

Machine Learning in Biomedical Science and Healthcare Informatics

# SEMANTIC WEB FOR EFFECTIVE HEALTHCARE SYSTEMS

**Edited by**  
VISHAL JAIN  
JYOTIR MOY CHATTERJEE  
ANKITA BANSAL  
ABHA JAIN

 Scrivener  
Publishing

WILEY

# Table of Contents

[Cover](#)

[Title page](#)

[Copyright](#)

[Preface](#)

[Acknowledgment](#)

[1 An Ontology-Based Contextual Data Modeling for Process Improvement in Healthcare](#)

[1.1 Introduction](#)

[1.2 Related Work](#)

[1.3 Motivation](#)

[1.4 Feature Extraction](#)

[1.5 Ontology Development](#)

[1.6 Dataset Description](#)

[1.7 Results and Discussions](#)

[1.8 Applications](#)

[1.9 Conclusion](#)

[1.10 Future Work](#)

[References](#)

[2 Semantic Web for Effective Healthcare Systems: Impact and Challenges](#)

[2.1 Introduction](#)

[2.2 Overview of the Website in Healthcare](#)

[2.3 Data and Database](#)

[2.4 Big Data and Database Security and Protection](#)

[References](#)

[3 Ontology-Based System for Patient Monitoring](#)

[3.1 Introduction](#)

[3.2 Literature Review](#)

[3.3 Architectural Design](#)

[3.4 Experimental Results](#)

[3.5 Conclusion and Future Enhancements](#)

[References](#)

[4 Semantic Web Solutions for Improvised Search in Healthcare Systems](#)

[4.1 Introduction](#)

[4.2 Background](#)

[4.3 Searching Techniques in Healthcare Systems](#)

[4.4 Emerging Technologies/Resources in Health Sector](#)

[4.5 Conclusion](#)

[References](#)

[5 Actionable Content Discovery for Healthcare](#)

[5.1 Introduction](#)

[5.2 Actionable Content](#)

[5.3 Health Analytics](#)

[5.4 Ontologies and Actionable Content](#)

[5.5 General Architecture for the Discovery of Actionable Content for Healthcare Domain](#)

[5.6 Conclusion](#)

[References](#)

[6 Intelligent Agent System Using Medicine Ontology](#)

[6.1 Introduction to Semantic Search](#)

[6.2 Semantic Search](#)

[6.3 Structural Pattern of Semantic Search](#)

[6.4 Implementation of Reasoners](#)

[6.5 Implementation and Results](#)

[6.6 Conclusion and Future Prospective](#)

[References](#)

## [7 Ontology-Based System for Robotic Surgery—A Historical Analysis](#)

[7.1 Historical Discourse of Surgical Robots](#)

[7.2 The Necessity for Surgical Robots](#)

[7.3 Ontological Evolution of Robotic Surgical Procedures in Various Domains](#)

[7.4 Inferences Drawn From the Table](#)

[7.5 Transoral Robotic Surgery](#)

[7.6 Pancreatoduodenectomy](#)

[7.7 Robotic Mitral Valve Surgery](#)

[7.8 Rectal Tumor Surgery](#)

[7.9 Robotic Lung Cancer Surgery](#)

[7.10 Robotic Surgery in Gynecology](#)

[7.11 Robotic Radical Prostatectomy](#)

[7.12 Conclusion](#)

[7.13 Future Work](#)

[References](#)

## [8 IoT-Enabled Effective Healthcare Monitoring System Using Semantic Web](#)

[8.1 Introduction](#)

[8.2 Literature Review](#)

[8.3 Phases of IoT-Based Healthcare](#)

[8.4 IoT-Based Healthcare Architecture](#)

[8.5 IoT-Based Sensors for Health Monitoring](#)

[8.6 IoT Applications in Healthcare](#)

[8.7 Semantic Web, Ontology, and Its Usage in Healthcare Sector](#)

[8.8 Semantic Web-Based IoT Healthcare](#)

[8.9 Challenges of IoT in Healthcare Industry](#)

[8.10 Conclusion](#)

[References](#)

[9 Precision Medicine in the Context of Ontology](#)

[9.1 Introduction](#)

[9.2 The Rationale Behind Data](#)

[9.3 Data Standards for Interoperability](#)

[9.4 The Evolution of Ontology](#)

[9.5 Ontologies and Classifying Disorders](#)

[9.6 Phenotypic Ontology of Humans in Rare Disorders](#)

[9.7 Annotations and Ontology Integration](#)

