been efforts in developing XML pull parsers [50] for both Simple API for XML (SAX) and DOM interfaces. Also, a new API [51] has been presented that is built just one level on top of the XML tokenizer, hence claiming to be the simplest, quickest, and most efficient engine for processing XML.

Another important direction related to XML is the definition of languages that represent semantics such as the RDF [18]. RDF provides a technique for describing resources on the Web. Hence, this development has spanned topics such as generation of metadata [52], storage and querying of RDF schemas [53], and use of RDF for network infrastructure [54].

1.2 XML versus HTML

Although both XML and HTML may look alike, there exist important differences between them. Both XML and HTML are derived from SGML. SGML is an older and more complex markup language, codified as an international standard by the International Organization for Standardization (ISO) as ISO 8879. HTML is, indeed, an application of SGML, and a new version of HTML

4, called XHTML, is an application of XML. Although SGML, HTML, XML, and XHTML are all markup languages, only SGML and XML can be considered metalanguages—they can be used to create new languages (HTML is a single and predefined markup language).

Figure 1.2 presents a sample HTML document showing information similar to that in Figure 1.1. Although the document looks similar to that in Figure 1.1, we observe that the set of tags used here is different: the *head* element in line 2 contains general information (metainformation) about the document, and the *body* element in line 6 contains all the contents in the document. The rest of the elements in the HTML document are presentation-oriented, and hence we have the elements *h1*, *h2*, and *h3* (lines 7, 8, and 28, respectively) that

```
1. <html xmlns="http://www.w3.org/1999/xhtml">
2. <head>
3.   <meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
4.   <title>Clinical Document</title>
5. </head>
6. <body>
7.   <h1>Clinical Document</h1>
8.   <h2>Patient</h2>
9.   <table border="1" cellspacing="0" cellpadding="0">
10.    <tr>
11.      <td width="120">Given Name</td>
12.      <td width="120"><strong>Peter</strong></td>
13.    </tr>
14.    <tr>
15.      <td>Family Name</td>
16.      <td><strong>Patient</strong></td>
17.    </tr>
18.    <tr>
19.      <td>Gender</td>
20.      <td><strong>MALE</strong></td>
21.    </tr>
22.    <tr>
23.      <td>Birth Time</td>
24.      <td><strong>09-24-2002</strong></td>
25.    </tr>
26.  </table>
27.  <h2>Clinical Encounter</h2>
28.  <h3>Medications</h3>
29.  <table width="640" border="1" cellspacing="0" cellpadding="0">
30.    <tr>
31.      <th>Illness</th>
32.      <th>Medication</th>
33.    </tr>
34.    <tr>
35.      <td>Asthma</td>
36.      <td>Theophylline</td>
37.    </tr>
38.    <tr>
39.      <td>Bronchitis</td>
40.      <td>Albuterol</td>
41.    </tr>
42.</table>
43.</body>
44.</html>
```

FIGURE 1.2
Sample HTML document.

correspond to different levels of header formatting in the document. HTML tags are predefined and have specific presentation meaning, whereas XML tags are defined by the user and have no specific presentation meaning.

We can observe how even when the captured information is similar, the HTML document does not describe any logical structure or semantics about what the document is about, whereas the XML document richly describes the data it contains. The tags in the XML document directly correspond to concepts in the domain of electronic health records.

It is important to say that XML is not a replacement for HTML. Both were designed with different goals: HTML's goal is to display data and it focuses on how data looks; XML's goal is to transport and store data, focusing on the data content and not on its presentation.

1.3 XML versus Relational Data Model

In database management, the *relational model* is, nowadays, the dominant model for commercial data processing applications. Originally proposed by E. F. Codd in 1970 [55], this model specifies that the data are stored in the database as a collection of Tables (formally, mathematical relations). Each Table (relation) can be seen as a set of records (tuples) [56]. The relational model uses a schema to model the data in terms of table name, name of each field, and type of each field.

For example, we can create a relation to store the information about patients in a hospital as follows:

> Patients(*patient_id*: integer, *first_name*: string, *last_name*: string, *date_of_birth*: date, *gender*: string)

This schema specifies that each tuple in the Patients relation has five fields, whose names and types are explicitly indicated. An example instance (content at a specific time) of this relation is depicted in table 1.1.

One of the advantages of XML over other data models is its ability to exchange data between heterogeneous platforms. In contrast to proprietary systems and

TABLE 1.1

Sample Instance of a *Patients* Relation

Patient_Id	First_Name	Last_Name	Date_of_Birth	Gender
60135	Jacqueline	Jones	2002-02-12	Female
76638	Andrew	Smith	2003-05-25	Male
76639	Jason	Smith	2003-06-18	Male
76640	Melinda	Galvin	2004-01-12	Female

formats, whose data are incompatible among others, XML data are stored in plain text in a standardized format, which provides software and hardware independence in storing and sharing data. Moreover, as discussed previously, XML combines the data schema and the instance of the data in the same file. This self-containment also makes XML more appropriate for data exchange than other data models such as the relational model. The structure of XML files also makes it a convenient means to represent complex hierarchical data.

1.4 XML Syntax

We now present the basic syntactic elements of XML. For a complete and detailed overview, see World Wide Web Consortium [14], Birbeck et al. [57], and W3Schools [12]. In addition, an annotated version of the first edition of the XML Recommendation is given by Bray [58].

As shown in Figures 1.1 and 1.2, XML elements and their content look very similar to those of HTML. But looking further, it becomes obvious that XML documents provide more information within the document, since the element types (tags) give additional information about the data.

All XML documents that follow certain basic rules specified in the XML 1.0 Recommendation [14] are known as *well formed* [57]. To be well formed, an XML document must follow more than 100 rules; however, most of them are trivial. To summarize, an XML document is well formed if it satisfies the following conditions:

- Every start-tag has a matching end-tag. Moreover, all elements must be properly nested (no overlapping in element definitions), and there are no instances of multiple attributes with the same name for one element.
- It conforms to all rules of XML specification, meaning that start-tags and end-tags are always matched, there is no overlapping in elements, attributes have unique names, and markup reserved characters are properly escaped.
- It has a unique root element, with all the elements forming a hierarchical tree under the root element.

Tree representation of XML. A special exception to the hierarchical tree property cited in the last bulleted item may be achieved when internal links are introduced to the file. XML has mechanisms to introduce internal and external pointers. For example, an ID/IDREF attribute combination can be used to establish a link from one element to another. For example, the document in Figure 1.1 includes an ID/IDREF link: the *content* element in line 39 has an ID attribute that is referenced by the IDREF attribute in line 23. In this case, the document cannot be represented as a hierarchical tree, but becomes a graph.

As discussed, the XML document can be represented as a hierarchical tree. In this model, every XML element is represented as a node, and the parent-child relationships between elements are captured as edges. We call these *containment edges*. The use of ID/IDREF attributes creates an additional edge between elements that are not directly connected by a parent–child relationship. We call these edges *ID/IRDEF edges*. This introduces a new edge into the representation and transforms the tree into a graph, since a cycle is created within the graph. ID-IDREF edges dramatically complicate the processing of XML data given that many algorithmic problems, such as shortest path and proximity search, become very expensive if we move from trees to graphs. Figure 1.3 shows the tree representation for the document in Figure 1.1. Note how we have an ID/IDREF edge between the *content* element (line 39) and the *reference* element (line 23).

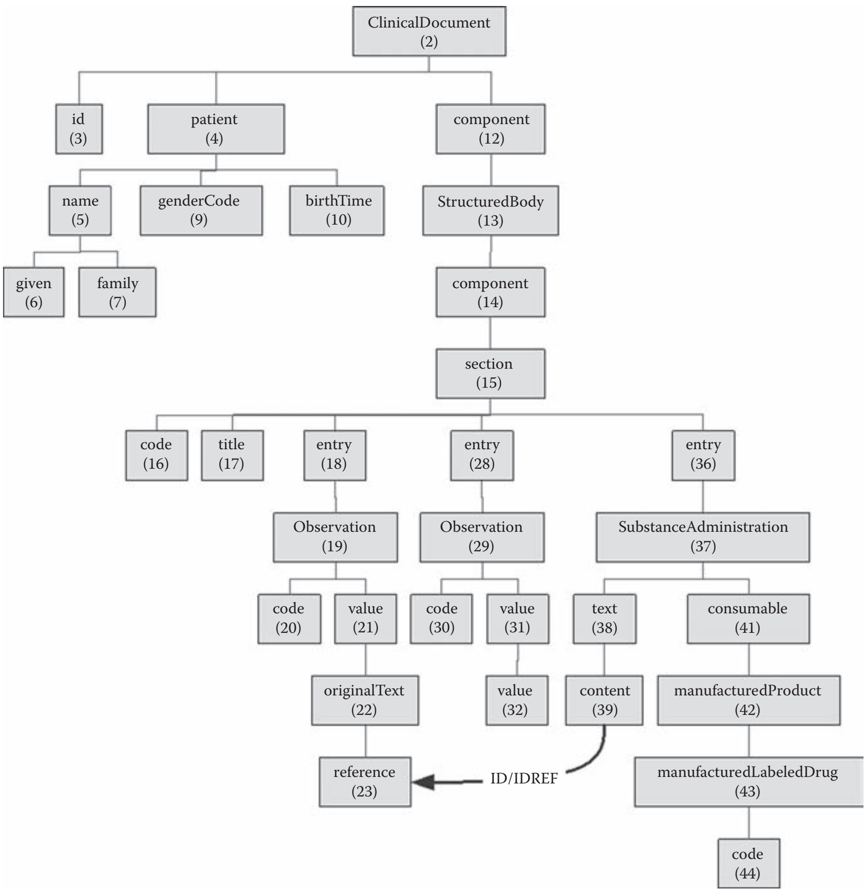

FIGURE 1.3
Tree representation of the XML document in Figure 1.1.

1.5 XML Schema

In general, we can think of data schema as the detailed description of rules and constraints that data instances have to comply in order to be valid. In addition to being well formed, an XML document can, in occasions, meet certain further validity rules. In this case, the document is said to be *valid*. A valid XML document is a well-formed document that also complies with a Document Type Definition (DTD; [59]) file or XML Schema file [60]. Note that the validity of an XML document can only be checked against an XML schema.

DTD was the first method used to specify the schema of XML documents. A DTD file specifies a set of rules that define how the data in the XML document should be structured, by defining a list of valid elements and attributes, what attributes can describe each element, and the nesting of the elements.

Figure 1.4 shows a fragment of the DTD document that specifies the validity of the XML document shown in Figure 1.1.

```
<!ELEMENT ClinicalDocument          (id, patient, component) >
<!ELEMENT id                        (#PCDATA) >
<!ATTLIST id                        extension   CDATA #REQUIRED
                                    root  CDATA #REQUIRED >
<!ELEMENT patient                   (name, genderCode, birthTime) >
<!ELEMENT component                 (StructuredBody, section) >
<!ELEMENT name                      (given, family) >
<!ELEMENT genderCode                EMPTY >
<!ATTLIST genderCode                code  CDATA #REQUIRED
                                    codeSystem  CDATA #REQUIRED >
<!ELEMENT birthTime                 EMPTY >
<!ATTLIST birthTime                 value CDATA #REQUIRED >
<!ELEMENT StructuredBody            (component) >
<!ELEMENT section                   (code, title, entry) >
<!ELEMENT given                     (#PCDATA) >
<!ELEMENT family                    (#PCDATA) >
<!ELEMENT code                      EMPTY >
<!ATTLIST code                      code  CDATA #REQUIRED
                                    codeSystem  CDATA #REQUIRED
                                    codeSystemName    CDATA #REQUIRED >
<!ELEMENT title                     (#PCDATA) >
<!ELEMENT entry                     (Observation, SubstanceAdministration) >
<!ELEMENT Observation               (code, value) >
<!ELEMENT SubstanceAdministration   (text, consumable) >
<!ELEMENT value                     (value, originalText, #PCDATA) >
<!ATTLIST value                     type  CDATA #REQUIRED
                                    code  CDATA #REQUIRED
                                    codeSystem  CDATA #REQUIRED
                                    codeSystemName    CDATA #REQUIRED
                                    displayName CDATA #REQUIRED >
<!ELEMENT text                      (content, #PCDATA) >
<!ELEMENT consumable                (manufacturedProduct) >
<!ELEMENT originalText              (reference) >
<!ELEMENT content                   (#PCDATA) >
<!ATTLIST content                   ID    ID #REQUIRED
<!ELEMENT manufacturedProduct       (manufacturedLabeledDrug) >
<!ELEMENT reference                 (EMPTY) >
<!ATTLIST reference                 IDREF value #REQUIRED
<!ELEMENT manufacturedLabeledDrug   (code) >
```

FIGURE 1.4
DTD specification for the XML document in Figure 1.1.

A more recent approach to specifying the structure of XML documents is XML Schema. This is a W3C Recommendation aimed to provide a more powerful and flexible language by which to define the XML document structure. XML Schema is more expressive than DTD, allowing new features such as richer specification of data types (e.g., nonNegativeInteger vs. PCDATA), namespaces and number, and order of child elements. XML Schemas are themselves XML documents, which is another advantage since there is no need to learn a new language to specify the structure of the document. XML Schema provides an object-oriented approach to defining the data schema.

For a detailed description of DTDs and XML Schema, see Birbeck et al. [57].

1.6 XML Parsing

In computer science and linguistics, the process of analyzing a sequence of tokens to determine the grammatical structure with respect to a given formal grammar is called *syntactic analysis* or *parsing*. In the case of XML, this means that the XML file is analyzed and the sequence of tokens is checked to validate that all the constraints noted in the previous section about XML syntax are satisfied. As the document is parsed, the data contained in the document is made available to the application that is parsing it [61].

The XML 1.0 Recommendation [14] defines two levels of parsing:

1. *Nonvalidating* makes sure that the document is well formed, but does not require an external schema to be present.
2. *Validating* Ensures that the document is both well formed and valid, according to a DTD or XML Schema.

Another distinction between parsers is the implementation that they use to process the data:

- *Tree-based parsers*. This class of parsers creates an in-memory representation of the XML tree. This allows user-friendly navigation of the tree, but may require large amounts of memory to represent the tree.
- *Event-driven parsers*. The data are processed sequentially, and the data component is handled one at a time. The memory requirements are minimal, but the interface may not be as user-friendly.

Two popular representatives of these two parsing implementations are the Document Object Model [15, 62] and the SAX [63], respectively. As with many other solutions to real-world problems, the vast number of possibilities and requirements make these two approaches necessary and compatible. Every different scenario can benefit from these implementations or a combination of both. In general, DOM is easier to program with, whereas SAX is more efficient and scalable.

1.7 XML Querying

Another mechanism of accessing XML data is to use query languages. Several query languages have been proposed, again covering a vast range of requirements. Two of the most popular XML query languages are XPath [16] and XQuery [17].

XPath is a language for selecting nodes from an XML document, and is based on the tree representation of the XML document, providing the ability to navigate the XML tree. XPath also provides a series of functions for manipulating strings, numbers, Booleans, and node sets.

XQuery, on the other hand, is a query language designed to access an XML document or a collection of XML documents in a manner similar to what a relational database does with relations. XQuery tries to exploit the flexibility and hierarchical structure of XML documents. By defining its own data model and algebra, XQuery uses path expressions (based on XPath), conditional expressions, and complex constructs, recursion, and other mechanisms to deliver a powerful, yet easy-to-learn query language. XQuery is generally more complex than SQL, which is used for querying relational databases, and hence it has so far not been widely accepted in practice.

1.8 XML Advantages and Disadvantages

Now that we have presented XML, we can summarize the advantages and disadvantages of this data model.

One of the advantages for XML that has majorly contributed to its popularity is its orientation to data exchange. XML has been designed to be platform independent by storing its contents as data files. This reduces the complexity of exchanging data, by allowing XML documents to be shared among incompatible platforms, making it resistant to software or hardware updates.

XML is also defined to be self-contained: both metadata and data are included in the XML document. Hence, there is no need to store any additional resources to interpret the data.

XML is standardized. It was created as a W3C Recommendation, backed up by the industry giants and academic researchers, and accepted by the community in general. This has also contributed to its quick popularization.

XML can represent complex, nested data in scenarios where representing the same on relational databases would be extremely cumbersome.

On the other hand, the expensive processing and querying of XML documents is also its major drawback. The need for large amounts of memory and processing power to parse and query XML data makes it unfeasible for some configurations.

Also, to date there is still no popular and efficient XML-native database systems. Instead, all the major RDBMS vendors, such as Oracle, IBM DB2, and Microsoft SQL Server, incorporate XML storage modules.

Moreover, in many cases, the complexity and overhead of XML makes it simply suboptimal for simple and small environments.

1.9 Chapter Summary

In this chapter, we have introduced the XML, which has revolutionized the manner in which data are stored, exchanged, and processed in distributed systems. We have reviewed the XML syntax, data model, and semantic aspects of its definition.

We reviewed some related work, both in industry and academia, which are based in XML or extend XML in new and more powerful directions. We also compared XML and HTML, outlining the differences in approach and syntax of these two languages.

We talked about XML storage, parsing, and querying, and based on this we identified the advantages and disadvantages of this metalanguage, which has become the *lingua franca* of the World Wide Web.

References

1. World Wide Web Consortium. W3C Math Home. 2008. http://www.w3.org/Math/(Accessed Aug. 2, 2008).
2. GraphML Working Group. 2007. GraphML. http://graphml.graphdrawing.org/(Accessed Aug. 2, 2008).
3. Open Geospatial Consortium. 2008. Geography Markup Language. http://www.opengeospatial.org/standards/gml, (Accessed Sept. 1, 2009).
4. World Wide Web Consortium. Scalable Vector Graphics. 2008. http://www.w3.org/Graphics/SVG/(Accessed Aug. 2, 2008).
5. FIX Protocol. 2008. FIXML Resources for FIX 4.4 Specification. http://www.fixprotocol.org/specifications/fix4.4fixml, (Accessed Sept. 1, 2009).
6. OASIS XML Cover Pages. 1999. FinXML—The Digital Language for Capital Markets. http://xml.coverpages.org/finXML.html, (Accessed Sept. 1, 2009).

