# CONTENT NETWORKING

## ARCHITECTURE, PROTOCOLS, AND PRACTICE

MARKUS HOFMANN & LELAND R. BEAUMONT

# Content Networking

Architecture, Protocols, and Practice

## The Morgan Kaufmann Series in Networking

Series Editor, David Clark, M.I.T.

*Content Networking: Architecture, Protocols, and Practice*
Markus Hofmann and Leland R. Beaumont

*Network Algorithmics: An Interdisciplinary Approach to Designing Fast Networked Devices*
George Varghese

*Network Recovery: Protection and Restoration of Optical, SONET-SDH, IP, and MPLS*
Jean Philippe Vasseur, Mario Pickavet, and Piet Demeester

*Routing, Flow, and Capacity Design in Communication and Computer Networks*
Michal Pióro and Deepankar Medhi

*Wireless Sensor Networks: An Information Processing Approach*
Feng Zhao and Leonidas Guibas

*Communication Networking: An Analytical Approach*
Anurag Kumar, D. Manjunath, and Joy Kuri

*The Internet and Its Protocols: A Comparative Approach*
Adrian Farrel

*Modern Cable Television Technology: Video, Voice, and Data Communications, 2e*
Walter Ciciora, James Farmer, David Large, and Michael Adams

*Bluetooth Application Programming with the Java APIs*
C Bala Kumar, Paul J. Kline, and Timothy J. Thompson

*Policy-Based Network Management: Solutions for the Next Generation*
John Strassner

*Computer Networks: A Systems Approach, 3e*
Larry L. Peterson and Bruce S. Davie

*Network Architecture, Analysis, and Design, 2e*
James D. McCabe

*MPLS Network Management: MIBs, Tools, and Techniques*
Thomas D. Nadeau

*Developing IP-Based Services: Solutions for Service Providers and Vendors*
Monique Morrow and Kateel Vijayananda

*Telecommunications Law in the Internet Age*
Sharon K. Black

*Optical Networks: A Practical Perspective, 2e*
Rajiv Ramaswami and Kumar N. Sivarajan

*Internet QoS: Architectures and Mechanisms*
Zheng Wang

*TCP/IP Sockets in Java: Practical Guide for Programmers*
Michael J. Donahoo and Kenneth L. Calvert

*TCP/IP Sockets in C: Practical Guide for Programmers*
Kenneth L. Calvert and Michael J. Donahoo

*Multicast Communication: Protocols, Programming, and Applications*
Ralph Wittmann and Martina Zitterbart

*MPLS: Technology and Applications*
Bruce Davie and Yakov Rekhter

*High-Performance Communication Networks, 2e*
Jean Walrand and Pravin Varaiya

*Internetworking Multimedia*
Jon Crowcroft, Mark Handley, and Ian Wakeman

*Understanding Networked Applications: A First Course*
David G. Messerschmitt

*Integrated Management of Networked Systems: Concepts, Architectures, and their Operational Application*
Heinz-Gerd Hegering, Sebastian Abeck, and Bernhard Neumair

*Virtual Private Networks: Making the Right Connection*
Dennis Fowler

*Networked Applications: A Guide to the New Computing Infrastructure*
David G. Messerschmitt

*Wide Area Network Design: Concepts and Tools for Optimization*
Robert S. Cahn

**For further information on these books and for a list of forthcoming titles, please visit our Website at http://www.mkp.com.**

# Content Networking

## Architecture, Protocols, and Practice

Markus Hofmann and Leland Beaumont

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
MORGAN KAUFMANN PUBLISHERS IS AN IMPRINT OF ELSEVIER



MORGAN KAUFMANN PUBLISHERS

*Publishing Director* Diane Cerra
*Senior Acquisitions Editor* Rick Adams
*Developmental Editor* Karyn Johnson
*Assistant Editor* Mona Buehler
*Publishing Services Manager* Simon Crump
*Project Manager* Justin R. Palmeiro
*Cover Design* Yvo Riezebos Design
*Composition* Kolam
*Copyeditor* Kolam USA
*Proofreader* Kolam USA
*Indexer* Kolam USA
*Interior printer* Maple Press
*Cover printer* Phoenix Color

Morgan Kaufmann Publishers is an imprint of Elsevier.
500 Sansome Street, Suite 400, San Francisco, CA 94111

This book is printed on acid-free paper.