[9.8 Precision Annotation and Integration](#)

[9.9 Ontology in the Contexts of Gene Identification Research](#)

[9.10 Personalizing Care for Chronic Illness](#)

[9.11 Roadblocks Toward Precision Medicine](#)

[9.12 Future Perspectives](#)

[9.13 Conclusion](#)

[References](#)

[10 A Knowledgebase Model Using RDF Knowledge Graph for Clinical Decision Support Systems](#)

[10.1 Introduction](#)

[10.2 Relational Database to Graph Database](#)

[10.3 RDF](#)

[10.4 Knowledgebase Systems and Knowledge Graphs](#)

[10.5 Knowledge Base for CDSS](#)

[10.6 Discussion for Further Research and Development](#)

[10.7 Conclusion](#)

[References](#)

[11 Medical Data Supervised Learning Ontologies for Accurate Data Analysis](#)

[11.1 Introduction](#)

[11.2 Ontology of Biomedicine](#)

[11.3 Supervised Learning](#)

[11.4 AQ21 Rule in Machine Learning](#)

[11.5 Unified Medical Systems](#)

[11.6 Performance Analysis](#)

[11.7 Conclusion](#)

[References](#)

[12 Rare Disease Diagnosis as Information Retrieval Task](#)

[12.1 Introduction](#)

[12.2 Definition](#)

[12.3 Characteristics of Rare Diseases \(RDs\)](#)

[12.4 Types of Rare Diseases](#)

[12.5 A Brief Classification](#)

[12.6 Rare Disease Databases and Online Resources](#)

[12.7 Information Retrieval of Rare Diseases Through a Web Search and Other Methods](#)

[12.8 Tips and Tricks for Information Retrieval](#)

[12.9 Research on Rare Disease Throughout the World](#)

[12.10 Conclusion](#)

[References](#)

[13 Atypical Point of View on Semantic Computing in Healthcare](#)

[13.1 Introduction](#)

[13.2 Mind the Language](#)

[13.3 Semantic Analytics and Cognitive Computing: Recent Trends](#)

[13.4 Semantics-Powered Healthcare SOS Engineering](#)

[13.5 Conclusion](#)

[References](#)

[14 Using Artificial Intelligence to Help COVID-19 Patients](#)

[14.1 Introduction](#)

[14.2 Method](#)

[14.3 Results](#)

[14.4 Discussion](#)

[14.5 Conclusion](#)

[Acknowledgment](#)

[References](#)

[Index](#)

[End User License Agreement](#)

## **List of Illustrations**

Chapter 1

[Figure 1.1 Decision-making process from social media reviews.](#)

[Figure 1.2 User-generated content analysis \(UCA\) model.](#)

[Figure 1.3 Term weighing schemes for feature extraction.](#)

[Figure 1.4 Synonymy and polysemy issues in English.](#)

[Figure 1.5 Approximated TD matrix by SVD.](#)

[Figure 1.6 LDA framework.](#)

[Figure 1.7 Plate notation of  \$CFS\_{LDA}\$  model.](#)

[Figure 1.8 Ontology-based semantic indexing \(OnSI\) model.](#)

[Figure 1.9 Domain ontology modeling for features.](#)

[Figure 1.10 Ontology mapping using OnSI model.](#)

[Figure 1.11 Spring view of domain ontology \(DS1\).](#)

[Figure 1.12 Precision vs recall curve for the Dataset DS1.](#)

[Figure 1.13 Hierarchy of MCDM problem.](#)

[Figure 1.14 Ranking of features by VIKOR method.](#)

## Chapter 2

[Figure 2.1 Sources of data in healthcare. Wikipedia contributors. \(2021, May 11\)...](#)

[Figure 2.2 Conceptual architecture of big data.](#)

## Chapter 3

[Figure 3.1 Conceptual diagram for patient's monitoring system.](#)

[Figure 3.2 Ontology for patient's monitoring system.](#)

[Figure 3.3 Parse tree for patient's monitoring system.](#)

[Figure 3.4 SWRL for patient's monitoring system.](#)

[Figure 3.5 SPQRAL query for patient's monitoring system.](#)

[Figure 3.6 Comparison between SQL and SAPRQL.](#)

## Chapter 5

[Figure 5.1 \(a\) Sequential composition model; \(b\) Parallel composition model.](#)

[Figure 5.2 Evolution of actionable content.](#)

[Figure 5.3 Data to actionable content triangle.](#)

[Figure 5.4 Spectrum of healthcare analytics.](#)