7. OASIS XML Cover Pages. 2001. SwiftML for Business Messages. http://xml.coverpages.org/swiftML.html, (Accessed Sept. 1, 2009).
8. Wikipedia. RSS Web feed format. 2008. http://en.wikipedia.org/wiki/RSS_(file_format), (Accessed Sept. 1, 2009).
9. World Wide Web Consortium. 2002. XHTML 1.0 The Extensible HyperText Markup Language (Second Edition). http://www.w3.org/TR/xhtml1/ (accessed Aug. 2, 2008).
10. Dolin, R. H., Alschuler, L., Boyer, S., Beebe, C., Behlen, F. M., Biron, P. V., and Shabo, A. 2006. HL7 Clinical document architecture, Release 2. *International Journal of the American Medical Informatics Association* 13(1):30–39.
11. World Wide Web Consortium. W3C Homepage. 2008. http://www.w3.org/ (Accessed Aug. 4, 2008).
12. W3 Schools. 2008. Extensible Markup Language. http://www.w3schools.com/XML (Accessed Aug. 2, 2008).
13. World Wide Web Consortium. 2004. Overview of SGML Resources. http://www.w3.org/MarkUp/SGML/(Accessed Aug. 2, 2008).
14. World Wide Web Consortium. 2006. Extensible Markup Language (XML) 1.0 (Fourth Edition). http://www.w3.org/TR/REC-xml/(Accessed Aug. 2, 2008).
15. World Wide Web Consortium. Document Object Model (DOM). http://www.w3.org/DOM/(Accessed Aug. 4, 2008).
16. World Wide Web Consortium. 1999. XML Path Language (XPath). http://www.w3.org/TR/xpath (Accessed Aug. 4, 2008).
17. World Wide Web Consortium. XQuery 1.0: An XML Query Language. 2007. http://www.w3.org/TR/xquery/(Accessed Aug. 4, 2008).
18. World Wide Web Consortium. 2004. Resource Descriptor Framework (RDF). http://www.w3.org/RDF/(Accessed Aug. 5, 2008).
19. Bhadkamkar, M., Farfán, F., Hristidis, V., and Rangaswami, R. 2009. Storing semi-structured data on disk drives. *ACM Transactions on Storage* 5(2):1–35.
20. McHugh, J., Abiteboul, S., Goldman, R., Quass, D., and Widom, J. 1997. Lore: A database management system for semistructured data. *ACM SIGMOD Record* 26(3):54–66.
21. Jagadish, H. V., Al-Khalifa, S., Chapman, A., et al. 2002. TIMBER: a native XML database. *VLDB Journal* 11(4):274–291.
22. Fiebig, T., Helmer, S., Kanne, C.-C., et al. 2002. Natix: a technology overview, revised papers from the NODe 2002 Web and Database-Related Workshops on Web, Web-Services, and Database Systems, pp. 12–33.
23. Data ex machina. 2008. The NATIX XML Repository. http://www.data-ex-machina.de/natix.html, (Accessed Sept. 1, 2009).
24. Meier, W. 2002. eXist: An open source native XML database. In *Web, Web-Services, and Database Systems*, E. R. Chaudri, M. Jeckle, and R. Unland, Eds., Springer LNCS Series.
25. Shanmugasundaram, J., Tufte, K., He, G., Zhang, C., Dewitt, D., and Naughton, J. 1999. Relational databases for querying XML documents: limitations and opportunities. In *Proceedings of the 25th VLDB Conference*.
26. Deutsch, A., Fernandez, M. F., and Suciu, D. 1999. Storing semistructured data with STORED. In *Proceedings of the ACM International Conference on Management of Data (SIGMOD)*.

27. Florescu, D., and Kossmann, D. 1999. Storing and querying XML data using an RDMBS. *IEEE Data Engineering Bulletin* 22:27–34.
28. Tatarinov, I., Beyer, K., and Shanmugasundaram, J. 2002. Storing and querying ordered XML using a relational database system. In *Proceedings of the ACM SIGMOD*.
29. Microsoft Developer Network. 2005. XML Support in Microsoft SQL Server 2005. http://msdn.microsoft.com/en-us/library/ms345117.aspx, (Accessed Sept. 1, 2009).
30. Oracle Corporation. XML Technology Center. http://www.oracle.com/technology/tech/xml/index.html, (Accessed Sept. 1, 2009).
31. IBM Corporation. 2008. pureXML. http://www-306.ibm.com/software/data/db2/xml/, (Accessed Sept. 1, 2009).
32. Bohannon, P., Freire, J., Roy, P., and Simeon, J. 2002. From XML schema to relations: a cost-based approach to XML storage. In *Proceedings of International Conference on Data Engineering*.
33. Carey, M. J., Florescu, D., Ives, Z. G. et al. 2000. XPERANTO: publishing object-relational data as XML. In International Workshop on the Web and Databases.
34. Lee, D., Mani, M., Chiu, F., and Chu, W. W. 2001. Nesting-based relational-to-XML schema translation. In *International Workshop on the Web and Databases WebDB*.
35. Chen, Y., Davidson, S. B., and Zheng. Y. 2004. BLAS: an efficient XPath processing system. In *Proceedings of ACM SIGMOD*.
36. Brantner, M., Helmer, S., Kanne, C.-C., and Moerkotte, G. 2005. Full-fledged algebraic XPath processing in Natix. In *Proceedings of the 21st International Conference on Data Engineering ICDE*.
37. Barton, C., Charles, P., Goyal, D., Raghavachari, M., and Fontoura, M. 2003. Streaming XPath processing with forward and backward axes. In *Proceedings of the International Conference on Data Engineering ICDE*.
38. May, N., Helmer, S., Kanne, C.-C., and Moerkotte, G. 2004. XQuery processing in Natix with an emphasis on join ordering. In *Proceedings of International Workshop on XQuery Implementation, Experience and Perspectives XIME-P*.
39. Zhang, X., Mulchandani, M., Christ, S., Murphy, B., and Rundensteiner, E. A. 2002. Rainbow: mapping-driven XQuery processing system. In *Proceedings of the ACM SIGMOD*.
40. Liu, Z. H., Krishnaprasad, M., and Arora, V. 2005. Native XQuery processing in oracle XMLDB. In *Proceedings of ACM SIGMOD*.
41. Nicola, M., and John, J. 2003. XML parsing: a threat to database performance. In *Proceedings of the 12th Conference on Information and Knowledge Management*.
42. World Wide Web Consortium. Extensible Stylesheet Language (XSL) 2007. http://www.w3.org/TR/xsl/(Accessed Aug. 5, 2008).
43. World Wide Web Consortium. XSL Transformations. 2007. http://www.w3.org/TR/xslt (Accessed Aug. 5, 2008).
44. Noga, M., Schott, S., and Löwe, W.. Lazy XML processing. In *ACM DocEng*, 2002.
45. Schott, S., and Noga, M. 2003. Lazy XSL transformations. In *ACM DocEng*.
46. Kenji, M., and Hiroyuki, S. 2005. Static optimization of XSLT Stylesheets: template instantiation optimization and lazy XML parsing. In *ACM DocEng*.

47. Farfán, F., Hristidis, V., and Rangaswami, R. 2007. Beyond lazy XML parsing. In *Proceedings of DEXA*.
48. Farfán, F., Hristidis, V., and Rangaswami, R. 2009. 2LP: A double-lazy XML parser. *Information Systems* 34:145–163.
49. Lu, W., Chiu, K., and Pan, Y. 2006. A parallel approach to XML parsing. In *IEEE/ACM International Conference on Grid Computing Grid*.
50. XML Pull Parsing. 2006. http://www.xmlpull.org/index.shtml (Accessed Aug. 5, 2008).
51. XML Pull Parser (XPP). 2004. http://www.extreme.indiana.edu/xgws/xsoap/xpp/(Accessed Aug. 5, 2008).
52. Jenkins, C., Jackson, M., Burden, P., and Wallis, J. 1999. Automatic RDF metadata generation for resource discovery. *International Journal of Computer and Telecommunications Networking* 31(11–16):1305–1320.
53. Broekstra, J., Kampman, A., and van Harmelen, F. 2002. Sesame: a generic architecture for storing and querying RDF and RDF Schema. In *Proceedings of the Semantic Web Conference ISWC*.
54. Nejdl, W., Wolf, B., Qu, C., et al. 2002. EDUTELLA: a P2P networking infrastructure based on RDF. In *Proceedings of the International WWW Conference*.
55. Codd, E. F. 1970. A relational model of data for large shared data banks. *Communications of the ACM* 13(6):377–387.
56. Ramakrishnan, R., and Gehrke, J. 2000. *Database Management Systems*, 3rd ed. New York, NY: McGraw-Hill Higher Education.
57. Birbeck, M., Duckett, J., and Gudmundsson, O. G. et al. 2001. *Professional XML*, 2nd edn. Birmingham, England: Wrox Press Inc.
58. Bray, T. 1998. Introduction to the Annotated XML Specification: Extensible Markup Language (XML) 1.0. http://www.xml.com/axml/testaxml.htm, (Accessed Sept. 1, 2009).
59. W3Schools. Introduction to DTD. 2003. http://www.w3schools.com/DTD/dtd_intro.asp (Accessed Aug. 4, 2008).
61. McLaughlin, B., and Loukides, M. 2001. *Java and XML. (O'Reilly Java Tools)*. O'Reilly & Associates.
60. World Wide Web Consortium. 2006. XML Schema. http://www.w3.org/XML/Schema (Accessed Aug. 4, 2008).
62. W3 Schools. 2008. XML DOM Tutorial. 2008. http://www.w3schools.com/dom/default.asp (Accessed Aug. 4, 2008).
63. Official SAX Website. 2004. http://www.saxproject.org/about.html, (Accessed Sept. 1, 2009).

References

1 Chapter 1. Overview of XML

1. World Wide Web Consortium. W3C Math Home. 2008. http://www.w3.org/ Math/(Accessed Aug. 2, 2008).

2. GraphML Working Group. 2007. GraphML. http://graphml.graphdrawing. org/(Accessed Aug. 2, 2008).

3. Open Geospatial Consortium. 2008. Geography Markup Language. http:// www.opengeospatial.org/standards/gml, (Accessed Sept. 1, 2009).

4. World Wide Web Consortium. Scalable Vector Graphics. 2008. http://www. w3.org/Graphics/SVG/(Accessed Aug. 2, 2008).

5. FIX Protocol. 2008. FIXML Resources for FIX 4.4 Specification. http://www.fixprotocol.org/specifications/fix4.4fixml, (Accessed Sept. 1, 2009).

6. OASIS XML Cover Pages. 1999. FinXML—The Digital Language for Capital Markets. http://xml.coverpages.org/finXML.html, (Accessed Sept. 1, 2009).

7. OASIS XML Cover Pages. 2001. SwiftML for Business Messages. http://xml. coverpages.org/swiftML.html, (Accessed Sept. 1, 2009).

8. Wikipedia. RSS Web feed format. 2008. http://en.wikipedia.org/wiki/RSS_ (file_format), (Accessed Sept. 1, 2009).

9. World Wide Web Consortium. 2002. XHTML 1.0 The Extensible HyperText Markup Language (Second Edition). http://www.w3.org/TR/xhtml1/ (accessed Aug. 2, 2008).

10. Dolin, R. H., Alschuler, L., Boyer, S., Beebe, C., Behlen, F. M., Biron, P. V., and Shabo, A. 2006. HL7 Clinical document architecture, Release 2. International Journal of the American Medical Informatics Association 13(1):30-39.

11. World Wide Web Consortium. W3C Homepage. 2008. http://www.w3.org/ (Accessed Aug. 4, 2008).

12. W3 Schools. 2008. Extensible Markup Language. http://www.w3schools.com/ XML (Accessed Aug. 2, 2008).

13. World Wide Web Consortium. 2004. Overview of SGML Resources. http:// www.w3.org/MarkUp/SGML/(Accessed Aug. 2, 2008).

14. World Wide Web Consortium. 2006. Extensible Markup Language (XML) 1.0 (Fourth Edition). http://www.w3.org/TR/REC-xml/(Accessed Aug. 2, 2008).

15. World Wide Web Consortium. Document Object Model (DOM). http://www. w3.org/DOM/(Accessed Aug. 4, 2008).

16. World Wide Web Consortium. 1999. XML Path Language (XPath). http://www. w3.org/TR/xpath (Accessed Aug. 4, 2008).

17. World Wide Web Consortium. XQuery 1.0: An XML Query Language. 2007. http://www.w3.org/TR/xquery/(Accessed Aug. 4, 2008).

18. World Wide Web Consortium. 2004. Resource Descriptor Framework (RDF). http://www.w3.org/RDF/(Accessed Aug. 5, 2008).

19. Bhadkamkar, M., Farfán, F., Hristidis, V., and Rangaswami, R. 2009. Storing semi-structured data on disk drives. ACM Transactions on Storage 5(2):1–35.

20. McHugh, J., Abiteboul, S., Goldman, R., Quass, D., and Widom, J. 1997. Lore: A database management system for semistructured data. ACM SIGMOD Record 26(3):54–66.

21. Jagadish, H. V., Al-Khalifa, S., Chapman, A., et al. 2002. TIMBER: a native XML database. VLDB Journal 11(4):274–291.

22. Fiebig, T., Helmer, S., Kanne, C.-C., et al. 2002. Natix: a technology overview, revised papers from the NODe 2002 Web and Database-Related Workshops on Web, Web-Services, and Database Systems, pp. 12–33.

23. Data ex machina. 2008. The NATIX XML Repository. http://www.data- ex- machina.de/natix.html, (Accessed Sept. 1, 2009).

24. Meier, W. 2002. eXist: An open source native XML database. In Web, Web-Services, and Database Systems, E. R. Chaudri, M. Jeckle, and R. Unland, Eds., Springer LNCS Series.

25. Shanmugasundaram, J., Tufte, K., He, G., Zhang, C., Dewitt, D., and Naughton, J. 1999. Relational databases for querying XML documents: limitations and opportunities. In Proceedings of the 25th VLDB Conference.

26. Deutsch, A., Fernandez, M. F., and Suciu, D. 1999. Storing semistructured data with STORED. In Proceedings of the ACM International Conference on Management of Data (SIGMOD).

27. Florescu, D., and Kossmann, D. 1999. Storing and querying XML data using an RDMBS. IEEE Data Engineering Bulletin 22:27–34.

28. Tatarinov, I., Beyer, K., and Shanmugasundaram, J. 2002. Storing and querying ordered XML using a relational database system. In Proceedings of the ACM SIGMOD.

29. Microsoft Developer Network. 2005. XML Support in Microsoft SQL Server 2005. http://msdn.microsoft.com/en-us/library/ms345117.aspx, (Accessed Sept. 1, 2009).

30. Oracle Corporation. XML Technology Center. http://www.oracle.com/technology/tech/xml/index.html, (Accessed Sept. 1, 2009).

31. IBM Corporation. 2008. pureXML. http://www-306.ibm.com/software/data/ db2/xml/, (Accessed Sept. 1, 2009).

32. Bohannon, P., Freire, J., Roy, P., and Simeon, J. 2002. From XML schema to relations: a cost-based approach to XML storage. In Proceedings of International Conference on Data Engineering.

33. Carey, M. J., Florescu, D., Ives, Z. G. et al. 2000. XPERANTO: publishing object-relational data as XML. In International Workshop on the Web and Databases.

34. Lee, D., Mani, M., Chiu, F., and Chu, W. W. 2001. Nesting-based relational-toXML schema translation. In International Workshop on the Web and Databases WebDB.

35. Chen, Y., Davidson, S. B., and Zheng. Y. 2004. BLAS: an ef�cient XPath processing system. In Proceedings of ACM SIGMOD.

36. Brantner, M., Helmer, S., Kanne, C.-C., and Moerkotte, G. 2005. Full-�edged algebraic XPath processing in Natix.

In Proceedings of the 21st International Conference on Data Engineering ICDE.

37. Barton, C., Charles, P., Goyal, D., Raghavachari, M., and Fontoura, M. 2003. Streaming XPath processing with forward and backward axes. In Proceedings of the International Conference on Data Engineering ICDE.

38. May, N., Helmer, S., Kanne, C.-C., and Moerkotte, G. 2004. XQuery processing in Natix with an emphasis on join ordering. In Proceedings of International Workshop on XQuery Implementation, Experience and Perspectives XIME-P.

39. Zhang, X., Mulchandani, M., Christ, S., Murphy, B., and Rundensteiner, E. A. 2002. Rainbow: mapping-driven XQuery processing system. In Proceedings of the ACM SIGMOD.

40. Liu, Z. H., Krishnaprasad, M., and Arora, V. 2005. Native XQuery processing in oracle XMLDB. In Proceedings of ACM SIGMOD.

41. Nicola, M., and John, J. 2003. XML parsing: a threat to database performance. In Proceedings of the 12th Conference on Information and Knowledge Management.

42. World Wide Web Consortium. Extensible Stylesheet Language (XSL) 2007. http://www.w3.org/TR/xsl/(Accessed Aug. 5, 2008).

43. World Wide Web Consortium. XSL Transformations. 2007. http://www.w3.org/ TR/xslt (Accessed Aug. 5, 2008).

44. Noga, M., Schott, S., and Löwe, W.. Lazy XML processing. In ACM DocEng, 2002.

45. Schott, S., and Noga, M. 2003. Lazy XSL transformations. In ACM DocEng.

46. Kenji, M., and Hiroyuki, S. 2005. Static optimization of XSLT Stylesheets: template instantiation optimization and lazy XML parsing. In ACM DocEng.

