

Internet and New Media

Course Notes

Version of 03 Nov 2014



Ao.Univ.-Prof. Dr. Keith Andrews

IICM
Graz University of Technology
Inffeldgasse 16c
A-8010 Graz
kandrews@iicm.edu

Copyright 2014 Keith Andrews.

Contents

Contents	iii
List of Figures	vii
List of Tables	ix
List of Listings	xi
Preface	xiii
Credits	xv
1 Introducing the Internet	1
1.1 Internet Statistics	3
1.2 Internet FAQs (Frequently Asked Questions)	14
1.3 Warriors of the Net	20
2 Newsgroups	25
2.1 Usenet News	26
2.2 Character Encodings and Plain Text	31
2.3 Configuring Your News Reader	40
2.4 Netiquette	47
3 Electronic Mail	57
3.1 Starting with Email	57
3.2 Different Kinds of Email Access	59
3.3 Configuring Your Email Reader	61
3.4 Using Email	62
3.5 Securing Your Email	69
4 Searching and Researching	73
4.1 Searching The Web	74
4.2 Academic Research	77
4.3 Browser Search Hacks	79
4.4 Desktop Search	80

5	Plagiarism and Copyright	81
5.1	Academic Integrity	83
5.2	Copyright Law	84
5.3	Quoting Text	86
5.4	A Worked Example	88
5.5	Quoting Images	90
5.6	Plagiarism Policy and Consequences	90
5.7	Breaches of Copyright	90
6	Getting Connected	93
6.1	Hooking Up	94
6.2	Types of Internet Connection	96
6.3	How the Internet Works	104
7	Staying Safe	113
7.1	The Bad Guys	114
7.2	Privacy	115
7.3	Prevention is Better than Cure	116
8	File Transfer	129
9	The Web	133
9.1	The World Wide Web	133
9.2	Configuring Your Web Browser	141
9.3	HTTPS, SSL, and TLS	148
10	Creating Your Own Web Site	151
10.1	Creating Web Pages	153
10.2	XHTML	154
10.3	Building a Polyglot XHTML5 Web Site	157
10.4	Cascading Style Sheets (CSS3)	164
10.5	Web Hosting	165
10.6	Using HTTPS on Your Web Site	166
11	Designing with Style	167
11.1	Structuring Your XHTML	168
11.2	Styling with CSS	170
11.3	Responsive Web Design	181
12	A Short History of the Internet	189
12.1	Remote Login (Telnet)	192
12.2	File Transfer (FTP)	192
12.3	Electronic Mail (Email)	192
12.4	Newsgroups (USENET News)	193
12.5	Internet Relay Chat (IRC)	193
12.6	Wide Area Information Servers (WAIS)	194
12.7	Gopher	194
12.8	The World Wide Web (WWW)	194
12.9	Hyper-G	198

13 Other Topics	203
13.1 Media Types	203
13.2 Dynamic Web Content	207
13.3 Services Running over The Internet	208
13.4 The Semantic Web	209
13.5 Web 2.0	209
13.6 Web Usability and Accessibility	212
13.7 Web Initiatives	212
Bibliography	215

List of Figures

1.1	A Global Network of Networks	3
1.2	Opte Map of the Internet	4
1.3	Internet Users by Year	6
1.4	Users by Region	6
1.5	Internet Host Count	9
1.6	OECD Fixed Broadband Subscribers Chart	11
1.7	OECD Wireless Broadband Subscribers Chart	11
1.8	Tweets by Source	13
1.9	Browser vs Native App	14
1.10	The Client-Server Model	17
1.11	Clients and Servers	18
1.12	A Firewall Machine	18
1.13	A Private Intranet Behind a Firewall	19
1.14	Warriors of the Net Story	21
1.15	Warriors of the Net: Clicking a Link in a Web Browser	21
1.16	Warriors of the Net: Filling an IP Packet	22
1.17	Warriors of the Net: Router	22
1.18	Warriors of the Net: Switch	23
1.19	Warriors of the Net: Transoceanic Cable	23
2.1	TU Graz Newsgroups in Thunderbird	27
2.2	Directory Listing of Plain Text Files and Word Documents	35
2.3	Plain Text File in PSPad Text Editor	36
2.4	Word Document in Word 2010	36
2.5	UTF-8 Encoded Plain Text File in HxD	37
2.6	7-Bit ASCII Encoded Plain Text File in HxD	37
2.7	ISO-8859-15 Encoded Plain Text File in HxD	38
2.8	Word Document (.doc) in HxD	38
2.9	Word Document (.docx) in HxD	39
2.10	Word Document (.docx) is Zip Folder of XML Files	39
2.11	Subscribing to TU Graz Newsgroups	42
2.12	Message Source	46
3.1	Email Message Passing	58
3.2	Spam Canned Meat	67
4.1	Searching in IEEE Explore	78

5.1	Two Famous Cases of Plagiarism: Guttenberg and Schavan	82
5.2	Originality Report from TurnItIn	84
6.1	The Last Mile	94
6.2	WiFi at Home	103
6.3	Internet and OSI Reference Models	104
6.4	IP Suite Stack	105
6.5	Packet Switching	106
6.6	The IPv4 Address Space	106
6.7	The Domain Name System	108
6.8	The Site www.mercedes.at Nov. 1998	110
6.9	The Site www.mercedes.at Nov. 2001	111
7.1	Phishing Email Displayed as Original HTML	123
7.2	Phishing Email Displayed as Plain Text	123
7.3	Skype.At Subscription Trap	126
8.1	FTP with FileZilla	130
8.2	An NcFTP Session	131
8.3	NcFTP Commands	131
8.4	An SFTP Session	132
9.1	HTTP 1.0 Connection Model	135
9.2	HTTP 1.1 Connection Model	136
9.3	Timeline of HTTP Requests	137
9.4	HTTP 1.1 Request and Response Headers	138
9.5	The HTML 2.0 source code of the W3 project page in 1994.	139
9.6	Proportional Versus Fixed Width Fonts	142
9.7	Serif and Sans Serif Fonts	142
9.8	Font Size Changes	143
9.9	A Typical Cookie from Amazon.Com	147
9.10	Chrome Security Warning	149
11.1	Marking Up a Shopping List	168
11.2	The Structure of a CSS Rule	171
11.3	The CSS Box Models	173
11.4	Floated Elements within Container (Desired)	174
11.5	Floated Elements within Container (Problem)	175
11.6	Inline-Block Layouts	176
11.7	Well-Structured Polyglot XHTML5 Markup	177
11.8	Southwest Airlines Desktop Web Site	182
11.9	Southwest Airlines Mobile Web Site	183
11.10	The Guardian Responsive Web Site (Wide)	184
11.11	The Guardian Responsive Web Site (Narrow)	184
11.12	The Graz Times at 320 × 480	185
11.13	The Graz Times at 640 × 480	185
11.14	The Graz Times at 1024 × 768	186
11.15	A Breakpoint Diagram	187

12.1	A Telnet Session	193
12.2	IRC Log from the Gulf War	195
12.3	WAIS Question	196
12.4	WAIS Document	196
12.5	Gopher Client-Server Interaction	196
12.6	A Gopher Client for Unix	197
12.7	The CERN NeXT Browser/Editor	199
12.8	The CERN Line Mode Browser	199
12.9	The Lynx Text Browser	200
12.10	Xmosaic Graphical Browser	200
12.11	The HGTV Terminal Viewer	201
12.12	Harmony Browser for Hyper-G	201
13.1	Sample SVG Graphic	205
13.2	Classic Tiger SVG Graphic	206

List of Tables

1.1	Internet Users by Year	5
1.2	Internet Users by Region	5
1.3	Internet Users in Europe	7
1.4	Internet Host Count	8
1.5	Internet Traffic Per Capita	9
1.6	OECD Fixed Broadband Subscribers	10
1.7	OECD Wireless Broadband Subscribers	12
1.8	Measured Download Speeds	15
2.1	Top-Level Newsgroup Hierarchies	30
2.2	US-ASCII Code Table	31
2.3	ISO-8859-15 Code Table	32
2.4	Common Emoticons (Smileys)	53
2.5	Common Acronyms	54
4.1	Most Popular Search Engines	74
6.1	Streaming Media Bandwidth Requirements	96
6.2	Typical Download Amounts	97
6.3	Example IP Addresses and Domain Names	109
7.1	High-Risk File Types	117
7.2	Lower-Risk File Types	118
10.1	Percent Encoding UTF-8 Characters	160
12.1	The Early History of the Web	198

List of Listings

7.1	Phishing Email with False URL	122
7.2	Phishing Email Purporting to be from TU Graz	124
13.1	Source Code of Sample SVG Graphic	205

Preface

These lecture notes are for an introductory course about the Internet, taught for the first time in October 2001 at Graz University of Technology.

The notes have evolved over several years and have benefitted from my experiences teaching courses at FH Technikum Kärnten in Villach, FH Joanneum in Graz, FH Hagenberg near Linz and numerous intensive courses at conferences and for industry.

I would like to thank my tutors for their many helpful ideas and comments over the years. I would also like to thank all my students past and present for their many suggestions and corrections which have helped to massage these notes into their current form.

Thanks and happy reading,

Keith

References in Association with Amazon

References with an ISBN number are linked to amazon.com (or amazon.co.uk or amazon.de) for quick, discounted purchasing. Amazon pay a small referral fee to the referrer (me) for each item you purchase after following such a link – the item itself does not cost you any more.

Credits

Much of the material in these lecture notes was inspired by Angus Kennedy's wonderful "Rough Guide to the Internet", now maintained by Peter Buckley and Duncan Clark [[Buckley and D. Clark, 2009](#)].

Chapter 1

Introducing the Internet

“ *I think there is a world market for maybe five computers.* ”

[Thomas J. Watson, founder and chairman of IBM, 1943.]

References

- + Peter Buckley and Duncan Clark; *The Rough Guide to the Internet*; 14th Edition. Rough Guides, Aug. 2009. ISBN 1848361068 (com, uk) [Buckley and D. Clark, 2009]
- + John R. Levine and Margaret Levine Young; *The Internet For Dummies*; 13th Edition. Wiley, 09 Dec 2011. ISBN 1118096142 (com, uk) [J. R. Levine and Young, 2011]
- Douglas E. Comer; *Computer Networks and Internets*; 6th Edition. Addison-Wesley, 02 Jan 2014. ISBN 0133587932 (com, uk) [Comer, 2014]
- James F. Kurose and Keith W. Ross *Computer Networking: A Top-down Approach*; 6th Edition. Pearson, 03 May 2012. ISBN 0273768964 (com, uk) [Kurose and Ross, 2012]
- Joseph B. Miller; *Internet Technologies and Information Services*; 2nd Edition. Libraries Unlimited, 26 Aug 2014. ISBN 1610694732 (com, uk) [Miller, 2014]
- Gary P. Schneider and Jessica Evans; *New Perspectives on the Internet: Comprehensive*; 9th Edition. Cengage Learning, 01 May 2012. ISBN 1111529116 (com, uk) [on the Internet: Comprehensive, 2012]
- Preston Gralla; *How the Internet Works*; 8th Edition. Que, Nov. 2006. ISBN 0789736268 (com, uk) [Gralla, 2006]
- ++ Simon May; *The Rough Guide to Windows 7*; Rough Guides, 01 Sept 2009. ISBN 1848362773 (com, uk) [May, 2009]
- + Peter Buckley and Duncan Clark; *The Rough Guide to Macs and OS X Snow Leopard*; 3rd Edition. Rough Guides, 09 Oct 2009. ISBN 1848362765 (com, uk) [Buckley, 2009]

References in German

- + John R. Levine et al; *Internet für Dummies*; 13. Auflage. Wiley, 03 Jul 2013. [In German] ISBN 352770986X (com, uk) [J. R. Levine, Young and Nagy, 2013]

- + Douglas Comer und Ralph Droms; *Computernetzwerke und Internets*; 3. Auflage. Pearson Studium, 2003. [In German] ISBN 382737149X (com, uk)
- James F. Kurose und Keith W. Ross; *Computernetzwerke*; 5. Auflage. Pearson Studium, 2012. [In German] ISBN 3868941851 (com, uk)

Online Resources

- ++ Wikipedia; *Internet*; <http://en.wikipedia.org/wiki/Internet>
- + Google Chrome Team; *20 Things I Learned about Browsers and the Web*; 20thingsilearned.com
- + William Stewart; *The Living Internet*; livinginternet.com
- + How Stuff Works; computer.howstuffworks.com
- BBC WebWise; <http://bbc.co.uk/webwise/>
- LearnTheNet; learntenet.com
- internet101; internet101.org
- About.com; *Internet for Beginners*; <http://netforbeginners.about.com/od/internet101/tp/Internet-Beginners-Quick-Reference-Guide.htm>
- Cisco; *Internetworking Technology Handbook*; http://docwiki.cisco.com/wiki/Internetworking_Technology_Handbook
- Raj Jain; *Computer Networking and Internet Protocols: A Comprehensive Introduction*. Dec. 1998. <http://www.cs.wustl.edu/~jain/bnr/>
- Warriors of the Net; 13-minute video animation. warriorsofthe.net

Online Resources in German

- ++ Bernd Zimmermann; *Internet + WWW Kurs*; www-kurs.de
- ++ *SelfHTML*; selfhtml.org
- Verein für Internet-Benutzer Österreichs; vibe.at

A Global Network of Networks

- The internet is a loose, international collection of networks.
- Millions of computers are connected to each another via cables and radio waves.
- Public backbones connect all the local networks together.
- Some companies and organisations operate their own long-distance private networks.
- Most traffic is local to an individual network.

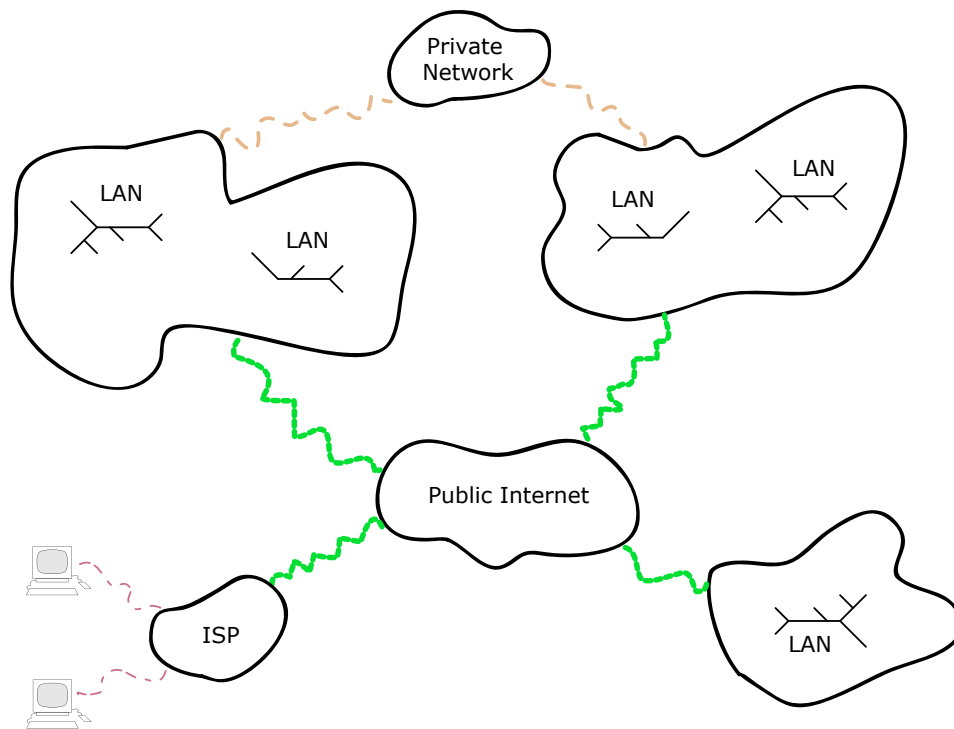


Figure 1.1: The Internet is a global network of networks.

Topological Map of the Internet

- Barrett Lyon's Opte Project [Lyon, 2009] gathers traceroute paths to class C networks.
- Around 13 million traceroute commands are issued (can be done in one day).
- The connectivity information is stored in a MySQL database.
- The very large graph of nodes (IP addresses) and links (connections) is then drawn using LGL or Graphviz.
- Colours encode the top-level domain name, for example say net, ca, and us are blue, com and org are green, mil, gov, and edu are red, white is for unknown, etc.
- Figure 1.2 shows an internet topology map using traceroute data collected on 15 Jan 2005.

1.1 Internet Statistics

Big and Getting Bigger

- The number of people online worldwide is thought to be around 2.8 billion (as of 31 Dec 2013), or around 39% of the world's population (see Table 1.1) [Internet World Stats, 2013a].
- The number of internet users worldwide is increasing by around 12% year-on-year (see Table 1.1 and Figure 1.3) [Internet World Stats, 2013b].
- Asian users now account for almost half of all internet users, easily outnumbering European and North American users combined (see Figure 1.4) [Internet World Stats, 2013a].

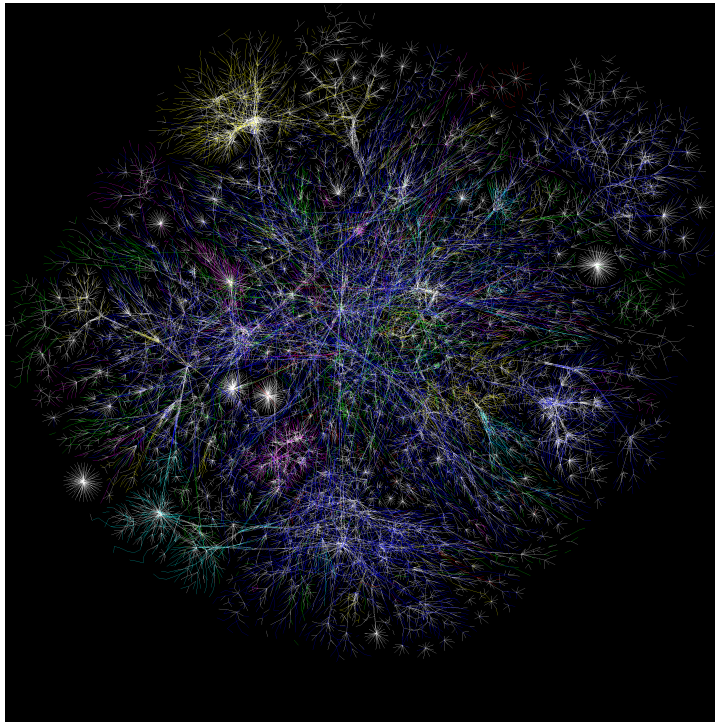


Figure 1.2: A topology map of the internet from the Opte project [Lyon, 2009]. The connectivity data was collected on 15 Jan 2005 by issuing traceroute commands to around 13 million class C networks. The colours encode the top-level domain. [Image obtained from the Opte Project [Lyon, 2009] and used under the terms of the Creative Commons Attribution-Noncommercial-Share Alike 1.0 Generic licence.]

- In Europe, Iceland (97.1%), Norway (96.9%), The Netherlands (92.9%), and Sweden (92.7%) have the highest rates of internet penetration (see Table 1.3) [Internet World Stats, 2013c].

Hosts, Traffic, and Web Pages

- The number of hosts (computers) connected to the internet stands at around 1 billion (Jul 2014), an increase of around 3.2% from Jul 2013 [Internet Software Consortium, 2014]. See Table 1.4 and Figure 1.5. [The Internet Domain Survey actually counts the number of IP addresses that have been assigned at least one domain name.]
- Internet traffic (the amount of bytes transferred) is increasing by roughly 40–50% every year (year-end-2009 data) [Odlyzko, 2009]. In Western Europe it is estimated that 5 gb of internet traffic is generated per capita per month (see Table 1.5) [Odlyzko, 2009].
- Google has reportedly now seen over a trillion unique URLs [Alpert and Hajaj, 2008].

Maurice de Kunder estimates there are around 35 billion web pages in the current Google index (based on the number of results when searching for common words) [de Kunder, 2010].

Date	Users (millions)	Users (% World)	% Increase YoY
2013-12	2,802	39.0%	12.2%
2012-12	2,497	35.7%	10.1%
2012-06	2,405	34.3%	14.0%
2011-12	2,267	32.7%	
2011-06	2,110	30.4%	7.3%
2010-06	1,966	28.7%	17.8%
2009-12	1,802	26.6%	14.5%
2009-06	1,669	24.7%	14.1%
2008-12	1,574	23.5%	19.3%
2008-06	1,463	21.9%	24.7%
2007-12	1,319	20.0%	20.7%
2007-06	1,173	17.8%	12.5%
2006-12	1,093	16.7%	7.4%
2006-06	1,043	16.0%	11.2%
2005-12	1,018	15.7%	24.6%
2005-06	938	14.6%	
2004-12	817	12.7%	13.6%
2003-12	719	11.1%	
2002-09	587	9.4%	
2001-12	553	9.3%	22.6%
2000-12	451	7.4%	81.9%
1999-12	248	4.1%	68.7%
1998-12	147	3.6%	110.0%
1997-12	70	1.7%	94.4%
1996-12	36	0.9%	125.0%
1995-12	16	0.4%	

Table 1.1: The estimated number of internet users worldwide since 1995. [Data as of 31 Dec 2013 from internetworldstats.com [Internet World Stats, 2013a] and other sources.]

Region	Population (2013-12-31)	Internet Users (2013-12-31)	% Population (Penetration)	% of World (Users)	% Growth (2011–2013)	Internet Users (2011-12-31)
North America	353,860,227	300,287,577	84.9%	10.7%	10.0%	273,067,546
Europe	825,802,657	566,261,317	68.6%	20.2%	13.1%	500,723,686
Oceania	36,724,649	24,804,226	67.5%	0.9%	3.7%	23,927,457
Latin America	612,279,181	302,006,016	49.3%	10.8%	28.1%	235,819,740
Middle East	231,062,860	103,829,614	44.9%	3.7%	34.8%	77,020,995
Asia	3,996,408,007	1,265,143,702	31.7%	45.1%	24.4%	1,016,799,076
Africa	1,125,721,038	240,146,482	21.3%	8.6%	71.7%	139,875,242
<i>World</i>	<i>7,181,858,619</i>	<i>2,802,478,934</i>	<i>39.0%</i>	<i>100.0%</i>	<i>23.6%</i>	<i>2,267,233,742</i>

Table 1.2: The number of internet users worldwide by region. The table is sorted in descending order of internet penetration. The percentage growth is shown for a 2-year period. [Data as of 31 Dec 2013 from internetworldstats.com [Internet World Stats, 2013a].]

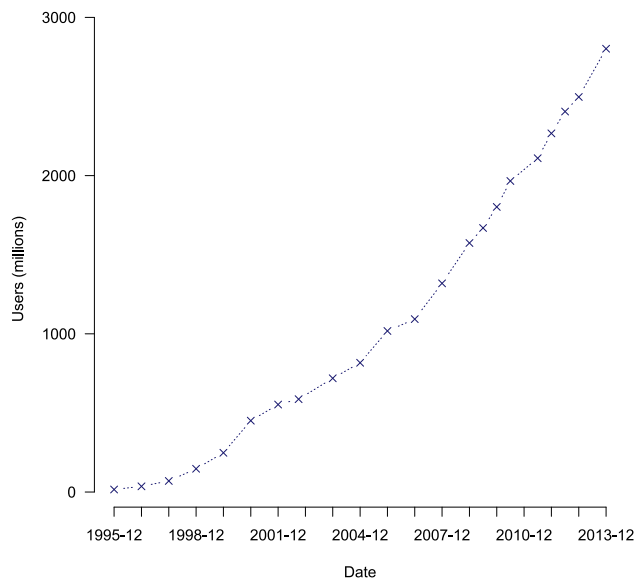


Figure 1.3: The estimated number of internet users worldwide since 1995. [Data as of 31 Dec 2013 from internetworldstats.com [Internet World Stats, 2013a] and other sources. Graphic created by Keith Andrews using R[R, 2009].]

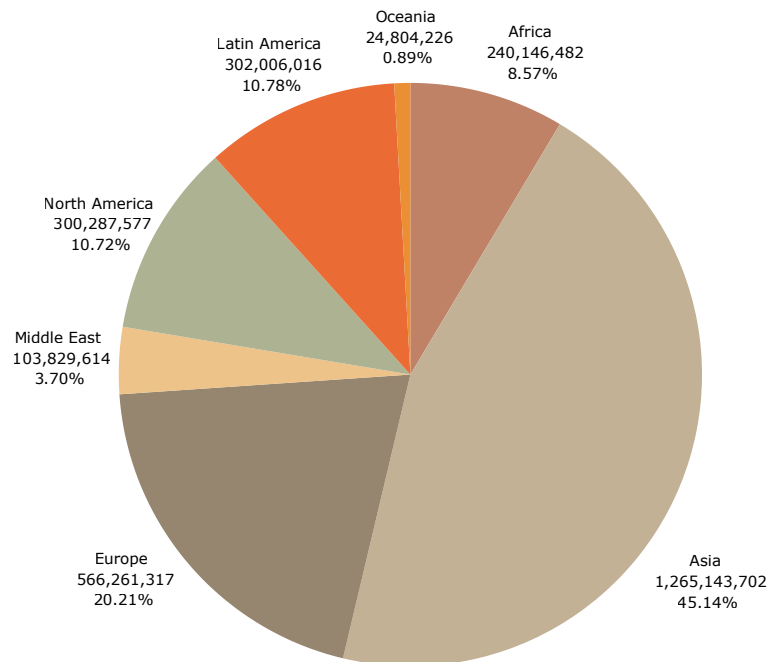


Figure 1.4: The proportion of internet users located in each of seven main geographical regions. [Data as of 31 Dec 2013 from internetworldstats.com [Internet World Stats, 2013a].]

Country	Population (2013-12-31)	Internet Users (2013-12-31)	% Population (Penetration)	% of Europe (Users)	% Growth 2011-2013	Internet Users (2011-12-31)
Monaco	30,510	30,700	100.6%	0.0%	33.5%	23,000
Iceland	313,183	304,129	97.1%	0.1%	0.0%	304,129
Norway	4,707,270	4,560,572	96.9%	0.9%	0.0%	4,560,572
Netherlands	16,730,632	15,549,787	92.9%	3.0%	3.2%	15,071,191
Sweden	9,103,788	8,441,718	92.7%	1.6%	0.0%	8,441,718
Luxembourg	509,074	462,697	90.9%	0.1%	0.6%	459,833
Denmark	5,543,453	4,989,108	90.0%	1.0%	1.3%	4,923,824
Vatican	535	480	89.7%	0.0%	0.0%	480
Finland	5,262,930	4,703,480	89.4%	0.9%	0.9%	4,661,265
Liechtenstein	36,713	31,206	85.0%	0.0%	8.3%	28,826
United Kingdom	63,047,162	52,731,209	83.6%	10.2%	0.0%	52,731,209
Germany	81,305,856	67,483,860	83.0%	13.0%	0.2%	67,364,898
Switzerland	7,925,517	6,509,247	82.1%	1.3%	1.2%	6,430,363
Belgium	10,438,353	8,489,901	81.3%	1.6%	0.0%	8,489,901
Andorra	85,082	68,916	81.0%	0.0%	0.3%	68,740
Faroe Islands	49,483	39,948	80.7%	0.0%	6.5%	37,500
Austria	8,219,743	6,559,355	79.8%	1.3%	6.8%	6,143,600
France	65,630,692	52,228,905	79.6%	10.1%	3.9%	50,290,226
Slovakia	5,483,088	4,337,868	79.1%	0.8%	0.0%	4,337,868
Estonia	1,274,709	993,785	78.0%	0.2%	0.0%	993,785
Ireland	4,722,028	3,627,462	76.8%	0.7%	16.2%	3,122,358
Guernsey	65,345	48,300	73.9%	0.0%	0.0%	48,300
Czech Republic	10,177,300	7,426,376	73.0%	1.4%	2.8%	7,220,732
Slovenia	1,996,617	1,440,066	72.1%	0.3%	1.4%	1,420,776
Latvia	2,191,580	1,570,925	71.7%	0.3%	2.0%	1,540,859
Gibraltar	29,034	20,660	71.2%	0.0%	2.3%	20,200
Croatia	4,480,043	3,167,838	70.7%	0.6%	19.3%	2,656,089
Malta	409,836	282,648	69.0%	0.1%	7.7%	262,404
Spain	47,042,984	31,606,233	67.2%	6.1%	3.1%	30,654,678
Hungary	9,958,453	6,516,627	65.4%	1.3%	0.0%	6,516,627
Lithuania	3,525,761	2,293,508	65.1%	0.4%	9.0%	2,103,471
Poland	38,415,284	24,940,902	64.9%	4.8%	4.6%	23,852,486
Bosnia	3,879,296	2,327,578	60.0%	0.4%	19.0%	1,955,277
Italy	61,261,254	35,800,000	58.4%	6.9%	0.0%	35,800,000
Cyprus	1,138,071	656,439	57.7%	0.1%	12.2%	584,863
Macedonia	2,082,370	1,180,704	56.7%	0.2%	10.4%	1,069,432
Serbia	7,276,604	4,107,000	56.4%	0.8%	0.0%	4,107,000
Portugal	10,781,459	5,950,449	55.2%	1.1%	9.1%	5,455,217
Greece	10,767,827	5,706,948	53.0%	1.1%	13.2%	5,043,550
San Marino	32,140	17,000	52.9%	0.0%	0.0%	17,000
Bulgaria	7,037,935	3,589,347	51.0%	0.7%	3.6%	3,464,287
Montenegro	657,394	328,375	50.0%	0.1%	0.0%	328,375
Albania	3,002,859	1,471,400	49.0%	0.3%	2.0%	1,441,928
Jersey	94,949	45,800	48.2%	0.0%	0.0%	45,800
Russia	142,517,670	67,982,547	47.7%	13.1%	10.6%	61,472,011
Isle of Man	85,421	39,460	46.2%	0.0%	10.8%	35,600
Belarus	9,643,566	4,436,800	46.0%	0.9%	0.0%	4,436,800
Turkey	79,749,461	36,455,000	45.7%	7.0%	4.2%	35,000,000
Moldova	3,656,843	1,639,463	44.8%	0.3%	14.7%	1,429,154
Romania	21,848,504	9,642,383	44.1%	1.9%	12.4%	8,578,484
Ukraine	44,854,065	15,300,000	34.1%	3.0%	0.0%	15,300,000
Kosovo	1,836,529	377,000	20.5%	0.1%	0.0%	377,000
<i>Europe</i>	<i>820,916,255</i>	<i>518,512,109</i>	<i>63.2%</i>	<i>100.0%</i>	<i>3.6%</i>	<i>500,723,686</i>

Table 1.3: The number of internet users in European countries. The table is sorted in descending order of internet penetration. The percentage growth is shown for a 2-year period. [Data as of 31 Dec 2013 from internetworldstats.com [Internet World Stats, 2013c].]

Date	Hosts	% Increase YoY
07/2014	1,028,544,414	3.2%
01/2014	1,010,251,829	4.9%
07/2013	996,230,757	9.6%
01/2013	963,518,598	8.5%
07/2012	908,585,739	6.9%
01/2012	888,239,420	8.5%
07/2011	849,869,781	10.5%
01/2011	818,374,269	11.7%
07/2010	768,913,036	12.9%
01/2010	732,740,444	17.2%
07/2009	681,064,561	19.3%
01/2009	625,226,456	15.4%
07/2008	570,937,778	16.6%
01/2008	541,677,360	25.0%
07/2007	489,774,269	
01/2007	433,193,199	9.7%
01/2006	394,991,609	24.3%
01/2005	317,646,084	36.3%
01/2004	233,101,481	35.8%
01/2003	171,638,297	16.5%
01/2002	147,344,723	34.5%
01/2001	109,574,429	51.3%
01/2000	72,398,092	67.5%
01/1999	43,230,000	45.7%
01/1998	29,670,000	36.0%
01/1997	21,819,000	52.0%
01/1996	14,352,000	145.5%
01/1995	5,846,000	163.7%
01/1994	2,217,000	68.8%
01/1993	1,313,000	80.6%
01/1992	727,000	93.4%
01/1991	376,000	
10/1989	159,000	
01/1989	80,000	
07/1988	33,000	
12/1987	28,174	
11/1986	5,089	
02/1986	2,308	
10/1985	1,961	91.5%
10/1984	1,024	
08/1983	562	
05/1982	235	
08/1981	213	

Table 1.4: The number of hosts (computers) connected to the internet stands at over 1 billion (Jul 2014), an increase of around 3.2% from Jul 2013. [Data as of Jul 2014 from Internet Software Consortium [2014].]

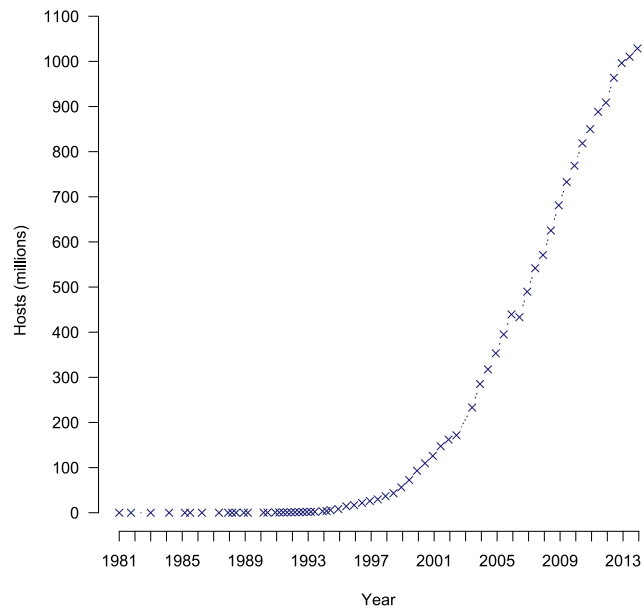


Figure 1.5: The number of hosts (computers) connected to the internet stands at over 1 billion (Jul 2014), an increase of around 3.2% from Jul 2013. [Data as of Jul 2014 from Internet Software Consortium [2014].]

Region	Monthly Traffic Per Capita (gb)
South Korea	30.0
Hong Kong	22.5
USA	7.0
Japan	5.0
Western Europe	5.0
Australia	1.5

Table 1.5: Estimated monthly internet traffic (gb) per capita in different regions of the world. [Data as of Dec 2009 from Odlyzko [2009].]

Country	DSL	Cable	Fibre/LAN	Other	Total	Total Subscriptions
Switzerland	27.9	13.2	3.4	0.3	44.9	3,597,000
Netherlands	18.6	18.7	3.2	0.0	40.4	6,794,000
Denmark	20.7	11.5	7.8	0.0	40.0	2,245,593
France	34.2	2.6	0.8	0.0	37.6	24,751,000
Korea	3.7	9.6	24.2	0.0	37.5	18,737,125
Norway	15.7	11.6	9.7	0.0	37.0	1,881,610
Iceland	27.8	0.0	7.9	0.0	35.8	115,826
United Kingdom	24.7	6.9	3.7	0.0	35.2	22,559,353
Germany	28.2	6.2	0.3	0.1	34.8	28,603,463
Belgium	16.8	17.6	0.0	0.0	34.4	3,819,393
Canada	13.5	18.8	1.1	0.0	33.5	11,675,481
Luxembourg	26.8	3.4	2.2	0.1	32.5	177,300
Sweden	14.0	6.0	12.4	0.1	32.4	3,113,400
Finland	18.9	5.8	0.9	5.2	30.8	1,676,400
New Zealand	28.3	1.5	0.5	0.0	30.2	1,341,846
United States	9.8	17.3	2.4	0.2	29.8	93,618,000
Japan	3.7	4.8	19.6	0.0	28.1	35,785,203
Spain	20.3	4.6	1.4	0.0	26.3	12,080,540
Greece	26.2	0.0	0.0	0.0	26.2	2,910,074
Austria	17.6	8.2	0.3	0.0	26.1	2,214,428
Australia	21.2	4.1	0.7	0.0	26.0	6,009,000
Estonia	10.7	5.8	8.6	0.4	25.5	341,465
Israel	16.0	9.1	0.0	0.0	25.1	2,024,000
Slovenia	12.3	7.5	5.2	0.1	25.1	517,249
Ireland	16.9	7.4	0.1	0.0	24.4	1,121,551
Portugal	10.5	9.3	4.4	0.0	24.1	2,528,604
Hungary	8.0	11.6	3.5	0.0	23.1	2,282,133
Italy	21.7	0.0	0.5	0.1	22.3	13,597,570
Czech Republic	9.2	4.9	3.3	0.0	17.4	1,826,726
Poland	7.7	5.7	0.6	1.7	15.6	6,022,651
Slovak Republic	8.1	2.6	4.9	0.0	15.6	845,997
Chile	5.5	6.6	0.3	0.5	12.9	2,271,240
Mexico	8.2	2.4	0.7	0.1	11.4	13,533,448
Turkey	8.9	0.6	1.6	0.1	11.2	8,382,811
<i>OECD</i>	<i>13.9</i>	<i>8.4</i>	<i>4.5</i>	<i>0.2</i>	<i>27.0</i>	<i>339,001,480</i>

Table 1.6: Fixed broadband subscribers per 100 inhabitants in the OECD countries by technology. Data from Dec 2013 [OECD, 2013].

OECD Broadband Subscribers

In the developed countries, internet users increasingly have access to both fixed and wireless broadband internet connections.

See Table 1.6 and Figure 1.6.

See Table 1.7 and Figure 1.7.

Mobile is Here

- Sales of new smartphones overtook sales of new PCs (desktop + laptop) in Q4 2010 [Ferguson, 2011].

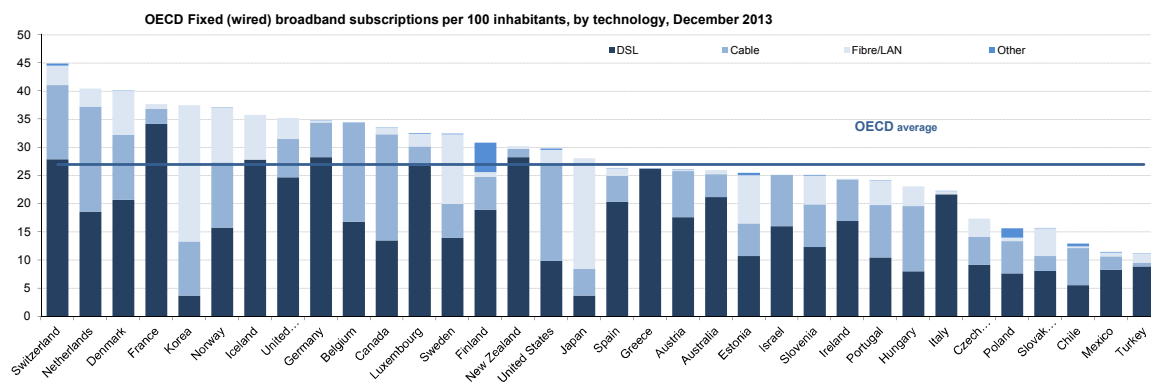


Figure 1.6: Fixed (wired) broadband subscribers per 100 inhabitants in the OECD countries. Data from Dec 2013. [Chart ©OECD 2013 [OECD, 2013].]

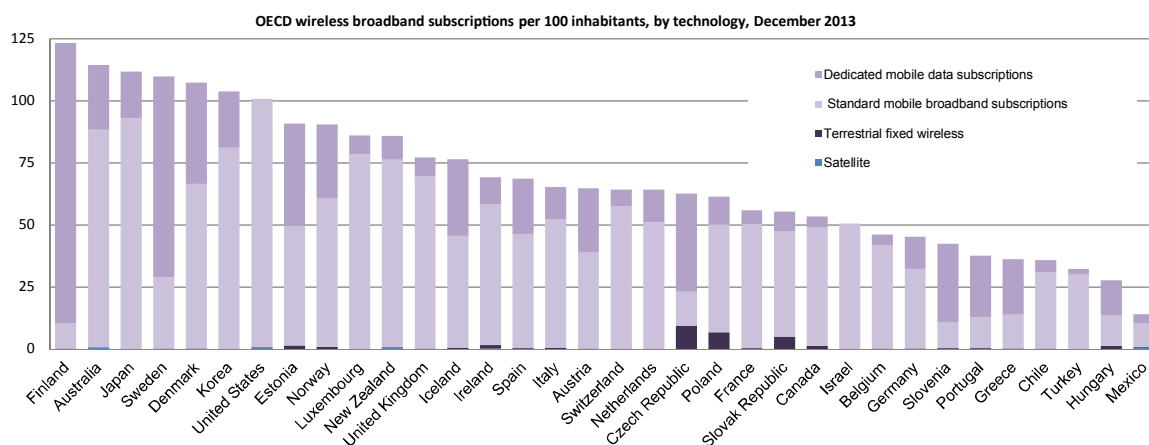


Figure 1.7: Wireless broadband subscribers per 100 inhabitants in the OECD countries. Data from Dec 2013. [Chart ©OECD 2013 [OECD, 2013].]

Country	Satellite	Terrestrial Fixed Wireless	Mobile Bundle	Mobile Internet	Total	Total Subscriptions
Finland	0.0	0.1	10.3	112.9	123.3	6,704,800
Australia	0.4	0.2	87.7	26.1	114.4	26,460,000
Japan	0.0	0.0	92.9	18.9	111.8	142,595,498
Sweden	0.0	0.0	28.8	81.0	109.8	10,545,000
Denmark	0.0	0.2	66.3	40.8	107.3	6,021,411
Korea	0.0	0.0	81.1	22.7	103.8	51,892,608
United States	0.6	0.3	99.8	0.0	100.7	316,440,000
Estonia	0.0	1.4	48.2	41.2	90.8	1,216,367
Norway	0.0	0.8	60.0	29.6	90.4	4,590,444
Luxembourg	0.0	0.0	78.4	7.7	86.1	469,300
New Zealand	0.2	0.4	75.8	9.6	85.9	3,816,733
United Kingdom	0.0	0.0	69.5	7.7	77.2	49,470,645
Iceland	0.0	0.5	45.0	31.0	76.5	247,690
Ireland	0.2	1.3	56.8	10.9	69.2	3,175,008
Spain	0.0	0.2	46.1	22.3	68.5	31,468,383
Italy	0.0	0.5	51.6	13.2	65.3	39,840,597
Austria	0.0	0.2	38.6	26.0	64.7	5,481,818
Switzerland	0.0	0.0	57.4	6.9	64.2	5,151,300
Netherlands	0.0	0.0	51.0	13.2	64.2	10,787,000
Czech Republic	0.0	9.2	13.8	39.5	62.5	6,574,264
Poland	0.0	6.7	43.4	11.1	61.3	23,594,131
France	0.0	0.3	50.0	5.5	55.9	36,733,000
Slovak Republic	0.0	4.9	42.4	8.0	55.3	2,992,742
Canada	0.0	1.2	47.8	4.3	53.3	18,581,921
Israel	0.0	0.0	50.5	0.0	50.5	4,070,000
Belgium	0.0	0.1	41.7	4.3	46.0	5,113,490
Germany	0.0	0.0	32.1	12.9	45.1	37,057,293
Slovenia	0.0	0.4	10.2	31.9	42.4	873,168
Portugal	0.0	0.3	12.4	24.8	37.5	3,932,347
Greece	0.0	0.0	13.7	22.5	36.2	4,016,512
Chile	0.0	0.1	30.8	4.9	35.8	6,282,872
Turkey	0.0	0.0	30.0	2.3	32.3	24,183,723
Hungary	0.0	1.2	12.4	14.1	27.7	2,738,282
Mexico	0.4	0.4	9.4	3.8	14.0	16,558,806
<i>OECD</i>	<i>0.2</i>	<i>0.5</i>	<i>61.4</i>	<i>10.2</i>	<i>72.4</i>	<i>909,677,153</i>

Table 1.7: Wireless broadband subscribers per 100 inhabitants in the OECD countries by technology. Data from Dec 2013 [OECD, 2013].

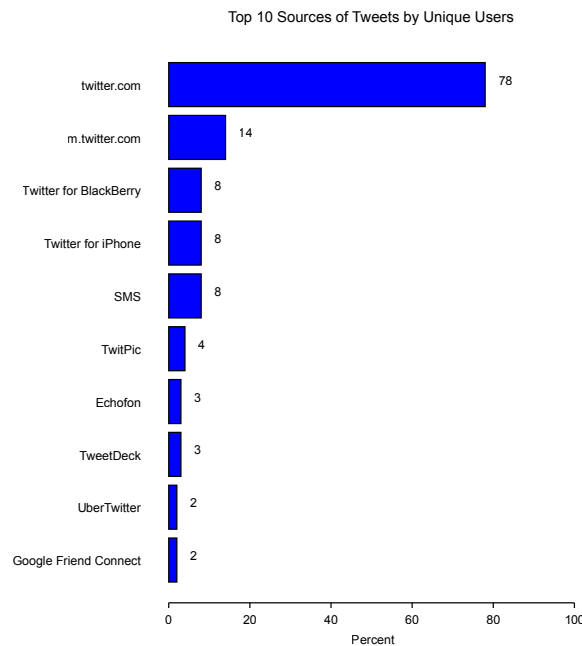


Figure 1.8: The top ten sources of tweets on twitter by unique user. TweetDeck and EchoFon both represent multiple clients on both mobile and desktop. Data over 30-day period in Aug 2010. [Data from twitter blog [Twitter, 2010].]

- Gartner predicted smartphones would overtake PCs as the most common web access device in 2013 [Gartner, 2010].
- As of Apr 2013 \approx 1.5 million Android devices activated daily [Statista, 2013].
- At CES 2011, Twitter CEO Dick Costolo said: “40% of all tweets now are created on mobile devices... only a year ago it was around 25%.” [Costolo, 2011, 00:12:14]
- Not just apps: increasing use of mobile web browser to access mobile web sites. See Figures 1.8 and 1.9.
- \approx 30% of page impressions (to ÖWA Basic web sites) in Austria are from mobile web browsers [ÖWA, 2014b]
- See talk [Wroblewski, 2011a] and book “Mobile First” [Wroblewski, 2011b] by Luke Wroblewski.

Measured Download Speeds

- Ookla’s web site speedtest.net [Ookla, 2011b] performs (and collects data from) about 20 million broadband speed tests around the world every month.
- These data are collated at netindex.com [Ookla, 2011a]. They raw data can be downloaded or explored with Google’s Public Data Explorer.
- Lithuania and South Korea have the highest measured average bandwidths.

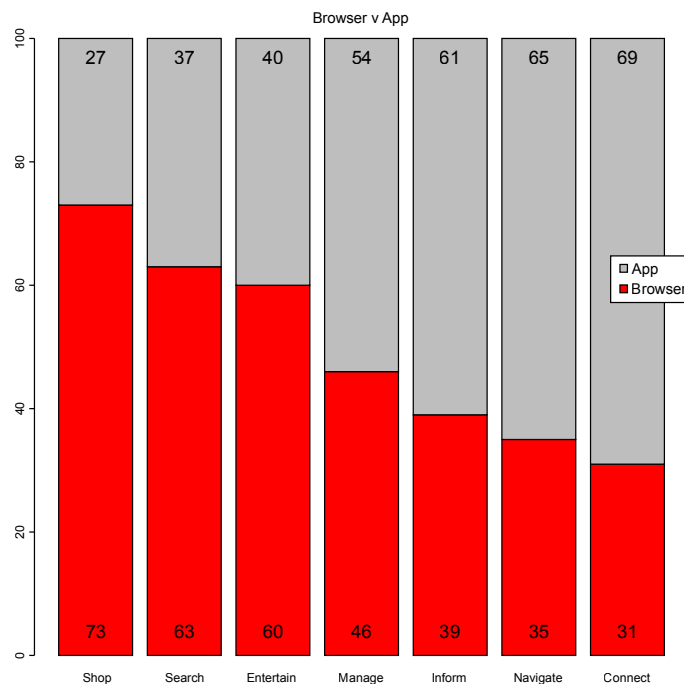


Figure 1.9: Smartphone user preference (browser or native app) for seven kinds of task. Data as of Aug 2011 from survey of 3,844 smartphone users. [Data from Yahoo Mobile Modes study [Yahoo, 2011].]

- See Table 1.8.

In Austria

- ≈ 6 million people (85%) aged 14 and older have access to the internet in Austria (Dec 2013) [ORF, 2013].
- There are ≈ 3.6 million hosts in the at domain (Jul 2014) [Internet Software Consortium, 2014].
- `willhaben.at`, `derstandard.at`, `news.at`, and `krone.at` are the most popular web sites in Austria (Aug 2014) by number of page impressions [ÖWA, 2014a].
- $\approx 30\%$ of page impressions (to ÖWA Basic web sites) in Austria are from mobile browsers [ÖWA, 2014b]

1.2 Internet FAQs (Frequently Asked Questions)

FAQ ... Frequently Asked Question

So, the Internet is Run by Google, Right?

- Well, not really... ;-)

Rank	Country	Speed (mbps)
1	Lithuania	33.32
2	South Korea	29.00
3	Sweden	26.09
4	Macau	25.77
5	Latvia	25.43
6	Romania	25.17
7	Netherlands	24.10
8	Andorra	22.25
9	Iceland	21.18
10	Bulgaria	20.93
11	Switzerland	19.96
12	Singapore	18.05
13	Moldova	17.73
14	Taiwan	17.55
15	Belgium	17.37
16	Denmark	17.15
17	Aland Islands	16.70
18	Portugal	15.97
19	Finland	15.65
20	Japan	15.61
21	Germany	15.53
22	Hungary	14.90
23	Czech Republic	14.76
24	Slovakia	14.71
25	Ukraine	14.56
26	Luxembourg	14.20
27	Norway	13.70
28	Malta	13.57
29	France	13.46
30	Estonia	13.29
31	United States	11.99
32	Monaco	11.63
33	Canada	11.59
34	Austria	10.95
35	United Kingdom	10.75
...		
160	Botswana	0.86
161	Cote D'Ivoire	0.83
162	Mali	0.81
163	Swaziland	0.76
164	Bhutan	0.74
165	Burkina Faso	0.74
166	Reunion	0.74
167	Sudan	0.70
168	Malawi	0.67
169	Bolivia	0.66
170	Iran	0.64

Table 1.8: Average download speeds as measured by NetIndex. Data as of 04 Oct 2011 from netindex.com [Ookla, 2011a].

- Google is one of the big commercial players, along with facebook, Amazon, ebay, Microsoft, Apple, Cisco, Oracle, Yahoo, AOL, etc.
- In terms of technical standards and administration:
 - The Internet Society (ISOC) develops and approves technical standards [ISOC, 2004].
 - ICANN coordinates domain names and IP addresses [ICANN, 2004].
 - Five RIRs (Regional Internet Registry) allocate blocks of IP addresses.
 - The World Wide Web Consortium (W3C) develops standards for the web [W3C, 2007a].
- Open Standards: IP, TCP, SMTP, NNTP, HTTP, ...
- Multiple platforms: Windows, Mac, Unix, PDA, WebTV, smartphone, ...
- Legal issues are governed by the laws of the territory you are in.

The Internet and the Web are the Same Thing, Right?

- Er, no, but that's OK, lots of people confuse the two.
- The web (WWW) is the popular face of the internet: billions of web sites, each comprising one or more web pages.
- However, the web is just one of many services running over the internet.
- Others services which run on the internet include email, newsgroups, file transfer, video streaming, instant messaging, file sharing (peer to peer), and voice calls (Skype, SIP).

What is a Web Page and a Web Site?

- A *web page* is a single page of text (in HTML), which may include other objects such as images and videos.
- A *web site* is a collection of web pages representing someone or something.
- A web site is usually based around a single domain name such as the hundreds of pages about Vienna at `wien.at`.

In German, the term “Webseite” is often (mis-)used to mean an entire web site.

What is a Home Page?

In English, the term *home page* has two (different) meanings:

1. A browser home page: The page which appears when a web browser is started or the browser's Home button is pressed. You can choose any web page as your browser's home page.
2. A web site home page: The “front page” of a web site. For example, the page which is displayed when you visit `wien.at`.

In German, the term “Homepage” is often (mis-)used to mean an entire web site.

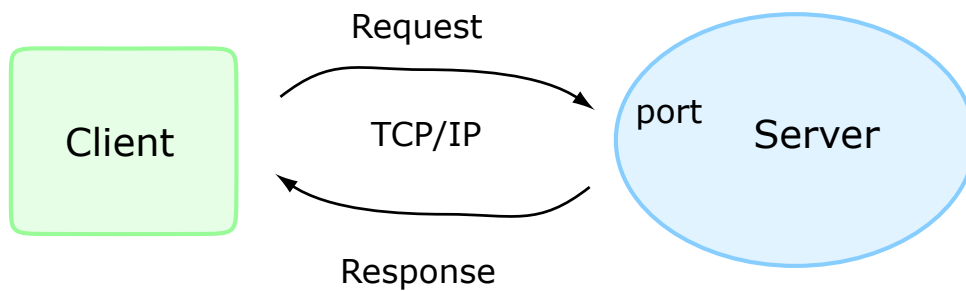


Figure 1.10: The client-server model. A server runs continuously and waits for connections from clients on a specific port.

What is an Internet Service Provider (ISP)?

- An ISP is a company which connects your computer to the rest of the internet.
- ISPs are often telephone or cable television companies, but there are also ISPs who specialise only in providing internet access.

What is a Local Area Network (LAN)?

A local area network (LAN) is the kind of network typically found *within* companies or organisations:

- Often based on Ethernet or ATM technology.
- A LAN typically also uses internet protocols (IP and TCP) and is usually connected through a *gateway* to the wider internet.

See Figures 1.1 and 1.13.

What are Hosts, Servers, Ports, and Clients?

- A computer which is open to external online access over the internet is called a *server* or *host*. A host typically has its own (domain) name.
- A *server* is (also) a program which runs continuously on a host, listening for requests on a specific *port*.
- One port is used for each kind of service. For example, WWW requests are sent to a server on port 80 by default.
- A *client* is a program started locally by a user, such as a web browser, which connects to a server to request some action (like asking for a web page), as shown in Figure 1.10.
- For example, you might use a web client such as Microsoft Internet Explorer 8.0 to access a web server running of port 80 of the host `kleine.at`, as shown in Figure 1.11.

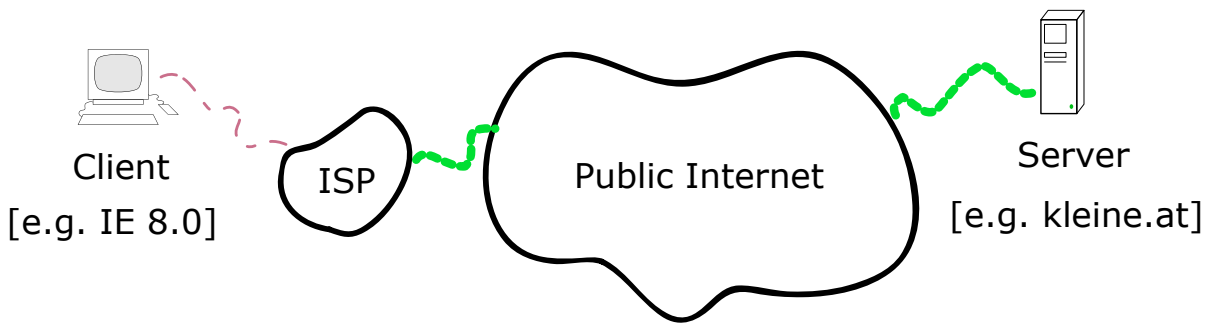


Figure 1.11: Client software on the user's computer (a web browser, such as IE 8.0) makes a request for a web page to a web server (kleine.at) on port 80.

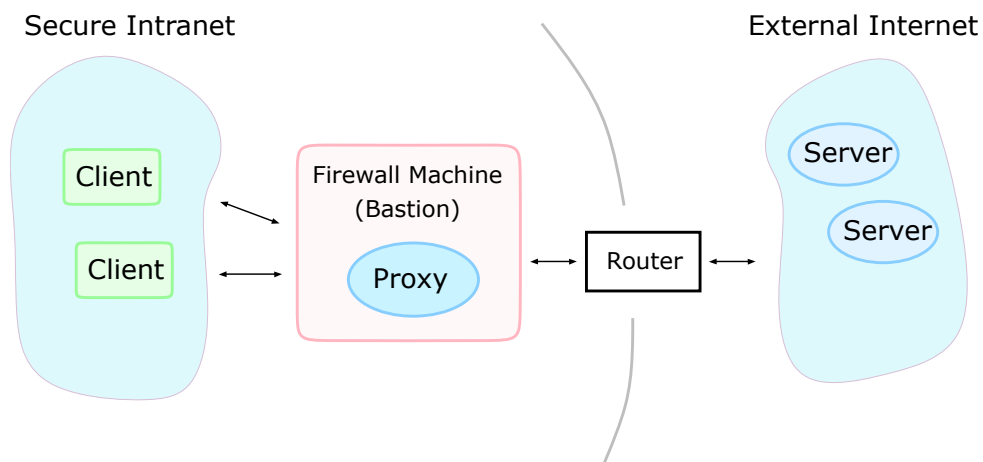


Figure 1.12: A firewall machine.

What are Firewalls?

A *firewall* is a (configurable) barrier between a LAN and the outside world.

- Often a single firewall computer serves as the gateway between a LAN and the outside world. See Figure 1.12.
- The firewall can monitor and block connections into and out from the LAN.
- Typically all connections into the LAN will be blocked and only specific services will be allowed out (for example HTTP traffic via a proxy).
- All connections and traffic are usually logged.

What are Intranets and Extranets?

- An *intranet* is a private internal network, usually behind a firewall, which runs IP and internet services. See Figure 1.13.
- An *extranet* specifically grants access to parts of the intranet to business partners or specific external users.