Figure credit: Image clips in Figure 6.9 used with permission.



**Library of Congress Cataloging-in-Publication Data**
Hofmann, Markus.
Content networking : architecture, protocols, and practice / Markus Hofmann and Leland Beaumont.
p. cm. — (The Morgan Kaufmann series in networking)
Includes bibliographical references and index.
ISBN 1-55860-834-6
1. Computer networks. I. Beaumont, Leland R. II. Title. III. Series.
TK5105.5.H63 2005
004.6—dc22
2005001732

ISBN: 1-55860-834-6

For information on all Morgan Kaufmann publications,
visit our Web site at www.mkp.com or www.books.elsevier.com

Printed in the United States of America
05 06 07 08 09 5 4 3 2 1

*Dedicated with great affection to my wife Bettina and our kids Jennifer, Dennis, and Kevin for their love and support, and to my parents for preparing me to take on such an endeavor.*

*– Markus Hofmann*

*Dedicated to my parents, who prepared me to write this, and to my wife Eileen, daughter Nicole, and son Rick, for their encouragement and support while writing it.*

*– Leland Beaumont*

# Contents

# Preface

## Why This Book?

People are sociable. They want to stay in touch with each other, share their experiences, and exchange information regarding their common interests. When Markus and his wife moved to the United States a few years ago, the Internet and the Web became their main means to stay in touch with family and friends back in Germany. E-mail, a Web page with guestbook, and instant messaging allowed timely and very effective exchange of the latest gossip. Photos from recent happenings were uploaded to a Web page and shared minutes later. A little later, the first personal video clip found its way from the digital camcorder onto the Web page, allowing even livelier information sharing across the continents.

Soon, however, the limitations of the underlying technology became obvious. Parents and friends back in Germany started to complain about long download times, unavailable Web servers, long playback delays, and the choppy quality of video clips. Knowing our research and work interests, they posed the challenge of helping to overcome these problems: "Hey, you are working on data networking and telecommunications—why can't you produce something useful and help solve these problems?" A team at Bell Labs/Lucent Technologies—our employer at that point in time—took the challenge and worked on designing and developing solutions to overcome the slowdown on the World Wide Web. It is a very exciting effort, which brought Bell Labs Researchers together with system engineers, developers, and sales personnel from Lucent Business Units—working hand in hand, collaborating very closely, and leveraging each other's experiences and strengths. This was also the time when Markus and Lee met, embarking on their very exciting journey into the space of Content Networking.

People are curious. They want to understand and learn about issues that affect and impact them. When we first demonstrated the exciting results of the team's work, people started to ask how it works, what was done, and how it will help improve the scalability and reliability of Internet services. Motivated by this interest, we wrote this book to help people understand the reasons for current problems in the Internet and to explain both the challenges and possible solutions for building a more reliable and scalable Internet. Markus has been working as a researcher in content delivery and related fields for more than 10 years and has gained valuable practical experience,

which he would like to pass on to the readers of this book. His colleague, Leland Beaumont, has 30 years of experience in developing data network systems—an invaluable asset when bringing ideas from the research lab into the real world.

## Audience

The Internet, and in particular the World Wide Web (WWW), have become an integral part of people's lives. With the increase in popularity, however, users face more and more problems when using the Internet, such as long access delays, poor quality of service, and unreliable services. This book is aimed at helping practitioners and researchers working with network service providers, software and hardware vendors, and content providers to understand the reasons for these problems. It explains the challenges in making content available on the WWW, describes basic concepts and principles for improving the current situation, and outlines possibilities for tapping into the huge potential of custom-tailored services over the Internet. In particular, the book describes the pressures that caused the Internet to evolve from the original End-to-End model to a more complex model that has intelligence embedded within various intermediaries placed throughout the network.

