Applied Computing

Springer-Verlag London Ltd.

The Springer-Verlag Series on Applied Computing is an advanced series of innovative textbooks that span the full range of topics in applied computing technology.

Books in the series provide a grounding in theoretical concepts in computer science alongside real-world examples of how those concepts can be applied in the development of effective computer systems.

The series should be essential reading for advanced undergraduate and postgraduate students in computing and information systems.

Books in this series are contributed by international specialist researchers and educators in applied computing who draw together the full range of issues in their specialist area into one concise authoritative textbook.

Titles already available:

Deryn Graham and Anthony Barrett
Knowledge-Based Image Processing Systems
3-540-76027-X

Derrick Morris, Gareth Evans, Peter Green, Colin Theaker
Object Orientated Computer Systems Engineering
3-540-76020-2

John Hunt
Java and Object Orientation: An Introduction
3-540-76148-9

David Gray
Introduction to the Formal Design of Real-Time Systems
3-540-76140-3

Mike Holcombe and Florentin Ipate
Correct Systems: Building A Business Process Solution
3-540-76246-9

Jan Noyes and Chris Baber
User-Centred Design of Systems
3-540-76007-5

Arturo Trujillo
Translation Engines: Techniques for Machine Translation
1-85233-057-0

Ulrich Nehmzow
Mobile Robotics: A Practical Introduction
1-85233-173-9

Fabio Paternò
Model-Based Design and Evaluation of Interactive Applications
1-85233-155-0

Tim Morris
Multimedia Systems: Delivering, Generating and Interacting with Multimedia
1-85233-248-4

Arno Scharl

Evolutionary Web Development

Springer

Arno Scharl
Department of Information Systems, University of Economics and Business Administration, Augasse 2-6, A-1090 Vienna, Austria

Series Editors

Professor Ray J. Paul, BSc MSc PhD
Dean of the Faculty of Science, Brunel University,
Uxbridge, Middlesex UB8 3PH, UK

Professor Peter J. Thomas, MIEE MBCS CEng FRSA
Centre for Personal Information Management, University of the West of England,
Frenchay Campus, Bristol BS16 1QY, UK

Dr Jasna Kuljis, PhD MS Dipl Ing
Department of Information Systems and Computing, Brunel University,
Uxbridge, Middlesex UB8 3PH, UK

ISBN 978-1-85233-310-2

British Library Cataloguing in Publication Data
Scharl, Arno
Evolutionary web development. - (Applied computing)
1.World wide web 2.Web-site design
I.Title
005.2'76
ISBN 978-1-85233-310-2 ISBN 978-1-4471-0517-6 (eBook)
DOI 10.1007/978-1-4471-0517-6

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

© Springer-Verlag London 2000
Originally published by Springer-Verlag London Berlin Heidelberg in 2000

The use of registered names, trademarks etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant laws and regulations and therefore free for general use.

The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this book and cannot accept any legal responsibility or liability for any errors or omissions that may be made.

Typesetting: PostScript files by author

34/3830-543210 Printed on acid-free paper SPIN 10765393

Preface

With the continuing evolution and convergence of previously disparate technologies around electronic commerce, the World Wide Web is increasingly pervasive in the corporate value chain. Most business processes, from procurement and inbound logistics to marketing and after-sales service, create and use information (Porter 1998, 167f.). They are connected via networked information systems, which have become the basic infrastructure for global transaction-oriented applications. Every transaction process occurring in such an electronic marketplace goes hand in hand with the access, absorption, arrangement, and selling of information in heterogeneous ways (Zakon 1999).

Despite these technological and organizational changes, the customers' information needs provide a uniform purpose for Web information systems. Accordingly, developers of such systems have to analyze the requirements of the newly empowered and technologically savvy customers in order to exploit the potential of online trading of information, services, and physical goods.