[Figure 5.5 Model of ontology.](#)

[Figure 5.6 Different types of ontologies.](#)

[Figure 5.7 Block diagram for actionable content discovery.](#)

[Figure 5.8 Overall architecture of ontology-driven actionable content discovery ...](#)

[Figure 5.9 Inner mechanism of proposed architecture.](#)

[Figure 5.10 Contextual dimension modeling diagram with data coming from various ...](#)

## Chapter 6

[Figure 6.1 Architecture diagram.](#)

[Figure 6.2 Parse tree of medicine.](#)

[Figure 6.3 Medicine ontology.](#)

[Figure 6.4 Rules of construction.](#)

[Figure 6.5 SPARQL query output.](#)

## Chapter 7

[Figure 7.1 The above figure represents a graph-based representation of meta-anal...](#)

[Figure 7.2 The above figure represents a graph-based representation of meta-anal...](#)

[Figure 7.3 The intraoperative results of the robotic-assisted mitral valve surge...](#)

[Figure 7.4 The encounters of complications faced during the surgery by both grou...](#)

[Figure 7.5 The postoperative results depict that the patients in Group II requir...](#)

## Chapter 8

[Figure 8.1 Phases of IoT process.](#)

[Figure 8.2 IoT-based healthcare architecture.](#)

[Figure 8.3 Semantic web-based IoT healthcare.](#)

## Chapter 9

[Figure 9.1 Examples of some ontologies.](#)

## Chapter 10

[Figure 10.1 CAP. Consistency and tolerance to network partition or availability ...](#)

[Figure 10.2 Stages in the proposed model.](#)

## Chapter 11

[Figure 11.1 Disorders of tyrosinemia type II.](#)

[Figure 11.2 UMLS metathesaurus relations.](#)

## Chapter 14

[Graph 14.1 Increase in life expectancy with the advancement in healthcare sector...](#)

[Figure 14.1 Distribution of COVID cases across the world.](#)

[Figure 14.2 Projections of the healthcare facilities present VS that we require.](#)

[Figure 14.3 Three-stage model for critical care patients.](#)

## List of Tables

### Chapter 1

[Table 1.1 Confusion matrix.](#)

[Table 1.2 Social media reviews for healthcare service \(DS1\).](#)

[Table 1.3 Number of reviews of features and hospitals \(DS1\).](#)

[Table 1.4 List of top five terms by LDA model.](#)

[Table 1.5 Sample correlated terms selected by  \$CFS\_{LDA}\$ .](#)

[Table 1.6 List of correlated feature terms selected by  \$CFS\_{LDA}\$  model.](#)

[Table 1.7 Performance evaluation.](#)

### Chapter 2

[Table 2.1 Growth of data \[15\].](#)

### Chapter 4

[Table 4.1 List of popular biomedical ontologies.](#)

[Table 4.2 List of popular biomedical libraries.](#)

## Chapter 7

[Table 7.1 The table shows counts of complications that were encountered intra- a...](#)

## Chapter 9

[Table 9.1 Different terms used in computational examination \[13\].](#)

[Table 9.2 Examples for interoperability standards and ontologies commonly used i...](#)

[Table 9.3 Resources and applications in human phenotype ontology contexts.](#)

## Chapter 10

[Table 10.1 Contingency matrix of positives and negatives.](#)

## Chapter 11

[Table 11.1 Patients characteristics in the final dataset \(n = 829\).](#)

[Table 11.2 ADL disabilities patients before and after cancer diagnosis.](#)

[Table 11.3  \$Q\(R, w^\*\)\$ , rule for ADL.](#)

## Chapter 12

[Table 12.1 List of various diagnostic tools with their webpages \[6\].](#)

## Chapter 14

[Table 14.1 Three-stage model of application of AI for patients in critical care.](#)

## **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](mailto:drvishaljain83@gmail.com) and/or [jyotirchatterjee@gmail.com](mailto:jyotirchatterjee@gmail.com)

*Publishers at Scrivener*

Martin Scrivener ([martin@scrivenerpublishing.com](mailto:martin@scrivenerpublishing.com))

Phillip Carmical ([pcarmical@scrivenerpublishing.com](mailto: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](http://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](http://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](http://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

# 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 bio-medical 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

# 1

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

A. M. Abirami<sup>1\*</sup> and A. Askarunisa<sup>2</sup>

<sup>1</sup>*Department of Information Technology, Thiagarajar College of Engineering, Madurai, Tamil Nadu, India*

