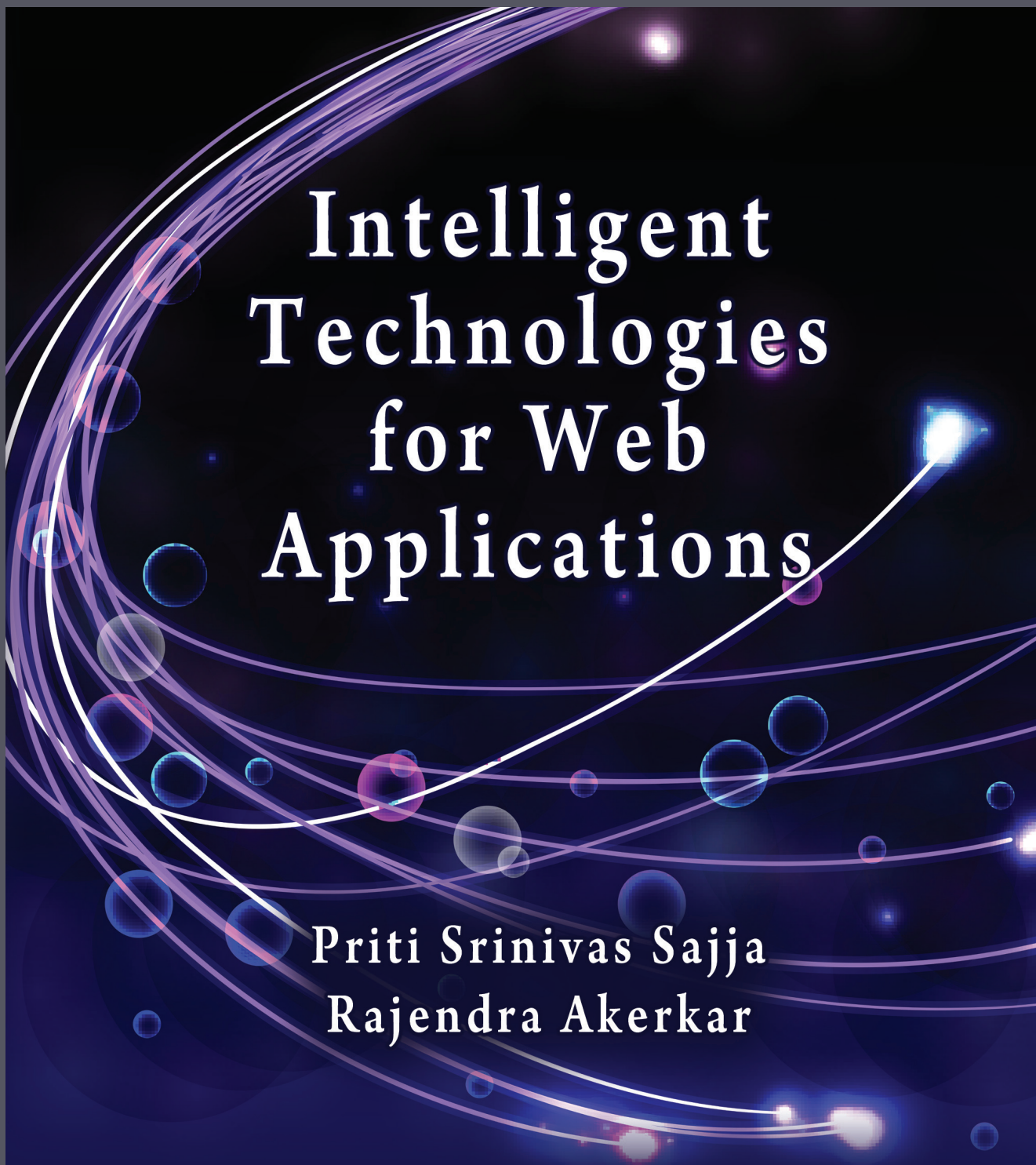
Chapman & Hall/CRC
Data Mining and Knowledge Discovery Series



Intelligent Technologies for Web Applications

Priti Srinivas Sajja
Rajendra Akerkar

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Intelligent Technologies for Web Applications

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

TEMPORAL DATA MINING
Theophano Mitsa

RELATIONAL DATA CLUSTERING: MODELS, ALGORITHMS, AND APPLICATIONS
Bo Long, Zhongfei Zhang, and Philip S. Yu