47. Farfán, F., Hristidis, V., and Rangaswami, R. 2007. Beyond lazy XML parsing. In Proceedings of DEXA.

48. Farfán, F., Hristidis, V., and Rangaswami, R. 2009. 2LP: A double-lazy XML parser. Information Systems 34:145–163.

49. Lu, W., Chiu, K., and Pan, Y. 2006. A parallel approach

to XML parsing. In IEEE/ ACM International Conference on Grid Computing Grid.

50. XML Pull Parsing. 2006. http://www.xmlpull.org/index.shtml (Accessed Aug. 5, 2008).

51. XML Pull Parser (XPP). 2004. http://www.extreme.indiana.edu/xgws/xsoap/ xpp/(Accessed Aug. 5, 2008).

52. Jenkins, C., Jackson, M., Burden, P., and Wallis, J. 1999. Automatic RDF metadata generation for resource discovery. International Journal of Computer and Telecommunications Networking 31(11-16):1305-1320.

53. Broekstra, J., Kampman, A., and van Harmelen, F. 2002. Sesame: a generic architecture for storing and querying RDF and RDF Schema. In Proceedings of the Semantic Web Conference ISWC.

54. Nejdl, W., Wolf, B., Qu, C., et al. 2002. EDUTELLA: a P2P networking infrastructure based on RDF. In Proceedings of the International WWW Conference.

55. Codd, E. F. 1970. A relational model of data for large shared data banks. Communications of the ACM 13(6):377-387.

56. Ramakrishnan, R., and Gehrke, J. 2000. Database Management Systems, 3rd ed. New York, NY: McGraw-Hill Higher Education.

57. Birbeck, M., Duckett, J., and Gudmundsson, O. G. et al. 2001. Professional XML, 2nd edn. Birmingham, England: Wrox Press Inc.

58. Bray, T. 1998. Introduction to the Annotated XML Speci�cation: Extensible Markup Language (XML) 1.0. http://www.xml.com/axml/testaxml.htm, (Accessed Sept. 1, 2009).

59. W3Schools. Introduction to DTD. 2003. http://www.w3schools.com/DTD/ dtd_intro.asp (Accessed Aug. 4, 2008).

61. McLaughlin, B., and Loukides, M. 2001. Java and XML. (O'Reilly Java Tools). O'Reilly & Associates.

60. World Wide Web Consortium. 2006. XML Schema.

http://www.w3.org/XML/ Schema (Accessed Aug. 4, 2008).

62. W3 Schools. 2008. XML DOM Tutorial. 2008. http://www.w3schools.com/ dom/default.asp (Accessed Aug. 4, 2008).

63. Official SAX Website. 2004. http://www.saxproject.org/about.html, (Accessed Sept. 1, 2009).

2 Chapter 2. Electronic Health Records

1. HL7 Clinical Document Architecture, Release 2.0 (2004). http://lists.hl7.org/read/attachment/61225/1/CDA-doc%20version.pdf (Accessed Sept. 25, 2008).

2. Chheda, N. C. 2005. Electronic Medical Records and Continuity of Care Records—the Utility Theory. http://www.emrworld.net/emr-research/articles/emr-ccr.pdf, (Accessed Sept. 1, 2009).

3. Garets, D., and Davis, M. 2006. Electronic medical records vs. electronic health records: yes, there is a difference. HIMSS Analytics White Paper.

4. Healthcare Information and Management Systems Society. http://www.himss. org/, (Accessed Sept. 1, 2009).

5. Wikipedia. 2008. Electronic Health Record. http://en.wikipedia.org/wiki/ Electronic_health_record, (Accessed Sept. 1, 2009).

6. Personal Health Records. http://en.wikipedia.org/wiki/Personal_health_ record, (Accessed Sept. 1, 2009).

7. Microsoft HealthVault. 2008. http://www.healthvault.com/, (Accessed Sept. 1, 2009).

8. Google Health. 2008. http://www.google.com/health/, (Accessed Sept. 1, 2009).

9. Records for Living. HealthFrame. http://www.recordsforliving.com/ HealthFrame/, (Accessed Sept. 1, 2009).

10. Kohn, L. T., Corrigan, J. M., and Donaldson, M. 1999. To Err is Human: Building a Safer Health System. Washington, DC: Institute of Medicine.

11. Blumenthal, D., DesRoches, C., Donelan, K. et al. 2006. Health Information Technology in the United States: The Information Base for Progress. Report for the Robert Wood Johnson Foundation.

12. Wikipedia. ISO/TC 215 Standard. http://en.wikipedia.org/wiki/ISO_TC_215, (Accessed Sept. 1, 2009).

13. Wikipedia. OpenEHR. 2008. http://en.wikipedia.org/wiki/Openehr, (Accessed Sept. 1, 2009).

14. Wikipedia. ASC X12. 2008c. http://en.wikipedia.org/wiki/ANSI_X12, (Accessed Sept. 1, 2009).

15. United States Department of Health and Human Services: Office of Civil Rights— HIPAA. 2006. http://www.hhs.gov/ocr/hipaa/, (Accessed Sept. 1, 2009).

16. European eHealth Continuity Site. 2009. CONTsys. http://www.contsys.eu/.

17. Miller, R. H., and Sim, I. 2004. Physicians use of electronic medical records. Barriers and solutions. Health Affairs 23:116–126.

18. Hillestad, R., Bigelow, J., Bower, A., Girosi, F., Meili, R., Scoville, R., and Taylor, R. 2005. Can electronic medical record systems transform health care? Potential health benefits, savings and costs. Health Affairs 24:1103–1117.

19. Bates, D. W., Ebell, M., Gotlieb, E., and Zapp, J. 2003. A proposal for electronic medical records in U.S. primary care. Journal of the American Medical Informatics Association 10:616.

20. Burke, R. P., and White, J. A. 2004. Internet rounds: A congenital heart surgeon's web log. Seminars in Thoracic and Cardiovascular Surgery 16(3):283–292.

21. Boulus, N., and Bjorn. P. 2008. A cross-case analysis of technology-in-use practices: EPR-adaptation in Canada and Norway. International Journal of Medical Informatics, in press. (Available online July 31, 2008.)

22. Ludwick, D. A., and Doucette, J. 2008. Adopting electronic medical records in primary care: lessons learned from health information systems implementation experience in seven countries. International Journal of Medical Informatics 78(1):22–31.

23. Open Clinical. 2007. Ontologies. http://www.openclinical.org/ontologies. html, (Accessed Sept. 1, 2009).

24. Gangemi, A., Pisanelli, D. M., and Steve, G. 1998.

Ontology integration: experiences with medical terminologies. In Formal Ontology in Information Systems, N. Guarino, ed. Amsterdam: IOS Press, pp. 163–178.

25. Medicomp Systems. MEDCIN. http://en.wikipedia.org/wiki/MEDCIN, (Accessed Sept. 1, 2009).

26. U.S. Food and Drug Administration. The National Drug Code Directory. http:// www.fda.gov/cder/ndc/, (Accessed Sept. 1, 2009).

26a. Hristidis, V., Farfán, F., Burke, R., Rossi, A., and White, J. 2007. Information discovery on electronic medical records. In National Science Foundation Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation, NGDM.

27. Eichelberg, M., Aden, T., Riesmeier, J., Dogac, A., and Laleci, G.B. 2005. A survey and analysis of Electronic Healthcare Record standards. ACM Computing Surveys 37(4):277–315.

28. Wikipedia. Medical Record. 2008. In http://en.wikipedia.org/wiki/Medical_ record, (Accessed Sept. 1, 2009).

29. HL7 Reference Information Model. 2008. http://www.hl7.org/library/ datamodel/RIM/C30204/rim.htm (Accessed Sept. 25, 2009).

30. HL7 CCOW Technical Committee. 2007. Clinical Context Object Workgroup (CCOW). http://www.hl7.org/special/Committees/ccow_sigvi.htm (Accessed Sept. 25, 2009).

31. Health Level Seven Inc. 2006. Health Level Seven (HL7). http://www.hl7.org/ (Accessed Sept. 1, 2008).

32. Unified Modeling Language. 2008. http://www.uml.org/, (Accessed Sept. 1, 2009).

33. DICOM Standards Committee. 2006. Digital Imaging and Communications in Medicine (DICOM). http://medical.nema.org/(Accessed Sept. 1, 2009).

34. JPEG Interactive Protocol (JPIP). 2006. http://www.jpeg.org/jpeg2000/ j2kpart9.html, (Accessed Sept. 1, 2009).

35. Integrating the Healthcare Enterprise. 2006. http://www.ihe.net/ (Accessed Sept. 25, 2009).

36. Hristidis, V., Clarke, P. J., Prabakar, N., Deng, Y., White, J. A., and Burke, R. P. 2006. A flexible approach for electronic medical records exchange. In Proceedings of the Workshop on Health and Information Knowledge Management (HIKM) 2006, in conjunction with CIKM 2006.

37. Wikipedia. Electronic Medical Record. 2008e. http://en.wikipedia.org/wiki/ Electronic_Medical_Record, (Accessed Sept. 1, 2009).

38. Waegemann, C. P. 2003. EHR vs. CCR: what is the difference between the electronic health record and the continuity of care record? Medical Records Institute, http:// www.medrecinst.com/pages/libArticle.asp?id=42 (Accessed Oct. 2, 2008).

39. Conn, J. 2006. Identity crisis? Renewed debate over national patient ID. Modern Healthcare Magazine, http://www.modernhealthcare.com/article.cms?articleId=39954, (Accessed Sept. 1, 2009).

40. U.S. House of Representatives. 2009. “The American Recovery and Reinvestment Act of 2009.” Committee on Rules.

41. American Health Information Management Association (AHIMA) 2008. http:// www.ahima.org/, (Accessed Sept. 1, 2009).

42. McDonald, C. J. 1997. The barriers to electronic medical record systems and how to overcome them. Journal of the American Medical Informatics Association 4:213–221.

43. Synamed. 2008. http://www.synamed.com/, (Accessed Sept. 1, 2009).

44. Teges Corporation. 2008. http://www.teges.com/, (Accessed Sept. 1, 2009).

45. MedcomSoft. 2008. http://www.medcomsoft.com/, (Accessed Sept. 1, 2009).

46. StreamlineMD. 2008. http://www.streamline-md.com/, (Accessed Sept. 1, 2009).

47 Practice Partner. 2008. http://www.practicepartner.com/(Accessed Aug. 2, 2008).

48. Practice Partner. 2008.
http://www.practicepartner.com/(Accessed Aug. 2, 2008).

48. eClinicalWorks. 2008. http://www.eclinicalworks.com/, (Accessed Sept. 1, 2009).

49. MediNotes. 2008. http://www.medinotes.com/, (Accessed Sept. 1, 2009).

50. Misys PLC Health Care. 2008.
http://www.misyshealthcare.com/, (Accessed Sept. 1, 2009).

51. NextGen. 2008. http://www.nextgen.com/, (Accessed Sept. 1, 2009).

52. Allscripts. 2008. http://www.allscripts.com/, (Accessed Sept. 1, 2009).

53. OmniMD. 2008. http://www.omnimd.com/, (Accessed Sept. 1, 2009).

54. GE HealthCare. 2008. http://www.gehealthcare.com/, (Accessed Sept. 1, 2009).

55. InteGreat. 2008. http://www.igreat.com/, (Accessed Sept. 1, 2009).

56. Cerner. 2008. http://www.cerner.com/, (Accessed Sept. 1, 2009).

57. AC Group website. 2008. http://www.acgroup.org (Accessed Aug. 2, 2008).

57a. Wang, A. Y., Barret, J. W., Bentley, T., Markwell, D., Price, C., Spackman, K. A. et al. 2001. Mapping between SNOMED RT and clinical terms, version 3: A key component of the SNOMED CT development process. In Proceedings/AMIA Annual Symposium.

58. SNOMED Clinical Terms (SNOMED CT). 2008.
http://www.snomed.org/ snomedct/index.html, (Accessed Sept. 1, 2009).

59. Spackman, K. A., Campbell, K. E., and Cote, R. A. 1997. SNOMED-RT: a reference terminology for health care. In Proceedings of the 1997 AMIA Annual Fall Symposium, pp. 640-644.

60. McDonald, C. J., Huff, S., Mercer, K., Hernandez, J.

A., and Vreeman, D., ed. 2009. Logical Observation Identifiers Names and Codes (LOINC) User's Guide. Indianapolis, IN: Regenstrief Institute.

61. Health Insurance Portability and Accountability Act. 2008. http://www.hipaa. org/, (Accessed Sept. 1, 2009).

62. McDonald, C. J., Huff, S. M., Suico, J. G. et al. 2003. LOINC, a universal standard for identifying laboratory observations: a 5-year update. Clinical Chemistry 49(4):624–633.

63. RxNorm. 2008. United States National Library of Medicine. http://www.nlm. nih.gov/research/umls/rxnorm/index.html (Accessed Sept. 1, 2008).

64. United States National Library of Medicine. Medical Subject Headings (MeSH) Fact List. http://www.nlm.nih.gov/pubs/factsheets/mesh.html (Accessed Sept. 1, 2008).

65. DICOM Files. DICOM Sample Image Sets. http://pubimage.hcuge.ch:8080/, (Accessed Sept. 1, 2008).

3 Chapter 3. Overview of Information Discovery Techniques on EHRs

Banko, M., Brill, E., Dumais, S., and Lin, J. 2002. AskMSR: question answering using the Worldwide Web. In Proceedings of 2002 AAAI Spring Symposium on Mining Answers from Texts and Knowledge Bases.

Berners-Lee, T., Hendler, J., and Lassila, O. 2001. The semantic Web. Scientific American 284(5):34–43.

Cios, K. J., and Moore, W. G. 2002. Uniqueness of medical data mining. Elsevier Artificial Intelligence in Medicine 26(1):1–24.

Coiera, E., Westbrook, J. I., and Rogers, K. 2008. Clinical decision velocity is increased when meta-search filters enhance an evidence retrieval system. Journal of the American Medical Informatics Association 15(5):638–646.

Duncan, J. C., and Ayache, N. 2000. Medical image analysis: progress over two decades and the challenges ahead. IEEE Transactions on Pattern Analysis Machine Intelligence 22(1):85–106. DOI:http://dx.doi.org/10.1109/34.824822.

Inokuchi, A., Takeda, K., Inaoka, N., and Wakao, F. 2007. MedTAKMI-CDI: interactive knowledge discovery for clinical decision intelligence. IBM Systems Journal 46(1):115–134.

Lassila, O., and Swick, R. R. 1999. Resource Description Framework (RDF) Model and Syntax Specification. W3C. http://www.w3.org/TR/1999/REC-rdf-syntax19990222/.

Manning, C., and Schütze, H. 1999. Foundations of statistical natural language processing. Cambridge, MA: MIT Press.

Pfeiffer, K. P., Goebel, G., and Leitner, K. 2003. Demand for intelligent search tools in medicine and health care. Lecture Notes in Computer Science 2818:5–18.

Proper, H. A., and Bruza, P. 1999. What is information discovery about? Journal of the American Society for Information Science and Technology 50(9):737–750.

Salton, G. 1989. Automatic text processing: the transformation, analysis, and retrieval of information by computer. Boston, MA: AddisonWesley Longman.

Singhal, A. 2001. Modern information retrieval: a brief

overview. Bulletin of the IEEE Computer Society Technical
Committee on Data Engineering 24(4):35-42.

4 Chapter 4. Data Quality and Integration Issues in Electronic Health Records

1. Komaroff, A. L. 1979. The variability and inaccuracy of medical data. Proceedings of the IEEE 67(9):1196–1207.

2. Hogan, W. R., and Wagner, M. M. 1997. Accuracy of data in computerbased patient records. Journal of the American Medical Informatics Association 4(5):342–355.

3. Wyatt, J. C., and Wright, P. 1998. Design should help use of patients' data. Lancet (British edition) 352(9137):1375–1378.

4. Coiera, E. 2003. Guide to Health Informatics. London: Arnold London.

5. Barnett, O. 1990. Computers in medicine. JAMA 263(19):2631.

6. Richart, R. H. 1970. Evaluation of a medical data system. Computers and Biomedical Research 3(5):415.

7. Audit Commission. 1995. For your information: a study of information management and systems in the acute hospital. London: HMSO.

8. Mamlin, J. J., and Baker, D. H. 1973. Combined time-motion and work sampling study in a general medicine clinic. Medical Care 11:449–456.

9. Korpman, R. A., and Lincoln. T. L. 1998. The computer-stored medical record: for whom? Journal of the American Medical Informatics Association 259:3454–3456.

10. Wyatt, J. C. 1994. Clinical data systems: Part 1. Data and medical records. The Lancet 344:1543–1547.

11. Dick, R. S., and Steen, E. B., eds. 1977. The Computer-based Patient Record: An Essential Technology for HealthCare. Washington, D.C.: National Academy Press.

12. Nygren, E., Wyatt, J. C., and Wright, P. 1998. Helping clinicians to find data and avoid delays. The Lancet, 352:1462–1466.

13. Hammond, K.W., Helbig, S. T., Benson, C. C., and Brathwaite-Sketoe, B. M. 2003. Are electronic medical records trustworthy? Observations on copying, pasting and duplication. In AMIA Annual Symposium Proceedings, pp.

269-73.

14. Hohnloser, J. H., Fischer, M. R., Konig, A., and Emmerich, B. 1994. Data quality in computerized patient records. Analysis of a haematology biopsy report database. International Journal of Clinical Monitoring and computing 11(4):233-240.

15. Weir, C. R., Hurdle, J. F., Felgar, M. A., Hoffman, J. M., Roth, B., and Nebeker, J. R. 2003. Direct text entry in electronic progress notes—an evaluation of input errors. Methods of Information in Medicine 42(1):61-67.