<sup>2</sup>*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 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

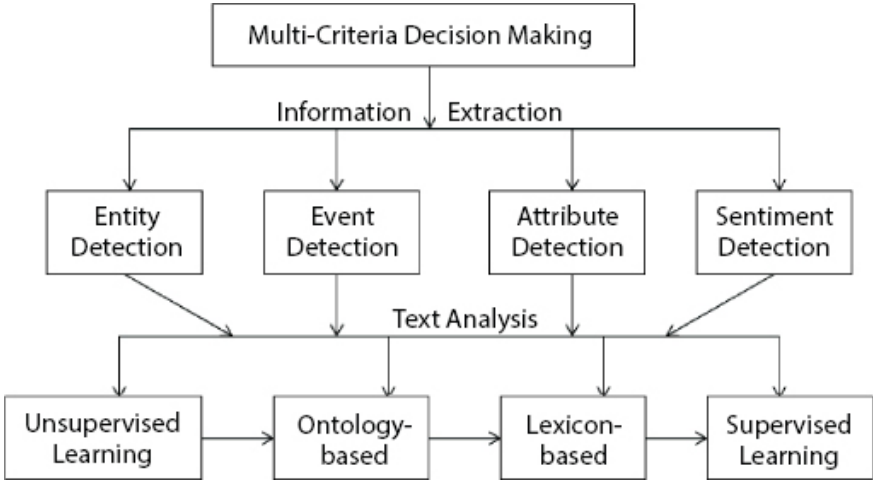
product business [1, 2], tourism [3, 4], and healthcare service [5] with the tremendous changes in the recent past years.

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.

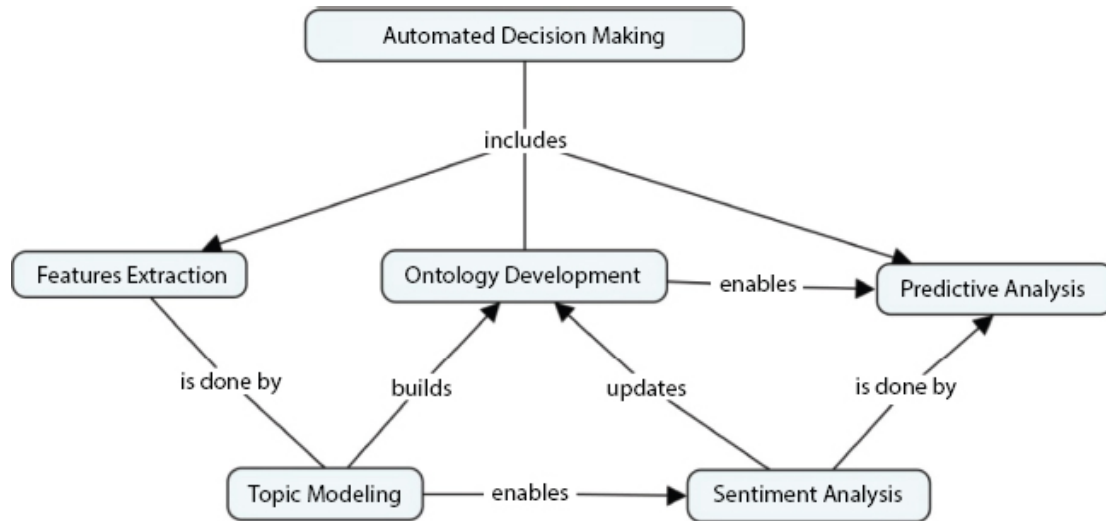
There exist many challenges while analyzing the social media text or user-generated content. In languages like English, the same word has multiple meaning (polysemy), and different words have same meaning (synonymy). People show “variety” and use heterogeneous words while expressing their views. It often 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 Processing (NLP) techniques are the main research focus in text data analysis.



**Figure 1.1** Decision-making process from social media reviews.

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](#).



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

### 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 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, or the translation of knowledge acquired by human [10]. It is the language by which human express their understanding about the concept. The concepts and the instances of a particular domain are expressed in the knowledge base also referred as the semantic knowledge dictionary. It is one of the most important techniques to represent the knowledge for a domain. Domain Ontology is developed to formally define the concepts, relationships, and rules so as to include the semantic content of the domain. The semantic approach uses the concepts in the documents to establish the contextual relationship rather than the terms. Issues like synonymy and polysemy may not be resolved if terms are used as indices while modelling the text documents. Various semantic-based information extraction approaches like Latent Semantic Indexing [11] and Latent Dirichlet Allocation [12, 13] techniques are used for building the relationship among the indexed terms, so as to represent the contexts between the concepts. This chapter focuses on developing domain Ontology to represent the features and their related terms mentioned in the product/service reviews generated in social media web sites.