KNOWLEDGE DISCOVERY FROM DATA STREAMS
João Gama

STATISTICAL DATA MINING USING SAS APPLICATIONS, SECOND EDITION
George Fernandez

INTRODUCTION TO PRIVACY-PRESERVING DATA PUBLISHING:
CONCEPTS AND TECHNIQUES
Benjamin C. M. Fung, Ke Wang, Ada Wai-Chee Fu, and Philip S. Yu

HANDBOOK OF EDUCATIONAL DATA MINING
Cristóbal Romero, Sebastian Ventura, Mykola Pechenizkiy, and Ryan S.J.d. Baker

DATA MINING WITH R: LEARNING WITH CASE STUDIES
Luís Torgo

MINING SOFTWARE SPECIFICATIONS: METHODOLOGIES AND APPLICATIONS
David Lo, Siau-Cheng Khoo, Jiawei Han, and Chao Liu

DATA CLUSTERING IN C++: AN OBJECT-ORIENTED APPROACH
Guojun Gan

MUSIC DATA MINING
Tao Li, Mitsunori Ogihara, and George Tzanetakis

MACHINE LEARNING AND KNOWLEDGE DISCOVERY FOR
ENGINEERING SYSTEMS HEALTH MANAGEMENT
Ashok N. Srivastava and Jiawei Han

SPECTRAL FEATURE SELECTION FOR DATA MINING
Zheng Alan Zhao and Huan Liu

ADVANCES IN MACHINE LEARNING AND DATA MINING FOR ASTRONOMY
Michael J. Way, Jeffrey D. Scargle, Kamal M. Ali, and Ashok N. Srivastava

FOUNDATIONS OF PREDICTIVE ANALYTICS
James Wu and Stephen Coggeshall

INTELLIGENT TECHNOLOGIES FOR WEB APPLICATIONS
Priti Srinivas Sajja and Rajendra Akerkar

Intelligent Technologies for Web Applications

Priti Srinivas Sajja

Rajendra Akerkar

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

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

Version Date: 20120424

International Standard Book Number-13: 978-1-4398-7164-5 (eBook - PDF)

Dedication

To Srinivas and Abhignya

Priti Srinivas Sajja

To my parents, Ashalata and Arvind Akerkar

Rajendra Akerkar

Contents

Part III: Semantic web and web knowledge management

Preface

The Web is becoming the largest data repository in the world and presents a key driving force for a large spectrum of information technology (IT). To develop effective and intelligent web applications and services, it is critical to discover useful knowledge through analyzing large amounts of content, hidden content structures, or usage patterns of web data resources. To achieve such a goal, a variety of techniques in diverse research areas need to be integrated, including natural language processing, information extraction, information retrieval, information filtering, knowledge representation, knowledge management, machine learning, databases, data mining, web mining, text mining, agent, human–computer interaction, and the semantic web. These integrated techniques should address the key challanges from the heterogeneous and dynamic nature of web contents and usage patterns.

Within the past ten years, the Web research community has brought to maturity a comprehensive set of foundational technology components, both at the conceptual level and in the form of prototypes and software.

Intended readers

This book describes the basics as well as the latest trends in the area of an integrated approach instead of an edited volume of papers/chapters. The book provides a detailed review of issues for web researchers. With extensive use of examples and more than 100 illustrations, as well as bibliographical notes, end-of-chapter exercises, and glossaries, to clarify complex material and demonstrate practical applications, this book can serve as a senior undergraduate-level book. It can also serve as a good reference for researchers and practitioners who deal with the various problems involving semantics, intelligent techniques for web ontologies, and the semantic web.

Understanding web-related concepts, studying the underlying standards and technical components, and putting all of this together into concrete terms require a substantial amount of effort. This book provides comprehensive and easy-to-follow coverage on both the "what-is" and "how-to" aspects of web-related technologies.

In particular, this book is written keeping the following readers in mind:

- Software engineers and developers who are interested in learning the intelligent and semantic web technology in general
- Web application developers who are interested in studying the intelligent web technologies and in constructing web applications
- Researchers who are interested in the research and development of intelligent and semantic technologies

- Undergraduate and graduate students in computer science departments, whose main area of focus is the intelligent and semantic web
- Practitioners in related engineering fields

The prerequisites needed to understand the concepts in this book include the following:

- Working knowledge of a programming language
- Basic understanding of the Web, including its main technical components such as URL, HTML, and XML

Salient features

This book has the following salient features:

- Makes all fundamental as well as in-depth material available at one place in an integrated manner
- Provides a more concrete organization than an edited volume
- Incorporates new topics on artificial intelligence (AI), thus making the book more effective and helpful in solving problems
- Integrates illustrations and examples to support pedagogical exposition
- Equips the reader with the necessary information in order to obtain hands-on experience of the topics of discussion
- Facilitates experimentation of the content discussed in the book by making available fundamental tools, research directions, practice questions, and additional reading material
- Integrates all material, yet allows each chapter to be used or studied independently
- Supplies further tools and information at the associated website for instructors and students

Outline of the chapters

The book is organized into four parts. Part I provides an introduction to the Web, machine learning, new AI techniques, and web intelligence.