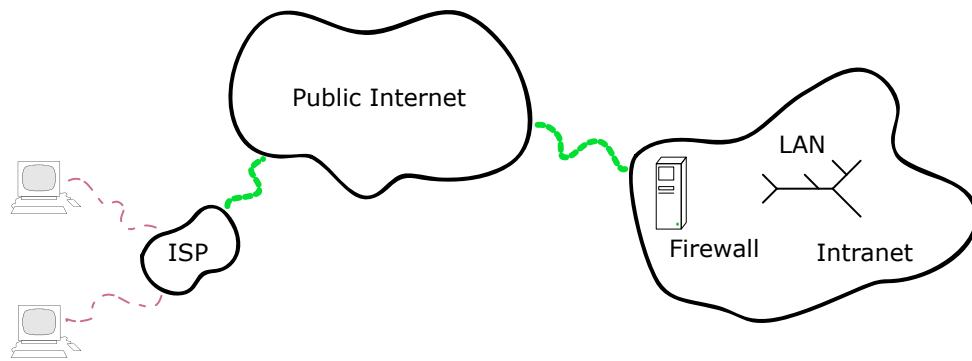


Figure 1.13: A private intranet behind a firewall.

What is a Proxy Server?

A *proxy server* forwards requests to their destination and passes responses back to the originator:

- At the destination, it appears that the proxy is the originator of the request.
- Access to external web servers from inside a firewall.
- Application-level proxy: document caching in the proxy. Frequently accessed objects are *cached* (a local copy with a certain life expectancy is made) by the proxy for all users of the proxy.

1.3 Warriors of the Net

- 13-minute video animation, illustrating how the internet works (see Figure 1.14).
- Gunilla Elam, Tomas Stephanson, and Niklas Hanberger; originally at Ericsson Medialab, 1999. Reworked version 2002 [Elam, Stephanson and Hanberger, 2002].
- Remastered in HD in 2011 [Elam, Stephanson and Hanberger, 2011].

Terminology

- *Routers and switches*: components which forward packets of information onward towards their destination. Routers (Figure 1.17) work at a higher level (network layer, OSI level 3) than switches (Figure 1.18) (data link layer, OSI level 2). Routers are able to learn routes from other routers and thus build up knowledge about network topology. See http://www.petri.co.il/csc_routers_switches_and_firewalls.htm
- *Ping of Death*: IP packet which exceeds the maximum legal length (65535 bytes) and used to cause a buffer overflow and crash on many operating systems. <http://www.insecure.org/spl0its/ping-o-death.html>

Warriors of the Net

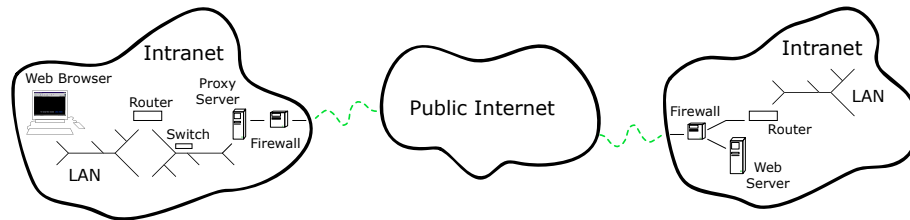


Figure 1.14: The Warriors of the Net film tells the story of a request sent from a web browser, making its way over the public internet to a web server, and a response being sent back.

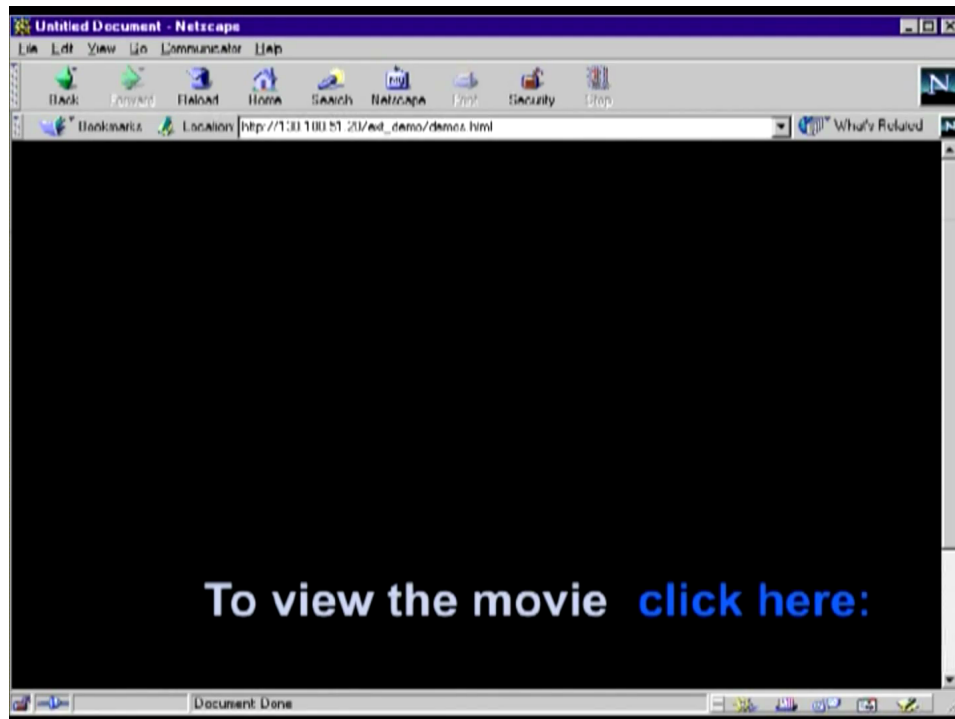


Figure 1.15: A still frame from the Warriors of the Net animation showing the user about to click on a link in the Netscape web browser. [Image copyright 2002 Gunilla Elam, Tomas Stephanson, Niklas Hanberger [Elam, Stephanson and Hanberger, 2002]. Used with permission.]



Figure 1.16: A still frame from the Warriors of the Net animation showing an IP packet being filled with bits of data. [Image copyright 2002 Gunilla Elam, Tomas Stephanson, Niklas Hanberger [Elam, Stephanson and Hanberger, 2002]. Used with permission.]



Figure 1.17: A router works at a higher level (OSI level 3, network layer) than a switch. It examines the destination IP address of each IP packet and sends it on its way towards its destination. A router can learn about routes from other routers and thus build up knowledge about network topology. [Image copyright 2002 Gunilla Elam, Tomas Stephanson, Niklas Hanberger [Elam, Stephanson and Hanberger, 2002]. Used with permission.]

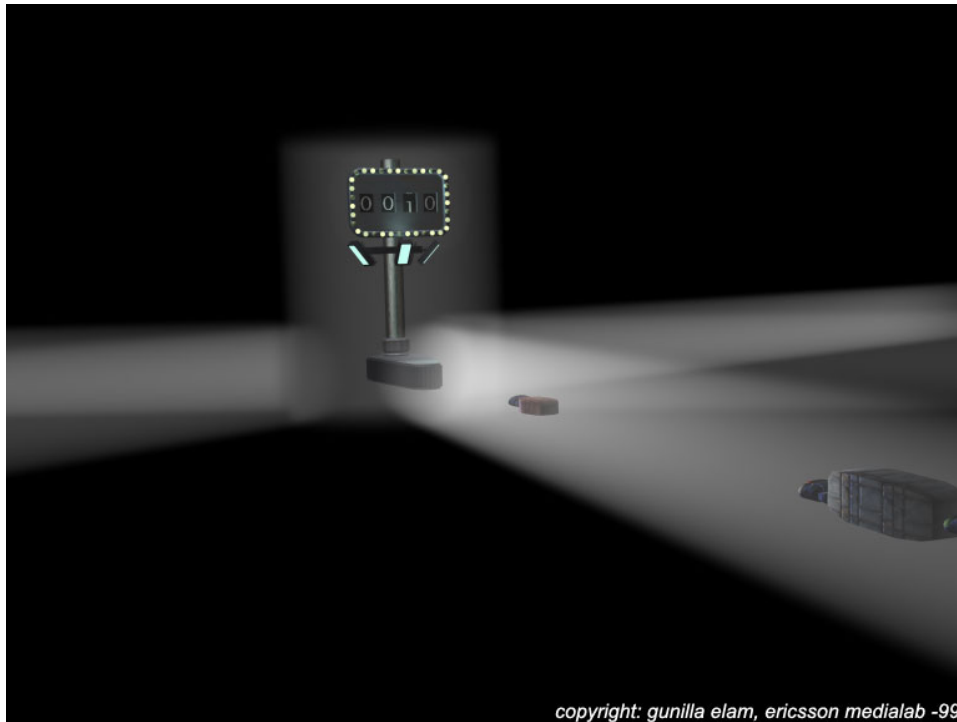


Figure 1.18: A switch works at a lower level (OSI level 2, data link layer) than a router, simply passing Ethernet frames onto the next network connection according to their destination MAC address. [Image copyright 2002 Gunilla Elam, Tomas Stephanson, Niklas Hanberger [Elam, Stephanson and Hanberger, 2002]. Used with permission.]

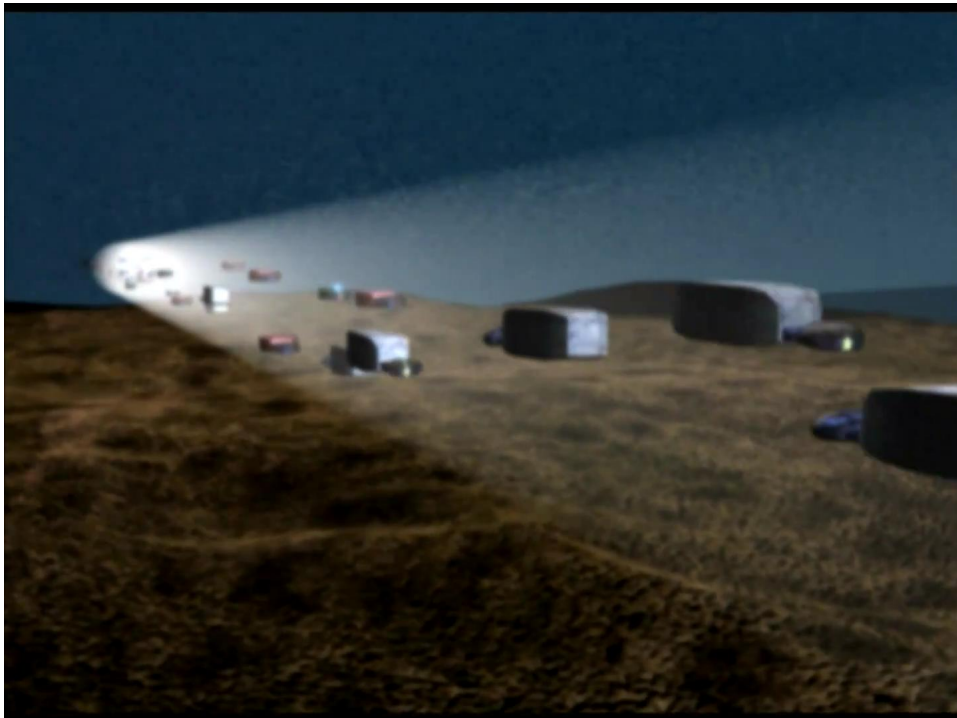


Figure 1.19: IP packets travelling through a transoceanic underwater cable. Still frame from Warriors of the Net. [Image copyright 2002 Gunilla Elam, Tomas Stephanson, Niklas Hanberger [Elam, Stephanson and Hanberger, 2002]. Used with permission.]

Chapter 2

Newsgroups

“ If you have nothing constructive to write, please have the courtesy to write nothing.
”

[J.R. Stockton, 2003.]

Resources

- ++ *Wikipedia Usenet*; <http://en.wikipedia.org/wiki/Usenet>
- ++ *Wikipedia Newsgroup*; <http://en.wikipedia.org/wiki/Newsgroup>
- ++ *NewsReaders.com*; <http://www.newsreaders.com/>
 - *news.newusers.questions*; <http://www.anta.net/misc/nnq/>
 - J.R. Stockton; *About News-Posting*; <http://www.merlyn.demon.co.uk/news-use.htm>
 - *Internet FAQ Archives*; <http://www.faqs.org/faqs/>
 - *Google Groups* (formerly DejaNews); <http://groups.google.com/>
 - *NetManners.com*; <http://netmanners.com/>
 - *Netiquette Home Page*; <http://www.albion.com/netiquette/>
 - RFC 1855 ; *Netiquette Guidelines*; <http://www.faqs.org/rfcs/rfc1855.html>
 - *EmoticonUniverse*; <http://www.emoticonuniverse.com/>
 - Ken Marsh; *The Unofficial Smilie Dictionary*; <http://www.charm.net/~kmarsh/smiley.html>
 - *Acronym Finder*; <http://www.acronymfinder.com/>
 - Sven Guckes; *Signatures*; <http://www.guckes.net/sig/>
 - Esther Filderman; *Esther's Massive Signature File Collection*; <http://www.contrib.andrew.cmu.edu/~moose/sigs.html>
 - Jon Bell and Thor Kottelin; *Quoting Style in Newsgroup Postings*; <http://www.anta.net/misc/nnq/nquote.shtml>

- *UseNet Tomfoolery*; <http://www.elsop.com/wrc/humor/usenet.htm>
- Roman Czyborra; *ISO 646 Good Old ASCII*; <http://czyborra.com/charsets/iso646.html>
- Roman Czyborra; *The ISO 8859 Alphabet Soup*; <http://czyborra.com/charsets/iso8859.html>
- RFC 977 ; *Network News Transfer Protocol*; <http://www.faqs.org/rfcs/rfc977.html>
- RFC 1036 ; *Standard for Interchange of USENET Messages*; <http://www.faqs.org/rfcs/rfc1036.html>
- RFC 2980 ; *Common NNTP Extensions*; <http://www.faqs.org/rfcs/rfc2980.html>

Resources in German

- ++ Michael Prokop; *Newsgroup/Usenet*; <http://www.michael-prokop.at/internet/newsgroup.html>
- ++ Volker Gringmuth; *Zitieren im Usenet*; <http://einklich.net/usenet/zitier.htm>
 - Dirk Nimmich; *Wie zitiere ich im Usenet?*; <http://learn.to/quote>
 - Heinz Mybach; *Newsserverguide*; <http://www.newsserverguide.de/>
 - Alexander Griesser and Christoph Pittracher; *Häufige Fehlverhalten im USENET*; <http://www.tuxx-home.at/usenet.php>
 - *Mini-FAQ: Falsche E-Mail-Adressen*; <http://www.gerlo.de/falsche-email-adressen.html>
 - *Newsgroups: Usenet-Server der TU Graz*; <http://portal.tugraz.at/portal/page/portal/zid/netzwerk/dienste/newsgroups/>
 - *FAQ die ersten Schritte im Usenet*; aus `news:at.usenet.infos`
 - *FAQ der Newsgroup at.usenet.einsteiger*; <http://pamer.net/usenet/aue-faq.html>
 - Wolfgang Schmidhuber; *Probleme mit OutlookExpress*; <http://www.wschmidhuber.de/oeprob/index.html>
 - Boris Piwinger; *Erste Schritte für die Benutzung von Outlook Express*; <http://piology.org/news/oe-erste-schritte.html>

2.1 Usenet News

Usenet, the users' network, is the internet's prime discussion area:

- It comprises some 85,000 discussion groups or *newsgroups* dedicated to a specific topic.
- Newsgroup names are classified hierarchically by subject.
- *Articles* or *messages* are *posted* to these newsgroups by people using appropriate newsreader client software.
- *Articles* are then broadcast from news server to news server.

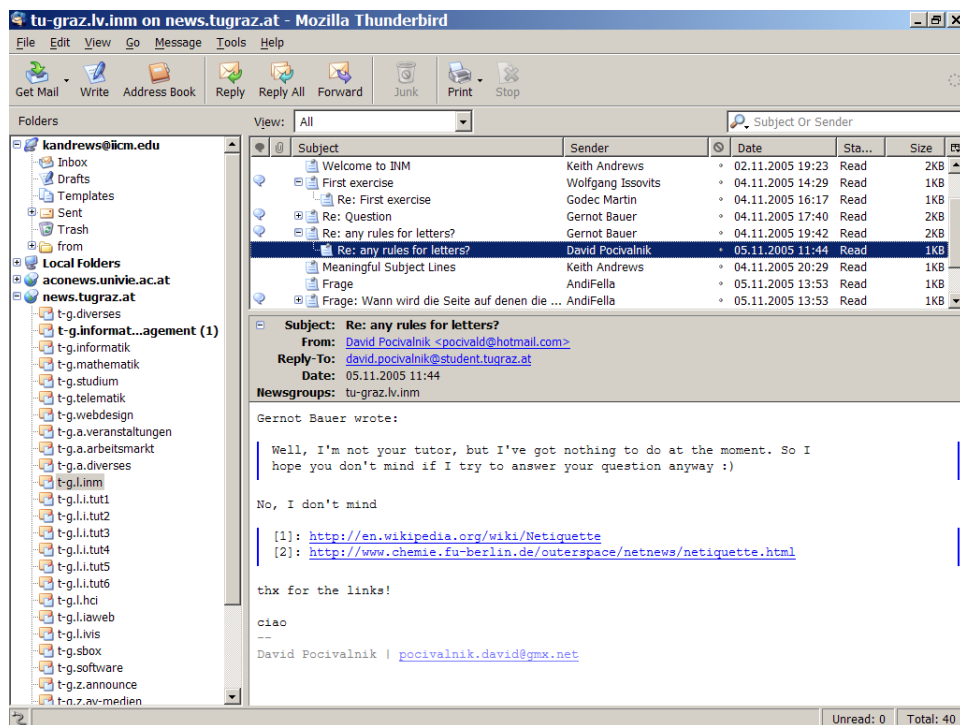


Figure 2.1: TU Graz Newsgroups in Thunderbird.

- Some newsgroups are *moderated*; in these newsgroups, articles are first sent to a human moderator for approval before appearing in the newsgroup.
- Some newsgroups are only *local* to a specific region or organisation and are not broadcast widely. Figure 2.1 shows the TU Graz newsgroups in the Thunderbird news reader.
- The set of freely available newsgroups is known as *Usenet*.

A Huge Resource for Discussion

With over two billion people online, it is potentially possible to get access to the world experts (and cranks) in every field:

- Want to know about recipes for Tibetan food? (`rec.food.cooking`)
- Or to ask about personal experiences with a particular internet service provider in Austria? (`at.internet.provider`)

Easy. Somebody else has probably already asked the same question and got several responses. Otherwise there is probably someone out there who can help you.

What Usenet Is Not

1. Usenet is not about delivering news (in the sense of Reuters or CNN).
2. Usenet is not an organisation.

3. Usenet is not a democracy.
4. Usenet is not fair.
5. Usenet is not a right.
6. Usenet is not a public utility.
7. Usenet is not an advertising medium.
8. Usenet is not the Internet.

What a Newsgroup Is

- A newsgroup is like a public notice board (say at your local supermarket or library), but on the internet.
- When you send (post) a message to a newsgroup, everyone who reads the group can see it.
- They can then contribute to the discussion publicly by posting a reply and/or contact you privately by email.
- Most newsgroups have some rules or conventions about what may or may not be posted.
- You have no idea who might be reading your postings, unless they themselves also post.
- It is possible to read any message in any group for as long as it remains on your news provider's server.
- This might be a few days or several months, depending on your news provider's policy for that newsgroup.
- Many newsgroups are archived externally for (almost) ever, for example by Google Groups (<http://groups.google.com/>).

Choosing a News Reader

A news reader is the client software which accesses a news server, allowing you to view and post to newsgroups.

There are many, many different news readers, check out http://en.wikipedia.org/wiki/Comparison_of_news_clients or <http://newsreaders.com/> for an overview.

For Windows:

- ++ Thunderbird <http://www.mozilla.com/thunderbird/>
- + 40tude Dialog <http://40tude.com/dialog/>
- + Newsman Pro Free <http://www.newsmanpro.com/>
 - Mahogany <http://mahogany.sourceforge.net/>
 - tin <http://www.tin.org/> (text-based news reader)
- Outlook Express is *not* a good news reader (see <http://www.wschmidhuber.de/oeprob/index.html> and <http://piology.org/news/oe-erste-schritte.html>)

- Opera Mail <http://opera.com/> is *not* a good news reader (it often produces invalid message ids and does not encode special characters in UTF-8 properly).

and many, many others.

For Mac:

- ++ Thunderbird <http://www.mozilla.com/thunderbird/>
- + MT-NewsWatcher <http://www.smfr.org/mtnw/>
- + Unison <http://www.panic.com/unison/> (not free).

and many others.

For Linux:

- ++ Thunderbird <http://www.mozilla.com/thunderbird/>
- Evolution <http://projects.gnome.org/evolution/>

and many others.

Web News Access

You can also access newsgroups with a web browser:

- Google Groups <http://groups.google.com/> provides web access to (and also archives) most of Usenet.
- WebNews is an alternative interface to the TU Graz newsgroups <http://webnews.tugraz.at/>.
- TUGnews provides web access to the TU Graz newsgroups <http://tugnews.tugraz.at/>. [This service is experimental.]

The Usenet Topic Hierarchy

Newsgroups are divided into topics using a simple hierarchical naming system, reading from left (broad) to right (more specific), for example:

- `comp.infosystems` deals with information systems in general.
- `comp.infosystems.www` deals specifically with the WWW.

Getting Access to Newsgroups

Your internet service provider (ISP) or university will typically provide access to a news server.

However, no news server carries *every* newsgroup:

- Some newsgroups are restricted to a local geographic area or a particular organisation.
- Others are banned for censorship reasons.
- Still others are deemed irrelevant and simply taking up resources.

alt.	Alternative, anarchic, and freewheeling.
comp.	Computing.
rec.	Hobbies and recreation.
sci.	Science.
soc.	Social, cultural, and religious.
misc.	Miscellaneous.
uk.	British topics.
de.	German-speaking newsgroups.
at.	Austrian topics.
tu-graz.	Graz University of Technology topics (local).

Table 2.1: Some of the top-level newsgroup hierarchies.

Newsgroups at TU Graz

- The TU Graz news server `news.tugraz.at` only carries the *local* TU Graz newsgroups.
- A list of all TU Graz Newsgroups is here: <http://news.tugraz.at/cgi-bin/usenet.csh>
- They are archived at `newsarchiv.tugraz.at`
- Remember that the TU Graz newsgroups are *public*. They can be read by anyone and are archived and indexed. Do not post anything which you may later regret.
- The central ACOnet news server `aconews.univie.ac.at` used to provide access to most of Usenet for all Austrian universities. It was turned off on 01 Jan 2012.
- For the time being, access to Usenet is possible from within TU Graz (log in via VPN) through the server `tugraz.zedat.de`.

See [ZID, 2010b] for more details.

Free and Commercial News Providers

Lists of news providers:

- `newsserverliste.cord.de`
- `newzbot.com`

Free news providers (usually initial registration):

- `albasani.net`
- `www.eternal-september.org`
- `http://www.visyn.net/newsserver/`

Pay-for news providers:

- `usenext.de`

Bin		0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
	Hex	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0000	0	nul	soh	stx	etx	eot	enq	ack	bel	bs	ht	nl	vt	np	cr	so	si
0001	1	dle	dc1	dc2	dc3	dc4	nak	syn	etb	can	em	sub	esc	fs	gs	rs	us
0010	2	sp	!	"	#	\$	%	&	'	()	*	+	,	-	.	/
0011	3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
0100	4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
0101	5	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
0110	6	'	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
0111	7	p	q	r	s	t	u	v	w	x	y	z	{		}	~	del

Table 2.2: The 128 characters of the 7-bit US-ASCII code table.

- `teranews.com`
- `individual.net`
- `giganews.com`
- `newsguy.com`
- `usenetserver.com`

2.2 Character Encodings and Plain Text

Character Encodings

- A *character encoding* assigns each particular character from a set to a unique sequence of bits or bytes.
- For example, Morse code encodes the latin letters, arabic digits, and some other characters each to a unique sequence of dots and dashes (the sequences vary in length).

See [Wikipedia, 2012b] for further details.

US-ASCII (7-Bit)

- The American Standard Code for Information Interchange (ASCII or better US-ASCII) [Wikipedia, 2012a] defines unique 7-bit codes for each of the 128 characters shown in Table 2.2.
- 95 of the 128 characters are *printable* characters (including space), the other 33 are non-printable *control characters*.
- ISO/IEC 646 [Wikipedia, 2012c; ISO, 1991] later defined many national variants of ASCII.

Bin		0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
	Hex	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0000	0	nul	soh	stx	etx	eot	enq	ack	bel	bs	ht	nl	vt	np	cr	so	si
0001	1	dle	dc1	dc2	dc3	dc4	nak	syn	etb	can	em	sub	esc	fs	gs	rs	us
0010	2	sp	!	"	#	\$	%	&	'	()	*	+	,	-	.	/
0011	3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
0100	4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
0101	5	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
0110	6	'	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
0111	7	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
1000	8																
1001	9																
1010	A	nbsp	ı	¸													
1011	B																
1100	C																
1101	D																
1110	E																
1111	F																

Table 2.3: The 256 characters of the 8-bit ISO-8859-15 code table.

ISO-8859 Character Encodings (8-Bit)

- ISO/IEC 8859 is a series of standards for 8-bit character encodings (upto 256 different characters).
- The two most important for Western Europe are:
 - ISO-8859-1: ISO/IEC 8859 Part 1 (Latin-1 Western European), which includes characters for most Western European languages.
 - ISO-8859-15: ISO/IEC 8859 Part 15 (Latin-9), which replaced some unused symbols and includes the Euro symbol ().
- The ISO-8859-15 code table is shown in Table 2.3.

Unicode

- The more recent Unicode standard [Wikipedia, 2012e; Unicode Consortium, 2012a] defines more than 110,000 characters, but uses between one and four bytes to represent each character. BabelMap [BabelStone, 2012a] gives a good overview of unicode characters and their various encodings.
- There are several character encodings for the Unicode character set:
 - UTF-8: usually one or two bytes per character, upto 4 bytes for more unusual characters [Wikipedia, 2012h].

- UTF-16: usually two bytes per character, sometimes 4 bytes for more unusual characters [Wikipedia, 2012g]. There are two possible byte orders, specified by UTF-16BE (big-endian) and UTF-16LE (little-endian).
- UTF-32: always four bytes per character [Wikipedia, 2012f].
- Note that many common fonts fail to provide shapes for every character in the Unicode character set: more unusual characters might not be available in a particular font.

UTF-8

- The first 128 characters of UTF-8 correspond exactly to US-ASCII. Hence, the first 128 characters (US-ASCII) need one byte each.
- The next 1,920 characters need two bytes each to encode. This covers most European languages and common symbols. For example, ü (u-umlaut) is encoded as 2 bytes (0xC3 0xBC).
- Rarer characters require either 3 or 4 bytes. For example, the Euro symbol € is encoded as 3 bytes (0xE2 0x82 0xAC) in UTF-8.
- The exact encoding scheme is quite complicated, see Wikipedia [2012h] for details.
- Tables showing the entire set of available characters are provided by Unicode Consortium [2012b].
- Individual characters can be looked up here: <http://www.fileformat.info/info/unicode/char/20ac/index.htm>

Plain Text Files

- *Plain text* is a sequence of displayable textual characters and control codes (such as linebreak or del).
- A plain text file is a sequence of characters (in the current character encoding) contained in a file.
- The file extension `.txt` is often used for plain text files.
- The only special-purpose characters usable for markup are the control characters present in the character set: such as newline, tab, and formfeed.
- Plain text files are human-readable as-is without any special decoding.
- Plain text files are used for program source code (`.c` or `.java`), config files (`.cfg`), news-group postings, email, web pages (`.html`), standardised markup (`.xml`), and many other things.

Plain Text Editors

A large number of plain text editors [Wikipedia, 2009c] are available:

- PSPad (Win) [Fiala, 2014]

- BabelPad (Win) [BabelStone, 2012b]
- Crimson Editor (Win) [Wikipedia, 2010a]
- Notepad++ (Win) [Wikipedia, 2009a]
- UltraEdit (Win, not free) [Mead, 2014]
- Sublime Text (Win, Mac, Linux, not free) [Sublime, 2014]
- RText (multi-platform, Java) [FifeSoft, 2014]
- Emacs (multi-platform) [GNU, 2014]
- Smultron (Mac) [Borg, 2014]
- BBedit (Mac, not free) [Bare Bones, 2014]

Word Processors

Word processors (like Microsoft Word or Open Office Writer) perform additional functions like layout, justification, and typesetting.

- Documents created by a word processor traditionally contain format-specific “control characters” beyond those defined in the character set which enable features like bold, italics, fonts, columns, tables, and so forth.
- Since raw word processor document files are no-longer human-readable, they are often referred to as binary files. A Microsoft Word `.doc` file is an example.
- Word processors can usually edit a plain text file and save in the plain text file format. However, care must be taken to tell the word processor application that this is what is wanted. Otherwise, the supposed text file may contain unwanted special characters.
- Microsoft Office 2007 introduced a new text document format (Office Open XML) based on standardised markup in XML. A `.docx` file is basically a zip compressed folder of XML files.

Plain Text vs Word Processor

- Figure 2.2 shows a directory listing of six files:
 - `my-name.doc` is a Microsoft word document in `.doc` format (22,016 bytes).
 - `my-name.docx` is a Microsoft word document in `.docx` format (13,140 bytes).
 - `my-name-7bit.txt` is a plain text file using 7-bit ASCII character encoding (36 bytes).
 - `my-name-iso8859-15.txt` is a plain text file using ISO-8859-15 character encoding (36 bytes).
 - `my-name-utf-16le.txt` is a plain text file using UTF-16LE character encoding (74 bytes).
 - `my-name-utf-8.txt` is a plain text file using UTF-8 character encoding (36 bytes).

Name *	Extension	Size
..		
my-name.doc	.doc	22.016
my-name.docx	.docx	13.140
my-name-7bit.txt	.txt	36
my-name-iso-8859-15.txt	.txt	36
my-name-utf-16le.txt	.txt	74
my-name-utf-8.txt	.txt	36

Figure 2.2: Directory listing of plain text files and Word documents. Word documents are *not* plain text files and are much larger, even if they contain the same textual content.

- All the files contain the same content, the text:

My name is Keith.
I live in Graz.

- Figure 2.3 shows the UTF-8 encoded plain text file in the PSPad text editor. All four plain text files look the same inside a text editor. They differ in the way they are stored on disk (and sometimes internally).
- Figure 2.4 shows the word document inside Word 2010.
- However, the different file types have very different sizes and internal structures when saved to disk.

Seeing the Differences Inside Plain Text and Word Processor Files

The internal differences in the various file types can best be seen using a Hex Editor such as HxD [Hörz, 2012]:

- Figure 2.5 shows the UTF-8 encoded plain text file byte for byte using the HxD hex editor. Note that the 34 bytes each encode one character of the text. Since no special characters have been used inside the document, the file is byte-for-byte identical to both the 7-bit ASCII (Figure 2.6) and the ISO-8859-15 (Figure 2.7) encoded files.
- Figure 2.8 shows the .doc format word document byte for byte using the HxD hex editor. Note that the word document contains many more bytes (22,016) than are necessary simply to encode the text string. Various settings and formatting are also specified. The string itself is somewhere in the middle of the binary file.
- Figure 2.9 shows the .docx format word document byte for byte using the HxD hex editor. A .docx file is actually a zipped folder of various XML files, as can be seen in Figure 2.10.

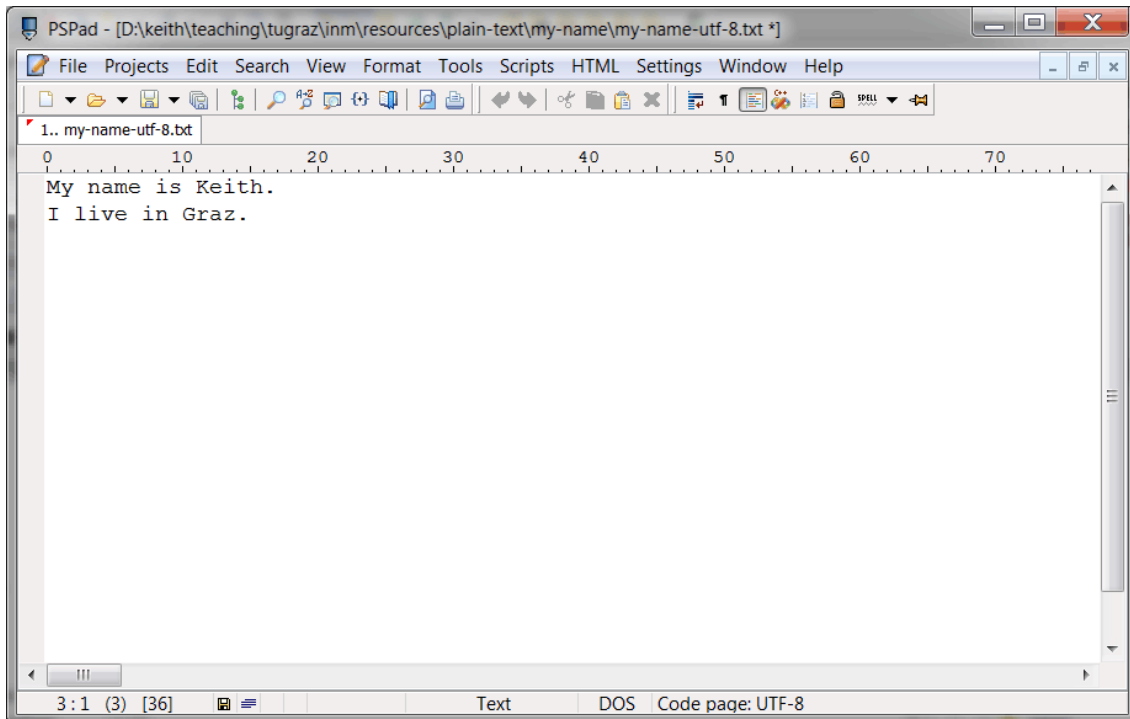


Figure 2.3: The UTF-8 encoded plain text file viewed using the PSPad Text Editor. All four plain text files actually look the same inside a text editor. They differ in the way they are stored on disk (and sometimes internally).

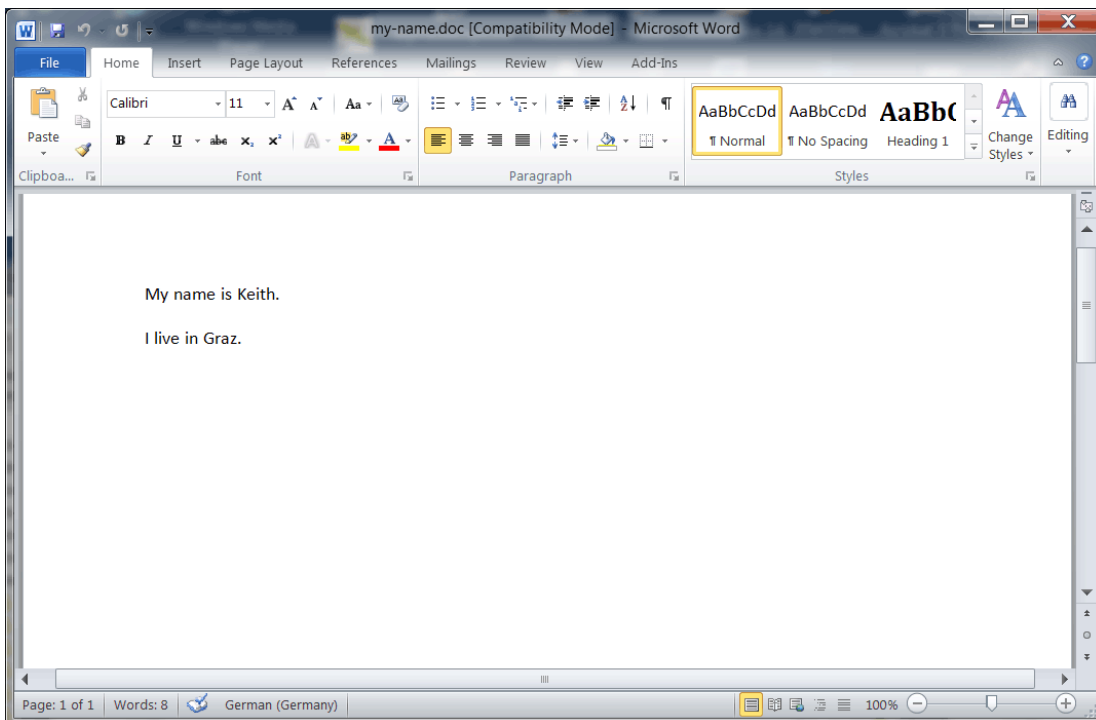


Figure 2.4: Word document shown in Word 2010.

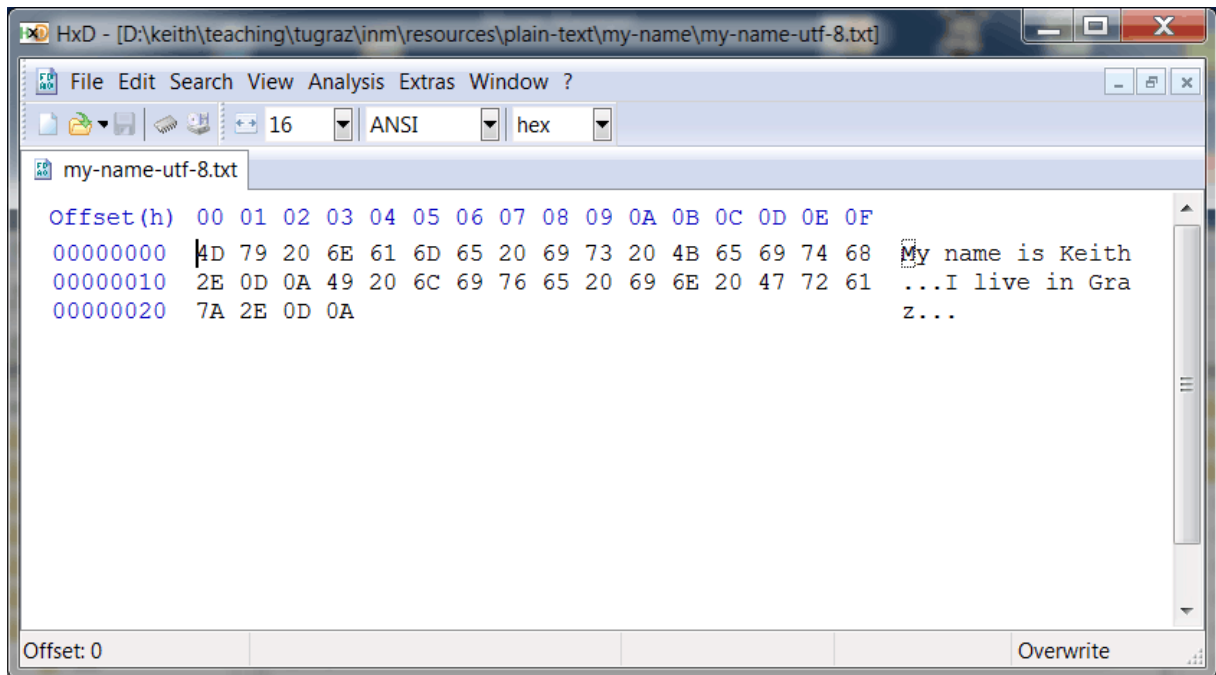


Figure 2.5: The UTF-8 encoded plain text file shown byte for byte in the HxD hex editor. Since no special characters have been used, it is, in fact, identical to the 7-bit ASCII (Figure 2.6) and ISO-8859-15 (Figure 2.7) encoded files.

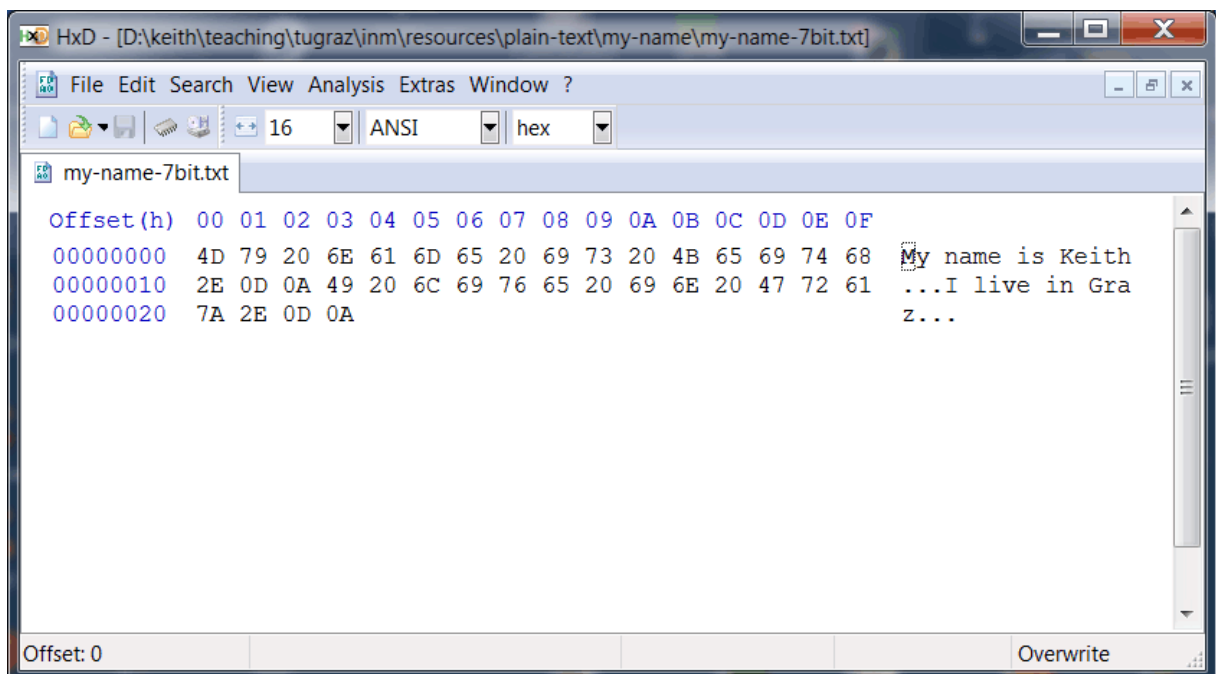


Figure 2.6: The 7-Bit ASCII encoded plain text file shown byte for byte in the HxD hex editor. Since no special characters have been used, it is, in fact, identical to the UTF-8 (Figure 2.5) and ISO-8859-15 (Figure 2.7) encoded files.

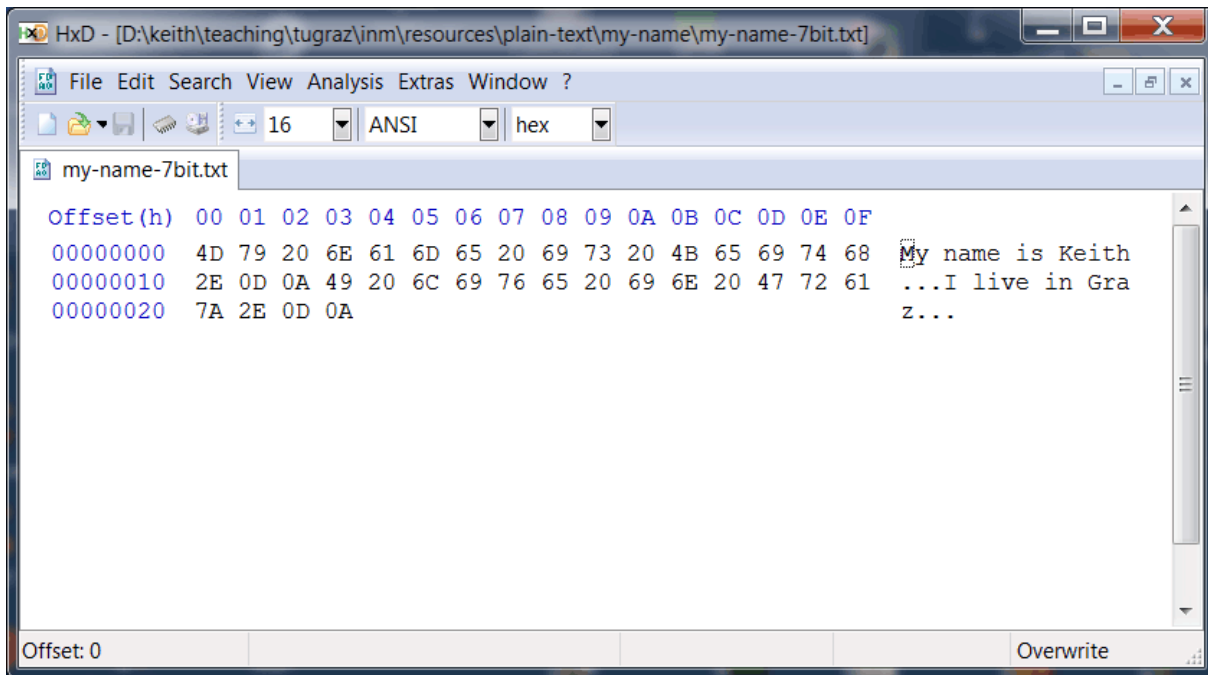


Figure 2.7: The ISO-8859-15 encoded plain text file shown byte for byte in the HxD hex editor. Since no special characters have been used, it is, in fact, identical to the UTF-8 (Figure 2.5) and 7-bit ASCII (Figure 2.6) encoded files.

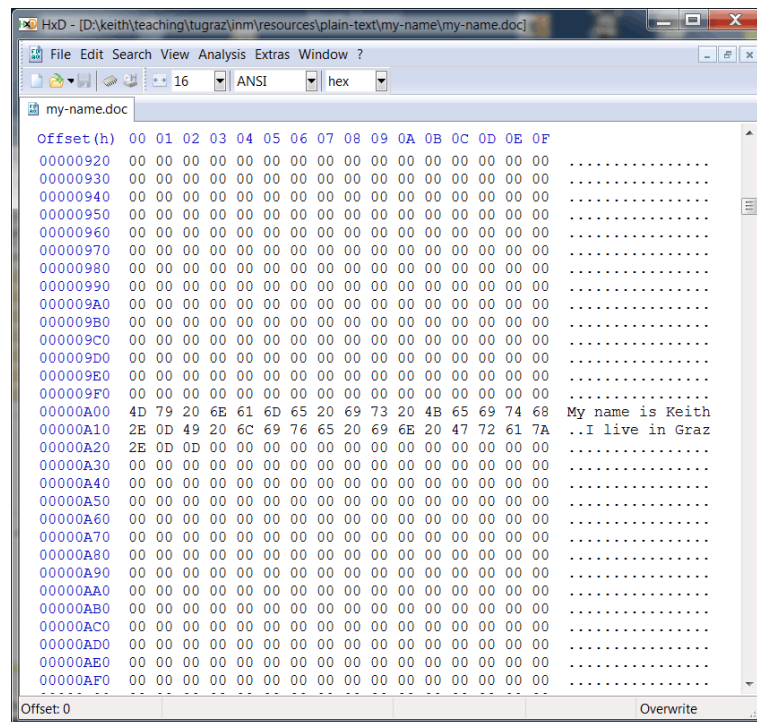


Figure 2.8: The .doc Word document shown byte for byte in the HxD hex editor. Many more bytes are contained in the file than are necessary to encode simply the text string. Various settings and formatting are also specified. The text string itself is somewhere in the middle of the binary document.

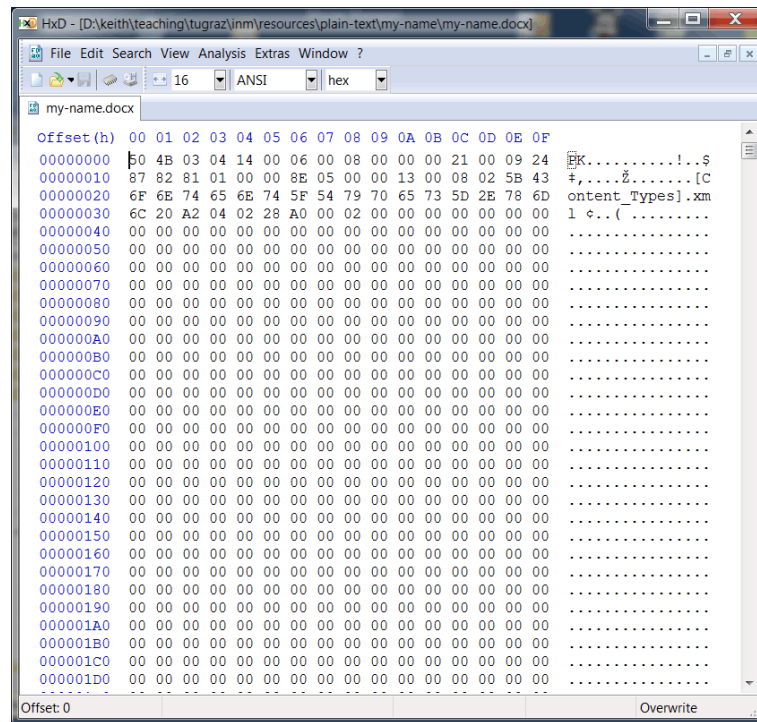


Figure 2.9: The .docx Word document (Office Open XML) shown byte for byte in the HxD hex editor. A .docx file is actually a zipped folder of various XML files, as can be seen in Figure 2.10.

Path	Name	Size	Ratio
	[Content_Types].xml	1.422	73%
_rels\	.rels	590	59%
docProps\	app.xml	984	52%
docProps\	core.xml	637	49%
word\	document.xml	2.005	67%
word\	fontTable.xml	1.186	62%
word\	settings.xml	3.421	65%
word\	styles.xml	15.015	87%
word\	stylesWithEffects.xml	15.768	87%
word\	webSettings.xml	450	40%
word_rels\	document.xml.rels	953	70%
word\theme\	theme1.xml	7.076	76%

Figure 2.10: A .docx file is actually a zipped folder of various XML files, shown here using WinZip. The text string can be found inside the XML file word/document.xml.

Binary Files

- Binary files (binaries) are sequences of bytes containing data which, in contrast to plain text, are not human-readable.
- Software has to read and decode the sequence of bytes, in order to reconstruct the original content.
- Binaries are generally large files such as images, sounds, video, word processor output (.doc, pdf), software packages, and executables.

2.3 Configuring Your News Reader

The most common settings include:

- *News Server*: the name of your news server (e.g. `news.tugraz.at`).
- *Name*: who should appear as the sender of the mail (e.g. Keith Andrews). Appears in the From: field.
- *Email Address*: where the message you send will appear to come from (e.g. `kandrews@iicm.edu`). Appears in the From: field.
- *Reply-To Address (optional)*: where private replies to your posting should be sent. Appears in the Reply-To: field.

Only use the Reply-To field if replies should go to a *different* email address than the one given in the From field. Specifying exactly the same email address in both the From and the Reply-To fields is redundant and is considered to be bad practice.

Using a Real Name

- Some users like to post under a pseudonym rather than their real name.
- However, the quality of discussion tends to suffer when people post anonymously.
- Many newsgroups, including all newsgroups in the `tu-graz.` and `de.` hierarchies, have a policy of using real names.

See <http://www.realname-diskussion.info/> and <http://www.wschmidhuber.de/realname/> for more information.

Using a Real Email Address

Since spammers regularly harvest all the email addresses from Usenet postings, some users prefer to post using a masked or fictitious email address:

- *Masking* means doctoring your email address in an obvious way so as to fool bulk mailing software but not human readers [Baseley, 1999], for example `kandrews at iicm dot edu`.
- The problem is that masking breaks the rules of Usenet and breaks the “Reply to Sender by Email” feature in newsreaders.

- A better solution is to use a real, working but never checked email address in the From: field and a real secondary email address in the Reply-To: field.

```
From: Keith Andrews <devnull@andrews.at>  
Reply-To: newsanswers@andrews.at
```

Most spammers only harvest email addresses from news posting headers, which contain the From: field but not the Reply-To: field.

- Or use your real email address and live with the spam.
- The TU Graz newsgroup charter [ZID, 2010a] requires that you post with a valid email address.

Finding and Subscribing to Newsgroups

Your news server probably provides in the order of 20,000 groups:

- In your news reader, browse through the hierarchy of folders, looking for newsgroups which might be appropriate. Figure 2.11 shows the subscription dialog in Thunderbird for the TU Graz newsgroups.
- Most news readers allow searching for groups on the server by name.
- You will not generally know which newsgroups are active or appropriate at first glance – you have to subscribe and read some of the postings to find out.
- You can also run a keyword search at Google Groups [Google, 2010] to locate newsgroups carrying postings of interest to you.
- Subscribe to the newsgroup by double-clicking or pressing “Subscribe”. The currently selected newsgroup is added to your personal list/folder of active newsgroups.

Reading Messages

Having subscribed to one or more newsgroups:

- Selecting a newsgroup name will generally start downloading the headers of current messages in that group.
- Selecting a message header will generally download its body into a separate window or panel.
- Some news readers support both downloading of message headers and bodies for offline reading of news.

Message Threads

- When a newsgroup message raises a new topic, it is said to *start a new thread*.
- Responses to the initial message, known as *followups*, are added to this thread.

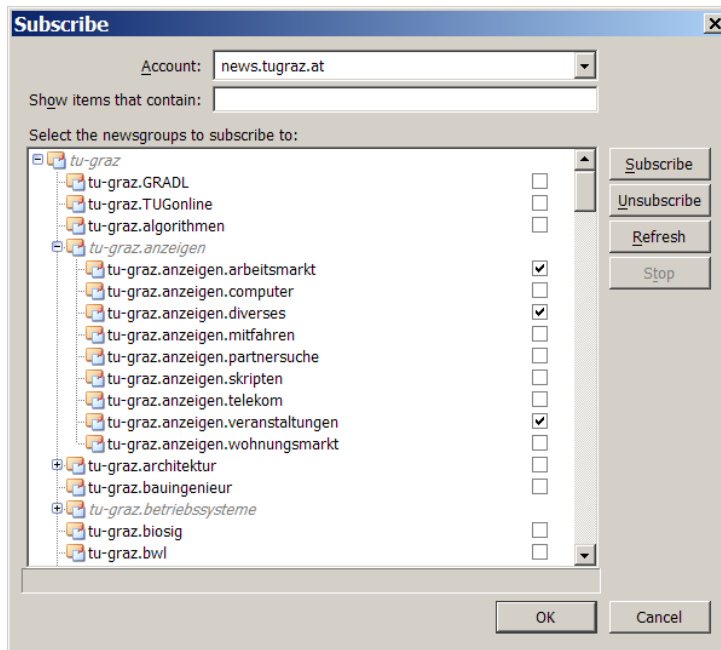


Figure 2.11: Subscribing to TU Graz Newsgroups. The dialog for choosing newsgroups in Thunderbird.

- News readers allow you to sort messages by threads, allowing you to follow the progress of a discussion.
- If you follow a group regularly, it is sometimes more convenient to sort messages by date instead, so you always see what is new.

Posting a Message

Posting a newsgroup message is like sending an email. You can either:

- start a new thread, or
- follow up an existing thread (and/or respond privately by email to the original sender).

It can take a few minutes for your message to appear on the news server, so wait a while before sending a duplicate message.

Post in Plain Text not HTML

- Posting anything other than *plain text* to standard newsgroups is offensive.
- Set your news client to send plain text messages by default.
- Not all news readers support HTML messages.
- HTML mail might look fancy, but detracts from the actual content, the words.

Prepare a Posting in a Plain Text Editor First

- For a longer posting, a good strategy is to compose your posting in a plain text editor first.
- When you are ready, copy and paste the posting into your newsreader for sending.
- Microsoft Word is a *word processor* [Wikipedia, 2010c], *not* a plain text editor. If you compose your posting in Microsoft Word and then paste it into your newsreader, you may introduce unwanted special characters.

Character Sets for Message Headers

- Message headers should only contain characters from plain old 7-bit US-ASCII, shown in Table 2.2.
- Full names in the From: field may contain only printing ASCII characters from space through tilde, except (,), <, and >.
- 8-bit characters, umlauts, etc. in message headers are not reliably interpreted, so stick to 7-bit ASCII.
- If your real name is “Jürgen Weiß”, write it as “Juergen Weiss”.
- Similarly, the message subject line should only use characters from 7-bit US-ASCII.

See <http://portal.tugraz.at/portal/page/portal/zid/netzwerk/dienste/email/nutzung/format#faq608>

For a full discussion of character set issues, and topics such as Unicode and UTF-8, see the excellent tutorial by Korpela [2007].

Character Sets for Message Bodies

- The only character set guaranteed to work fine in all newsgroups worldwide is 7-bit US-ASCII, shown in Table 2.2.
- Most newsreaders allow you to select character encodings for both message display and message composition.
- ISO-8859-1 (Latin 1) includes German umlauts, but not the Euro symbol.
- ISO-8859-15 (Latin 9) includes German umlauts and the Euro symbol.
- UTF-8 includes pretty much any character you would ever need and is now widely supported. I suggest you set your newsreader to send UTF-8. [In Thunderbird, newsgroup account - View settings for this account - Server Settings - Default Character Encoding]
- Opera Mail does not encode special characters as UTF-8 (even when UTF-8 is set in the Options), but escapes them.

See also <http://www.cs.tut.fi/~jkorpela/chars.html>

Posting Binaries

- Binary files are (generally large) files such as images, sounds, word processor output (doc, pdf), and software packages and executables.
- There are special newsgroups for posting binaries, which have `.binaries` in their name.
- Many news servers block the posting of binaries (they are always large, and often pirated or pornographic).
- Be extremely wary of installing software and programs posted to `.binaries` newsgroups, for fear of viruses.
- *Never* post binary files to standard newsgroups.
- *Never* attach binary files to postings to standard newsgroups. That includes, say a JPEG photo of an object you are trying to sell or a job description in PDF. [The conventional workaround is to upload your photo or PDF to an ftp server or a web server and include the URL to it in your posting.]

Posting URLs

- Manually shorten the URL to the shortest form that still works:
 - Try removing `www.` and see if it still works.
 - Try removing parameters at the end of a URL after `?`.
 - Do not remove `http://`, since newsreaders often recognise this to make links clickable.
- Do not use URL shortening services such as `tinyurl.com`, `goo.gl`, etc. They mask the original URL, so you have no idea where you will go when you click on the link. There is also no guarantee as to how long they will exist. [See [bit.do \[2014\]](#) for a list of URL shortening services which no longer exist.]

