



SEMANTIC WEB FOR EFFECTIVE HEALTHCARE SYSTEMS

Edited by

VISHAL JAIN

JYOTIR MOY CHATTERJEE

ANKITA BANSAL

ABHA JAIN

Semantic Web for Effective Healthcare

Scrivener Publishing
100 Cummings Center, Suite 541J
Beverly MA, 01915-6106

Machine Learning in Biomedical Science and Healthcare Informatics

Series Editors: Vishal Jain and Jyotir Moy Chatterjee

In this series, the focus centers on the various applications of machine learning in the biomedical engineering and healthcare fields, with a special emphasis on the most representative learning techniques, namely deep learning-based approaches. Machine learning tasks typically classified into two broad categories depending on whether there is a learning "label" or "feedback" available to a learning system: supervised learning and unsupervised learning. This series also introduces various types of machine learning tasks in the biomedical engineering field from classification (supervised learning) to clustering (unsupervised learning). The objective of the series is to compile all aspects of biomedical science and healthcare informatics, from fundamental principles to current advanced concepts.

Submission to the series: Please send book proposals to drvishaljain83@gmail.com and/or jyotirchatterjee@gmail.com

Publishers at Scrivener
Martin Scrivener (martin@scrivenerpublishing.com)
Phillip Carmical (pcarmical@scrivenerpublishing.com)

Semantic Web for Effective Healthcare

Edited by

Vishal Jain

Sharda University, India

Jyotir Moy Chatterjee

Lord Buddha Education Foundation, Nepal

Ankita Bansal

Netaji Subhas University of Technology, India

and

Abha Jain

Shaheed Rajguru College of Applied Sciences for Women,

Delhi University, India



WILEY

This edition first published 2022 by John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA and Scrivener Publishing LLC, 100 Cummings Center, Suite 541J, Beverly, MA 01915, USA

© 2022 Scrivener Publishing LLC

For more information about Scrivener publications please visit www.scrivenerpublishing.com.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, except as permitted by law. Advice on how to obtain permission to reuse material from this title is available at <http://www.wiley.com/go/permissions>.

Wiley Global Headquarters

111 River Street, Hoboken, NJ 07030, USA

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Limit of Liability/Disclaimer of Warranty

While the publisher and authors have used their best efforts in preparing this work, they make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives, written sales materials, or promotional statements for this work. The fact that an organization, website, or product is referred to in this work as a citation and/or potential source of further information does not mean that the publisher and authors endorse the information or services the organization, website, or product may provide or recommendations it may make. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for your situation. You should consult with a specialist where appropriate. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read.

Library of Congress Cataloging-in-Publication Data

ISBN 978-1-119-76229-4

Cover image: Pixabay.Com

Cover design by Russell Richardson

Set in size of 11pt and Minion Pro by Manila Typesetting Company, Makati, Philippines

Printed in the USA

10 9 8 7 6 5 4 3 2 1

Contents

Preface	xv
Acknowledgment	xix
1 An Ontology-Based Contextual Data Modeling for Process Improvement in Healthcare	1
<i>A. M. Abirami and A. Askarunisa</i>	
1.1 Introduction	1
1.1.1 Ontology-Based Information Extraction	3
1.1.2 Ontology-Based Knowledge Representation	4
1.2 Related Work	5
1.3 Motivation	8
1.4 Feature Extraction	9
1.4.1 Vector Space Model	10
1.4.2 Latent Semantic Indexing (LSI)	11
1.4.3 Clustering Techniques	12
1.4.4 Topic Modeling	12
1.5 Ontology Development	17
1.5.1 Ontology-Based Semantic Indexing (OnSI) Model	17
1.5.2 Ontology Development	18
1.5.3 OnSI Model Evaluation	19
1.5.4 Metrics Analysis	23
1.6 Dataset Description	24
1.7 Results and Discussions	25
1.7.1 Discussion 1	29
1.7.2 Discussion 2	29
1.7.3 Discussion 3	30
1.8 Applications	31
1.9 Conclusion	32
1.10 Future Work	33
References	33