## Approach

The book starts with a discussion of fundamental techniques and protocols for moving content on the Internet, followed by an introduction to content replication and Web caching. From there, the book outlines the evolution from traditional Web caching towards a flexible and open architecture to support a variety of content-oriented services. Evolutionary steps include support for streaming media, systems for global request routing, and the design of APIs and protocols that enable value-added services, such as compression, filtering, and transformation. Content navigation, peer-to-peer networks, instant messaging, content services, standards, and future directions are all discussed. The book also explains how the different components interact with each other and how they can be used to build complex content delivery networks.

We hope the reader will learn how the technology evolved from traditional Web caching to more sophisticated content delivery services. The reader will get a better understanding of the key components in modern content delivery networks and the protocols that make the components interact with each other. Various examples are provided to help the reader to better understand how this technology can be deployed and how it could help their business.

The book concentrates mainly on underlying principles, concepts, and mechanisms and tries to explain and evaluate them. While the specific protocols, interfaces, and languages used in content networking will continue to evolve and change, it is expected that the core principles and concepts underlying content networks will remain valid for a long time. As such, the book focuses on principles and attempts

to explain and evaluate them. Specific protocols and languages are selected as examples of how the concepts and mechanisms can be incorporated into real-life networks. It uses many examples and case studies for illustration. The book is *not* intended as a reference guide to Web-related protocols, but as a guide providing a systematic and architectural view of the content delivery and content services field. It helps the reader to understand the overall picture and how all the components fit together. The examples are timely and the principles remain timeless.

Much of the design of the Internet is described in freely available documents known as RFC, Requests for Comment. These are relied on heavily as references through the book. RFCs are dynamic. Some classics remain as useful, accurate, and pertinent as they were when they were written a decade or more ago. Others may be superseded before this book completes its first printing. Readers working in the field need to stay abreast of changes as they unfold.

## Content

The first chapter serves as introduction to the remainder of the book. It explains the notion of content networking and establishes the key concepts. A brief look at the early days of information access over the Internet establishes the roots of modern content networking—the World Wide Web. The chapter continues with a flashback to the first half of the 1990s, with a history of the Web setting the stage for a discussion of fundamental concepts and principles.

The rest of the book takes us on a journey that follows the evolution of content networks.

Chapter 2 explains the core principles that guided the design of the Internet, leading into a discussion of how content is transported over the Internet. The focus is on the *Hypertext Transfer Protocol (HTTP)* and some of the features that will be important in later chapters of the book.

Chapter 3 shows how Web caching is used to bring static content closer to the users and how this helps in improving content delivery over the Internet. These first three chapters form the foundation for the balance of the book.

Chapter 4 stands alone and extends Web caching to include streaming media such as audio and video. Optimized techniques are introduced that take into account the special characteristics of time-constrained streaming media.

Chapter 5 deals with the question of how user requests actually get to the server or Web cache best suited to serve each user. Different metrics for evaluating closeness in content networks are introduced and different mechanisms for request routing are explained.

Chapter 6 introduces the new concept of peer-to-peer networks, in which the traditional client-server model of the Web is replaced with a federation of end-systems that help each other in delivering content. Chapters 4, 6, and 7 each stand alone and may be reading any order, or skipped entirely at the discretion of the reader.

Chapter 7 extends the notion of content networking to include delivery of interactive media, such as instant messaging. The chapter explains a variety of standards-

based and proprietary approaches that enable people to interact with each other in (near) real-time.

Chapter 8 is the centerpiece of the book, describing Content Services. After developing an architecture for content services, two similar approaches are introduced. These are the Internet Content Adaptation Protocol (ICAP) and the Open Pluggable Edge services (OPES), the latter one being standardized in the IETF. The W3C sponsored approach to Web-based services is then described. Finally, the wide range of services made possible by the convergence of Web services and traditional telephony are described.

Chapter 9 brings the various technologies and network elements together, and explains how they can be deployed to build content networks for specific needs.

Chapter 10 provides an overview of the various standards activities relevant to the field on content networking, and explains which efforts are of interest for each specific area.

Chapter 11 finally summarizes our journey through the evolution of content networks and attempts to provide an outlook of what the future might bring.