Message IDs

- Each news message must have a globally unique message id of the form `unique-part@FQDN`, where FQDN means a fully qualified domain name.
- Some news readers generate their own message ids. Forte Free Agent generates ids in the domain `4ax.com`, for example:

```
s6oum11noebcjmievjccqetj1q2jg0gd77b@4ax.com
```

- It is probably safer to let your news server generate the message id for you. This is the default in Thunderbird, for example:

```
j6l2uc$gns$1@news.tugraz.at
```

- Opera Mail generates its own message ids, but does not always produce valid FQDNs.
- See also <http://www.michael-prokop.at/internet/newsgroup.html#message-id>

Sending a Test Post

Once you have your news reader configured, you are probably itching to experiment:

- Do not rush into posting to real, active newsgroups!
- Try posting to `alt.test` or `tu-graz.test` first.

Viewing the Message Source

- In most news readers, you can view the message source to see what exactly was posted, including all of the headers and fields.
- In Thunderbird: View - Message Source (or Control-u on Windows).
- In Outlook Express: Properties - Details tab - Source.
- Use this feature to check exactly what users of a text-based news reader will see.
- Figure 2.12 shows the source of the message selected in Figure 2.1.

Canceling a Message

- If you have second thoughts about a message you have just posted, you can usually cancel it. In Thunderbird, select the message in the newsgroup and choose “Message - Cancel Message” (“Nachricht - Beitrag zurückziehen”). [Do *not* use “Delete Message”, found in the Edit menu and the right-mouse context menu of Thunderbird, which only deletes a message locally for you in your browser.]
- Your news reader will then send a cancel message to the news server. If your original message has already been sent on to other news servers, then the cancel message will be too.
- After a certain time, your message will have been removed from most news servers, but it is still possible someone might have seen and/or archived it.
- Some news servers do not honour Cancel messages, so you may not be able to cancel all instances of your message on all servers.
- If you wait too long, and someone else has already followed-up to your message, think twice before canceling it. If you do, you will leave any followups dangling as orphans.
- Under certain circumstances, it is also possible for other people to cancel your messages, see [Skirvin, 1999] for details.

[Note: “Delete Message” can be found in the Edit menu and in the right-mouse context menu of Thunderbird. It is usually disabled for newsgroup messages, because it only deletes a message locally in *your* copy of Thunderbird. It does *not* send a cancel message to the news server. It can be activated for newsgroup messages by setting the variable `news.allow_delete_with_no_undo` to true in “Tools - Options - Advanced - Config Editor”. However, it still only deletes the message locally in Thunderbird. Use “Message - Cancel Message” to actually cancel the message on the server.]

```

Path: news.tugraz.at!not-for-mail
From: David Pocivalnik <pocivald@hotmail.com>
Newsgroups: tu-graz.lv.inm
Subject: Re: any rules for letters?
Date: Sat, 05 Nov 2005 11:44:22 +0100
Organization: Technische Universitaet Graz
Lines: 15
Message-ID: <dki2e0$14j$1@fstgss00.tu-graz.ac.at>
References: <dkg6id$2hd$1@fstgss00.tu-graz.ac.at>
           <dkga2h$p3c$1@fstgss00.tu-graz.ac.at>
Reply-To: david.pocivalnik@student.tugraz.at
NNTP-Posting-Host: teacherads11.eduhi.at
Mime-Version: 1.0
Content-Type: text/plain; charset=us-ascii; format=flowed
Content-Transfer-Encoding: 7bit
X-Trace: fstgss00.tu-graz.ac.at 1131187456 21651 193.170.68.246
        (5 Nov 2005 10:44:16 GMT)
X-Complaints-To: news@tugraz.at
NNTP-Posting-Date: Sat, 5 Nov 2005 10:44:16 +0000 (UTC)
User-Agent: Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.7.12)
        Gecko/20050915
X-Accept-Language: en-us, en
In-Reply-To: <dkga2h$p3c$1@fstgss00.tu-graz.ac.at>
Xref: news.tugraz.at tu-graz.lv.inm:1142

```

Gernot Bauer wrote:

```
> Well, I'm not your tutor, but I've got nothing to do at the moment. So I
> hope you don't mind if I try to answer your question anyway :)
```

No, I don't mind

```
> [1]: http://en.wikipedia.org/wiki/Netiquette
> [2]: http://www.chemie.fu-berlin.de/outerspace/netnews/netiquette.html
```

thx for the links!

ciao
--

David Pocivalnik | pocivalnik.david@gmx.net

Figure 2.12: The message source of the message selected in Figure 2.1, including all of the various message headers.

Kill Files

- If you dislike a certain person on Usenet, you can “kill” their postings (for yourself).
- Add their email address to your news reader’s *kill file*. or use the Filters.
- You will never see their stuff again. Other people will though, of course, and their postings might quote from messages you have not seen.
- Be careful not to use too broad a filter, otherwise you might miss interesting stuff.

2.4 Netiquette

- *Netiquette* (net-etiquette) is a set of informal codes of conduct applying to behaviour on Usenet [Hambridge, 1995].
- Some additional rules or conventions may apply in particular newsgroups.
- If you participate in a community, you must follow the rules of that community.
- People who behave really badly on Usenet sometimes get reported to their news provider or ISP.
- But mostly they just get abused or ignored.
- On Usenet, personal abuse is called *flaming*.
- You don’t have to misbehave to get flamed, simply expressing a contrary point of view often does the trick.
- When a thread degenerates into nothing but name-calling, it is called a *flame war*.

Find the Right Group

- Search Google Groups and local news archives (<http://newsarchiv.tugraz.at/>) to find the most appropriate newsgroup(s).
- It is wise to *lurk* (reading to get the feel of a newsgroup) for a while, before posting yourself.
- Read the newsgroup’s charter.
- Posting a message not relevant to a particular group is known as posting *off-topic* (OT), and is frowned upon. For example, posting that you are looking for a flatmate to a newsgroup for a university course would be off-topic.

Look Before You Ask

Do not be lazy or waste other people’s time.

Before you post a question to a newsgroup:

- Read previous postings: to see if somebody else has already asked the same question.

- Read the FAQ: Many newsgroups have at least one FAQ (Frequently Asked Question) document:
 - The FAQ will describe the charter of the newsgroup, give guidelines for posting, and collect answers to frequently asked questions.
 - FAQs are usually posted to the newsgroup periodically but can also be found at Google Groups[Google, 2010].
 - Always read the FAQ before you post. Posting a question which is already answered in the FAQ is considered very uncool.
- RTFM (Read the F***ing Manual): read the manual and any documentation.
- STFW (Search the F***ing Web): or more mildly “Google is your friend”. Search before you ask. <http://lmgty.com/?q=tu+graz+vpn+client+download>

See “How To Ask Questions The Smart Way” [Raymond and Moen, 2012] for more very good advice.

Construct a Good Subject Line for New Threads

When starting a new thread, construct a good *subject heading* (or *subject line*):

- A meaningful, one line summary of the content of your message or question.
- The subject heading is used to identify the entire thread from there on.
- The subject heading lets people know what the thread is about and if it is worth their time downloading and/or reading it.
- Summaries of newsgroup messages usually *only* show the subject heading of each thread.
- The subject line should stand on its own, i.e. contain the key points from your posting in a single line.
- The subject line should be distinguishable from other subject lines (topics) in the same newsgroup.
- For example:

```
Bad:   Subject: Question
Good:  Subject: Q: Must We Use Our TU Graz Email Address?
Bad:   Subject: Exercise 1
Good:  Subject: Smoking Ban in Enclosed Public Spaces?
```

Do Not Multipost

- *Multiposting* means posting separate physical messages containing the same question or content to multiple groups.
- Multiposting annoys people, because they see the same message over and over again in different groups.

- Also, separate followups and discussions can start in the separate groups.
- Never multipost.

Crossposting

- You can *crosspost* (post a single physical copy of a message) to several groups by adding the other groups after the first group.
- Crossposting to two or three groups is generally considered acceptable.
- If a user reads a crossposted message in one newsgroup, they will no longer see it when reading any of the other newsgroups to which it was posted.
- When you crosspost you should set the Followup-To field to the most appropriate group:

```
From: Keith Andrews <kandrews@iicm.edu>
Subject: English Name for TU Graz
Newsgroups: tu-graz.studium, tu-graz.diverses
Followup-To: tu-graz.diverses
```

That way, any followups and discussion will be directed to one (and only one) newsgroup.

- Crossposting to more than three or four groups is considered spamming. (Within the tu-graz. hierarchy, it is not allowed to crosspost to more than two newsgroups).

Following Up

A *followup* is a new message posted to an existing thread:

- Like in email, you can quote parts or all of the original message(s).
- It does not hurt to respond by email as well as post a followup. The original poster will then get your reply instantly (don't forget to unmask their email address if masked).
- If you are following up a message, the original subject heading is usually retained and prefixed with "Re:" (automatically by your news reader) by default.

```
Subject: Re: English Name for TU Graz
```

- Sometimes, in the course of a thread, the subject under discussion changes emphasis. A convention is to change the subject line, but to include the old subject, along the lines of:

```
Subject: Native Speakers in Graz (was: English Name for TU Graz)
```

[In Thunderbird, when responding to newsgroup messages, be careful with the "Reply" buttons. The "Reply" button in the message pane (middle right) initiates a reply by personal email to the author of the newsgroup posting. Use the "Follow-Up" button in the message pane to send a follow-up to the message in the newsgroup thread.

The "Reply" button in the toolbar at the top of the main window actually initiates a follow-up (not personal email). "Reply All" in either location sends your reply both by email and to the newsgroup.

Check the heading fields in the message window which appears, to make sure you are sending to the right place.]

Quoting

- When following up, it is a convention to *quote* (include) the appropriate parts of the original message(s) to provide context or to comment upon.
- Depending on your settings, your new message will initially (automatically) contain a full copy of the original message with each line preceded by *quote tags* (>).
- In your followup, you can elect to keep the whole original message, parts of the original message, or delete the whole lot.
- The best strategy is *selective quoting*: selectively delete any non-relevant parts of quoted text and insert (interleave) your new text immediately following the relevant snippet (portion) of quoted text (see [Bell and Kottelin, 2014] or <http://einklich.net/usenet/zitier.htm>).

Subject: Re: English Name for TU Graz

Max Mustermann wrote:

> Is there an official translation for TU Graz in English?

Yes, there is: Graz University of Technology.

> And in French?

Not that I know of.

- Always quoting the entire original in every post is called a *full quote*. This practice makes followups grow larger and larger, wastes the time of your readers, and is frowned upon. Do not do it. [In German TOFU: Text Oben Fullquote Unten].
- Outlook Express makes it easy to full quote by starting your reply above the quoted text.
- Outlook Express fails to strip the signature away from any quoted text. You must do it manually.

Attribution when Quoting

- Make sure you preserve the *attributions* saying who wrote each of the quoted snippets.
- In some newsreaders, you can configure the format of the attribution string. See <http://www.guckes.net/faq/attribution.html>
- The attribution string should not be longer than one line, so it does not wrap.
- The name of the author is fine:

Keith Andrews wrote:

> Yes, there is: Graz University of Technology.

[This can be set in Thunderbird 17 using the Config Editor (Tools - Options - Advanced - General - Config Editor) to change the value of the variable `mailnews.reply_header_type` from 2 (default) to 1.]

- The name of the author plus their email address is best:

Keith Andrews <kandrews@iicm.edu> wrote:

> Yes, there is: Graz University of Technology.

[You can configure this in Thunderbird 17 by installing the SmartTemplate4 extension. In the extension's

Options, under the Reply tab, tick “Apply the following template to the new message”. Enter the string `On %X:=sent%%d% %b% %Y% %from% wrote:in` in the Quote Header text box. Leave the Template text box empty. Also tick “Apply the following template to the new message” under the Write tab, leaving the Template text box empty, so that any signature is applied to new messages.]

- Outlook Express (unnecessarily) appends the message-id of the original news posting to the attribution line, which generally causes it to be annoyingly wrapped over two lines:

```
In article news:dkftt4$r98$1@fstgss00.tu-graz.ac.at
Keith Andrews <kandrews@iicm.edu> writes:
> Yes, there is: Graz University of Technology.
```

In this case, you should manually remove the message-id before posting.

Signatures

- Most news readers let you add a personal signature (signature file, .sig) a few lines long at the bottom of your postings.
- Your signature might typically include your name, email, and home page address, and is often rounded off with a witty comment or saying.
- The signature is usually appended automatically to each message you compose.
- The convention [Gellens, 1999] for separating a signature from the rest of your text is to use a single line containing only dash-dash-space, “-- ” (the space is important!).
- Signatures should be 4 lines or less [Hambridge, 1995].
- See <http://www.newsreaders.com/guide/sigs.html> for an overview of newsreader support for signatures.
- Outlook Express places your signature *before* any quoted text. This is not good.
- For many years, Outlook Express produced an invalid signature separator (dash-dash without a space). The problem was fixed in Windows XP SP2. See <http://home.in.tum.de/~jain/software/oe-quotefix/>
- Do not confuse these signatures (.sig) with cryptographic signatures. Some newsreaders have problems dealing with PGP and GnuPG cryptographic signatures. [For our purposes in this course, please turn off cryptographic signatures when posting to newsgroups.]
- Thunderbird has a bug whereby it automatically appends a single blank line to your signature. This extra (possibly 5th) blank line, will be tolerated for the practical exercises.

Example of Newsgroup Signature

Here is an example of a typical newsgroup signature (the contents of a .sig file):

```
Keith Andrews                IICM, Graz University of Technology, Austria
http://www.iicm.tugraz.at/keith      "No wild kangaroos in Austria"
```

Note the sig separator (“-- ”) is usually inserted by the news reader, it should not be inside your .sig file.

Turn Off vCards

If your news reader lets you specify a vCard (or “Address Book Card”) to be attached to each message, turn it off:

- A vCard is sent as an attachment of type `text/x-vcard`.
- A message with a vCard attachment is sent as a multipart message containing two parts.
- The message is no longer simply plain text. Users of a text-based news reader will see lots of extraneous headers and clutter.
- Use a plain text signature instead.

See <http://www.cs.tut.fi/~jkorpe/usenet/vcard.html> for a longer discussion.

Line Lengths and Format=Flowed

- Do not use lines longer than 72 characters to avoid line wrapping when others quote your message.
[Thunderbird 2: Tools - Options - Composition - 523 Wrap plain text messages at 72 characters.]
[Thunderbird 3: Tools - Options - Advanced - General - Config Editor - mailnews.wraplength.]
- Check the line length by looking at the message source of a message you post (Control-u in Thunderbird).
- A Content-Type header option of `format=flowed`

```
Content-Type: text/plain; charset=ISO-8859-1; format=flowed
```

indicates that a news reader *may* re-flow the text.

- A news reader which understands `format=flowed` suppresses the display of quote characters (>). They are still there, but hidden.
- The news reader then re-flows the text according to the current window size and displays the quote hierarchy in some other way, usually with nested vertical bars to the left of quoted text.
- See Gellens [1999] and <http://www.joeclark.org/ffaq.html> for more details.
- You can turn off re-flowing in Thunderbird under Tools - Options - Advanced - General - Config Editor...

Search for Preference Names containing “flowed” and set as below:

```
mailnews.display.disable_format_flowed_support true
```

: -)	The basic smiley. Used to inflect a sarcastic or joking statement.
; -)	Winking. I just made a flirtatious and/or sarcastic remark. "Don't hit me for what I just said".
: -(Frowning. I am sad or upset about the last remark.
: -o	Shock.
X=	Fingers crossed.
: -X	My lips are sealed. I'll say nothing.
: *	Kiss.

Table 2.4: Common emoticons (smileys) used in (English) newsgroup postings and email to convey emotional connotations. They are read by tilting your head 90 degrees to the left.

Good Writing Style

- Use asterisks for light emphasis:

Only use Outlook Express if you **really** have to.

- Only use all upper case when you are shouting:

I explained that in my last posting.
WHY DON'T YOU READ BEFORE YOU POST?

Keep Your Cool

- Never post in anger, especially when everything you ever post is likely to be archived somewhere.
- Do not respond to flame-bait. Let it pass.
- Never post when you are tired. Get some sleep and post in the morning.
- If you have nothing constructive to say, please say nothing.

Beware of Humour

- Other cultures may have a differently developed sense of humour.
- Tread carefully with sarcasm, not everyone will get it.
- Even if you use a smiley... ;-)
- Smileys were suggested by Scott Fahlman in 1982 to specifically mark posts which were not intended to be taken seriously (see Fahlman [2007]).
- In some cultures, it is even necessary to explain that you are being ironic by appending the single-word sentence "Not." to your witty remark.

AFAIK	As far as I know.
ASAP	As soon as possible.
A/S/L	Age/Sex/Location
BFN	Bye for now.
BTW	By the way.
FYI	For your information.
g	Grin.
HTH	Hope this helps / Happy to help.
IMHO	In my humble/honest opinion.
LOL	Laughing out loud [at what you wrote].
RTFM	Read the f***ing manual [before you ask].
ROTFL	Rolling on the floor laughing [at what you wrote].
RSN	Real soon now.
SCNR	Sorry, could not resist.
STFW	Search the f***ing web [before you ask].
TIA	Thanks in advance.
WRT	With respect to.
YMMV	Your mileage may vary.

Table 2.5: Common acronyms and abbreviations used in (English) newsgroup postings and email.

Be a Good Netizen

- Post positively and invite discussion.
- Don't be a loudmouth.
- Remember the human – “Do unto others as you would have others do unto you.”
- On Usenet, people only know you from the quality of your postings.
- Do not post adverts or product endorsements.
- Never post personal email from someone else without their prior consent.

The TU Graz Cancelbot

- The TU Graz Cancelbot enforces many of the rules in the `tu-graz.*` charter [ZID, 2010a].
- Postings which break the rules are automatically cancelled and a cancel report is made to the newsgroup `news:tu-graz.cancel-reports`.
- The cancelbot also checks for spam. If you repeatedly post slightly different versions of the same posting from the same IP address (say, while testing), the TU Graz Cancelbot may think you are spamming and cancel your postings automatically. If that happens, try posting from a different IP address, or wait half a day or so for the Cancelbot's working set to be filled with newer postings.
- If your posting has gone missing, check for a message in `news:tu-graz.cancel-reports`.

- For example:

```
cancel reason(s):  
* binary  
* From: missing proper name
```

- For further details, read the TU Graz Cancelbot FAQ [ZID, 2010c] or Tim Skirvin's more general Cancel FAQ [Skirvin, 1999].

Chapter 3

Electronic Mail

This mail is a natural product. The slight variations in spelling and grammar enhance its individual character and beauty and in no way are to be considered flaws or defects.

[Hermann Maurer, email signature, 2001.]

Resources

- Wikipedia; *E-mail*; <http://en.wikipedia.org/wiki/E-mail>
- Kaitlin Sherwood; *A Beginner's Guide to Effective Email*; <http://webfoot.com/advice/email.top.php>
- David Shipley and Will Schwalbe; *Send: Why People Email So Badly and How to Do It Better*; 2nd Revised Edition, Vintage 24 Aug 2010. ISBN 030727599X (com, uk) [Shipley and Schwalbe, 2010]
- Terry Gray; *Comparing Two Approaches to Remote Mailbox Access: IMAP vs. POP*; <http://staff.washington.edu/gray/papers/imap.vs.pop.brief.html>
- Daniel Tobias; *Dan's Mail Format Site*; <http://mailformat.dan.info/>
- Wikipedia; *Internet Message Access Protocol*; http://en.wikipedia.org/wiki/Internet_Message_Access_Protocol
- Wikipedia; *Post Office Protocol*; http://en.wikipedia.org/wiki/Post_Office_Protocol

3.1 Starting with Email

Electronic mail (or *email*) is better than the post (“snailmail”), fax, and sometimes even the telephone:

- Extremely easy. Enter an address (or choose one from an address book), enter a short message, and press Send.
- Almost instantaneous, worldwide.

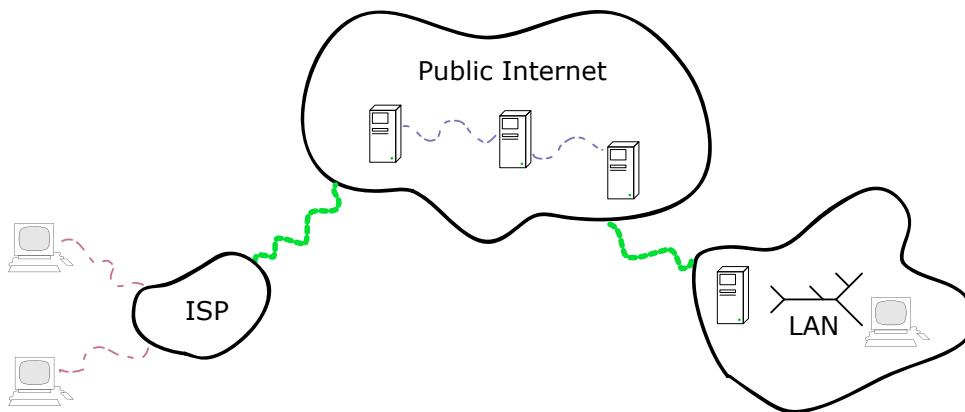


Figure 3.1: Electronic mail is passed from one routing point to the next until it reaches its destination.

- You get the actual text and not a photocopy. Plus no paper jams and busy signals.
- No synchronising of phone calls. People on the other side of the world answer while you are asleep.
- Can attach any computer file (image, spreadsheet, document).
- *Everything* you send and receive can be archived and is searchable.

Store and Forward

- Email messages are passed from computer to computer until their destination is reached (“store-and-forward”).
- Your mail server looks at the recipient’s address and passes the message on to its most appropriate neighbour, which does the same.
- This usually means messages are first routed towards the backbone, the chain of high-speed links which carry most of the internet’s long-haul traffic.
- Email is generally transmitted in the clear (i.e. unencrypted). Anyone along the route can take a look inside.

How Do I Read an Email Address?

- Email addresses all take the same form `someone@somewhere`.
- For example `kandrews@iicm.edu` says that the user’s name is `kandrews` and he is somehow affiliated with IICM, which is an educational (`.edu`) institution.
- The `somewhere` part is the *domain name* of the internet host which handles someone’s email. More on domain names later.
- The `someone` part usually identifies a person at that host address, but is sometimes a functional alias such as `help` or `office`.

Email Aliases

- An email alias is a virtual email address which forwards incoming email to one or multiple email addresses.
- For example, mails to `webmaster@iicm.edu` might actually go to `kandrews@iicm.edu` (they don't).
- Mails to `help@iicm.edu` might actually go to two people who run the help desk.
- Mails to `vrweb-project@iicm.edu` might go to all members of the VRweb project team.

Brief and Intimate

- Emails should be simple text files.
- The words are important, not the formatting.
- Do not worry about font sizes, logos, justification, or embossed paper.
- Email distills correspondence down to its essence, the *words*.

3.2 Different Kinds of Email Access

There are basically two kinds of email access:

- *Webmail*: access to email through a standard web browser.
- *Email reader* (or client): access to email using dedicated email software installed on your own computer.

Webmail

Webmail accounts work through the web:

- You use a web browser rather than an email client.
- You log into a web server to collect, write, and send your email.
- Your mail is physically stored on the provider's server:
 - + You can access your mail from any web browser anywhere, perfect for the traveller.
 - It is difficult to maintain a local archive of your email (and make backups).
 - You are dependent on the privacy policy of your provider.
[See EPIC [2007] for a discussion of gmail privacy issues.]
- There are dozens of free webmail providers:
 - Hotmail <http://hotmail.com/>
 - Gmail <http://mail.google.com/>
 - AOL Mail <http://mail.aol.com/>

- GMX <http://gmx.at/>

- ...

- Many include spam and virus scanning.
- Some webmail providers deactivate your account after a period of inactivity (Hotmail: 3 months).
- Since you can give any details you like when you register, webmail accounts are perfect for secondary email accounts or anonymous mail (although abusers can still be traced by their IP address).

Email Readers

A mail reader is specialised client software for reading and sending email. It connects to an:

- *Incoming* mail server: to receive email, usually either a POP3 or an IMAP4 server. POP (Post Office Protocol) [Myers and M. T. Rose, 1996; Wikipedia, 2008c] and IMAP (Internet Message Access Protocol) [Crispin, 2003; Wikipedia, 2008b; University of Washington, 2009] are protocols for receiving email.
- *Outgoing* mail server: to send email, usually an SMTP server. SMTP (Simple Mail Transfer Protocol) [Klensin, 2008; Wikipedia, 2008d] is the standard protocol for sending outgoing mail.

Note that receiving mail (POP or IMAP) and sending mail (SMTP) are *completely separate* services.

Choosing a Mail Reader

For Windows:

++ Thunderbird <http://getthunderbird.com/>

- + Mulberry <http://mulberrymail.com/>

- Eudora <http://eudora.com/>

- Opera <http://opera.com/>

- Windows Live Mail

- Outlook

- Outlook Express

- and many, many others.

Offline versus Online

In general, there are three paradigms for accessing email:

- *Offline*: The mail client fetches messages, messages are deleted from the incoming mail server and manipulated locally. Mail is composed offline and the mail reader connects to the outgoing mail server to send all unsent mail at once.
 - on-demand retrieval to a *single* client machine.
 - maintain your own local mail folder archive (and do your own backups).
 - + minimum connect time.
- *Online*: Messages are left on the incoming mail server and manipulated remotely. You must stay connected to the Internet the whole time.
 - + available from anywhere.
 - mail folders are kept on the server (and must be backed up there).
 - - connected for whole session.
- *Disconnected*: “cache” copy made on client before disconnecting, client reconnects and *synchronises* with the server. Like online but without needing to stay connected the entire time.

POP and IMAP

- POP (Post Office Protocol): is the original protocol supporting offline access to incoming mail, now in version 3 (POP3) [Myers and M. T. Rose, 1996; Wikipedia, 2008c].
- IMAP (Internet Message Access Protocol): newer and supports offline, online, and disconnected operation [Crispin, 2003; Wikipedia, 2008b; University of Washington, 2009]. It is also richer, supporting for example multiple mailboxes, shared mailboxes, and message status flags (seen, answered, deleted, etc.).

3.3 Configuring Your Email Reader

Probably the most difficult part about email is configuring your email client the first time.

The most common settings include:

- *Name*: whoever should appear as the sender of the mail (e.g. Keith Andrews). Appears in the From: field.
- *Email Address*: where the mail you send will appear to come from (e.g. kandrews@iicm.edu). Appears in the From: field.
- *Reply-To Address (optional)*: where replies to your mail should be sent. By default, replies will go to your regular email address (above), but you might want to divert replies to a second account (e.g. kandrews@tugraz.at). Appears in the Reply-To: field.

[Note that specifying the exact same email address in both the From: and Reply-To: fields is redundant and is considered bad practice.]
- *Outgoing Mail Server (SMTP)*: usually your provider (e.g. mail.iicm.edu).
- *Incoming Mail Server (POP3 or IMAP)*: where your mail is stored (e.g. pop.iicm.edu).

- *User Name*: sometimes, but not always, the first part of your email address (e.g. kandrews).
- *Password*: *****

Email at TU Graz

For students at TU Graz:

- *Incoming IMAP server*: `sbox.tugraz.at` with SSL/TLS and TUGrazOnline login details.
- *Outgoing SMTP server*: `mailrelay.tugraz.at` with SSL/TLS and TUGrazOnline login details.

Note that your TUGrazOnline user name must be written in *lower case* letters! Security should be “SSL/TLS” and Authentication “Normal password”.

See also <http://email.tugraz.at/> or <http://www.michael-prokop.at/faq/#mail> for details.

Plain Text versus HTML Mail

- Always send email as simple, plain text.
- Set your mail client to send Plain Text mail by default.
- HTML mail might look fancy, but detracts from the actual content, the words.
- Not all mail readers support HTML mail.

Character Encoding for Email

- Set your mail client to send UTF-8 by default. [In Thunderbird: Tools - Options - Display - Formatting - Advanced - Character Encodings. Set UTF-8 for Outgoing Mail and Incoming Mail. Tick the checkbox “When possible, use the default character encoding in replies”.]
- UTF-8 contains most characters you will ever need and is now widely supported.

Delivery and Read Receipts

- *Delivery receipts* acknowledge the safe arrival of your message on the recipient’s mail server, but only if it supports the Delivery Service Notification (DSN) standard.
- *Read receipts* notify you that your message has been opened, but only if the recipient’s mail program supports Message Disposition Notification (MDN) and it has been turned on.

In all, neither are particularly reliable or worth the bother.

3.4 Using Email

The Address Book

- Most mail readers allow you to maintain an address book with names, nicknames, email addresses, and other contact information.

- The simplest way to address a message, however, is to reply to a previous one.

Sending to Multiple People (To, CC, BCC)

- *To*: At least one email address should go into the To field. The To field indicates the main recipient(s) of the email, i.e. whom you would like to get a reply from.
- *CC*: The *carbon copy* field is used to inform further people of the email. Everyone in the To and CC fields is visible as a recipient of the email.
- *BCC*: The *blind carbon copy* field is used to send a copy of the mail to somebody, without other recipients knowing. Addresses in the BCC field are masked from all other recipients.
- To send a bulk mail without disclosing the list of recipients, put yourself in the To field and everybody else in the BCC field. Note however that some users filter out such emails, assuming them to be spam.

Forwarding a Message

- If you get an email you would like to pass on to someone else, you can *forward* it.
- Forwarded messages are like replies except they are not addressed to the original sender.
- You can tell if a message has been forwarded to you, because the Subject line will generally start with “Fwd”.
- It is generally better to forward messages *inline*, that is beneath a dashed line as part of the email itself, rather than as an attachment.
- Do not forward or distribute personal email you have received from someone else without first asking permission!
- And do not forward virus warnings, pyramid schemes, or any other form of chain letter to anyone else.

Sending Attachments

Files and documents can be sent via email as an *attachment*.

- The file should be encoded using an encoding standard which both the sending and receiving mail readers support (MIME is now widespread [Hood, 1998]).
- Never send an attachment larger than 100 kb without prior agreement. Large attachments fill up mailbox quotas and take ages to download.
- Before you send on a program or an Office document to others, make real sure that it does not contain a virus. That is a great way to lose friends.
- If the file you are sending is a plain text file, then do not send it as an attachment. Pasting it straight into the mail itself saves your recipients time and effort.
- Do not send Word documents and other proprietary formats unless you are sure your recipients can read them. Not everybody uses Microsoft Windows!

- Do not send vCards unless you are sure your recipients can read them. They cause the mail to be sent in two parts, which looks awful in a text-based email reader.

How Can I Cancel the Mail I Just Sent?

- You can't.
- Always check the recipients fields (To, CC, BCC) *before* you press "Send".
- Be careful about hitting "Reply All" when intending only "Reply".
- Never send a mail when you are tired or angry. An hour or two of reflection can save you a lot of grief.

Bouncing Mail

- If you send an email with a misspelled or non-existent email address, it will normally bounce back to you (from your outgoing mail server), with a message saying what went wrong.
- Sometimes this also happens if the recipient's mailbox is full and they are exceeding their quota.

Use a Meaningful Subject

- Like in newsgroup postings, the subject heading should be a meaningful one line (a few words) summary of the content.
- If you are replying to a mail, the original subject heading is usually retained and prefixed with "Re:" by default.

Subject: Re: Review OK by Fri.?

Sorry, Bill, but I **really** need the review by Weds.

- For time-critical mails, start the subject line with "URGENT: ", but do not overuse it.

Subject: URGENT: Need Review TODAY

Hi Bill, I need the review **TODAY**... otherwise I can't include it.

Replying

- "Reply" addresses a reply only back to the sender (unless the Reply-To field is set).
- "Reply All" addresses your reply to all recipients of the original email, which may or may not be what you want.
- Be careful when replying to a mail you received from a mailing list. You may intend to send a reply just to the author, but end up sending your (sometimes rather personal) message to everyone on the list. [Mailing lists sometimes set the Reply-To field to send replies to the entire mailing list.]

- For example, my good friend Ben. He intended to send a private reply only to Martin, but instead sent some rather personal comments about his ex-girlfriend to everyone on the mailing list, *including* his ex-girlfriend. . . and it is archived forever on Yahoo Groups [Jones, 2001].

Quoting

- Depending on your settings, your reply will automatically contain a copy of the original mail with each line preceded by *quote tags* (>).
- In your reply, you can keep parts of original mail, the whole original mail, or delete the whole lot.
- Your mail reader will usually automatically recognise a signature and not include it in the quoted material. If the signature of the original mail gets quoted, you should delete it by hand.
- The best strategy is *selective quoting*: selectively editing the quoted copy of the original mail to keep the relevant parts and interleaving your responses at the appropriate place(s).
- Quoting the entire original in your reply is called *full quoting*. It wastes the time of your recipients and is frowned upon.

Email Etiquette

- Most of the rules of usenet netiquette apply equivalently to email.
- Signatures are used analogously.
- The conventions for writing style, abbreviations, and smileys apply.

Good Email Style

- If you want something from the recipient, put separate requests into separate emails. Then the recipient can selectively reply to easier or more important points first and leave the others in their inbox.
- Use lines under 72 characters to avoid line wrapping when others quote parts of your mail.
- Use asterisks for light emphasis.

Please only send me personal email if you **really** have to.

Turn Off Virus Scanner Certification

- Some virus scanners append a short text to the bottom of incoming and/or outgoing emails, saying that they have been successfully scanned.
- Unfortunately, this text often breaks the rules for signatures.
- If your virus scanner does this, turn off certification for both incoming and outgoing email.
- In other words, still scan for viruses, but do not append any text to the emails.

- Some virus scanners only insert an extra header line, that is OK.
- In AVG under E-mail Scanner - Configure, uncheck the option Certify mail for both incoming and outgoing mail [Check yes, Certify no].

Primary Email Accounts

- A primary email account is a main email account, which you use and check regularly.
- Keep your primary email address to yourself and your friends.
- Do not use your ISP (internet provider) account as a primary email account. That way you have no ties and can change provider without having to change your email address if a better offer comes along.
- Most email accounts allow you to set up *forwarding*, so that all email arriving at that address is automatically sent on (forwarded) to another address. That way you can publish a particular email address and have all mail forwarded to where you currently read mail.

Secondary Email Accounts

- A secondary email account is one you only use and check occasionally.
- You might use a secondary email address when registering for a free service at a web site or when posting to newsgroups.
- Set up a (permanent) secondary email address at a free webmail provider.
- Services like 10 Minute Mail <http://10minutemail.com/> and spambox <http://spambox.us/> allow you to set up temporary secondary email accounts.

Filtering

Most mail programs can filter incoming mail into folders for you, based on finding specific phrases in the address or subject.

- Great for collecting mail from a mailing list into a designated folder.
- Not so good for blocking spam.

Spam

- *Spam* means unsolicited or undesired bulk electronic messages.
- Around 86% of all email sent in Sept 2009 was spam [Symantec, 2009].
- Around 4.5% of spam contained malware [Symantec, 2009].
- Originally, spam was (and is) a canned processed meat (like Extrawurst), and the name meant “SPiced hAM”. See Figure 3.2.
- During World War 2, spam was one of the few kinds of meat not subject to rationing in Britain, and people rapidly became fed up of eating it.



Figure 3.2: Spam canned processed meat. Spam is made from pork and the name means “SPiced hAM”. [Image (cc) Matthew W. Jackson, from the Wikimedia Commons [Wikimedia Commons, 2009]]

- This was the basis for the Monty Python comedy sketch, first broadcast on 15 Dec 1970.
- The term spam in computer usage is derived from the practice of driving unwelcome users out of early chat rooms by flooding the chat screen with quotes from the Monty Python Spam sketch.

See Wikipedia [2009b] for more details.

Avoiding Spam

- The best way to avoid spam is to avoid exposing your real email address in the first place.
- Mask your primary email address in your email signature and web pages (e.g. kandrews@snip-this-bit.iicm.edu or kandrews at iicm dot edu).
- On a web page you can also use a PNG image to display your contact details, which is easy for a human to read but hard to parse automatically.
- Use a secondary email account when posting to newsgroups or registering for web sites.

Coping with Spam

- Once spam starts arriving, there is not much you can really do about it, apart from changing your email address once in a while and notifying all of your contacts.
- Filters do not work 100% so you have to keep checking that legitimate mail did not get filtered out.
- Better to just delete spam on sight.
- Never *ever* reply to a spammer’s invitation to be removed from further mail – you are simply verifying that your email address is real and in use. It is then worth more to the spammers than an unverified email address.

Managing Your Email

- “Keep the Inbox Empty” strategy by Mark Hurst [Hurst, 2003; Hurst, 2007]
- Also advocated by Merlin Mann as “Inbox Zero” [Mann, 2010; Mann, 2007]

Organising Your Mail Archives

Archive your incoming and outgoing mail into folders so you can find things again.

- Think of a convention for naming folders and files, say by topic or by person. (I actually maintain the same topical hierarchy for my email folders, file directories, and web bookmarks).
- Some email may fit more than one category, file it under each (so you can find it again).
- Periodically transfer all of your mail folders into a longer term archive (make a backup on CDROM or DVD).
- When naming attachments or downloads to include a date in the file name, use ISO 8601 [ISO, 2004] format for the date (yyyy-mm-dd), for example `times-2006-09-29.pdf`. They are internationally unambiguous and can be sorted chronologically by file name.

Mailing Lists

- Automated list of email addresses.
- Like a newsgroup, but messages are sent by email.
- Subscribe to a mailing list, then send and receive messages.
- The owner of a mailing list knows exactly who is subscribed to the list (and can block, add, or remove subscribers).
- Mailing lists are sometimes *moderated*: postings and/or subscription requests are subject to approval by a list moderator to reduce spam.
- Web-based mailing list providers make it easy to set up and administer a list: see Yahoo Groups groups.yahoo.com, FreeLists freelists.org, Listbox listbox.com, or YLMP ymlp.com.
- Or run majordomo <http://greatcircle.com/majordomo/> on a (Unix) machine of your own.
Or ListMessenger <http://listmessenger.com>, Mailman <http://list.org/>, or SmartList <http://procmail.org/>.
- TU Graz runs a Mailman server for TU-related mailing lists at <http://mlist.tugraz.at/>

3.5 Securing Your Email

[Disclaimer: I am not an expert in either security or cryptography. Please let me know if you spot any glaring omissions or mistakes.]

Privacy

- Always assume your boss can read your email!
- Anyone who has access to your incoming mail server can read your email (sysadmins, university authorities, super users, ...).
- Anyone who has access to your computer can read your email.
- In fact, anyone with some technical competence who can tap into a server or network router along the route, can read your email!
- → Always assume everybody can read your email!

Cryptography Can Help

Cryptography can help secure your privacy in two ways:

- *Authentication*: ensuring the identity of your counterpart through a cryptographic signature.
- *Encryption*: scrambling the contents of a connection (or email), so (hopefully) only the intended recipient can unscramble it.

Public-Key Cryptography

In public-key (or asymmetric) cryptography, each user (or use case) has two separate but linked *keys*:

- *Public key*: made public to a greater or lesser extent by the user (handed out personally, sent out by email, or possibly uploaded to a key server, web site, TU Graz Visitenkarte, etc.).
- *Private key*: known only by the user and kept secret (but do not lose it!).

Think of a key as a very large number (or very long sequence of digits).

Using Public-Key Cryptography

- *Authentication*: encryption with private key, decryption with public key. [The sender encrypts a hash (or fingerprint) of the message with their private key. The recipient(s) decrypt the hash with the sender's public key, make their own hash of the message, and verify that they are the same.]
- *Encryption*: encryption with public key, decryption with private key. [The sender encrypts the message with the recipient's public key. The recipient decrypts the message with their private key.]

If someone encrypts an email with your public key, you can decrypt it with your private key (so long as you do not lose your private key!).

See ZID [2014] for a good overview in German.

Two Different Models of Trust

The problem is being able to trust that the public key you have for a user, *really* does belong to that user, and not to some imposter.

Apart from directly exchanging public keys, there are two different models of trust:

- *Hierarchy of Trust*: One central authority issues digitally signed certificates verifying the authenticity of lower-level authorities (Certificate Authorities, CAs), who issue certificates verifying the authenticity of owners and their public keys: forming a Public Key Infrastructure (PKI) such as X.509 [Wikipedia, 2014j; IETF, 2014]. [X.509 certificates are used to encrypt connections between client and server, such as by TLS/SSL for HTTPS, SMTP over TLS, IMAP over TLS, and are also used for end-to-end email encryption by S/MIME. The problem is: a signed certificate issued by a well-known Certificate Authority (CA) usually costs money, a self-signed certificate cannot be reliably validated.]
- *Web of Trust*: Another user checks your identity and your public key and signs a certificate endorsing the association between them. For a higher level of trust, this is often done by meeting in person (for example at a key-signing party) with photo id. A decentralised chain or web of trust is built up: A trusts B, and B trusts C, so A can probably trust C to some extent [Wikipedia, 2014i]. [This model is used for end-to-end email encryption by OpenPGP (GPG or GnuPG [Wikipedia, 2014d; GnuPG, 2014] is a widely-used implementation of OpenPGP [OpenPGP, 2014; Callas et al., 2007]).]

Most activists prefer the decentralised web of trust model.

Securing the Connection Between Mail Server and Client

- Most incoming (IMAP, POP) and outgoing (SMTP) mail servers now support some kind of secure connection to an email client.
- Transport Layer Security (TLS) [Wikipedia, 2014h] first authenticates its counterpart, then encrypts the data over the connection. [Secure Sockets Layer (SSL) was the predecessor to TLS and is still sometimes used.]
- TLS is usually assigned its own special port, to indicate that TLS is being used (port 465 for TLS-encrypted SMTP, port 993 for TLS-encrypted IMAP).
- STARTTLS [Wikipedia, 2014g] is a variant of TLS which can encrypt an existing clear text connection, without requiring a dedicated port.
- To offer TLS, the owner of a server must first obtain an X.509 [Wikipedia, 2014j; IETF, 2014] certificate.
- Note that setting your email client to use TLS only secures the communication between your email client and its incoming and/or outgoing mail servers. SMTP communication between mail servers along the route is not necessarily secure (although often is).

- Also note that email headers are always unsigned and unencrypted: hence liable to interception and/or manipulation.

Signing and Encrypting Messages in Thunderbird (End-to-End Encryption)

- GPG and Enigmail (web of trust): Install GPG [GnuPG, 2014] on your machine, install the Enigmail add-on for Thunderbird [Enigmail, 2014], create OpenPGP keys (key length at least 2048, better is 4096 bits), then configure Thunderbird to use them. At some point, go to a key-signing party to have your public key signed (cryptoparty.at). Make sure to backup your private key and make a revocation certificate in case your private key is lost or compromised.
- X.509 and S/MIME (hierarchy of trust): Obtain an X.509 certificate (say from TU Graz or Commodo), install the certificate in Thunderbird, then configure Thunderbird to use it (Account Settings – Security). [Thunderbird has built-in support for S/MIME, no add-on is necessary.]

See ZID [2014] for a good overview in German.

See Bleich and Neuhaus [2012] for a good guide (in German) to S/MIME for email encryption.

There is No 100 Percent Security

Even if you protect your email with the latest and best encryption, you may still be compromised in a variety of ways:

- There are bugs in software implementations, which cause vulnerabilities.
- There are backdoors built into software implementations, which leave deliberate vulnerabilities.
- Your computer may be stolen.
- Your passwords and private keys may be stolen in other ways (worm, virus, tojan horse, ...).
- With appropriate data mining, even your email metadata (when, to/from whom, subject) can reveal much about you. [Email metadata is typically not encrypted.]
- and so forth.

See also Chapter 7 on Staying Safe.

Chapter 4

Searching and Researching

“ There is nothing more perilous at a border crossing than a Google-happy border guard. Over the past year, two Canadians reported they were denied entry into the U.S. after a border guard Googled their names and decided, based on the search results, that they were undesirables. ”

[Samer Elatrash, Montreal Mirror, 05 Jul 2007.]

References

- + Milstein, Biersdorfer, and MacDonald; *Google: The Missing Manual*, 2nd Edition, O’Reilly, March 2006. ISBN 0596100191 (com, uk)
- + Dornfest and Calishain; *Google Hacks*; 3rd Edition, O’Reilly, Aug. 2006. ISBN 0596527063 (com, uk)
- Sergey Brin and Lawrence Page; *The Anatomy of a Large-Scale Hypertextual Web Search Engine*; Proc. WWW7, Brisbane, Australia, April 1998. [Brin and Page, 1998]

Online Resources

- Search Engine Watch; searchenginewatch.com
- Search Engine Guide; searchengineguide.com
- Michael Bergman; *Guide to Effective Searching of the Internet*; Dec 2004. <http://brightplanet.com/images/uploads/SearchEngineTutorialFormatted041218.pdf>
- Robert Harris; *Internet Search Tips and Strategies*; 06 Jul 2000. <http://www.virtualsalt.com/howlook.htm>
- Ian Rogers; *The Google Pagerank Algorithm and How It Works*; <http://sirgroane.net/google-page-rank/>
- Matt Cutts, *How Does Google Search Work?*; Google Webmaster Help video, 23 Apr 2012 <http://youtube.com/watch?v=KyCYyoGusqs> [Cutts, 2012]

Rank	Search Engine	Searches (millions)	% Share
1	Google	11,317	66.4%
2	Bing	2,710	15.9%
3	Yahoo	2,177	12.8%
4	Ask	550	3.2%
5	AOL	292	1.7%
<i>All US Searches</i>		<i>17,046</i>	<i>100.0%</i>

Table 4.1: The top five search providers in the USA. [Data as of Aug 2012 from comscore qSearch [comscore, 2012]]

4.1 Searching The Web

Directories (Catalogs)

Categories of web sites, organised *manually*:

- Yahoo Directory dir.yahoo.com
- dmoz dmoz.org
- Google Directory <http://google.com/dirhp/> (based on dmoz) was shut down in 2011.

Search Engines

Large, searchable indices of the web, constructed by *automated* crawlers:

- Google google.com
- Microsoft Bing bing.com
- Ask ask.com
- Yahoo Search search.yahoo.com
- ixquick ixquick.com promises privacy.
- DuckDuckGo duckduckgo.com promises privacy.

Human-Augmented Search Engines

Editors or end users provide input to make the search results better (writing summary answers for a particular search term, re-ordering result sets, etc.):

- ChaCha chacha.com
- Mahalo mahalo.com
- Blekko blekko.com
- Wolfram Alpha wolframalpha.com
- Wikianswers answers.wikia.com

Search Engine Mechanics

A search engine has three basic components:

1. *Crawling*: A robot or *crawler* visits web pages in sequence, storing a local copy of the page in a database.
2. *Indexing*: All of the words appearing in every page in the database are analysed and an index is created: for each word, a list of pages containing that word.
3. *Ranking*: For a given query, the set of best matching pages is determined and then ranked. Many factors are used to determine the ranking: relevance to the query, reputation of the page (PageRank), etc.

Why Google Returns the Most Relevant Results

Google uses a weighted combination of over 200 factors (ranking signals) [Cutts, 2012, at 05:45] to rank its organic search results.

Two such factors:

- *PageRank Algorithm*: Google's patented PageRank algorithm considers a link to a page from another page as a "vote" of support. Pages with more votes are more important [Brin and Page, 1998; Page, 2001; Rogers, 2002].
- *Anchor Text*: Google uses the link source anchor text (the blue underlined text) as a "description" of the web page pointed to by the link. The sum of anchor texts describing a page are considered in addition to the actual content of the page [Slegg, 2007].

Web Search Tips

- Think up a search term which is narrow enough to get rid of the junk, but broad enough not to miss anything useful.
- If you know an exact phrase, use exact phrase search (enclose the phrase in inverted commas).
- Remember that much content on the web is in English.
- Try both US and British spelling, e.g. visualisation and visualization, colour and color.

Google Search Tips

- keith andrews
Finds pages containing both the terms keith and andrews, not necessarily next to each other.
- "keith andrews"
Finds pages containing the *exact phrase* keith andrews.
- keith OR andrews
Finds pages containing either keith, or andrews, or both.

- `keith -andrews`
Finds pages containing keith but not containing andrews.
- `keith +andrews`
Finds pages containing keith which *must* contain andrews.
- `~pda`
Finds pages containing also synonyms and related terms to pda, for example palm and pocket pc.
- `"keith * andrews"`
An asterisk substitutes for any word in a phrase. Finds pages containing the keith thomas andrews as well as keith andrews.
- `keith andrews site:at`
Finds pages containing keith and/or andrews, but only in web sites in the at domain.
- `keith andrews site:infovis.org`
Finds pages containing keith and/or andrews, but only in the infovis.org domain.
- `train bristol intitle:timetable`
Finds pages containing train and/or bristol, but only with timetable in the title.
- `page rank filetype:pdf`
Finds only documents of type pdf containing page and/or rank
- `link:tugraz.at`
Finds pages which contain a link to tugraz.at

Image Search

- Google Image Search <http://images.google.com/>
- Flickr <http://www.flickr.com/>

Audio and Video Search

- Google Video Search <http://video.google.com/>
e.g. only videos which have subtitles
- YouTube <http://youtube.com/>
- YouTube Interactive Transcript
Can browse or search for words *inside* audio and video clips and jump to them. <http://www.youtube.com/watch?v=IY3U2GXhz44>

Back in Time

- Google Cache

Google keeps a copy of everything it indexes. If the page you are looking for has been deleted, Google may still have a copy in its cache.

The general syntax is:

```
http://webcache.googleusercontent.com/search?q=cache:www.something.com/path
```

For example:

```
http://webcache.googleusercontent.com/search?q=cache:courses.iicm.tugraz.at/inm/
```

- Internet Archive <http://www.archive.org/>

The Internet Archive keeps snapshots of web sites over time. For example:

- Enter [amazon.com](http://www.amazon.com) to see the progression of Amazon designs from 1996 to today.
- The Internet Archive still has a copy of the Managing Incoming Email PDF [Hurst, 2003], which is no longer available at <http://goodexperience.com/>

4.2 Academic Research

Book Search

- Amazon.com <http://amazon.com/>
 - Search by author, title, or ISBN.
 - Look at the reviews by other readers.
 - Look at the sales ranking.
 - Look Inside - full-text search of book contents.
- Google Book Search <http://books.google.com/>
 - Search by author, title, or ISBN.
 - Full-text search of book contents.

ISBNs

International Standard Book Number (ISBN):

- A unique identifier for (a particular edition of) a book.
- Used to be 10 digits (ISBN-10): for example 1843538393 for Buckley and D. Clark [2009].
- Since 01 Jan 2007 an ISBN has 13 digits (ISBN-13): for example 978-1843538394.
- Always give the ISBN if you list a book as a reference.

See Wikipedia [2008a] for more details.



Figure 4.1: Searching for variants of Edi Gröller’s name to find his papers using the Advanced Search of IEEE Explore.

Academic Research in Computer Science

To find research papers and articles in the area of computer science:

- ACM Digital Library <http://acm.org/dl>
[free from IP addresses within TU Graz (use VPN from off campus)]
- IEEE Explore <http://ieeexplore.ieee.org/>
[free from IP addresses within TU Graz (use VPN from off campus)]
- CiteSeer <http://citeseer.ist.psu.edu/>
CiteSeer collects, indexes, and cross-references articles and papers which are publicly available.
- Google Scholar <http://scholar.google.com>

The majority of good research papers in computer science are published with either ACM or IEEE, so having access to both their digital libraries is essential.

Figure 4.1 shows an example of searching for variants of Edi Gröller’s name to find his papers using the Advanced Search of IEEE Explore.

DOIs

Digital Object Identifier (DOI):

- A unique identifier given to an electronic document.

- Like a URL but independent of location.
- A DOI looks something like this: 10.1145/1168149.1168151.
- Any preceding “http://...” prefix is not part of the DOI.
- A DOI can be converted to a working URL by prepending the prefix “http://dx.doi.org/” to the DOI. For example: <http://dx.doi.org/10.1145/1168149.1168151>. [This should always work. If it does not, report the DOI in question to <http://doi.org/>.]
- Most published academic papers are now assigned DOIs by the publisher.
- If you cannot find a DOI for a paper anywhere else, try crossref.org.
- Always give the DOI if you list a paper as a reference (if the paper has a DOI).

See DOI [2010] and Wikipedia [2010b] for more details.

4.3 Browser Search Hacks

Firefox Search Engine List

- At the top right of the Firefox browser window, there is a search box with a configurable set of search engines.
- Open the drop-down and click on “Manage Search Engines” to configure the list of search engines.
- I have entries for: Google, Wikipedia (en), Wikipedia (de), Amazon (de), Amazon (co.uk), Amazon (com), Geizhals.at, and DOI Lookup, among others.
- To add a search engine to this list, someone must have previously prepared a search plugin for the search engine you want to add.
- A similar facility is available in IE.

Firefox Search Keywords

- The Firefox browser allows you to name a search box which appears on any web page and then automate searches using it.
- For example, go to <http://streetmap.co.uk/>, right-click on the search box, and choose “Add a Keyword for this Search...”.
- Give it the name “Streetmap UK” and keyword “uk”.
- Then a query on a UK address or postcode can be entered into the Firebox address field simply as “uk B73 6PJ”.
- This technique will work for most HTML search boxes.
- Make a search keyword “g” for quick Google searches.

4.4 Desktop Search

A desktop search program is installed locally on your PC:

- Incrementally maintains a searchable index of files and directories on your computer.
- Takes quite a long time (many hours) to construct the index the first time.
- Can then search for text phrases *inside* your documents, emails, and files.
- Different providers:
 - Copernic Desktop Search <http://copernic.com/>
 - Windows Search is built into Windows 7 and Vista.
 - Spotlight is built into Mac OS X.
 - Lucene is an open source software library for search and retrieval.
 - Google Desktop Search was discontinued on 14 Sept 2011.

Chapter 5

Plagiarism and Copyright

“ *Imitation is the sincerest flattery.* ”

[Charles Caleb Colton, English writer, 1780–1832.]

[Disclaimer: I am not a lawyer. The material in this chapter reflects the current situation as I understand it, but does not constitute legal advice and carries no guarantees of accuracy.]

References

- Justin Zobel; *Writing for Computer Science*; Second Edition, Springer, 2004. ISBN 1852338024 (com, uk) [Zobel, 2004]
- Lyn Dupré; *Bugs in Writing: A Guide to Debugging Your Prose*; Second Edition, Addison-Wesley, 1998. ISBN 020137921X (com, uk) [Dupré, 1998]
- Debora Weber-Wulff; *False Feathers: A Perspective on Academic Plagiarism*; Springer, 05 Mar 2014, ISBN 3642399606 (com, uk) [Weber-Wulff, 2014]
- Judy Anderson; *Plagiarism, Copyright Violation and Other Thefts of Intellectual Property*; McFarland, 1998, ISBN 0786404639 (com, uk)
- Maurer et al; *Plagiarism - A Survey*; J.UCS, August 2006. http://www.jucs.org/jucs_12_8/plagiarism_a_survey/

Online Resources

- Keith Andrews; *Writing a Thesis: Guidelines for Writing a Master's Thesis in Computer Science*; <http://ftp.iicm.edu/pub/keith/thesis/>
- Wikipedia; *Plagiarism*; <http://en.wikipedia.org/wiki/Plagiarism>
- *Center for Academic Integrity*; <http://academicintegrity.org/>
- iParadigms; *Plagiarism.org*; <http://plagiarism.org/>
- Theodore Frick; *Understanding Plagiarism*; <http://education.indiana.edu/~frick/plagiarism/>
- Hugh Levinson; *Brains for Sale*; BBC Radio 4 investigation, April 2005. http://news.bbc.co.uk/2/hi/uk_news/education/4445357.stm

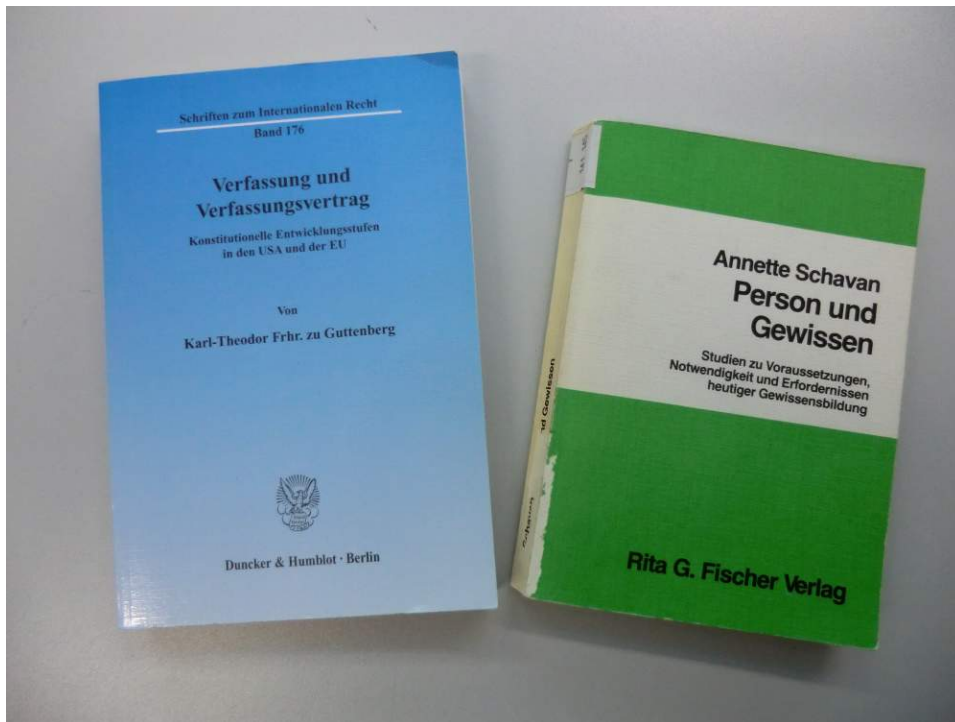


Figure 5.1: Two famous cases of plagiarism in Germany: the PhD theses of the (former) defence minister ~~Dr.~~ Karl-Theodor Frhr. von und zu Guttenberg and the (former) education minister ~~Dr.~~ Annette Schavan.