2	Semantic Web for Effective Healthcare Systems: Impact and Challenges	39
	<i>Hemendra Shankar Sharma and Ashish Sharma</i>	
2.1	Introduction	40
2.2	Overview of the Website in Healthcare	45
2.2.1	What Is Website?	45
2.2.2	Types of Website	45
2.2.2.1	Static Website	45
2.2.2.2	Dynamic Website	46
2.2.3	What Is Semantic Web?	46
2.2.4	Role of Semantic Web	47
2.2.4.1	Pros and Cons of Semantic Web	49
2.2.4.2	Impact on Patient	51
2.2.4.3	Impact on Practitioner	52
2.2.4.4	Impact on Researchers	52
2.3	Data and Database	53
2.3.1	What Is Data?	54
2.3.2	What Is Database?	54
2.3.3	Source of Data in the Healthcare System	54
2.3.3.1	Electronic Health Record (EHR)	55
2.3.3.2	Biomedical Image Analysis	56
2.3.3.3	Sensor Data Analysis	57
2.3.3.4	Genomic Data Analysis	57
2.3.3.5	Clinical Text Mining	58
2.3.3.6	Social Media	59
2.3.4	Why Are Databases Important?	60
2.3.5	Challenges With the Database in the Healthcare System	61
2.4	Big Data and Database Security and Protection	61
2.4.1	What Is Big Data	61
2.4.2	Five V's of Big Data	62
2.4.2.1	Volume	62
2.4.2.2	Variety	63
2.4.2.3	Velocity	63
2.4.2.4	Veracity	64
2.4.2.5	Value	65
2.4.3	Architectural Framework of Big Data	65
2.4.4	Data Protection Versus Data Security in Healthcare	67
2.4.4.1	Phishing Attacks	67
2.4.4.2	Malware and Ransomware	67

2.4.4.3	Cloud Threats	67
2.4.5	Technology in Use to Secure the Healthcare Data	68
2.4.5.1	Access Control Policy	69
2.4.6	Monitoring and Auditing	69
2.4.7	Standard for Data Protection	70
2.4.7.1	Healthcare Standard in India	70
2.4.7.2	Security Technical Standards	71
2.4.7.3	Administrative Safeguards Standards	71
2.4.7.4	Physical Safeguard Standards	71
	References	71
3	Ontology-Based System for Patient Monitoring	75
	<i>R. Mervin, Tintu Thomas and A. Jaya</i>	
3.1	Introduction	76
3.1.1	Basics of Ontology	77
3.1.2	Need of Ontology in Patient Monitoring	78
3.2	Literature Review	78
3.2.1	Uses of Ontology in Various Domains	78
3.2.2	Ontology in Patient Monitoring System	80
3.3	Architectural Design	80
3.3.1	Phases of Patient Monitoring System	82
3.3.2	Reasoner in Patient Monitoring	87
3.4	Experimental Results	88
3.4.1	SPARQL Results	89
3.4.2	Comparison Between Other Systems	89
3.5	Conclusion and Future Enhancements	90
	References	91
4	Semantic Web Solutions for Improvised Search in Healthcare Systems	95
	<i>Nidhi Malik, Aditi Sharan and Sadika Verma</i>	
4.1	Introduction	95
4.1.1	Key Benefits and Usage of Technology in Healthcare System	96
4.2	Background	97
4.2.1	Significance of Semantics in Healthcare Systems	97
4.2.2	Scope and Benefits of Semantics in Healthcare Systems	98
4.2.3	Issues in Incorporating Semantics	98
4.2.4	Existing Semantic Web Technologies	99
4.3	Searching Techniques in Healthcare Systems	100