A glossary at the back of the book includes terms that are unique to this content area.

The focus of this book is on the architectures and protocols specific to content networks. It cannot address every single topic in depth. Therefore, the book does not address other relevant topics such as security issues surrounding content networks or the operation and management of content networks.

A companion Web site for this book exists at http://www.content-networking.com/ or at www.mkp.com/?isbn=1558608346. At this site, you will find additional support material to enhance reading of the book. We suggest that you visit the page for this book every so often, as we will be adding and updating material and establishing new links to content networking related sites on a regular basis.

In this spirit, start the engines, get rolling, and have fun!

## Acknowledgments

Clearly, an undertaking such as this book is impossible without the support of others. First and foremost, we thank our families—Bettina with Jennifer, Dennis and Kevin, and Eileen with Nicole and Rick—for their patience, their sympathy, and their continued encouragement during the sometimes stressful period of writing this book. The book was written in addition to the commitments of our daytime jobs, written exclusively on personal time—nights, early mornings, and weekends. It is time now to make up for some of the lost weekends with beach visits, hikes, biking tours and canoe and kayak trips.

Many colleagues and co-workers have given us inspiration—too numerous to mention individually. However, we would like to say special thanks to Wayne Hatter, whose calm, yet determined and highly motivational leadership helped transition some of the concepts presented in this book into real-world products. Wayne represents an entire team of excellent and bright developers that we had the pleasure to

work with. We also thank our management at Bell Labs Research, in particular Krishan Sabnani and Sudhir Ahuja, for their encouragement in finishing this book. A special thanks also to Sanjoy Paul, who has been key in starting our efforts around Content Networking.

The book was made possible only by our own excitement and enthusiasm for the Content Networking space, fueled even more by active participation and involvement in several international standardizations efforts—not a trivial task! Our colleague Igor Faynberg provided helpful hints and tips on how to move our work into the respective standards bodies. Guidance from Allison Mankin ensures that the work on content services being done by others and ourselves is sensitive to the existing Internet architecture. Acknowledgments also go to Michael Condry and Hilarie Orman, who inspired much of the work in the content services field.

The thoughtful and detailed comments of our manuscript reviewers—including Mark Nottingham, Alex Rousskov, Michael Vernick, and Martin Stecher—have greatly strengthened the final result. Their critique and suggestions have prompted improvement in the structure of the book and addition of new subjects. A big "Thank you!" for their help.

We also wish to acknowledge the editorial staff at Morgan Kaufman/Elsevier for a great job in giving the book this professional touch. Karyn Johnson has to be thanked for her extreme patience and persistence getting work back on track after deadlines slipped. Rick Adams deserves much credit for having the courage to ask us to write this book.

The content of this book is based on several tutorials and graduate lectures, which Markus gave before and during the preparation of this book. Notably, we thank the tutorial chairs and organizers of ACM Multimedia, NGC Workshop, World Wide Web Conference, and IEEE ICNP for the opportunity to present tutorials accompanying this book. Likewise, we thank the professorship (in particular Martina Zitterbart) and the administration of University of Braunschweig, Germany, and University of Karlsruhe, Germany, for the opportunity to present two 5-day graduate lectures based on the content of this book.

## Growing Together

Most of the book was written in the wonderful and vibrant shore region of New Jersey. Other parts were written during trips in places around the world, including Karlsruhe (Germany), Juan Les Pins (France), Yokohama (Japan), San Jose (Costa Rica), London (England), Boulder (Colorado, USA), Los Angeles, San Francisco, San Diego (California, USA), Atlanta (Georgia, USA), and mid-air between several of these places. Hopefully, the technology described in this book will help people in all these places and around the world to grow together even stronger.