Resources in German

- Stefan Weber; *Das Google-Copy-Paste-Syndrom*; dpunkt, Nov. 2006. [In German] ISBN 3936931372 (com, uk) [Weber, 2006]
- Debora Weber-Wulff; *Fremde Federn: Plagiat Ressourcen*; [In German] <http://plagiat.htw-berlin.de/> [Weber-Wulff, 2010]
- Stefan Weber; *Blog für wissenschaftliche Redlichkeit*; [In German] <http://plagiatsgutachten.de/blog.php/>
- ~~Dr.~~ Karl-Theodor Frhr. zu Guttenberg; *Verfassung und Verfassungsvertrag. Konstitutionelle Entwicklungsstufen in den USA und der EU*. Duncker & Humblot, 2009. ISBN 3428125347 (com, uk) [Guttenberg, 2009] [In German] No longer available.
- ~~Dr.~~ Annette Schavan; *Person und Gewissen: Studien zu Voraussetzungen, Notwendigkeit und Erfordernissen heutiger Wissensbildung* Rita G. Fischer Verlag, 1980. ISBN 3883232203 (com, uk) [Schavan, 1980] [In German] No longer available.
- *GuttenPlag Wiki*; [In German] <http://de.guttenplag.wikia.com/>
- *Interaktiver Guttenberg Report*; [In German] <http://gut.greasingwheels.org/>
- *VroniPlag Wiki*; [In German] <http://de.vroniplag.wikia.com/>
- *PlagiPedi Wiki*; [In German] <http://de.plagipedi.wikia.com/>
- *Initiative Transparente Wissenschaft*; [In German] <http://de.antiplagaustria.wikia.com/>

- Der Standard; *Googler, Copy-Paster und Plagiatsjäger* Der Standard, 18 Mar 2007. [In German] <http://derstandard.at/2808256> [Standard, 2007]
- Debora Weber-Wulff; *Eine Professorin auf Plagiat-Jagd* Der Spiegel, 06 Nov 2002. [In German] <http://www.spiegel.de/unispiegel/studium/0,1518,221507,00.html>

5.1 Academic Integrity

Work which you submit as your own, must actually be your own!

You must take care to avoid both plagiarism and breach of copyright:

- *Plagiarism*: using the work of others *without acknowledgement*.
- *Breach of copyright*: using the work of others *without permission*.

The Google Copy + Paste Society

- It is very easy to find helpful material on the web.
- Do not be tempted to copy such material verbatim into your work and pass it off as your own.
- It is just as easy for your advisor or anyone else to check the originality of your work by copying a passage into Google or services such as TurnItIn [iParadigms, 2010], Docoloc [IFALT, 2010], or Ephorus [Ephorus, 2010], or Copyscape [Indigo Stream Technologies, 2010].
- Figure 5.2 shows an originality report from TurnItIn.

Plagiarism

- Plagiarism is a violation of intellectual honesty.
- Plagiarism means copying other people's work or ideas without due acknowledgement, giving the impression that these are your own work and ideas.
- Plagiarism is the most serious violation of academic integrity and can have dire consequences, including suspension and expulsion [Reisman, 2005].

The Concise Oxford Dictionary, 8th Edition, defines plagiarism as:

- plagiarise** **1** take and use (the thoughts, writings, inventions, etc. of another person) as one's own.
- 2** pass off the thoughts etc. of (another person) as one's own.

Helping Other People

- In a university environment, it is collegial to help other people by explaining things which you already understand.
- However, in the context of university practical work, be careful to explain things in general terms and to use a *different* example to the ones to be handed in.

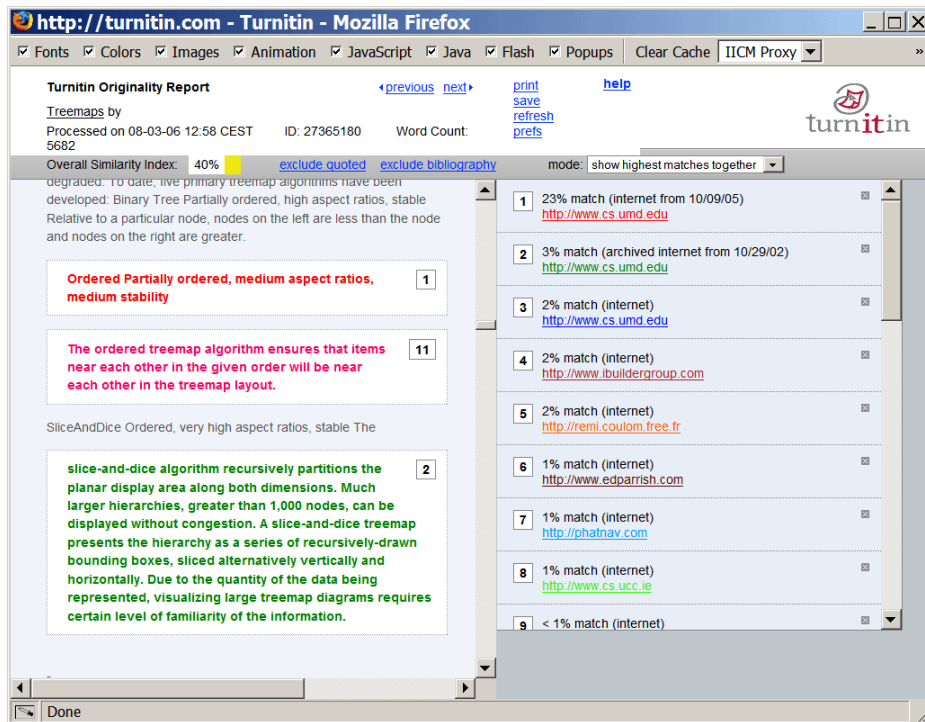


Figure 5.2: Part of an Originality Report from TurnItIn [iParadigms, 2010]. Copied material is colour-coded by source. On the left, part of the student report on treemaps as handed in. On the right, the sources from where text passages were copied verbatim.

- If you give your exercise solution to someone else, it is very tempting for them to hand in (almost) identical work, even if they promise and cross their heart not to.
- When two or more people (two or more groups for group work) hand in (almost) identical work, it is normal practice for everyone involved to be given 0 points, regardless of who did the work and who copied. You have been warned.

5.2 Copyright Law

Copyright law varies in detail from country to country, but certain aspects are internationally accepted:

- In general, the creator of a work, say a piece of writing, a diagram, a photograph, or a screenshot, *automatically* has copyright of that work.
- It is *not* necessary for the owner to explicitly claim copyright, for example by writing the phrase “Copyright Keith Andrews 2007” or by using the © symbol.
- Copyright usually expires 50 or 70 years after the creator’s death.
- The copyright holder can grant the right for others to use or publish their work on an exclusive or non-exclusive basis.

Fair Use

The copyright laws of most countries have provisions for *fair use*, for example:

- It is allowable to quote small parts of a work verbatim, providing the original source is cited.
- It is allowable to reformulate the work of others in your own words, providing the original source is cited.
- It is allowable to make copies at schools and universities for classroom teaching.

Austrian Copyright Law

- Austrian copyright law [UrhG, 2010, § 46] distinguishes between small quotes (Kleinzitaten) and large quotes (Großzitaten).
- Small quotes of published works are generally allowed, providing the original source is cited.
- Large quotes of published works, including quoting whole passages of text and using entire images and diagrams, are allowed in academic works, providing the original source is always cited.
- Austrian copyright law also makes certain exemptions for teaching materials used in schools and universities.

There are two good online guides to Austrian copyright law [FNM Austria, 2008; Nentwich, 2008].

Acceptable Academic Practice

Academic work almost always builds upon the work of others, and it is appropriate, indeed essential, that you discuss the related and previous work of others in your work.

However, this must be done according to the rules of acceptable use.

Acceptable Use

The two forms of acceptable use of material or ideas from somewhere else are:

- *Paraphrasing* with attribution (*indirect quote*).
- *Quoting* with attribution (*direct quote*).

Attribution means that the original source must be cited. Give as much detail as possible: author, title, name of publication, date, page number(s), DOI or ISBN, . . .

For further information on acceptable and non-acceptable academic practice see [Wikipedia, 2007c; Procter, 2010; Weber-Wulff, 2010].

5.3 Quoting Text

Paraphrasing (Indirect Quotation)

Paraphrasing means closely summarising and restating the ideas of another person, but in your own words.

When doing a literature survey, for example, you will generally want to *paraphrase* (parts of) each relevant paper or source.

My own tried and trusted technique for paraphrasing is:

1. Read the original source.
2. Put it down away from view.
3. *Without referring to the original*, summarise it in my own words.

Whenever you paraphrase someone else's ideas, you must cite the original source!

If your summary covers multiple paragraphs, include the citation of the source at the end of the first sentence of your summary.

Directly Quoting Text (Direct Quotation)

In some circumstances, you may want to *directly quote* small parts of text (typically upto a few paragraphs) from a relevant source:

- When quoting text, you must copy the original *exactly*, including all punctuation and even any spelling mistakes.
- Directly quoted text must be clearly distinguishable from text which you have written yourself, either inside quotation marks (inline quote) or inside an indented block (block quote).
- You must *always* cite the original source!
- It is *not* acceptable to simply copy and paste the words of others verbatim, if they are not clearly distinguished from your own words. That would give the impression that you wrote those words, when you did not. This applies even if an alibi source is given somewhere in the vicinity.
- In general, use direct quotations only if you have good reason.
- Most of your work should be written in your own words.

There are two ways to include directly quoted text into your own document: *inline quoting* for shorter pieces of text and *block quoting* for longer pieces of text (more than say 3 lines).

Inline Quoting of Text

Use inline quoting for direct quotations of shorter pieces of text, upto say 3 lines. For example:

Winston Churchill urged the country to stand firm, even after the fall of France: “Let us therefore [...] so bear ourselves that, if the British Empire and its Commonwealth last for a thousand years, men will still say, ‘This was their finest hour.’” [Churchill, 1940]

Block Quoting of Text

Use block quoting for direct quotations of longer pieces of text, more than say 3 lines. For example:

Winston Churchill urged the country to stand firm, even after the fall of France:

Hitler knows that he will have to break us in this Island or lose the war. If we can stand up to him, all Europe may be free and the life of the world may move forward into broad, sunlit uplands. But if we fail, then the whole world, including the United States, including all that we have known and cared for, will sink into the abyss of a new Dark Age made more sinister, and perhaps more protracted, by the lights of perverted science. Let us therefore brace ourselves to our duties, and so bear ourselves that, if the British Empire and its Commonwealth last for a thousand years, men will still say, “This was their finest hour.” [Churchill, 1940]

Conventions for Directly Quoting Text

- Different publishers have different conventions as to when to switch from inline to block quotations, often depending on the length of the quoted text: 3 lines, 4 lines, 5 lines, multiple paragraphs.
- American quotation style: double quotation marks “...” enclose quotations. Single quotation marks ‘...’ enclose a quotation within a quotation. At the end of the quotation, periods and commas are always placed inside the quotation marks.
- Traditional British quotation style: single quotation marks ‘...’ enclose quotations. Double quotation marks “...” enclose a quotation within a quotation. At the end, periods and commas are placed inside the quotation marks only if they are part of the quoted material.
- Block quotations are indented and set in a smaller font. The block itself is usually not enclosed in quotation marks.

I prefer a mixture: double outer quotation marks and single inner quotation marks (American), but punctuation placed inside the quotation marks only if is part of the quoted material (See NIVA [1996]).

5.4 A Worked Example

Original Piece of Text

The information pyramids approach utilises three dimensions to compactly visualise large hierarchies. A plateau represents the top of the hierarchy (or root of the tree). Other, smaller plateaus arranged on top of it represent its subtrees. Separate icons are used to represent non-subtree members of a node such as files or documents. The general impression is that of pyramids growing upwards as the hierarchy grows deeper.

An original piece of text about the Information Pyramids visualisation technique, taken from Andrews [2002, page 794].

Verbatim Copy without Attribution (Plagiarism)

The information pyramids approach utilises three dimensions to compactly visualise large hierarchies. A plateau represents the top of the hierarchy (or root of the tree). Other, smaller plateaus arranged on top of it represent its subtrees. Separate icons are used to represent non-subtree members of a node such as files or documents. The general impression is that of pyramids growing upwards as the hierarchy grows deeper.

A verbatim copy without attribution. The original passage has been copied word for word. There are no quotation marks or indentation to indicate a direct quote and no attribution of the source. This is a clear case of plagiarism.

Verbatim Copy with Attribution but No Distinction (Plagiarism)

The information pyramids approach utilises three dimensions to compactly visualise large hierarchies. A plateau represents the top of the hierarchy (or root of the tree). Other, smaller plateaus arranged on top of it represent its subtrees. Separate icons are used to represent non-subtree members of a node such as files or documents. The general impression is that of pyramids growing upwards as the hierarchy grows deeper. [Andrews, 2002]

A verbatim copy with attribution but no distinction. The original passage has been copied word for word. There are no quotation marks or indentation to indicate a direct quote. The impression is given that these words were written by the author, not copied from somewhere else. An alibi reference to the original source has been given, but this is not enough. This is a clear case of plagiarism.

Paraphrasing without Attribution (Plagiarism)

Information pyramids utilise three dimensions to compactly visualise large hierarchies. A plateau represents the root of the tree. Smaller plateaus arranged on top of it represent its subtrees. Separate icons represent non-subtree members of a node such as files or documents. The general idea is of pyramids growing upwards as the hierarchy grows deeper.

Paraphrasing without attribution. A few words have been changed or rearranged, but the original source has not been cited. A clear case of plagiarism.

Paraphrasing with Attribution (OK)

Information pyramids [Andrews, 2002, page 794] are a 3d technique for visualising large hierarchies. The root of the tree is represented by the lowermost plateau. Sub-directories sit as pedestals atop their parent's plateau. Icons are placed to represent documents and files. As the hierarchy grows deeper the pyramids grow upwards.

Paraphrasing with attribution. The original text has been completely rewritten in new words, but still conveys the original ideas. The original source is cited. This is standard academic practice.

Direct Inline Quote with Attribution (OK)

Information pyramids is an approach which: “[...] utilises three dimensions to compactly visualise large hierarchies.” [Andrews, 2002, page 794] Similar to treemaps in 2d, information pyramids uses the concept of successive inclusion to implicitly represent hierarchical structure.

A direct inline quote, enclosed in quotation marks, with the original source cited. This is acceptable academic practice for shorter quotations.

Direct Block Quote with Attribution (OK)

As Andrews [2002, page 794] states in his seminal work:

The information pyramids approach utilises three dimensions to compactly visualise large hierarchies. A plateau represents the top of the hierarchy (or root of the tree). Other, smaller plateaus arranged on top of it represent its subtrees. Separate icons are used to represent non-subtree members of a node such as files or documents. The general impression is that of pyramids growing upwards as the hierarchy grows deeper. [Andrews, 2002, page 794]

A direct block quote. The exact wording of the original is preserved. The quoted text is indented and set in a smaller font. The original source is cited. This is acceptable practice for longer quotations.

5.5 Quoting Images

Directly Quoting Images

Often, for example as part of a literature survey of related work, you will want to use photographs, diagrams, or screenshots taken from the internet or from another work:

- The safest policy is to ask permission from the copyright owner.
- If that is not possible, attempt to make a similar image yourself:
 - Download the software and make your own screenshot. Cite the source of the software.
 - Draw a similar diagram yourself using a drawing editor such as Adobe Illustrator or Inkscape. Cite the original source.
 - Use gnuplot, R, or a similar tool to produce your own graph of published tables of data. Cite the original source of the data.

5.6 Plagiarism Policy and Consequences

Plagiarism Policy at TU Graz

- The university has a Commission for Scientific Integrity and Ethics which formulates policy and oversees cases of suspected unethical conduct.
- Two relevant documents were published in the University Gazette (Mitteilungsblatt) of 01 Oct 2008 [TUGraz, 2008c]:
 - Code of Ethics [TUGraz, 2008a].
 - Guidelines for Scientific Integrity [TUGraz, 2008b].

Consequences of Plagiarism

If you are found to have committed plagiarism in a university setting, the consequences can (typically) range from:

- Receiving 0 points for that exercise.
- Receiving a fail for the entire course.
- Your thesis being rejected and a new topic being assigned.
- Your academic title being revoked.

5.7 Breaches of Copyright

Illegal Image Use

- Never copy an image from a web site (or anywhere else) and use it on your own web site (or post it on a forum, etc.).

- Images are always copyrighted. Initially by the creator, although the creator may then transfer the rights to a third party.
- Even if the web site site says something like “free recipes with images”, the images are only free to *look at*, not free to use!!
- You must obtain explicit permission from the copyright holder to use an image:
 - Sometimes blanket usage rights may be given, for example by the owner placing the image under a Creative Commons licence [Creative Commons, 2008].
 - Sometimes you can email the owner to obtain specific permission.

Marions Kochbuch

- In Germany, Folkert and Marion Knieper of Marions Kochbuch [M. Knieper and F. Knieper, 2008] are infamous [ARD, 2008; Kremer, 2007] for prosecuting hundreds of people who used (stole) photographs from their recipe web site.
- Many photographs from Marions Kochbuch are in the top ten results of German Google Image Search. For example, “Tomate” is number 1, “Banane” is number 2 (27 Oct 2008).
- Thousands of links and several indices increase the Google ranking of the Marions Kochbuch images.

Consequences of Illegal Image Use

- The Kniepers demanded € 8,632.60 in compensation and legal costs from Anja and Jens Reichmann for using several images on their bird web site [ARD, 2008].
- Hermann Kroppenberg had to pay almost € 7,000 for illegally using an image of Pumuckl (a cartoon character) on his cafe web site, which he copied from a seemingly free clip-art collection [handwerk.com, 2007].
- Sebastian Truxius was asked to pay € 2,000 for illegally using an official image of a t-shirt he was selling on ebay. He was later able to settle for EUR 300.
- Carsten Grentrup was asked to pay € 2,591.80 for 5 pictures of bread rolls and pretzels, which a user had posted in his web forum <http://www.abnehmen.com/> [ARD, 2008].
- Jürgen Bechstein was asked to pay around € 3,000 for one picture of a glass of ice tee, which an anonymous user had posted in his web forum <http://www.bundesligaforen.de/> [ARD, 2008].

Illegal File Sharing

- Most p2p file sharing networks (Kazaa, Limewire, Gnutella, eMule, BitTorrent, ...) work by simultaneously downloading and uploading [Wikipedia, 2007a; Taylor, 2007].
- While you are downloading a file, someone else is typically also simultaneously uploading (part of) it from you.
- In Austria (and most other countries), providing copyrighted material via the internet to other people (uploading) is illegal [UrhG, 2010, §18a], [Höhne, 2006].

- In Austria, simply downloading copyrighted material from the internet could be argued to be legal under the provisions for making a private copy [Schmidbauer, 2007].
- In Austria, IFPI Austria [IFPI, 2007] takes legal action on behalf of the music industry [Konsument, 2006].

Consequences of Illegal File Sharing

- In the US, in the first case to go to court there (rather than be settled out of court) Jammie Thomas was convicted of sharing 24 specific songs online and was fined \$ 220,000 [BBC, 2007].
- In the UK, on 22 Jul 2008 Isabella Barwinska was ordered to pay over £ 16,000 in compensation for illegally sharing the computer game Dream Pinball 3D [BBC, 2008].
- At TU Graz, the computer service department (ZID) receives several requests a year from lawyers regarding illegal file sharing.
- If a court order is presented and the data are available, the user data behind a specific file sharing session over TUGnet have generally been disclosed.
- In one case (in 2007), a student in Graz was offered out-of-court settlement terms of around € 4,000 for sharing around 1000 songs in the Gnutella network [Peter, 2007] (or otherwise legal action for around € 10,000 in damages).
- However, a recent (14 Jul 2009) Austrian supreme court decision ([OGH, 2009]) ruled that Internet Service Providers do not have to disclose the user details behind a particular dynamic IP address (and should not even be logging such data).

Chapter 6

Getting Connected

“ *There is no such thing as virtual beer.* ”

[Robert Cailliau, co-founder of the web, talking about why the web will not replace real meetings, at a conference in Graz, June 1994.]

References

- + Douglas E. Comer; *Computer Networks and Internets*; 6th Edition. Addison-Wesley, 02 Jan 2014. ISBN 0133587932 (com, uk) [Comer, 2014]
- James F. Kurose and Keith W. Ross *Computer Networking: A Top-down Approach*; 6th Edition. Pearson, 03 May 2012. ISBN 0273768964 (com, uk) [Kurose and Ross, 2012]
- Gary P. Schneider and Jessica Evans; *New Perspectives on the Internet: Comprehensive*; 9th Edition. Cengage Learning, 01 May 2012. ISBN 1111529116 (com, uk) [on the Internet: Comprehensive, 2012]
- + BBC; *The Web Weavers*; BBC / Open University [BBC, 1998], 30 minute video.

Online Resources

- + Wikipedia; *Internet Access*; [Wikipedia, 2014e]
- + Justin Ellingwood; *An Introduction to DNS Terminology, Components, and Concepts*; [Ellingwood, 2014]
- Daniel Tobias; *Dan's Domain Site*; domains.dan.info

Online Resources in Austria

- Geizhals; *ISP-Preisvergleich für Festverbindungen* <http://geizhals.at/isp/>
- AK Wien; *Konsumentenschutz - Handy & Internet* <http://wien.arbeiterkammer.at/beratung/konsumentenschutz/handyundinternet/index.html>
- AK Wien; *Tarifwegweiser Internet ADSL / xDSL / Breitband*; 10 Oct 2014 <http://mobilfunkrechner.de/akwien/pdf/internetadsl.pdf>

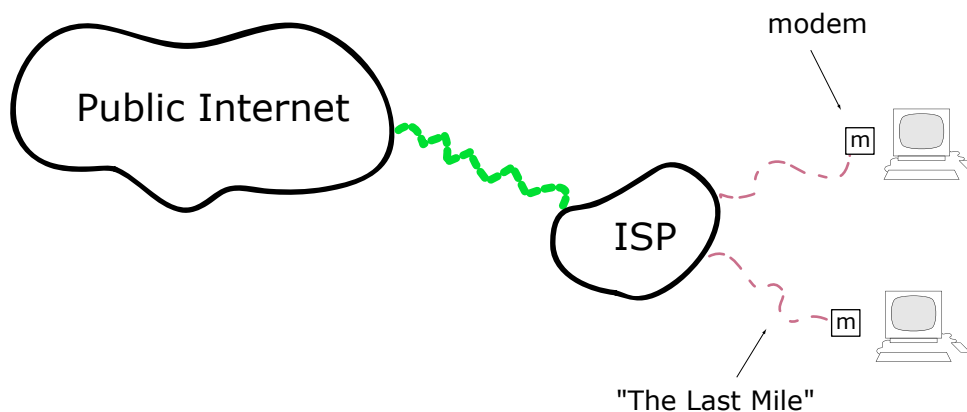


Figure 6.1: The last mile: the final part of the link to your computer.

- AK Wien; *Tarifwegweiser Mobiles Breitband*; 10 Oct 2014 <http://mobilfunkrechner.de/akwien/pdf/internetmobile.pdf>
- Newsgroup `at.internet.provider`
- Internet Service Providers Austria; `ispa.at`

6.1 Hooking Up

To get yourself connected, you will need:

- Device: computer, tablet, smartphone.
- Modem (modulator/demodulator): usually built in to device.
- Internet Service Provider (ISP).

The modem covers “the last mile” to your ISP and thence to the wider internet. See Figure 6.1

Bandwidth

The speed of computer connections is called *bandwidth*.

- Measured in bps (*bits* per second), kbps (kilo bps), mbps (mega bps), etc. [not bytes!]
- There are 8 bits (binary digits) in a *byte*.
- A theoretical speed of 56 kbps means max. 7 kb (7 kilo bytes) of data per second, but in practice it is usually less.
- A theoretical speed of 1 mbps means max. 125 kb of data per second, but in practice it is usually less.
- The term *broadband* is used to refer to high-speed internet connections of at least 256 kbps (using the OECD’s definition [OECD, 2014; Wikipedia, 2014e]).
- The term *narrowband* is used to refer to internet connections slower than 256 kbps.

- Test your bandwidth at:
 - speedtest.net,
 - speed.io,
 - speedtest.at,
 - <http://www.wieistmeineip.at/speedtest/>,
 - <http://www.bandwidthplace.com/speedtest/>,
 - <http://thinkbroadband.com/speedtest.html>,
 and many others.
- Overview of measured speeds around the world at netindex.com,

Latency

- Latency is the time delay between issuing a command (or clicking a link) and receiving (the start of) a response.
- Systems with high latency feel unresponsive and are not suited to real-time activities such as gaming or video conferencing.
- Typical internet connections over DSL or cable might have latencies around 20 ms or 50 ms.
- Mobile phone internet connections have fairly high latency around 100–150 ms (the new LTE connections have latency around 10 ms).
- Satellite internet connections have very high latency around 300–400 ms.
- You can use the ping program to check latency times, for example: `ping www.tugraz.at`

Typical Streaming Bandwidth Requirements

- Listening to radio, say BBC Radio 1 on RealPlayer at 44 kbps <http://bbc.co.uk/radio1/>, works out as:

44000 bits per second =
 5500 bytes per second =
 330000 bytes per minute =
 330 kb per minute =
 19800 kb per hour =
 20 mb per hour

The quality at 44 kbps is not particularly good, CDs are much better (1.2 mbps standard read speed).

- Watching standard definition (SD) TV, say Freeview (DVB-T) freeview.co.uk in the UK, which transmits at 4 mbps (MPEG-2 compression), works out at around 1.8 gb per hour.
- Watching HDTV, depending on the resolution and compression, transmits at around 10 mbps, which works out at around 4.5 gb per hour.

<i>Media</i>	<i>Bandwidth</i>	<i>Per Hour</i>
Radio (RealPlayer)	44 kbps	20 mb
SDTV 720x576i/50 MPEG-2	4 mbps	1.8 gb
HDTV 1280x720p/50 MPEG-4	10 mbps	4.5 gb
HDTV 1920x1080i/25 MPEG-4	13 mbps	5.8 gb

Table 6.1: The bandwidth requirements for streaming media such as radio and TV.

See Table 6.1.

6.2 Types of Internet Connection

- *Analog Dialup*: 56 kbps download/upload, per minute.
- *ISDN*: 2×64 kbps download/upload, per minute.
- *ADSL*: typical download 8 mbps, upload 1 mbps, flat rate (fair use) or transfer limit.
- *SDSL*: typical download and upload both 2 mbps, flat rate or transfer limit.
- *Cable*: download typically around 8 mbps, upload 768 kbps, flat rate (fair use) or transfer limit.
- *Fibre Optic*: download around 20 mbps, upload 10 mbps, flat rate (fair use) or transfer limit.
- *Wireless*: IEEE 802.11 standard, 10 mbps (802.11b), 54 mbps (802.11g) or 150 mbps (802.11n) download/upload, flat rate or per gigabyte.
- *Mobile Phone*:
 - LTE upto 150 mbps download.
 - HSPA+ upto 42 mbps download.
 - HSDPA: upto 7.2 mbps (but in reality around 1 mbps).
 - UMTS: 384 kbps.
 - EDGE: 200 kbps.
 - GPRS (HSCSD): 9.6 kbps or 14.4 kbps (per GSM channel). per minute or per megabyte.
- *Satellite*: download upto 6 mbps, upload via phone connection or via satellite.

Table 6.2 shows typical download amounts and times for a variety of internet connection types.

Analog Dialup

- 56 kbps download/upload.

<i>Connection</i>	<i>Bandwidth</i>	<i>Per Second</i>	<i>Per Minute</i>	<i>Per Hour</i>	<i>1 mb Download</i>
Analog dialup	56 kbps	7 kb	420 kb	25 mb	140 s
Mobile phone HSDPA	1 mbps	140 kb	8.4 mb	500 mb	7 s
ADSL	8 mbps	1.0 mb	60 mb	3.6 gb	1 s
WLAN 802.11b	10 mbps	1.4 mb	84 mb	5 gb	1 s
WLAN 802.11g	54 mbps	7.5 mb	450 mb	27 gb	0.1 s

Table 6.2: Typical download amounts per second, minute, and hour, and download times in seconds for a 1 mb download over various kinds of internet connection.

- Monthly fee plus metered (per-minute) call charges.
- Traditional analog telephone line and modem.
- Telekom Austria: need Telekom Austria land line, standard telephone monthly line rental from € 17.44, plus Telekom Austria internet tarif of € 0.0274 (peak) or € 0.0144 (off-peak) per minute. Discontinued 2012 ?
- selfnet free dialup: 01-962962 (Vienna) with user=selfnet and password=selfnet. See selfnet.at Discontinued 2011 ?

Integrated Services Digital Network (ISDN)

- 2×64 kbps download/upload, instantaneous connection.
- Monthly fee plus per-minute call charges.
- A digital phone line rather than analog.
- AonFlash Easy: Telekom Austria standard ISDN monthly line rental € 28,68, plus Telecom Austria internet tarif € 0.0274 (peak) or € 0.0144 (off-peak) per minute. Discontinued on 22 Apr 2011.

Asymmetric Digital Subscriber Line (ADSL)

- Typical download 10 mbps, upload 1 mbps.
- Flat monthly rate, but often transfer limit or “fair use”.
- High speed connection over old analog phone lines.
- Must be within 3–4 km of an ADSL-enabled digital telephone exchange.
- *Asymmetric*: different upload and download speeds.
- Either *bundled*: the last mile is provided by the monopoly telco (Telekom Austria monthly fee € 17.44 + ISP fee).
- Or *unbundled*: the last mile is provided by the ISP directly (no Telekom Austria fee, just ISP fee).

- upc Take IT max upto 20 mbps / 1 mbps: € 26,15 per month, unlimited. <http://www.upc.at/internet/dsl-internet/dsl-uebersicht/take-it-max/> (includes telephone connection and free calls to Austrian land lines).
- Tele2 Internet und Telefon XS upto 8 mbps / 768 kbps: € 19,05 per month, unlimited. tele2.at (includes telephone connection).

Symmetric Digital Subscriber Line (SDSL)

- Typical download 2 or 4 mbps, upload 2 or 4 mbps.
- Monthly rate, but often transfer limit or “fair use”.
- High speed connection in *both* directions over dedicated phone line.
- High upload speeds are necessary, for example, when running a web server at home.
- Must be within 3–4 km of an SDSL-enabled digital telephone exchange.
- *Symmetric*: the same upload and download speeds.
- Silver:SDSL:2048: 2048/2048 € 99.00 per month, unlimited. sil.at
- comteam SDSL 4096: 4096/4096 € 150.00 per month, unlimited. comteam.at

Cable

- Download typically around 10 mbps, upload 1 mbps.
- Flat rate, but often transfer limit or “fair use”.
- Area must be serviced by cable television.
- UPC fix internet upto 10 mbps / 1 mbps: € 17.15 per month, unlimited. <http://upc.at/internet/fix/>

Fibre Optic

- Download typically around 20 or 40 mbps (but potentially much higher, upto 100 mbps), upload around 10 mbps.
- Flat rate, but sometimes transfer limit or “fair use”.
- Area must be serviced by fibre optic provider.
- d-light.net fiber 50: upto 50 mbps / 50 mbps, € 32.90 per month, unlimited. If connection via Citycom Graz possible. <http://d-light.at/privat/internet/>
- d-light.net fiber 100: upto 100 mbps / 100 mbps, € 34.90 per month, unlimited. In Vienna. <http://d-light.at/privat/internet/>
- UPC Fiber Power Ultra: upto 150 mbps / 15 mbps, € 51.15 per month, unlimited. http://upc.at/internet/fiberpower_ultra/

Wireless LAN (WLAN or WiFi)

- WLAN ... wireless local area network.
- IEEE 802.11 Standard, upto 10 mbps (802.11b) or 54 mbps (802.11g) download and upload.
- Needs either close proximity to or line of sight to base station (typical range is 20–100 m).
- Uses unlicensed radio spectrum.
- Flat rate with transfer limit or per gigabyte.
- WestNet Privat Standard: download 8 mbps, upload 1 mbps, € 24.90, fair use. Directional radio. westnet.at
- T-Mobile HotSpot Anytime: 60 mins. for € 8. Network of hotspots in hotels, cafes, etc. around Austria and worldwide. <http://t-mobile.at/hotspot/Hotspot.php>

Free Hotspots

- There are many (deliberately) freely accessible wifi hotspots, some provided by businesses such as hotels and cafes, others by private individuals willing to share their bandwidth.
- WiFi hotspot finders, for example Wlanmap wlanmap.com, JiWire <http://v4.jiwire.com/search-hotspot-locations.htm>, or Boingo wifi.boingo.com.
- Fon is a worldwide wifi sharing initiative fon.com.
- Funkfeuer is a free wifi initiative in Austria funkfeuer.at.
- Many wireless networks are unknowingly or carelessly left open by their owners. Do not use such networks, it is illegal in many countries.
- Many cities and towns now provide free wireless hotspots, including: Linz hotspotlinz.at, Norwich, Oulu, Paris, Austin, and Mountain View.
- Graz currently has 32 free hotspots under the banner “Cityaccess”. freegrazwlan.at [Graz, 2014; Daublebsky, 2014]
- There is a list of free WiFi hotspots in Graz: wlangraz.at.

Mobile Phone (GPRS, UMTS, HSDPA)

- Data transfer over mobile phone network.
- GPRS upto 40 kbps, UMTS upto 384 kbps, HSDPA upto 14.4 mbps.
- Mobile phone internet connections usually have fairly high latency (100 or 150 ms is not uncommon).
- T-mobile All Inclusive Internet: upto 30 mbps / 5 mbps, € 16,67 per month, 10 gb (then 256 kbps). t-mobile.at
- Drei Hui 9gb: upto 4 mbps / 2 mbps, € 10.67 per month, 9 gb limit (then € 4,00 per gb). drei.at

- Drei Hui SIM 9gb: upto 4 mbps / 2 mbps, € 9 for 9 gb, credit expires after 30 days. [drei.at](#)
- Yesss Diskont-Surfen: upto 7.2 mbps, € 20 for 2 gb, credit expires after 12 months. Discontinued 15 Apr 2013. [yesss.at](#)
- Bob Breitband 1 gb: upto 3.6 mbps, € 5 for 1 gb, credit expires after 30 days. [bob.at](#)
- Mobile internet connections are often in practice much slower than advertised. [A consumer test in Austria in Sept 2008 found average download speeds of around 1 mbps for connections advertised as upto 7.2 mbps [AK Wien, 2008].]

Satellite

- Download upto 6 or 10 mbps, upload via standard telephone (one-way systems) or special i-LNB (two-way systems).
- Satellite dish must be installed with line of sight to satellite (usually Astra 1D at 23.5° East of South).
- Satellite internet connections have very high latency (400 ms is not uncommon).
- Some providers still require standard telephone line for up connection. New system allows two-way satellite connection with special LNB and satellite modem.
- Filiago SAT 10000 Speed (ASTRA2Connect, via Astra 1D). Two-way 10 mbps / 1.5 mbps, € 19.95 per month, plus one-off cost (dish, i-LNB, modem) of € 275. [filiago.de](#)
- sosat dsDSLcompact 1024 (ASTRA2Connect, via Astra 1D). Two-way 5 mbps / 1 mbps, fair use (progressively slower after 2 gb) € 24.90 per month, plus one-off cost (dish, i-LNB, modem) of € 499. [sosat.at](#)

WiMAX

- IEEE 802.16 Standard.
- Like wifi but longer range (several km).
- Typically 2–4 mbps download and 2–4 mbps upload within a few km, but much higher closer to cell tower (upto 30 mbps).
- Typically uses licensed radio spectrum.
- Needs a special external or internal WiMAX modem.
- Russia, Korea, and Pakistan have the largest Wimax networks (Oct 2010).
- In Europe, for example, Iberbanda runs a Wimax network in parts of Spain. AVI Pack Plus: 3 mbps / 3 mbps, free calls to land lines, € 47.53 per month. [iberbandainenglish.com](#)

See <http://en.wikipedia.org/wiki/WiMAX>

LTE (Long Term Evolution)

- The next big thing...
- 4G mobile communications technology.
- Download upto 150 mbps, upload upto 75 mbps, and low latency of around 10 ms.
- A1 Mobil Breitband Premium: upto 150 mbps / 50 mbps, € 59.90 per month, 35 gb (then 64 kbps). Available in the main cities. <http://a1.net/a1-mobil-breitband-premium>
- t-mobile My Net2Go Extreme i: upto 150 mbps / 50 mbps, € 32.99 per month, 20 gb (then 256 kbps). http://t-mobile.at/pdf/Entgeltbestimmungen_My_Net2Go_Extreme_i.pdf
- Drei Hui LTE Premium: upto 150 mbps / 50 mbps, € 46.67 per month, unlimited. <https://drei.at/portal/de/privat/tarife/internet/hui/>

See http://en.wikipedia.org/wiki/LTE_%28telecommunication%29

Virtual Private Network (VPN)

- A VPN is a private network carried (partly) over some larger (public) network.
- The connections are said to be *tunneled* through the larger network.
- One common application is to provide a secure connection to a home organisation over the public internet.
- A variety of standards exist, often involving encryption and authentication.

See http://en.wikipedia.org/wiki/Virtual_private_network

Choosing an Internet Service Provider (ISP)

Questions to ask your ISP:

- Is there a limit to the numbers of hours connection per month?
- Is there a download limit per month?
- Is there (free) phone support?
- Does it run a Usenet server (newsgroups)?
- How many email accounts are included?
- Is some free web space included?
- Can you register your own domain name?
- Is the ISP reliable and does it have a good reputation?

Connecting via the University

At Graz University of Technology:

- Wireless LAN on campus of TU Graz in various hot spots.
- Many student halls are connected to the Virtual Campus.
- Since 25 May 2010 TU Graz participates in eduroam (eduroam.org), the worldwide roaming infrastructure provided by educational organisations. See eduroam.tugraz.at for details.
- The university's analog and ISDN dial-in service closed down on 01 Jun 2006.

See tugnet.tugraz.at.

WiFi at TU Graz

- If you have a student or employee account at TU Graz, connect to the wifi network `tug-wpa` using an 802.1X client (these are integrated by default in current versions of Windows and MacOS) and enter your account details.
- If you have a guest account at TU Graz, connect to the wifi network `tug`, open a web browser, then enter your account details (“captive portal”).

See <http://tugnet.tugraz.at/zugang/wlan>.

VPN at TU Graz

TU Graz runs a Cisco VPN.

- From outside the TU Graz campus, you can use a VPN client to “log in” to the university's network.
- Once connected to the VPN, you are assigned an IP address from the university's pool (TUGnet, 129.27.*).
- You can then use services which are only allowed from within the university, such as downloading full papers from the ACM Digital Library <http://acm.org/dl>.
- For example, if I connect to the web site <http://whatismyip.com/> directly from home, I have the IP address 91.128.188.26. If I log in to the TU Graz VPN and then go to the same web site, I have the IP address 129.27.12.11.
- Using a reverse lookup service such as iplocation.net or ipaddresslocation.org indicates that (for example) 91.128.188.26 is assigned to Tele2, Austria and 129.27.12.11 is assigned to Technische Universitaet Graz, Austria.

Configuring a VPN Client for TU Graz

- The Cisco VPN client for Windows is available from the TU Graz Software Server swsrv.tugraz.at (log in with your TU Graz account).



Figure 6.2: The bottom device is the broadband (xDSL) modem provided by the internet provider (inode). The top device is a wireless router providing wifi throughout the flat.

- Many other clients are also available, see www.vpn.tugraz.at/#faq3656.
- Enter the following settings:
 VPN Server: 129.27.200.3
 Group: default
 Password: default
 Allow local LAN access
- Start the VPN client and enter your TUGrazOnline account name (lower case) and password.
- Full instructions are at www.vpn.tugraz.at.

WiFi at Home

- Once you have a broadband internet connection, a good way to share it between several computers is to set up your own home wifi network (WLAN).
- A wireless router (around € 30 or 40) is attached to your broadband provider's modem, as shown in Figure 6.2.
- Any computer with built-in wifi connectivity (or a wifi adapter) can now connect to the router (and thus to the internet) wirelessly.
- Network-enabled printers (or printers with a print server) and other computers can be connected directly to the router with an Ethernet (patch) cable.
- For better security, make sure to enable WPA2 encryption (WEP encryption is rather weak and can be cracked in a couple of minutes).
- Consider also using an access control list: a list of MAC addresses of devices which are allowed to connect to the router.

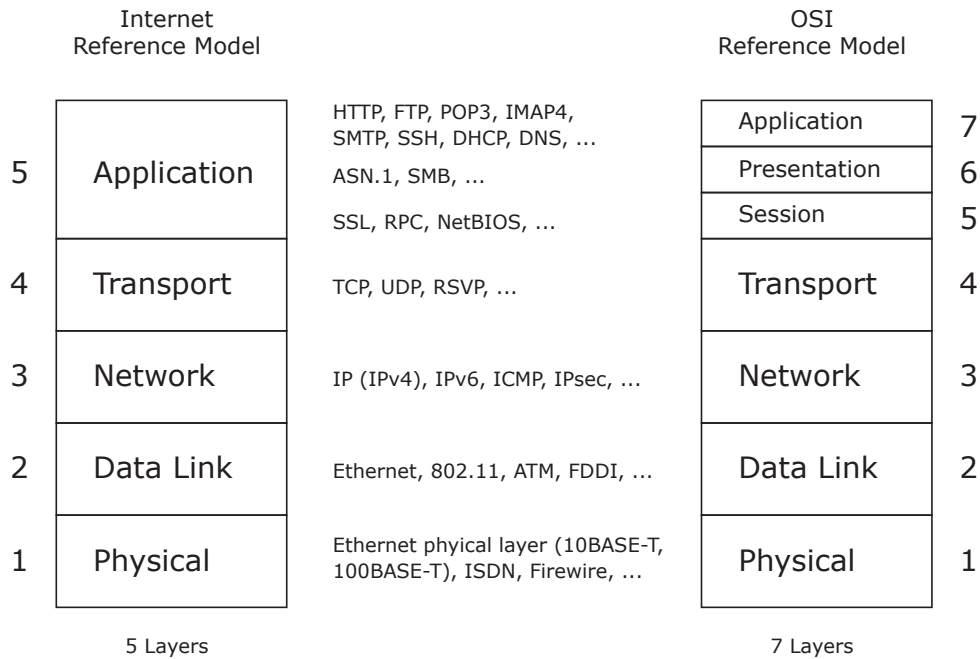


Figure 6.3: On the left, the five-layer Internet Reference Model (or TCP/IP Model). In comparison, on the right, the OSI Reference Models has seven layers. The top three layers of the OSI Reference Model correspond roughly to the top layer of the Internet Reference Model.

6.3 How the Internet Works

The Internet Reference Model (TCP/IP Model)

The Internet Reference Model (or TCP/IP Model) describes five layers of protocols, as shown in Figure 6.3.

- Each higher layer builds on the functionality of the layer below.
- The Internet Reference Model was developed before the similar OSI Reference Model (the OSI Reference Model defined seven layers, but is no longer used in practice).

For more details see Kozierok [2005], Stevens [1993], Matthews [2005], or Wikipedia [2007d].

Internet Protocol (IP)

- IP defines the addressing and routing mechanism for the internet.
- The current version is IPv4. It will be replaced by IPv6 over the next few years (IPv6 and IPv4 are already running in parallel in many places).
- Data is put into small packets (like envelopes) of upto 1500 bytes.
- The internet is a so-called *packet switching* network, analogous to the postal service rather than telephone network. See Figure 6.5.
- Packets are marked with source and destination addresses and sent on their way.

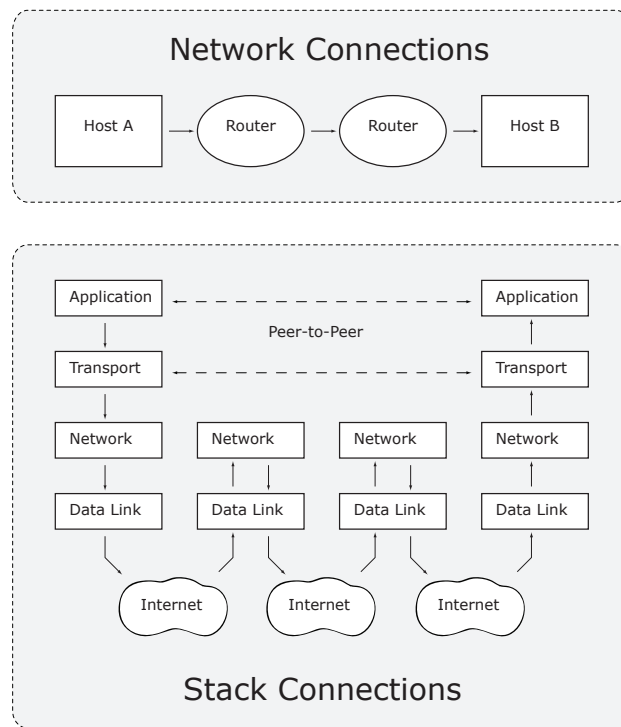


Figure 6.4: IP suite stack showing the connection of two hosts via two routers and the corresponding layers used at each hop. Routers use the network layer in order to route packets based on the IP address. Hosts use all layers. [Adapted from [Burnett, 2007]]

- They are not guaranteed to arrive in order, or even to arrive at all!

IP Addresses

- In IPv4, addresses are unique 32-bit numbers, usually written as four 8-bit parts, e.g. 129.27.153.10
- The first one, two, or three bytes indicate the network, the remaining bits the host on that network. See Figure 6.6.
- You obtain an IP address (or block of addresses) from your Internet Service Provider (ISP).
- The ISP obtains blocks of addresses from their upstream registry or their appropriate regional registry.

IPv4 Address Scarcity

- IPv4 addresses are 32 bits (4 bytes).
- There are (potentially) $2^{32} = 4,294,967,296 \approx 4$ billion IPv4 addresses.
- On 03 Feb 2011, IANA assigned its remaining free blocks of IPv4 addresses to the five RIRs (regional internet registries). [NRO, 2011; Mahloujian, 2011]

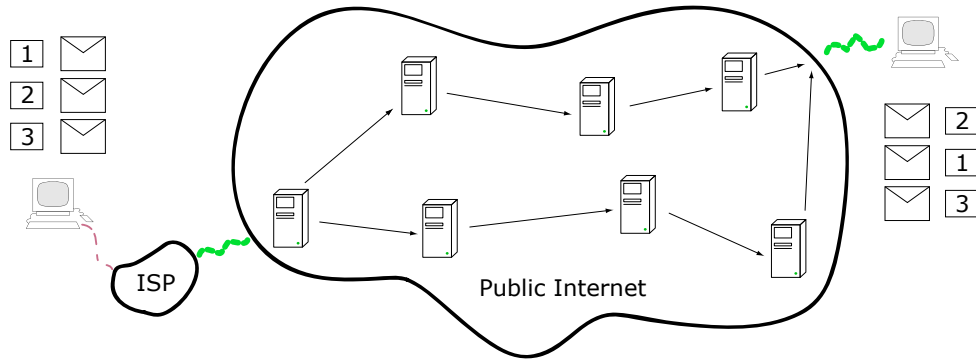


Figure 6.5: Packet switching.

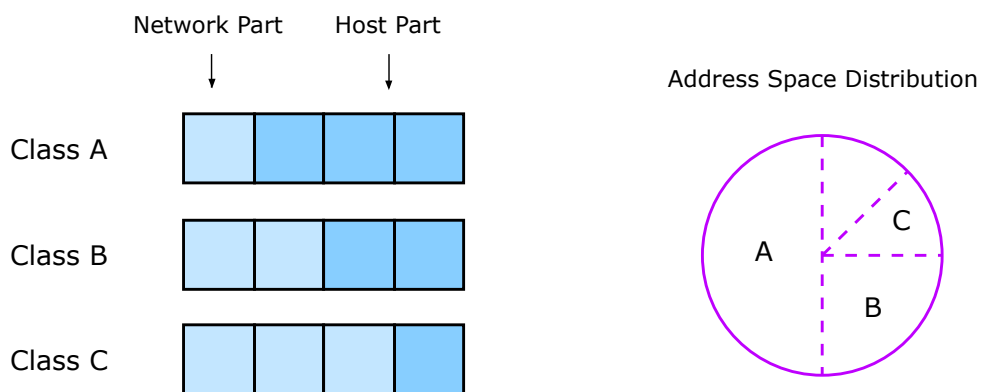


Figure 6.6: The IPv4 address space. IPv4 addresses comprise 4 bytes of 8 bits (32 bits total). The first one, two, or three bytes identity the network.

- The individual RIRs and ISPs have some reserves of IPv4 addresses, but it is only a matter of time before ISPs start denying requests for IPv4 addresses.

See http://en.wikipedia.org/wiki/IPv4_address_exhaustion.

IPv6 Address Space

- IPv6 addresses are 128 bits (16 bytes).
- They comprise 8 fields of 16-bits, each represented by 4 hex digits, for example:

2001:0000:3247:0000:0000:D11E:CDAB:0362

- Typically, the first 64 bits identify the network, and the second 64 bits the device (interface id).
- There are (potentially) $2^{128} = 340,282,366,920,938,463,463,374,607,431,768,211,456 \approx 3.4 \times 10^{38} \approx 340$ trillion, trillion, trillion IPv6 addresses.
- So they should not run out anytime soon...

See <http://en.wikipedia.org/wiki/IPv6>.

Good tutorials include 9tut [2011]; Broersma [2010].

Transition to IPv6

IPv4 and IPv6 will probably run in parallel for many years:

- Dual Stack: networks and computers operate both IPv4 and IPv6 at the same time.
- Tunneling: IPv6 packet is encapsulated in IPv4 header.
- Translation: local IPv6 hosts can access IPv4 servers through router translation (NAT64 and DNS64).

But at some point, new entrants will only be able to get hold of IPv6 addresses.

<http://google.com/ipv6/statistics.html> shows current adoption rates by country.

For further information, see Phillips [2013]; Google [2013]; Deloitte [2013]

Transmission Control Protocol (TCP)

- Large information is broken into pieces with sequence numbers.
- TCP packets are placed in IP packets for transmission.
- Packets are collected, placed in order, and the original data extracted.
- Missing and corrupted packets are re-transmitted.

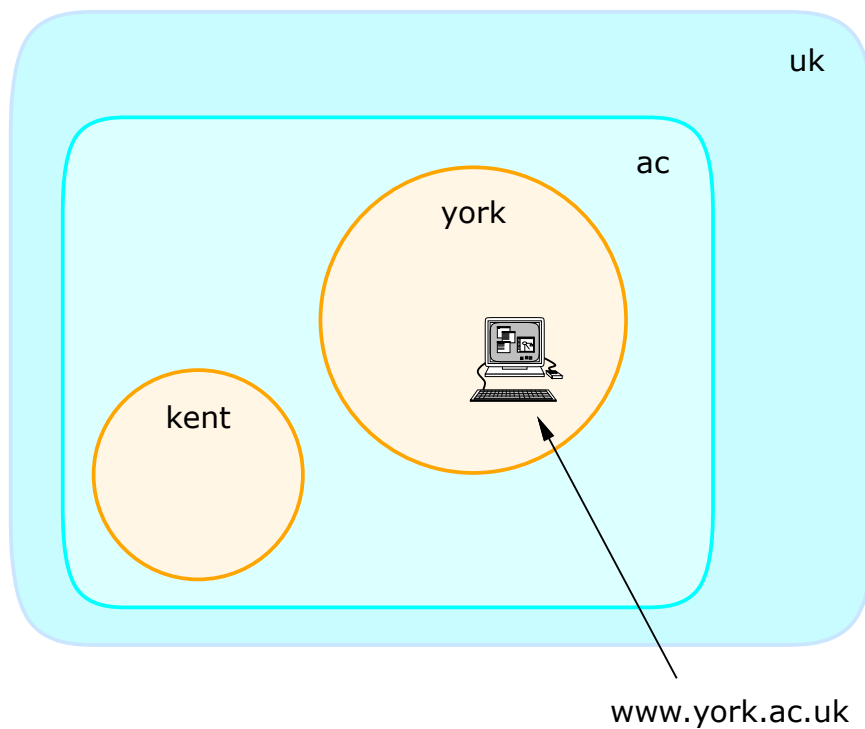


Figure 6.7: The domain name system.

Domain Name System (DNS)

- Names are easier to remember than 32-bit numbers.
- Name server mechanism translates a name into an IP address.
- Official name plus any number of aliases. See Table 6.3.
- Domain hierarchy, responsibility at each level (typically 3–5 levels). See Figure 6.7.
- Originally 6 top-level domains (TLDs) in US (com, edu, gov, mil, org, and net).
- Then >300 country codes (e.g. at for Austria)
- Now there are numerous new TLDs such as biz, info, name, pro, aero, eu, asia, ...

See http://en.wikipedia.org/wiki/List_of_Internet_top-level_domains

Domain Name Registration

- ICANN (Internet Corporation for Assigned Names and Numbers) <http://www.icann.org/> oversees policy and top-level domains.
- The IANA Root Zone Database contains a list of all top-level domains: <http://iana.org/domains/root/db>
- Accredited registrars carry out the registration (for example, a .com domain costs € 9,90 per year through sprit.org).

<i>Address</i>	<i>Name</i>	<i>Aliases</i>
192.48.153.1	sgi.com	
18.26.0.36	lcs.mit.edu	
36.190.0.136	www.stanford.edu	
207.237.99.100	www.aiesec.org	
129.27.3.20	www.tugraz.at	
129.27.3.20	www.tu-graz.ac.at	
144.32.128.78	www.york.ac.uk	web-lb
129.27.153.17	www.iicm.edu	info hyperg
129.27.153.17	fiicmhp01.tu-graz.ac.at	mars

Table 6.3: Some IP addresses and their corresponding domain names.

- Country code top-level domains (ccTLDs) are maintained by each country.
- .at domains are managed by `nic.at`.
- .at domains can be registered at `nic.at` (€ 72 for the first year, € 36 thereafter) or various other accredited registrars such as `sprit.org` (€ 24.90 per year).
- A web hosting package often includes a domain name.
- You can check who, if anyone, owns a domain name using a service such as `domaintools.com` or `http://nic.at/domainsuche/` (for .at domains).
- ICANN started accepting applications for generic top-level domains (gTLDs) in Jan 2012. A company or organisation can apply for its own TLD, say `.ibm` or `.tesco`, for an initial cost of around US\$ 185,000 (and possibly much more, depending on disputes and challenges). [Schroeder, 2011]
- Over new 400 gTLDs have been delegated as of Oct 201415.
- For example, `sprit.org` offers `.wien` domains for € 36 per year, `.company` domains for € 90 per year, etc.

Domain Name Lookup

The local name server:

1. Looks up local names in local database.
2. Looks in cache for recently used names.
3. Asks name server of *top-level* domain for address of server of next lower domain, and so forth, until desired address is found.

Common Misunderstandings About DNS

- Parts of a domain name do *not* correspond directly to parts of an IP address: `www.tu-graz.ac.at` and `iicm.tu-graz.ac.at` may be on different networks!



Figure 6.8: The site www.mercedes.at used to be a second-hand lorry dealer! Screenshot taken in Nov. 1998.

- Top-level domain does not necessarily tell you where a machine is located: av1.com is in Austria not the USA!
- An IP address may have several names: www.tugraz.at, news.tugraz.at, www.tu-graz.ac.at, and fstgss00.tu-graz.ac.at all map to the IP address 129.27.3.20.
- Names are not necessary for communication, only IP addresses: <http://194.232.106.11/> will work fine (even if DNS is broken).
- Remember names rather than IP addresses. When a machine or service is moved, it will probably keep its name, but have a different address.

Grab Your Domain Name!

- All 2, 3, and 4 digit .com domains are already assigned, as well as pretty much all proper words/nouns [DeletedDomains.com, 2002].
- Domain name hijacking: mcdonalds.com was grabbed in 1994 by a Wired Magazine journalist as a joke! [Quittner, 1994]
- mercedes.at used to be a second-hand lorry dealer! (see Figures 6.8 and 6.9)
- Disputes are resolved through the courts, depending on the trademark laws of individual countries.



Figure 6.9: It is now what you would expect. Screenshot taken in Nov. 2001.

Dynamic Host Configuration Protocol (DHCP)

Protocol for dynamic allocation of (re-usable) IP addresses:

- “Plug-and-play” Internet. A temporary IP address and configuration parameters are obtained from a DHCP server.
- Allows organisations to share a limited pool of IP addresses among a group of clients not needing permanent IP addresses.
- Dynamic IP addresses are fine for browsing, but not suitable for example for running a web server.
- DHCP servers can also be configured to allocate static IP addresses.
- RFC 2131 [Droms, 1997]

MAC Addresses

MAC address . . . media access control address

- A unique identifier for a network access device such as a LAN card or WLAN card.
- 48-bit (6-byte) address space maintained by IEEE.
- The first three bytes identify the organisation (usually manufacturer) which issued the MAC address.

- A computer might have multiple network devices, each with its own MAC address:

```
WLAN 00-13-02-8F-62-47
```

```
LAN 00-13-A9-3D-42-99
```

- Under Windows open a command shell and type:

```
ipconfig /all
```

to list available network devices and their MAC addresses.

See http://en.wikipedia.org/wiki/MAC_address for more details.

Chapter 7

Staying Safe

“ *The only thing the NSA respects is the power of mathematics (cryptography).* ”

[Jacob Appelbaum, speaking at Elevate 2013 in Graz.]

```
Subject: ILOVEYOU
-- body --
kindly check the attached LOVELETTER coming from me.
-- attachment --
LOVE-LETTER-FOR-YOU.TXT.vbs
```

[The ILOVEYOU worm, May 2000.]

