

## Contents

<b>1</b>	<b>Web Archiving: Issues and Methods</b> .....	<b>1</b>
	<i>Julien Masanès</i>	
1.1	Introduction .....	1
1.2	Heritage, Society, and the Web .....	2
1.3	Web Characterization in Relation to Preservation .....	11
1.4	New Methods for a New Medium .....	18
1.5	Current Initiatives Overview .....	40
1.6	Conclusion .....	46
	References .....	46
<b>2</b>	<b>Web Use and Web Studies</b> .....	<b>55</b>
	<i>Steve Jones and Camille Johnson</i>	
2.1	Summary .....	55
2.2	Content Analysis .....	56
2.3	Surveys .....	58
2.4	Rhetorical Analysis .....	59
2.5	Discourse Analysis .....	60
2.6	Visual Analysis .....	61
2.7	Ethnography .....	63
2.8	Network Analysis .....	64
2.9	Ethical Considerations .....	65
2.10	Conclusion .....	66
	References .....	67
<b>3.</b>	<b>Selection for Web Archives</b> .....	<b>71</b>
	<i>Julien Masanès</i>	
3.1	Introduction .....	71
3.2	Defining a Selection Policy .....	72
3.3	Issues and Concepts .....	76
3.4	Selection Process .....	82
3.5	Documentation .....	89
3.6	Conclusion .....	89
	References .....	90

---

<b>4. Copying Websites .....</b>	<b>93</b>
<i>Xavier Roche</i>	
4.1 Introduction – The Art of Copying Websites .....	93
4.2 The Parser.....	95
4.3 Fetching Document .....	102
4.4 Create an Autonomous, Navigable Copy .....	107
4.5 Handling Updates.....	109
4.6 Conclusion.....	112
Reference .....	112
<b>5 Archiving the Hidden Web.....</b>	<b>115</b>
<i>Julien Masanès</i>	
5.1 Introduction .....	115
5.2 Finding At Least One Path to Documents.....	116
5.3 Characterizing the Hidden Web .....	119
5.4 Client Side Hidden Web Archiving.....	121
5.5 Crawler-Server Collaboration .....	123
5.6 Archiving Documentary Gateways .....	125
5.7 Conclusion.....	127
References.....	128
<b>6 Access and Finding Aids .....</b>	<b>131</b>
<i>Thorsteinn Hallgrímsson</i>	
6.1 Introduction .....	131
6.2 Registration .....	133
6.3 Indexing and Search Engines .....	135
6.4 Access Tools and User Interface .....	137
6.5 Case Studies .....	146
6.6 Acknowledgements .....	151
References.....	151
<b>7 Mining Web Collections.....</b>	<b>153</b>
<i>Andreas Aschenbrenner and Andreas Rauber</i>	
7.1 Introduction .....	153
7.2 Material for Web Archives.....	155
7.3 Other Types of Information.....	160
7.4 Use Cases .....	161
7.5 Conclusion .....	172
References.....	174

---

<b>8</b>	<b>The Long-Term Preservation of Web Content.....</b>	<b>177</b>
	<i>Michael Day</i>	
8.1	Introduction .....	177
8.2	The Challenge of Long-Term Digital Preservation.....	178
8.3	Developing Trusted Digital Repositories .....	181
8.4	Digital Preservation Strategies .....	184
8.5	Preservation Metadata .....	189
8.6	Digital Preservation and the Web.....	193
8.7	Conclusion .....	194
8.8	Acknowledgements .....	194
	References.....	194
<b>9</b>	<b>Year-by-Year: From an Archive of the Internet to an Archive on the Internet.....</b>	<b>201</b>
	<i>Michele Kimpton and Jeff Ubois</i>	
9.1	Introduction .....	201
9.2	Background: Early Internet Publishing .....	202
9.3	1996: Launch of the Internet Archive .....	202
9.4	1997: Link Structure and Tape Robots.....	203
9.5	1998: Getting Archive Data Onto (Almost) Every Desktop ..	204
9.6	1999: From Tape to Disk, A New Crawler, and Moving Images .....	205
9.7	2000: Building Thematic Web Collections .....	206
9.8	2001: Public Access with the Wayback Machine: The 9/11 Archive .....	207
9.9	2002: The Library of Alexandria, The Bookmobile, and Copyrights .....	208
9.10	2003: Extending Our Reach via National Libraries and Educational Institutions .....	210
9.11	2004: And the European Archive and the Petabox .....	211
9.12	The Future .....	211
	References.....	212
<b>10</b>	<b>Small Scale Academic Web Archiving: DACHS.....</b>	<b>213</b>
	<i>Hanno E. Lecher</i>	
10.1	Why Small Scale Academic Archiving? .....	213
10.2	Digital Archive for Chinese Studies.....	214
10.3	Lessons Learned: Summing Up .....	223
10.4	Useful Resources.....	224
	<b>List of Acronyms.....</b>	<b>227</b>
	<b>Index.....</b>	<b>229</b>



<http://www.springer.com/978-3-540-23338-1>

Web Archiving

Masanès, J. (Ed.)

2006, VII, 234 p., Hardcover

ISBN: 978-3-540-23338-1