Chapter 1 describes introductory concepts such as a brief history of the Web and the Internet. It also discusses the latest trends on the Web such as blogs, tweets, wikis, etc. Collaborative mapping, aggregation technologies, open platforms, tools, and application programming interfaces (APIs) are discussed in this chapter. The chapter also describes the organization of the content.

Chapter 2 reviews machine learning that has made its way from AI into web applications and technologies. It presents the capabilities of machine learning methods and provides ideas on how these methods could be useful for web intelligence. The chapter establishes fundamentals such as linear regression, estimation, generalization, supervised learning, unsupervised learning, reinforcement learning, hidden Markov models, and Bayesian networks.

Chapter 3 covers the new AI and knowledge-based system (KBS) and discusses the limitations of the typical symbolic AI and the need of bio-inspired AI for the Web. The most essential and widely employed material pertaining to neural networks, genetic algorithms, fuzzy systems, and rough sets are discussed in brief with their possible advantages.

Chapter 4 explores the basic roles as well as practical impacts of artificial intelligence and advanced information technology for the next generation of web-based systems, services, and environments. The chapter also presents the concept of wisdom web.

Part II is dedicated to information retrieval, mining, and extraction of content from the Web.

Web information retrieval is another important aspect linked to web intelligence. Web spiders, distributed spiders, focused spiders, search engine mechanisms, personalized search techniques, and natural language processing (NLP) in conjunction with effective retrieval are discussed in Chapter 5. This chapter also presents architectures of knowledge-based systems for information retrieval from the Web.

Web mining is the application of machine learning (especially data mining) techniques to web-based data for the purpose of learning or extracting knowledge. Web mining methodologies can generally be classified into one of three distinct categories: web usage mining, web structure mining, and web content mining. Chapter 6 discusses these methodologies along with suitable applications.

Chapter 7 introduces the concept of information extraction to facilitate structured data extraction. Information extraction (IE) is a technology enabling relevant content to be extracted from textual information available electronically. It plays a crucial role for researchers and professionals as well as for other end users who have to deal with vast amounts of information from the Internet. This chapter focuses on wrapper induction as well as semiautomatic and automatic wrapper generation along with a suitable case study.

Part III is dedicated to the semantic web and web knowledge management.

Chapter 8 establishes the semantic web as an immediate extension of the Web in which the meaning (semantics) of content and services on the Web is defined along with the content. Embedding of such semantics makes it possible for the Web to "understand" the content and satisfy the requests of people and machines to use the Web. The chapter discusses metadata, metadata standards, layered architecture of semantic web, and tools and ontology constructs such as resource description framework (RDF), web ontology language (OWL), and extensible markup language (XML). Ontology spectrum, meta-ontology, editors, inference and annotation tools, etc., are also included. It also discusses web applications such as semantic search, social communities, and semantic web research issues.

The Web encompasses a large amount of content organized heterogeneously. For effective retrieval and better access of the content available on the Web, it is necessary to use suitable knowledge representation, knowledge use, and knowledge-sharing techniques. Chapter 9 discusses various knowledge management techniques for the Web. It also suggests a generic architecture on the top of the semantic web for knowledge management.

Chapter 10 combines the concepts and the methods of two fields, namely, the semantic web and social networks, which, together, aid in the analysis of the social web and in the design of a new class of applications that combine human intelligence with machine processing. The chapter presents the application of semantic web technologies to the social web that forms a network of interlinked and semantically enabled content and knowledge. It also provides readers with an understanding of the key concepts and methods of both the fields and describes a simple real-world application incorporating social and semantic metadata.

Part IV discusses additional topics such as agent-based web, security issues, and human–computer interaction.