Resources

- + *Malware*; <http://en.wikipedia.org/wiki/Malware>
- ++ *The Intercept*; <https://firstlook.org/theintercept/>
- + Lorenzo Franceschi-Bicchierai; *The 10 Biggest Revelations From Edward Snowden's Leaks*; Mashable, 05 Jun 2014 <http://mashable.com/2014/06/05/edward-snowden-revelations/>
- + Ted Gioia; *162 Things We've Learned About the NSA*; 07 Aug 2014 http://tedgioia.com/nsa_facts.html
- ++ Bruce Schneier; *Keeping Track of All the Snowden Documents*; https://schneier.com/blog/archives/2013/12/keeping_track_o.html
- + Wikipedia; *Global Surveillance Disclosures (2013–present)*; https://en.wikipedia.org/wiki/Global_surveillance_disclosures_%282013%E2%80%93present%29
- + Chaos Computer Club; *How To Survive*; https://events.ccc.de/congress/2013/wiki/Static:How_To_Survive
- + Bruce Schneier; *NSA Surveillance: A Guide to Staying Secure*; The Guardian, 06 Sept 2013 <http://www.theguardian.com/world/2013/sep/05/nsa-how-to-remain-secure-surveillance>
- + *Get Safe Online*; <http://www.getsafeonline.org/>

- *Spybot Search and Destroy*; <http://spybot.info/>
- *FILEExt - The File Extension Source*; filext.com
- Symantec; *Symantec Global Internet Security Threat Report: Trends for 2009*; Volume XV, Apr 2010. http://eval.symantec.com/mktginfo/enterprise/white_papers/b-whitepaper_internet_security_threat_report_xv_04-2010.en-us.pdf

Resources in German

- + *IT-Security*; <http://www.zid.tugraz.at/security/>
- + *Den Computer halbwegs sicher betreiben*; <http://karl-voit.at/2014/03/23/Computer-halbwegs-sicher/>
- + *CryptoParty Graz*; <https://cryptoparty.at/graz>
- + *Newsgroup tu-graz.security*; tu-graz.security
- + *Watchlist Internet*; watchlist-internet.at
- *Safer Internet*; saferinternet.at
- *Austrian Internet Ombudsmann*; ombudsmann.at

7.1 The Bad Guys

The two main threats to your computer are *malware* and *crackers*.

Malware – Malicious Software

Malware is a piece of software designed to take over and/or damage a computer, without the owner's knowledge or approval.

There are several kinds of malware:

- A *virus* is a program which infects (inserts itself into) other program files. To catch a virus you must run an infected program (or boot from an infected CD-ROM, USB stick, etc.). A virus spreads passively, when the software it is attached to is transferred from one computer to another (by download, email attachment, file sharing, etc.).
- A *worm* actively replicates itself over a network, for example by sending itself to everyone in your email address book or by seeking out unprotected computers attached to the internet. A worm can carry other malware as a load.
- A *trojan horse* is a program hidden inside another program. The host program works normally while the trojan horse does something in the background (such as opening a back door allowing hackers into your system while you are online).
- A *macro virus* spreads by infecting macro-programmable document formats such as Microsoft Word or Excel. The virus is run when the document is opened (or sometimes when a link within the document is clicked).

- *Spyware* is software designed to collect and send on information about a user's activity, for example by logging web sites visited, passwords, or keystrokes.
- *Drive-by downloads*: the client computer is compromised merely by visiting the malicious web server with a (vulnerable) web browser.

See <http://cisco.com/web/about/security/intelligence/virus-worm-diffs.html>.

Hackers and Crackers

- Originally, a *hacker* meant a skilled programmer able to hack together lines of code.
- In more recent popular usage, a hacker is someone who breaks into computer systems without permission (sometimes also called a *cracker*).
- The security services routinely hack into computer systems without permission.

7.2 Privacy

Publishing Information

When you write a comment in a forum or a newsgroup, upload a photo to a social network, post a blog entry, etc. you are *publishing* that information *for ever*:

- Even if you delete it a few hours later, chances are that it has already been archived or someone else has already saved a copy and possibly reposted it elsewhere.
- Employers routinely do research on job applicants by searching Google, Facebook, and co. That drunken photo of you half-naked at a party may not be good for your career. . . [Finder, 2006]
- If you make, say, libellous, racist, or other illegal comments online, you may well be pursued by the authorities or in a civil case [O'Toole, 2010; APA, 2011].

State Surveillance

State security agencies such as NSA, GCHQ, BND, and HNaA routinely collect, copy, archive, index, and analyse your communications and data:

- NSA (USA) has many programs (PRISM, Fairview, XKeyscore) which intercept and store vast amounts of communications. <http://mashable.com/2014/06/05/edward-snowden-revelations/>
- GCHQ (UK) under TEMPORA program taps into fibre-optic cables and stores huge volumes of data for 30 days (and shares with NSA) <http://theguardian.com/uk/2013/jun/21/gchq-cables-secret-world-communications-nsa>
- Austria, Iraq, Bahamas, Afghanistan, and Somalia are among six NSA "Full Take Countries" (MYSTIC, SOMALGET), where all communications, both metadata and content, are intercepted and stored for at least 30 days [Horcicka, 2014]. <http://pcworld.com/article/2158980/assange-names-country-targeted-by-nsas-mystic-mass-phone-tapping-program.html>

Watch these Guarnieri and Marquis-Boire [2013]; Appelbaum [2013] and other videos from the 30C3 conference to be blown away.

Facebook

Facebook is not private:

- When you sign up for Facebook, you grant Facebook worldwide rights to use, publish, and even sell anything you do, say, or upload on Facebook, forever [Consumerist, 2009].
- Private chats on Facebook are saved and indexed. Facebook routinely passes such material on to the authorities upon request [APA, 2011].

Skype

- Since Microsoft acquired Skype, you can safely assume that Skype chat, audio, and video content is being saved, analysed, and passed to the NSA [Heise, 2013; Anderson et al., 2013].
- The content of Skype audio and video streams is still hard to intercept without knowledge of the encryption keys [Economist, 2008].

7.3 Prevention is Better than Cure

Some tips to help ensure that your data and privacy remain intact. Note that there is no such thing as 100% security and never will be.

Choose Secure Software

The best way to avoid malware is not to use Microsoft software:

- Microsoft Windows is notorious for its vulnerabilities and is the main target for hackers and malware writers largely because it is the predominant operating system.
- Real computer scientists use Linux or other open-source Unix variants. These also have vulnerabilities, but since the source code is freely available, they can be spotted and fixed more easily.
- Mac users face significantly less threat than Windows users, largely because they are a small minority and thus make a less interesting target.
- Firefox and Thunderbird are inherently safer than Internet Explorer, Outlook, and Outlook Express, because they are open-source and are less of a target.

Avoid Running Dodgy Software

- Only run programs which come directly from a trustworthy source (distribution CDROM or publisher's web site).

.bas	BASIC program
.bat	DOS batch file script (program)
.com	Command file (program)
.exe	Executable file (program)
.inf	package information file
.js	Javascript file
.jse	Fichier encoded Javascript
.pif	Program information file (Win 3.1)
.reg	Registration file
.scr	Screen saver
.shs	Shell scrap file
.vbe	Visual Basic related
.vbs	Visual Basic program
.wsf	Windows script file
.wsh	Windows script host settings file

Table 7.1: These types of file are considered high-risk under Microsoft Windows, because they can contain pieces of executable code.

- Never accept to run an office (Word, Excel, ...) document macro, unless the creator assures you that it is there for a good reason.
- If you are surfing the web and are offered an ActiveX Control, do *not* accept it, unless you know and trust the company offering it.
- *Never* open a file or attachment sent by email, newsgroups, icq, chat, etc.
- ... especially if it has one of the extensions in Table 7.1. [Note that you need to explicitly turn on display of file extensions in Windows.] You can check the meaning of common file extensions at [FILEExt, 2005].
- ... even if it appears to come from your best friend! [Email senders can easily be doctored]
- Instead, save it to disk, check that the icon is the correct one for the type of file it appears to be, and scan it with a virus scanner.
- A common trick is to insert dozens of spaces between a fake extension and the real extension, for example:

myphoto.jpg

.pif

- The types of file in Table 7.2 are considered lower risk, because they do not generally contain executable code. However, vulnerabilities have also been reported in some of these file types, including PDF (CVE-2013-0640 , CVE-2013-0641 , CVE-2005-2470 , and CVE-2007-5020) [Naraine, 2006] and JPEG (CVE-2005-1988 and CVE-2006-3198) [BBC News, 2004], and also in PowerPoint (CVE-2006-0009 and CVE-2006-0022).
- Note that an infected file does not have to arrive as an email attachment. You might have grabbed it from a web site, a peer to peer network (such as Kazaa), or copied it from a USB stick or CD-ROM.

.txt	Plain text
.pdf	Portable document format
.jpg	JPEG image
.gif	Graphic interchange format
.bmp	Bitmap graphics
.tif	Tagged image format
.png	Portable network graphic
.mp3	MPEG audio stream, layer 3
.wav	Waveform audio
.mpg	MPEG video
.avi	Audio video interleave
.mov	QuickTime movie
.wmv	Windows media file

Table 7.2: These types of file are considered lower-risk under Microsoft Windows, because they generally do not contain pieces of executable code. However, vulnerabilities have also been reported in some of these file types, including JPEG and PDF.

Turn On File Extensions

Under Windows (XP, Vista, 7) file extensions are *hidden* by default.

- You must explicitly turn on the display of file extensions [Stockdale, 2007].
- Otherwise, a file you think is called `myphoto.jpg` might just be called `myphoto.jpg.exe`.
- Or a file which you see (with extensions hidden) as `myphoto` and assume is `myphoto.jpg` might be actually called `myphoto.JPG`. [Upper and lower case matters, for example, when you upload your local web pages to a web server running under Unix.]

Turn On Hidden Files

By default Windows (XP, Vista, 7) hides certain files from being seen in Windows Explorer.

- This is supposedly done to protect system files and the like from being accidentally modified or deleted by the user.
- Unfortunately, viruses and spyware often hide files this way.
- You should explicitly turn on the display of hidden files [Abrams, 2007].

Hide Behind a Firewall

- A firewall protects from worms and hackers probing for unsecured ports when your computer is connected to the internet.
- Never connect an unprotected (Windows) computer to the internet! It can be attacked by worms within as little as 10 seconds (I have experienced this myself on Chello). See also [Ward, 2006]

- The best is a standalone firewall, i.e. a piece of hardware acting as a barrier between you and the internet. Many router or wireless routers have a built-in firewall and NAT (Network Address Translation). Some people use an old PC running Linux to act as a firewall.

Install a Personal Firewall (Software Firewall)

You should probably install a personal firewall, particularly if you have a laptop, which moves around and is connected to numerous external networks at one time or another.

- A personal firewall is a piece of software installed on your PC, which monitors and controls communications to and from your PC.
- A personal firewall usually operates according to a security policy and displays an “alert” when an incoming connection request is received, or when a program requests to open an outgoing connection.
- Install a personal firewall onto your computer from CD-ROM or USB stick *before* you connect it to the internet.
- Good and free personal firewalls include:
 - Comodo <http://personalfirewall.comodo.com/>
 - ZoneAlarm <http://zonelabs.com/>
 - PC Tools Firewall Plus <http://www.pctools.com/firewall/>
 - NetVeda Safety.Net <http://www.netveda.com/consumer/safetynet.htm>
- Windows XP SP2 now has a basic firewall which is activated by default (but ZoneAlarm or Comodo are better).
- Windows Vista has a basic firewall which is activated by default (but by default does not block outgoing connections).
- The Windows 7 firewall is turned on by default and is much more usable.
- Some experts believe personal firewalls give users a false sense of security, because they offer no real additional security over simply turning off unused services and ports [Wikipedia, 2007b; Atkins, 2007; Lugo and Parker, 2005].
- You can check which ports are open on your PC using a free port scanner such as <http://www.heise.de/security/dienste/portscan/>

Use a Virus Scanner

- Install a free virus scanner such as:
 - avast! <http://avast.com/>
 - AVG Free Edition <http://free.grisoft.com/>
 - AntiVir PersonalEdition Classic <http://free-av.de/>
 - Microsoft Security Essentials http://www.microsoft.com/security_essentials/

- If you use AVG with a mail reader, make sure it really is scanning incoming emails: a small window should pop up in the bottom right corner of the display whenever you fetch mail. [If you have had AVG installed for a longer period of time and a more recent install of Thunderbird, you may need to de-install and re-install AVG afresh.]

Use a Spyware Scanner

- Install a free spyware scanner such as:
 - Spybot Search and Destroy [Kolla, 2006]
 - AVG Ewido Anti-Spyware Free [AVG, 2006]

Keep Your Software Up-to-Date

- Install the latest software versions and security patches for your operating system.
- Install the latest updates to your firewall, virus scanner, and web browser.
- Much software now has an auto-update facility, to detect and install updates (semi-)automatically over the internet.

Secure Your Wireless Network

- If you have a wireless LAN at home, make sure to turn on encryption.
- WEP (Wireless Encryption Protocol) encryption is not very secure. [An intruder can crack a WEP password in a few minutes by sniffing sufficient traffic and running a key cracking program [Cheung, 2005].]
- Use WPA2 (Wi-Fi Protected Access) if available. WPA2 and WPA are more secure than WEP.
- If you have an older device on your network, it may only support WEP, but you might be able to download new firmware to upgrade it to WPA or WPA2.
- Choose a good network password and change it on occasion.
- Make sure to change the default admin password on your router (the defaults for various routers are well-known).
- Use an access control list (ACL) in your router to only allow devices with specific MAC addresses to connect to your WLAN.
- Do not use a network name (SSID) which gives away your name or address (such as “AndrewsFlat3”).
- Even better, tell your router to stop broadcasting the network name (SSID) altogether, to make your network invisible to casual would-be intruders.
- Turn off your router when you are not using it (or when you are going away).

Unsecured WLANs, Warchalking, and Wardriving

- Many wireless networks are unknowingly or carelessly left open by their owners. Do not use such networks, it is illegal in many countries.
- *Warchalking* is the drawing of symbols in public places to indicate open wireless networks, inspired by gypsy chalk symbols.
- *Wardriving* involves searching for wireless networks by driving around in a vehicle with a laptop. <http://wgv.at/>

Beware of Phishing Scams

- *Phishing* (password fishing) is a form of social engineering with the goal of acquiring sensitive information, such as passwords and credit card details [Moore and Clayton, 2007].
- For example, you may receive an email purporting to be from your bank asking you to confirm your account details and directing you to an official-looking web site.
- However, on closer inspection, the web site will probably not be hosted by your bank...
- Listing 7.1 shows the source code of the email. Figure 7.1 shows how the email is displayed in Original HTML. Figure 7.2 shows how the email is displayed in Plain Text.
- It is best to access your online banking system using only your own saved bookmark.
- Spoofed domain names can also be constructed to look at first glance surprisingly like the original, for example `googIe.com` and `rnozilla.org`.
- Gabrilovich and Gontmakher [2002] describe using international domain names to construct spoof domains, for example `microsoft.com` with Cyrillic `c` and `o` characters.

Phishing of TU Graz Accounts

- Phishers regularly attempt to scam the login details for TU Graz accounts.
- Students and staff at TU Graz regularly fall for it and hand over their login details.
- Listing 7.2 shows the source code of a typical phishing email aimed at collecting TU Graz account details.
- Never respond to such a mail.
- Check with the list of known phishing attempts at TU Graz: <http://portal.tugraz.at/portal/page/portal/zid/security/informationssicherheit/phishing> whether or not the attack has already been logged.

Beware of Malicious Web Sites

Malicious web sites (either knowingly or unknowingly) host a web page containing malicious code [Moshchuk et al., 2006; Seifert et al., 2007; Seifert, 2007; Wang et al., 2006]:

```

Date: Sun, 13 Feb 2005 00:23:51 +0900
From: "Visa" <visaverified@visa.com>
Reply-To: <visaverified@visa.com>
Message-ID: <147917190.20040830181650@visa.com>
To: "Visa Customer" <kandrews@iicm.edu>
Subject: [ SPAM? ] =?koi8-r?B?VmVyaWZpZWQgQnkgVmlzYQ==?=
-----0388FOCBEB37
Content-Type: text/html; charset=koi8-r
Content-Transfer-Encoding: 8bit

<table width=500>
<tr><td bgcolor="#FFFFFF" height=50>
<img src=http://www.usa.visa.com/img/home/logo_visa.gif>
<b><font face=verdana color=#878787>Verified By Visa</font></b>
</td></tr>
<tr><td>
<p>
<span style="font-size: 12.5px; font-family: Verdana; color:#373737">
<br>
<!--fY-->
<b>Dear Visa customer,</b><br><br>

Before activating your card, read this important information for
cardholders!<br>

You have been sent this invitation because the records of Visa
Corporate indicate you are a current or former Visa card holder. To
ensure your Visa card's security, it is important that you protect
your Visa card online with a personal password. Please take a moment,
and activate for Verified by Visa now.<br>

Verified by Visa protects your existing Visa card with a password you
create, giving you assurance that only can use your Visa card
online.<br>

Simply activate your card and create your personal password. You'll
get added confidence that your Visa card is safe when you shop at
participating online stores.<br>

Activate Now for Verified by Visa, click on the link below:<br><br>
<a href="http://66.232.142.24/personal/security/vbv/index.html">
https://usa.visa.com/personal/security/vbv/index.html</a><br><br>

Sincerly, Visa Inc.
<br><br>
<b>© Copyright 2005, Visa Inc. All rights reserved</b>
</td></tr>

</table>

-----0388FOCBEB37--

```

Listing 7.1: A typical phishing email. The From, Reply-To, and Message-ID fields are all faked. The logo is actually fetched from the official VISA site. The link leads to a dubious destination 66.232.142.24, which was probably a fake clone of the official VISA web site.

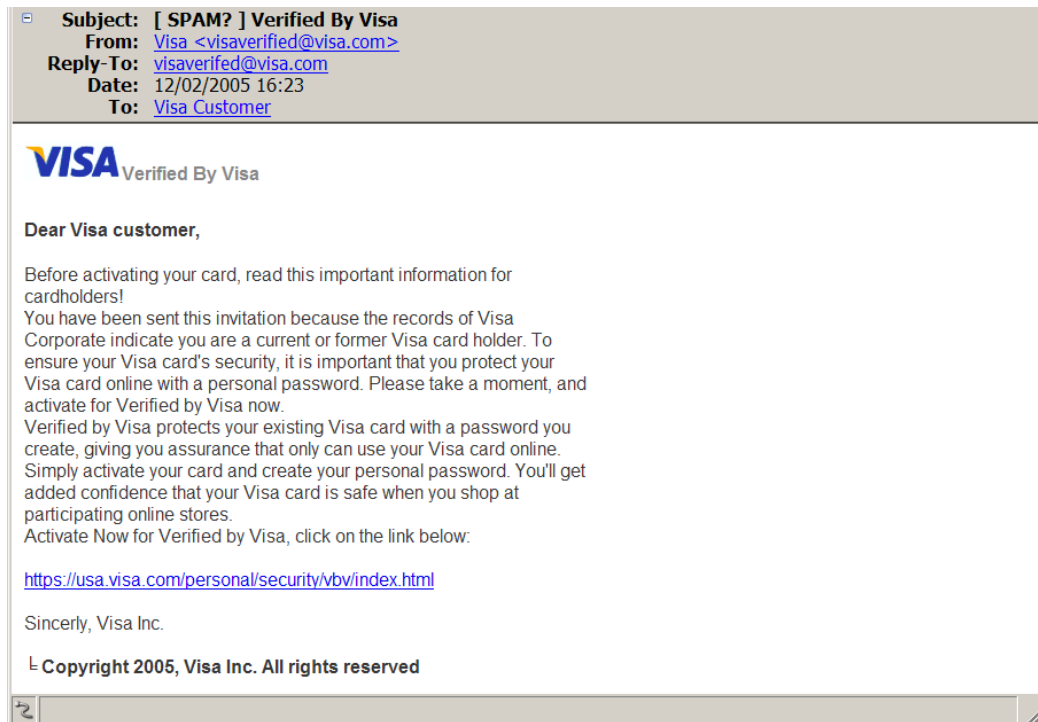


Figure 7.1: Phishing Email Displayed in Thunderbird as Original HTML. The dubious destination of the link is not apparent.

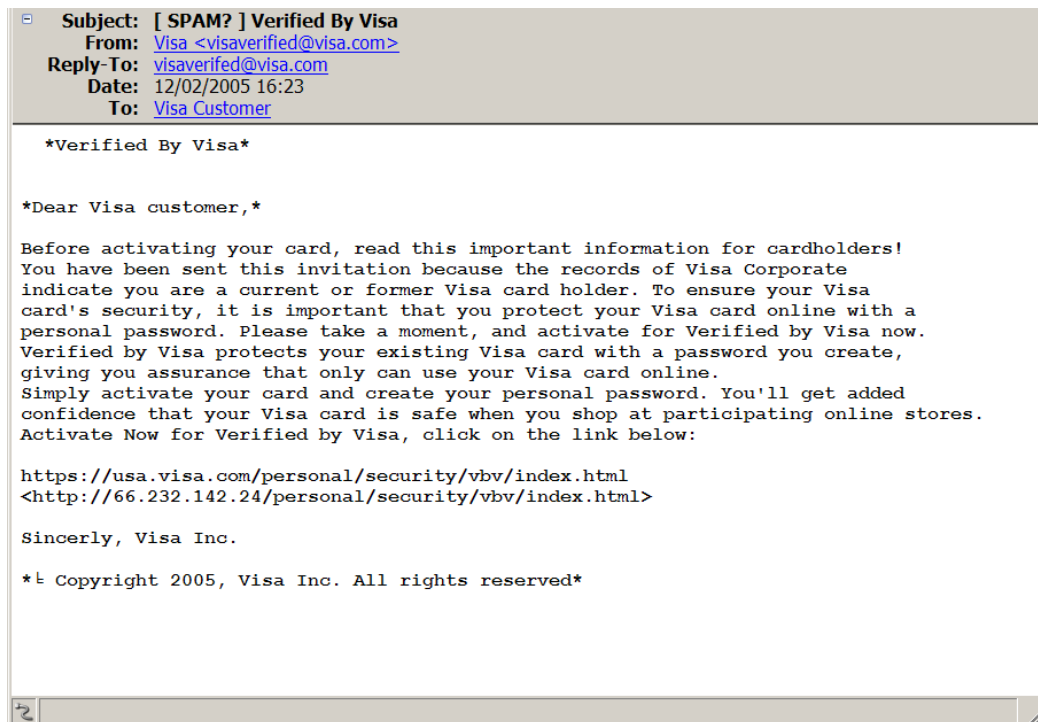


Figure 7.2: Phishing Email Displayed in Thunderbird as Plain Text. The dubious destination of the link is very apparent. One more reason to view your email in plain text format.

```

Date: Thu, 21 Mar 2013 17:17:24 +0700
From: TU-GRAZ Webmail Upgrade Service <info1@tu-graz.ac.at>
To: undisclosed-recipients;;
Subject: Sehr geehrte TU-Graz Mitarbeiter/Student Konto User
Organization: TU-GRAZ Webmail Upgrade Service
Reply-To: <upgradeteam101@live.com>
Mail-Reply-To: <upgradeteam101@live.com>
Message-ID: <8e62a59b3016278a78ba0a304534267f@dof.in.th>
X-Sender: info1@tu-graz.ac.at
X-User-Sender: sf-phichit@dof.in.th
User-Agent: MailGoThai/3.0
--=_ff007e026793daa367c34799765ba2de
Content-Transfer-Encoding: 8bit
Content-Type: text/plain; charset=UTF-8

Sehr geehrte TU-Graz Mitarbeiter/Student Konto User,

Wir sind
derzeit Modernisierung unserer Datenbank und in der Mitte TU-Graz
Webmail-Konten dh Homepage-Ansicht. Wir löschen E-Mail Accounts, die
nicht mehr aktiv, um mehr Platz für neue Konten Benutzer erstellen sind.
Wir haben auch eine systemweite Security-Audit untersucht zu verbessern
und zu unseren aktuellen Sicherheitseinstellungen.

Um weiterhin unsere
Dienste sind Sie verpflichtet, zu aktualisieren und bestätigte erneut
Ihre Kontodaten angeforderte E-Mail unten.

Um Ihr neues Konto
Bestätigung abzuschließen, müssen Sie auf diese E-Mail sofort antworten
und geben Sie Ihre Kontodaten wie gewünscht unten.

User Name:
.....
E-Mail-Benutzername: .....
Kennwort:
.....
Passwort bestätigen: .....
Geburtsdatum:
.....

Ansonsten Ihr Account wird sofort aus unserer Datenbank
deaktiviert und der Service ist nicht so wichtig, Nachrichten
unterbrochen sein sowie durch reduzierte Ihre Kontodaten für uns erneut
bestätigt verloren.

Wir entschuldigen uns für die Unannehmlichkeiten
während dieser Zeit entschuldigen, aber sein Vertrauen, dass wir hier
sind, um Ihnen einen besseren Service und die Bereitstellung von mehr
Technologie, um E-Mail und Internet dreht.

Es ist auch relevant, Sie
verstehen, dass unser Hauptanliegen für unsere TUGraz Mitarbeiter /
Student ist, und für die Sicherheit Ihrer Dateien und
Daten.

https://webmail.tu-graz.ac.at/login.php
https://webmail.tugraz.at/login.php

Bestätigungscode:
-/93-1A388-480 Webmail Technical Support Team.
Copyright (c) 2013. Graz
University of Technology. Alle Rechte vorbehalten.

```

Listing 7.2: A typical phishing email attempt to collect TU Graz account details.

- A *client-side* attack, which targets vulnerabilities in a web browser (or plug-in).
- *Drive-by downloads*: the client computer is compromised merely by visiting the malicious web server with a (vulnerable) web browser.
- In the 2007 HoneyNet study [Seifert et al., 2007], 306 malicious URLs were found among the 302,812 which were tested with an unpatched IE 6 SP2.
- The malicious code attempted to exploit the following vulnerabilities (amongst others): ADODB (BID 10514), WScript.Shell (BID 10652), QuickTime plugin (CVE-2007-0015), Microsoft Web View (CVE-2006-3730 and CVE-2006-4690), and WinZip plugin (CVE-2006-3890 and CVE-2006-5198).
- None of the attacks worked on a fully patched IE 6 (10 May 2007).
- None of the attacks worked on Firefox 1.5 or Opera 8.0.0.
- The best protection seems to be to keep your software (browser and plug-ins) fully patched and up-to-date.

If It Looks Free . . .

One common scam is to offer services on the web which appear at first glance to be free, but are not (subscription trap):

- An online IQ test for which you are then charged € 30 (<http://iqfight.de/>).
- Information on apprenticeships (say at <http://lehrstellen.de/>), which actually costs € 7 per month for 24 months.
- An online route planner (<http://www.routenplanung-heute.com/>), which actually costs € 7 per month for 24 months.
- A download site for free software (<http://www.opendownload.de/>), which actually costs € 8 per month for 24 months.
- The web site <http://skype.at/> used to redirect to <http://www.opendownload.de/> and charge € 192 to download software which elsewhere is free! [Abzocknews.de, 2009] The colour scheme is similar to original Skype and the notification of cost is contained in an unobtrusive image (see Figure 7.3). The real Skype forced them to abandon this and other similar trap sites [Claudius, 2009].
- A family name research site (<http://genealogie.de/>), which actually costs € 60.

If this happens to you, do not panic. Check out the advice and sample letters from consumer organisations [Europakonsument, 2009; Internet Ombudsmann, 2009; Verbraucherrecht.at, 2009].

Further stories at [c't-TV, 2007; Hantke, 2006].

Turn Off AutoRun (AutoPlay)

- AutoRun (or AutoPlay) is a “feature” of Microsoft Windows which allows programs in the autorun.inf file to be started automatically when external media (CD-ROM, DVD, USB stick, FireWire drive) are inserted into or connected to your computer.

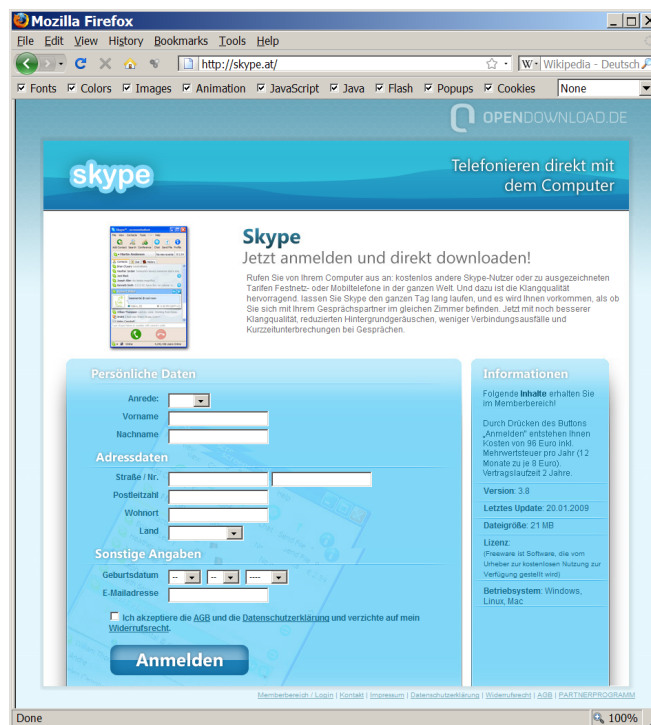


Figure 7.3: The web site <http://skype.at> used to offer otherwise free Skype software for download with a 2-year € 192 subscription (by redirecting to opendownload.de). The web site used similar colours to the real Skype site. The cost information is contained in an unobtrusive graphic. The real Skype forced them to abandon this and other trap sites.

- It allows, for example, music CDs to start playing automatically, and install screens to be displayed for software CD-ROMs.
- Unfortunately, it allows pretty much anything to be started. So, if a bad guy prepares a CD-ROM or USB stick, and you unwittingly insert it into your drive... bang [Garfinkel, 2006].
- The infamous Sony Rootkit [Russinovich, 2005; BBC News, 2005] was also installed as part of an AutoRun script when users inserted certain music CDs into their PC drives.
- Under Windows XP, to disable AutoRun, install the Windows XP PowerToy called Tweak UI [Microsoft, 2005]. Under AutoPlay and Drives, uncheck each drive letter.
- Under Windows Vista, go to Control Panel - Hardware and Sound - AutoPlay and turn off AutoPlay for all media and devices.
- Under Windows 7, go to Control Panel - AutoPlay and turn off AutoPlay for all media and devices.

Good Password Construction Strategies

- Use a three-phase strategy for constructing passwords:
 1. Think of a naming scheme such as types of cheese, brands of beer, or cities in Scotland.

2. Spell the name wierdly: gowda, chedda, brieh
 3. Insert one or two non-alphabetic characters: gowda666, ++chedda, b++rieh
- Your passwords are now fairly easy to remember, but much harder to guess and less susceptible to a dictionary attack.
 - An alternative strategy is to use a phrase such as “My kids names are Susan and Tom”, which you can easily remember, and take the first letters of each word as your password: MknaSaT. Even better, replace “and” with an ampersand: MknaS&T.

Levels of Password Security

You probably have several dozen passwords and PIN codes. Divide them, say, into three levels:

1. One low-level password: for bog-standard web site registration, which you re-use on multiple occasions.
2. Several medium-level passwords: for your mail server, web hoster, and so forth.
3. One or two high-level passwords: for your bank account and credit card, for example.

Safe Shopping

- Only enter your credit card details on a secure webpage.
- The web page address should begin with `https://`
- Your browser should display a closed padlock symbol.
 - The closed padlock actually only indicates the presence of a valid certificate, it could be *any* valid certificate.
 - To be absolutely safe, double-click on the closed lock symbol and view the certificate to check that it belongs to the organisation you expect.
- Never send your credit card or bank details by standard email! Email can be intercepted far too easily. Give the details by fax or phone.

Backing Up

Sometimes, hard disks do crash or may be wiped by malware. It is important to make regular *backups* (redundant, safe copies) of your data.

- Back up anything into which you have put time and effort.
- When you first install a new (Windows) computer, create a separate (D:) partition for your data. The operating system and applications (on C:) can then be re-installed from scratch, without needing to touch the data partition.
- Use an application such as TaskZip <http://www.softpedia.com/get/Compression-tools/TaskZip.shtml> to make daily copies of your web browser’s bookmarks (favourites) and email client’s address book.

- Back up your email archive as well. I use Thunderbird and keep my email in a local directory on my data partition (D:\keith\mail), which is backed up.
- I have several large external hard disks, which I use alternately every week or so, to copy the entire D: partition from my laptop.
- Professional data recovery services (for example <http://www.computerrepairs.at/> or <http://www.eticon.at/>) charge hundreds or thousands of € to retrieve data from a crashed hard disk, with no guarantee of success.

Chapter 8

File Transfer

“ Thomas Watson is famously reported to have said in the Fifties, that the world would only need 5 computers. Actually, there is no evidence he said that, but if he did, he would have been terribly wrong. He would have been off by 4. Because essentially we have a giant, global, computational platform that enables self-organisation. ”

[Don Tapscott, Enterprise 2.0 Conference, 20 Jun 2007 [Tapscott, 2007].]

Online Resources

- Daniel Bernstein; *FTP: File Transfer Protocol*; <http://cr.yp.to/ftp.html> [Bernstein, 2014]
- File Transfer Protocol (FTP); http://en.wikipedia.org/wiki/File_Transfer_Protocol [Wikipedia, 2014a; J. Postel and Reynolds, 1985-10]
- SSH File Transfer Protocol (SFTP); http://en.wikipedia.org/wiki/SSH_file_transfer_protocol [Wikipedia, 2014f; IETF, 2006]
- FTP over TLS (FTPS); [Wikipedia, 2014b; Ford-Hutchinson, 2005]

File Transfer Protocol (FTP)

- FTP is used to transfer files between computers on a network.
- An ftp client connects to an ftp server.
- Dedicated ftp clients handle both download and upload of files.
- Most web browsers support ftp download, some support upload.

Dedicated FTP Clients

For Windows:

- ++ FileZilla; <http://filezilla.sourceforge.net/> (supports the SFTP protocol and HTTP proxies) See Figure 8.1.
- WinSCP; winscp.net (supports SFTP)

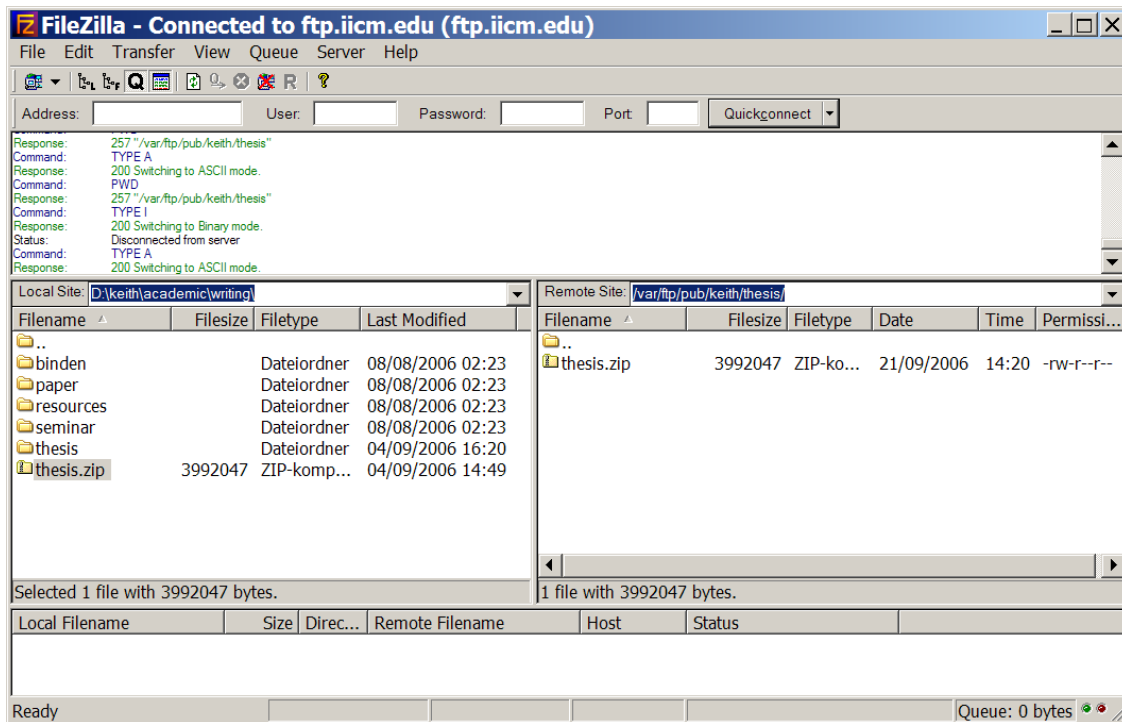


Figure 8.1: An FTP session with FileZilla.

- CoreFTP; coreftp.com (supports SFTP)
- NcFTP; ncftp.com (does not support SFTP)
- and many, many others.

See http://en.wikipedia.org/wiki/List_of_SFTP_clients

Anonymous FTP

- Most public ftp servers allow anonymous ftp: you are allowed to log in without an account and password.
- You have read-only access.
- You can look through the contents of a limited number of directories and download the contents.
- You cannot generally delete, rename, or upload anything.

Uploading Files

- If you have full access (an account and password), you can also rename, delete, and upload files.
- FTP or SFTP are often used to upload a local working copy of a web site to the actual web server.

```

fiicm61:~ 166>ncftp ftp://ftp.iicm.edu/pub/keith/phd/
NcFTP 3.0.2 (October 19, 2000) by Mike Gleason (ncftp@ncftp.com).
Connecting to ftp.iicm.edu via fiicmfw01.tu-graz.ac.at...

fiicm61 FTP proxy (Version V2.1) ready.
Logging in...

-----
|
|           Welcome to the IICM ftp server "ftp.iicm.edu"
|
|           (also known as ftp.iicm.tu-graz.ac.at)
|
|-----

Guest login ok, access restrictions apply.
Logged in to ftp.iicm.edu.

Current remote directory is /pub/keith/phd.
ncftp /pub/keith/phd > dir
drwxrwxr-x  2 1001  406      4096  Okt 23  2000  pdf
drwxrwxr-x  2 1001  406      4096  Okt 23  2000  ps
ncftp /pub/keith/phd > cd pdf
ncftp /pub/keith/phd/pdf > dir
-rw-r--r--  1 1001  406      52933  Nov 11  1997  ch00.pdf
-rw-r--r--  1 1001  406       8738  Nov 11  1997  ch01.pdf
-rw-r--r--  1 1001  406     56540  Nov 11  1997  ch02.pdf
-rw-r--r--  1 1001  406     55428  Nov 11  1997  ch03.pdf
-rw-r--r--  1 1001  406    878463  Nov 11  1997  ch04.pdf
-rw-r--r--  1 1001  406   2034720  Nov 11  1997  ch05.pdf
-rw-r--r--  1 1001  406   3786072  Nov 11  1997  ch06.pdf
-rw-r--r--  1 1001  406    20259  Nov 11  1997  ch07.pdf
-rw-r--r--  1 1001  406     4949  Nov 11  1997  ch08.pdf
-rw-r--r--  1 1001  406    67104  Nov 11  1997  ch99.pdf
ncftp /pub/keith/phd/pdf > get ch00.pdf
ch00.pdf:                               51,69 kB    1,21 MB/s
ncftp /pub/keith/phd/pdf > bye
fiicm61:~ 167>

```

Figure 8.2: FTP allows files to be uploaded to and downloaded from a remote computer. The `dir` command lists files and directories, the `get` command fetches one or more files.

```

ascii      cd          lcd          lrm          pls          rmdir
bgget      chmod       lchmod       lrmdir       put          set
bgput      close       lls          ls           pwd          show
bgstart    debug       lmkdir       mkdir        quit         site
binary     dir          lookup       open         quote        type
bookmark   get          lpage        page         rename       umask
bookmarks  help         lpwd         passive      rhelp        version
cat        jobs        lrename      pdir         rm

```

Figure 8.3: Command-line ftp clients support a number of Unix-like commands.

```
sftp user@pluto.tugraz.at
Connecting to pluto.tugraz.at...
The authenticity of host 'pluto.tugraz.at (129.27.41.4)' can't be established.
RSA key fingerprint is 6e:3e:60:da:e0:0c:93:42:be:48:cb:ff:a5:a4:4b:98.
Are you sure you want to continue connecting (yes/no)? yes
Password:
> cd MS-P
> mkdir inm
> cd inm
> put index.html
> put posting.html
> put research.html
```

Figure 8.4: An SFTP Session. SFTP allows files to be uploaded to and downloaded from a remote computer securely.

SSH File Transfer Protocol (SFTP)

- SFTP is a protocol in its own right, see <http://tools.ietf.org/wg/secsh/draft-ietf-secsh-filexfer/>
- It provides secure file transfer.
- SFTP uses SSH (secure shell) version 2 to provide its security.
- The very first time you connect to a new SFTP server, you will see a message about the authenticity of the server. That is OK, the SFTP client will save the RSA key of the server in its cache for next time. See Figure 8.4.

FTP over TLS (FTPS)

- FTPS secures FTP transmissions using SSL/TLS over the standard FTP ports.
- It is specified in RFC 4217 [Ford-Hutchinson, 2005; Ford-Hutchinson, 2006].
- The client can request that the session be encrypted by sending the "AUTH TLS" command.

Chapter 9

The Web

Q: “If you look back over the last 13 or 14 years, is there any one thing that you regret?”

TBL: “*Yeah, the slash slash...*”

[Tim Berners-Lee, inventor of the web, answering questions after his talk to the Royal Society on 22 Sep 2003
[Berners-Lee, 2003, 01:16:16]]

References

- Berners-Lee et al; *The World-Wide Web*; Communications of the ACM, 1994.[Berners-Lee, Cailliau et al., 1994]

Online Resources

- W3C; w3.org

9.1 The World Wide Web

The World Wide Web (WWW or W3):

- Distributed, heterogeneous, hypermedia information system.
- Initiated by Tim Berners-Lee at CERN, Geneva in 1989.
- Client-server model across Internet.
- Supports hypertext pages and multimedia documents.
- Hierarchies are simulated by lists of hyperlinks.
- Search facilities are provided externally.

Design Decisions

- Platform independence.

- Data format independence.
- Protocol independence.
- Mandatory universal address space.
- Link consistency dropped (for scalability).
- Links transmitted with document.
- Restricted subset of SGML.

W3 Protocols

Three key specifications:

- Universal Resource Identifier (URI)
- Hypertext Transfer Protocol (HTTP)
- Hypertext Markup Language (HTML)

Specifications are available at w3.org.

Universal Resource Identifier (URI)

A URI uniquely identifies a resource:

```

http:// hostport [ / path ] [ ? search ]
gopher:// hostport [ /gopher-path ]
wais:// hostport / database [ ? search ]
ftp:// login / path [ ftptype ]
news: * | group | article
telnet:// [ user [ : password ] @ ] hostport
mailto:user@address
tel:number

```

whereby: *hostport* = *host* [: *port*]

A Universal Resource Locator (URL) is that kind of URI which not only uniquely identifies a resource, but also indicates how to access it (for example, which protocol to use).

See http://en.wikipedia.org/wiki/URI_scheme for a list of official and unofficial URI schemes.

Example URIs

```

http://www.w3.org/hypertext/WWW/TheProject.html
gopher://gopher.ora.com/00/feature_articles/pgp
ftp://ftp.ncsa.uiuc.edu/Web/html/hotmetal/FAQ
news:comp.infosystems.www.browsers.x

```

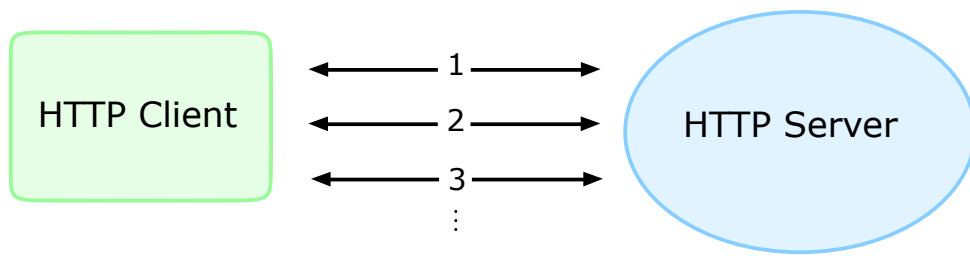



Figure 9.1: The HTTP 1.0 connection model. A new TCP connection is opened for *every* request.

```
telnet://tub@fstgvs05.tu-graz.ac.at
mailto:kandrews@iicm.edu
tel:+43-316-873-5610
```

Hypertext Transfer Protocol (HTTP) 1.0

- ASCII protocol atop TCP/IP; data may be binary.
- Based loosely on MIME Internet mail conventions.
- Stateless transactions:
 - connection** – client connects to server (port 80)
 - request** – client sends request
 - response** – server sends response
 - close** – either party closes connection
- The web server listens for HTTP connections on port 80 by default.
- New TCP connection for *every* request, including each inline image! [Establishing a TCP connection is time-consuming]

Figure 9.1 shows the HTTP 1.0 connection model.

Some HTTP Commands

GET	retrieve document identified by URL
HEAD	retrieve document headers only
DELETE	request server delete particular document
POST	create new object from data part of request
LINK	add link to specified object
UNLINK	remove link from specified object

Example

Client Request:

```
GET /httpd_3.0/Proxies/Proxies.html HTTP/1.0
```

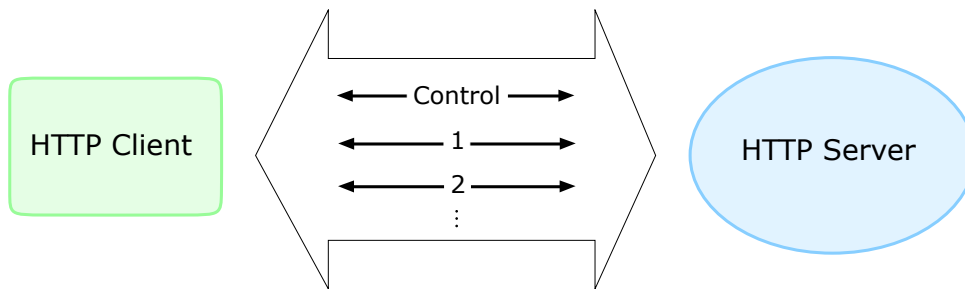


Figure 9.2: The HTTP 1.1 connection model. A *single* TCP connection is divided into multiple virtual channels.

Server Response:

```
HTTP/1.0 200 Document follows
MIME-Version: 1.0
Server: CERN/3.0
Date: Saturday, 19-Nov-94 18:07:47 GMT
Content-Type: text/html
Content-Length: 5799
Last-Modified: Tuesday, 15-Nov-94 19:09:22 GMT

<HTML><HEAD><TITLE>The World Wide Web Initiative:
The Project</TITLE><NEXTID N="z183"></HEAD><BODY>
...
</BODY>
</HTML>
```

HTTP 1.1

- *Single* TCP connection divided into virtual channels.
- Binary ASN.1 encoding.
- Simplified format negotiation.
- Message wrapper with fields for arbitrary security information.

Figure 9.2 shows the HTTP 1.1 connection model.

Viewing HTTP Connections

- Web browser developer tools allow you to examine HTTP requests and responses in some detail. [In Chrome: Tools - Developer Tools - Network. In Firefox: Tools - Web Developer - Network.]
- Figure 9.3 shows that 368 HTTP requests (totalling 2.8 mb) are necessary to download the entire home page of the *Kleine Zeitung* newspaper kleinezeitung.at.
- The Web-Sniffer web service lets you initiate and observe HTTP requests and responses, as shown in Figure 9.4.

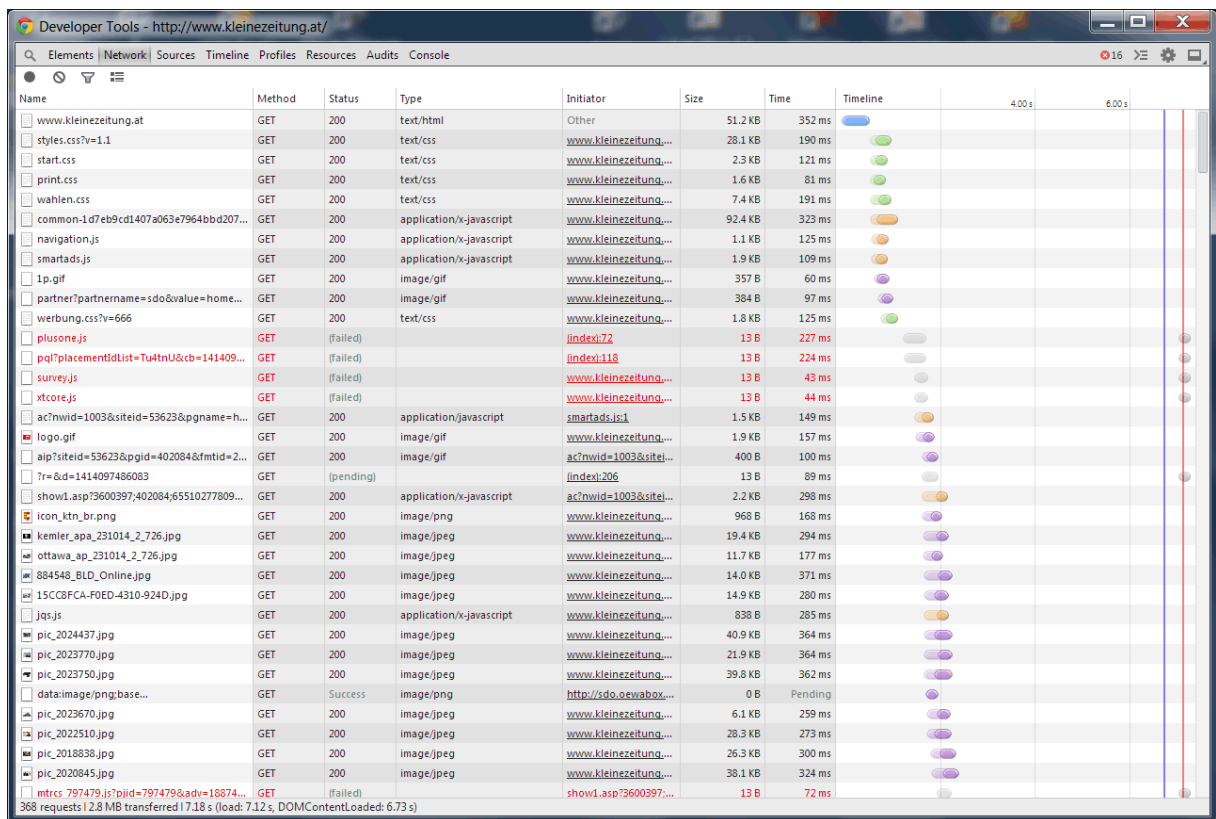


Figure 9.3: The Chrome Developer Tools Network tab showing a timeline of the 368 HTTP requests made when downloading the home page of the Kleine Zeitung newspaper `kleinezeitung.at`.

- The Web-Sniffer application [web-sniffer, 2009] allows you to send HTTP requests directly to a server from outside the browser.
- The Live HTTP Headers extension for Firefox lets you see the HTTP requests issued by your browser in real time livehttpheaders.mozdev.org

HTML (Hypertext Markup Language)

- SGML representation for W3 text documents:
 - marked-up text
 - embedded links
 - inline graphics
- W3C; *HTML Working Group*; <http://w3.org/html/wg/>

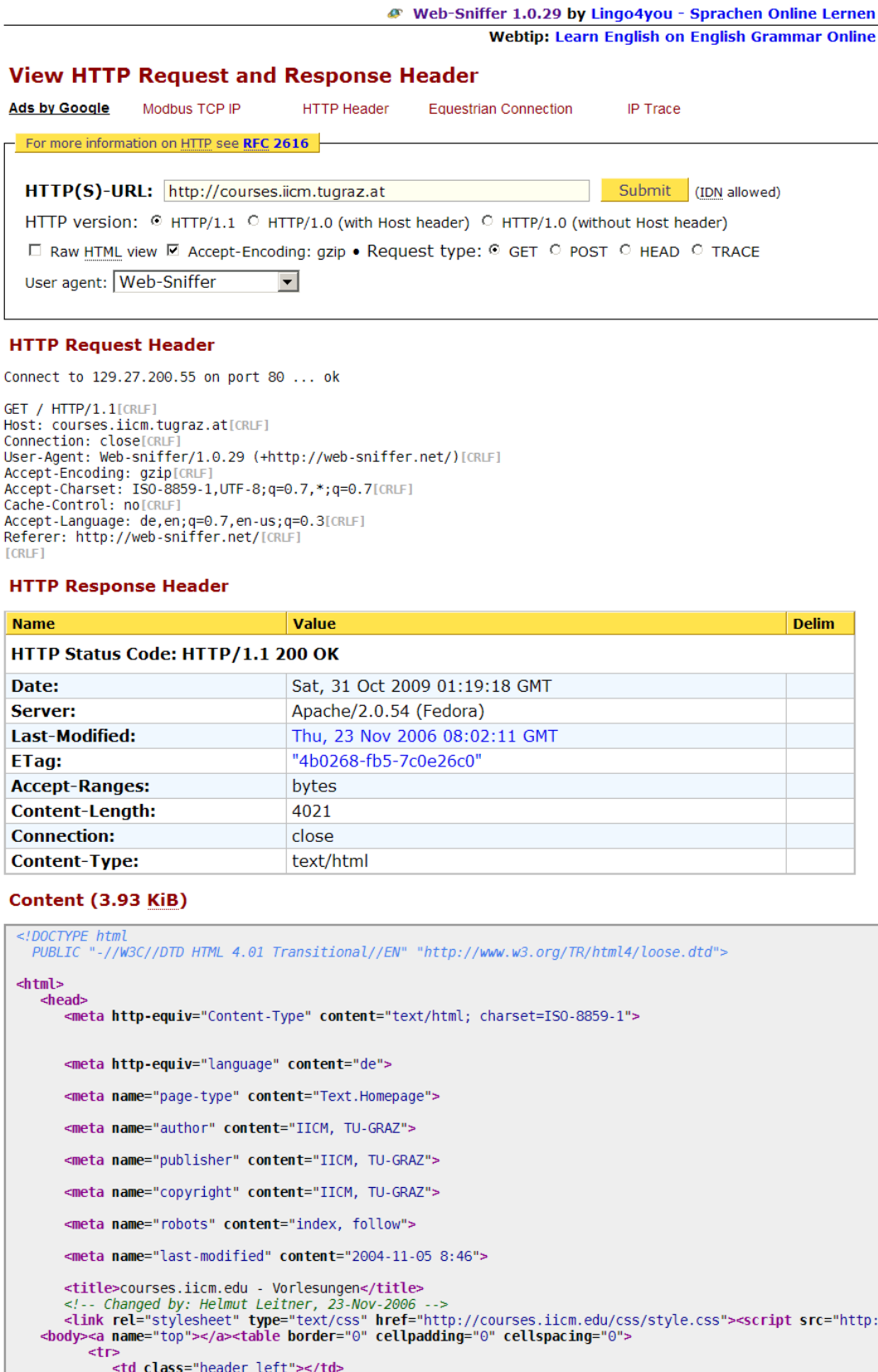


Figure 9.4: HTTP 1.1 Request and Response Headers. A web-sniffer [web-sniffer, 2009] trace of a connection to `http://courses.iicm.tugraz.at`.

```
<HTML><HEAD>
<TITLE>The World Wide Web Initiative: The Project</TITLE>
</HEAD>

<BODY>
<!-- Changed by: , 9-Nov-1994 -->
<!-- Changed by: Arthur Secret, 1-Nov-1994 -->
<!-- Changed by: Arthur Secret, 24-Oct-1994 -->

<H1><IMG ALT="W3Ologo" SRC="Icons/WWW/w3o96x.gif">
World Wide Web Initiative</H1>

<HR>

The WorldWideWeb (W3) is the universe of network-accessible
information, an embodiment of human knowledge. It is an
initiative started at
<A HREF="http://www.cern.ch/">CERN</A>,
now with many participants. It has a body of software, and
a set of protocols and conventions. W3 uses hypertext and
multimedia techniques to make the web easy for anyone to
roam, browse, and contribute to. Future evolution of W3 is
coordinated by the
<A HREF="Organization/Consortium/W3OSignature.html">
W3 Organization</A>.
<P>
Everything there is to know about W3 is linked directly or
indirectly to this document.

...stuff deleted ...

<address><a href=" ../TBL_Disclaimer.html">Tim BL</a></address>
</BODY>
</HTML>
```

Figure 9.5: The HTML 2.0 source code of the W3 project page in 1994.

HTML Features

<i>Level</i>	<i>Year</i>	<i>Features</i>
HTML 0		Anchors, lists, special characters.
HTML 1		Highlighting, inline images.
HTML 2	1995	Forms (user input).
HTML 3.2	1997	Figures (flowing text), tables.
HTML 4.01	1998	Style sheets.
XHTML 1.0	2000	XML reformulation of 4.01, lower case tags, nesting. Transitional or Strict.
XHTML 1.1	2001	XML document, strict checking.
(X)HTML5	2012	<section>, <header>, <footer>, <audio>, <video>, inline SVG, standardised DOM, error handling, JavaScript APIs.

Web Browsers (Clients)

For Windows:

++ Firefox; firefox.com

++ Chrome; <http://google.com/chrome>

+ Opera; opera.com

+ IE; <http://microsoft.com/windows/ie/>

◦ Safari; <http://apple.com/safari/> (Mac and Windows)

◦ Lynx; lynx.browser.org (text-only browser)

See http://en.wikipedia.org/wiki/List_of_web_browsers

Web Servers

- There are about 1 billion web sites (Oct 2014) <http://news.netcraft.com/archives/category/web-server-survey/> [Netcraft, 2011].
- About 37.5% are running Apache, about 33.5% Microsoft server software (Oct 2014) [Netcraft, 2011].
- Kleine Zeitung (kleinezeitung.at) runs Apache on Linux: http://toolbar.netcraft.com/site_report?url=kleinezeitung.at
- Microsoft Austria (microsoft.at) runs Microsoft-IIS/6.0 on Windows Server 2003: http://toolbar.netcraft.com/site_report?url=microsoft.at
- See [ServerWatch, 2011] for comparisons and reviews of different server software.

Web Proxy Servers

- Access to external web servers from inside firewall.

- Document caching (if application-level proxy): proxy keeps local copies (with an expiry date) of often requested pages to speed up access.
- Can provide some protection against drive-by-downloads by blocking access to suspect sites.
- The TU Graz shut down its web proxy server on 31 Dec 2008 due to lack of use.

9.2 Configuring Your Web Browser

Fixed Width Fonts

- *Fixed width* (or monospace) fonts allocate every character exactly the same width, regardless of its shape.

For example, `iii` and `www`.

- Fixed width fonts are used for program source code and listings, where columns and indentation should be preserved.
- Examples of fixed width fonts include Courier and Lucida Console.

Proportional Fonts

- *Proportional* fonts allocate less space for narrower characters and more for wider characters.

For example, `iii` and `www`.

- For normal passages of text, proportional fonts consume less space and are easier to read than fixed width fonts.
- Examples of proportional fonts include Georgia (serif) and Verdana (sans serif).