Markus Hofmann and Leland Beaumont,
New Jersey, USA, September, 2004

# About the Authors

*Markus Hofmann* is Director of Services Infrastructure Research at Bell Labs/Lucent Technologies. He received his PhD in Computer Engineering from University of Karlsruhe, Germany, in 1998 and joined Bell Labs Research the same year. Currently, he is also an Adjunct Professor at Columbia University in New York, USA. Markus is known for his pioneering work on reliable multicasting over the Internet and for defining and shaping fundamental principles on content networking. He is Chair of the Open Pluggable Edge Services (OPES) Working Group in the IETF since it has been chartered in 2002. More recently, Markus' work has extended into the areas of VoIP and converged communications. Markus is on the Editorial Board of the Computer Communications Journal, has recently been elected chair of the Internet Technical Committee (ITC), and has published numerous papers in the multicasting and content delivery area. His PhD thesis won the 1998 GI/KuVS Award for best PhD thesis in Germany in the area of Telecommunications, and also the 1998 FZI Doctoral Dissertation price awarded by the German Research Center for Computer Science. More information is available at www.mhof.com.

*Leland Beaumont* consults on quality management and product development. Prior to that, he was responsible for specification and verification of content delivery products at Lucent, including Web caching and content network navigation. After graduating with highest honors from Lehigh University, he received his Master of Science degree in Electrical Engineering from Purdue University. He has worked in the data communications product development industry for over 30 years.

CHAPTER 1

# Introduction

Over the last few decades, the Internet has revolutionized our society and our economy. It has changed the way people communicate with each other and the way business is conducted. The Internet has created a global environment that is drawing people from all over the world closer together. Collaboration and interaction of individuals through their networked computers have been main applications on the Internet since the beginning. Electronic mail and Internet chat rooms are just two examples of popular applications. Over the last decade, the Internet has been used ever more as a mechanism for information dissemination and broadcasting, mainly driven by the emergence of the World Wide Web—also referred to as WWW or the Web. The Web forms a universe of information accessible via networked computers, offering *content* in the form of Web pages, images, text, animations, or audio and video streams. This book examines the technical concepts and the challenges of distributing, delivering, and servicing content over the Internet. Business-related aspects are considered when they have impact on the underlying technology. The focus is on fundamental principles and concepts rather than providing a reference for specific communication protocols or implementation details.

The first chapter serves as an introduction, explaining the notion of content networking and establishing the underlying key concepts. A brief look at the early days of information access over the Internet segues to the roots of modern content networking—the World Wide Web. The chapter continues with a flashback to the first half of the 1990s, with a history of the Web setting the stage for a discussion of underlying concepts and principles. These include the *representation*, *identification,* and *transport* of Web objects, which are most often referred to as Hypertext Markup Language (HTML), Universal Resource Identifier (URI), and Hypertext Transport Protocol (HTTP), respectively. The power of URIs and hyperlinks allows a variety of protocols to link new content types together and add richness to the original WWW. For example, other protocols such as RTSP and RTP allow other object types, such as multimedia streams, to be

linked into the WWW. The chapter continues looking at Web applications as a driving force for the evolution of the Web and for adopting new technology. It identifies the shortcomings of today's Web architecture and outlines an evolutionary path toward advanced communication architectures of the future. The technology-focused part is complemented with a description of the various Web beneficiaries and their diversity of interests. The chapter concludes with a tour through the book that outlines the remaining ten chapters.

## 1.1 The Early Days of Content Delivery over the Internet

Until about a decade ago, most of the world knew little or nothing about the Internet. It was used largely by the scientific community for sharing resources on computers and for interacting with colleagues in their respective research fields. When work on the ARPANET—the origin of today's Internet—started in the late 1960s and the 1970s, the prevailing applications were as follows: access to remote machines, exchange of e-mails, and copying files between computers. Electronic distribution of documents soon gained importance, as it became apparent that the traditional academic publication process was too slow for the fast-paced information exchange essential for creating the Internet. When the *File Transfer Protocol (FTP)* [Bhu71, RFC 959] came into use in the early 1970s, documents were prepared as online files and made accessible on servers via FTP. Interested parties used an FTP client program to establish a connection to the server for downloading the document. Over the years, FTP evolved into the primary means for document retrieval and software distribution over the Internet. In the early 1990s, FTP accounted for almost half of the Internet traffic [Mer1].