An agent is an entity that is autonomous, independent, and cooperative. It does intended work on behalf of the user. To carry out various web activities and support web

functionalities in a structured manner, one may take the help of agents. Chapter 11 discusses agent typology, intelligent agents, agents for the Web, web services, and case studies. Considering the technologies discussed within the aforementioned chapters, some agents can be designed to fit into the framework of a multi-agent web. One such possible framework of a multi-agent system is discussed in this chapter. The chapter also elaborates on applications suitable for the framework suggested.

Chapter 12 discusses issues related to web security. It reviews different AI and machine learning methods concerning security, privacy, and reliability issues of cyberspace. It also enables readers to discover the types of methods at their disposal, summarizes the state of the practice in this important area, and provides a classification of existing work. The topics include security management and governance, network security and authentication, intrusion detection, trust management, access control, and privacy.

The expectations from the Web are ever increasing, and the Web will also evolve accordingly. However, the facilities offered by such a giant organization would be made more effective with better interface. Chapter 13 focuses on human–web interactions. It defines web interaction and identifies interaction applications. Topics such as interactive information search/retrieval, interactive query expansion, personalization, user profiling, visualization, user interfaces, usability, web adaptation, and interactive authoring/annotation for the semantic web are discussed in this chapter along with other similar applications.

Use as a book

The book can be covered in a total of approximately 40–45 lecture hours (plus 20–30 hours dedicated to exercises and hands-on practice).

Parts I and II can be covered as a complete course in about 30 taught hours. Such a course requires a significant amount of additional practical activity, normally consisting of several exercises from each chapter and a project involving the design and implementation of a web application.

Parts III and IV can be covered in a second course. They can alternatively be integrated in part within an extended first course. In advanced, project-centered courses, the study of current technology can be accompanied by a project dedicated to the development of technological components. The advanced course can be associated with further readings or with a research-oriented seminar series.

Acknowledgments

The organization and the contents of this book have benefited from our experience in teaching the subject in various contexts. All the students attending those courses, dispersed over many schools and countries (Sardar Patel University, the International School of Information Management, Saint Mary's University, American University of Armenia, and SIBER-India), deserve our deepest gratitude. Some of these students have class-tested rough drafts and incomplete notes and have contributed to their development, improvement, and correction. Similarly, we would like to thank the staff from IT companies and government organizations who attended our courses for professionals and helped us learn the practical aspects that we have tried to convey in this book. We would also like to thank all the colleagues who have contributed, directly or indirectly, to the development of this book, through discussions on course organization or the actual revision of drafts. They include Pawan Lingras, Terje Aaberge, Svein Ølnes, David Camacho, Henry Hexmoor, and Darshan Choksi.

We thank the reviewers of this edition for a number of very useful suggestions concerning the organization of the book and the specific content of chapters.

We also thank Aastha Sharma, David Fausel, Sarah Morris, the staff at CRC Press, and Remya Divakaran (SPi Global) who have contributed to the birth of this book.

Finally, we express our gratitude to our families for their love, support, and patience during the preparation of the book. We also thank our families for reminding us that there are things in life beyond writing books.

Priti Srinivas Sajja

Rajendra Akerkar

Authors

Priti Srinivas Sajja joined the faculty of the Department of Computer Science, Sardar Patel University, Gujarat, India, in 1994 and is presently working as an associate professor. She received her MS (1993) and PhD (2000) in computer science from Sardar Patel University. Her research interests include knowledge-based systems, soft computing, multiagent systems, and software engineering. She has more than 100 publications in books, book chapters, journals, and in the proceedings of national and international conferences. Three of her publications have won best research paper awards. Dr. Sajja is the coauthor of *Knowledge-Based Systems*. She supervises the work of seven doctoral research students. She is also the principal investigator of a major research project funded by the University Grants Commission (UGC), India. She serves as a member on the editorial boards of many international science journals and has served as a program committee member for various international conferences.

Rajendra Akerkar is a professor/senior researcher at Western Norway Research Institute (Vestlandsforsking), Norway. His research and teaching experience includes over 20 years in academia, spanning different universities in Asia, Europe, and North America. As the founder of Technomathematics Research Foundation (TMRF), he is instrumental in ensuring that the organization lends a platform for research in India. Under his leadership, TMRF has become a well-known organization among the research community worldwide.