16. Berner, E., and Moss, J. 2005. Informatics challenges for the impending patient information explosion. Journal of the American Medical Informatics Association 12(6):614-617.

17. Hogan, W. R., and Wagner, M. M. 1997. Accuracy of data in computerbased patient records. Journal of the American Medical Informatics Association 4(5):342-355.

18. Savage, A. M. 1999. Framework for characterizing data and identifying anomalies in health care databases. In Proceedings of the AMIA Symposium, p. 374. American Medical Informatics Association.

19. Muthukrishnan, S. 2005. Data streams: Algorithms and Applications. New York, NY: Now Publishers Inc.

20. Johnson, S. B. 1996. Generic data modeling for clinical repositories. Journal of the American Medical Informatics Association 3(5):328-339.

21. Van Ginneken, A. M., Stam, H., and Duisterhout, J. S. 1994. A powerful macro-model for the computer patient record. In Proceedings of the Annual Symposium on Computer Application in Medical Care, p. 496. American Medical Informatics Association.

22. Los, R. K. 2006. Supporting Uniform Representation of Data. PhD thesis, Department of Medical Informatics, Erasmus Medical Center, Rotterdam, the Netherlands.

23. Shortliffe, E. H., and Cimino, J. J. 2006. Biomedical Informatics—Computer Applications in Health Care and Biomedicine, 3rd edn. New York, NY: Springer.

24. Shortliffe, E. H., Perreault, L. E., Wiederhold, G., and Fagan, L. M. 1990. Medical Informatics: Computer

Applications in Health Care. Boston, MA: Addison-Wesley Longman Publishing Co.

25. Wyatt, J. C., and Sullivan, F. 2005. ABC of Health Informatics. Blackwell Publishing, Malden, MA: BMJ Books.

26. Riva, G. 2003. Ambient intelligence in health care. Cyberpsychology & Behavior 6(3):295–300.

27. Anfindsen, O. J. 2000. Database management system and method for combining meta-data of varying degrees of reliability. US Patent 6,044,370.

28. Tayi, G. K., and Ballou, D. P. 1998. Examining data quality. Communications of the ACM 41(2):54–57.

29. Wyatt, J. C., and Liu, J. L. Y. 2002. Basic concepts in medical informatics. British Medical Journal 56(11):808–812.

30. Wang, R. Y. 1998. Total data quality. Communications of the ACM 41(2):58–65.

31. Gertz, M., Ozsu, T., Saake, G., and Sattler, K. 2003. Data quality on the Web. In Dagstuhl Seminar, Dagstuhl, Germany.

32. Strong, D. M. Lee, Y. W., and Wang, R. Y. 1997. Data quality in context. Communications of the ACM 40(5):103–110.

33. Orr, K. 1998. Data quality and systems theory. Communications of the ACM 41(2):66–71

34. Pourasghar, F., Malekafzali, H., Kazemi, A., Ellenius, J., and Fors, U. 2008. What they fill in today, may not be useful tomorrow: lessons learned from studying medical records at the women's hospital in Tabriz, Iran. BMC Public Health 8(1):139.

35. Jaspers, M. W., Knaup, P., and Schmidt, D. 2006. The computerized patient record: where do we stand. Methods of Information in Med, 45(Suppl 1):29–39.

36. Soto, C. M., Kleinman, K. P., and Simon, S. R. 2002. Quality and correlates of medical record documentation in the ambulatory care setting. BMC Health Services Research 2(1):22.

37. Roberts, C. L., Algert, C. S., and Ford, J. B. 2007.

Methods for dealing with discrepant records in linked population health datasets: a cross-sectional study. BMC Health Services Research 7:12.

38. Arts, D. G. T., de Keizer, N. F., and Scheffer, G. J. 2002. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. Journal of the American Medical Informatics Association 9(6):600–611.

39. Oliveira, P., Rodrigues, F., and Henriques, P. 2005. A formal definition of data quality problems. In IQ, F. Naumann, M. Gertz, and S. Madnick, eds. Cambridge, MA: MIT Press.

40. Maletic, J. I., and Marcus, A. 2000. Data cleansing: beyond integrity analysis. In Proceedings of the Conference on Information Quality, pp. 200–209.

41. Kamber, M., and Han, J. 2001. Data Mining: Concepts and Techniques. San Francisco, CA: Morgan Kaufmann Publishers.

42. Koh, H. C., and Tan, G. 2005. Data mining applications in healthcare. Journal of Healthcare Information Management 19(2):64–72.

43. Na, K. S., Baik, D. K., and Kim, P. K. 2001. A practical approach for modeling the quality of multimedia data. In Proceedings of the 9th ACM international conference on Multimedia, pp. 516–518. New York, NY: ACM Press.

44. Hodge, V., and Austin, J. 2004. A survey of outlier detection methodologies. Artificial Intelligence Review 22(2):85–126.

45. Lee, A. H., Xiao, J., Vemuri, S. R., and Zhao, Y. 1998. A discordancy test approach to identify outliers of length of hospital stay. Statistics in Medicine, 17(19):2199–2206.

46. Podgorelec, V., Hericko, M., and Rozman, I. 2005. Improving mining of medical data by outliers prediction. In 18th IEEE Symposium on Computer-Based Medical Systems, 2005. Proceedings, pp. 91–96.

47. Ramaswamy, S., Rastogi, R., and Shim, K. 2000. Efficient algorithms for mining outliers from large data sets. In Proceedings of the 2000 ACM SIGMOD International Conference on Management of data, pp. 427–438. New York,

NY: ACM Press.

48. Network, I. 2002. Population and health in developing countries: volume 1. Population, Health, and Survival at INDEPTH sites. IDRC, Ottawa, ON, CA.

49. Palmblad, M., and Tiplady, B. 2004. Electronic diaries and questionnaires: designing user interfaces that are easy for all patients to use. Quality of Life Research 13(7):1199–1207.

50. Hyeoneui, K., Harris, M. R., Savova, G. K., and Chute, C. G. 2008. The first step toward data reuse: disambiguating concept representation of the locally developed ICU nursing flowsheets. Computers, Informatics, Nursing 26(5):282.

51. Connell, F. A., Diehr, P., and Hart, L. G. 1987. The use of large data bases in health care studies. Annual Review of Public Health 8(1):51–74.

52. Ola, B., Khan, K. S., Gaynor, A. M., and Bowcock, M. E. 2001. Information derived from hospital coded data is inaccurate: the Birmingham Women’s Hospital experience. Journal of Obstetrics and Gynaecology 21(2):112–113.

53. Cohen, J. 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20(1):37.

54. Ellis, J., Mulligan, I., Rowe, J., and Sackett, D. L. 1995. Inpatient general medicine is evidence based. A-team, Nuffield Department of Clinical Medicine. Lancet 346(8972):407.

55. Gorelick, M. H., Knight, S., Alessandrini, E. A., Stanley, R. M., Chamberlain, J. M., Kuppermann, N., and Alpern, E. R. 2007. Lack of agreement in pediatric emergency department discharge diagnoses from clinical and administrative data sources. Academic Emergency Medicine 14(7):646–652.

56. Icen, M., Crowson, C. S., McEvoy, M. T., Gabriel, S. E., and Kremers, H. M. 2008. Potential misclassification of patients with psoriasis in electronic databases. Journal of the American Academy of Dermatology 59(6):981–985.

57. Silva-Costa, T., Freitas, A., Jácome, J., Lopes, F., and Costa-Pereira, A. 2007. A eficicia de uma ferramenta de validao na melhoria da qualidade de dados hospitalares. In

CISTI 2007—2 Conferncia Ibrica de Sistemas e Tecnologias de Informao, Portugal, June.

58. Keren, R., Wheeler, A., Cof�n, S. E., Zaoutis, T., Hodinka, R., and Heydon, K. 2006. ICD-9 codes for identifying in�uenza hospitalizations in children. Emerging Infectious Disease 12(10):1603–1604.

59. Goldstein, L. B. 1998. Accuracy of ICD-9-CM coding for the identi�cation of patients with acute ischemic stroke: effect of modi�er codes. Stroke 29(8):1602–1604.

60. Rassinoux, A. M., Miller, R. A., Baud, R. H., and Scherrer, J. R. Modeling concepts in medicine for medical language understanding. Methods of Information in Medicine 37(4–5):361–372.

61. Kalra, D., Beale, T., and Heard, S. 2005. The openEHR Foundation. Studies in Health Technology and Informatics 115:153.

62. Synchronize timepieces in your trauma room. 2000. ED Management, 12(2):23–24.

63. McCartney, P. R. 2003. Synchronizing with standard time and atomic clocks. MCN: The American Journal of Maternal Child Nursing 28(1):51.

64. Ornato, J. P., Doctor, M. L., Harbour, L. F., Peberdy, M. A., Overton, J., Racht, E. M., Zauhar, W. G., Smith, A. P., and Ryan, K. A. 1998. Synchronization of timepieces to the atomic clock in an urban emergency medical services system. Annals of Emergency Medicine 31(4):483–487.

65. Kaye, W., Mancini, M. E., and Truitt, T. L. 2005. When minutes count: the fallacy of accurate time documentation during in-hospital resuscitation. Resuscitation 65(3):285–290.

66. Ferguson, E. A., Bayer, C. R., Fronzeo, S., Tuckerman, C., Hutchins, L., Roberts, K., Verger, J., Nadkarni, V., and Lin, R. 2005. Time out! Is timepiece variability a factor in critical care? American Journal of Critical Care 14(2):113.

67. Neumann, P. 1995. Computer-Related Risks. New York, NY: ACM Press.

68. Institute of Medicine. 2001. Crossing the Quality Chasm: A New Health System for the 21st Century.

Washington, D. C.: National Academy Press.

69. Herzlinger, R. E. 2004. Consumer-Driven Health Care: Implications for Providers, Payers, and Policy-Makers. San Francisco, CA: Jossey-Bass.

70. MacStravic, S. 2004. What good is an EMR without a PHR? HealthLeaders, September 3.

71. Kukafka, R., and Morrison, F. 2006. Patients' needs. In Aspects of Electronic Health Record Systems, H. P. Lehmann et al., eds. pp. 47–64. Calgary: Springer.

72. Kalra, D. 2006. Electronic health record standards. Methods of Information in Medicine 45(1):136–144.

73. Land, R., and Crnkovic, I. 2003. Software systems integration and architectural analysis—a case study. In Proceedings of the International Conference on Software Maintenance, ICSM 2003, pp. 338–347.

74. Heathfield, H., Pitty, D., and Hanka, R. 1998. Evaluating information technology in health care: barriers and challenges. British Medical Journal 316:1959–1961.

75. Berg, M. 2001. Implementing information systems in health care organizations: myths and challenges. International Journal of Medical Informatics 64(2–3):143–156.

76. Littlejohns, P., Wyatt, J. C., and Garvican, L. 2003. Evaluating computerised health information systems: hard lessons still to be learnt. British Medical Journal 326:860–863.

77. Lenz, R., and Kuhn, K. A. 2002. Integration of heterogeneous and autonomous systems in hospitals. Business Briefing: Data management & Storage Technology.

78. Ferranti, J., Musser, C., Kawamoto, K., and Hammon, E. 2006. The clinical document architecture and the continuity of care record: a critical analysis. Journal of the American Medical Informatics Association 13(3):245–252.

79. Cruz-Correia, R. J., Vieira-Marques, P., Ferreira, A., Almeida, F., Wyatt, J. C., and Costa-Pereira, A. 2007. Reviewing the integration of patient data: how systems are evolving in practice to meet patient needs. BMC Medical Informatics and Decision Making 7(1):14.

80. Chen, R., Enberg, G., and Klein, G. O. 2007. Julius—a template based supplementary electronic health record system. BMC Medical Informatics and Decision Making 7(1):10.

81. Los, R. K., van Ginneken, A. M., and van der Lei, J. 2005. OpenSDE: a strategy for expressive and flexible structured data entry. International Journal of Medical Informatics 74(6):481–490.

82. Hoya, D., Hardikerb, N. R., McNicollc, I. T., Westwelld, P., and Bryana, A. 2008. Collaborative development of clinical templates as a national resource. Internation Journal of Medical Informatics 78(1):95–100.

83. Quantin, C., Binquet, C., Bourquard, K., Pattisina, R., Gouyon-Cornet, B., Ferdynus, C., Gouyon, J. B., and Allaert, F. A. 2004. A peculiar aspect of patients' safety: the discriminating power of identifiers for record linkage. Studies in Health Technology and Informatics 103:400–406.

84. Arellano, M. G., and Weber, G. I. 1998. Issues in identification and linkage of patient records across an integrated delivery system. Journal of Healthcare Information Management 12(3):43–52.

85. Cruz-Correia, R., Vieira-Marques, P., Ferreira, A., Oliveira-Palhares, E., Costa, P., and Costa-Pereira, A. 2006. Monitoring the integration of hospital information systems: how it may ensure and improve the quality of data. Studies in Health Technology and Informatics 121:176–82.

86. Dunn, H. L. 1946. Record linkage. American Journal of Public Health 36(12):1412.

87. Blakely, T., and Salmond, C. 2002. Probabilistic record linkage and a method to calculate the positive predictive value. International Journal of Epidemiology 31(6):1246–1252.

88. Scheuren, F. 1997. Linking health records: human rights concerns. In Record Linkage Techniques—1997: Proceedings of an International Workshop and Exposition, March 20–21, 1997, Arlington, VA, p. 404. Federal Committee on Statistical Methodology, Office of Management and Budget.

89. Evans, J. M. M., and MacDonald, T. M. 1999. Record-linkage for pharmacovigilance in Scotland. British Journal of Clinical Pharmacology 47(1):105–110.

90. Karmel, R., and Gibson, D. 2007. Event-based record linkage in health and aged care services data: a methodological innovation. BMC Health Services Research 7(1):154.

91. Hägglund, M., Chen, R., Scandurra, I., and Koch, S. 2009. Modeling shared care plans using CONTsys and openEHR to support shared homecare of elderly. Submitted.

92. Chen, R., Hemming, G., and A . hlfeldt, H. 2009. Representing a chemotherapy guideline using openEHR and rules. Studies in Health Technology and Informatics 150: 653–657.

93. Chen, R., Klein, G., Sundvall, E., Karlsson, D., and A . hlfeldt, H. 2009, July. Archetype-based import and export of EHR content models: pilot experience with a regional EHR system. BMC Medical Informatics and Decision Making 9:33.

94. Pearson, R. K. 2005. Mining Imperfect Data: Dealing with Contamination and Incomplete Records. Philadelphia, PA: Society for Industrial and Applied Mathematics.

95. Hawkins, D. M. 1980. Identification of Outliers. London: Chapman and Hall.

96. Breiman, L., Friedman, J. H., Olshen, R. A., and Stone. C. J. 1984. Classification and Regression Trees. New York, NY: Chapman and Hall/CRC.

97. Rodrigues, P. P., Gama, J., and Bosnizc ´, Z. 2008. Online reliability estimates for individual predictions in data streams. In Proceedings of the 8th International Conference on Data Mining Workshops (ICDM Workshops'08), pp. 36–45. Pisa, Italy, December, IEEE Computer Society Press.

98. Efron, B. 1979. Bootstrap methods: another look at the jackknife. Annals of Statistics 7:1–26.

99. Hastie, T., Tibshirani, R., and Friedman, J. 2000. The Elements of Statistical Learning: Data Mining, Inference and Prediction. New York, NY: Springer Verlag.

100. Breiman, L. 1996. Bagging predictors. Machine Learning 24:123–140.

101. Drucker, H. 1997. Improving regressors using boosting

techniques. In Machine Learning: Proceedings of the 14th International Conference, pp. 107-115.

102. Gama, J., and Rodrigues, P. P. 2007. Data stream processing. In Learning from Data Streams—Processing Techniques in Sensor Networks, J. Gama and M. Gaber, eds., chapter 3, pp. 25–39. Berlin: Springer Verlag.

103. Gama, J., and Gaber, M., eds. 2007. Learning from Data Streams—Processing Techniques in Sensor Networks. Berlin: Springer Verlag.

104. Gama, J., Medas, P., Castillo, G., and Rodrigues, P. P. 2004. Learning with drift detection. In Proceedings of the 17th Brazilian Symposium on Artificial Intelligence (SBIA 2004), volume 3171 of Lecture Notes in Artificial Intelligence, A. L. C. Bazzan and S. Labidi, eds., pp. 286–295, São Luiz, Maranhão, Brazil, October 2004. Springer Verlag.

105. Böttcher, M., Höppner, F., and Spiliopoulou, M. 2008. On exploiting the power of time in data mining. SIGKDD Explorations 10(2):3–11.

5 Chapter 5. Ethical, Legal, and Social Issues for EHR Data Protection

Advisory Committee on Automated Personal Data Systems. 1973. Records, Computers, and the Rights of Citizens. Washington, DC: US Government Printing Office.

Agre, P. E., and Rotenberg, M. 1997. Technology and Privacy: The New Landscape. Cambridge: MIT Press.

American Hospital Association. 1992. A Patient's Bill of Rights. Chicago IL: American Hospital Association.

American Medical Association. 2009. Code of Medical Ethics: Current Opinions with Annotations, 2008–2009. Chicago IL: American Medical Association.

Anderson, J. G., and Goodman, K. W. 2002. Ethics and Information Technology. New York: Springer-Verlag.

Anderson, R. 2001. Security Engineering: A Guide to Building Dependable Distributed Systems. Indianapolis, IN: John Wiley and Sons, Inc.

Bainbridge, D. 2005. Data Protection Law. St. Albans, UK: XPL Publishing.

Barrows, R., and Clayton, P. 1996. Privacy, confidentiality and electronic medical records. Journal of the American Medical Informatics Association 3:139–148.