See Figure 9.6.

Serif and Sans Serif Fonts

- A *serif* is a slight projection or embellishment at the end of a letter stroke, as shown in Figure 9.7.
- Examples of serif fonts include Times Roman and Georgia.
- A *sans serif* (French = without serif) font does not have such embellishments.
- Examples of sans serif fonts include Arial, Helvetica, and Verdana.
- When reading passages of text on paper (very high resolution), serif text may be easier to read than sans serif, but the evidence is not conclusive [Schriver, 1997, Chapter 5].

See Figure 9.7

Proportional fonts allocate varying amounts of space to characters depending on their shape. For normal passages of text, they consume less space and are easier to read than fixed width fonts.

Fixed width or monospace fonts allocate exactly the same amount of space to every character regardless of its shape. They are used in cases where it is useful to preserve column alignment and indentation, such as source code listings:

```
c1, 11, 11, 11
c2, AA, AA, AA
c3, FF, FF, FF
```

Proportional fonts allocate varying amounts of space to characters depending on their shape. For normal passages of text, they consume less space and are easier to read than fixed width fonts.

Fixed width or monospace fonts allocate exactly the same amount of space to every character regardless of its shape. They are used in cases where it is useful to preserve column alignment and indentation, such as source code listings:

```
c1, 11, 11, 11
c2, AA, AA, AA
c3, FF, FF, FF
```

Figure 9.6: Proportional versus fixed width fonts.

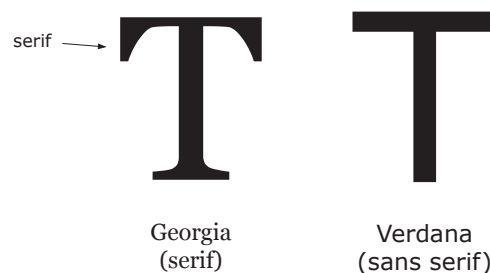


Figure 9.7: Georgia is a serif font, Verdana is a sans serif font.

2 Sixteen Point Fonts Might Be Used for Titles

2.1 Fourteen Point Fonts for Section Headings

Twelve point is great for flowing text such as this¹. Remember that font size changes should be differentiated by at least two points.

¹10 pt might be used for subscripts and footnotes.

Figure 9.8: Font size changes should be differentiated by at least two points.

Font Sizes

Font sizes are traditionally expressed in printers' points (pt):

$$\begin{aligned} 1 \text{ pt} &= \frac{1}{72} \text{ inch} = \approx 0.35 \text{ mm} \\ 1 \text{ pica} &= \frac{1}{6} \text{ inch} = \approx 4.20 \text{ mm} = 12 \text{ pt} \end{aligned}$$

- 10 pt is legible, 11 pt or 12 pt is better.
- The distinction between font sizes should be at least 2 pt. [smaller changes cannot be discriminated by the eye]

Example Font Sizes

For example, use:

- 12 pt for flowing text
- 10 pt for subscripts and footnotes
- 14 pt for section headings
- 16 pt or larger for titles

as shown in Figure 9.8.

Fonts for Online Text

Research on reading online text [Bernard et al., 2002] suggests that:

- Times and Arial are read faster than Courier, Schoolbook, and Georgia.
- Verdana is the most preferred font, is perceived as being legible, and is read fairly quickly.
- Fonts at 12 pt size are read faster than fonts at 10 pt size.

Many original studies are available at SURL [2006].

Setting Your Web Browser Fonts

- In most web browsers, you can set default fonts to be used.
- You can also choose to override the fonts suggested by web pages.
- As a rule of thumb, set a proportional font such as Times or Verdana at 12 pt or higher as the standard font.
- The No Squint extension for Firefox controls text and page zooming on a site-by-site basis. <http://urandom.ca/nosquint/>

Colours

- In most web browsers, you can set default colours for text and hyperlinks.
- You can also choose to override the settings suggested by web pages.
- Black text on a white or almost white background (dark on light) works well.
- You can choose the colours for unvisited and visited links. The defaults are generally blue and purple respectively.
- You can also choose whether links should be underlined.

Plug-Ins

- A *plug-in* (or helper application) is an auxiliary program which works alongside a browser.
- You download and install the plug-in and the browser will call it when it is needed.
- For example, to display some funky new multimedia format.
- In Firefox, enter `about:plugins` into the address bar to see a list of installed plug-ins.
- ActiveX stuff only runs in IE for Windows.

JavaScript

- JavaScript is a programming language interpreted directly by the browser.
- It adds dynamic programmability to HTML pages.
- For example, you can ask what resolution is currently set in the browser window.
- Or write some code to check the answers to web forms.
- JavaScript implementations can differ between browser and browser version!

See [Koch, 2006a; Koch, 2006b] for more information.

Java

- Java is a secure, object-oriented, programming language (similar to C++).
- Compiled to byte-codes which are interpreted.
- Virtual machine ported to several platforms (Windows, Unix, portables, rings, ...).
- Java *applications* are standalone Java programs.
- Java *applets* are distributable across the web and run in a secure area *inside* the web browser.
- Check your browser's Java settings!

See <http://java.sun.com/>

Bookmarks (Favourites)

- A hierarchical folder system for saving and managing URLs.
- Allows you to thematically organise pointers to web sites.
- Saving a bookmark does *not* save the content of a web page.
- If the page changes, you will only see the new version. If the page disappears, you will get an error message.

Browsing History

- The browser's History mechanism keeps a record of which web pages you have seen and when (In Firefox: History - Show All History).
- If you forgot to bookmark a page, you may be able to locate it again by looking through your browsing history.
- The entries can be sorted by date and site.
- For privacy reasons, you may want to set your browser to clear its history upon exiting.
- The History Submenus extension for Firefox collects the history into submenus per day.
<https://addons.mozilla.org/en-US/firefox/addon/682>

Saving and Downloading

- Firefox and IE allow you to save the content of a web page including all of its components (Save As, Web Page complete).
- A Firefox extension will save one or more pages in a special ZIP file (Mozilla Archive Format). <http://maf.mozdev.org/>
- `wget` (wget.org) and other download managers allow you grab entire hierarchies and web sites.

- You can “print” a web page to PDF, by installing a PDF printer such as PDFCreator [pdfforge, 2010] or CutePDF Writer [Acro Software, 2010].
- I create PDFs of interesting web pages I think I may need to refer to later (articles, papers, etc.) and save them in my personal archive.
- The Firefox extension DownThemAll helps you download multiple files simultaneously. downthemall.net

Disable the Browser Cache

The browser cache temporarily stores web pages and other objects (images, etc.) locally on your computer.

- This saves time, if objects do not change very often.
- Freshness and expiry are controlled by HTTP headers and/or the browser’s cache policy.
- If an object changes frequently, you might sometimes be seeing an old version from the browser cache.
- To ensure you always fetch a page (and its components) from the server directly, disable the browser cache:
 - In Firefox: Enter `about:config` into URL bar, search for `cache`, set the variable `browser.cache.check_doc_frequency` to 1 (3 is default).
 - In IE: Tools - Internet Options - General - Browsing History - Settings - Every time I visit the web page.

See http://web-caching.com/mnot_tutorial/ for more details.

Cookies

- A cookie is a piece of text sent by a web server and saved by your browser on your hard disk.
- Figure 9.9 shows a typical cookie sent to my web browser by <http://amazon.com/>.
- Cookies allow a Web site to store information on a user’s machine and later retrieve it.
- For example, to remember who you are next time you visit.
- A cookie is simply a name-value pair (Firefox: Tools - Options - Privacy - Cookies).
- Cookies are not programs! They do not *do* anything.

See [Whalen, 2002; Brain, 2002] for more information.

Saved Form Fields

- Some web browsers allow you to save common entries to web form fields such as name, address, and telephone number.

```
Name:      session-id
Content:   002-0436454-1596866
Domain:    .amazon.com
Path:      /
Send For:  Any type of connection
Expires:   24 October 2006 09:00:09
```

Figure 9.9: One of the cookies sent to my web browser by amazon.com. The cookie's name is session-id, its value is 002-0436454-1596866.

- Very useful and often saves repeated typing.
- In Firefox under Tools - Options - Privacy - Private Data.
- Caution: your credit card details can also be saved! All other users of this computer/browser can potentially see them.

Saved Passwords

- Some web browsers allow you to save passwords.
- Very useful and often saves repeated typing and remembering.
- In Firefox under Tools - Options - Security - Passwords.
- Caution: All other users of this computer/browser can potentially log in to your various accounts!

Firefox Extensions

- Firefox has a whole suite of useful extensions. Click on Tools - Extensions - Get More Extensions.
- Only install extensions from a trusted source.
- My favourite Firefox extensions are:
 - DownThemAll! downthemall.net
 - PrefBar prefbar.mozdev.org
 - Firebug getfirebug.com
 - DownloadHelper downloadhelper.net
 - No Squint <http://urandom.ca/nosquint/>
 - Print Edit <https://addons.mozilla.org/en-US/firefox/addon/print-edit/>
 - Live HTTP Headers livehttpheaders.mozdev.org
 - oldbar <https://addons.mozilla.org/en-US/firefox/addon/6227>
 - ViewSourceWith <http://dafizilla.sourceforge.net/viewsourcewith/>

- Add Bookmark Here 2 <https://addons.mozilla.org/en-US/firefox/addon/add-bookmark-here-2/>
- AdBlockPlus adblockplus.org
- Ghostery ghostery.com
- Total Validator totalvalidator.com
- Page Speed <https://developers.google.com/speed/pagespeed/>
- YSlow <http://developer.yahoo.com/yslow/>

9.3 HTTPS, SSL, and TLS

- Secure Sockets Layer (SSL), invented by Netscape. Now superseded by Transport Layer Security (TLS) [Wikipedia, 2014h].
- Denoted by `https` URLs, by default on port 443.
- Uses encryption to set up a secure channel between web client and web server.
- Very difficult (but not impossible) to crack [Doligez, 1995; Murray, 2001; Lee, Malkin and Nahum, 2007].

SSL Certificates

- An SSL certificate (more correctly X.509 certificate) verifies that the web server is who they say they are [ZyTrax, 2014].
- Your browser and/or OS comes with a pre-filled *certificate store* of “trusted CA certificates”. [Chrome and IE use the operating system’s cert store, Firefox uses its own.]
- If you go to a web site which uses a self-signed certificate, your browser will typically not trust that certificate and will display a security warning, like the one shown in Figure 9.10.

HTTPS Everywhere

Use HTTPS whenever possible:

- You can install the HTTPS Everywhere [EFF, 2014] extension to your web browser (Firefox, Chrome, Opera), to take advantage of HTTPS wherever it is available.
- To make sure you are using HTTPS, look for the closed padlock symbol to the left of the URL in your web browser’s address bar.
- The closed padlock actually only indicates the presence of a valid certificate, it could be *any* valid certificate. [To be absolutely safe, click on the closed padlock symbol and view the certificate to check that it actually does belong to the organisation you expect.]

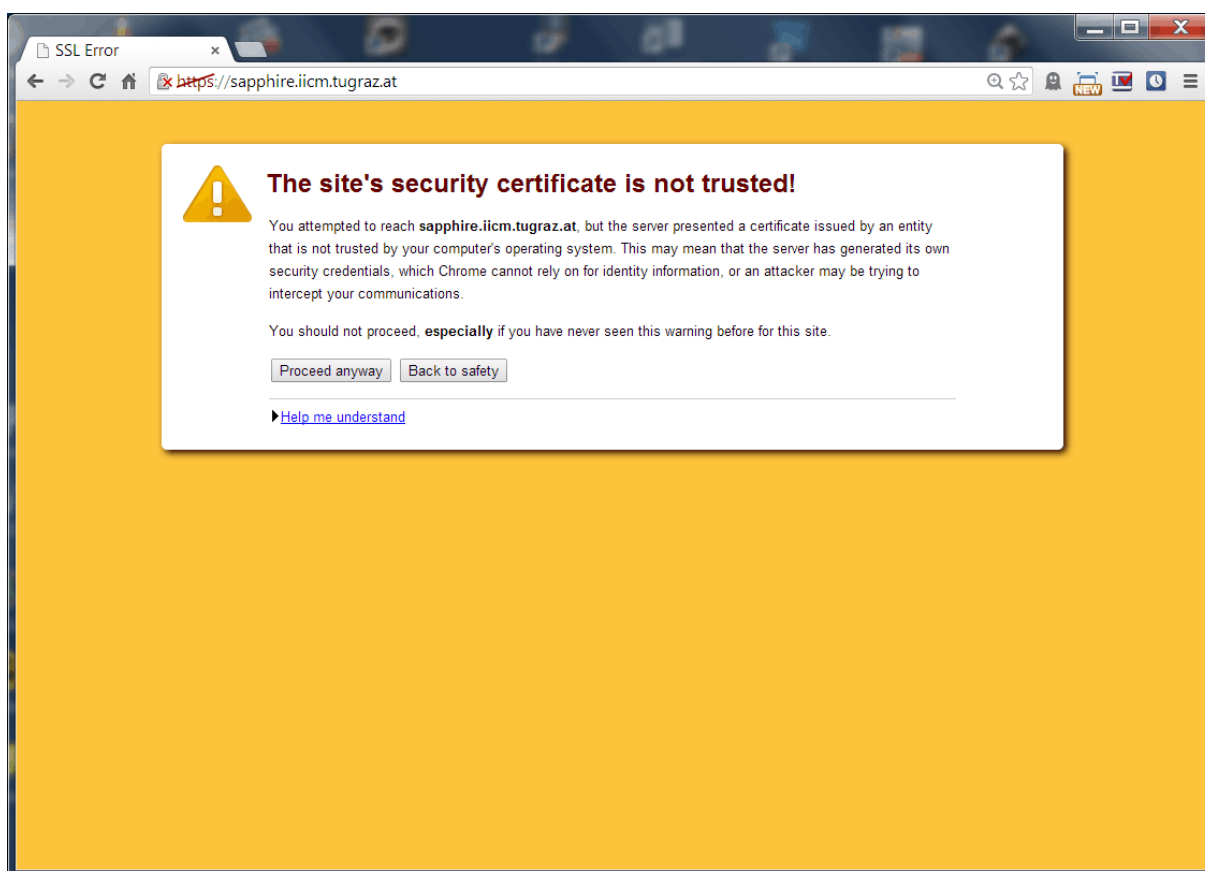


Figure 9.10: The security warning displayed by Chrome when it encounters a HTTPS connection with an untrusted certificate. This is often the case when the web server uses a self-signed certificate.

Chapter 10

Creating Your Own Web Site

“ *Structure now, style later.* ”

[Dan Cederholm, *Web Standards Solutions*, 2004 [Cederholm, 2004, page 19].]

References

- ++ John Duckett; *HTML & CSS: Design and Build Web Sites*; Wiley, 18 Nov 2011. ISBN 1118008189 (com, uk) [Duckett, 2011]
- + Patrick Griffiths; *HTML Dog: The Best-Practice Guide to XHTML and CSS*; New Riders, 22 Nov 2006. ISBN 0321311396 (com, uk) [Griffiths, 2006] {The best introduction to XHTML 1.1 and CSS 2.1.}
- ++ Jeremy Keith; *HTML5 for Web Designers*; A Book Apart, 2010. ISBN 0984442502 (com, uk) [Keith, 2010]
- + Andy Clarke; *Hardboiled Web Design*; Five Simple Steps, 19 Oct 2010. ISBN 1907828001 (com, uk) [Clarke, 2010]
- + Ben Frain; *Responsive Web Design with HTML5 and CSS3*; Packt Publishing, 10 Apr 2012. ISBN 9350237881 (com, uk) [Frain, 2012]
- ++ Dan Cederholm; *CSS3 For Web Designers*; A Book Apart, 2011. ISBN 0984442529 (com, uk) [Cederholm, 2011]
 - Jeffrey Zeldman; *Designing with Web Standards*; New Riders, 3rd Edition, Oct 2009. ISBN 0321616952 (com, uk) [Zeldman and Marcotte, 2009]

Online Resources

- ++ Patrick Griffiths; *HTML Dog*; htmldog.com [Griffiths, 2013] {HTML Dog is a great resource for learning HTML and CSS. The tutorials are HTML5 and CSS3. Unfortunately, the reference sections are still XHTML 1.1 and CSS 2.1. Start here and then move on to Polyglot XHTML5 with CSS3.}
- ++ W3C; *HTML5*; W3C Recommendation, 28 Oct 2014 <http://w3.org/TR/2014/REC-html5-20141028/>
- ++ Accilent; *HTML 5 Reference for Web Designers*; html-5.com

- + Wikipedia; *HTML5*; <http://en.wikipedia.org/wiki/HTML5>
- + Wikipedia; *CSS3*; http://en.wikipedia.org/wiki/CSS3#CSS_3
- + Marcin Wichary; *HTML5 Presentation*; slides.html5rocks.com [Wichary, 2012] {A HTML5 slide set best viewed with Chrome. A great overview of HTML5, JS APIs, and CSS3 features.}
- + Google; *HTML5rocks*; html5rocks.com [Google, 2012]
- + Jakob Jenkov; *HTML5 Tutorial* <http://tutorials.jenkov.com/html5/index.html> [Jenkov, 2012]
 - Duncan Aitken; *HTML5 Tutorial* html-5-tutorial.com [Aitken, 2012]
 - W3C; *WebPlatform.org* webplatform.org [W3C, 2012c]
- + Dave Shea; *Zen Garden: The Beauty of CSS Design*; csszengarden.com [Shea, 2012]
- ++ W3C; *Polyglot Markup: HTML-Compatible XHTML Documents*; <http://www.w3.org/TR/html-polyglot/> [W3C, 2012b]
- ++ Jesper Tverskov; *Benefits of Polyglot XHTML5*; <http://www.xmlplease.com/xhtml/xhtml5polyglot/> [Tverskov, 2012]
 - W3C; *HTML5: A Vocabulary and Associated APIs for HTML and XHTML*; W3C Candidate Recommendation, 29 Oct 2013. <http://www.w3.org/html/wg/drafts/html/CR/> [W3C, 2014c]
 - W3C; *XHTML 1.1 Specification*; <http://www.w3.org/TR/xhtml11/> [W3C, 2007b]
 - W3C; *XHTML 1.0 Specification*; <http://www.w3.org/TR/xhtml1/> [W3C, 2002]
 - W3C; *HTML 4.01 Specification*; <http://www.w3.org/TR/html401/> [W3C, 1999]
 - W3C; *CSS Current Work* <http://www.w3.org/Style/CSS/current-work.en.html> [W3C, 2014a]
 - W3C; *Cascading Style Sheets, CSS 2.1 Specification* <http://www.w3.org/TR/CSS21/> [W3C, 2011a]
 - W3C; *Nu Markup Validation Service* <http://validator.w3.org/nu/> [W3C, 2012a]
 - Henri Sivonen; *Validator.nu* <http://validator.nu/>
 - W3C; *CSS Validation Service* <http://jigsaw.w3.org/css-validator/>
 - W3C; *W3C Link Checker* <http://validator.w3.org/checklink>

Resources in German

- + Stefan Münz and Clemens Gull; *HTML5 Handbuch*; Franzis Verlag, 2nd Edition, 28 Nov 2011 ISBN 3645601511 (com, uk) [Münz and Gull, 2011]
- + Stefan Münz and Clemens Gull; *HTML5 Handbuch*; Online Version <http://webkompetenz.wikidot.com/docs:html-handbuch>
- + Peter Kröner; *HTML5. Webseiten innovativ und zukunftssicher*; Open Source Press, 2nd Edition, 28 Feb 2011 ISBN 3941841343 (com, uk) [Kröner, 2011]
 - Wikipedia; *HTML5*; <http://de.wikipedia.org/wiki/HTML5>
 - Daniel Würstl; *selfhtml5.org*; <http://selfhtml5.org/>

Web Space for Students at TU Graz

- Homepage für Studierende <http://portal.tugraz.at/portal/page/portal/zid/netzwerk/dienste/homepage/studierende>
- BBO des ZIDs der TU Graz <http://portal.tugraz.at/portal/page/portal/zid/richtlinien>
- If your TU Graz account name were `koarl.tester1`, the URL of your web space would likely be: <http://www.student.TUGraz.at/koarl.tester1/>
For example: <http://www.student.tugraz.at/matthias.link/>

10.1 Creating Web Pages

Work Locally First Then Upload

- Work locally at first.
- Keep a local working copy of your entire web site hierarchy.
- Update locally, then upload your files (by FTP or SFTP) to your web site hoster.

Use a Plain Text Editor

- Use a plain text editor such as Notepad, Crimson Editor, Emacs, PSPad, or UltraEdit to write your HTML.
- Microsoft Word is a word processor, not a text editor. Its HTML export produces *horrible* HTML.
- FrontPage and other web editors also often produce horrible, bloated HTML.
- Dreamweaver is excellent, Adobe Edge and Microsoft Expression are reasonably good, but for this course we want you to code your XHTML and CSS by hand in a plain text editor.

Major Previous Versions of HTML

- HTML 1.1 was released as an Internet Engineering Task Force (IETF) Internet Draft in Jun 1993.
- HTML 2.0 was released as an IETF Request for Comments (RFC 1866) in Nov 1995.
- HTML 3.2 became a W3C Recommendation on 14 Jan 1997. It added fonts, tables, applets, superscripts, and other things.
- HTML 4.0 became a W3C Recommendation on 18 Dec 1997. It separated content (HTML) from presentation (style sheets).
- HTML 4.01 Strict does not support old deprecated tags, but HTML 4.01 Transitional still does.
- XHTML 1.0 reformulates HTML 4.01 in XML.

- XHTML 1.1 no longer supports old deprecated tags.
- Work on XHTML 2.0 was effectively stopped on 02 Jul 2009.

HTML5

The latest and greatest version of HTML is HTML5:

- Much larger standard than HTML 4.01 or XHTML 1.1.
- Standardises three new areas:
 - the document object model (DOM)
 - numerous JavaScript APIs
 - browser error handling behaviour.
- Introduces new input elements:
 - range sliders
 - date choosers
 - colour pickers
- Introduces new semantic elements:
 - audio, video
 - header, footer, nav
 - section
 - progress bars

10.2 XHTML

- XHTML: The Extensible HyperText Markup Language.
- A stricter version of HTML, compliant with XML syntax.
- XHTML tags (mark-up) are enclosed between less than ('<') and greater than ('>') characters.
- XHTML encodes the structure and content of a web page, CSS then specifies the presentation.
- General information and attributes go between the head tags.
- The actual content goes between the body tags.

Main Differences Between HTML and XHTML

In XHTML (compared to HTML):

- All tags must be in lower case.
- All documents must have a doctype.
- All documents must be properly formed .
- All tags must be closed.
- All attributes must be in quotation marks.
- All tags must be properly nested.

Serving HTML or XML to the Browser

The way a browser parses a (X)HTML file depends on the *content type* sent with it by the web server:

- `text/html`: the browser assumes HTML and parses loosely with liberal error handling.
- `application/xhtml+xml`: the browser assumes XHTML and parses strictly with an XML parser.

Usually, if your (X)HTML file has the extension:

- `.html`: the server will send it as `text/html`.
- `.xhtml`: the server will send it as `application/xhtml+xml`.

XHTML5

- HTML5 has an XML serialisation (stricter subset) called XHTML5, which complies with the stricter rules for XML.
- For example, in XHTML5, all tags and attributes must be written entirely in lower case. In HTML5 upper case, lower case, or a mixture can be used.
- There are only five named entity references in XML: `&`, `<`, `>`, `'`, and `"`. Do not use any others.
- Although hex character entities (such as `ä` for ä) are allowed in XML, only use them for non-printing characters such as non-breaking space (` `).
- For all other characters, for reasons of readability, use the UTF-8 character itself: Männer instead of `Männer`.
- Every XHTML5 file is also an HTML5 file.

Polyglot XHTML5

- Polyglot XHTML5 [W3C, 2012b; Tverskov, 2012] takes this a step further: the same file can be served as either HTML5 (`text/html`) or XHTML5 (`application/xhtml+xml`) *and works exactly the same in the browser*.
- For this to work, a few extra rules (conventions) are necessary, so that the DOM created inside the browser will be the same in both cases (and thus JavaScript scripts will work the same).

In this course, we will write platform-independent, resolution-independent, polyglot XHTML5.

Extra Rules for Polyglot XHTML5

The following sum up most of the extra rules necessary for Polyglot XHTML5 [W3C, 2012b]:

- Use both the `lang` and `xml:lang` attribute.
- Use `<meta charset="UTF-8"/>`.
- Use `tbody`, `thead`, or `tfoot` in tables (at least one of these must be present).
- When `col` is used in a table, also use `colgroup`.
- Do not use the `noscript` element.
- Do not start `pre` and `textarea` elements with newline.
- Use `innerHTML` instead of `document.write()`.
- In a `script` element, wrap JavaScript in a commented-out CDATA section.
- Many names in SVG and one name in MathML use lowerCamelCase.

Separating Content from Form

Separation of content and form (presentation):

- The XHTML specifies the *content*, a Cascading Style Sheet (CSS) specifies the *form*.
- The same content (an XHTML file) can be *repurposed* using different CSS style sheets:
 - Extra-large, high-contrast display for sight-impaired users.
 - Audio rendering for blind users, or people driving their car.
 - etc.
- Since XHTML is also valid XML [Wikipedia, 2012i], the content can be transformed automatically into different formats using XSLT (Extensible Stylesheet Language Transformations) [Wikipedia, 2012j]:
 - PDF
 - HTML

- plain text
- etc.

Validation Services

Make sure both your XHTML and CSS is *valid* and that your links all work:

- W3C Nu Markup Validation Service: <http://validator.w3.org/nu/> [W3C, 2012a].
To validate XHTML5:
 - Use file upload, make sure your file has the extension `.xhtml`, so that the stricter XML parser is used.

Note that the Nu validator does not check for Polyglot rules, such as the use of at least one of `tbody`, `thead`, or `tfoot` in tables.
- W3C CSS Validation Service: <http://jigsaw.w3.org/css-validator/> [W3C, 2008].
- W3C Link checker: <http://validator.w3.org/checklink>

Browser Validation Extensions

- Total Validator comprises an installable program and an extension for Firefox and Chrome, supporting within-browser validation totalvalidator.com [Total Validator, 2012].
- Total Validator can also check the rules for valid Polyglot XHTML5.

10.3 Building a Polyglot XHTML5 Web Site

A Skeleton Polyglot XHTML5 Web Page

```
<!DOCTYPE html>
<html xmlns="http://www.w3.org/1999/xhtml" lang="en" xml:lang="en">

<head>
<meta charset="UTF-8"/>
<meta name="viewport" content="width=device-width"/>

<title>TU Graz: Keith Andrews: Web Site</title>
</head>

<body>

<h1>Keith Andrews</h1>
<p>
This is Keith's web site.
</p>

</body>
</html>
```

Character Sets

- The character set defines which characters may appear inside the XHTML file.
- Versions of HTML before 4.0 supported only ISO-8859-1 (Latin-1), the standard 8-bit character set for Western European languages.
- ISO-8859-15 is a revamped version of ISO-8859-1, also containing the Euro symbol.
- UTF-8 is the default character encoding for XML. Use UTF-8 for XHTML5 web pages.

Headings and Paragraphs

- The h1, h2, h3, . . . , h6 elements define decreasing levels of headings.
- The p tag encloses paragraphs of text.

Emphasis

- The em element is used for emphasis, which is often (but not necessarily) rendered in italic.

Void Elements

- Most XHTML5 elements are enclosed in both starting and ending tags: they typically have some inner contents:

```
<p>Some text ... in a paragraph.</p>
```

- Some elements (such as br, col, and img) *never* have inner contents. They are called *void* elements.
- Void elements do not have closing tags, they are written as a single atomic tag ending with />. For example:

```
<br/>  

```

Line Breaks

- The br tag is used to start the next line on a new line.
- It is a void element (has no inner content) and is written
.
- *You generally do not need to use line breaks!* The browser will break lines automatically depending on the font size and available space, which generally looks much better than enforced line breaks.

Lists

```
<p>
Three desirable attributes:
</p>
<ul>
<li>cheap</li>
<li>fast</li>
<li>good</li>
</ul>
<p>
Choose any two.
</p>
```

- The `ul` tag is used for unordered (bulleted) lists.
- The `ol` tag is used for ordered (numbered) lists.
- Within a list, the `li` tag defines each list item.

Links

The `a` element links to another resource:

- *Absolute* links: an absolute link gives the entire URL of the destination:

```
Keith lives in <a href="http://graz.at/">Graz</a>, Austria.
```

For links to other web sites, use absolute URLs.

- *Relative* links: a *relative* link gives the relative path to the destination:

```
More detailed <a href="../detail/notes.html">notes</a>.
```

For links to local files on the same web site, use relative URLs.

- The text between the `<a>` and `` tags is known as the *source anchor* text.

URL Encoding

When certain special characters are used within a URL inside a Polyglot XHTML5 file, they have to be encoded specially.

To conform with XML rules:

- Use `&` instead of ampersand.

To conform with URL specifications [Berners-Lee, Fielding and Masinter, 2006; Berners-Lee, Fielding and Masinter, 1998], use percent encoding [Wikipedia, 2012d] for other UTF-8 characters, such as:

- Use `%20` instead of space.
- Use `%25` instead of percent.
- Use `%7E` instead of tilde.

Character	UTF-8 Encoding	Percent Encoding
space	0x20	%20
%	0x25	%25
~	0x7E	%7E
ä	0xC3 0xA4	%C3%A4
ö	0xC3 0xB6	%C3%B6
ü	0xC3 0xBC	%C3%BC
Ä	0xC3 0x84	%C3%84
Ö	0xC3 0x96	%C3%96
Ü	0xC3 0x9C	%C3%9C
ß	0xC3 0x9F	%C3%9F

Table 10.1: Percent encoding for some common UTF-8 characters when they appear inside URLs in Polyglot XHTML5 documents.

- Use %C3%BC instead of ü.

Table 10.1 lists some of the characters which should be percent encoded.

For example, the URLs:

```
<a href="space in name.html">
<a href="http://www.cs.tut.fi/~jkorpela/tilde.html">
<a href="http://google.com/search?hl=en&q=url+encoding">
<a href="http://google.com/search?hl=de&q=Jürgen">
```

should be encoded in XHTML5 as:

```
<a href="space%20in%20name.html">
<a href="http://www.cs.tut.fi/%7Ejkorpela/tilde.html">
<a href="http://google.com/search?hl=en&amp;q=url+encoding">
<a href="http://google.com/search?hl=de&amp;q=J%C3%BCrgen">
```

See <http://en.wikipedia.org/wiki/Percent-encoding>.

Tel URIs

Make telephone numbers clickable (very useful on mobile browsers!) using tel: URIs:

- It works like mailto: is used for email addresses.
- Specify full international numbers, so that they will work from any location.
- Use a hyphen(“-”) as a visual separator, if necessary.
- See RFC 3966 [Schulzrinne, 2004] for more details.

```
<a href="tel:+43-316-873-5610">873-5610</a>
```

Image Formats

Use the appropriate format for the type of image:

- Use JPEG for photographic images.
- Use PNG for icons, drawings, diagrams, screen shots.
- Use SVG for vector graphics.

Note that:

- PNG is *lossless* and has good compression.
- JPEG is *lossy*: it smudges lines and hard edges, so is not suitable for diagrams and screen shots containing text.
- SVG is freely scalable to any size without loss of quality.

Inline Images

Places an image inside an XHTML page:

```

```

```

```

Specify the Width and Height of Images Up Front

The width and height of every image (raster or vector) should be made known to the browser up front (ahead of time):

- If the size of an image is not given up front, the browser has to first fully download and then decode the image, in order to know its size.
- If image sizes are known up front, browsers can lay out the textual content of a page and leave slots for the images to fill in as they arrive.
- Users perceive the site to be faster, since they can begin reading the text straight away (and the text does not slide around as image sizes become known).

This can either be done:

- in XHTML by specifying pixel values in the width and height attributes.
- in CSS using the width and height in any units (including relative units such as % and em) [much better].

Scaling Images in the Browser

- Using a width and/or height different from the actual dimensions of the image causes the browser to scale the image.
- If the proportions (aspect ratio) of the image are changed, the image will be skewed.
- Browsers are much better now at scaling raster images than they used to be.
- Vector graphics (svg) scale perfectly to different sizes without loss of quality.
- In responsive web design, images are rescaled or recropped dynamically (using CSS) to better suit the available space.

Always Use Alternative Text

Although it is no longer always mandatory in XHTML5, you should always specify alternative text (the `alt` attribute) for an image:

- If an image has moved or for some reason cannot be fetched, the text indicates what would have been there.
- Alt text is indexed by some of the search engines.
- Alternative text can be rendered by text-only browsers and read out to sight-impaired users.

Acronyms and Abbreviations

An *abbreviation* is a shortened form of a phrase. An *acronym* is an abbreviation made up of the first letters of the words of a phrase.

- HTML is an abbreviation for HyperText Markup Language.
- CSS is an acronym for Cascading Style Sheets.
- A screen reader would attempt to pronounce the word “css”.
- Use the `abbr` tag to mark up abbreviations and acronyms (the `abbr` tag is obsolete in XHTML5).
- Mousing over will reveal the full form of the abbreviation or acronym.

```
<abbr title="HyperText Markup Language">HTML</abbr>  
<abbr title="Cascading Style Sheets">CSS</abbr>
```

Tables

- Use tables to lay out tabular information in rows and columns.
- In Polyglot XHTML5, tables *must* use one of the `thead`, `tbody`, or `tfoot` elements, so that the DOMs created when parsing as XML and HTML5 are the same. For example:

```

<table>
<thead>
<tr><th>Beer</th><th>Price €</th></tr>
</thead>
<tbody>
<tr><td>Puntigamer</td><td>2,60</td></tr>
<tr><td>Gösser</td><td>2,60</td></tr>
<tr><td>Guinness</td><td>4,35</td></tr>
</tbody>
</table>

```

- Do *not* misuse tables to do page layout. Ever!

Elements Historically Associated with Misuse

Although the following elements are valid in XHTML5, they have a history of being abused for presentational rather than structural markup:

- `` (bold)
- `<i>` (italic)
- `<u>` (underline)
- `<small>` (small font)
- `` (small font)
- `<hr>` (horizontal rule)

Do *not* use them!

The first five can be achieved using `` and style sheets.

Drawing a horizontal line can be achieved with the CSS properties `border-top` or `border-bottom` (or `border`).

[The HTML5 draft standard [W3C, 2014c] attempts to assign new semantic meaning to these old and often-abused tags. In INM their use remains banned.]

Obsolete Attributes of Table

The following attributes of `<table>` are obsolete in XHTML5 and should not be used:

- `border`
- `cellpadding`
- `cellspacing`
- `frame`
- `rules`
- `width`

Browser Compatibility

Browsers generally do not support all of XHTML5 or CSS3.

A great resource to check which browsers support a particular feature is caniuse.com [Deveria, 2013].

Browsers also have bugs, like validators too, so expect there to be occasional issues.

10.4 Cascading Style Sheets (CSS3)

- Style sheets control the presentation of the elements in the XHTML file.
- Style parameters can be set in the head of the XHTML file.
- But they are usually contained in a separate external style file (.css) which is referenced with the link tag.
- This way, a single style sheet can be maintained for multiple web pages or entire web sites.

Using an External Style Sheet

```
<!DOCTYPE html>
<html xmlns="http://www.w3.org/1999/xhtml" lang="en" xml:lang="en">

<head>
<meta charset="UTF-8"/>
<link rel="stylesheet" href="./inm.css"/>

<title>Keith Andrews Web Page</title>
</head>

<body>
<h1>Keith Andrews</h1>

<p>
Keith lives in <a href="http://graz.at/">Graz</a>.
</p>
</body>
</html>
```

A Simple External Style Sheet

```
body { color: black; background-color: silver; }

body { margin-left: 5%; margin-right: 5%; }

img { border: none; }

h1,h2 { font-family: Verdana, sans-serif; }

table {
border-spacing: 0.25em 0.5em;
}
```

```

th { border: 0.3em solid #ff0000; }
td { border: 0.3em solid #ff0000; }

footer {
  text-align: center;
  margin: 2em 0 0 0;
  background-color: #f4f4f4;
  border-radius: 1em;
}

```

10.5 Web Hosting

- Web hosting providers (web hosts) provide you with space on a web server and connectivity to the internet for a few Euros per month.
- Thousands of web hosts worldwide: <http://software-pointers.com/de-austria-webhosting.html>, thehostingchart.com, or hosting-review.com.
- For recommendations, see newsgroups at.internet.sonstiges or tu-graz.anzeigen.diverses (search <http://newsarchiv.tugraz.at/> first).

Web Host Packages

Data as of Oct 2013:

- fatcow: US\$ 3.00 per month, unlimited space, unlimited traffic, unlimited MySQL databases, PHP5, unlimited email accounts (POP/IMAP), free .com domain. fatcow.com (discount price via hosting-review.com)
- hostmonster: US\$ 3.95 per month, unlimited space, unlimited traffic, 100 MySQL databases, PHP5, unlimited email accounts (POP/IMAP), free .com domain. hostmonster.com
- justhost: € 3.00 per month, unlimited space, unlimited traffic, unlimited MySQL databases, PHP5, unlimited email accounts (POP/IMAP), 1 free domain. justhost.com
- hoststar.at starentry: € 3.90 per month, 30 gb space, unlimited traffic, 50 MySQL databases, PHP5, unlimited email accounts (POP/IMAP), no free domain. hoststar.at
- easyserver.at Start: € 2.40 per month, 3 gb space, traffic fair use, no MySQL databases, PHP5, unlimited email accounts, max 1 domain, no free domain. easyserver.at
- World4You DomainServer Basic: € 2,50 per month, 2 gb space, unlimited traffic, MySQL fair use, 20 POP3 accounts, max 1 domain, no free domain. world4you.com
- all-inkl privat: € 4.95 per month, 10 gb space, 50 gb traffic per month, 5 MySQL databases, PHP5, 500 email accounts (POP/IMAP), 1 free domain (at, de, com, ...). all-inkl.com

...and many, many more.

Questions to Ask Your Web Host

- How much web space?
- Transfer limits?
- Price and length of contract.
- Unix or Windows host?
- Connectivity (bandwidth) to the internet?
- PHP support (which version)?
- MySQL support (how many databases)?
- How many email accounts?
- Is a domain name included?
- How many subdomains?
- Technical support (free, 24-hour, email or hotline)?

10.6 Using HTTPS on Your Web Site

If you can, provide HTTPS access to your web site as well as HTTP:

- Indeed, make HTTPS the default [Palmer and Zhu, 2013].
- You will need to obtain a signed X.509 server certificate (known colloquially as an SSL certificate) [ZyTrax, 2014].
- *Self-signed certificate*: signing your server certificate yourself is free, but your users will typically see a warning in their web browser about an untrusted connection. [Users then have to either add a security exception to the browser, or manually import your certificate into their browser.]
- *CA-signed certificate*: having a trusted Certificate Authority (CA) sign your server certificate will generally cost money. However, web browsers come with (or use) a store of trusted certificates of CAs and can validate your CA-signed certificate automatically (no need for the user to pre-import your self-signed certificate).
- If you run a web site inside the `tugraz.at` domain, you can obtain a CA-signed certificate through the university (for free). See `tcs.tugraz.at` for details.
- You receive a CA-signed server certificate, together with a chain of intermediate certificates and the root CA certificate (COMODO CA Limited GB in our case at TU Graz). These have to be installed on your web server (say, Apache).

Chapter 11

Designing with Style

“ *Look mom, no tables.* ”

[Keith Andrews, 2004 (and apparently many others independently).]

Use CSS style sheets to create fluid, responsive, scalable web pages.

References

- ++ John Duckett; *HTML & CSS: Design and Build Web Sites*; Wiley, 18 Nov 2011. ISBN 1118008189 (com, uk) [Duckett, 2011]
- ++ Dan Cederholm; *CSS3 For Web Designers*; A Book Apart, 2011. ISBN 0984442529 (com, uk) [Cederholm, 2011]
- ++ Jeremy Keith; *HTML5 for Web Designers*; A Book Apart, 2010. ISBN 0984442502 (com, uk) [Keith, 2010]
- ++ Jeffrey Zeldman; *Designing with Web Standards*; New Riders, 3rd Edition, Oct 2009. ISBN 0321616952 (com, uk) [Zeldman and Marcotte, 2009]

Online Resources

- + Wikipedia; *CSS3*; http://en.wikipedia.org/wiki/CSS3#CSS_3
- ++ Patrick Griffiths; *HTML Dog*; htmldog.com [Griffiths, 2013] {HTML Dog is a great resource for learning HTML and CSS. The tutorials are HTML5 and CSS3. Unfortunately, the reference sections are still XHTML 1.1 and CSS 2.1. Start here and then move on to Polyglot XHTML5 with CSS3.}
- + Chris Coyier; *CSS Almanac*; <http://css-tricks.com/almanac/>
- + webplatform.org; *CSS*; <https://docs.webplatform.org/wiki/css>
- + ALA; *A List Apart CSS*; <http://alistapart.com/topics/code/css/>
 - *Zen Garden: The Beauty of CSS Design*; csszengarden.com
- ++ Alexis Deveria; *Can I use... Compatibility tables for support of HTML5, CSS3, SVG and more in desktop and mobile browsers.*; caniuse.com [Deveria, 2013]

```

biscuits <br/>
tea <br/>
bananas <br/>
milk <br/>

<ul class="shopping">
<li>biscuits</li>
<li>tea</li>
<li>bananas</li>
<li>milk</li>
</ul>

```

Figure 11.1: Two ways of marking up a shopping list in XHTML5. Both are valid XHTML5, but the second is much more flexible in terms of later applying style.

- + *I Want to Use*; iwanttouse.com
- + W3C Core Mobile Web Platform Community Group; *Ringmark*; rng.io
 - W3C; *CSS Current Work* <http://www.w3.org/Style/CSS/current-work.en.html> [W3C, 2014a]
- + W3C; *W3C CSS Validation Service*; <http://jigsaw.w3.org/css-validator/>

Resources in German

- + Heiko Stiegert; *Modernes Webdesign mit CSS*; Galileo Design, 2011. ISBN 3836216663 (com, uk) [Stiegert, 2011]
- + Kai Laborenz; *CSS: Das umfassende Handbuch.*; Galileo Computing, Sept 2011. ISBN 3836217252 (com, uk) [Laborenz, 2011]

11.1 Structuring Your XHTML

Mark Up Your XHTML By Its Structure

- XHTML is about applying meaning to content.
- Mark up your XHTML *structurally*.
- Figure 11.1 shows two ways of marking up a shopping list. Both are valid XHTML5, but the second is much more flexible in terms of later applying style.

Marking Up Header Levels

- The first level of headings should be coded as h1.
- The second level of headings (subheadings) as h2, etc.
- Do not arbitrarily use a h3 somewhere, because you want it to be *smaller* than a h1.
- That would be marking up the presentation rather than the structure.

- In HTML5, sectioning elements like `section` and `article` have their own internal heading structure. Within each section, you can start again with `h1` and HTML5 will get the outline structuring correct.

Block-Level Elements

- A *block-level* element is a larger block of material which may contain other blocks, inline elements, and simple text.
- A block level element is generally displayed on a line of its own, with a line break before and after it.
- Examples of block level elements include `p`, `table`, and `h1` to `h6`.
- HTML5 defines new *content models*. Block-level elements are further divided into:
 - *grouping elements*: `p`, `ul`, `div`, etc.
 - *sectioning elements*: `section`, `article`, `aside`, `header`, `footer`, and `nav`.

Inline Elements (Text-Level Elements)

- An *inline* element is a smaller piece of material, which contains only simple text or possibly other inline elements.
- An inline element is generally displayed within the flow of another element, at the next character position, without any line breaks.
- Examples of inline elements include `em`, `a`, `q`, `img`, and `code`.
- In HTML5, the content model *embedded content* now covers the `img`, `audio`, `video`, and `canvas` elements.

Styling Elements

- The browser applies its own default style to all XHTML elements.
- Individual elements unique on a page can be named with `id`. Use `id` if there is at most one element like this on any one XHTML page.
- Classes of elements can be created with `class`. Use `class` if there might be many such elements on the same XHTML page and all should be styled the same way.
- An element can be a member of multiple classes.

```
<p class="important deadline">  
The deadline is 31 Oct 2005 23:59.  
</p>
```

Generic Block-Level Chunks (`div`)

- Use `div` (for division) to mark a larger chunk of material as a block level element, so that it can be styled as a unit.

- Name the chunk using `id` and/or assign it to one or more classes.
- Assign the name according to the `div`'s meaning and function, not its intended final formatting:

```
Bad: <div id="rightcolumn">
Good: <div id="related">
```

```
<div id="related">
<h2>Related Content</h2>
<p>
Lorem ipsum dolor sit amet, consectetur
adipiscing elit. Etiam congue posuere nibh.
</p>
</div>
```

Generic Inline Chunks (`span`)

- Use `span` to mark a short piece of text within a block, so that it can be styled as a unit.
- Name the `span` using `id` and/or assign it to one or more classes.
- Assign the name according to the `span`'s meaning and function, not its intended final formatting:

```
Bad: <span class="bold">
Good: <span class="variablename">
```

```
...costs <span class="price">&euro; 266</span>, which is...
```

11.2 Styling with CSS

First Make Sure Your XHTML is Valid

- For style to be applied properly, your XHTML needs to be valid.
- Otherwise the browser might not be able to construct the DOM (internal data structure) properly, and style can not be applied properly.
- There is no point debugging your CSS, without first checking that your XHTML validates.

The New CSS3

- CSS 2.1 was a single (quite large) specification document. CSS3 would have been too large and monolithic to attempt in one go.
- CSS3 is split into a variety of modules, each at different stages of development.

Structure of CSS Rules

- A *stylesheet* is a collection of style rules.



Figure 11.2: The Structure of a CSS Rule.

- A rule consists of a *selector* and a *declaration*.
- The declaration says what to do.
- The selector says where to do it.

See Figure 11.2.

```
h1 {color: green;}
```

A Simple Style Sheet

```
body { color: black; background-color: silver; }
body { margin-left: 10%; margin-right: 10%; }
img { border: none; }
h1, h2 { font-family: Arial, Verdana, sans-serif; }
```

CSS Selectors

```
#tagline {...}                      /* the element with id="tagline" */
.important {...}                    /* any element with class="important" */
p.important {...}                   /* all p with class="important" */
h1 {...}                            /* all h1 elements */

div.story p {...}                   /* all p which are inside a div of class story */
div.story > p {...}                /* all p which are direct children of div of class story */
h2 + p {...}                       /* all p which directly follow a h2 */
li ~ li {...}                       /* all li which are siblings of and follow a li */
li:nth-child(2) {...}              /* every second li */
a[href^=tel] {...}                 /* all a with href starting with tel */
```

A good overview of all the various selectors is provided by http://w3schools.com/cssref/css_selectors.asp

Absolute Sizes

- Inches (in)
- Centimetres (cm)
- Millimetres (mm)

- Points (pt)
- Picas (pc)
- Pixels (px)

Do not use them. Absolute sizes mean that your pages do not scale freely.

Possible exceptions might be:

- setting font sizes in points inside a print style sheet.
- specifying the (exact) size of an image in pixels within a style sheet, so the browser does not scale it.

Relative Sizes

- Em (em): the width (and height) of the character box (also called the em-square or quad-width) of the current font. Traditionally, the width of a capital M.
- Rem (rem): root em, like em but referring to the initial default font-size of the root element (html). Using rem solves the problem of the compounding of font-sizing inside nested elements when using em or %.
- Ex-height (ex): the height of a lower case x in the current font.
- Percent (%): a value relative to another value.

For freely scalable web pages, specify sizes in rem, em, ex, and %.

For sizes of 0, the units are optional.

CSS 2.1 Box Model (Content-Box)

- The width and height of an element specify the dimensions of the *inner* content box in which the element is placed.
- Padding, border, and margin are *in addition* to the width and height.

See Figure 11.3.

CSS3 Box-Sizing Property

- CSS3 introduces the box-sizing property.
- If box-sizing is set to content-box, the width and height of an element specify the dimensions of the *inner* content box in which the element is placed (as in CSS 2.1).
- If box-sizing is set to border-box, the width and height of an element specify the dimensions of the border box, i.e. padding and border are considered inside the box. Margin is still added to the box.

See Figure 11.3.

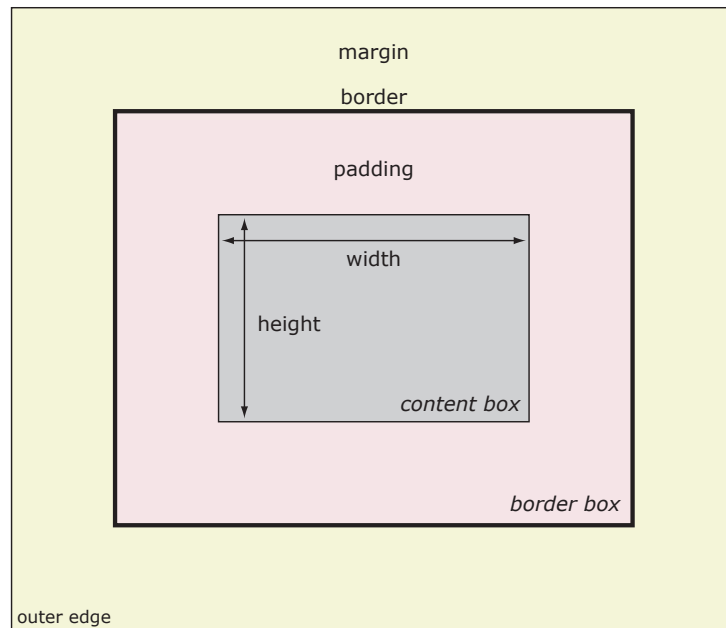


Figure 11.3: The CSS 2.1 box model specifies the inner content box. Padding, border, and margin are *in addition* to the width and height specified. In CSS3, the border box can be specified instead, which includes padding and border (but not margin).

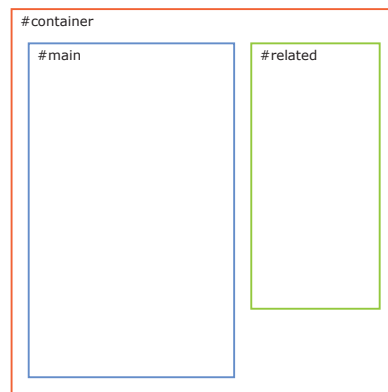
Positioned Elements

- Positioned elements can be put anywhere in relation to the containing block.
- Positioned elements let anything slide underneath them (i.e. content can be obscured).
- Positioned elements can be overlapped in layers.

```
nav#localnav {
  position: absolute;
  top: 0;
  left: 0;
  width: 14%;
}
```

Floated Elements

- You can float any element left or right.
- The element aligns itself to that side of whatever element it is contained within.
- To create multi-column layouts, float the column elements to the same side and they will line up beside each other (as long as they fit).
- The order of floated elements depends on the order of the elements in the XHTML file.
- Floated elements permit background and borders of other elements to slide beneath them, but not the content.



Text which should be below container.

Figure 11.4: When placing floated elements within a container, this is usually the desired outcome. [This diagram was adapted from Vernacchia [2011].]

```
nav#localnav {
  float: left;
  width: 14%;
}

div#content {
  float: left;
  width: 64%;
}

div#related {
  float: left;
  width: 16%;
}
```

Clearfix for Floated Elements

- When elements are floated within a container box, the designer usually desires a result like that shown in Figure 11.4.
- Unfortunately, the container’s height is not adjusted to the height of the floated elements. Figure 11.5 illustrates the result.
- The solution is to use a “clearfix” [Vernacchia, 2011; A.M.K, 2013], the cleanest of which is shown below.

```
#container:after {
  content: " ";
  display: block;
  height: 0;
  clear: both;
}
```

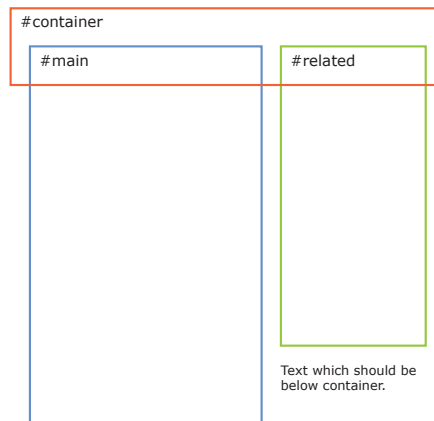



Figure 11.5: Unfortunately, this is what the CSS standard specifies should happen. To achieve the desired result, it is necessary to use a “clearfix”. [This diagram was adapted from Vernacchia [2011].]

Inline Elements as Block Level

- Inline elements can be forced to be displayed as block level elements.
- For example, to display a simple sequence of a tags as a vertical list of links.

```
nav#localnav a {
  display: block;
  text-decoration: none;
  text-align: right;
}
```

- `img` is an inline element. To centre an image within its containing block, make its class say centered and display it as a block:

```
img.centered
{
  display: block;
  text-align: center;
}
```

Block Level Elements as Inline

- Block elements can be forced to be displayed inline.
- For example, to display a list of links coded as a `ul` horizontally.

```
nav#globalnav ul li {
  margin-left: 0;
  border-left: 0.1em solid #000;
  list-style: none;
  display: inline;
}
```

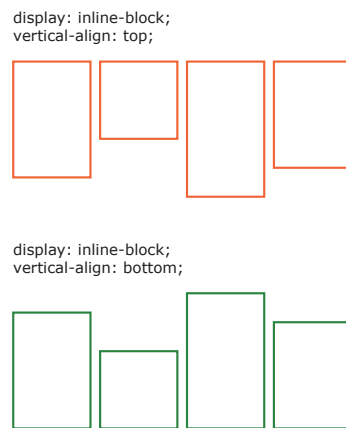


Figure 11.6: An example of inline-block layouts, with vertical-align set to top and bottom respectively. Note that inline block elements are always separated by a single space (like characters on a line), if there is any whitespace between them in the HTML source code.

Inline-Block Elements

- Using `display: inline-block;`, blocks are lined up in a row, as if they were characters [Tetlaw, 2011].
- Align vertically by setting `vertical-align` to top, middle, or bottom.
- Figure 11.6 shows two example inline-block layouts.
- Inline-block elements are always separated by a single space (like characters on a line), if there is any whitespace between them in the HTML source code. See Johnson [2012] and Coyier [2012] for ways to overcome this.

CSS3 Flexbox

CSS3 Flexbox [W3C, 2014b] is now available in all modern browsers <http://caniuse.com/flexbox>.

- A container is given `display: flex;`
- Its children can then be arranged horizontally `flex-flow: row;` or vertically `flex-flow: column;`
- Alignment, spacing, ordering can all be configured.
- Grid layouts, vertical centering, and sticky footers are very easy with flexbox [Walton, 2014].
- Holy Grail Layout (3 columns with header and footer, centre column appears first in source code) [M. Levine, 2006] is now possible with flexbox [Walton, 2013].

Well-Structured XHTML5

- Figure 11.7 shows some well-structured XHTML5.

```

<!DOCTYPE html>
<html xmlns="http://www.w3.org/1999/xhtml" lang="en" xml:lang="en">

<head>
<meta charset="UTF-8"/>
<meta name="viewport" content="width=device-width"/>

<link rel="stylesheet" href="reset.css"/>
<link rel="stylesheet" href="normalise.css"/>
<link rel="stylesheet" href="base.css"/>

<link rel="stylesheet" media="screen" href="./standard.css" />
<link rel="stylesheet" media="print" href="print.css" />

<title>Well-Structured Markup</title>
</head>

<body>

<header>
<h1>Well-Structured Markup</h1>
<p id="tagline">
A witty tagline goes here
</p>
</header>

<nav id="globalnav">
<ul>
<li><a href="#chap1">First Chapter</a></li>
<li><a href="#chap2">Second Chapter</a></li>
<li><a href="#chap3">Third Chapter</a></li>
</ul>
</div>

<nav id="localnav">
<ul>
<li><a href="#sec1">First Section</a></li>
<li><a href="#sec2">Second Section</a></li>
<li><a href="#sec3">Third Section</a></li>
</ul>
</div>

<div id="content">
<h2>Main Content</h2>
<p>
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam congue
posuere nibh.
</p>

<p>
...
</p>
</div>

<aside>
<h2>Related Content</h2>
<p>
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam congue
posuere nibh.
</p>
</aside>

<footer>
<p>
...
</p>
</footer>

</body>
</html>

```

Figure 11.7: Well-structured Polyglot XHTML5 markup. The style sheets can style each element separately. The order of elements means float can be used for positioning.

Why Cascading?

Style definitions are *inherited* and can be overridden:

- The browser uses its own default style sheet, in case no other style sheet is specified.
- The user can provide a local style sheet to override the browser defaults.
- The author of a web page can provide an explicit style sheet for the page.
- The user can also override the page author's style definitions.

Reset Style Sheets

- Browsers have their own defaults for the various style settings.
- A *reset* style sheet typically resets margins, padding, and border to 0 for all elements, font size to 100%, and some other initial settings, so the starting point is the same regardless of browser. Good examples are Meyer [2011]; R. Clark [2010].
- A *normalise* style sheet typically fixes inconsistencies in browsers and browser versions. A good example is Gallagher [2013].
- Appendix D of the CSS 2.1 Specification contained a sample default style sheet for HTML 4 elements in CSS 2.1 [W3C, 2011a].
- I like to collect my own sensible default styling into a style sheet `base.css`.

```
<head>
...
<link rel="stylesheet" href="reset.css"/>
<link rel="stylesheet" href="normalise.css"/>
<link rel="stylesheet" href="base.css"/>

<link rel="stylesheet" media="screen" href="./standard.css" />
...
</head>
```

Alternate Style Sheets

- It is possible to define several alternate style sheets for a web page.
- In this case, the browser (user agent) can allow the user to choose between them:
 - Firefox 33: View - Page Style.
 - IE 11: View - Style.
 - Opera: Opera - Page - Style.
 - Chrome: requires an extension be installed.