Akerkar's current research agenda focuses on learning and language—how each works in the human and how they can be replicated in a machine. He received a DAAD fellowship in 1990 and was also awarded the prestigious BOYSCASTS Young Scientist Award of the Department of Science & Technology, Government of India, in 1997. He is the editor in chief of the *International Journal of Computer Science & Applications* and an associate editor of the *International Journal of Metadata, Semantics, and Ontologies*. Akerkar serves as a member of the scientific committees of several international conferences and also serves on the editorial boards of international journals in computer science. He has authored 12 books, more than 90 research papers, and has edited 5 volumes of international conferences. He initiated the International Conference Series on Web Intelligence, Mining and Semantics (WIMS). Akerkar has been actively involved in many industrial research and development projects for more than 14 years.

part one

Introduction to the Web, machine learning, new AI techniques, and web intelligence

chapter one

Introduction to World Wide Web

1.1 Brief history of the Web and the Internet

The World Wide Web, abbreviated as WWW and commonly known as the Web, has been weaving a variety of solutions for different problems and meeting information requirements of a global audience. It is a system of interlinked hypertext documents in multimedia accessed via Internet, which is defined as network of networks. The dream was conceived by Tim Berners-Lee, who is now director of the World Wide Web Consortium and extending the dream project further in a form of semantic web by adding semantics to the existing web. The Web is developed to be a pool of information to allow collaborators from remote sites to share their ideas and information.

During the year 1980, Tim Berners-Lee built ENQUIRE as a personal database of people using hypertext and software utilities to access the database. The objective was to share data globally without common machines and presentation software. He implemented this system on a newly acquired NeXT workstation. After the invention of supporting hypertext transfer protocol (HTTP) and a web browser named World Wide Web, the first web server and page were created that described the project itself. It was further modified to be used on any machine rather than NeXT. On August 6, 1991, Berners-Lee posted a short summary of the World Wide Web project on the alt.hypertext newsgroup. This date also marked the debut of the Web as a publicly available service on the Internet. According to the summary, the World Wide Web (WWW) project aimed to allow all links to be made to any information anywhere. He invited high-energy physicists and other experts to share data, news, and documentation. Inspired from the message, university-based scientific departments and physics laboratories adopted the concept developed such as Fermilab (Fermi National Accelerator Laboratory for high-energy physics, Batavia, IL) and SLAC (Stanford Linear Accelerator Center, Stanford University, Menlo Park, CA).

There was still no graphical browser available for computers besides the NeXT. This gap was filled in 1992 with the release of Erwise (an application developed at Helsinki University of Technology, Finland) and ViolaWWW (created by Pei-Yuan Wei, which included advanced features such as embedded graphics, scripting, and animation). This gave rise to the development of different web browsers. Some prominent early browsers are Mosaic (now Netscape Navigator) and Cello. Immediately after that, in 1994, the World Wide Web Consortium was founded at Massachusetts Institute of Technology (MIT) with the support of Defense Advance Research Project Agency (DARPA) and European Commission. Berners-Lee made the Web available freely, with no patent and no royalties due. By the end of 1994, the total number of websites has increased, however, the increase is minute (in comparison with) the present standards of 15 million index pages approximately. However, by 1996, the usage of the Web was no longer optional. Earlier, people had identified the possibilities of free publishing and instant worldwide information, but, at present, the Web has opened up the possibility of direct web-based commerce and instantaneous group communications worldwide. The innovation of protocols, standards, and utilities like search engine and e-mail made the Web ubiquitous and fall within the reach

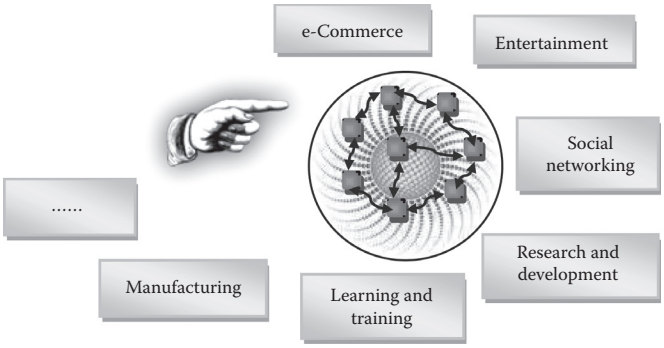


Figure 1.1 Ubiquitous applications of the Web.

of a common man. New ideas such as blogs and social networking are also welcomed in recent times. Some typical application areas of the Web include e-commerce, decision support in various businesses, e-learning, social networking, training, information repository services, manufacturing, and research and development. Figure 1.1 represents ubiquitous application of the Web. Some of the recent innovations are discussed in the next section.