Beauchamp, T. L., and Childress, J. F. 1994. Principles of Biomedical Ethics, 4th edn. New York NY: Oxford University Press.

Bennett, B. 1995. Medical Records: Sweeping Reforms to Ensure Privacy of Personal Medical Records. Washington, DC: Federal Document Clearinghouse Press, Release, 24 October.

Bennett, C. J. 1992. Regulating Privacy: Data Protection and Public Policy in Europe and the United States. Ithaca NY: Cornell University Press.

Bennett, C. J. 2008. The Privacy Advocates: Resisting the Spread of Surveillance. Cambridge: Massachusetts Institute of Technology Press.

Bishop, M. 2003. Computer Security: Art and Science. London, UK: Pearson Education, Inc.

Centers for Disease Control and Prevention. 1996. Legislative Survey of State Confidentiality Laws, with Specific Emphasis on HIV and Immunization.

Code of Federal Regulations, Current Titles. U.S. Government Printing Office. http:// www.gpoaccess.gov/cfr/.

Committee on the Role of Institutional Review Boards in Health Services Research Data Privacy Protection. 2000. Protecting Data Privacy in Health Services Research. Institute of Medicine. Washington, DC: National Academy Press.

Council of Europe. 1981. Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data.

Department of Health and Human Services. 1993. Health Records: Social Needs and Personal Privacy. Washington, DC: US Government Printing Office.

Dick, R. S., Steen, E. B., and Detmer, D. E. (eds.) 1991. The Computer-Based Patient Record: An Essential Technology for Health Care. Institute of Medicine. Washington, DC: National Academy Press.

Donaldson, M. S., and Lohr, K. (eds). 1994. Health Data in the Information Age: Use, Disclosure and Privacy. Institute of Medicine. Washington, DC: National Academy Press.

Electronic Privacy Information Center and Privacy International. 2006. Privacy and Human Rights 2006: An International Survey of Privacy Laws and Developments. Washington, DC: EPIC.

Etzioni, A. 2000. The Limits of Privacy. New York, NY: Basic Books.

Etziony, M. B. 1973. The Physician's Creed: An Anthology of Medical Prayers, Oaths and Codes of Ethics Written and Recited by Medical Practitioners through the Ages. Springfield, IL: Charles C. Thomas.

European Union. 1995. Directive 95/46/EC on the Protection of Individuals With Regard to the Processing of Personal Data.

Fairchild, A. L., Bayer, R., and Colgrove, J. 2007.

Privacy, the State and Disease Surveillance in America. Berkeley, CA: Milbank Foundation.

Flaherty, D. H. 1989. Protecting Privacy in Surveillance Societies: The Federal Republic of Germany, Sweden, France, Canada, and the United States. Chapel Hill, NC: University of North Carolina Press.

Gavison, R. 1984. Privacy and the limits of law. In Schoeman, F. (ed), Philosophical Dimensions of Privacy: An Anthology. Cambridge UK: Cambridge University Press.

Gellman, R. M. 1996. Can privacy be regulated effectively on a national level? Thoughts on the possible need for international privacy rules. Villanova Law Review 41(1):129–165.

Goldman, J. 1995. Statement before the Senate Committee on Labor and Human Resources on S.1360. Federal Document Clearinghouse Congressional Testimony, 14 November.

Gollmann, D. 1999. Computer Security. New York, NY: John Wiley and Sons, Inc.

Goodman, K. W. 1998. Ethics, Computing and Medicine. Cambridge UK: Cambridge University Press.

Gostin, L. O. et al. 1993. Privacy and security of personal information in a new health care system. Journal of the American Medical Association 270:2487–2493.

Gostin, L. O. 1994. Health information privacy. Cornell Law Review 80:101–132.

Harris Interactive. 2007. Harris Poll #127—Health Information Privacy Survey, http://www.harrisinteractive.com.

Holtzman, D. H. 2006. Privacy Lost: How Technology Is Endangering Your Privacy. San Francisco, CA: Jossey-Bass.

Humbler, J. M., and Almeder, R. F. (eds.) 2001. Privacy and Health Care (Biomedical Ethics Reviews). Totowa, NJ: Humana Press.

International Organization for Standardization (ISO). 2005. Information Technology— Security Techniques—Code of Practice for Information Security Management. Geneva: ISO Publications.

Jha, As. K., Ferris, T. G., Donelan, K. et al. 2006. How common are electronic health records in the United States? A summary of the evidence. Health Affairs 25(6):496–507.

Joint Commission. 2005. Comprehensive Accreditation Manual for Hospitals. Oakbrook Terrace, IL: Joint Commission on Accreditation of Healthcare Organizations.

Nass, S. J., Levit, L. A., and Gostin, L. G. (eds.) 2009. Committee on Health Research and the Privacy of Health Information: Beyond the HIPAA Privacy Rule—Enhancing Privacy, Improving Health through Research. Institute of Medicine. Washington, DC: National Academy Press.

Ness, R. B., for the Joint Policy Committee, Societies of Epidemiology. 2007. Influence of the HIPAA privacy rule on health research. JAMA 298:2164–2170.

National Committee on Vital and Health Statistics (NCVHS). 1997. Hearings of the Subcommittee on Health Data Needs, Standards and Security, and of the Subcommittee on Privacy and Confidentiality, under the Health Insurance Portability and Accountability Act (PL 104–191).

NCQA. 2008. Physician and Hospital Quality Standards and Guidelines.

National Research Council. 1972. Databanks in a Free Society: Computers, Record-Keeping, and Privacy. Washington, DC: National Academy Press.

National Research Council. 1991. Computers at Risk: Safe Computing in the Information Age. Washington, DC: National Academy Press.

National Research Council. 1997. For the Record: Protecting Electronic Health Information. Washington, DC: National Academy Press.

Neumann, P. 1995. Computer-Related Risks. Reading, MA: Addison-Wesley.

Office of Technology Assessment. 1993. Protecting Privacy in Computerized Medical Information. Washington, DC: US Government Printing Office.

Office of Technology Assessment. 1995. Bringing Health Care Online: The Role of Information Technologies. Washington, DC: US Government Printing Office.

Organization for Economic Cooperation and Development. 1980. Guidelines on the Protection of Privacy and Transborder Flows of Personal Data.

Oxford English Dictionary, 2nd edn. 1989. Oxford, UK: Oxford University Press.

Privacy/Data Protection Project. 2005. University of Miami School of Medicine and University of Miami Ethics Programs. http://privacy.med.miami.edu.

Privacy Protection Study Commission. 1977. Personal Privacy in an Information Society. Washington, DC: US Government Printing Office.

Prosser, W. O. 1960. Privacy. California Law Review 48(3):383–423.

Regan, P. M. 1995. Legislating Privacy: Technology, Social Values and Public Policy. Chapel Hill, NC: University of North Carolina Press.

Roberts, C. 1995. Statement before the Senate Committee on Labor and Human Resources on S.1360. Federal Document Clearinghouse Congressional Testimony, 14 November.

Room, S. 2007. Data Protection and Compliance in Context. Swindon, UK: British Computer Society.

Russell, D., and Gangemi, G. T., Sr. 1991. Computer Security Basics. Sebastopol, CA: O'Reilly and Associates.

Schneier, B. 2000. Secrets and Lies: Digital Security in a Networked World. New York, NY: Wiley Computer Publishing.

Schneier, B. 2003. Beyond Fear: Thinking Sensibly About Security in an Uncertain World. New York, NY: Copernicus Books.

Schwartz, P. 1995a. Privacy and participation: personal information and public sector regulation in the United States. Iowa Law Review 80:553–618.

Schwartz, P. 1995b. European data protection law and restrictions on international data flows. Iowa Law Review 80:471–499.

Solove, D. J. 2008. Understanding Privacy. Cambridge, MA: Harvard University Press.

van den Hoven, M. J. 1995. Information Technology and Moral Philosophy: Philosophical Explorations in Computer Ethics. Rotterdam, Netherlands: Ridderprint BV.

Waldo, J., Lin, H. S., and Millet, L. I. 2007. Engaging Privacy and Information Technology in a Digital Age. National Research Council. Washington, DC: National Academy Press.

Warren, S. D., and Brandeis, L. D. 1890. The right to privacy. Harvard Law Review 4:193.

Westin, A. F. 1977. Computers Health Records and Citizen's Rights. New York, NY: Petrocelli Books.

Workgroup for Electronic Data Interchange. 1992. Report to the Secretary of the Department of Health and Human Services. Washington, DC: US Government Printing Office.

6 Chapter 6. Searching Electronic Health Records

1. Health Level Seven Group. http://www.hl7.org/, 2007.

2. HL7 Clinical Document Architecture, Release 2.0. http://lists.hl7.org/read/attachment/61225/1/CDA-doc%20version.pdf, 2007.

3. HL7 Reference Information Model. 2007. http://www.hl7.org/library/datamodel/RIM/C30204/rim.htm.

4. U.S. National Library of Medicine. 2004. SNOMED Clinical Terms® (SNOMED CT).

5. Logical Observation Identi�ers Names and Codes (LOINC). http://www. regenstrief.org/medinformatics/loinc/, 2006.

6. RxNorm. 2007. United States National Library of Medicine. http://www.nlm. nih.gov/research/umls/rxnorm/index.html.

7. Buckley, C., Salton, G., and Allan, J. 1993. Automatic retrieval with locality information using SMART. In Proceedings of the First Text REtrieval Conference (TREC-1), pp. 59–72. NIST Special Publication, 500-207.

8. Rocchio, J. J. 1971. Relevance feedback in information retrieval. In The SMART Retrieval System—Experiments in Automatic Document Processing, G. Salton, ed., pp. 313–323, Englewood Cliffs, NJ: Prentice Hall.

9. Salton, G., ed. 1971. The SMART Retrieval System—Experiments in Automatic Document Retrieval. Englewood Cliffs, NJ: Prentice Hall.

10. Salton, G., Wong, A., and Yang, C. S. 1975. A vector space model for information retrieval. Communications of the ACM, 18(11):613–620.

11. Salton, G., and Buckley, C. 1988. Term-weighting approaches in automatic text retrieval. Information Processing and Management 24(5):513–523.

12. Robertson, S. E. 1997. The probabilistic ranking principle in IR. Journal of Documentation 33:294–304.

13. Harman, D. K. 1993. Overview of the �rst Text REtrieval Conference (TREC-1). In Proceedings of the First Text REtrieval Conference (TREC-1), pp. 1–20. NIST Special

Publication 500-207, March.

14. Singhal, A. 2001. Modern information retrieval: a brief overview. Google. IEEE Data Engineering Bulletin 24(4):35–43.

15. Salton, G., and McGill, M. J. 1983. Introduction to Modern Information Retrieval. New York, NY: McGraw Hill Book Co.

16. Savoy, J. 1992. Bayesian inference networks and spreading activation in hypertext systems. Information Processing and Management 28(3):389–406.

17. Brin, S., and Page, L. 1998. The anatomy of a large-scale hypertextual web search engine. Computer Networks 30(1–7):107–117.

18. Haveliwala, T. 2002. Topic-sensitive PageRank. In Proceedings of the 11th World Wide Web Conference (WWW), Honolulu, ACM Press.

19. Kleinberg, J. M. 1999. Authoritative sources in a hyperlinked environment. Journal of the ACM 46(5):604–632.

20. Balmin, A., Hristidis, V., and Papakonstantinou, Y. 2004. ObjectRank: authoritybased keyword search in databases. In Proceedings of the 30th international conference on Very Large Data Bases.

21. Guo, L., Shao, F., Botev, C., and Shanmugasundaram, J. 2003. XRANK: ranked keyword search over XML documents. In Proceedings of ACM SIGMOD International Conference on Management of Data.

22. Huang, A., Xue, Q., and Yang, J. 2003. TupleRank and implicit relationship discovery in relational databases. In Proceedings of the International Conference on Web-Age Information Management, pp. 445–457.

23. Cohen, S., Kanza, Y., Kogan, Y., Nutt, W., Sagiv, Y., and Serebrenik, A. 2002. EquiX: a search and query language for XML. Journal of the American Society for Information Science and Technology, 53(6):454–466.

24. Carmel, D., Maarek, Y. S., Mandelbrod, M., Mass, Y., and Soffer, A. 2003. Searching XML documents via XML fragments. In Proceedings of SIGIR, CACM, pp. 151–158.

25. Fuhr, N., and Großjohann, K. 2001. XIRQL: a query

language for information retrieval in XML documents. In Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.

26. Hristidis, V., Koudas, N., Papakonstantinou, Y., and Srivastava, D. 2006. Keyword proximity search in XML trees. IEEE Transactions on Knowledge and Data Engineering 18(4):525–539.

27. Hristidis, V., Papakonstantinou, Y., and Balmin, A. 2003. Keyword proximity search on XML graphs. In Proceedings of the 19th International Conference on Data Engineering ICDE.

28. Li, Y., Yu, C., and Jagadish, H. V. 2004. Schema-free XQuery. In Proceedings of the Conference on Very Large Data Bases Conference.

29. Xu, Y., and Papakonstantinou, Y. 2005. Efficient keyword search for smallest LCAs in XML databases. In Proceedings of SIGMOD 2005.

30. Cohen, S., Mamou, J., Kanza, Y., and Sagiv, Y. 2003. XSEarch: a semantic search engine for XML. In Proceedings of the 29th International Conference on Very Large Data Bases.

31. Cohen, S., Kanza, Y., and Kimelfeld, B. 2005. Interconnection semantics for keyword search in XML. In Proceedings of the CIKM, pp. 389–396.

31a. Li, G., Ooi, B. C., Feng, J., Wang, J., and Zhou, L. 2008. EASE: efficient and adaptive keyword search on unstructured, semi-structured and structured data. In Proceedings of the 35th SIGMOD International Conference on Management of Data, pp. 681–694.

32. Aguilera, V., Cluet, S., and Wattez, F. 2001. Xyleme query architecture. In Proceedings of the 10th World Wide Web Conference, Hong Kong.

33. Bohm, K., Aberer, K., Neuhold, E. J., and Yang, X. 1997. Structured document storage and refined declarative and navigational access mechanisms in HyperStorM. VLDB Journal 6(4):296–311.

34. Brown, L. J., Consens, M. P., Davis, I. J., Palmer, C. R., and Tompa, F. W. 1998. A structured text ADT for object-relational databases. Theory and Practice of Object

Systems 4(4):227-244.

35. Florescu, D., Kossmann, D., and Manolescu, I. 2000. Integrating keyword search into XML query processing. International Journal of Computer and Telecommunications Networking 33(1):119-135.

36. Schmidt, A., Kersten, M., and Windhouwer, M. 2001. Querying XML documents made easy: nearest concept queries. In Proceedings of the ICDE Conference.

37. Christophides, V., Abiteboul, S., Cluet, S., and Schollt, M. 1994. From structured documents to novel query facilities. In Proceedings of the 1994 ACM SIGMOD Conference, Minneapolis, MN.

38. Dao, T., Sacks-Davis, R., and Thom, J. 1997. An indexing scheme for structured documents and their implementation. Conference on Database Systems for Advanced Applications.

39. Lee, Y. K., Yoo, S. J., Yoon, K., and Berra, P. B. 1996. Index structures for structured documents. In Proceedings of the 1st ACM International Conference on Digital Libraries, pp. 91-99.

40. Agrawal, S., Chaudhuri, S., and Das, G. 2002. DBXplorer: A System for KeywordBased Search over Relational Databases. In Proceedings of the 18th International Conference on Data Engineering ICDE.

41. Hristidis, V., and Papakonstantinou, Y. 2002. DISCOVER: keyword search in relational databases. In Proceedings of the 28th international conference on Very Large Data Bases.

42. Bhalotia, G., Nakhe, C., Hulgeri, A., Chakrabarti, S., and Sudarshan, S. 2002. Keyword searching and browsing in databases using BANKS. In Proceedings of the 18th International Conference on Data Engineering ICDE.

43. Dar, S., Entin, G., Geva, S., and Palmon, E. 1998. DTL's DataSpot: database exploration using plain language. In Proceedings of the VLDB Conference, pp. 645-649.

44. Goldman, R., Shivakumar, N., Venkatasubramanian, S., and Garcia-Molina, H. 1998. Proximity search in databases. In Proceedings of the 24th International Conference on Very Large Data Bases, New York, NY, pp. 26-37.

45. Chakrabarti, S., Joshi, M., and Tawde, V. 2001. Enhanced topic distillation using text, markup, tags and hyperlinks. In Proceedings of the SIGIR Conference, pp. 208–216.

46. Barg, M., and Wong, R. 2001. Structural proximity searching for large collections of semi-structured data. In Proceedings of the 2001 ACM CIKM International Conference on Information and Knowledge Management, pp. 175–182, Atlanta, GA, USA: ACM Press.

47. Kanza, Y., and Sagiv, Y. 2001. Flexible queries over semistructured data. In Proceedings of the 20th Symposium on Principles of Database Systems, pp. 40–51, Santa Barbara, California, USA: ACM Press, May.

48. Myaeng, S. H., Jang, D. H., Kim, M. S., and Zhoo, Z. C. 1998. A flexible model for retrieval of SGML documents. In Proceedings of the SIGIR Conference.

49. Theobald, A., and Weikum, G. 2002. The index-based XXL search engine for querying XML data with relevance ranking. In Proceedings of the 8th International Conference on Extending Database Technology, EDBT 2002, pp. 477–495, Prague, Czech Republic, Springer-Verlag.

50. Luk, R., Chan, A., Dillon, T., and Leong, H. V. 2000. A survey of search engines for XML documents. In Proceedings of the SIGIR Workshop on XML and IR, Athens.

50a. Fikes, R., Hayes, P. J., and Horrocks, I. 2004. OWL-QL—A language for deductive query answering on the semantic web. Journal of Web Semantics, 2(1): 19–29.