See <http://richinstyle.com/guides/advanced2.html>

Specifying Alternate Style Sheets

```

<head>
...

<link rel="stylesheet" href="reset.css"/>
<link rel="stylesheet" href="normalise.css"/>
<link rel="stylesheet" href="base.css"/>

<link rel="stylesheet" media="screen"
      title="Standard Style" href="./standard.css" />

<link rel="alternate stylesheet" media="screen"
      title="Kids Style" href="./kids.css" />

<link rel="stylesheet" media="print"
      href="./print.css" />

...
</head>

```

Media Types

Different style sheets can be specified for different kinds of media. For example:

- all: suitable for all devices.
- speech: for speech synthesisers, screen readers, etc.
- handheld: for handheld devices.
- projection: for projected presentations.
- screen: for color computer screens.
- print: for paged, opaque material and for documents viewed on screen in print preview mode.

Print Style Sheets

A print style sheet is used to specify styles for printing. For example:

- Printing is black on white.
- Navigation elements are usually not printed:


```

nav#localnav {
  display: none;
}

```
- Links are sometimes printed blue and underlined:

```

a:link, a:visited {
  color: blue;
  text-decoration: underline;
}

```

- Links are often printed with their destination URLs immediately following them:

```
a:after {
  content: " (" attr(href) ")";
}
```

- Serif fonts might be used for the main text and sans-serif for headings (are easier to read on paper).
- Font sizes are often specified in points (pt).

See <http://alistapart.com/stories/goingtoprint/> for more details.

Speech Style Sheets

CSS3 defines a media type “speech” and a CSS Speech Module:

- Headings are read in a rich male voice:

```
h1, h2, h3, h4
{
  voice-family: male;
  richness: 80;
}
```

- Emphasised text is read in a loud female child’s voice:

```
em {
  voice-family: child female;
  volume: 90;
}
```

- Links are preceded and followed by specific sounds:

```
a {
  cue-before: url("bell.mp3");
  cue-after: url("dong.mp3")
}
```

- Unfortunately, none of the major browsers currently support the CSS3 Speech Module [Gibbins, 2010].

Web Browser CSS3 Compatibility

Not all browsers support all aspects of CSS3:

- Can I Use has a good overview of browser support for CSS3 and HTML5 [Deveria, 2013].
- I Want to Use shows you what % of users you might be shutting out with a particular feature mix. iwanttouse.com
- There are several CSS test suites:
 - Official W3C CSS Test Suites [W3C, 2011b].
 - Ringmark for mobile web browsers [Coremob, 2012].
 - The Web Standards Project [WaSP, 2011], Acid Tests [Hickson, 2011].

11.3 Responsive Web Design

A single web site design which dynamically responds (adapts) to the characteristics (screen width, resolution, capabilities) of the display device.

Responsive web design is a combination of:

- Responsive layout: dynamically adjusting the layout to best suit the available window size (reflow at breakpoints, rescale in between).
- Responsive content: dynamically adjusting embedded content (tables, images, videos, charts, adverts, etc.) to best suit the device characteristics (space, resolution, etc.).
- Responsive interaction: dynamically adjusting interaction and navigation elements (menus, buttons, search boxes, etc.) to best suit the device characteristics (space, touch, etc.).
- Progressive enhancement: selectively adding support for additional features such as location, orientation, tilt, and gestures. [Detect *features* not devices!]

“Don’t think pages, think components.”

Components which (re-)arrange themselves according page constraints.

References

- ++ Ethan Marcotte; *Responsive Web Design*; A Book Apart, 2011. ISBN 098444257X (com, uk) [Marcotte, 2011]
- + Ben Frain; *Responsive Web Design with HTML5 and CSS3*; Packt Publishing, 10 Apr 2012. ISBN 9350237881 (com, uk) [Frain, 2012]
- + Stephen Hay; *Responsive Design Workflow*; New Riders, 05 Apr 2013. ISBN 0321887867 (com, uk) [Hay, 2013]
- + Scott Jehl; *Responsible Responsive Design*; A Book Apart; Dec 2014. [Jehl, 2014]
- + Tom Barker; *High Performance Responsive Design*; O’Reilly; Dec 2014. [Barker, 2014]

Online Resources

- ++ Matt Andrews; *Responsive Design at the Guardian or How I learned to Stop Worrying and Love the Mobile Web*; <http://mattandrews.info/talks/port80-2013/>
- + BBC News; *Cutting the Mustard*; Responsive News, 08 Mar 2012. <http://responsivenews.co.uk/post/18948466399/cutting-the-mustard>
- + Brad Frost; *This Is Responsive.*; <http://bradfrost.github.io/this-is-responsive/>
- + Brad Frost; *WTF Mobile Web*; <http://wtfmobileweb.com/>

Separate Mobile Web (m.) Sites

Examples include:



Figure 11.8: The Southwest Airlines desktop web site at width 1056 pixels. Screenshot taken by the author on 12 Apr 2012.

- Southwest Airlines' desktop web site southwest.com (Figure 11.8) vs. their mobile web site m.southwest.com (Figure 11.9).

Examples of Responsive Web Sites

Lists of responsive web sites:

- Eivind Uggedal; mediaqueri.es
- DesignModo; <http://designmodo.com/responsive-design-examples/>

Individual examples of responsive web sites:

- *The Boston Globe*; bostonglobe.com
- *Smashing Magazine*; smashingmagazine.com
- *Time*; time.com
- *engadget*; engadget.com
- *BBC One*; <http://bbc.co.uk/bbcone/>
- *Channel 4 News*; <http://channel4.com/news/>
- *The Guardian*; <http://theguardian.com/uk?view=mobile>
- *Gov.UK*; gov.uk

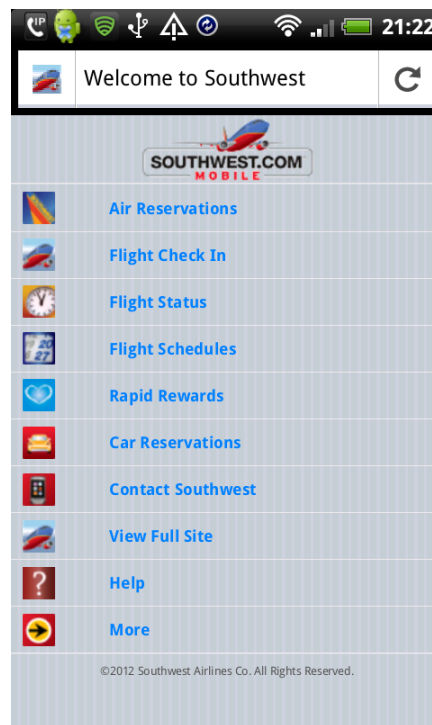


Figure 11.9: The Southwest Airlines mobile web site at width 480 pixels. Screenshot taken by the author on 12 Apr 2012. The constraints of mobile screen widths force designers to focus on what is really important.

- *Andy Clarke*; stuffandnonsense.co.uk
- *SparkBox*; seesparkbox.com
- *FoodSense*; foodsense.is
- *New Adventures in Web Design conference*; 2013.newadventuresconf.com

Figures 11.10 and 11.11 show the same responsive web site (The Guardian) at wide and narrow screen widths.

The Graz Times Responsive Design

A responsive design for The Graz Times web site:

- Figure 11.12 shows the page at resolution 320×480 .
- Figure 11.13 shows the page at resolution of 640×480 .
- Figure 11.14 shows the page at resolution of 1024×768 .

Responsive Layout with CSS3 Media Queries

- First, appropriate breakpoints for the content are determined and specified in a breakpoint diagram as shown in Figure 11.15.



Figure 11.10: The Guardian responsive web site at width 1365 pixels, as it might be seen in a desktop web browser. Screenshot taken by the author on 31 Oct 2014.



Figure 11.11: The same Guardian responsive web site at width 498 pixels, as it might be seen on a smartphone web browser. Screenshot taken by the author on 31 Oct 2014.



Figure 11.12: The Graz Times web page at a resolution of 320 × 480.



Figure 11.13: The Graz Times web page at a resolution of 640 × 480.

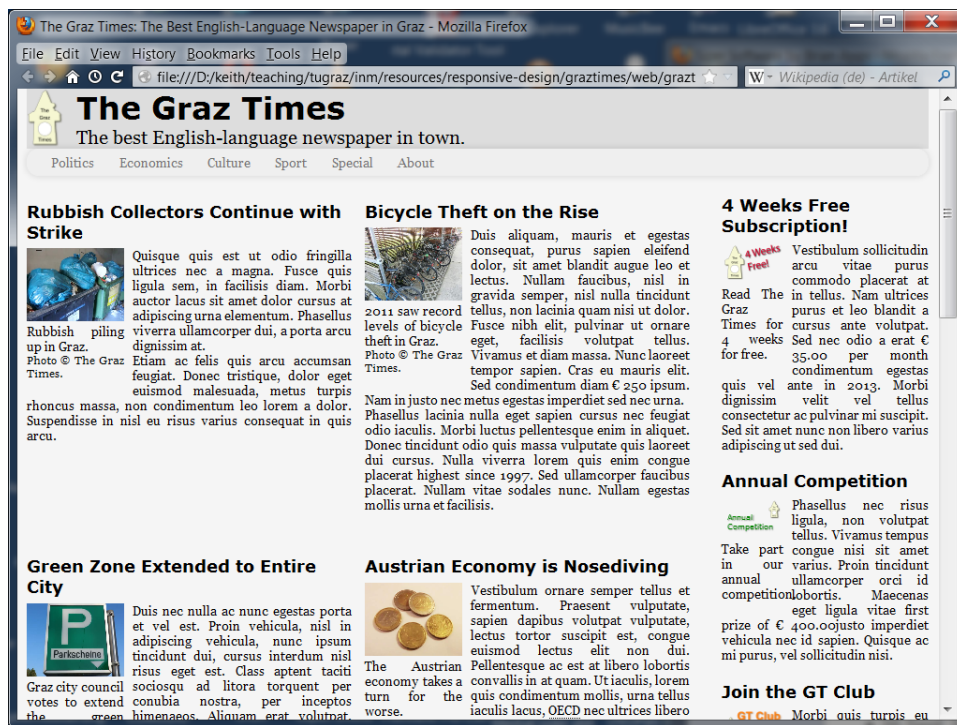


Figure 11.14: The Graz Times web page at a resolution of 1024×768 .

- It is good practice to specify the breakpoints in em or rem rather than in px, so they also respond when users set larger fonts (or text zooming). [Anyhow, at the root element, 1 em = 1 rem.]
- Most of the CSS code can be shared between all window widths.
- CSS3 media queries are used to distinguish between the different screen widths.

Specifying Breakpoints with CSS3 Media Queries

```

/* settings for window-width: narrow */
@media only all and (max-width: 40rem) {
  ...
}

/* settings for window-width: medium */
@media only all and (min-width: 40rem) and (max-width: 60rem) {
  ...
}

/* settings for window-width: wide */
@media only all and (min-width: 60rem) {
  ...
}

```

Breakpoint Diagram for The Graz Times

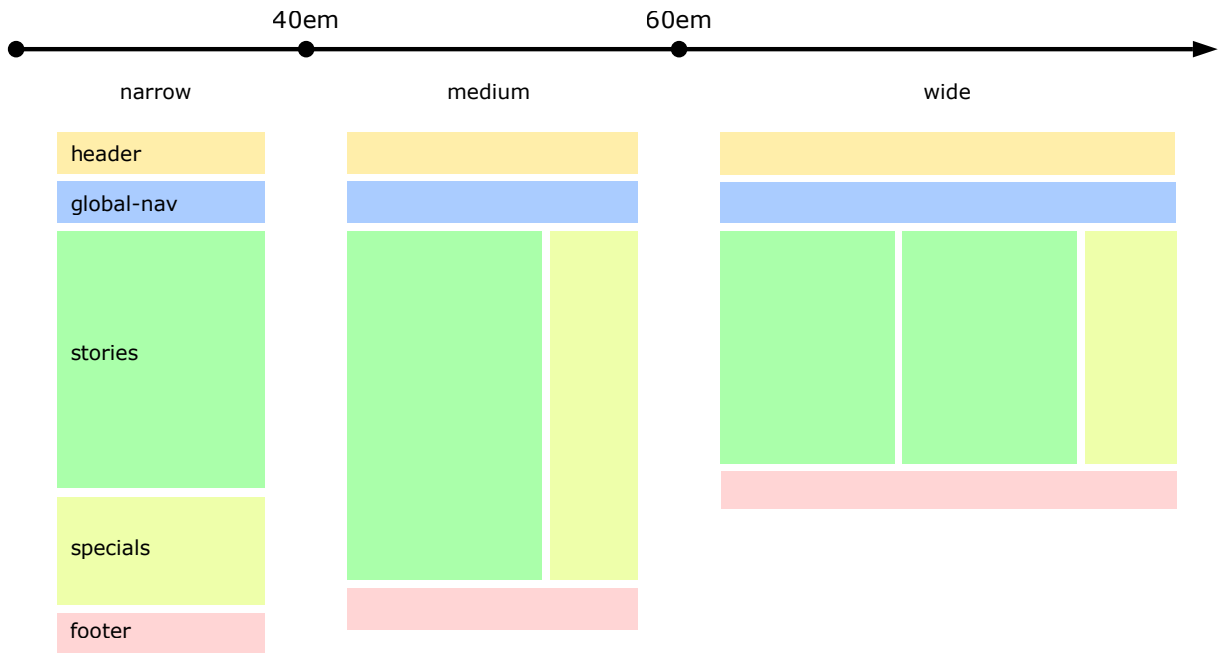


Figure 11.15: A breakpoint diagram.

Checking the Responsiveness of a Design

- Try at various font sizes and various window sizes.
- Grab the corner of the window and move in and out. [Only web designers ever do this!]
- Change the font size up and down (text zoom).
- In IE and Firefox, Control-Mousewheel zooms the entire page (page zoom) by default, rather than increasing or decreasing the font size (text zoom).
- You have to choose Text Zoom under View - Zoom or install the Firefox No Squint extension.
- In Firefox, under Tools - Web Developer - Responsive Design View you can test your web page at various window sizes.
- Under Windows, the Sizer software <http://www.brianapps.net/sizer/> allows you to resize any window to an exact size.

Chapter 12

A Short History of the Internet

“During my service in the United States Congress, I took the initiative in creating the Internet.”

[Al Gore, mistakenly claiming to have invented the internet, in a CNN interview, March 1999]

References

- ++ Katie Hafner and Matthew Lyon; *Where Wizards Stay Up Late: The Origins of the Internet*; Touchstone Books, 1998. ISBN 0684832674 (com, uk)
- + Tim Berners-Lee; *Weaving the Web*; Collins, 2000. ISBN 006251587X (com, uk)
- + James Gillies and Robert Cailliau; *How the Web was Born*; Oxford University Press, 2000. ISBN 0192862073 (com, uk)
- o Katie Hafner; *The Well: A Story of Love, Death & Real Life in the Seminal Online Community* Carroll & Graf Publishers, 2001. ISBN 0786708468 (com, uk)

Online Resources

- ++ Computer History Museum; *A History of the Internet: 1962-1992*. http://www.computerhistory.org/internet_history/
- o Bill Stewart; *Living Internet*. <http://livinginternet.com/>
- o Robert Zakon; *Hobbes' Internet Timeline*. <http://www.zakon.org/robert/internet/timeline/>
- o Richard Griffiths; *History of the Internet*; <http://www.let.leidenuniv.nl/history/ivh/>
- o Leiner et al; *A Brief History of the Internet*. <http://www.isoc.org/internet/history/brief.shtml>
- o Andrew Odlyzko; *The History of Communications and its Implications for the Internet*. <http://www.dtc.umn.edu/~odlyzko/doc/history.communications0.pdf>
- o Internet Archive; <http://www.archive.org>
- o Katie Hafner; *The Epic Saga of The Well*; Wired 5.05, May 1997. http://www.wired.com/wired/archive/5.05/ff_well.html

- RFC Editor; <http://www.rfc-editor.org/>

Videos

- + PBS; *Nerds 2.0.1: A Brief History of the Internet*;
Oregon Public Broadcasting, 1998. [PBS, 1998] 3× 1-hour videos.
<http://pbs.org/opb/nerds2.0.1/>
<http://amazon.com/exec/obidos/ASIN/6305128235/keithandrewshcic>
<http://archive.org/search.php?query=subject%3A%22Nerds+2.0.1%22>
http://youtube.com/watch?list=PL_I1I1rxhtPPHU4tNvoaE2c20DEalP4B0&v=d9CaPat78p8
- + OU; *The Web Weavers*
BBC / Open University, 1998. [BBC, 1998] 30-minute video.

Internet Timeline 1960s

- 1961 Leonard Kleinrock, MIT. PhD thesis proposal, “Information Flow in Large Communication Nets”. First paper on packet-switching network theory.
- 1962 J.C.R Licklider writes memos describing a *galactic network* of people using an interconnected set of computers.
- 1963 The ASCII (American Standard Code for Information Interchange) code is standardised, 128 7-bit strings.
- 1964 Paul Baran, RAND. “On Distributed Communications Networks”, package switching networks and their redundancy in case of attack.
- 1965 First two computers (TX-2 at MIT and AN/FSQ-32 at System Development Corporation in Santa Monica) connected by dedicated 1200 bps phone line and acoustic modem.
- 1966 Lawrence G. Roberts, MIT. “Towards a Cooperative Network of Time-Shared Computers”, first ARPANET plan.
- 1967 Donald Davies, NPL, England, coins the term “packet switching”.
The NPL network, an experimental packet-switching network, runs on 768 kbps lines.
- 1969 ARPANET runs on 4 nodes (UCLA, SRI, UCSB, Uni. Utah). First attempt at LOGIN.
First RFC (Request for Comment) issued RFC 1 : *Host Software* [Crocker, 1969].

Internet Timeline 1970s

- 1970 ARPANET hosts start using Network Control Protocol (NCP).
- 1971 15 nodes on ARPANET.
Ray Tomlinson of BBN invents first email program.
- 1972 RFC 318 : *Telnet Protocol* [J. B. Postel, 1972].
- 1973 First international nodes: NORSAR in Norway and UCL in London.
Bob Metcalfe outlines idea for Ethernet.
RFC 454 : *File Transfer Protocol* [McKenzie, 1973-02-16].

1974 Vint Cerf and Bob Kahn publish Transmission Control Program (TCP) proposal.

1975 First ARPANET mailing list, MsgGroup, is created by Steve Walker.

1976 UUCP email developed between Unix machines.

1978 TCP split into TCP and IP.

1979 USENET news established using UUCP. All original newsgroups were under net.* hierarchy.

ARPA establishes the Internet Configuration Control Board (ICCB).

Internet Timeline 1980s

1983 ARPANET switches over from NCP to TCP/IP on 01 Jan 1983, the start of the “Internet” as we know it.

1984 Domain Name System (DNS) introduced.

More than 1000 Internet hosts.

1985 First registered domain name (01 Jan 1985): nordu.net.

The WELL (Whole Earth 'Lectronic Link) is started.

First registered .com domain name (15 Mar 1985): symbolics.com.

1986 Network News Transfer Protocol (NNTP).

NSFNET created, highspeed backbone between universities.

IETF established.

1987 More than 10,000 hosts.

1988 First *worm* affects 10% of the 60,000 Internet hosts.

IANA established.

Internet Relay Chat (IRC) developed in Finland.

1988 More than 100,000 hosts.

Internet Timeline 1990s

1991 WAIS released.

Gopher released.

WWW released.

Hyper-G released.

NSFNET traffic passes 1 trillion bytes/month.

1992 TU Graz web site (TUGinfo) goes live, number 6 on CERN's list of WWW servers.

Internet Society formed.

More than 1,000,000 hosts.

1993 NCSA Mosaic released.

- 1994 Pizza Hut online pizza ordering.
Netscape released.
First banner ads on hotwired.com.
WebCrawler, first web search engine.
- 1995 Sun launches Java.
Real Audio.
Netscape goes public.
AltaVista.
VRML.
- 1996 Browser wars: Netscape vs. IE.

RFCs

- Request for Comment (RFC) is discussion document proposing a new technical standard.
- Archived at <http://www.faqs.org/rfcs/> and <http://www.rfc-editor.org/>
- For example, RFC 1081 *Post Office Protocol - Version 3* [M. Rose, 1988].

12.1 Remote Login (Telnet)

- Log in to a remote computer somewhere on the Internet.
- First proposal for a telnet protocol in RFC 97 [Melvin, 1971], in Feb 1971.
- First specific telnet protocol in RFC 318 [J. B. Postel, 1972], in April 1972.

See Figure 12.1.

12.2 File Transfer (FTP)

- Fetch and/or upload files from/to remote server.
- RFC 454 *File Transfer Protocol* [McKenzie, 1973-02-16].

12.3 Electronic Mail (Email)

- First email program invented in 1971 by Ray Tomlinson of BBN.
- Derived from two other programs: SENDMSG (intra-machine email) and CPYNET (experimental file transfer).
- The @ sign was chosen from Tomlinson's teletype terminal meaning "at".
- Larry Roberts wrote first email management program (RD) in 1972 for listing, reading, saving, forward, and replying.

```

fiicmpc61:~ 159>telnet fiicmss02
Trying 129.27.153.30 port 23...
Connected to fiicmss02.

SunOS 5.7

login: kandrews
Password:
Last login: Mon Oct 29 08:22:25 from fiicmpc61
Sun Microsystems Inc.   SunOS 5.7       Generic October 1998
Running .cshrc
Fri Nov  2 19:20:16 MET 2001
fiicmss02:~ 1>ps
  PID TTY          TIME CMD
 21407 pts/18    0:00 tcsh
fiicmss02:~ 2>logout
Vergiss nicht auf die Datensicherung!
Remote server has closed connection
Connection closed by foreign host.
fiicmpc61:~ 160>

```

Figure 12.1: Telnet allows a connection to be made for logging in to a remote computer.

- SMTP (Simple Mail Transfer Protocol) specified in RFC 788 in 1981 for switch over to TCP/IP [J. B. Postel, 1981].

12.4 Newsgroups (USENET News)

- Initial hierarchy of news topics called USENET, distributed from server to server over UUCP (1979).
- All original newsgroups were under net.* hierarchy.
- NNTP (Network News Transfer Protocol) developed in 1986 as RFC 977 for more efficient distribution of newsgroups over TCP/IP [Kantor and Lapsley, 1986].

12.5 Internet Relay Chat (IRC)

- Multi-user chat system, developed by Jarkko Oikarinen in Aug 1988.
- People convene on "channels" (virtual places, usually with a specific topic of conversation, such as #cricket).
- Talk in groups or privately.
- Gained international fame during the 1991 Persian Gulf War, live updates from the region.
- RFC 1459 *Internet Relay Chat Protocol* [Oikarinen and Reed, 1993] standardised in May 1993.
- There are now many (web-based) alternatives.

- IRC itself has split into four major and several dozen smaller networks (see <http://www.irchelp.org/irchelp/networks/>).

See Figure 12.2.

12.6 Wide Area Information Servers (WAIS)

- Search and retrieval of information across Internet.
- Client/server system with binary protocol.
- April 1991: freeware UNIX server and Mac client.
- “Natural language” queries.
- Full-text search with normalised ranking (scores from 0 . . . 1000).
- Relevance feedback.
- Multimedia documents indexed by title or descriptive tags.
- WAIS clients send queries to (multiple) servers, display (combined) ranked search results, and fetch and display desired documents.

See Figure 12.3 and Figure 12.4.

12.7 Gopher

- Hierarchically structured information system.
- Developed in 1990-91 as campus info system at University of Minnesota.
- Purpose: “go fer” things across the Internet.
- Largely ASCII client-server protocol (see Figure 12.5).
- RFC 1436 *The Internet Gopher Protocol* [Anklesaria et al., 1993].

See Figure 12.6.

12.8 The World Wide Web (WWW)

Very Early History of the Web (as told by Tim BL)

The Plan

1. Make sexy hypertext software package (NeXT).
2. Give it scaling property.
3. Sit back and watch.

```

IRC Log started Thu Jan 17 19:43
<EricBlade> Sorry about this, need to test a new client function to make
sure I can send to this channel.
-> *banshee* thank you
-> *ericblade* it came through.
*EricBlade* Thanks.
<Python:+report> palestinia sources say 5 explosions...
1st one 2:20 israel time (cnn)
-> *nati* Hi, I am chris.... I am sorry to hear about this ...
I pray for you and your family
<Enigma:+report> saudi arabian f-15 missed
> nbc says major attack
<cos> Israel says one chemical warhead hit Tel Aviv
<bunny:+report> there is also a thunder storm accompanying the missile attack
> from pentagon says israel has "jerico" [sp] to respond immediately.
<Ryman:+report> NBC: Pentagon says its a major attack
*Swan* scott@!! this is swan could you make me a moderator?
<Mustang:+report> is anyone watching cnn?
*cos* "Jericho"
> victims of chemical warfare have been taken to hospital
> I dont have cnn
<Mustang:+report> scot: so it was chemical then?
*Swan* scott! helpppp meeee
<bunny:+report> i have CNN on right now
<EricBlade> Is Nati gone?
<Swan> test
<Python:+report> cnn is trying to talk to security chief at sheraton in
tel aviv (unsuccessfully)
<Python:+report> nati is still on, but not saying anything
> nbc says pentagon has verified 1 chemical missile has hit tel aviv
> tom B. is speechless.
<Mustang:+report> Nati??
!Tatsuya! I can handle some traffic here in hamblin ...
<bunny:+report> no further explosions in the past 20 minutes
-> *nati* hello?
*manson* that's not a bad thing.
<tiger> CNN has problems with audio link to Jerusalem.
-> *manson* no... but everyone expects israel to attack now.. immediately.
the entire nation is on alert.
<bunny:+report> a scud missile landed in south tel aviv
> 3:50am in iraq...
> 2:50 in israel[?]
*manson* this is getting more and more complicated...
<Ryman:+report> NBC: One large building collapsed - people taken to hospital
<bunny:+report> second missile in suburban area SE of Tel Aviv
<Mustang:+report> I hope Nati is ok
<bunny:+report> confirmed 2 missiles landed south of tel aviv

```

Figure 12.2: Part of an IRC log documenting live reports from the Gulf War. Taken from [Scott, 1991].

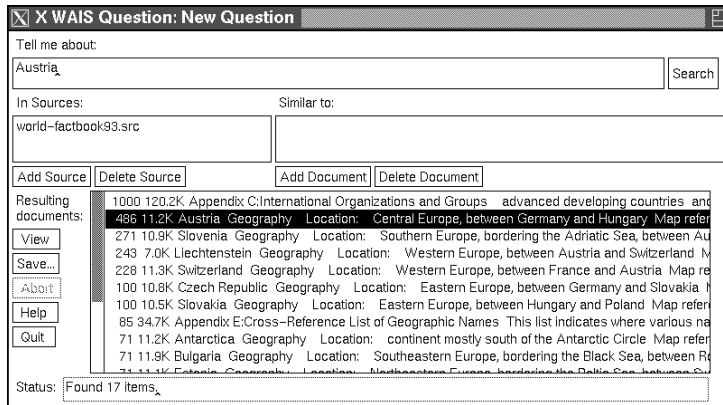


Figure 12.3: Querying a WAIS server. The server world-factbook93 has been given the query term “Austria” and has responded with a list of 17 matching documents ranked by estimated relevance.



Figure 12.4: A document delivered by WAIS, this case “Austria Geography Location”.

```

C:                                     [connects to gopher.micro.umn.edu at port 70]
S:                                     [accepts connection but says nothing]
C: <CR><LF>                             [empty line, meaning “list what you have”]
S:                                     [series of lines ending with CR LF, Δ ≡ TAB]
0About Internet GopherΔStuff:About usΔgopher.micro.umn.eduΔ70
1Around the Uni of MinnesotaΔZ,5692,AUMΔdog.micro.umn.eduΔ70
1Microcomputer News & PricesΔPrices/Δbookstore.umn.eduΔ70
1Courses, Schedules, CalendarsΔΔevents.ais.umn.eduΔ120
1Student-Staff DirectoriesΔΔuinfo.ais.umn.eduΔ70
1Departmental PublicationsΔStuff:DP:Δgopher.micro.umn.eduΔ70
[etc...]
.                                       [period on line by itself]
                                         [server closes connection]
    
```

Figure 12.5: Example of an interaction between Gopher client and server.

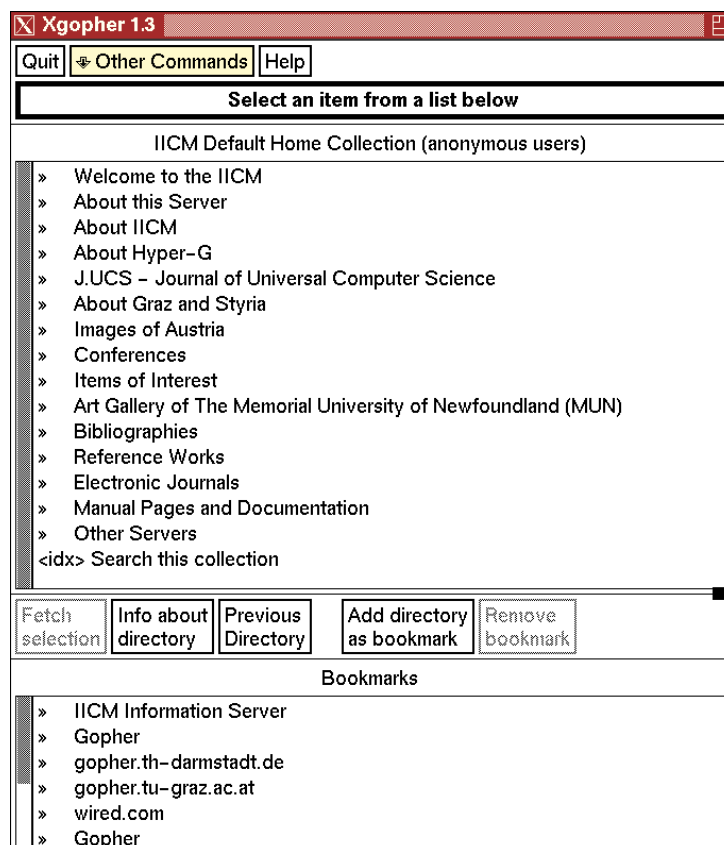


Figure 12.6: Gopher, a predecessor of WWW.

Reaction at CERN

Nobody bit!

- “wrong platform”.
- “too fancy”.
- “should use existing products”.
- “SGML too slow”.

Plan 2

1. Really crude interface (vt100).
2. Bootstrap the web with data (CERN phone book).
3. Post sample source code on net.
4. Persuade volunteers to port to X, Mac, PC.
5. Sit back and watch.

Mar 89	W3 project proposed to CERN management.
Nov 90	Initial prototype on NeXT (see Figure 12.7).
Mar 91	VT100 line-mode browser (see Figure 12.8).
Aug 91	Code posted to <code>alt.hypertext</code> .
Oct 91	Demo of NeXT client at Hypertext 91.
Nov 92	Lynx text browser (see Figure 12.9).
Jan 93	XMosaic alpha released by NCSA (see Figure 12.10).
May 94	First WWW Conference, CERN.
Sep 94	Tim BL leaves CERN for MIT.
Oct 94	Netscape 0.9 beta released. MIT W3 Consortium founded.
Apr 95	VRML and Java demos at WWW'95.

Table 12.1: The early history of the web.

Reaction

Started to generate interest:

- “WWW is a way of accessing information by typing numbers” (quote from newsgroup).

12.9 Hyper-G

- Large-scale, multi-user, distributed, *structured*, hypermedia information system.
- Initiated 1989 by Hermann Maurer, at Graz University of Technology.
- Server for Unix workstations released Dec. 1991, serves Gopher, WWW, and native Hyper-G clients.
- Native clients (authoring tools) for Unix and Windows.

See Figure 12.11 and Figure 12.12

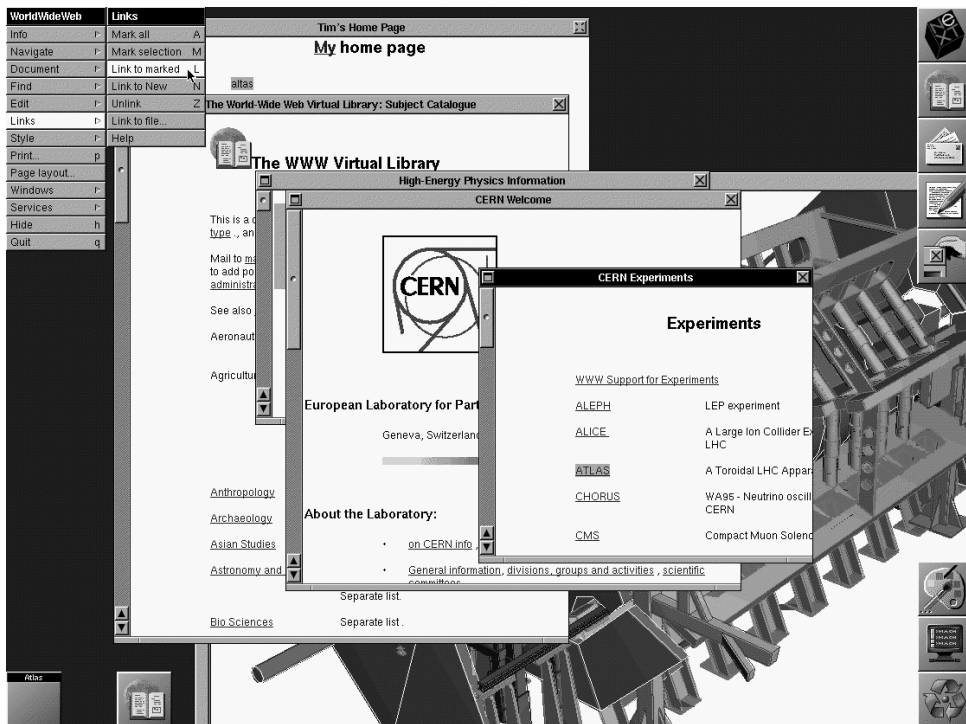


Figure 12.7: The first web client was the CERN NeXT browser/editor, a fully-featured graphical browser and editor, but it only ran on the NeXT workstation.

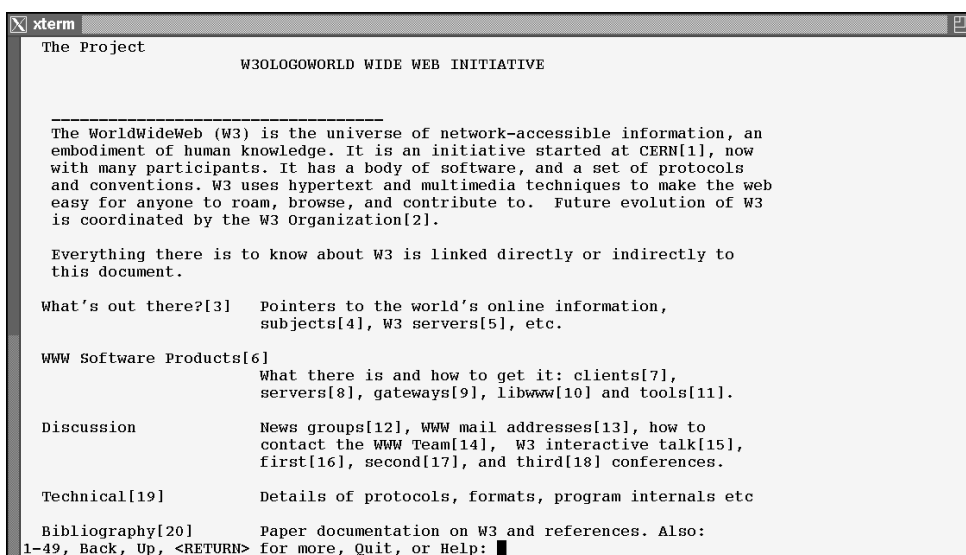


Figure 12.8: The CERN line mode browser.

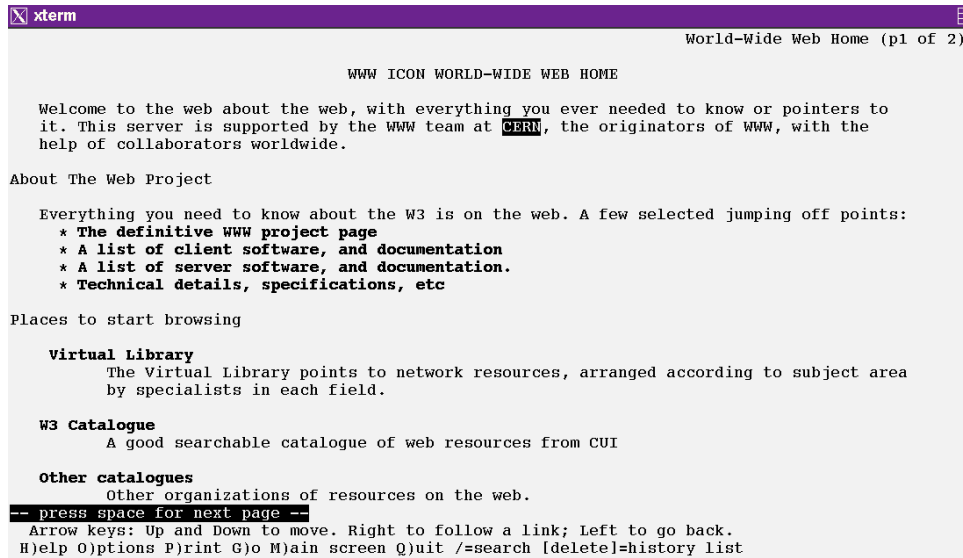


Figure 12.9: The Lynx text browser. Lynx 2.3 beta, VT100 full screen client.



Figure 12.10: The Xmosaic graphical browser.



Figure 12.11: The HGTV terminal viewer (client) for Hyper-G.

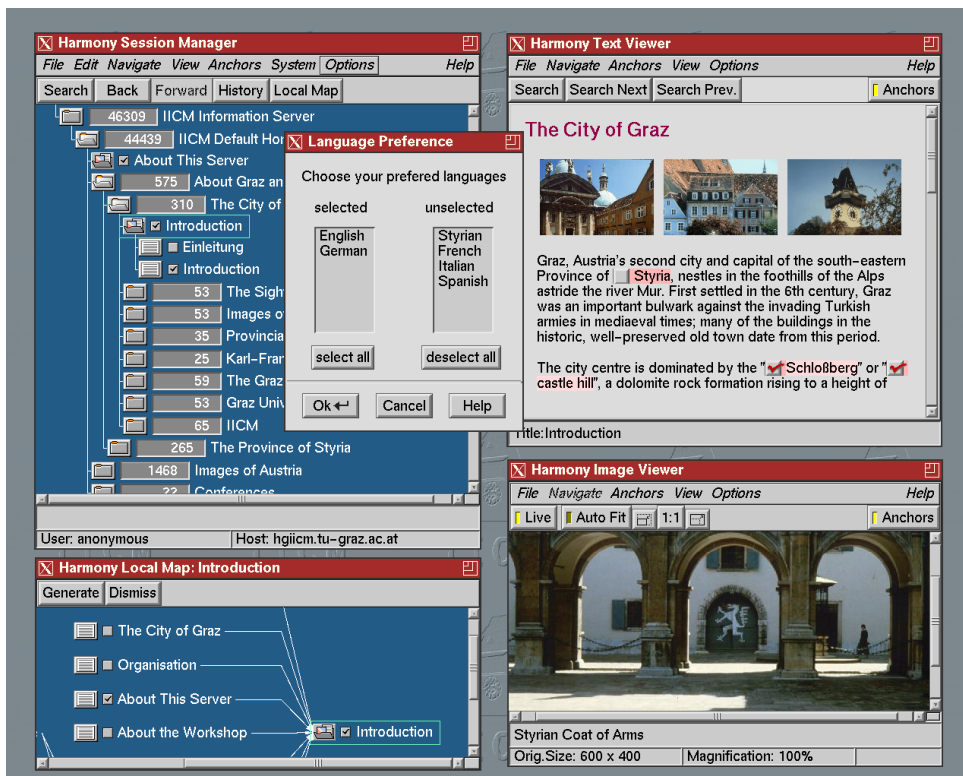


Figure 12.12: The Harmony browser and authoring tool for Hyper-G.

Chapter 13

Other Topics

“ Somehow, there is always a category called ‘miscellaneous’. ”

[Keith Andrews, talking about how users sort concept cards into piles, IAweb lecture, Nov. 2005.]

13.1 Media Types

XML (Extensible Markup Language)

- XML is a markup language for documents containing structured information.
- XML allows new markup languages (with their own sets of tags) to be defined:

```
<shopping>
<item>biscuits</item>
<item>tea</item>
<item>bananas</item>
<item>milk</item>
</shopping>
```

- XML Stylesheets transform XML documents from one format to another. Extensible Stylesheet Language (XSL).
- XHTML is just one example of a markup language formulated in XML.

See <http://www.w3.org/XML/> and <http://www.xml.com/>

Microformats

- Standardised sets of (X)HTML class and rel names.
- See <http://en.wikipedia.org/wiki/Microformats> and <http://microformats.org/wiki/>.

- For example, the hCard microformat is used to mark up contact details in a standard way:

```
<div class="vcard">
  <div class="fn">Keith Andrews</div>
  <div class="org">IICM, Graz University of Technology</div>
  <div class="tel">+43-316-875-5610</div>
  <a class="url" href="http://www.iicm.tugraz.at/keith">
    http://www.iicm.tugraz.at/keith</a>
</div>
```

See also <http://en.wikipedia.org/wiki/HCard>

Raster Image Formats

Raster images store a raster (grid) of pixels (picture elements), each in a particular colour.

- *GIF*: The original format used for most web graphics. Lossless, but typically only capable of handling upto 256 colours. Supports single-bit transparency and short animated sequences. Now replaced by PNG.
- *PNG*: The new open standard for raster images. Lossless and very good compression. Upto 24 bits of colour and 8 bits of transparency per pixel. PNG is now widely supported.
- *JPEG*: A *lossy* format, well-suited to continuous tone images such as photographs. Used for the final finished version of a photograph to place on the web.
- *TIFF*: Lossless image format, capable of storing many different types of image. Used by image editing software. Now largely replaced by PNG.

Rule of thumb: use PNG for almost everything, use JPEG for the final version of a photograph to place on the web.

SVG (Scalable Vector Graphics)

SVG is the new open standard for vector graphics [W3C, 2011c; W3C, 2013; Eisenberg, 2002]:

- Vector graphics store images as a series of objects such as rectangles, circles, lines, curves, and text strings.
- Vector graphics are freely scalable to any pixel resolution.
- SVG is formulated in XML. An example can be seen in Figure 13.1 and the corresponding Listing 13.1.
- A more complex example can be seen in Figure 13.2.
- The major browsers now have native support for SVG 1.1: <http://caniuse.com/svg>.

Portable Document Format (PDF)

- PDF is a file format for printable documents developed by Adobe. http://en.wikipedia.org/wiki/Portable_Document_Format



Figure 13.1: A sample SVG graphic, containing an outer rectangle (with rounded corners), a three-sided polygon (with an inner colour gradient), a red circle, and a piece of text. The source code is given in Listing 13.1.

```
<?xml version="1.0" standalone="no"?>
<!-- Hand-Edited by Keith Andrews 2013-11-12 -->
<!DOCTYPE svg PUBLIC "-//W3C//DTD SVG 1.1//EN"
"http://www.w3.org/Graphics/SVG/1.1/DTD/svg11.dtd">

<svg viewBox="0 0 100 100" version="1.1"
  xmlns="http://www.w3.org/2000/svg"
  xmlns:xlink="http://www.w3.org/1999/xlink">

<g>
<defs>
  <linearGradient id="orange-gradient">
    <stop offset="5%" stop-color="#f60" />
    <stop offset="95%" stop-color="#ff6" />
  </linearGradient>
</defs>

<rect x="0" y="0" width="100" height="100" rx="5" ry="5"
  fill="rgb(240,240,240)" stroke-width="0.5" stroke="black"/>

<circle cx="50" cy="50" r="10" stroke="black"
  stroke-width="2" fill="red"/>

<polygon points="10,10 10,50 50,30"
  fill="url(#orange-gradient)" stroke="purple" stroke-width="1"/>

<text x="10" y="90" font-family="Verdana" font-size="10"
  fill="black" xml:space="default">
Some Text</text>
</g>

</svg>
```

Listing 13.1: The source code of the sample SVG graphic shown in Figure 13.1. The graphic contains an outer rectangle with rounded corners, a three-sided polygon (with an inner colour gradient), a red circle, and a piece of text.



Figure 13.2: The classic tiger vector graphic, originally one of the Ghostscript example images, here as SVG. [Image from Commons [1990] and used under the terms of the GNU General Public License version 3.]

- PDF version 1.0 was released in 1993. PDF was initially proprietary, but Adobe decided to open PDF and it became an ISO standard on 01 Jul 2008. [ISO, 2008]
- Free Acrobat Reader for viewing and printing PDF files is available for most platforms <http://www.adobe.com/products/acrobat/>
- The full version of Acrobat can edit PDFs and allows you to save PDF from Word and other applications.
- Foxit Reader 2 is a free alternative http://www.foxitsoftware.com/pdf/rd_intro.php
- PDFCreator [pdfforge, 2010] and CutePDF Writer [Acro Software, 2010] are free tools which install as a printer and create PDF.
- For other PDF creation software, check out <http://www.pdfzone.com/>.
- For anything longer than a letter, use LaTeX (pdflatex) to create your PDF document. See <http://www.tug.org/begin.html> or <http://latex.tugraz.at/>

Flash

- Binary format (.swf) for animated multimedia (vector graphics, raster images, video, etc.). See http://en.wikipedia.org/wiki/Adobe_Flash.
- Invented by FutureWave (1995), bought by Macromedia (1996), then bought by Adobe (2005).
- Flash player plugin for web browsers on various platforms.
- Originally proprietary, the SWF file format specification was published by Adobe without licence restrictions on 01 May 2008. [Adobe, 2008]
- Dominant format for animated vector graphics on the web.
- Often abused for bloated, lengthy intro pages on web sites.

- New lease of life with Flash video for streaming video.

Audio and Video

- MPEG (MPEG-1, MPEG-2, MPEG-4) is a format for video clips.
- MP3 is a file format for audio clips (MPEG-1 Audio Layer 3).
- Streaming formats for audio and video broadcasts, e.g. Real <http://www.real.com/>
- SMIL (Synchronized Multimedia Integration Language) initiative from W3C <http://www.w3.org/AudioVideo/>

13.2 Dynamic Web Content

Server-side scripting and programming allows web pages to be generated dynamically.

CGI (Common Gateway Interface)

- Convention for running external programs, e.g.:
 - scripts to handle image maps and form requests;
 - access to external databases (SQL, WAIS, archie, etc.).
- Gateway programs can be in any language (C++, Perl, csh).
- *Interface* is standardised through set of environment variables, e.g.:
 - QUERY_STRING passes string following ? in URL
 - PATH_INFO passes extra info and parameters
- Program returns full MIME document or URL reference.

See CGI 1.1 specification at <http://hoohoo.ncsa.uiuc.edu/cgi/>

PHP: Hypertext Preprocessor

- PHP is a server-side, cross-platform, HTML embedded scripting language.

```
<html>
<head><title>PHP Test</title>
</head>
<body>
<?php echo "Hello World<p>"; ?>
</body>
</html>
```

- <http://www.php.net/>
- ASP is the Microsoft version of PHP.
- JSP is the Java version of PHP.

13.3 Services Running over The Internet

Secure Shell (ssh)

- A secure version of Telnet.
- All communication is encrypted.
- Allows you to log into a remote machine (running an ssh server).
- For example, OpenSSH for Unix <http://www.openssh.com/>, PuTTY for Windows/Mac <http://www.chiark.greenend.org.uk/~sgtatham/putty/>

Instant Messaging

- Online interactive, text-based chat between specific, known users (buddy list, friend list, contact list).
 - Unix talk (BSD v4.2 in 1983).
 - ICQ (Internet-based, 1996).
 - MSN Messenger
 - AIM
 - Skype Chat
 - Google Talk
 - social network messaging

See http://en.wikipedia.org/wiki/Instant_messaging

Peer to Peer (p2p) File Sharing

- People who want a file and people who have (parts of) the file are connected together.
- Your p2p client fetches parts of the file you still need, while simultaneously giving parts of the file you already have to other people, until you have a complete copy.
- File sharing services differ in how they locate and trade parts of the files.
- File sharing services are often used to (illegally) share copyrighted material and users are sometimes pursued by copyright owners.
- Some file sharing services:
 - Napster
 - Kazaa
 - Gnutella
 - BitTorrent
- Some file sharing clients:

- LimeWire (Gnutella and BitTorrent)
- Shareaza (Gnutella and BitTorrent)
- Vuze (BitTorrent)
- Phex (Gnutella)

Voice Over IP (VoIP)

- Voice over Internet Protocol (VoIP).
- Telephone calls over the internet.
- Session Initiation Protocol (SIP)
- Skype (2003) uses proprietary protocols.
- VoipStunt
- Google Talk

13.4 The Semantic Web

- The future direction for the web is towards more *semantic* meaning.
- Moving from HTML towards XML.
- The “data” web.
- I recommend you watch the lecture given by Tim Berners-Lee to the Royal Society in London in 2003 [Berners-Lee, 2003].

13.5 Web 2.0

Tim O'Reilly coined the term Web 2.0 to characterise the new breed of web sites which:

- Encourage users to generate content (“harness collective intelligence”). For example, Amazon users add reviews which add value to bare isbn database.
- Continually update their interface (“the perpetual beta”).
- Re-use the data and services of others and open up their own data and service for re-use (“small pieces loosely joined” and “innovation in assembly”).
- Use open APIs and protocols.
- See data as the new source of competitive advantage (“data is the new Intel Inside”). Tele Atlas owns the data behind Google Maps.

See http://en.wikipedia.org/wiki/Web_2 for more details.

Wikis

- A wiki is a read-write web site.
- Users can edit the content of the site as well as read it.
- Usually registration is required to edit pages, and sometimes an editorial process is involved.
- Ward Cunningham developed the first Wiki software (WikiWikiWeb) in 1994.
- Wikipedia <http://www.wikipedia.org/> is the most well-known example of a wiki web site.

Blogs

- A weblog or blog is an online “diary” or journal, typically published by an individual.
- An online interface simplifies the publication process.
- Articles are typically organised by topic and date.
- Blog Software includes:
 - Movable Type
 - Wordpress
- Blog hosting services include:
 - Blogger <http://blogger.com/> (1999, now owned by Google).
 - WordPress.com <http://wordpress.com/>
 - TypePad <http://www.typepad.com/>

See <http://en.wikipedia.org/wiki/Blog>.

Social Bookmarks

Shared collections of bookmarks and tags:

- Del.icio.us url <http://del.icio.us/>
- Users can comment on and add descriptive *tags* (keywords) to web resources.
- Tag clouds.

Social Network Web Sites

- Facebook <http://www.facebook.com/>
- MySpace <http://www.myspace.com/>
- Bebo <http://www.bebo.com/>
- Orkut <http://www.orkut.com/>

- Xing <http://www.xing.com/> (formerly openBC)
- LinkedIn <http://www.linkedin.com/>
- StudiVZ <http://www.studivz.net/>
- SchülerVZ <http://www.schuelervz.net/>

See also http://en.wikipedia.org/wiki/Social_network_service

Microblogging

- Short text updates are distributed along named channels (feeds), often associated with individuals.
- Users can “follow” (subscribe to) one or more feeds.
- Microblogging services include:
 - Twitter <http://www.twitter.com/>
 - Tumblr <http://www.tumblr.com/>
 - Jaiku <http://www.jaiku.com/>

See also <http://en.wikipedia.org/wiki/Microblogging>

RSS Feeds

RSS (originally Rich Site Summary) now stands for Really Simple Syndication:

- XML format for publishing updates (lists of new or changed items and their metadata).
- For example: a feed of news stories, new blog entries, or new podcasts.
- Client software called a *feed reader* or *feed aggregator* reads the RSS XML file and presents a more comfortable interface to the user.
- Most news organisations offer news feeds in RSS, including the BBC and CNN: <http://www.bbc.co.uk/news/10628494>, <http://edition.cnn.com/services/rss/>

Podcasting

A *podcast* is an audio or video item, often a report or an interview, usually one of a series released periodically and publicised via syndication (RSS).

- Sometimes the term *netcast* or *webcast* is used instead of podcast.
- Individuals and organisations can self-publish their own audio or video programmes as podcasts.
- Traditional broadcasters also provide their programmes as podcasts. <http://www.bbc.co.uk/radio4/podcasts/>
- Juice is a popular feed reader for podcasts (podcast receiver) <http://juicereceiver.sourceforge.net/>

13.6 Web Usability and Accessibility

Web Usability

- How to make web sites easier to use.
- Steve Krug; *Don't Make Me Think! A Common Sense Approach to Web Usability*, 2nd Edition, New Riders, Aug. 2005. ISBN 0321344758 (com, uk)
- Martina Manhartsberger and Sabine Musil; *Web Usability - Das Prinzip des Vertrauens*; Galileo Press, ISBN 3898421872 (com, uk) .
- See appropriate chapters from my HCI and Web Usability courses <http://courses.iicm.edu/hci>.

Web Accessibility

- How to make web sites accessible.
- <http://www.w3.org/WAI/>
- John Slatin and Sharron Rush; *Maximum Accessibility: Making Your Web Site More Usable for Everyone*; Addison-Wesley, 2002 ISBN 0201774224 (com, uk) .
- Jim Thatcher, et al; *Constructing Accessible Websites*; APress, 2003. ISBN 1590591488 (com, uk)

13.7 Web Initiatives

Net Neutrality

- Should all internet traffic be treated equally?
- Some ISPs might want to be able to prioritise certain internet traffic.
- Or block certain content or services (say to their competitors).
- Or provide free (or outside of data limit) access to certain web sites.
- According to <http://savetheinternet.com/>:

Net Neutrality simply means no discrimination. Net Neutrality prevents Internet providers from blocking, speeding up or slowing down Web content based on its source, ownership or destination.

See <http://savetheinternet.eu/> and http://en.wikipedia.org/wiki/Network_neutrality.

Be In Control of Your Own Material

Stay in control of your own material:

- If you publish your blog on blogger.com, you hand over control to Google.

- If you upload your photos to Flickr, you hand over control to Yahoo.
- If you upload your videos to Youtube or Vimeo, you hand over control.
- If you save your data to Dropbox or the “Cloud”, you hand over control over it.

GeoCities, once the third most visited web site in the world (in 1999 behind AOL and Yahoo), was acquired by Yahoo in May 1999 for US\$3.57 billion and was shut down on 26 Oct 2009 [Wikipedia, 2014c] eliminating at least 38 million user-built pages.

Indie Web

The Indie Web movement [Çelik, 2014; Novak and Walters, 2014] encourages people to own their own content, self-publish, and syndicate:

- Get your own domain name.
- Run your own web site.
- Run your own cloud (owncloud.org), email server, etc.
- And stay in control of your own data and material.

See Finley [2013]; Gillmor [2014].

Bibliography

- 9tut [2011]. *IPv6 Tutorial*. CCNA Training. 24th Jun 2011. <http://www.9tut.com/ipv6-tutorial> (cited on page 107).
- Abrams, Lawrence [2007]. *How to Show Hidden Files in Windows*. Bleeping Computer. Nov 2007. <http://www.bleepingcomputer.com/tutorials/tutorial62.html> (cited on page 118).
- Abzocknews.de [2009]. *Warnung vor Download-Abzocke auf Skype.at*. 22nd Jan 2009. <http://www.abzocknews.de/2009/01/22/warnung-vor-download-abzocke-auf-skypeat/> (cited on page 125).
- Acro Software [2010]. *CutePDF Writer*. Oct 2010. <http://www.cutepdf.com/> (cited on pages 146, 206).
- Adobe [2008]. *SWF File Format Specification, Version 9*. 1st May 2008. http://www.adobe.com/devnet/swf/pdf/swf_file_format_spec_v9.pdf (cited on page 206).
- Aitken, Duncan [2012]. *HTML5 Tutorial*. 5th Nov 2012. <http://html-5-tutorial.com/> (cited on page 152).
- AK Wien [2008]. *Praxistest Mobiles Breitband*. Sep 2008. <http://www.arbeiterkammer.at/pictures/d77/MobilesBreitband08.pdf> (cited on page 100).
- Alpert, Jesse and Nissan Hajaj [2008]. *We Knew the Web was Big...* 25th Jul 2008. <http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html> (cited on page 4).
- A.M.K [2013]. *How do you keep parents of floated elements from collapsing?* stackoverflow. 8th May 2013. <http://stackoverflow.com/questions/218760/how-do-you-keep-parents-of-floated-elements-from-collapsing> (cited on page 174).
- Anderson, Collin D. et al. [2013]. *Open Letter to Skype from Concerned Privacy Advocates, Internet Activists, Journalists & Other Organizations*. Open Letter. 24th Jan 2013. <http://skypeopenletter.com/> (cited on page 116).
- Andrews, Keith [2002]. “Visual Exploration of Large Hierarchies with Information Pyramids”. In: *Proc. Sixth International Conference on Information Visualisation (IV'02)*. (London, England). IEEE Computer Society Press, Jul 2002, pages 793–798. doi:10.1109/IV.2002.1028871 (cited on pages 88, 89).
- Anklesaria, Farhad et al. [1993]. *The Internet Gopher Protocol*. RFC 1436. Mar 1993 (cited on page 194).
- APA [2011]. *Facebook kooperierte bei Neonazi-Jagd mit Verfassungsschutz*. Der Standard. 30th Dec 2011. <http://derstandard.at/1324501617480/Kaernten-Facebook-kooperierte-bei-Neonazi-Jagd-mit-Verfassungsschutz> (cited on pages 115, 116).