1.2 Blogs

A blog, derived from the term web log (or weblog), is a type of website that is usually maintained by an individual to share details regarding personal views, events, or multimedia materials such as graphics or videos. The term weblog was coined by Jorn Barger on December 17, 1997, to describe the list of links on his Robot Wisdom website that "logged" his Internet wanderings. The short form "blog" was coined by Peter Merholz, who jokingly broke the word *weblog* into the phrase *we blog* in the sidebar of his blog Peterme.com in April–May, 1999. Blog entries are commonly displayed in reverse-chronological order, from most recent entries to the oldest. Blog can also be used as a verb, meaning to maintain or add content to a blog. Many blogs are interactive and allow visitors to leave comments and messages. Such instructiveness makes the blog different from other static websites. Further, a typical website encompasses plenty of web pages, each may have some subpages under a home page following a well-organized approach. A blog is normally a single page of entries by a single author in the most personalized way. There may be archives of older entries optionally to manage the large stuff in a well-organized way. A blog is normally public and accessible to anyone without any formality such as registration. However, to post a personal comment that may require to be processed (replied) further, there is a need of user information such as an e-mail ID. Figure 1.2 represents a typical blog.

Early blogs were simply manually updated components of common websites. However, the evolution of tools to facilitate the production and maintenance of blog is much easier and less technical. Blogs can be hosted by dedicated blog hosting services or on regular web-hosting services. Alternatively, a blog can be developed and maintained using blog software (also called as blogware) like Wordpress (which is a third-party software, see http://wordpress.org/) and Blogsmith (http://www.blogsmith.com/). The blog software supports the authoring, editing, and publishing of blog posts and comments, with special functions for image management, moderation of posts, and comments. All blog software support authoring, editing, and publishing of entries in the following format:

Figure 1.2 Example of a blog.

- Title or headline of the post
- Body, main content of the post
- Reference or link to other article
- Date and time on which the blog is published
- Blog entries can optionally include comments and categories (or tags)

Generally blogs publish focused content on a particular topic such as home design, sports, mobile systems technology, etc. However, there may be some blogs that provide a variety of content and links to plenty of other locations. Most of the blogs have a few things in common. These include

- Heading and main content
- Archives
- Facility for the readers to comment and contact
- Some useful related links (blogrolls)
- Feeds like really simple syndication (RSS), resource description framework (RDF), or atoms
- Excerpts (summary) and plugins for readymade additional functionalities

A blog may have features like trackbacks and pingbacks in order to allow users to comment on blog posts and link to the posts and comment on and recommend them further. Track back is a way to provide notification between different websites. If a person finds the specific blog content interesting, he may send trackback *ping* to the Web author. Pingback offers the advanced facility of automatically notifying the author that the other person referred the author's post. Pingbacks can be automated and do not send content, whereas trackbacks are manual and send an excerpt of the initiating post. The facility of trackback and pingback aim to provide some control on comments on blog content and hence help in extending authority to blog commenting. Some tools offer a feature of

comment moderation to monitor and control the comments on the different article posts. Here authors (blog publishers) are given rights to manage the comment spam, delete harsh comments, and approve cool comments. Finally, it must be remembered that blogs require continuous, regular, and meaningful update.

1.3 Tweets

A tweet is a small message, post, or status update on some network. The size of a tweet is comparatively smaller than the typical e-mail. The "tweet" originally means a sound of a bird or whistle, which sweetly says something. Tweet is also considered as a real-time microblogging service. The term tweet became popular by a social networking website called Twitter (Twitter Inc., San Francisco, CA) created in 2006 by Jack Dorsey who is an American software architect and businessman. This site offers facility to send and receive tweets up to 140 characters. That is why it is sometimes described as the short message service (SMS) of the Internet. Generally, tweets are displayed on the user's home page or profile page of the website and are visible by default free of cost. However, it may be restricted to the user's friends list. Other users may choose to read and subscribe (opt) for some specific users tweets. These subscribers are known as followers. The process of subscribing a user's tweet is known as "following." Alternatively, such tweets can be followed on compatible external applications such as smart phones or by SMS available. It is reported that Twitter currently has more than 175 million users. The original project code name for the Twitter service was twttr, inspired by Flickr and the five-character length of American SMS short codes. Flickr is an image-hosting and video-hosting website for online community created by Ludicorp and later acquired by Yahoo! It is a popular website to share and embed personal photographs and images. The basic advantage of using the facility to tweet a message or an item is that one can "post once" and the service will redistribute the item to multiple followers.