This book presents a methodology for analyzing and developing Web information systems that considers the structural changes in electronic markets outlined above. Chapter 1 defines the field and introduces the term "ergodic literature" for textual material whose access and utilization requires non-trivial efforts by the reader. It identifies navigational and textual features of interactive Web applications, a subset of ergodic literature, and compares these features with those of printed media. An analysis of the multifaceted term "interactivity" then sets the stage for Chapter 2, which investigates the life-cycle economies and diffusion characteristics of traditional and Web-based applications. The subsequent sections delineate the evolution of electronic markets in general and the World Wide Web in particular from both Darwinian and methodological perspectives. To investigate this process, an evolutionary framework based on system adaptivity and the underlying communication patterns is introduced. It classifies Web information systems and analyzes their ability to support the various phases of electronic business transactions. Only advanced system architectures can fully leverage the potential of the World Wide Web to deliver additional customer value. The framework specifies the following four categories: Static Web information systems that provide basic information of limited value for the average customer (↪ Chapter 3), interactive systems that enable explicit customer feedback and transaction processing (↪ Chapter 4), adaptive systems that instantly and automatically generate the hyper-

text structure according to embedded user models (↪ Chapter 5), and agent-mediated architectures that allow individual negotiations regarding product and non-product attributes (↪ Chapter 6).

Chapter 3 summarizes the characteristics of static Web information systems. These collections of rudimentary hypertext documents are still common for many small and medium-sized companies. More in line with current business practices of larger organizations, Chapter 4 focuses on gathering customer feedback and its immediate analysis within an iterative cycle of design, implementation, usage, and analysis. The emergent attributes of modern organizations suggest such an evolutionary approach to Web development. Analytic activities are no longer captured within the early stages of a system's life-cycle but represent a continuing task of system maintenance. Analysis, operation, and maintenance activities become parallel but highly interrelated processes. Cyclical planning is rendered obsolete, since the results of the ongoing analysis are continuously fed into the maintenance phase. Section 4.2 outlines the current state of Web engineering with special regard to design methodologies, conceptual modeling approaches, and related visualization techniques. It introduces a symbolic modeling language for the construction of both reference and customized models during the development process of commercial Web information systems (↪ Section 4.3). This graphical notation helps analysts to visualize individual and aggregated access patterns of online customers derived from log file data of corporate Web servers. It enhances the limited, statistically oriented representations of commercially available Web-tracking software with a map-like overview similar to customer tracking in traditional retailing outlets (↪ Section 4.4).

Numerous tools support the structured design of Web information systems, ranging from rudimentary layout products for individual documents to sophisticated Web site management solutions that facilitate conceptual authoring-in-the-large. Structured analysis, by contrast, has largely been neglected by both theory and practice but belongs to the most relevant questions in formulating business strategies for electronic commerce. This comes as a surprise, since analysis is less constrained by the technical limitations of existing architectures. Section 4.5 intends to fill this gap by suggesting a methodology for the automated analysis of Web information systems. The chapter summarizes empirical results covering several business sectors, namely information technology, travel and tourism, retail banking, and environmentally oriented non-profit organizations. Classifications, comparative sector assessments, longitudinal studies, and exploratory textual analyses help practitioners chart the industry evolution and compare the performance of their own Web information system with those of competing organizations.

Chapter 5 introduces adaptive Web information systems that promise a sustainable competitive advantage in an environment where information redundancy becomes increasingly evident. Such a competitive advantage can only be achieved through customizing products, services, and communications. Key elements of customization are the separation of product and process life-cycles as described in Section 5.1. The resulting "classic loop of adaptation", delineated in Section 5.2, is founded on the continuing gathering of information regarding the customers' current needs and preferences. Sections 5.3 and 5.4 compare explicit and implicit acquisition methods and describe ways to combine the various information sources and transform the resulting body of knowledge into consistent user models. On the basis of such embedded user models, adaptive Web information systems can automatically generate the content of documents, the primary navigational system comprising links between and within these documents, and supplemental navigational systems such as index pages, trails, guided tours, or interactive site maps. Many prototypes that incorporate adaptive components are developed without a clear model of the components' functionality. Therefore, Section 5.5 provides a conceptual guideline for the development process by classifying the various mechanisms into content-level, link-level, and meta-level adaptation.

In the fourth and last stage of Web evolution, which is portrayed in Chapter 6, efforts to automate and optimize electronic business-to-consumer transactions gradually transfer certain tasks from adaptive Web information systems to agent-mediated architectures. Many deployed applications use content-based information agents and collaborative filtering systems for general information retrieval or specific product recommendations (↪ Sections 6.1 and 6.2). By contrast, most transaction agents are only available as prototypical implementations up until now (↪ Section 6.3). Transaction agents allow multidimensional negotiations regarding a variety of product and non-product attributes. Once the necessary infrastructure is in place (↪ Section 6.4), their customizability and remarkable flexibility promise to change the inherent characteristics of doing business electronically.