- Appelbaum, Jacob [2013]. *To Protect And Infect, Part 2*. Talk at 30C3 conference. 30th Dec 2013. http://c3media.vsos.ethz.ch/congress/2013/webm/30c3-5713-en-de-To_Protect_And_Infect_Part_2_webm.webm (cited on page 116).
- ARD [2008]. *Abzocke im Internet*. Plusminus TV Broadcast 05 Feb 2008. 5th Feb 2008. <http://www.youtube.com/watch?v=dmZAuv1zrdo> (cited on page 91).
- Atkins, Steve [2007]. “*Personal Firewalls*” are Mostly Snake-Oil. SamSpade.org. Oct 2007. <http://www.samspace.org/d/firewalls.html> (cited on page 119).
- AVG [2006]. *AVG Ewido Anti-Spyware Free*. Oct 2006. <http://free.grisoft.com/doc/ewido-anti-spyware-free/> (cited on page 120).
- BabelStone [2012a]. *BabelMap*. Oct 2012. <http://www.babelstone.co.uk/software/babelmap.html> (cited on page 32).
- BabelStone [2012b]. *BabelPad: Unicode Text Editor for Windows*. Oct 2012. <http://babelstone.co.uk/software/BabelPad.html> (cited on page 34).
- Bare Bones [2014]. *BEdit*. 1st Oct 2014. <http://barebones.com/products/bbedit/> (cited on page 34).
- Barker, Tom [2014]. *High Performance Responsive Design*. O’Reilly, Dec 2014. ISBN 1491949988 (com, uk) (cited on page 181).
- Baseley, W. D. [1999]. *Address Munging FAQ*. Aug 1999. <http://faqs.org/faqs/net-abuse-faq/munging-address/> (cited on page 40).
- BBC [1998]. *The Web Weavers*. 30 minute video. Open University Learning Resources M206/AIP/03, 1998 (cited on pages 93, 190).
- BBC [2007]. *Jury Penalises Music File-Sharer*. BBC News. 5th Oct 2007. <http://news.bbc.co.uk/1/hi/technology/7029229.stm> (cited on page 92).
- BBC [2008]. *Game Sharers Face Legal Crackdown*. Aug 2008. <http://news.bbc.co.uk/1/hi/technology/7568642.stm> (cited on page 92).
- BBC News [2004]. *Image Flaw Exposes Windows PCs*. 16th Sep 2004. <http://news.bbc.co.uk/2/hi/technology/3661678.stm> (cited on page 117).
- BBC News [2005]. *Viruses Use Sony Anti-Piracy CDs*. 11th Nov 2005. <http://news.bbc.co.uk/2/hi/technology/4427606.stm> (cited on page 126).
- Bell, Jon and Thor Kottelin [2014]. *Quoting Style in Newsgroup Postings*. 10th Oct 2014. <http://anta.net/misc/mnq/nquote.shtml> (cited on page 50).
- Bernard, Michael et al. [2002]. “A Comparison of Popular Online Fonts: Which Size and Type is Best?” *Usability News* 4.1 (Jan 2002). <http://psychology.wichita.edu/surl/usabilitynews/41/onlinetext.htm> (cited on page 143).
- Berners-Lee, Tim [2003]. *WWW Past and Future*. 1-hour talk plus 25 minutes of questions. The Royal Society, London, Sep 2003. <http://www.royalsoc.ac.uk/live/> (cited on pages 133, 209).
- Berners-Lee, Tim, Robert Cailliau et al. [1994]. “The World-Wide Web”. *Communications of the ACM* 37.8 (Aug 1994), pages 76–82. doi:10.1145/179606.179671 (cited on page 133).
- Berners-Lee, Tim, Roy T. Fielding and Larry Masinter [1998]. *Uniform Resource Identifiers (URI): Generic Syntax*. RFC 2396. Aug 1998 (cited on page 159).
- Berners-Lee, Tim, Roy T. Fielding and Larry Masinter [2006]. *Uniform Resource Identifier (URI): Generic Syntax*. RFC 3986. Jan 2006 (cited on page 159).

- Bernstein, Daniel J. [2014]. *FTP: File Transfer Protocol*. 23rd Oct 2014. <http://cr.yp.to/ftp.html> (cited on page 129).
- bit.do [2014]. *List of URL Shorteners*. 10th Oct 2014. <http://bit.do/list-of-url-shorteners.php> (cited on page 44).
- Bleich, Holger and Sven Neuhaus [2012]. “Brief mit Siegel”. *c’t Magazin* 2012.18 (13th Aug 2012), pages 136–140. <http://heise.de/-1911842> (cited on page 71).
- Borg, Peter [2014]. *Smultron*. 1st Oct 2014. <http://smultron.sourceforge.net/> (cited on page 34).
- Brain, Marshall [2002]. *How Internet Cookies Work*. Oct 2002. <http://www.howstuffworks.com/cookie1.htm> (cited on page 146).
- Brin, Sergey and Lawrence Page [1998]. “The Anatomy of a Large-Scale Hypertextual Web Search Engine”. *Computer Networks and ISDN Systems* 30.1–7 (Apr 1998). Proceedings of the Seventh International World Wide Web Conference (WWW7), Brisbane, Australia, pages 107–117. doi:10.1016/S0169-7552(98)00110-X. <http://infolab.stanford.edu/pub/papers/google.pdf> (cited on pages 73, 75).
- Broersma, Ron [2010]. *IPv6 Tutorial*. DREN Networkers 2010. 16th Aug 2010. <http://www.hpcmo.hpc.mil/IPV6/wp-content/networkers-IPv6-Tutorial-2010.pdf> (cited on page 107).
- Buckley, Peter [2009]. *The Rough Guide to Macs and OS X Snow Leopard*. 3rd edition. Rough Guides, 9th Oct 2009. ISBN 1848362765 (com, uk) (cited on page 1).
- Buckley, Peter and Duncan Clark [2009]. *The Rough Guide to the Internet*. 14th edition. Rough Guides, 3rd Aug 2009. ISBN 1848361068 (com, uk) (cited on pages xv, 1, 77).
- Burnett, Colin [2007]. *IP Stack Connections*. 24th Mar 2007. http://commons.wikimedia.org/wiki/Image:IP_stack_connections.svg (cited on page 105).
- Callas, J. et al. [2007]. *OpenPGP Message Format*. RFC 4880. Nov 2007 (cited on page 70).
- Creative Commons [2008]. *Creative Commons*. Oct 2008. <http://creativecommons.org/> (cited on page 91).
- Cederholm, Dan [2004]. *Web Standards Solutions*. friends of ED, 2004. ISBN 1590593812 (com, uk) (cited on page 151).
- Cederholm, Dan [2011]. *CSS3 for Web Designers*. A Book Apart, 2011. ISBN 0984442529 (com, uk). <http://www.abookapart.com/products/css3-for-web-designers> (cited on pages 151, 167).
- Çelik, Tantek [2014]. *IndieWebCamp*. 14th Oct 2014. <http://indiewebcamp.com/> (cited on page 213).
- Cheung, Humphrey [2005]. *The Feds Can Own Your WLAN Too*. Tom’s Networking. 31st Mar 2005. http://www.tomsnetworking.com/2005/03/31/the_feds_can_own_your_wlan_too/ (cited on page 120).
- Churchill, Winston [1940]. *Their Finest Hour*. Speech to the House of Commons 18 Jun 1940. 18th Jun 1940. <http://www.winstonchurchill.org/learn/speeches/speeches-of-winston-churchill/122-their-finest-hour> (cited on page 87).
- Clark, Richard [2010]. *HTML5 Reset Stylesheet*. 17th Sep 2010. <http://html5doctor.com/html-5-reset-stylesheet/> (cited on page 178).
- Clarke, Andy [2010]. *Hardboiled Web Design*. Five Simple Steps, 19th Oct 2010. ISBN 1907828001 (com, uk). <http://hardboiledwebdesign.com/> (cited on page 151).

- Claudius [2009]. *Knapp vorbei ist auch daneben*. 20th Jan 2009. http://share.skype.com/sites/de/2009/01/knapp_vorbei_ist_auch_daneben.html (cited on page 125).
- Comer, Douglas E. [2014]. *Computer Networks and Internets*. 6th edition. Addison Wesley, 2nd Jan 2014. 672 pages. ISBN 0133587932 (com, uk) (cited on pages 1, 93).
- Commons [1990]. *Tiger from the Ghostscript Examples*. Wikimedia Commons. 12th Apr 1990. http://commons.wikimedia.org/wiki/File:Ghostscript_Tiger.svg (cited on page 206).
- comscore [2012]. *August 2012 U.S. Search Engine Rankings*. 12th Sep 2012. http://www.comscore.com/Press_Events/Press_Releases/2012/9/comScore_Releases_August_2012_U.S._Search_Engine_Rankings (cited on page 74).
- Consumerist [2009]. *Facebook's New Terms Of Service: "We Can Do Anything We Want With Your Content. Forever."* Consumerist. 15th Feb 2009. <http://consumerist.com/2009/02/15/facebook-new-terms-of-service-we-can-do-anything-we-want-with-your-content-forever/> (cited on page 116).
- Coremob [2012]. *Ringmark*. W3C Core Mobile Web Platform Community Group. 16th Nov 2012. <http://rng.io/> (cited on page 180).
- Costolo, Dick [2011]. *Full DCES Interview Video: Twitter CEO Dick Costolo*. Interview with Kara Swisher at CES 2011. 7th Jan 2011. <http://allthingsd.com/20110117/full-dces-interview-video-twitter-ceo-dick-costolo/> (cited on page 13).
- Coyier, Chris [2012]. *Fighting the Space Between Inline Block Elements*. css-tricks. 21st Apr 2012. <http://css-tricks.com/fighting-the-space-between-inline-block-elements/> (cited on page 176).
- Crispin, Mark R. [2003]. *Internet Message Access Protocol - Version 4rev1*. RFC 3501. Mar 2003 (cited on pages 60, 61).
- Crocker, Steve [1969]. *Host Software*. RFC 1. 7th Apr 1969 (cited on page 190).
- c't-TV [2007]. *Abofallen, Gewinnversprechen, kostenlose Warenproben*. 22nd Sep 2007. <http://www.heise.de/ct/tv/artikel/96119> (cited on page 125).
- Cutts, Matt [2012]. *How Does Google Search Work?* Google Webmaster Help Video. 23rd Apr 2012. <http://www.youtube.com/watch?v=KyCYyoGusqs> (cited on pages 73, 75).
- Daublebsky, Max [2014]. *WLAN: Hier surfen Sie durch die Stadt*. Woche Graz Ost. 8th Oct 2014, pages 8–9. <http://meinbezirk.at/epaper/woche-graz-ost-ausgabe-412014-e17340.html> (cited on page 99).
- De Kunder, Maurice [2010]. *The Size of the World Wide Web*. 30th Sep 2010. <http://www.worldwidewebsite.com/> (cited on page 4).
- DeletedDomains.com [2002]. *Deleted Domain Lists*. Oct 2002. <http://deleteddomains.com/> (cited on page 110).
- Deloitte [2013]. *Start Making the Move to IPv6*. CIO Journal. 5th Sep 2013. <http://deloitte.wsj.com/cio/2013/09/05/start-making-the-move-to-ipv6/> (cited on page 107).
- Deveria, Alexis [2013]. *Can I Use... Compatibility tables for support of HTML5, CSS3, SVG and more in desktop and mobile browsers*. 15th Oct 2013. <http://caniuse.com/> (cited on pages 164, 167, 180).
- DOI [2010]. *The Digital Object Identifier System*. International DOI Foundation. 8th Oct 2010. <http://doi.org/> (cited on page 79).

- Doligez, Damien [1995]. *I Broke Hal's SSL Challenge*. 15th Aug 1995. <http://pauillac.inria.fr/~doligez/ssl/index.html> (cited on page 148).
- Droms, Ralph [1997]. *Dynamic Host Configuration Protocol*. RFC 2131. Mar 1997 (cited on page 111).
- Duckett, John [2011]. *HTML & CSS: Design and Build Web Sites*. Wiley, 18th Nov 2011. ISBN 1118008189 (com, uk) (cited on pages 151, 167).
- Dupré, Lyn [1998]. *Bugs in Writing: A Guide to Debugging Your Prose*. 2nd edition. Addison-Wesley, 1998. ISBN 020137921X (com, uk) (cited on page 81).
- Economist [2008]. *Bugging the Cloud*. The Economist. 6th Mar 2008. <http://www.economist.com/node/10789393> (cited on page 116).
- EFF [2014]. *HTTPS Everywhere*. 15th Oct 2014. <http://eff.org/https-everywhere> (cited on page 148).
- Eisenberg, J. David [2002]. *SVG Essentials*. O'Reilly, 12th Feb 2002. ISBN 0596002238 (com, uk). http://commons.oreilly.com/wiki/index.php/SVG_Essentials (cited on page 204).
- Elam, Gunilla, Tomas Stephanson and Niklas Hanberger [2002]. *Warriors of the Net*. 13-minute animation. 2002. <http://www.warriorsofthenet/> (cited on pages 20–23).
- Elam, Gunilla, Tomas Stephanson and Niklas Hanberger [2011]. *Warriors of the Net HD*. 13-minute animation. 2011. http://www.youtube.com/watch?v=PBWhzz_Gn10 (cited on page 20).
- Ellingwood, Justin [2014]. *An Introduction to DNS Terminology, Components, and Concepts*. DigitalOcean Tutorial. 30th May 2014. <https://digitalocean.com/community/tutorials/an-introduction-to-dns-terminology-components-and-concepts> (cited on page 93).
- Enigmail [2014]. *The Enigmail Project*. 29th Aug 2014. <http://enigmail.net/> (cited on page 71).
- Ephorus [2010]. *Ephorus*. 2010. <http://ephorus.com/> (cited on page 83).
- EPIC [2007]. *Gmail Privacy Page*. Electronic Privacy Information Center. Oct 2007. <http://www.epic.org/privacy/gmail/faq.html> (cited on page 59).
- Europakonsument [2009]. *Europäisches Verbraucher Zentrum*. 16th Oct 2009. <http://www.europakonsument.at/> (cited on page 125).
- Fahlman, Scott E. [2007]. *Smiley Lore :-)*. Sep 2007. <http://www.cs.cmu.edu/~sef/sefSmiley.htm> (cited on page 53).
- Ferguson, Jeremy [2011]. *Smartphones Sales Pass PC Sales for the First Time in History!* SmartOnline. 10th Feb 2011. <http://www.smartonline.com/smarton-products/smarton-mobile/smartphones-pass-pc-sales-for-the-first-time-in-history/> (cited on page 10).
- Fiala, Jan [2014]. *PSPad*. 1st Oct 2014. <http://pspad.com/> (cited on page 33).
- FifeSoft [2014]. *RText*. 1st Oct 2014. <http://fifesoft.com/rtext/> (cited on page 34).
- FILExt [2005]. *FILExt – The File Extension Source*. Oct 2005. <http://filext.com/> (cited on page 117).
- Finder, Alan [2006]. *For Some, Online Persona Undermines a Résumé*. New York Times. 11th Jun 2006. <http://www.nytimes.com/2006/06/11/us/11recruit.html> (cited on page 115).
- Finley, Klint [2013]. “Meet the Hackers Who Want to Jailbreak the Internet”. *Wired Magazine* (14th Aug 2013). <http://wired.com/2013/08/indie-web/> (cited on page 213).

- FNM Austria [2008]. *eLearningrechtsfragen*. Forum Neue Medien in der Lehre Austria. Oct 2008. <http://elearningrechtsfragen.at/> (cited on page 85).
- Ford-Hutchinson, Paul [2005]. *Securing FTP with TLS*. RFC 4217. Oct 2005 (cited on pages 129, 132).
- Ford-Hutchinson, Paul [2006]. *ftps - FTP-SSL and FTP-TLS (RFC4217) - the state of play*. 16th Feb 2006. <http://ford-hutchinson.com/~fh-1-pfh/ftps-ext.html> (cited on page 132).
- Frain, Ben [2012]. *Responsive Web Design with HTML5 and CSS3*. Packt Publishing, 10th Apr 2012. ISBN 1849693188 (com, uk) (cited on pages 151, 181).
- Gabrilovich, Evgeniy and Alex Gontmakher [2002]. "The Homograph Attack". *Communications of the ACM* 45.2 (Feb 2002). doi:10.1145/503124.503156. <http://www.cs.technion.ac.il/~gabr/papers/homograph.html> (cited on page 121).
- Gallagher, Nicolas [2013]. *About normalize.css*. 26th Aug 2013. <http://nicolasgallagher.com/about-normalize-css/> (cited on page 178).
- Garfinkel, Simson [2006]. *Attack of the iPods!* CSO Online. May 2006. <http://www.csoonline.com/read/050106/ipods.html> (cited on page 126).
- Gartner [2010]. *Gartner Highlights Key Predictions for IT Organizations and Users in 2010 and Beyond*. Press Release. 13th Jan 2010. <http://www.gartner.com/it/page.jsp?id=1278413> (cited on page 13).
- Gellens, Randall [1999]. *The Text/Plain Format Parameter*. RFC 2646. Aug 1999 (cited on pages 51, 52).
- Gibbins, Jon [2010]. *Aural CSS / Speech*. dotjay. 12th May 2010. <http://lab.dotjay.co.uk/tests/css/aural-speech/> (cited on page 180).
- Gillmor, Dan [2014]. *Why the Indie Web Movement is so Important*. 4th May 2014. <http://dangillmor.com/2014/04/25/indie-web-important/> (cited on page 213).
- GNU [2014]. *Emacs*. 1st Oct 2014. <http://gnu.org/software/emacs/> (cited on page 34).
- GnuPG [2014]. *GNU Privacy Guard*. 10th Oct 2014. <https://gnupg.org/> (cited on pages 70, 71).
- Google [2010]. *Google Groups*. 1st Oct 2010. <http://groups.google.com/> (cited on pages 41, 48).
- Google [2012]. *HTML5rocks*. 5th Nov 2012. <http://html5rocks.com/> (cited on page 152).
- Google [2013]. *IPv6*. 18th Oct 2013. <http://www.google.com/intl/en/ipv6/> (cited on page 107).
- Gralla, Preston [2006]. *How the Internet Works*. 8th edition. Que, 30th Nov 2006. ISBN 0789736268 (com, uk) (cited on page 1).
- Graz [2014]. *Wlan Hotspots in Graz*. Holding Graz. 9th Oct 2014. <http://freegrazwlan.at/> (cited on page 99).
- Griffiths, Patrick [2006]. *HTML Dog: The Best-Practice Guide to XHTML and CSS*. New Riders, Nov 2006. ISBN 0321311396 (com, uk). <http://htmldog.com/> (cited on page 151).
- Griffiths, Patrick [2013]. *HTML Dog*. Apr 2013. <http://htmldog.com/> (cited on pages 151, 167).
- Guarnieri, Claudio and Morgan Marquis-Boire [2013]. *To Protect And Infect - The Militarization of the Internet*. Talk at 30C3 conference. 29th Dec 2013. http://c3media.vsos.ethz.ch/congress/2013/webm/30c3-5439-en-To_Protect_And_Infect_webm.webm (cited on page 116).

- Guttenberg Freiherr zu, Karl-Theodor [2009]. *Verfassung und Verfassungsvertrag. Konstitutionelle Entwicklungsstufen in den USA und der EU*. German. No longer available. Duncker & Humblot, 2009. ISBN 3428125347 (com, uk) (cited on page 82).
- Hambridge, Sally [1995]. *Netiquette Guidelines*. RFC 1855. Oct 1995 (cited on pages 47, 51).
- handwerk.com [2007]. *Bei Konditormeister Abgesahnt*. 29th Jan 2007. <http://www.handwerk.com/service/archiv/abmahnung-konditor.htm> (cited on page 91).
- Hantke, Manfred [2006]. *Per Google und über AOL in der Abo-Falle Gelandet*. Schwäbisches Tagblatt. 25th Oct 2006. http://www.tagblatt.de/index.php?artikel_id=35615056 (cited on page 125).
- Hay, Stephen [2013]. *Responsive Design Workflow*. New Riders, 5th Apr 2013. ISBN 0321887867 (com, uk) (cited on page 181).
- Heise [2013]. *Skype with Care – Microsoft is Reading Everything You Write*. Heise Online. 14th May 2013. <http://www.h-online.com/security/news/item/Skype-with-care-Microsoft-is-reading-everything-you-write-1862870.html> (cited on page 116).
- Hickson, Ian [2011]. *Acid Tests*. The Web Standards Project. Oct 2011. <http://www.acidtests.org/> (cited on page 180).
- Höhne, Thomas [2006]. *Finger weg von Tauschbörsen*. Interview with Konsument magazine. Oct 2006. <http://www.konsument.at/konsument/document.asp?id=28818> (cited on page 91).
- Hood, Earl [1998]. *Multipurpose Internet Mail Extensions (MIME)*. Jul 1998. <http://www.oac.uci.edu/indiv/ehood/MIME/MIME.html> (cited on page 63).
- Horcicka, Florian [2014]. *NSA zielt auf Österreich*. Format. 7th Apr 2014. <http://format.at/articles/1414/524/374054/nsa-oesterreich> (cited on page 115).
- Hörz, Maël [2012]. *HxD*. Oct 2012. <http://mh-nexus.de/en/hxd/> (cited on page 35).
- Hurst, Mark [2003]. *Managing Incoming Email*. Good Experience. May 2003. <http://web.archive.org/web/20060127013559/http://www.goodexperience.com/reports/e-mail/email-report-goodexperience.pdf> (cited on pages 68, 77).
- Hurst, Mark [2007]. *Bit Literacy: Productivity in the Age of Information and E-mail Overload*. Good Experience Press, Jun 2007. ISBN 0979368103 (com, uk). <http://bitliteracy.org/> (cited on page 68).
- ICANN [2004]. *The Internet Corporation for Assigned Names and Numbers*. Oct 2004. <http://www.icann.org/> (cited on page 16).
- IETF [2006]. *SSH File Transfer Protocol*. IETF Draft. 10th Jul 2006. <https://tools.ietf.org/html/draft-ietf-secsh-filexfer-13> (cited on page 129).
- IETF [2014]. *Public-Key Infrastructure (X.509)*. 10th Oct 2014. <http://datatracker.ietf.org/wg/pkix/> (cited on page 70).
- IFALT [2010]. *Docoloc*. Institut für angewandte Lerntechnologie (IFALT). 2010. <http://www.docoloc.com/> (cited on page 83).
- IFPI [2007]. *IFPI Austria – Verband der Österreichischen Musikwirtschaft*. Oct 2007. <http://www.ifpi.at/> (cited on page 92).
- Internet Ombudsmann [2009]. *Musterbriefe*. 16th Oct 2009. <http://www.ombudsmann.at/schlichtung.php/cat/39/title/Musterbriefe> (cited on page 125).

- iParadigms [2010]. *turnitin*. 2010. <http://turnitin.com/> (cited on pages 83, 84).
- Internet Software Consortium [2014]. *Internet Domain Survey*. Jul 2014. <https://isc.org/solutions/survey/> (cited on pages 4, 8, 9, 14).
- ISO [1991]. *Information Technology — ISO 7-bit Coded Character Set for Information Interchange*. Version ISO/IEC 646:1991. first edition in 1972. International Organization for Standardization (ISO). 1991. http://www.iso.org/iso/catalogue_detail.htm?csnumber=4777 (cited on page 31).
- ISO [2004]. *Data elements and interchange formats – Information interchange – Representation of dates and times (ISO 8601:2004)*. <http://www.w3.org/TR/NOTE-datetime>. International Organization for Standardization (ISO). 2004. <http://iso.org/> (cited on page 68).
- ISO [2008]. (*ISO 32000-1:2008*). http://www.adobe.com/devnet/acrobat/pdfs/PDF32000_2008.pdf. International Organization for Standardization (ISO). 1st Jul 2008. <http://iso.org/> (cited on page 206).
- ISOC [2004]. *The Internet Society*. Sep 2004. <http://www.isoc.org/> (cited on page 16).
- Indigo Stream Technologies [2010]. *Copyscape*. 2010. <http://copyscape.com/> (cited on page 83).
- Internet World Stats [2013a]. *World Internet Users and Population Stats*. 31st Dec 2013. <http://internetworldstats.com/stats.htm> (cited on pages 3, 5, 6).
- Internet World Stats [2013b]. *Internet Growth Statistics*. 31st Dec 2013. <http://internetworldstats.com/emarketing.htm> (cited on page 3).
- Internet World Stats [2013c]. *Internet Usage in Europe*. 31st Dec 2013. <http://internetworldstats.com/stats4.htm> (cited on pages 4, 7).
- Jehl, Scott [2014]. *Responsible Responsive Design*. A Book Apart, Dec 2014. <http://www.abookapart.com/products/responsible-responsive-design> (cited on page 181).
- Jenkov, Jakob [2012]. *HTML5 Tutorial*. 5th Nov 2012. <http://tutorials.jenkov.com/html5/index.html> (cited on page 152).
- Johnson, Joshua [2012]. *What's the Deal With Display: Inline-Block?* design shack. 29th Feb 2012. <http://designshack.net/articles/css/whats-the-deal-with-display-inline-block/> (cited on page 176).
- Jones, Ben [2001]. *RE: [aac-students] Party! Party! Party!* 21st May 2001. <http://groups.yahoo.com/group/anglo-austrian/message/34> (cited on page 65).
- Kantor, Brian and Phil Lapsley [1986]. *Network News Transfer Protocol*. RFC 977. Feb 1986 (cited on page 193).
- Keith, Jeremy [2010]. *HTML5 for Web Designers*. A Book Apart, 2010. ISBN 0984442502 (com, uk). <http://www.abookapart.com/products/html5-for-web-designers> (cited on pages 151, 167).
- Klensin, John C. [2008]. *Simple Mail Transfer Protocol*. RFC 5321. Oct 2008 (cited on page 60).
- Knieper, Marion and Folkert Knieper [2008]. *Marions Kochbuch*. Oct 2008. <http://www.marions-kochbuch.de/> (cited on page 91).
- Koch, Peter-Paul [2006a]. *ppk on JavaScript*. New Riders, Jul 2006. ISBN 0321423305 (com, uk). <http://www.quirksmode.org/book/> (cited on page 144).
- Koch, Peter-Paul [2006b]. *Quirksmode JavaScript*. Oct 2006. <http://www.quirksmode.org/js/contents.html> (cited on page 144).

- Kolla, Patrick M. [2006]. *Spybot Search and Destroy*. Oct 2006. <http://spybot.info/> (cited on page 120).
- Konsument [2006]. *Streitgespräch: Pro und kontra Musikdownloads*. Oct 2006. <http://www.konsument.at/konsument/document.asp?id=28958> (cited on page 92).
- Korpela, Jukka [2007]. *A Tutorial on Character Code Issues*. 13th Jul 2007. <http://www.cs.tut.fi/~jkorpela/chars.html> (cited on page 43).
- Kozierok, Charles M. [2005]. *The TCP/IP Guide*. No Starch Press, Oct 2005. ISBN 159327047X (com, uk). <http://www.tcpipguide.com/> (cited on page 104).
- Kremer, Sascha [2007]. *Abmahnung für Brötchen und Long Island Ice Tea*. 14th Jan 2007. <http://www.kremer-legal.com/2007/01/14/abmahnung-fur-brotchen-und-long-island-ice-tea/> (cited on page 91).
- Kröner, Peter [2011]. *HTML5. Webseiten innovativ und zukunftssicher*. 2nd edition. Open Source Press, 28th Feb 2011. ISBN 3941841343 (com, uk) (cited on page 152).
- Kurose, James F. and Keith W. Ross [2012]. *Computer Networking: A Top-down Approach*. 6th edition. Pearson, 3rd May 2012. ISBN 0273768964 (com, uk) (cited on pages 1, 93).
- Laborenz, Kai [2011]. *CSS: Das umfassende Handbuch. Aktuell zu CSS3 und HTML5*. Galileo Computing, 28th Sep 2011. ISBN 3836217252 (com, uk) (cited on page 168).
- Lee, Homin K., Tal Malkin and Erich Nahum [2007]. “Cryptographic Strength of SSL/TLS Servers: Current and Recent Practices”. In: *Proc. 7th ACM SIGCOMM Conference on Internet Measurement (IMC’07)*. ACM. San Diego, California, USA, Oct 2007, pages 83–92. doi:10.1145/1298306.129831. <http://www.imconf.net/imc-2007/papers/imc130.pdf> (cited on page 148).
- Levine, John R. and Margaret Levine Young [2011]. *The Internet For Dummies*. 13th edition. Wiley, 9th Dec 2011. ISBN 1118096142 (com, uk) (cited on page 1).
- Levine, John R., Margaret Levine Young and Ferenc Nagy [2013]. *Internet für Dummies*. German. 13th edition. Wiley, 3rd Jul 2013. ISBN 352770986X (com, uk) (cited on page 1).
- Levine, Matthew [2006]. *In Search of the Holy Grail*. A List Apart. 30th Jan 2006. <http://alistapart.com/article/holygrail> (cited on page 176).
- Lugo, Israel G. and Don Parker [2005]. *Software Firewalls: Made of Straw?* Security Focus. 8th Jun 2005. <http://www.securityfocus.com/infocus/1839/> (cited on page 119).
- Lyon, Barrett [2009]. *The Opte Project*. 28th Sep 2009. <http://www.opte.org/maps/> (cited on pages 3, 4).
- Mahloujian, Mohammad [2011]. *IPv4 addresses? Sorry, We are closed!* 5th Feb 2011. <http://amirit.wordpress.com/2011/02/05/ipv4-addresses-sorry-we-are-closed/> (cited on page 105).
- Mann, Merlin [2007]. *Inbox Zero*. Google Tech Talk. 23rd Jul 2007. <http://video.google.com/videoplay?docid=973149761529535925> (cited on page 68).
- Mann, Merlin [2010]. *Original Inbox Zero Articles*. 43 Folders Series. 8th Oct 2010. <http://inboxzero.com/articles/> (cited on page 68).
- Marcotte, Ethan [2011]. *Responsive Web Design*. A Book Apart, 2011. ISBN 098444257X (com, uk). <http://www.abookapart.com/products/responsive-web-design> (cited on page 181).
- Matthews, Jeanna [2005]. *Computer Networks: Internet Protocols in Action*. Wiley, Jan 2005. ISBN 0471661864 (com, uk) (cited on page 104).

- May, Simon [2009]. *The Rough Guide to Windows 7*. Rough Guides, 1st Sep 2009. ISBN 1848362773 (com, uk) (cited on page 1).
- McKenzie, Alexander A. [1973-02-16]. *File Transfer Protocol*. RFC 454. 1973-02-16 (cited on pages 190, 192).
- Mead, Ian D. [2014]. *UltraEdit*. 1st Oct 2014. <http://ultraedit.com/> (cited on page 34).
- Melvin, John T. [1971]. *A First Cut at a Proposed Telnet Protocol*. RFC 97. 15th Feb 1971. <http://asg.andrew.cmu.edu/rfc/rfc97.pdf> (cited on page 192).
- Meyer, Eric [2011]. *Reset CSS*. 26th Jan 2011. <http://meyerweb.com/eric/tools/css/reset/> (cited on page 178).
- Microsoft [2005]. *Tweak UI*. Microsoft PowerToys for Windows XP. Nov 2005. <http://www.microsoft.com/windowsxp/downloads/powertoys/xppowertoys.msp> (cited on page 126).
- Miller, Joseph B. [2014]. *Internet Technologies and Information Services*. 2nd edition. Libraries Unlimited, 26th Aug 2014. ISBN 1610694732 (com, uk) (cited on page 1).
- Moore, Tyler and Richard Clayton [2007]. “An Empirical Analysis of the Current State of Phishing Attack and Defence”. In: *Proc. Sixth Workshop on the Economics of Information Security (WEIS 2007)*. Pittsburgh, Pennsylvania, USA, Jun 2007. <http://www.cl.cam.ac.uk/~twm29/weis07-phishing.pdf> (cited on page 121).
- Moshchuk, Alexander et al. [2006]. “A Crawler-based Study of Spyware on the Web”. In: *Proc. 13th Annual Network and Distributed System Security Symposium (NDSS 2006)*. San Diego, California, USA: Internet Society, –32nd Feb 2006. <http://www.isoc.org/isoc/conferences/ndss/06/proceedings/papers/spycrawler.pdf> (cited on page 121).
- Münz, Stefan and Clemens Gull [2011]. *HTML5 Handbuch*. 2nd edition. Franzis Verlag, 28th Nov 2011. ISBN 3645601511 (com, uk). <http://webkompetenz.wikidot.com/docs:html-handbuch> (cited on page 152).
- Murray, Eric [2001]. *Changes in Deployment of Cryptography*. Usenix Security Symposium 2001. 16th Aug 2001. <http://www.usenix.org/events/sec01/murray/> (cited on page 148).
- Myers, John G. and Marshall T. Rose [1996]. *Post Office Protocol - Version 3*. RFC 1939. May 1996 (cited on pages 60, 61).
- Naraine, Ryan [2006]. *Hacker Discovers Adobe PDF Back Doors*. eWeek.com. 15th Sep 2006. <http://www.eweek.com/article2/0,1895,2016606,00.asp> (cited on page 117).
- Nentwich, Michael [2008]. *KB:Law Knowledge Base Copyright Law*. Demokratiezentrum Wien. Oct 2008. <http://kb-law.info/kbc/kbc.php> (cited on page 85).
- Netcraft [2011]. *Web Server Survey*. Oct 2011. <http://news.netcraft.com/archives/category/web-server-survey/> (cited on page 140).
- NIVA [1996]. *Placement of Punctuation and Quotation Marks*. Writer’s Block. May 1996. <http://www.writersblock.ca/tips/monthtip/tipmay96.htm> (cited on page 87).
- Novak, Brennan and Barnaby Walters [2014]. *Indie Webify.Me*. 14th Oct 2014. <http://indiewebify.me/> (cited on page 213).
- NRO [2011]. *Free Pool of IPv4 Address Space Depleted*. Number Resource Organization. 3rd Feb 2011. <http://www.nro.net/news/ipv4-free-pool-depleted> (cited on page 105).
- Odlyzko, Andrew M. [2009]. *Minnesota Internet Traffic Studies (MINTS)*. 31st Dec 2009. <http://www.dtc.umn.edu/mints/> (cited on pages 4, 9).

- OECD [2013]. *Broadband Subscribers per 100 Inhabitants in OECD Countries*. Dec 2013. <http://oecd.org/sti/ict/broadband> (cited on pages 10–12).
- OECD [2014]. *OECD Broadband Portal*. 22nd Jul 2014. <http://oecd.org/sti/ict/broadband> (cited on page 94).
- OGH [2009]. *OGH Entscheidung 4Ob41/09x*. 14th Jul 2009. <http://www.ris.bka.gv.at/Jus/> (cited on page 92).
- Oikarinen, Jarkko and Darren Reed [1993]. *Internet Relay Chat Protocol*. RFC 1459. May 1993 (cited on page 193).
- On the Internet: Comprehensive, New Perspectives [2012]. *Gary P. Schneider and Jessica Evans*. 9th edition. Cengage Learning, 1st May 2012. ISBN 1111529116 (com, uk) (cited on pages 1, 93).
- Ookla [2011a]. *Net Index*. 5th Oct 2011. <http://netindex.com/> (cited on pages 13, 15).
- Ookla [2011b]. *Speedtest.net*. 5th Oct 2011. <http://speedtest.net/> (cited on page 13).
- OpenPGP [2014]. *The OpenPGP Alliance*. 10th Oct 2014. <http://openpgp.org/> (cited on page 70).
- ORF [2013]. *Austrian Internet Monitor*. Dec 2013. <http://mediaresearch.orf.at/internet.htm> (cited on page 14).
- O’Toole, Catie [2010]. *Seventh-Grade North Syracuse Student Suspended, 25 Others Disciplined for Facebook Page About Teacher*. The Syracuse Post-Standard. 24th Jan 2010. http://www.syracuse.com/news/index.ssf/2010/01/seventh-grade_north_syracuse_s.html (cited on page 115).
- ÖWA [2014a]. *Österreichische Webanalyse (ÖWA)*. Aug 2014. <http://oewa.at/> (cited on page 14).
- ÖWA [2014b]. *ÖWA Basic Browserstatistik*. Aug 2014. <http://oewa.at/index.php?id=14885> (cited on pages 13, 14).
- Page, Lawrence [2001]. *Method for node ranking in a linked database*. US Patent 6285999. Filed 09 Jan 1998, issued 04 Sep 2001. Stanford University, Sep 2001 (cited on page 75).
- Palmer, Chris and Yan Zhu [2013]. *How to Deploy HTTPS Correctly*. EFF. 12th Dec 2013. <http://eff.org/https-everywhere/deploying-https> (cited on page 166).
- PBS [1998]. *Nerds 2.0.1: A Brief History of the Internet*. Oregon Public Broadcasting. 1998. <http://www.pbs.org/opb/nerds2.0.1/> (cited on page 190).
- pdfforge [2010]. *PDFCreator*. Oct 2010. <http://www.pdfforge.org/products/pdfcreator> (cited on pages 146, 206).
- Peter, Reinfried [2007]. *Urheberrecht: Musikuploads, Filme, ...* Posting to Newsgroup tu-graz.network. 8th Mar 2007. <http://newsarchiv.tugraz.at/browse/tu-graz.network/msg03754.html> (cited on page 92).
- Phillips, Chris [2013]. *Six Things Small Networks Need to Know About IPv6*. VentureBeat. 17th Aug 2013. <http://venturebeat.com/2013/08/17/ipv4-ipv6-apocalypse/> (cited on page 107).
- Postel, Jon and Joyce K. Reynolds [1985-10]. *File Transfer Protocol (FTP)*. RFC 959. 1985-10 (cited on page 129).
- Postel, Jonathan B. [1972]. *Telnet Protocol*. RFC 318. 3rd Apr 1972 (cited on pages 190, 192).
- Postel, Jonathan B. [1981]. *Simple Mail Transfer Protocol*. RFC 788. Nov 1981 (cited on page 193).

- Procter, Margaret [2010]. *How Not to Plagiarize*. University of Toronto. 14th Jul 2010. <http://www.writing.utoronto.ca/images/stories/Documents/how-not-to-plagiarize.pdf> (cited on page 85).
- Quittner, Joshua [1994]. “Billions Registered”. *Wired Magazine* 2.10 (Oct 1994). <http://wired.com/wired/archive/2.10/mcdonalds.html> (cited on page 110).
- R [2009]. *The R Project for Statistical Computing*. 2009. <http://www.r-project.org/> (cited on page 6).
- Raymond, Eric Steven and Rick Moen [2012]. *How To Ask Questions The Smart Way*. 19th Jun 2012. <http://www.catb.org/esr/faqs/smart-questions.html> (cited on page 48).
- Reisman, Sorel [2005]. “Plagiarism or Ignorance? You Decide”. *IT Professional* 7.1 (Jan 2005), pages 7–8. ISSN 1520-9202. doi:10.1109/MITP.2005.16 (cited on page 83).
- Rogers, Ian [2002]. *The Google Pagerank Algorithm and How It Works*. May 2002. <http://www.ianrogers.net/google-page-rank/> (cited on page 75).
- Rose, Marshall [1988]. *Post Office Protocol - Version 3*. RFC 1081. Nov 1988 (cited on page 192).
- Russinovich, Mark [2005]. *Sony, Rootkits and Digital Rights Management Gone Too Far*. SysInternals. 31st Oct 2005. <http://www.sysinternals.com/blog/2005/10/sony-rootkits-and-digital-rights.html> (cited on page 126).
- Schavan, Annette [1980]. *Person und Gewissen: Studien zu Voraussetzungen, Notwendigkeit und Erfordernissen heutiger Gewissensbildung*. German. No longer available. Rita G. Fischer Verlag, 1980. ISBN 3883232203 (com, uk) (cited on page 82).
- Schmidbauer, Franz [2007]. *FAQ zum Urheberrecht, Tauschbörsen*. Internet4Jurists. Oct 2007. http://www.internet4jurists.at/urh-marken/faq_urh1a.htm (cited on page 92).
- Schriver, Karen A. [1997]. *Dynamics in Document Design*. Wiley, 1997. ISBN 0471306363 (com, uk) (cited on page 141).
- Schroeder, Stan [2011]. *ICANN Approves New Top-Level Domains, So Prepare For .whatever*. Mashable. 20th Jun 2011. <http://mashable.com/2011/06/20/icann-top-level-domains/> (cited on page 109).
- Schulzrinne, Henning [2004]. *The tel URI for Telephone Numbers*. RFC 3966. Dec 2004 (cited on page 160).
- Scott [1991]. *IRC Log*. Jan 1991. <http://www.ibiblio.org/pub/academic/communications/logs/Gulf-War/Jan-18-Scott> (cited on page 195).
- Seifert, Christian [2007]. *Know Your Enemy: Behind the Scenes of Malicious Web Servers*. The HoneyNet Project. 7th Nov 2007. <http://www.honeynet.org/papers/wek/> (cited on page 121).
- Seifert, Christian et al. [2007]. *Know Your Enemy: Malicious Web Servers*. The HoneyNet Project. 9th Aug 2007. <http://www.honeynet.org/papers/mws/> (cited on pages 121, 125).
- ServerWatch [2011]. *ServerWatch*. Oct 2011. <http://serverwatch.com/> (cited on page 140).
- Shea, Dave [2012]. *Zen Garden: The Beauty of CSS Design*. 5th Nov 2012. <http://csszengarden.com/> (cited on page 152).
- Shipley, David and Will Schwalbe [2010]. *Send: Why People Email So Badly and How to Do It Better*. Revised. Vintage, 24th Aug 2010. ISBN 030727599X (com, uk) (cited on page 57).

- Skirvin, Tim [1999]. *Cancel Messages: Frequently Asked Questions*. 30th Sep 1999. <http://wiki.killfile.org/projects/usenet/faqs/cancel/> (cited on pages 45, 55).
- Slegg, Jennifer [2007]. *Choosing Your Anchor Text for Incoming Links*. 17th May 2007. <http://www.jenniferslegg.com/2007/05/17/choosing-your-anchor-text-for-incoming-links/> (cited on page 75).
- Standard [2007]. *Googler, Copy-Paster und Plagiatsjäger*. 18th Mar 2007. <http://derstandard.at/2808256> (cited on page 83).
- Statista [2013]. *Number of Daily Activations of Android Devices from August 2010 to April 2013*. Apr 2013. <http://statista.com/statistics/278305/daily-activations-of-android-devices/> (cited on page 13).
- Stevens, W. Richard [1993]. *TCP/IP Illustrated, Volume 1, The Protocols*. Addison-Wesley, 1993. ISBN 0201633469 (com, uk) (cited on page 104).
- Stiegert, Heiko [2011]. *Modernes Webdesign mit CSS*. Galileo Design, 28th Jul 2011. ISBN 3836216663 (com, uk) (cited on page 168).
- Stockdale, Terry A. [2007]. *Hidden File Extensions in Windows*. Terry's Computer Tips. Nov 2007. <http://www.terryscomputertips.com/computers/hidden-file-extensions.php> (cited on page 118).
- Sublime [2014]. *Sublime Text*. 1st Oct 2014. <http://sublimetext.com/> (cited on page 34).
- SURL [2006]. *Software Usability Research Laboratory (SURL)*. Oct 2006. <http://psychology.wichita.edu/surl/> (cited on page 143).
- Symantec [2009]. *The State of Spam Monthly Report*. Oct 2009. http://eval.symantec.com/mktginfo/enterprise/other_resources/b-state_of_spam_report_10-2009.en-us.pdf (cited on page 66).
- Tapscott, Don [2007]. *Wikinomics - Winning the Enterprise 2.0*. Talk at Enterprise 2.0 2007. 19th Jun 2007. <http://enterprise2conf.vportal.net/> (cited on page 129).
- Taylor, Dave [2007]. *What's the difference between BitTorrent, Limewire and Kazaa?* 5th Feb 2007. http://www.askdavetaylor.com/difference_between_bittorrent_limewire_and_kazaa.html (cited on page 91).
- Tetlaw, Andrew [2011]. *Give Floats the Flick in CSS Layouts*. sitepoint. 24th Mar 2011. <http://www.sitepoint.com/give-floats-the-flick-in-css-layouts/> (cited on page 176).
- TUGraz [2008a]. *Ethischer Kodex der TU Graz*. 1st Oct 2008. http://mibla.tugraz.at/08_09/Stk_1/EthikKodex1_0_1008.pdf (cited on page 90).
- TUGraz [2008b]. *Richtlinie des Rektorates der Technischen Universität Graz zur Sicherung guter wissenschaftlicher Praxis und zur Vermeidung von Fehlverhalten in der Wissenschaft*. 1st Oct 2008. http://mibla.tugraz.at/08_09/Stk_1/GuteWissPraxisV1_1_1008.pdf (cited on page 90).
- TUGraz [2008c]. *University Gazette*. 1st Oct 2008. http://mibla.tugraz.at/08_09/Stk_1/1.html (cited on page 90).
- Total Validator [2012]. *Total Validator*. 5th Nov 2012. <http://totalvalidator.com/> (cited on page 157).
- Tverskov, Jesper [2012]. *Benefits of Polyglot XHTML5*. 27th Sep 2012. <http://www.xmlplease.com/xhtml/xhtml5polyglot/> (cited on pages 152, 156).

- Twitter [2010]. *The Evolving Ecosystem*. Twitter Blog. 2nd Sep 2010. <http://blog.twitter.com/2010/09/evolving-ecosystem.html> (cited on page 13).
- Unicode Consortium [2012a]. *Unicode*. Oct 2012. <http://unicode.org/versions/Unicode6.2.0/> (cited on page 32).
- Unicode Consortium [2012b]. *Unicode 6.2 Character Code Charts*. Oct 2012. <http://unicode.org/charts/> (cited on page 33).
- University of Washington [2009]. *The IMAP Connection*. 2nd Oct 2009. <http://www.imap.org/> (cited on pages 60, 61).
- UrhG [2010]. *Urheberrechtsgesetz, Fassung vom 14.01.2010*. Rechtsinformationssystem des Bundes (RIS). 2010. <http://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10001848> (cited on pages 85, 91).
- Verbraucherrecht.at [2009]. *Tipps gegen Internet Abzocke*. 16th Oct 2009. <http://www.verbraucherrecht.at/development/typo/test/index.php?id=718> (cited on page 125).
- Vernacchia, Alex [2011]. *Using CSS Clearfix*. 18th Nov 2011. <http://babble.byvernacchia.com/2011/11/18/using-css-clearfix.html> (cited on pages 174, 175).
- W3C [1999]. *HTML 4.01 Specification*. World Wide Web Consortium. 24th Dec 1999. <http://w3.org/TR/html401/> (cited on page 152).
- W3C [2002]. *XHTML 1.0 The Extensible HyperText Markup Language*. World Wide Web Consortium. 1st Aug 2002. <http://w3.org/TR/xhtml1/> (cited on page 152).
- W3C [2007a]. *The World Wide Web Consortium (W3C)*. Oct 2007. <http://w3.org/> (cited on page 16).
- W3C [2007b]. *XHTML 1.1 - Module-Based XHTML*. World Wide Web Consortium. 16th Feb 2007. <http://w3.org/TR/xhtml11/> (cited on page 152).
- W3C [2008]. *The W3C CSS Validation Service*. Jan 2008. <http://jigsaw.w3.org/css-validator/> (cited on page 157).
- W3C [2011a]. *Cascading Style Sheets Level 2 Revision 1 (CSS 2.1) Specification*. W3C Recommendation. 7th Jun 2011. <http://w3.org/TR/2011/REC-CSS2-20110607> (cited on pages 152, 178).
- W3C [2011b]. *Official W3C CSS Test Suites*. 4th Nov 2011. <http://www.w3.org/Style/CSS/Test/> (cited on page 180).
- W3C [2011c]. *Scalable Vector Graphics (SVG) 1.1*. W3C Recommendation. 16th Aug 2011. <http://w3.org/TR/SVG11/> (cited on page 204).
- W3C [2012a]. *Nu Markup Validation Service*. 5th Nov 2012. <http://validator.w3.org/nu/> (cited on pages 152, 157).
- W3C [2012b]. *Polyglot Markup: HTML-Compatible XHTML Documents*. 25th Oct 2012. <http://www.w3.org/TR/html-polyglot/> (cited on pages 152, 156).
- W3C [2012c]. *WebPlatform.org*. 5th Nov 2012. <http://webplatform.org/> (cited on page 152).
- W3C [2013]. *Scalable Vector Graphics (SVG) 2*. W3C Working Draft. 18th Jun 2013. <http://w3.org/TR/SVG2/> (cited on page 204).
- W3C [2014a]. *CSS Current Work*. 31st Oct 2014. <http://w3.org/Style/CSS/current-work.en.html> (cited on pages 152, 168).

- W3C [2014b]. *CSS Flexible Box Layout Module Level 1*. 28th Oct 2014. <http://dev.w3.org/csswg/css-flexbox/> (cited on page 176).
- W3C [2014c]. *HTML5: A Vocabulary and Associated APIs for HTML and XHTML*. W3C Recommendation. 28th Oct 2014. <http://w3.org/TR/html5> (cited on pages 152, 163).
- Walton, Philip [2013]. *Holy Grail Layout*. Solved by Flexbox. 16th Nov 2013. <http://philipwalton.github.io/solved-by-flexbox/demos/holy-grail/> (cited on page 176).
- Walton, Philip [2014]. *Solved by Flexbox*. 28th Oct 2014. <http://philipwalton.github.io/solved-by-flexbox/> (cited on page 176).
- Wang, Yi-Min et al. [2006]. “Automated Web Patrol with Strider HoneyMonkeys: Finding Web Sites That Exploit Browser Vulnerabilities”. In: *Proc. 13th Annual Network and Distributed System Security Symposium (NDSS 2006)*. San Diego, California, USA: Internet Society, – 32nd Feb 2006. <http://www.isoc.org/isoc/conferences/ndss/06/proceedings/papers/honeymonkeys.pdf> (cited on page 121).
- Ward, Mark [2006]. *Tracking Down Hi-Tech Crime*. BBC News. 8th Oct 2006. <http://news.bbc.co.uk/2/hi/technology/5414502.stm> (cited on page 118).
- WaSP [2011]. *The Web Standards Project*. Oct 2011. <http://www.webstandards.org/> (cited on page 180).
- Weber, Stefan [2006]. *Das Google-Copy-Paste-Syndrom*. dpunkt, Nov 2006. ISBN 3936931372 (com, uk) (cited on page 82).
- Weber-Wulff, Debora [2010]. *Fremde Federn: Plagiat Ressourcen*. 2010. <http://plagiat.htw-berlin.de/> (cited on pages 82, 85).
- Weber-Wulff, Debora [2014]. *False Feathers: A Perspective on Academic Plagiarism*. Springer, 5th Mar 2014. 200 pages. ISBN 3642399606 (com, uk) (cited on page 81).
- web-sniffer [2009]. *View HTTP Request and Response Header*. 31st Oct 2009. <http://web-sniffer.net/> (cited on pages 137, 138).
- Whalen, David [2002]. *The Unofficial Cookie FAQ*. Oct 2002. <http://www.cookiecentral.com/faq/> (cited on page 146).
- Wichary, Marcin [2012]. *HTML5 Presentation*. 5th Nov 2012. <http://slides.html5rocks.com/> (cited on page 152).
- Wikimedia Commons [2009]. *Image:Spam with cans.jpeg*. Oct 2009. http://commons.wikimedia.org/wiki/Image:Spam_with_cans.jpeg (cited on page 67).
- Wikipedia [2007a]. *File Sharing*. Oct 2007. http://en.wikipedia.org/wiki/File_sharing (cited on page 91).
- Wikipedia [2007b]. *Personal Firewall*. Oct 2007. http://en.wikipedia.org/wiki/Personal_firewall (cited on page 119).
- Wikipedia [2007c]. *Plagiarism*. Oct 2007. <http://en.wikipedia.org/wiki/Plagiarism> (cited on page 85).
- Wikipedia [2007d]. *TCP/IP model*. Sep 2007. http://en.wikipedia.org/wiki/TCP/IP_model (cited on page 104).
- Wikipedia [2008a]. *International Standard Book Number*. Oct 2008. <http://en.wikipedia.org/wiki/ISBN> (cited on page 77).

- Wikipedia [2008b]. *Internet Message Access Protocol*. Oct 2008. http://en.wikipedia.org/wiki/Internet_Message_Access_Protocol (cited on pages 60, 61).
- Wikipedia [2008c]. *Post Office Protocol*. Oct 2008. http://en.wikipedia.org/wiki/Post_Office_Protocol (cited on pages 60, 61).
- Wikipedia [2008d]. *Simple Mail Transfer Protocol*. Oct 2008. http://en.wikipedia.org/wiki/Simple_Mail_Transfer_Protocol (cited on page 60).
- Wikipedia [2009a]. *Notepad++*. Oct 2009. <http://en.wikipedia.org/wiki/Notepad++> (cited on page 34).
- Wikipedia [2009b]. *Spam*. Oct 2009. <http://en.wikipedia.org/wiki/Spam> (cited on page 67).
- Wikipedia [2009c]. *Text Editor*. Oct 2009. http://en.wikipedia.org/wiki/Text_editor (cited on page 33).
- Wikipedia [2010a]. *Crimson Editor*. Oct 2010. http://en.wikipedia.org/wiki/Crimson_Editor (cited on page 34).
- Wikipedia [2010b]. *Digital Object Identifier*. 8th Oct 2010. http://en.wikipedia.org/wiki/Digital_object_identifier (cited on page 79).
- Wikipedia [2010c]. *Word Processor*. Oct 2010. http://en.wikipedia.org/wiki/Word_processor (cited on page 43).
- Wikipedia [2012a]. *ASCII*. Oct 2012. <http://en.wikipedia.org/wiki/ASCII> (cited on page 31).
- Wikipedia [2012b]. *Character Encoding*. Oct 2012. http://en.wikipedia.org/wiki/Character_encoding (cited on page 31).
- Wikipedia [2012c]. *ISO/IEC 646*. Oct 2012. http://en.wikipedia.org/wiki/ISO/IEC_646 (cited on page 31).
- Wikipedia [2012d]. *Percent-Encoding*. 6th Nov 2012. <http://en.wikipedia.org/wiki/Percent-encoding> (cited on page 159).
- Wikipedia [2012e]. *Unicode*. Oct 2012. <http://en.wikipedia.org/wiki/Unicode> (cited on page 32).
- Wikipedia [2012f]. *Unicode*. Oct 2012. <http://en.wikipedia.org/wiki/Utf-32> (cited on page 33).
- Wikipedia [2012g]. *UTF-16*. Oct 2012. <http://en.wikipedia.org/wiki/Utf-16> (cited on page 33).
- Wikipedia [2012h]. *UTF-8*. Oct 2012. <http://en.wikipedia.org/wiki/Utf-8> (cited on pages 32, 33).
- Wikipedia [2012i]. *XML*. 9th Nov 2012. <http://en.wikipedia.org/wiki/XML> (cited on page 156).
- Wikipedia [2012j]. *XSLT*. 7th Nov 2012. <http://en.wikipedia.org/wiki/XSLT> (cited on page 156).
- Wikipedia [2014a]. *File Transfer Protocol*. 22nd Oct 2014. http://en.wikipedia.org/wiki/File_Transfer_Protocol (cited on page 129).
- Wikipedia [2014b]. *FTP over SSL*. 26th Sep 2014. <http://en.wikipedia.org/wiki/FTPS> (cited on page 129).
- Wikipedia [2014c]. *GeoCities*. 28th Oct 2014. <http://en.wikipedia.org/wiki/GeoCities> (cited on page 213).
- Wikipedia [2014d]. *GNU Privacy Guard*. 9th Sep 2014. http://en.wikipedia.org/wiki/GNU_Privacy_Guard (cited on page 70).

- Wikipedia [2014e]. *Internet Access*. 15th Oct 2014. http://en.wikipedia.org/wiki/Internet_access (cited on pages 93, 94).
- Wikipedia [2014f]. *SSH File Transfer Protocol (SFTP)*. 3rd Oct 2014. http://en.wikipedia.org/wiki/SSH_file_transfer_protocol (cited on page 129).
- Wikipedia [2014g]. *STARTTLS*. 20th Sep 2014. <http://en.wikipedia.org/wiki/STARTTLS> (cited on page 70).
- Wikipedia [2014h]. *Transport Layer Security*. 8th Oct 2014. http://en.wikipedia.org/wiki/Transport_Layer_Security (cited on pages 70, 148).
- Wikipedia [2014i]. *Web of Trust*. 6th Oct 2014. http://en.wikipedia.org/wiki/Web_of_trust (cited on page 70).
- Wikipedia [2014j]. *X.509*. 7th Oct 2014. <http://en.wikipedia.org/wiki/X.509> (cited on page 70).
- Wroblewski, Luke [2011a]. *Mobile First*. WebExpo Prague. 23rd Sep 2011. <http://vimeo.com/25548033> (cited on page 13).
- Wroblewski, Luke [2011b]. *Mobile First. A Book Apart*, 18th Oct 2011. ISBN 1937557022 (com, uk). <http://abookapart.com/products/mobile-first> (cited on page 13).
- Yahoo [2011]. *Mobile Modes*. Aug 2011. http://l.yimg.com/a/i/us/ayc/pdf/mobile_modes_whitepaper.pdf (cited on page 14).
- Zeldman, Jeffrey and Ethan Marcotte [2009]. *Designing with Web Standards*. 3rd edition. New Riders, Oct 2009. ISBN 0321616952 (com, uk) (cited on pages 151, 167).
- ZID [2010a]. *Newsgroup-Charta der TU Graz*. ZID der TU Graz. 1st Oct 2010. <http://portal.tugraz.at/portal/page/portal/zid/netzwerk/dienste/newsgroups/charta> (cited on pages 41, 54).
- ZID [2010b]. *Newsgroups: Usenet-Server der TU Graz*. ZID der TU Graz. 1st Oct 2010. <http://portal.tugraz.at/portal/page/portal/zid/netzwerk/dienste/newsgroups/> (cited on page 30).
- ZID [2010c]. *TU Graz Cancelbot*. ZID der TU Graz. 1st Oct 2010. <http://portal.tugraz.at/portal/page/portal/zid/netzwerk/dienste/newsgroups/charta/verstoss/cancelbot> (cited on page 55).
- ZID [2014]. *E-Mail und digitale Signatur*. ZID der TU Graz. 10th Oct 2014. <http://email.tugraz.at/zertifikat> (cited on pages 70, 71).
- Zobel, Justin [2004]. *Writing for Computer Science*. 2nd edition. Springer, 2004. ISBN 1852338024 (com, uk) (cited on page 81).
- ZyTrax [2014]. *TLS/SSL and SSL (X.509) Certificates*. ZyTrax Survival Guide. 4th Jul 2014. <http://zytrax.com/tech/survival/ssl.html> (cited on pages 148, 166).