## **1.2 Related Work**

Ontology facilitates the shared understanding among the people by formalizing the conceptualization of a specific domain. The contextual representation of data semantics is well described by the Ontology [8]. Ontology defines concepts (domain) by using the common vocabulary and describes attributes, behavior, relationships and constraints. The UML diagrams along with The interactions between proteins and genes are well explained by Ontology representation which would support the biologists for classification [9]. Reviews on hotels and movies are classified using the rule-based systems and Ontology [14-16]. Document annotation and rules were used to create knowledge base of web documents from the extraction of semantic data like named entities [14, 17, 18]. Ontology learning and RDF

Social media, [2](#)  
SPARQL, [77](#), [81](#), [89](#), [216](#), [217](#), [234](#), [235](#), [237](#), [245](#)  
SPARQL inquiry, [254](#)  
Spring view, [28](#)  
SQL, [219](#)-221, [225](#), [235](#), [240](#), [241](#)  
Standardized resource identifier (URI), [254](#)  
Standards, [197](#)  
Starting index (si), [143](#)  
Statistical classification, [196](#)  
Stop word data, [146](#)  
Stores, [218](#), [222](#), [229](#), [230](#), [241](#)  
Structure, [198](#)  
Supervised learning method, [147](#)

Supervised learning ontologies, for accurate data analysis, [249](#)-265

AQ21 rule in machine learning, [256](#)-259

community-based development, [255](#)

learning approaches, [255](#)

ontology,

of biomedicine, [251](#)-255

resource open sharing, [254](#)-255

overview, [249](#)-251

performance analysis, [262](#)-265

search database solution, [251](#)

UMLS, [259](#)-262

AQ21 rule, [250](#), [256](#), [257](#), [258](#), [259](#)

bioinformatic experts, [259](#)-260

cross-references external, [261](#)

data access, [262](#)

objective of, [252](#)

terminological incorporation principles, [260](#)-261

Support vector machines (SVMs)

techniques, [145](#), [147](#), [151](#)-152

Surgical robot,

Puma 560, [160](#)-161, [163](#)

Robodoc, [160](#)-162

Vinci, [162](#), [164](#), [167](#)

Zeus, [162](#), [164](#), [170](#)

SWRL, [81](#), [86](#)

Syntax, [193](#)

Systematized nomenclature in medicine (SNOMED), [251](#)-[252](#), [260](#)

Systems-of-systems, [302](#)

Task ontology, [77](#)

T-box, [88](#)

Telehealth, [320](#)

Term document matrix, [9](#)

Term weighing, [10](#)

    inverse document frequency, [10](#)

    term frequency, [10](#)

Terminological incorporation principles, UMLS, [260](#)-261

Terminology, [192](#)

TF-IDF, [79](#), [81](#), [86](#)

Thesaurus, [196](#)

Tips and tricks for information retrieval, [287](#)-288

Topic modeling, [12](#)-14

    Gibb's parameters, [15](#)

    latent dirichlet allocation, [13](#)

    plate notation, [15](#)

Tracheostomy, [317](#)

Transoral robotic surgery, [166](#)

Triple, [226](#), [229](#), [237](#), [238](#)

Types of rare diseases, [273](#)-276

UMLS (unified medical language system), [146](#), [199](#)

UMLS. *see* [Unified medical framework \(UMLS\)](#).

Unified medical framework (UMLS), [259](#)-262

AQ21 rule, [250](#), [256](#), [257](#), [258](#), [259](#)

bioinformatic experts, [259](#)-260

cross-references external, [261](#)

data access, [262](#)

objective of, [252](#)

terminological incorporation principles, [260](#)-261

URI, [254](#)

VATS, [170](#), [172](#)

Vector space model, [10](#)

Verb, [85](#)

Virtuoso triple store, [254](#)

Visualizations, [207](#), [234](#), [245](#)

Vocabulary, [196](#)

Web, [194](#), [220](#), [221](#), [225](#), [227](#)-229, [239](#), [241](#)

Web ontology language (OWL), [149](#)-150, [250](#), [254](#)

Whale optimization support vector machine (OWSVM) method, [145](#)

World wide web consortium (W3c), [149](#), [250](#)