50b. Prud'Hommeaux, E. and Seaborne, A. 2006. SPARQL Query Language for RDF, W3C Working Draft. http://www.w3.org/TR/rdf-sparq-l-query/

51. Pfeiffer, K. P., Göbel, G., and Leitner, K. 2003. Demand for intelligent search tools in medicine and health care. In Intelligent Search on XML Data, LNCS 2818, pp. 5–18. Berlin: Springer-Verlag, 2003.

52. Inokuchi, A., Takeda, K., Inaoka, N., and Wakao, F. 2007. MedTAKMI-CDI: interactive knowledge discovery for clinical decision intelligence. IBM Systems Journal 46(1):115–134.

53. Maron, M. E., and Kuhns, J. L. 1960. On relevance,

probabilistic indexing and information retrieval. Journal of the ACM 7:216–244.

54. Sparck Jones, K. 1972. A statistical interpretation of term specificity and its application in retrieval. Journal of Documentation 28:11–21.

55. Singhal, A., Buckley, C., and Mitra, M. 1996. Pivoted document length normalization. In Proceedings of ACM SIGIR'96, New York: Association for Computing Machinery, pp. 21–29.

56. Salton, G., and Buckley, C. 1999. Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science 41(4):288–297.

57. Buckley, C., Salton, G., and Allan, J. 1994. The effect of adding relevance information in a relevance feedback environment. In Proceedings of the 17th International Conference on Research and Development in Information Retrieval SIGIR.

58. Ruthven, I., and Lalmas, M. 2003. A survey on the use of relevance feedback for information access systems. The Knowledge Engineering Review 18(2):94–145.

59. Hristidis, V., Farfán, F., Burke, R. P., Rossi, A. F., and White, J. A. 2008. Challenges for information discovery on electronic medical records. In Next Generation of Data Mining, ed. H. Kargupta, J. Han, P. S. Yu, R. Motwani, and V. Kumar, Chapman & Hall/CRC Data Mining, and Knowledge Discovery Series.

60. Amer-Yahia, S., Botev, C., and Shanmugasundaram, J. 2004. TeXQuery: a fulltext search extension to XQuery. In Proceedings of the 13th International Conference on World Wide Web.

61. Xu, J., Lu, J., Wang, W., and Shi, B. 2006. Effective keyword search in XML documents based on MIU. In Proceedings of the 11th International Conference of Database Systems for Advanced Applications DASFAA.

61a. Salton, G. and Buckley, C. 1995. Optimization of relevance feedback weights. In Proceedings of ACM Special Interest Group in Information Retrieval (SIGIR).

62. Mitchell, T. M. 1997. Machine Learning. New York, NY: McGraw-Hill Higher Education.

63. Salton, G. 1989. Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer. Reading, MA: Addison Wesley.

64. Baeza-Yates, R., and Ribeiro-Neto, B. 1999. Modern Information Retrieval. New York, NY: ACM Press.

65. Hristidis, V., Gravano, L., and Papakonstantinou, Y. 2003. Efficient IR-style keyword search over relational databases. In Proceedings of the 29th International Conference on Very Large Data Bases.

66. Dong, X., Halevy, A., and Madhavan, J. 2005. Reference reconciliation in complex information spaces. In Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data.

67. Hernández, M. A., and Stolfo, S. J. 1995. The merge/purge problem for large databases. In Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data SIGMOD, 1995.

68. McCallum, A. K., Nigam, K., and Ungar, L. H. 2000. Efficient clustering of Highdimensional data sets with application to reference matching. In Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

69. McCallum, A., and Wellner, B. 2003. Toward conditional models of identity uncertainty with application to proper noun coreference. In IJCAI Workshop on Information Integration on the Web IIWEB.

70. Sarawagi, S., and Bhamidipaty, A. 2002. Interactive deduplication using active learning. In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining SIGKDD.

71. Tejada, S., Knoblock, C., and Minton, S. 2002. Learning domain-independent string transformation weights for high accuracy object identification. In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD.

72. Winkler, W. E. 1999. The State of Record Linkage and Current Research Problems. Technical report, U.S. Bureau of the Census, Washington, DC, 1999.

73. McCarthy, J. F., and Lehnert, W. G. 1995. Using decision trees for coreference resolution. In Proceedings

of the International Joint Conference on Artificial Intelligence IJCAI.

74. Ng, V., and Cardie, C. 2002. Improving machine learning approaches to coreference resolution. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics ACL.

75. Zelenko, D., Aone, C., and Richardella, A. 2003. Kernel methods for relation extraction. Journal of Machine Learning Research 3:1083–1106.

76. Elkin, P. L., Brown, S. H., Bauer, B. A. et al. 2005. A controlled trial of automated classification of negation from clinical notes. BMC Medical Informatics and Decision Making 5:13.

77. Ceusters, W., and Smith, B. 2005. Tracking referents in electronic health records. In Proceedings of the 19th International Congress of the European Federation for Medical Informatics MIE.

78. Li, Y., Yu, C., and Jagadish, H. V. 2004. Schema-Free XQuery. In Proceedings of the 30th International Conference on Very Large Data Bases.

79. World Wide Web Consortium. 2006. XQuery and XPath Full-Text Requirements. W3C Working Draft. http://www.w3.org/TR/xmlquery-full-textrequirements/.

80. Richardson, M., and Domingos, P. 2002. The intelligent surfer: probabilistic combination of link and content information in PageRank. Advances in Neural Information Processing Systems 14. Cambridge, MA: MIT Press.

81. Hristidis, V., Hwang, H., and Papakonstantinou, Y. 2008. Authority-based keyword search in databases. ACM Transactions on Database Systems (TODS) 33(1):1–40.

82. Motwani, R., and Raghavan, P. 1995. Randomized Algorithms. London: Cambridge University Press, 1995.

83. Varadarajan, R., Hristidis, V., and Raschid, L. 2008. Explaining and reformulating authority flow queries. In Proceedings of the 24th International Conference on Engineering, ICDE, pp. 883–892.

84. Uschold, N. and Gruninger, M. 1996. Ontologies: Principles, methods, and applications. The Knowledge Engineering Review, 11, 93–136.

85. Xu, J., and Croft, W. B. 1996. Query expansion using local and global document analysis. In Proceedings of the Annual ACM SIGIR International Conference on Research and Development in Information Retrieval SIGIR.

86. Wollersheim, D., and Rahayu, W. J. 2005. Using medical test collection relevance judgements to identify ontological relationships useful for query expansion. In ICDEW ’05: Proceedings of the 21st International Conference on Data Engineering Workshops, Washington, DC, USA: IEEE Computer Society, p. 1160.

87. Theobald, A. 2003. An ontology for domain-oriented semantic similarity search on XML data. In BTW, ser. LNI, G. Weikum, H. Schoning, and E. Rahm, eds., vol. 26. GI, pp. 217–226.

88. Schenkel, R., Theobald, A., and Weikum, G. 2003. Ontology-enabled XML Search. In Intelligent Search on XML Data, Ser. Lecture Notes in Computer Science, H. M. Blanken, T. Grabs, H.-J. Schek, R. Schenkel, and G. Weikum, eds., vol. 2818, pp. 119–131. Berlin: Springer.

89. Schenkel, R., Theobald, A., and Weikum, G. 2005. Semantic similarity search on semistructured data with the XXL search engine. Information Retrieval 8(4):521–545.

90. Kim, M. S., and Kong, Y.-H. 2005. Ontology-DTD matching algorithm for efficient XML query. In FSKD (2), Series Lecture Notes in Computer Science, L. Wang and Y. Jin, eds., vol. 3614. Springer, pp. 1093–1102, 2005.

91. Kim, M. S., Kong, Y.-H., and Jeon, C. W. 2006. Remote-specific XML query mobile agents. In DEECS, Series Lecture Notes in Computer Science, J. Lee, J. Shim, S. Goo Lee, C. Bussler, and S. S. Y. Shim, eds., vol. 4055. Springer, pp. 143–151, 2006.

92. Ide, N. C., Loane, R. F., and Demner-Fushman, D. 2007. Essie: a concept-based search engine for structured biomedical text. Journal of the American Medical Informatics Association 14(3):253–263.

93. Farfán, F., Hristidis, V., Ranganathan, A., and Burke, R. P. 2008. Ontology-aware search on XML-based electronic medical records. In Proceedings of the IEEE International Conference on Data Engineering ICDE, Poster paper.

94. Farfán, F., Hristidis, V., Ranganathan, A., and Weiner,

M. 2009. XOntoRank: ontology-aware search of electronic medical records. In Proceedings of the IEEE International Conference on Data Engineering (ICDE).

95. Kalpakis, K., Gada, D., and Puttagunta, V. 2001. Distance measures for effective clustering of ARIMA time-series. In Proceedings of the IEEE International Conference on Data Mining ICDM.

96. Price, S. L., Hersh, W. R., Olson, D. D., and Embi, P. J. 2002. SmartQuery: context-sensitive links to medical knowledge sources from the electronic patient record. In Proceedings of the 2002 Annual AMIA Symposium, pp. 627–631.

97. Reichert, J. C., Glasgow, M., Narus, S. P., and Clayton, P. D. 2002. Using LOINC to link an EMR to the pertinent paragraph in a structured reference knowledge base. In Proceedings of the AMIA Symposium.

97. Cimino, J. J. 1996. Linking patient information systems to bibliographic resources. Methods of Information in Medicine 35:122–126.

98. Kazai, G., Lalmas, M., and de Vries, A. P. 2004. The overlap problem in contentoriented XML retrieval evaluation. In Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval SIGIR.

99. Taylor, J. R. 1982. An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements. University Science Books, pp. 128–129.

100. Altman, D. G. and Bland, J. M. 1994. Statistics notes: Diagnostic tests 1: Sensitivity and speci�city. British Medical Journal, 308: 1552.

101. Fagin, R., Kumar, R., Mahdian, M., Sivakumar, D., and Vee, E. 2004. Comparing and aggregating rankings with ties. ACM Symposium on Principles on Database Systems (PODS ’04), pp. 47–58.

102. Fagin, R., Kumar, R. Mahdian, M., Sivakumar, D., and Vee, E. 2006. Comparing partial rankings. SIAM Journal on Discrete Mathematics 20(3):628–648.

103. Fagin, R., Kumar, R., and Sivakumar, D. 2003. Comparing Top-k lists. SIAM Journal on Discrete Mathematics 17:134–160.

7 Chapter 7. Data Mining and Knowledge Discovery on EHRs

Agrawal, R., and Srikant, R. 1994. Fast algorithms for mining association rules in large databases. In Proceedings of the 20th International Conference on Very Large Data Bases, pp. 487–499, Santiago, Chile, Morgan Kaufmann Publishers Inc.

Berndt, D. J., Hevner, A. R. et al. 2003. The CATCH data warehouse: support for community health care decision-making. Decision Support Systems 35(3):367.

Berry, M. J. A., and Linoff, G. 2004. Data mining Techniques: for Marketing, Sales, and Customer Relationship Management. Indianapolis, IN: Wiley.

Bishop, C. M. 2006. Pattern Recognition and Machine Learning. New York, NY: Springer.

Chen, D.-R., Chang, R.- F. et al. 2003. Computer-aided diagnosis for 3-dimensional breast ultrasonography. Archives of Surgery 138(3):296–302.

Chu, A., Ahn, H. et al. 2008. A decision support system to facilitate management of patients with acute gastrointestinal bleeding. Artificial Intelligence in Medicine 42(3):247–259.

Codd, E. F. 1983. A relational model of data for large shared data banks. Communications of the ACM 26(1):64–69.

Date, C. J. 2000. An Introduction to Database Systems. Boston, MA: Addison Wesley Longman.

Decoste, D., and Schölkopf, B. 2002. Training invariant support vector machines. Machine Learning Journal 46(1–3):161–190.

Deerwester, S. 1990. Indexing by latent semantic analysis. Journal of the American Society for Information Science 41(6):391–407.

Delgado, M., Sanchez, D. et al. 2001. Mining association rules with improved semantics in medical databases. Artificial Intelligence in Medicine 21(1):241–245.

Efron, B., and Tibshirani, R. J. 1993. An Introduction to the Bootstrap. London: Chapman & Hall.

Fayyad, U., Piatetsky-Shapiro, G. et al. 1996. The KDD process for extracting useful knowledge from volumes of data. Communications of the ACM 39(11):27–34.

Fletcher, R. 1987. Practical Methods of Optimization. Chichester: Wiley.

Fonarow, G. C., Adams, K. F., Jr. et al. 2005. Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis. JAMA 293(5):572–580.

Friedman, N., Geiger, D. et al. 1997. Bayesian network classifiers. Machine Learning 29:31–163.

Gerth, C., Zawadzki, R. J. et al. 2008. Retinal morphological changes of patients with x-linked retinoschisis evaluated by Fourier-domain optical coherence tomography. Archives of Ophthalmology 126(6):807–811.

Goodwin, L., Vandyne, M. et al. 2003. Data mining issues and opportunities for building nursing knowledge. Journal of Biomedical Informatics 36(4–5):379–388.

Hand, D. J. 1998. Data mining: statistics and more? The American Statistician 52(2):112.

Hastie, T., Tibshirani, R. et al. 2003. The Elements of Statistical Learning. New York, NY: Springer.

Haykin, S. S. 1999. Neural Networks: a Comprehensive Foundation. Upper Saddle River, NJ: Prentice Hall.

Hinton, G. E. 1989. Connectionist learning procedures. Artificial Intelligence in Medicine 40:185–234.

Hofmann, T. 1999. Probabilistic latent semantic indexing. In Proceedings of the 22nd Annual International SIGIR Conference on Research and Development in Information Retrieval, pp. 35–44.

Hopfield, J. J., and Tank, D. W. 1985. Computing with neural circuits: a model. Science 233:625–633.

Jain, A. K., and Dubes, R. C. 1988. Algorithms for Clustering Data. Englewood Cliffs, NJ: Prentice Hall.

Kaufman, L., and Rousseeuw, P. J. 1990. Finding Groups in Data: An Introduction to Cluster Analysis. New York, NY: John Wiley & Sons.

Kecman, V. 2001. Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models. Cambridge, MA: MIT Press.

Kimball, R., Ross, M. et al. 2008. The Data Warehouse Lifecycle Toolkit, 2nd edn. New York, NY: Wiley.

Lovell, B. C., and Walder, C. J. 2006. Support vector machines for business applications. In Business Applications and Computational Intelligence. K. Voges and N. Pope, eds. Hershey, PA: Idea Group Publishing.

Maglogiannis, I. G., and Zafiropoulos, E. P. 2004. Characterization of digital medical images utilizing support vector machines. BMC Medical Informatics and Decision Making 4(4).

Marmor, M. F., Choi, S. S. et al. 2008. Visual insignificance of the foveal pit: reassessment of foveal hypoplasia as fovea plana. Archives of Ophthalmology 126(7):907–913.

McCulloch, W. S., and Pitts, W. 1943. A logical calculus of the ideas of immanent in nervous activity. Bulletin of Mathematical Biophysics 5:115–133.

Medeiros, F. A., Zangwill, L. M. et al. 2004a. Comparison of scanning laser polarimetry using variable corneal compensation and retinal nerve fiber layer photography for detection of glaucoma. Archives of Ophthalmology 122(5):698–704.

Medeiros, F. A., Zangwill, L. M. et al. 2004b. Comparison of the GDx VCC scanning laser polarimeter, HRT II confocal scanning laser ophthalmoscope, and stratus OCT optical coherence tomograph for the detection of glaucoma. Archives of Ophthalmology 122(6):827–837.

Minsky, M. L., and Papert, S. 1969. Perceptrons. Cambridge, MA: MIT Press.

Mitchell, T. M. 1997. Machine Learning. New York, NY: McGraw-Hill.

Morales, D. A., Bengoetxea, E. et al. 2008. Selection of human embryos for transfer by Bayesian classifiers. Computers in Biology and Medicine 38(11–12):1177–1186.

Ordonez, C., Santana, C. A. et al. 2000. Discovering interesting association rules in medical data. In ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery.

Osuna, E., Freund, R. et al. 1997. Training support vector machines: an application to face detection. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'97).

Quinlan, J. R. 1993. C4.5: Programs for Machine Learning. San Mateo, CA: Morgan Kaufmann.

Rosenblatt, F. 1962. Principles of Neurodynamics. Washington, DC: Spartan Books.

Shannon, C. E. 1948. A mathematical theory of communication. Bell System Technical Journal 27:379–423, 623–656.

Smalheiser, N., Torvik, V. et al. 2006. Collaborative development of the Arrowsmith two node search interface designed for laboratory investigators. Journal of Biomedical Discovery and Collaboration 1(1):8.

Smalheiser, N. R., and Swanson, D. R. 1994. Assessing a gap in the biomedical literature: magnesium deficiency and neurologic disease. Neuroscience Research Commun 15:1–9.

Smalheiser, N. R., and Swanson, D. R. 1996a. Indomethacin and Alzheimer's disease. Neurology 46:583.

Smalheiser, N. R., and Swanson, D. R. 1996b. Linking estrogen to Alzheimer's disease: an informatics approach. Neurology 47:809–810.

Smalheiser, N. R., and Swanson, D. R. 1998a. Calcium-independent phospholipase A 2 and schizophrenia. Archives of Genenal Psychiatry 55:752–753.

Smalheiser, N. R., and Swanson, D. R. 1998b. Using ARROWSMITH: a computerassisted approach to formulating and assessing scientific hypotheses. Computer Methods and Programs in Biomedicine 57:149–153.

Swanson, D. R. 1986a. Fish oil, Raynaud's syndrome, and undiscovered public knowledge. Perspectives in Biology and Medicine 30:7–18.

Swanson, D. R. 1986b. Undiscovered public knowledge. Library Quarterly 56:103–118.