However, FTP did not solve all the problems related to information retrieval over the Internet—it enabled downloading files from remote machines, but it did not support users facing the daunting task of navigating through the Internet and in locating relevant resources. Retrieving documents via FTP required users to know in advance the exact server to contact and the name of the file to download. Knowing just the title and the authors of a research paper, for example, was not sufficient for retrieving an electronic copy of the paper. Moreover, the user was required to figure out which FTP server was storing the paper and which file name had been used. The Internet worked very much like a library without a catalog or index cards—users had to know where to look to find the content they needed.

Locating relevant files on the Internet was simplified to some extent with the introduction of *archie* in 1991 [ED92]. The archie system made use of a special "anonymous" account on FTP servers, which gave arbitrary users limited access without having to enter a password. Using these "anonymous" accounts, archie servers periodically searched FTP servers throughout the Internet and recorded the names of files they found. This information was used to create and maintain a global catalog of files available for download. Users could use this catalog to search for file names matching certain patterns. When matches were found, archie also indicated the FTP servers on which the files were available.

A major restriction of archie was its limitation to pattern matching on file names rather than the actual content of the files. The *Wide Area Information Server (WAIS)* project [KM91] implemented a more powerful concept by searching through the text of documents in addition to their file names or titles. Suppose you are interested in finding articles on Michael Jordan's second comeback to professional basketball, and you perform an archie search using "Jordan" as your keyword. Even if the file named "NBA-News-September-2001.txt" includes a story covering Jordan's comeback, it would not turn up under an archie search. As WAIS digs through the entire text of the article, that file would appear with a WAIS search. Moreover, the WAIS mechanism provided a scored response, ranking retrieved information based on the quantity of keyword appearances in the text and on how close to the document's beginning they turned up. WAIS was originally developed at the beginning of the 1990s by a consortium of companies that included Thinking Machines Inc., Dow Jones, Apple Computer, and KPMG Peat Marwick. The first version of WAIS was available in the public domain in 1991. By summer 1992, the project had evolved into a separate company called—not surprisingly—WAIS Inc. This company can be considered the first to commercialize technology related to content retrieval over the Internet.

However, the WAIS system was not perfect—the user interface was relatively difficult to use and the search capabilities were initially limited to text documents. Besides, it scored documents based on the absolute number of keyword appearances rather than the density of their appearance. As a result, long documents were more likely than short documents to end up at the top of the list. WAIS further lacked the capability for hierarchical organization of content resources—a feature introduced by the *Gopher* system [RFC 1436].

Gopher was developed at the University of Minnesota in 1991 and named after the school's furry mascot. It let users retrieve data over the Internet without using complicated commands and addresses. Gopher servers searched the Internet using WAIS and arranged the results in hierarchical menus, using plain language. As users selected menu items, they were lead to other menus, files, or images, which might not even have resided on the local Gopher server. References could move users to remote servers or fetch files from distant locations. Gopher significantly simplified information retrieval on the Internet. It handled the details of actually getting requested information, without requiring users to know how and from where to retrieve those resources. Initially deployed only on the University of Minnesota campus, other institutions quickly discovered Gopher's versatility and set up their own Gopher servers. At one time, there were a few thousand Gopher servers registered with the top-level server "Gopher Central" at the University of Minnesota or its counterparts in other countries.

Archie, WAIS, and Gopher emerged in the same era and coexisted for some time. They all had their advantages and disadvantages, and occasionally, they are still used today. Nevertheless, in the course of the 1990s, they all were subsumed into yet another system—the World Wide Web (WWW).

## 1.2 The World Wide Web—Where It Came From and What It Is

The *World Wide Web* is an Internet facility that links information accessible via networked computers. This information is typically represented in the form of Web pages, which can contain text, graphics, animations, audio/video, and hyperlinks. Embedding hyperlinks in documents is an important feature of the Web and differentiates it clearly from Gopher and other approaches. Embedded hyperlinks connect a Web page to other resources either locally or on remote computers. Users can follow the links and access referenced resources simply by pointing to the hyperlink and clicking a mouse button. This intuitive mechanism allows browsing through a collection of information resources without having to worry about their actual location or their format.