4.3.1	Keyword-Based Search	100
4.3.2	Controlled Vocabularies Based Search	101
4.3.3	Improvising Searches With Semantic Web Solutions	101
4.3.4	Health Domain-Specific Resources for Semantic Search	102
4.3.4.1	Ontologies	103
4.3.4.2	Libraries	103
4.3.4.3	Search Engines	103
4.4	Emerging Technologies/Resources in Health Sector	108
4.4.1	Elasticsearch	109
4.4.2	BioBERT	109
4.4.3	Knowledge Graphs	110
4.5	Conclusion	110
	References	111
5	Actionable Content Discovery for Healthcare	115
	<i>Ujwala Bharambe and Anuradha Srinivasaraghavan</i>	
5.1	Introduction	116
5.2	Actionable Content	117
5.2.1	Actionable Content in Theory	117
5.2.2	Actionable Content in Practice	122
5.3	Health Analytics	124
5.3.1	Artificial Intelligence/Machine Learning-Based Predictive Analytics	125
5.3.2	Semantic Technology for Prescriptive Health Analytics	126
5.4	Ontologies and Actionable Content	127
5.4.1	Ontologies in Healthcare Domain	129
5.5	General Architecture for the Discovery of Actionable Content for Healthcare Domain	130
5.5.1	Ontology-Driven Actionable Content Discovery in Healthcare Domain	131
5.5.2	Case Study for Actionable Content Discovery in Cancer Domain	134
5.6	Conclusion	136
	References	136
6	Intelligent Agent System Using Medicine Ontology	139
	<i>Tintu Thomas and R. Mervin</i>	
6.1	Introduction to Semantic Search	140

6.1.1	What Is an Ontology in Terms of Medicine?	140
6.1.2	Needs and Benefits of Ontology in Medical Search	141
6.2	Sematic Search	142
6.2.1	How NLP Works in Sematic Search?	142
6.2.2	Part of Speech Tagging and Chunking	142
6.2.3	Sentence Parsing	143
6.2.4	Discussion About the Various Semantic Search in Medical Databases	144
6.2.5	Discussion About the Retrieval Tools Used in Sematic Search in Medline	145
6.3	Structural Pattern of Semantic Search	146
6.3.1	Architectural Diagram	147
6.3.2	Agent Ontology	148
6.3.3	Rule-Based Approach	149
6.3.4	Reasoners-Based Approach	151
6.4	Implementation of Reasoners	152
6.5	Implementation and Results	153
6.6	Conclusion and Future Prospective	153
	References	154
7	Ontology-Based System for Robotic Surgery—A Historical Analysis	159
	<i>Ajay Agarwal and Amit Kumar Mishra</i>	
7.1	Historical Discourse of Surgical Robots	160
7.2	The Necessity for Surgical Robots	162
7.3	Ontological Evolution of Robotic Surgical Procedures in Various Domains	163
7.4	Inferences Drawn From the Table	164
7.5	Transoral Robotic Surgery	166
7.6	Pancreatoduodenectomy	167
7.7	Robotic Mitral Valve Surgery	168
7.8	Rectal Tumor Surgery	170
7.9	Robotic Lung Cancer Surgery	170
7.10	Robotic Surgery in Gynecology	171
7.11	Robotic Radical Prostatectomy	171
7.12	Conclusion	172
7.13	Future Work	172
	References	172