Swanson, D. R. 1988. Migraine and magnesium: eleven neglected connections. Perspectives in Biology and Medicine 31:526–557.

Swanson, D. R. 1993. Intervening in the life cycles of scientific knowledge. Library Trends 41:606–631.

Swanson, D. R., and Smalheiser, N. R. 1997. An interactive system for finding complementary literatures: a stimulus to scientific discovery. Artificial Intelligence 91:183–203.

Swanson, D. R., and Smalheiser, N. R. 1999. Implicit text linkages between Medline records: using Arrowsmith as an aid to scientific discovery. Library Trends 48:48–59.

Tangri, N., Ansell, D. et al. 2008. Predicting technique survival in peritoneal dialysis patients: comparing artificial neural networks and logistic regression. Nephrology Dialysis Transplantation 23(9):2972–2981.

Tremblay, M. C., Berndt, D. et al. 2005. Utilizing text mining techniques to identify fall related injuries. In Proceedings of the 11th Americas Conference on Information Systems (AMCIS 2005), Omaha, NE.

Tremblay, M. C., Berndt, D. J. et al. 2006. Feature selection for predicting surgical outcomes. In Proceedings of the 29th Annual Hawaii International Conference on System Sciences, Hawaii.

Tremblay, M. C., Fuller, R. et al. 2007. Doing more with more information: changing healthcare planning with OLAP tools. Decision Support Systems 43(4):1305–1320.

Uzuner, Ö., Sibanda, T. C. et al. 2008. A de-identifier for medical discharge summaries. Artificial Intelligence in Medicine 42(1):13–35.

Vapnik, V. N. 1996. The Nature of Statistical Learning Theory. New York, NY: SpringerVerlag.

Viveros, M. S., Nearhos, J. P. et al. 1996. Applying data mining techniques to a health insurance information system. In Proceedings of the 22th International Conference on Very Large Data Bases, Morgan Kaufmann Publishers.

Weiss, S. M., and Indurkhya, N. 1998. Predictive Data Mining: A Practical Guide. San Francisco, CA: Morgan Kaufmann.

Witten, I. H., and Frank, E. 2005. Data Mining: Practical Machine Learning Tools and Techniques. San Francisco, CA: Morgan Kaufmann.

Yang, Y., and Chute, C. G. 1993. An application of least squares fit mapping to text information retrieval. In Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Pittsburgh, PA: ACM.

Zhang, J., Z. Ghahramani, et al. 2008. Flexible latent variable models for multitask learning. Machine Learning 73(3):221-242.

1. Sweeney, L. 2002. k-Anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-Based Systems 10(5):557–570.

2. Machanavajjhala, A., Gehrke, J., Kifer, D., and Venkitasubramaniam, M. 2006. l-Diversity: privacy beyond k-anonymity. In Proceedings of the IEEE International Conference on Data Engineering.

3. Truta, T. M., and Vinay, B. 2006. Privacy protection: p-sensitive k-anonymity property. In Proceedings of the ICDE Workshops.

4. Aggarwal, G., Feder, T., Kenthapadi, K., Khuller, S., Panigrahy, R., Thomas, D., and Zhu, A. 2006. Achieving anonymity via clustering. In Proceedings of PODS, pp. 153–162.

5. Li, N., and Li, T. 2007. t-Closeness: privacy beyond k-anonymity and l- diversity. In Proceedings of the 23rd IEEE International Conference on Data Engineering (ICDE).

6. Xiao, X., and Tao, Y. 2007. M-invariance: towards privacy preserving republication of dynamic datasets. In Proceedings of the SIGMOD Conference, pp. 689–700.

7. Martin, D. J., Kifer, D., Machanavajjhala, A., Gehrke, J., and Halpern, J. Y. 2007. Worst-case background knowledge for privacy-preserving data publishing. In Proceedings of the ICDE.

8. Nergiz, M. E., Atzori, M., and Clifton, C. 2007. Hiding the presence of individuals from shared databases. In Proceedings of the ACM SIGMOD.

9. Jajodia, S., and Sandhu, R. 1991. Toward a multilevel secure relational data model. In Proceedings of the ACM SIGMOD.

10. Agrawal, R., Kieman, J., Srikant, R., and Xu, Y. 2002. Hippocratic databases. In Proceedings of the 28th International Conference on Very Large Data Bases.

11. LeFevre, K., Agrawal, R., Ercegovac, V., Ramakrishnan, R., Xu, Y., and DeWitt, D. 2004. Limiting disclosure in Hippocratic databases. In Proceedings of the 30th International Conference on Very Large Data Bases.

12. Agrawal, R., Bird, P., Grandison, T., Kieman, J., Logan, S., and Rjaibi, W. 2005. Extending relational database systems to automatically enforce privacy policies. In Proceedings of the 21st IEEE International Conference on Data Engineering.

13. Byun, J.-W., Bertino, E., and Li, N. 2005. Purpose based access control of complex data for privacy protection. In ACM Symposium on Access Control Models, and Technologies (SACMAT).

14. Byun, J., and Bertino, E. 2006. Micro-views, or on how to protect privacy while enhancing data usability—concept and challenges. SIGMOD Record 35(1): 9–13.

15. Adams, N. R., and Wortman, J. C. 1989. Security-control methods for statistical databases: a comparative study. ACM Computing Surveys 21(4):515–556.

16. Fung, B. C. M., Wang, K., Chen, R., and Yu, P. S. 2010. Privacy-preserving data publishing: a survey on recent developments. ACM Computing Surveys, 42 (4).

17. Iyengar, V. S. 2002. Transforming data to satisfy privacy constraints. In Proceedings of the 8th ACM International Conference on Knowledge Discovery and Data, pp. 279–288.

18. Wang, K., Yu, P. S., and Chakraborty, S. 2004. Bottom-up generalization: a data mining solution to privacy protection. In Proceedings of ICDM.

19. Meyerson, A., and Williams, R. 2004. On the complexity of optimal k- anonymity. In Proceedings of ACM Symposium on Principles of Database Systems, pp. 223–228.

20. Bayardo, R. J., and Agrawal, R. 2005. Data privacy through optimal k-anonymization. In Proceedings of the ICDE 2005.

21. Aggarwal, C. C. 2005. On k-anonymity and the curse of dimensionality. In Proceedings of the 31st International Conference Very Large Data Bases, pp. 901–909.

22. Fung, B. C. M., Wang, K., and Yu, P. S. 2005. Top-down specialization for information and privacy preservation. In Proceedings of ICDE.

23. Bertino, E., Ooi, B. C., Yang, Y., and Deng, R. H.

2005. Privacy and ownership preserving of outsourced medical data. In Proceedings of the ICDE 2005.

24. Zhong, S., Yang, Z., and Wright, R. N. 2005. Privacy-enhancing k-anonymization of customer data. In Proceedings of PODS, pp. 139–147.

25. LeFevre, K., Dewitt, D., and Ramakrishnan, R. 2005. Incognito: efficient fulldomain k-anonymity. In Proceedings of the ACM SIGMOD International Conference on Management of Data.

26. LeFevre, K., Dewitt, D., and Ramakrishnan, R. 2006. Mondrian multidimensional k-anonymity. In Proceedings of the IEEE ICDE.

27. LeFevre, K., Dewitt, D., and Ramakrishnan, R. 2006. Workload-aware anonymization. In Proceedings of the ACM SIGKDD.

28. Wang, K., and Fung, B. C. M. 2006. Anonymizing sequential releases. In Proceedings of the ACM SIGKDD.

29. Kifer, D., and Gehrke, J. 2006. Injecting utility into anonymized datasets. In Proceedings of the SIGMOD Conference, pp. 217–228.

30. Xiao, X., and Tao, Y. 2006. Anatomy: simple and effective privacy preservation. In Proceedings of the 32nd International Conference on Very Large Data Bases, pp. 139–150.

31. Zhang, Q., Koudas, N., Srivastava, D., and Yu, T. 2007. Aggregate query answering on anonymized tables. In Proceedings of ICDE.

32. Sweeney, L. 1996. Replacing personally-identifying information in medical records, the scrub system. Journal of the American Informatics Association, 333–337.

33. Sweeney, L. 1997. Guaranteeing anonymity when sharing medical data, the datafly system. In Proceedings of AMIA Annual Fall Symposium.

34. Thomas, S. M., Mamlin, B., Schadow, G., and McDonald, C. 2002. A successful technique for removing names in pathology reports. In Proceedings of the Annual AMIA Symposium, pp. 777–781.

36. Gupta, D., Saul, M., and Gilbertson, J. 2004.

Evaluation of a de-identification (De-id) software engine to share pathology reports and clinical documents for research. American Journal of Clinical Pathology 121(2):176–186.

37. Sibanda, T., and Uzuner, O. 2006. Role of local context in de-identification of ungrammatical fragmented text. In Proceedings of the North American Chapter of Association for Computational Linguistics/Human Language Technology.

38. Beckwith, B. A., Mahaadevan, R., Balis, U. J., and Kuo, F. 2006. Development and evaluation of an open source software tool for de-identification of pathology reports. BMC Medical Informatics and Decision Making 42(1):13–35.

39. Agrawal, R., and Srikant, R. 2000. Privacy-preserving data mining. In Proceedings of the ACM SIGMOD Conference on Management of Data, pp. 439–450.

40. Verykios, V. S., Bertino, E., Fovino, I.N., Provenza, L. P., Saygin, Y., and Theodoridis, Y. 2004. State-of-the-art in privacy preserving data mining. ACM SIGMOD Record 33(1):50–57.

41. Bu, S., Lakshmanan, L. V. S., Ng, R. T., and Ramesh, G. 2007. Preservation of patterns and input-output privacy. In Proceedings of the ICDE.

42. Rizvi, S., and Haritsa, J. R. 2002. Maintaining data privacy in association rule mining. In Proceedings of the 28th International Conference on Very Large Data Bases, pp. 682–693.

43. Evfimievski, A. V., Gehrke, J., and Srikant, R. 2003. Limiting privacy breaches in privacy preserving data mining. In Proceedings of PODS, pp. 211–222.

44. Evfimievski, A. V., Srikant, R., Agrawal, R., and Gehrke, J. 2004. Privacy preserving mining of association rules. Information Systems 29(4):343–364.

45. Kargupta, H., Datta, S., Wang, Q., and Sivakumar, K. 2003. On the privacy preserving properties of random data perturbation techniques. In Proceedings of the ICDM.

46. Huang, Z., Du, W., and Chen, B. 2005. Deriving private information from randomized data. In Proceedings of the SIGMOD Conference, pp. 37–48.

47. Teng, Z., and Du, W. 2006. Comparisons of

k-anonymization and randomization schemes under linking attacks. In Proceedings of the ICDM.

48. Lindell, Y., and Pinkas, B. 2002. Privacy preserving data mining. Journal of Cryptology 15(3):177–206.

49. Vaidya, J., and Clifton, C. 2002. Privacy preserving association rule mining in vertically partitioned data. In Proceedings of the 8th ACM International Conference on Knowledge Discovery and Data Mining.

50. Kantarcioglu, M., and Clifton, C. 2004. Privacy preserving data mining of association rules on horizontally partitioned data. IEEE Transactions on Knowledge and Data Engineering 16(9):1026–1037.

51. Kantarcoglu, M., and Vaidya, J. 2003. Privacy preserving naive Bayes classifier for horizontally partitioned data. In Proceedings of the ICDM Workshop on Privacy Preserving Data Mining.

52. Vaidya, J., and Clifton, C. 2003. Privacy preserving naive Bayes classifier for vertically partitioned data. In Proceedings of the 9th ACM International Conference on Knowledge Discovery and Data Mining.

53. Yang, Z., Zhong, S., and Wright, R. N. 2005. Privacy-preserving classification of customer data without loss of accuracy. In Proceedings of SIAM SDM.

54. Vaidya, J., and Clifton, C. 2003. Privacy-preserving k-means clustering over vertically partitioned data. In Proceedings of SIGKDD.

55. Aggarwal, G., Mishra, N., and Pinkas, B. 2004. Secure computation of the kth ranked element. In Proceedings of the IACR Conference on Eurocrypt.

56. Agrawal, R., Evfimievski, A., and Srikant, R. 2003. Information sharing across private databases. In Proceedings of SIGMOD.

57. Vaidya, J., and Clifton, C. 2005. Privacy-preserving top-k queries. In Proceedings of ICDE.

58. Kantarcioglu, M., and Clifton, C. 2005. Privacy preserving k-NN classifier. In Proceedings of the ICDE.

59. Xiong, L., Chitti, S., and Liu, L. 2005. Top-k queries across multiple private databases. In Proceedings of the

25th International Conference on Distributed Computing Systems (ICDCS 2005).

60. Bhowmick, S. S., Gruenwald, L., Iwaihara, M., and Chatvichienchai, S. 2006. Private-iye: A framework for privacy preserving data integration. In Proceedings of the ICDE Workshops.

61. Xiong, L., Chitti, S., and Liu, L. 2007. Mining multiple private databases using a knn classi�er. In Proceedings of the ACM Symposium on Applied Computing (SAC), pp. 435–440.

62. Xiong, L., Chitti, S., and Liu, L. 2007. Preserving data privacy for outsourcing data aggregation services. ACM Transactions on Internet Technology (TOIT) 7(3):17.

63. Xiao, X., and Tao, Y. 2006. Personalized privacy preservation. In Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data.

64. Gardner, J., and Xiong, L. 2008. HIDE: an integrated system for health information de-identi�cation. In Proceedings of the IEEE CBMS.

65. Gardner, J., and Xiong, L. 2009. An integrated framework for anonymizing unstructured medical data. Data and Knowledge Engineering (DKE), in press. Available online at doi:10.1016/j.datak.2009.07.006.

66. Ruch, P., Baud, R. H., Rassinoux, A. M., Bouillon, P., and Robert, G. 2000. Medical document anonymization with a semantic lexicon. In Proceedings of the AMIA Symposium.

67. Berman, J. J. 2003. Concept-match medical data scrubbing: how pathology text can be used in research. Archives of Pathology and Laboratory Medicine 127(66):680–686.

68. Douglass, M., Clifford, G. D., Reisner, A., Long, W. J., Moody, G. B., and Mark, R. G. 2005. De-identi�cation algorithm for free-text nursing notes. Computers in Cardiology 32:331–334.

69. Lafferty, J., McCallum, A., and Pereira, F. 2001. Conditional random �elds: probabilistic models for segmenting and labeling sequence data. In Proceedings of the 18th International Conference on Machine Learning.

70. Goldwasser, S. 1997. Multi-party computations: past and

present. In Proceedings of the ACM Symposium on Principles of Distributed Computing.

71. Goldreich, O. 2001. Secure multi-party computation. Working Draft, Version 1.3.

72. Jiang, W., and Clifton, C. 2006. A secure distributed framework for achieving k-anonymity. The VLDB Journal 15(4):316–333.

73. Zhong, S., Yang, Z., and Wright, R. N. 2005. Privacy-enhancing k-anonymization of customer data. In Proceedings of the 24th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, pp. 139–147. New York, NY: ACM Press.

74. Jurczyk, P., and Xiong, L. 2009. Distributed anonymization: achieving anonymity for both data subjects and data providers. In Annual IFIP WG 11.3 Working Conference on Data and Applications Security (DBSec).

75. Aggarwal, C., and Yu, P. S., eds. 2008. Privacy-Preserving Data Mining: Models and Algorithms. New York, NY: Springer.

76. Clifton, C., Kantarcioglu, M., Lin, X., Vaidya, J., and Zhu, M. 2003. Tools for privacy preserving distributed data mining. In Proceedings of SIGKDD Explorations.

77. Wang, K., Fung, B. C. M., and Dong, G. 2005. Integrating private databases for data analysis. In Proceedings of the IEEE ISI.

78. Kargupta, H., Datta, S., Wang, Q., and Sivakumar, K. 2003. On the privacy preserving properties of random data perturbation techniques. In Proceedings of the IEEE International Conference on Data Mining, p. 99, Melbourne, FL, November.

79. Liu, K., Kargupta, H., and Ryan, J. 2006. Random projection-based multiplicative data perturbation for privacy preserving distributed data mining. IEEE Transactions on Knowledge and Data Engineering 18(1):92–106.

80. Uzuner, O., Szolovits, P., and Kohane, I. 2006. i2b2 Workshop on natural language processing challenges for clinical records. In Fall Symposium of the American Medical Informatics Association (AMIA).

9 Chapter 9. Real-Time and Mobile Physiological Data Analysis

1. Varshney, U. 2006. Managing wireless health monitoring for patients with disabilities. IT Professional 8(6):12–16.

2. Podgorelec Vili et al., 2005. Some applications of intelligent systems in medicine. In Proceedings of the 3rd IEEE International Conference on Computational Cybernetics, Budapest, Hungary.

3. Varshney, U., and Sneha, S. 2006. Patient monitoring using ad hoc wireless networks: reliability and power management. IEEE Communications Magazine 1:63–68.

4. Wu, H.-C. et al. 1999. A mobile system for real-time patient-monitoring with integrated physiological signal processing. In Proceedings of the 1st Joint BMES/ EMBS Conference, p. 712.

5. Lin, Y.-H. et al. 2004. A wireless PDA-based physiological monitoring system for patient transport. IEEE Transactions on Information Technology in Biomedicine 8(4):439–447.

6. Lee, R.-G., Chen, K.-C., Hsiao, C.-C., and Tseng, C.-L. 2007. A mobile care system with alert mechanism. IEEE Transactions on Information Technology in Biomedicine, 11(5): 507–517.

7. Saranummi, N. 2002. Information technology in biomedicine. IEEE Transactions on Biomedical Engineering 49(12):1385–1386.

8. Skyaid Watch. Available at: http://Tinyurl.com/MXNTC9htm. (Accessed Sept. 23, 2009).

9. Lorincz, K. et al. 2004. Sensor networks for emergency response: challenges and opportunities. IEEE Pervasive Computing 3(4):16–23.