This section will briefly describe the origin of the Web, where it came from and why it has been so successful. A description of the architectural components will help in the understanding of the fundamental design of the Web and, at the same time, motivate the evolution of the Web. A detailed introduction to the Web can be found in [KR01].

### 1.2.1 The Origin of the World Wide Web

The World Wide Web has its origin at the European Organization for Nuclear Research (CERN) near Geneva, Switzerland. It was initially proposed by Tim Berners-Lee in 1989 to improve information access and help communication within the particle physics community [Ber89]. The community included several hundred members all scattered among various research institutes and universities. Although the groups were formally organized into a hierarchical management structure, the actual working and communication structure looked more like a loosely coupled mesh whose linkages evolved over time. A researcher looking for specific information was typically given a few references to experts who may prove helpful. In order to get the desired information, the researcher used the provided information to contact the respective colleagues. While this communication scheme was principally working fine, a high turnover of people made project record keeping and locating expertise increasingly difficult. A solution was required that would support dynamic, non-centralized interaction and quick access to documents stored at secluded locations.

In this situation, Tim Berners-Lee proposed to his management the idea of using *hypertext* for linking information available on individual computers [Ber89]. The hypertext concept had been envisioned earlier as a method for making computers respond to the way humans think and require information [Bus45, Nel67, EE68]. Hypertext documents embed so-called hyperlinks, which can be represented as underlined text or as icons in any size and shape. By selecting and clicking on a hyperlink, associated information is loaded and displayed. Tim's proposal extended the hypertext concept to allow linking of information not only on a single local machine, but also of information that can be stored on

remote computers connected via a network. Retrieving the associated information over the network is transparent to the user, without burdening the user with having to know the resource location and the network protocol to be used for retrieval. This scheme proved to be very powerful as it allows users transparent accesses to documents on remote computers with a click of the mouse.

The CERN management approved the proposal and launched the project in the second half of 1990. Tim started implementing a hypertext browser/editor and finished the first version at the end of 1990. The program was running on a NeXT computer and offered a graphical user interface. It was called WorldWideWeb but later renamed Nexus to avoid confusion with the abstract concept of the World Wide Web itself. At the same time, the implementation was complemented with a separate line-mode browser written by CERN student Nicola Pellow. Other people soon started implementing browsers on different platforms. By 1992, first versions of Erwise, ViolaWWW, and MidasWWW were introduced for the X/Motif system, followed by a CERN implementation for the Apple Macintosh in 1993.

At that time, there were around 50 known Web servers deployed, and the WWW was accounting for about 0.1% of the Internet traffic. It was a promising approach, but the real breakthrough came with the creation of Mosaic, the first widespread graphical Web browser. Mosaic development was started at the National Center for Supercomputing Applications (NCSA) by Marc Andreesen and Eric Bina. They realized that broad acceptance of Web technology would require a more user-friendly interface. Their browser software added clickable buttons for easy navigation and controls that let users scroll through text. More important, Marc and Eric were the first ones to get embedded images working. Earlier browsers allowed viewing of pictures only in separate windows, while Mosaic made it possible for images and text to appear in the same window. The application was trivial to install and the team followed up coding with very fast customer support. Overall, Mosaic drastically simplified the first step onto the Web and even allowed beginners to take advantage of the new, exciting Web technology. The Unix version of Mosaic was available for download from NCSA in early 1993. The software was provided free of charge and within weeks tens of thousands of people had downloaded it. Software versions for the PC and Macintosh followed later the same year, boosting its popularity even further. The Web started eclipsing competing systems, as it subsumed their main features and functionality. Users could conveniently access FTP servers as well as Archie, WAIS, and Gopher from their Web browsers, thus eliminating the need for these specialized applications.

By 1994, Marc and Eric had graduated and headed for Silicon Valley to commercialize their software. Initially called Mosaic Communications Corporation, their company was soon renamed Netscape Communications Corporation—the birthplace of the famous Netscape browser family, also known as Netscape Navigator and Netscape Communicator. The Web's popularity increased, and the number of Web sites grew from approximately 500 in 1994 to nearly 10,000 by the beginning of 1995. Netscape quickly became the dominant browser and by 1996,