Twitter's Application Programming Interface (API) is based on the representational state transfer (REST) architecture. REST is a collection of network design principles and guidelines that define resources and ways to address and access data. With the REST architecture, the Twitter works with most web syndication formats that help in gathering information from one source and sends it out to various destinations. Twitter is compatible with two of them—RSS and atom syndication format (atom). Both formats retrieve data from one resource and send it to another. A web page administrator can embed RSS/atom code into the code of his or her site. Visitors can subscribe to the syndication service—called feed—and receive an update every time the administrator updates the web page. Twitter uses this feature to allow members to post messages to a network of other Twitter members. In effect, Twitter members subscribe to other members' feeds. By allowing third-party developers partial access to its API, Twitter allows them to create programs that incorporate Twitter's services.

Current third-party applications include the following applications:

- *Twitterlicious* and *Twitterific*, two applications that allow users to access Twitter through desktop applications on PCs and MACs, respectively
- *OutTwit*, a Windows application that allows users to access Twitter through the Outlook e-mail program
- *Tweet Scan*, which allows users to search public Twitter posts in real time
- *Twessenger*, which integrates with the Windows Live Messenger 8.1 instant messenger program

- *Twittervision*, which integrates a Twitter feed into Google Maps
- *Flotzam*, which integrates Twitter with Facebook, Flickr, and blogs
- *iTunes to Twitter*, an application that broadcasts the title of the song currently playing in the user's iTunes to his or her network
- *TwitterBox*, a Twitter application that works inside the virtual community of Second Life

The ability to send and receive multiple tweets without any limit and sophisticated interface made twitter very popular in societies and industries. Tweeting is a tool for accessing opinions, decisions, and market information. Tweeting can be used as a very good communication and business tool. It can be used in the field of education for classroom community and collaborative writing within and across schools and institutes.

1.4 Wikis

A Wiki is a collaborative web platform that allows multiple users from different locations to add information at a centralized place in an interactive fashion. The interface for interaction is created in such a way that users do not require any training. The wiki concept was first introduced by Ward Cunningham as "the simplest online database that could possibly work" in the year 1995. This is also considered as writable web or open editing concept. Wiki looks like a simple website with an edit facility/link. Readers of the page can modify it or add content directly through this link. Wikis can be used as a centralized repository for many applications and provides efficient document management. Wiki can be used as a website, as a knowledge base for FAQ system, and as an information sharing utility. As many authors contribute and evaluate information for a specific topic, wiki can enhance the quality and quantity of information. Table 1.1 shows some typical applications that can be benefited by wikis.

On request, wiki shares available information and allows editing of the information through any browser. It also invites users to create new pages and promotes a specified topic. Wiki can create links to new empty pages that do not yet exist to invite users to share their opinion. There are many different ways in which users can edit the wiki content. Figure 1.3 gives an idea how users can work with wiki.

The content of wiki is normally specified with a simplified markup language, sometimes known as *"wiki text*." Some wikis allow dedicated content such as a repository of images or collections of audio. Common facilities that a wiki typically provides is creating new pages, editing content, navigation, and searching. Sometimes external search engine facility can be embedded within a wiki. Majority of wikis provide limited access to hypertext markup language (HTML) and cascade style sheet (CSS). This is the prime reason that generally wikis look simple. Many modern wikis are making "What You See Is What You Get" (WYSIWYG) editing available to users through scripts and controls

Table 1.1 Applications Benefited by Wikis

Project management and up-to-date project status information
Tutorials and e-learning
Internal notice board and institutional news circulation
FAQ management
Online document management and information sharing
Managing groups and social interactions

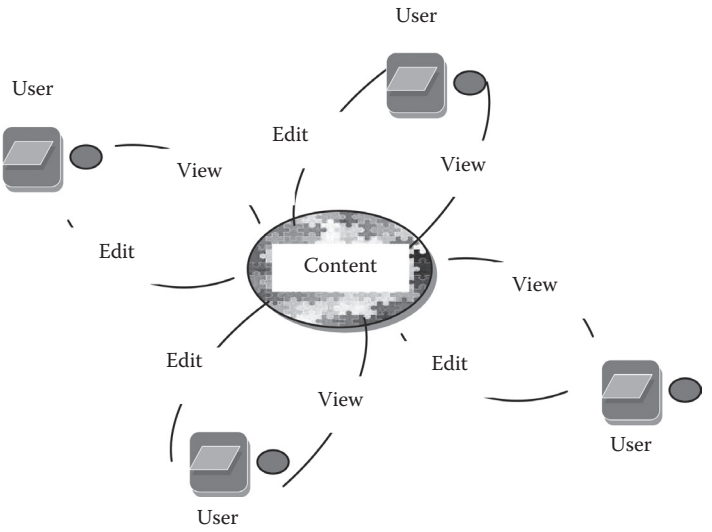


Figure 1.3 Working with wikis.