8	IoT-Enabled Effective Healthcare Monitoring System Using Semantic Web	175
	<i>Sapna Juneja, Abhinav Juneja, Annu Dhankhar and Vishal Jain</i>	
8.1	Introduction	176
8.2	Literature Review	177
8.3	Phases of IoT-Based Healthcare	178
8.4	IoT-Based Healthcare Architecture	179
8.5	IoT-Based Sensors for Health Monitoring	180
8.6	IoT Applications in Healthcare	182
8.7	Semantic Web, Ontology, and Its Usage in Healthcare Sector	183
8.8	Semantic Web-Based IoT Healthcare	183
8.9	Challenges of IoT in Healthcare Industry	185
8.10	Conclusion	186
	References	186
9	Precision Medicine in the Context of Ontology	191
	<i>Rehab A. Rayan and Imran Zafar</i>	
9.1	Introduction	192
9.2	The Rationale Behind Data	195
9.3	Data Standards for Interoperability	197
9.4	The Evolution of Ontology	198
9.5	Ontologies and Classifying Disorders	199
9.6	Phenotypic Ontology of Humans in Rare Disorders	201
9.7	Annotations and Ontology Integration	202
9.8	Precision Annotation and Integration	203
9.9	Ontology in the Contexts of Gene Identification Research	204
9.10	Personalizing Care for Chronic Illness	207
9.11	Roadblocks Toward Precision Medicine	208
9.12	Future Perspectives	209
9.13	Conclusion	209
	References	210
10	A Knowledgebase Model Using RDF Knowledge Graph for Clinical Decision Support Systems	215
	<i>Ravi Lourdasamy and Xavierlal J. Mattam</i>	
10.1	Introduction	216
10.2	Relational Database to Graph Database	217
10.2.1	Relational Database for Knowledge Representation	218
10.2.2	NoSQL Databases	220
10.2.3	Graph Database	223

10.3	RDF	225
10.3.1	RDF Model and Technology	226
10.3.2	Metadata and URI	226
10.3.3	RDF Stores	228
10.4	Knowledgebase Systems and Knowledge Graphs	230
10.4.1	Knowledgebase Systems	230
10.4.2	Knowledge Graphs	232
10.4.3	RDF Knowledge Graphs	233
10.4.4	Information Retrieval Using SPARQL	234
10.5	Knowledge Base for CDSS	235
10.5.1	Curation of Knowledge Base for CDSS	236
10.5.2	Proposed Model for Curation	236
10.5.3	Evaluation Methodology	238
10.6	Discussion for Further Research and Development	239
10.7	Conclusion	239
	References	240
11	Medical Data Supervised Learning Ontologies for Accurate Data Analysis	249
	<i>B. Tarakeswara Rao, R. S. M. Lakshmi Patibandla, V. Lakshman Narayana and Arepalli Peda Gopi</i>	
11.1	Introduction	250
11.2	Ontology of Biomedicine	251
11.2.1	Ontology Resource Open Sharing	254
11.3	Supervised Learning	255
11.4	AQ21 Rule in Machine Learning	256
11.5	Unified Medical Systems	259
11.5.1	Note of Relevance to Bioinformatic Experts	259
11.5.2	Terminological Incorporation Principles	260
11.5.3	Cross-References External	261
11.5.4	UMLS Data Access	262
11.6	Performance Analysis	262
11.7	Conclusion	265
	References	265
12	Rare Disease Diagnosis as Information Retrieval Task	269
	<i>Jaya Lakkakula, Rutuja Phate, Alfiya Korbu and Sagar Barage</i>	
12.1	Introduction	270
12.2	Definition	271
12.3	Characteristics of Rare Diseases (RDs)	272