Acknowledgements

I have been helped in various ways with the conceptual planning of this book, the gathering of the empirical data, the compilation of results, and with the production of the final manuscript. My first debt of gratitude is to Hans Robert Hansen and Hannes Werthner for encouraging and helping me through the process of writing this book. Their suggestions and critical observations have led to innumerable improvements. I

also owe special thanks to Bernard Glasson and Arie Segev for their generous support during my time as visiting research fellow at the Curtin University of Technology and the University of California at Berkeley, respectively.

The successful completion of a project such as this is not possible without the feedback of many colleagues. First of all, I would like to mention Christian Bauer, to whom I am greatly indebted for his invaluable help. Several analyses presented in this book originate from our joint research efforts over the last two years. I would also like to thank my home institution, the Information Systems Department of the Vienna University of Economics and Business Administration, for granting me the resources and intellectual support necessary to finish this project. The cooperation with my colleagues Martin Bichler, Roman Brandtweiner, and Marion Kaukal contributed to important aspects of my research. Thanks also go to Judith Gebauer, with whom I have been fortunate to work at the University of California at Berkeley.

I apologize to the people I have omitted to acknowledge, and also – in the event that their ideas have been misrepresented – to those whom I have acknowledged. I would also like to recognize the considerable financial support from the Austrian Science Fund and the Austrian National Bank. Finally, I would like to thank Rebecca Mowat and the staff at Springer London for their support and help in the materialization of this book.

Vienna, July 2000 *Arno Scharl*

Table of Contents

1 Introduction 1

1.1 Formal Definitions of Text 2

1.1.1 The Metaphysics of Written Text 2

1.1.2 Technological Implications for the Discursive Space 4

1.2 Ergodic Literature 5

1.3 Hypertext and Hypermedia 6

1.3.1 History 7

1.3.2 Adaptability versus Adaptivity 9

1.4 Web Information Systems 10

1.4.1 Theories of Communication 11

1.4.2 Media Characteristics of the World Wide Web 13

1.4.3 Evolving Web Genres 14

1.4.4 Structure and Navigation 15

1.4.4.1 The Myth about Linearity 16

1.4.4.2 The Labyrinth Metaphor 17

1.4.5 Textual Characteristics 18

1.4.5.1 Differences between Printed and Electronic Media 20

1.4.5.2 Mapping Textons to Scriptons 22

1.4.6 Interactivity: An Ambiguous Term 23

1.4.6.1 Media Permeability 24

1.4.6.2 Communication Patterns 25

1.4.6.3 Discourse Structure 25

2 The Evolution of Electronic Markets 27

2.1 Darwinian versus Technological Evolution 28

2.1.1 Alternative Life-Cycle Economies 32

2.1.1.1 Prototyping 32

2.1.1.2 Organizational Emergence 33

2.1.1.3 Types of Maintenance 34

2.1.1.4 Stable System Drag 36

2.1.2 Electronic Business Ecosystems 37

2.1.2.1 Autopoietic Self-Reference 38

2.1.2.2 Chronological Development and Demarcation 38

2.1.2.3 Critical Mass of Content and Adopters 40

2.2 A Brief History of Web Adaptivity 40
2.2.1 Stage Models for Describing Web Evolution 42
2.2.2 Diffusion of Web Technology 45
2.2.2.1 Temporal Pattern of the Diffusion Process 45
2.2.2.2 Rate of Adoption 47
2.3 Electronic Market Transactions 49
2.3.1 Evaluating Transactional Infrastructures 53
2.3.2 Redefining the Customer-Delivered Value 55
2.3.3 Extending the Evaluation Model 59
2.3.4 Electronic Customer Relations 61
2.3.4.1 Customizing Corporate Communication 61
2.3.4.2 Gathering and Representing Customer Information 63