(like JavaScript or ActiveX controls). The scripts and controls translate the formatting instructions into appropriate wiki text. The translation is transparent and hidden from users to avoid complexity.

Most wikis maintain a log of changes made to wiki pages and maintain versions of the pages. The administrator or author may revert to the older version. With the help of such a revert facility, authors can revert to an older version of the page. Some wikis like MediaWiki insist contributors to provide separate edit summary when they edit a page. The edit summary is an important parameter that justifies the editing and helps in taking the decision whether the page is to be reverted back or not.

For implementation of such collaborative and sharable platform, wiki software is used. The wiki software is implemented on a server and can be executed on different web servers. The content is stored in file systems and changes are stored in a relational database for version management. Some software that are helpful in creating wikis are given in Table 1.2.

Wiki can be developed and used as a standalone system. Such wikis are known as personal wikis. The example of such a personal wiki is the Wikidpad.

1.4.1 Improving wiki content reliability, quality, and security

In spite of its simplicity and advantages, wiki is not considered a reliable source of information. As anybody can contribute in such a system, the content of the system is easily tampered. However, there are genuine contributors and editors, who catch such malicious tampering and correct the mistakes and destructive content for the benefit of others. Many times, users contribute the content that they thought correct and genuine, such as ideas, research information, and experiments. It is not advisable to change such information; however, in this situation, one or more versions showing suggestions and changes can be prepared and a link can be provided adjacent to this. This strategy is also applicable if contributor to wiki does not have rights to edit the content. To improve the reliability, the content pages are ranked by administrator, readers, and contributors. There are

Table 1.2 List of Software for Wiki

ConcourseConnect is a freely available J2EE application with social networking, online community, business directory, and customer relationship management capabilities. This tool supports features like wiki, blog, document management, reviews, online advertising, and project management modules
DokuWiki is a simple-to-use Wiki to facilitate documentation management needs within a small institution. It uses plain text files and has a simple but powerful syntax, which ensures the data files remain readable outside the Wiki
MediaWiki is a popular free web-based wiki software application. It is developed by, and it runs, all the projects of the Wikimedia Foundation, including Wikipedia, Wiktionary, and Wikinews. It is written in the hypertext processor (PHP) programming language and uses a backend database
TiddlyWiki is a HTML/JavaScript-based wiki in which the entire site/wiki is contained in a single file. This tool does not require any server support
Wikidpad is a freeware (open source) personal-use wiki with native support of international characters (Unicode). This software is executable on a single machine. It helps in the implementation of features like storing thoughts, ideas, to-do lists, contacts, and other notes with wiki-like linking between pages
Windows SharePoint Server 2010 has built-in Wiki support. It is built on ASP.Net, C#, and Microsoft SQL Server
Wikia (formerly known as Wikicities) is a free web-hosting service for wikis

sophisticated software programs that help in improving reliability, security, and quality of wiki content by checking and imposing constraints on content and contributors. Table 1.3 suggests some ideas to improve trustworthiness and quality of wiki content.

Wikis that allow only registered users are known as closed wikis. In general, wiki allows anonymous editing by just recording IP address of the machine from which editing is done. In closed wikis, only registered users, whose information (like name and biography) is formally recorded, can edit the wiki content. With this strategy, wiki content is reliable and controllable; however, growth of content on wiki is slower. The popular Wikipedia is an open wiki that allows anonymous editing and records only IP addresses. The countermeasures shown in Table 1.3 help in preventing attacks from malicious contributors, vandalism content, harmful code, and bugs. Besides these, an edit war may occur

Table 1.3 Improving Reliability, Quality, and Security of Wiki Content

Improvement	Countermeasures
Improving reliability of content	By correcting the content manually with the help of experts and administrators
	By providing links to the edited version of the content
	By accreditation of users as well as content
Improving quality	By frequent editing, ranking, and filtering the content
Improving security	By allowing only registered users to edit the content
	By imposing requirement of additional waiting period before allowing any edit
	By dedicated software support (such as JavaScript) to automatically find vandalism, bugs, and harmful content
	By preparing a list of malicious sites, if any, within the content pointing to any of the sites from the list