12.4	Types of Rare Diseases	273
12.4.1	Genetic Causes	274
12.4.2	Non-Genetic Causes	275
12.4.3	Pathogenic Causes (Infectious Agents)	275
12.4.4	Toxic Agents	275
12.4.5	Other Causes	276
12.5	A Brief Classification	276
12.6	Rare Disease Databases and Online Resources	277
12.6.1	European Reference Network: ERN	277
12.6.2	Genetic and Rare Diseases Information Center: GARD	278
12.6.3	International Classification of Diseases, 10th Revision: ICD-10	279
12.6.4	Orphanet-INSERM (Institut National de la Santé et de la Recherche Médicale)	280
12.6.5	Medical Dictionary for Regulatory Activities: MedDRA	280
12.6.6	Medical Subject Headings: MeSH	281
12.6.7	Online Mendelian Inheritance in Man: OMIM	282
12.6.8	Orphanet Rare Disease Ontology: ORDO	282
12.6.9	UMLS: Unified Medical Language System	282
12.6.10	SNOMED-CT: Systematized Nomenclature of Human and Veterinary Medicine—Clinical Terms	283
12.7	Information Retrieval of Rare Diseases Through a Web Search and Other Methods	284
12.7.1	What Is Information Retrieval (IR)?	284
12.7.2	Listed Below Are Some of the Methods for Information Retrieval	284
12.7.2.1	Web Search for a Diagnosis	284
12.7.2.2	Cause of Diagnostic Errors in Web-Based Tools	285
12.7.2.3	Nonprofessional Use of Web Tool for Diagnosis	285
12.7.2.4	Performance of Web Search Tools	285
12.7.2.5	Design of Watson	286
12.8	Tips and Tricks for Information Retrieval	287
12.9	Research on Rare Disease Throughout the World	288
12.10	Conclusion	290
	References	290

13 Atypical Point of View of Semantic Computing in Healthcare	293
<i>L. Mayuri and K. M. Mehata</i>	
13.1 Introduction	294
13.2 Mind the Language	295
13.2.1 Why Words Matter	296
13.2.2 What Words Matter	296
13.2.3 How Words Matter	297
13.3 Semantic Analytics and Cognitive Computing: Recent Trends	297
13.3.1 Semantic Data Analysis	298
13.3.2 Semantic Data Integration	299
13.3.3 Semantic Applications	300
13.4 Semantics-Powered Healthcare SOS Engineering	302
13.5 Conclusion	303
References	304
14 Using Artificial Intelligence to Help COVID-19 Patients	309
<i>Ayush Hans</i>	
14.1 Introduction	310
14.2 Method	313
14.3 Results	314
14.4 Discussion	315
14.4.1 What is the Use of AI in Healthcare?	315
14.4.2 How to Use AI for Critical Care Units	315
14.4.2.1 Input Stage	315
14.4.2.2 Process Stage	316
14.4.2.3 Output Stage	317
14.5 Conclusion	320
Acknowledgment	321
References	321
Index	325

Preface

The tremendous amount of data being generated on a daily basis in hospitals and other medical institutions needs to be properly harnessed and analyzed in order to gain useful insights. In other words, the health-related data needs to be explored in order to uncover valuable information that could lead to improved healthcare practices and the development of better biomedical products. However, there are many challenges, which need to be addressed before this goal is reached. One of the major challenges is interoperability of health and medical data. The data generated not only comes from different sources but also has inconsistencies in naming, structure, and format. An important requirement is to capture relevant data and also make it widely available for others to use. In addition to the data integration problem, user interaction with the data is another challenge. The difficulty lies in search handling, data navigation and data presentation. Finally, another challenge is how to use this huge amount of data to find valuable new patterns and transform such data into valuable knowledge, leading to potential improvement of resource utilization and patient health, and the development of biomedical products. There is a vast potential for data mining and data analytics tools in healthcare that could lead to useful information for decision making. In recent years, the Semantic Web has been gaining ground in addressing these challenges. The aim of this book is to analyze the current status on how the Semantic Web can be used to solve various challenges and enlighten readers with key advances in ontology-based information retrieval techniques in the healthcare domain. The following is a brief summary of the wide range of subjects covered throughout the book.

- Chapter 1 discusses various information extraction techniques used to model the documents product/service reviews. The advantages of using the Semantic Web to ease communication between businesses and improve processes are also discussed.