3 A Static World 67

4 The Emergence of Interactivity 71

4.1 Overview 72
4.1.1 Development Process 72
4.1.2 Database Connectivity 73
4.1.3 Presentational versus Semantic Markup Languages 74
4.1.4 Traversal Functions 76
4.2 Design: The Current State of Web Engineering 77
4.2.1 Processing and Visualizing Complex Information 77
4.2.2 Cooperative Web Development 80
4.2.3 Review of Existing Design Methodologies 82
4.2.4 The Extended World Wide Web Design Technique 84
4.2.4.1 Information Object Types 87
4.2.4.2 Diagram Structuring 89
4.2.4.3 Navigation Design 90
4.2.5 Application of eW3DT (Electronic Shopping Mall) 91
4.3 Implementation: Industry-Specific Reference Models 94
4.3.1 Definition and Overview 94
4.3.2 Deducing Customized and Technical Models 94
4.3.3 Removing Communication Barriers 96
4.4 Usage: Visualizing Topology and Access Patterns 98
4.4.1 Categories of Perceptualization 98
4.4.2 Observing and Visualizing Human Behavior 99
4.4.3 Units of Analysis 100
4.4.4 Tool Support 103

4.5 Analysis: Gathering, Extracting, and Processing of Multidimensional Web Data 106
4.5.1 Overview and Analytical Objectives 107
4.5.2 Manual versus Automated Evaluation 109
4.5.2.1 Previous Web Evaluation Projects 109
4.5.2.2 Web Classification Frameworks 113
4.5.2.3 Integrative Analysis of Content and Form 115
4.5.3 Methodology 116
4.5.3.1 Sampling Process 118
4.5.3.2 Mirroring and Archiving Web Data 121
4.5.3.3 Structural Parameters 122
4.5.3.4 Textual Parameters 126
4.5.3.5 Combining and Pre-Processing the Extracted Parameters 127
4.5.4 Validating Manual Evaluations of the LSE/Novell Survey 131
4.5.4.1 Introduction to Neural Networks 131
4.5.4.2 Test Design 133
4.5.4.3 Results 135
4.5.4.4 Discussion 137
4.5.5 Snapshot Analysis (Classification) 139
4.5.5.1 Partitioning Clustering 140
4.5.5.2 Results 142
4.5.5.3 Discussion 145
4.5.5.4 Hierarchical Clustering 146
4.5.6 Cross-Sectional Analysis 147
4.5.6.1 Structure and Navigation 148
4.5.6.2 Interface Representation 150
4.5.6.3 Content and Strategy 152
4.5.7 Longitudinal Analysis 154
4.5.8 Exploratory Textual Analysis 156
4.5.8.1 Types of Content 159
4.5.8.2 Common Methods for Describing and Analyzing Textual Data 160
4.5.8.3 Word List Generation 161
4.5.8.4 Lemmatization 164
4.5.8.5 Keyword Analysis 165
4.5.8.6 Concordance Analysis and Collocation 170
4.5.8.7 Analysis of Contingency Tables 173

5 Adaptive Solutions 183

5.1 Separating Product and Process Life-Cycles 185

5.2 Classic Loop of Adaptation 188
5.3 Collecting Information About Users 190
5.3.1 Available Online Sources 192
5.3.2 Explicit Acquisition Methods 193
5.3.3 Implicit Acquisition Methods 194
5.4 User Modeling 197
5.4.1 Characteristics of User Models 199
5.4.2 Typology of the Extra-Linguistic Context 202
5.4.3 Building and Initializing User Models 204
5.5 Adaptive Web Presentation Techniques 206
5.5.1 Annotations 207
5.5.1.1 Historical Background 208
5.5.1.2 Automated Document Annotation 209
5.5.2 Content-Level Adaptation 210
5.5.2.1 Deictic Expressions 211
5.5.2.2 Expressing Shared Meaning 211
5.5.2.3 Techniques for Generating and Arranging Scriptons 212
5.5.3 Link-Level Adaptation 214
5.5.4 Meta-Level Adaptation 217
5.5.4.1 Site Maps 219
5.5.4.2 View Types 221
5.5.4.3 View Transformations 226
5.5.4.4 Implications for Hypertext Navigation 227
6 Agent-Mediated Architectures 233
6.1 Content-Based Information Agents 234
6.2 Collaborative Filtering 236
6.3 Transaction Agents 238
6.4 Developing Digital Agents 242
6.4.1 Agent Communication Models 242
6.4.2 Traversal Functions of Agent-Mediated Architectures 244
6.4.3 Modeling Multilateral Interactions 245
7 Conclusion and Outlook 247
8 References 251
Index 293