10. Jones, V. et al. 2006. Mobihealth: mobile health services based on body area networks. Technical Report TR-CTIT-06-37 Centre for Telematics and Information Technology, University of Twente, Enschede.

11. Karantonis, D. M., Narayanan, M. R., Mathie, M., Lovell, N. H., and Celler, B. G. 2006. Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. IEEE Transactions

on Information Technology in Biomedicine 10:156–167.

12. Varady, P., Benyo, Z., and Benyo, B. 2002. An open architecture patient monitoring system using standard technologies. IEEE Transactions on Information Technology in Biomedicine 6:95–98.

13. Sharshar, S., Allart, L., and Chambrin. M. C. 2005. A new approach to the abstraction of monitoring data in intensive care. Lecture Notes in Computer Science 3581:13–22.

14. Axisa, F., Dittimar, A., and Delhomme, G. 2003. Smart clothes for the monitoring in real time and conditions of physiological, emotional and sensorial reaction of human. In Proceedings of the 25th Conference IEEE Engineering Medicine and Biology Society, pp. 3744–3747.

15. Cheng, P.-T., Tsai, L.-M., Lu, L.-W., and Yang, D.-L. 2004. The design of PDAbased biomedical data processing and analysis for intelligent wearable health monitoring systems. In Proceedings of the 4th Conference on Computer and Information Technology.

16. Branche, P., and Mendelson, Y. 2005. Signal quality and power consumption of a new prototype re�ectance pulse oximeter sensor. In Proceedings of the 31st Northeast Bioengineering Conference, pp. 42–43.

17. Weber, J. L. et al. 2004. Telemonitoring of vital parameters with newly designed biomedical clothing. Studies in Health Technology and Informatics 108:260–265.

18. Manders, E., and Dawant, B. 1996. Data acquisition for an intelligent bedside monitoring system. In Proceedings of the 18th Conference IEEE Engineering in Medicine and Biology Society, 1987–1988.

19. Gupta, S., and Ganz, A. 2004. Design considerations and implementation of a cost-effective, portable remote monitoring unit using 3G wireless data networks. In Proceedings of the 26th Conference IEEE Engineering in Medicine and Biology Society, pp. 3286–3289.

20. Ermes, M., Pärkkä, J., Mäntyjärvi, J., and Korhonen, I. 2008. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. IEEE Transactions on Information Technology in Biomedicine 12(1):20–26.

21. Iakovidis, I. 1998. Towards personal health record: current situation, obstacles and trends in implementation of electronic healthcare records in Europe. International Journal of Medical Informatics 52(128):105 –117.

22. Apiletti, D., Baralis, E., Bruno, G., and Cerquitelli, T. 2008. Real-time individuation of global unsafe anomalies and alarm activation. In Intelligent Techniques and Tools for Novel System Architectures, Vol. 109, P. Chountas, I. Petrounias, and J. Kacprzyk, Eds., Studies in Computational Intelligence. Springer Verlag. ISBN: 978-3-540-77621-5.

23. Anliker, U., et al. 2004. AMON: a wearable multiparameter medical monitoring and alert system. IEEE Transactions on Information Technology in Biomedicine 8(4):415–427.

24. IEEE 802.15 WPAN Task Group 1. Available at: http://www.ieee802.org/15/ pub/TG1.html. (Accessed Sept. 23, 2009).

25. ZigBee Alliance. Available at: http://www.zigbee.org/.

26. IEEE 802.11 Wireless Local Area Networks. Available at: http://ieee802.org/11/. (Accessed Sept. 23, 2009).

27. ITU Recommendation (standard) ITU-T G.992.1. Available at: http://www.itu. int/rec/T-REC-G.992.1-200207-I!Cor2/en. (Accessed Sept. 23, 2009).

28. Global System for Mobile communications. Available at: http://www. gsmworld.com/. (Accessed Sept. 23, 2009).

29. GSM Facts and Figures. Available at: http://www.gsmworld.com/news/ newsroom/marketdata/marketdata_summary.htm (Accessed Sept. 23, 2009).

30. Technical Speci�cations and Technical Reports for a UTRAN-based 3GPP system. Available at: http://www.3gpp.org/ftp/Specs/html-info/21101.htm. (Accessed Sept. 23, 2009).

31. Aminian, K., Naja�, B., Bla, C., Leyvraz, P. F., and Robert, P. 2001. Ambulatory gait analysis using gyroscopes. 25th Annual Meeting of the American Society of Biomechanics, San Diego, CA.

32. Pentland, A. 2004. Healthwear: medical technology

becomes wearable. Computer 37(5):42–49.

33. Lee, S.-W., and Mase, K. 2002. Activity and location recognition using wearable sensors. Pervasive Computing, IEEE 1(3):24–32.

34. Apiletti, D., Baralis, E., Bruno, G., and Cerquitelli, T. 2007. SAPhyRA: stream analysis for physiological risk assessment. In Proceedings of the IEEE CBMS, pp. 193–198.

35. Bao, L., and Intille, S. S. 2004. Activity recognition from user-annotated acceleration data. In Proceedings of the 2nd International Conference on Pervasive Computing, pp. 1–17.

36. Ravi, N., Dandekar, N., Mysore, P., and Littman, M. L., 2005. Activity recognition from accelerometer data. In Proceedings of the National Conference on Artificial Intelligence, American Association for Artificial Intelligence, pp. 1541–1546.

37. Maurer, U., Smailagic, A., Siewiorek , D. P., and Deisher, M. 2006. Activity recognition and monitoring using multiple sensors on different body positions. In Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks (BSN'06).

38. Yang, J.-Y., Chen, Y.-P., Lee, G.-Y., Liou, S.-N., and Wang, J.-S. 2007. Activity recognition using one triaxial accelerometer: a neuro-fuzzy classifier with feature reduction. In Proceedings of the ICEC 2007, LNCS 4740, pp. 395–400.

39. Tapia, E. M., Intille, S. S., Haskell, W., Larson, K., Wright, J., King, A., and Friedman, R. 2007. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. In Proceedings of the International Symposium on Wearable Computers, IEEE Press, Boston, MA.

40. Lee, R.-G., Chen, K.-C., Hsiao, C.-C., and Tseng, C.-L. 2007. A mobile care system with alert mechanism. IEEE Transactions on Information Technology in Biomedicine 11(5):507–517.

41. Dai, S., and Zhang, Y. 2006. A wireless physiological multi-parameter monitoring system based on mobile communication networks. In Proceedings of tthe IEEE CBMS'06.

42. Rodriguez, J., Goi, A., and Illarramendi, A. 2005. Real-time classiﬁcation of ECGs on a PDA. IEEE Transactions on Information Technology in Biomedicine 9(1):23–34.

43. Salvador, C. H., Carrasco, M. P., de Mingo, M. A. G., Carrero, A. M., Montes, J. M., Martin, L. S., Cavero, M. A., Lozano, I. F., and Monteagudo, J. L. 2005. Airmed-cardio: a GSM and Internet services-based system for out-of-hospital follow-up of cardiac patients. IEEE Transactions on Information Technology in Biomedicine 9(1):73–85.

44. Lorenz, A., and Oppermann, R., In press. Mobile health monitoring for elderly: design for diversity. Pervasive and Mobile Computing.

45. Corchado, J. M., Bajo, J., de Paz, Y., and Tapia, D. I. 2008. Intelligent environment for monitoring Alzheimer patients, agent technology for health care. Decision Support Systems 44(2):382–396.

46. Imhoff, M., and Kuhls, S. 2006. Alarm algorithms in critical care monitoring. Anesthesia & Analgesia 102:1525–1537.

47. Adhikari, N., and Lapinsky, S. E. 2003. Medical informatics in the intensive care unit: overview of technology assessment. Journal of Critical Care 18:41–47.

48. Wu, W. H., Bui, A. A. T., Batalin, M. A., Liu, D., and Kaiser, W. J. 2007. Incremental diagnosis method for intelligent wearable sensor systems. IEEE Transactions on Information Technology in Biomedicine 11(5):553–562.

49. Alves, J. B. M., da Silva, J. B., and Paladini, S. 2006. A low cost model for patient monitoring in intensive care unit using a micro web-server. IADIS Virtual Multi Conference on Computer Science and Information Systems, MCCSIS 2006.

50. Han, J., and Kamber, M. 2000. Data mining: concepts and techniques. Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann.

51. MIMICDB. Available at: http://www.physionet.org/physiobank/database/ mimicdb.

52. Shahar, Y., and Musen, M. 1996. Knowledge-based temporal abstraction in clinical domains. Artificial

Intelligence in Medicine 8(3):267-298.

53. Varshney, U., and Sneha, S. 2006. Patient monitoring using ad hoc wireless networks: reliability and power management. IEEE Communications Magazine 44:49–55.

1. Shapiro, L. G., and Stockman, G. C. 2001. Computer Vision. Upper Saddle River, NJ: Prentice Hall.

2. Adams, R., and Bischof, L. 1994. Seeded region growing. IEEE Transactions on Pattern Analysis and Machine Intelligence 16(6):641-647.

3. Canny, J. 1986. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 8:679-714.

4. Manjunath, B., and Chellapa, R. 1991. Unsupervised texture segmentation using Markov random field models. IEEE Transactions on Pattern Analysis and Machine Intelligence 13:478-482.

5. Kass, M., Witkin, A., and Terzopoulos, D. 1987. Snakes: active contour models. International Journal of Computer Vision 1:321-331.

6. Mumford, D., and Shah, J. 1989. Optimal approximations by piecewise smooth functions and associated variational problems. Communications on Pure and Applied Mathematics 42(5):577-685.

7. Caselles, V., Kimmel, R., and Sapiro, G. 1995. Geodesic active contours. IEEE International Conference on Computer Vision.

8. Malladi, R., Sethian, J., and Vemuri, B. 1995. Shape modeling with front propagation: a level set approach. IEEE Transactions on Pattern Analysis and Machine Intelligence 17(2):158-175.

9. Shi, J., and Malik, J. 2000. Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(8):888-905.

10. Comaniciu, D., and Meer, P. 2002. Mean shift: a robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(5):603-619.

11. Li, S. Z. 1995. Markov Random Field Modeling in Computer Vision. London, UK: Springer-Verlag.

12. Staib, L. H., and Duncan, J. S.. Boundary finding with parametrically deformable models. IEEE Transactions on

Pattern Analysis and Machine Intelligence
14(11):1061–1075.

13. Metaxas, D. 1996. Physics-Based Deformable Models. Norwell, MA: Kluwer Academic Publishers.

14. Huang, X., and Metaxas, D. 2008. Metamorphs: Deformable shape and appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(8): 1444–1459.

15. Yezzi, A. J., Tsai, A., and Willsky, A. 1999. A statistical approach to snakes for bimodal and trimodal imagery. Proceedings of IEEE International Conference on Computer Vision 2:898–903.

16. Vese, L. A., and Chan, T. F. 2002. A multiphase level set framework for image segmentation using the Mumford and Shah model. International Journal of Computer Vision 50(3):271–293.

17. Zhu, S., and Yuille, A. 1996. Region competition: unifying snakes, region growing, and Bayes/MDL for multi-band image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 18(9):884–900.

18. Cohen, L. D., and Cohen, I. 1993. Finite-element methods for active contour models and balloons for 2-D and 3-D Images. IEEE Transactions on Pattern Analysis and Machine Intelligence 15:1131–1147.

19. McInerney, T., and Terzopoulos, D. 1996. Deformable models in medical image analysis: a survey. Medical Image Analysis 1(2):91–108.

20. Xu, C., and Prince, J. L. 1998. Snakes, shapes and gradient vector flow. IEEE Transactions on Image Processing 7(3):359–369.

21. Ronfard, R. 1994. Region-based strategies for active contour models. International Journal of Computer Vision 13(2):229–251.

22. Huang, R., Pavlovic, V., and Metaxas, D. 2004. A graphical model framework for coupling MRFs and deformable models. IEEE Conference on Computer Vision and Pattern Recognition 2:739–746.

23. Florin, C., Williams, J., and Paragios, N. 2006. Globally optimal active contours, sequential Monte Carlo and on-line learning for vessel segmentation. European

Conference on Computer Vision.

24. Huang, D., Swanson, E. A., Lin, C. P., Schuman, J. S., Stinson, W. G., Chang, W., Hee, M. R., Flotte, T., Gregory, K., Puliafito, C. A., and Fujimoto, J. G. 1991. Optical coherence tomography. Science 254:1178–1181.

25. Paragios, N., and Deriche, R. 2002. Geodesic Active regions and level set methods for supervised texture segmentation. International Journal of Computer Vision 46(3):223–247.

26. Tsechpenakis, G., Lujan, B., Martinez, O., Gregori, G., and Rosenfeld, P. J. 2008. Geometric deformable model driven by CoCRFs: application to optical coherence tomography. International Conference on Medical Image Computing and Computer Assisted Intervention, New York, NY, September.

27. Jiao, S., Knighton, R., Huang, X., Gregori, G., and Puliafito, C. 2005. Simultaneous acquisition of sectional and fundus ophthalmic images with spectral-domain optical coherence tomography. Optics Express 13(2):444–452.

28. Smith, W., Assink, J., Klein, R., Mitchell, P., Klaver, C. C., Klein, B. E., Hofman, A., Jensen, S., Wang, J. J., and de Jong, P. T. 2001. Risk factors for age-related macular degeneration: pooled findings from three continents. Ophthalmology 108(4):697–704.

29. Jones, T., and Metaxas, D. 1997. Automated 3D segmentation using deformable models and fuzzy affinity. Proceedings of the 15th International Conference on Information Processing in Medical Imaging, pp. 113–126.

30. Chen, T., and Metaxas, D. 2000. Image segmentation based on the integration of markov random fields and deformable models. Proceedings of the International Conference on Medical Imaging Computing and Computer-Assisted Intervention, pp. 256–265.

31. Samson, C., Blanc-Feraud, L., Aubert, G., and Zerubia, J. 2000. A level set model for image classification. International Journal of Computer Vision 40(3):187–198.

32. Chan, T., and Vese, L. 2001. Active contours without edges. IEEE Transactions on Image Processing 10(2):266–277.

33. Huang, R., Pavlovic, V., and Metaxas, D. 2006. A

tightly coupled region-shape framework for 3D medical image segmentation. Proceedings of the 2006 IEEE International Symposium on Biomedical Imaging, Arlington, VA, pp. 2121–2124.

34. Lafferty, J., McCallum, A., and Pereira, F. 2001. Conditional random fields: probabilistic models for segmenting and labeling sequence data. International Conference on Machine Learning.

35. Kumar, S., and Hebert, M. 2004. Discriminative fields for modeling spatial dependencies in natural images, Advances in Neural Information Processing Systems 16:1351–1358.

36. He, X., Zemel, R., and Carreira-Perpinan, M. 2004. Multiscale conditional random fields for image labelling. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 695–702.

37. Tsechpenakis, G., and Metaxas, D. 2007. CRF-driven implicit deformable model. IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8.

38. Tsechpenakis, G., Wang, J., Mayer, B., and Metaxas, D. 2007. Coupling CRFs and deformable models for 3D medical image segmentation. IEEE Mathematical Methods in Biomedical Image Analysis, IEEE International Conference on Computer Vision, Rio de Janeiro, Brazil.

39. Tsechpenakis, G., and Wang, J. 2007. CRF-based segmentation of human tear meniscus obtained with optical coherence tomography. IEEE International Conference on Image Processing, pp. 509–512.

40. Osher, S., and Sethian, J. 1988. Fronts propagating with curvature- dependent speed: algorithms based on the Hamilton-Jacobi formulation. Journal of Computational Physics 79:12–49.

41. Sederberg, T. W., and Parry, S. R. 1986. Free-form deformation of solid geometric models. In Proceedings of the 13th Annual Conference on Computer Graphics, pp. 151–160.

42. Faloutsos, P., van de Panne, M., and Terzopoulos, D. 1997. Dynamic free-form deformations for animation synthesis. IEEE Transactions on Visualization and Computer Graphics 3:201–214.

43. Huang, X., Paragios, N., and Metaxas, D. 2006. Shape registration in implicit spaces using information theory and free form deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(8):1303–1318.

44. Akgul, Y. S., and Kambhamettu, C. 2003. A coarse-to-fine deformable contour optimization framework. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(2):174–186.

45. Duda, R. O., and Hart, P. 1973. Pattern Classification and Scene Analysis. New York, NY: Wiley.

46. Elgammal, A., Duraiswami, R., and Davis, L. S. 2003. Efficient kernel density estimation using the fast Gauss transform with applications to color modeling and tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(11):1499–1504.

47. Haralick, R. M., and Shapiro, L. 1992. Computer and Robot Vision. New York, NY: Addison-Wesley.

48. Paragios, N., Rousson, M., and Ramesh, V. 2002. Matching distance functions: a shape-to-area variational approach for global-to-local registration. European Conference on Computer Vision, pp. 775–790.

49. Chan, T., and Zhu, W. 2005. Level set based shape prior segmentation. IEEE Conference on Computer Vision and Pattern Recognition.

50. Martinez, O., and Tsechpenakis, G. 2008. Integration of active learning in a collaborative CRF. IEEE Online Learning for Classification, Computer Vision and Pattern Recognition, Anchorage, AK, June.

51. Cristianini, N., and Shawe-Taylor, J. 2000. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge, UK: Cambridge University Press.

52. Sonka, M., Hlavac, V., and Boyle, R. 1999. Image Processing, Analysis and Machine Vision, 2nd edn. Pacific Grove, CA: PWS Publishing.

53. Berthod, M., Kato, Z., Yu, S., and Zerubia, J. 1996. Bayesian image classification using Markov random fields. Image and Vision Computing 14:285–295.