- Chapter 2 explores the impact of Semantic Web technologies and the challenges associated with their use in effective healthcare systems and also proposes solutions, which can be achieved with the present technology. In addition to this, some algorithms, frameworks, and real-time database systems realized with the help of artificial intelligence and web technology-based solutions are also discussed.
- Chapter 3 focuses on the importance of an ontology-based system for a patient monitoring system. A domain ontology has been constructed to preserve the details of patient health issues. With the support of ontology, a patient monitoring system is constructed wherein data concerning every detail about the patient and their health is stored.
- Chapter 4 highlights the role of Semantic Web technologies in improving services provided by healthcare systems. It elaborates on the search techniques used by researchers in the field to find the desired information. The role of semantics, how they are beneficial in the search process, and domain-specific resources are presented in detail. The latest technological advancements and resources from the biomedical field are also discussed.
- Chapter 5 discusses what actionable content should look like in practice and how it can become more efficient by aiding in clinical decision-making and administrative capability. The chapter renders various definitions of actionable content, and also focuses on the stages of health analytics and how ontology can be used for prescriptive health analytics.
- Chapter 6 depicts the retrieval of ontology-based information from the medical literature database MEDLINE. The main focus of the chapter is to enhance the retrieval of information from the medical literature database and conduct the search with more clarity. The approach discussed to achieve this is the preliminary design and execution of an ontology-based intelligent agent system that applies Semantic Web language, which benefits efficient systematic retrieval of medical information.
- Chapter 7 presents a historical analysis of an ontology-based system for robotic surgery and documents the most significant interventions of robots in medical surgery. The chapter discusses how the academic field has embraced this new discipline and how inclusive research on a worldwide scale

has honed the design and method of robotic procedures, all while maintaining an impeccable metric.

- Chapter 8 presents the applications of IoT in healthcare and how these applications can be used with the help of various sensors. It discusses the established strategies used by IoT-based devices to deal with patients, doctors, and hospitals in order to provide smarter and faster services. The authors propose an IoT-based architecture for monitoring the health of patients remotely.
- Chapter 9 discusses the use of precision medicine in the context of ontology. It explains ontologies and their application in computational reasoning to promote an accurate classification of patients' diagnoses and managing care, and for translational research.
- Chapter 10 discusses the use of knowledge graphs for knowledge representation. A model for such a knowledgebase is proposed that makes use of open information extraction systems to capture relevant knowledge from medical literature and curate it in the knowledgebase of the clinical decision support system.
- Chapter 11 covers all aspects related to the successful customization of data semantics, ontologies, clinical jobs, and free learning, and depicts the Unified Medical Language System (UMLS) framework used inside AQ21 rule learning programming. Ontologies are the quality systems for expressive genuine variables in clinical and flourishing fields.
- Chapter 12 provides information on rare diseases and explores the relationship between rare diseases, diagnoses, and information retrieval. In particular, it illustrates the history, characteristics, types, and classification along with databases of rare disease information. It also explores the challenges faced by researchers in rare disease information retrieval and how they can be resolved by search query optimization.
- Chapter 13 reviews the recent advances in medical terminology tools and application strategies currently in use for semantic reasoning and interoperability in healthcare. Common terminology standards used in health information and technology, such as SNOMED CT, RxNorm, LOINC, ICD-x-CM, and UCUM, are discussed. Also discussed are the current reference terminology mapping solutions that

enable semantic interoperability of data between health systems.

- Chapter 14 builds upon the existing AI-based model in order to discover a new model to improve healthcare facilities for the faster recovery of COVID-19 patients. The chapter discusses different AI-related solutions for the healthcare industry.

In conclusion, we are grateful to all those who directly and indirectly contributed to this book. We are also grateful to the publisher for giving us the opportunity to publish it.

Vishal Jain
Jyotir Moy Chatterjee
Ankita Bansal
Abha Jain
September 2021

Acknowledgment

I would like to acknowledge the most important people in my life—my late grandfather Shri Gopal Chatterjee, my late grandmother Smt. Subhankori Chatterjee, my late mother Nomita Chatterjee, my uncle Shri. Moni Moy Chatterjee, and my father Shri. Alope Moy Chatterjee. The book has been my long-cherished dream, which would not have become a reality without the support and love of these amazing people. They continued to encourage me despite my failing to give them the proper time and attention. I am also grateful to my friends, who have encouraged and blessed this work with their unconditional love and patience.

Jyotir Moy Chatterjee

Department of IT

Lord Buddha Education Foundation

(Asia Pacific University of Technology & Innovation)

Kathmandu, Nepal

An Ontology-Based Contextual Data Modeling for Process Improvement in Healthcare

A. M. Abirami^{1*} and A. Askarunisa²

¹*Department of Information Technology, Thiagarajar College of Engineering, Madurai, Tamil Nadu, India*

²*Department of Computer Science and Engineering, KLN College of Information Technology, Madurai, Tamil Nadu, India*

Abstract

The internet world contains large volume of text data. The integration of web sources is required to derive needed information. Human annotation is much difficult and tedious. Automated processing is necessary to make these data readable by machines. But mostly they are available in unstructured format, and they need to be formatted into structured form. Structured information is retrieved from unstructured or semi-structured text which is defined as text analytics. There are many Information Extraction (IE) techniques available to model the documents (product/service reviews). Vector space model uses only the content but not the contextual representation. This complexity is resolved by Semantic web, the initiative of WWW Consortium. The advantage of the use of Semantic web enables the ease of communication between Businesses and in process improvement.

Keywords: Ontology, semantic-web, decision making, healthcare, service, reviews

1.1 Introduction

Text analysis is defined as deriving structured data from unstructured text. Additional information like customer insight about the product or service can be retrieved from the unstructured data sources using text

*Corresponding author: abiramiam@tce.edu

analytics techniques. Its techniques have different applications such as insurance claims assessment, competitor analysis, sentiment analysis and the like. Many industries use text analytics for their business improvement. Social media impacts different industries like product business [1, 2], tourism [3, 4], and healthcare service [5] with the tremendous changes in the recent past years.

Retrieving and summarizing web data, which are dispersed in different web pages, are difficult and complex processes; also, they consume most of the manual effort and time. No standard data model exists for web documents. This increases the necessity of annotating the huge number of text documents that exist in the World Wide Web (WWW). Extracting and collating the information from these text is a complex task. Unlike numerical dataset, text documents contain more number of features. The amount of resources required to represent big dataset may be improved by representing the text documents with most needed and non-redundant features. Classification or clustering algorithms may be used for identifying the features from the text documents. The documents are analyzed, modeled and then used in the process of business improvement or for personal interest. Thus, the annotated text improves automated decision-making process, which in turn reduces the manual effort and time required for text analysis.

The report from British Columbia Safety and Quality Council says when patients and healthcare service entities are engaged in online platform, then there would be greater improvement in offering healthcare services. Improvement in healthcare services is visible when insights from the experience of patients are analyzed [5]. Hence, it becomes necessary to consolidate the opinions from the customers or clients so as to improve business, decision-making and increase revenue. Figure 1.1 gives the overview of decision-making process from the online product/service reviews, using different information extraction and text analysis techniques.

There exist many challenges while analyzing the social media text or user-generated content. In languages like English, the same word has multiple meaning (polysemy), and different words have same meaning (synonymy). People show “variety” and use heterogeneous words while expressing their views. It often leads to complication in processing the textual data. Most of the feature extraction techniques do not consider the semantic relationships between the terms. Subjectivity that exists in text processing techniques adds complexity to the process, which in turn impacts the evaluation of results. Also, the rare availability of gold-standard or annotated text data for different domains add more challenges to text analysis [6]. Hence, the identification and application of suitable Natural Language

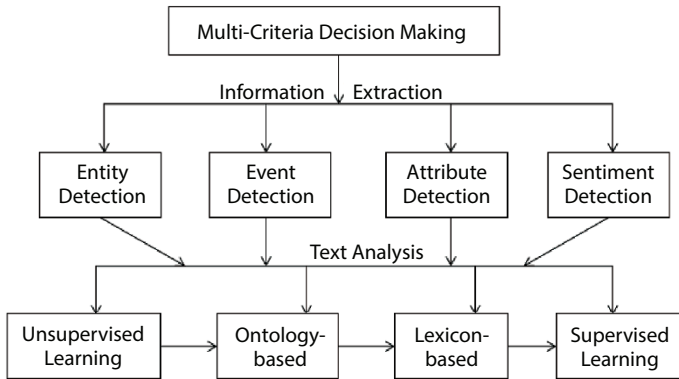


Figure 1.1 Decision-making process from social media reviews.

Processing (NLP) techniques are the main research focus in text data analysis.

Text analytics supports the context matching between the reader and the writer. This challenge can be managed if different vocabularies of features and their relationship are well represented in the data model. For example, content based contextual user feedback analysis enables the users to buy new products or avail any service by highlighting the best features of products or services. Challenges and issues in information retrieval problems are overcome if Ontology representation and topic modelling techniques are used for modeling the text documents. The chapter focuses on extracting relevant features from the set of documents and building domain ontology for them. The Ontology helps in building the predictive or sentiment analysis model by using suitable information retrieval (IR) techniques and contextual representation of data, so as to enable automated decision-making process, before buying a new product or availing a new service, as shown in Figure 1.2.

1.1.1 Ontology-Based Information Extraction

Ontology describes a domain of classes. It is defined as a conceptual model of knowledge representation. The concepts of the domain (classes), their attributes, their properties and their relationships are well described by the Ontology model. It also explains the meanings of the terms applicable to the domain. Ontology is one of the key components of semantic web technology. The semantic web technologies like Ontology, RDF and Sparql are

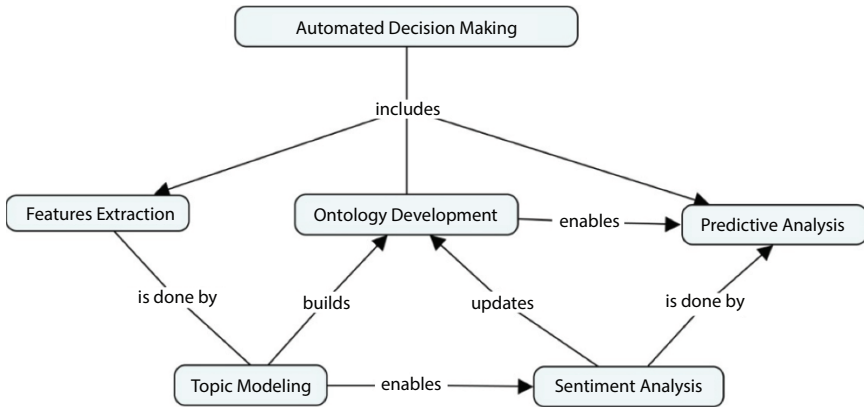


Figure 1.2 User-generated content analysis (UCA) model.

used in describing different words and their dependencies by modeling the textual data. Components of Ontology include:

- Concepts are also known as Classes. It is a unit of knowledge, shared among identified group of persons for the concept's domain. There exists relationship among concepts.
- Instances are individuals of concepts. They represent specific elements attached to the domain ontology. Instances are the “thing” represented by a concept.

Information Extraction (IE) and Ontology are related with one another like: Ontology is used in information extraction as part of understanding process of the domain; on the other hand, IE is used to design and enrich Ontology [7]. Common vocabulary and shared understanding among different people are enabled by Ontology. The contextual representation of data semantics is well described by the Ontology [8]. The UML diagrams along with Ontology support the biologists by classifying the entities and interactions between proteins and genes [9]. The terms (vocabularies) and the concepts (classes) in the source Ontology are used in term matching, thereby used in tagging the text documents. Thus the Ontology and their specifications are used in the information extraction process.

1.1.2 Ontology-Based Knowledge Representation

Knowledge is data that represents the outcome of computer-based cognitive processes such as perception, learning, association, and reasoning,

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley's ebook EULA.

