ABC (Australian Broadcasting Corporation), 106
abortion drug, 160–1, 186
abortion service, 161, 181, 183, 186, 188
.ac.uk, 24, 29–34, 34–42, 43
access, barriers to, 246
ACM Web Science conference, 238
ADHO (Alliance of Digital Humanities Organisations), 238
Adobe 177
advocacy, digital, 239
Africa, 3, 7, 15
Al-Thawra 215
Alexa, 6
algorithms, 99
America Online, 175
Apache Hive, 28, 43
API (Application Program Interface), 189, 246
apps, 245
Arab Spring, 215–6
ARC, 7
Archbishop of Canterbury, 191, 193–202
Archbishop of York, 198–9, 202
archive, 170, 189
Archive Team, 137, 140
Geocities Snapshot, 9, 10
Archive-it, 9, 11
Archivethe.Net, 9
Armitage, David, 241
Aust, Rowan, 221–2, 233
Australia, 7
authenticity, 205, 210, 212
awards, role in community, 153, 154
Bangemann, 118
BBC (British Broadcasting Corporation)
  Jimmy Savile scandal, 221–2, 233
  news website, 102, 105, 107–15
  online, 106–7
  radio, 194
  television, 19
Beat literature, 222–3, 231
Belgium, 8
Beverly Hills Internet, 138
bias, 45, 56, 59–61
Bibliothèque Nationale de France Web Archives, 8
Birt, John, 106
Bitly, 59
blogosphere, 164, 188
blog (weblog), 88, 89, 162–4, 188, 197, 205, 212–5, 218
born digital big data, 247–8
Brazil, 3
British Library
  homepage of, 240, 245
  Internet Archive data, 28, 47, 52, 220
  own crawls of UK websites, 59
  SHINE, 192, 200
  UK Web Archive, 1, 8, 10, 26, 191, 192, 200, 246
  visualization, 242–3
British National Party, 197, 199–202
britishblogs.co.uk, 197
broad crawl, 66–7, 73–8
BUDDAH (Big UK Domain Data for the Arts and Humanities) project, 239, 247
Burroughs, William, 223
Canada, 6
Carey, George (See also Archbishop of Canterbury), 194
cat, image of, 243
Caterpillar boots, image of, 243
ccTLD (Country code Top-Level Domain), 63–7, 76–9, 192
chav, 242
China, 3
Church of England, 190, 191, 193, 196, 201
  collective action, 163
  collective identity, 163
Columbia University Libraries, 8
  comment threads, 201
commerce, 186
Common Crawl, 9, 246
  community leaders, 151, 152
  community, 137–8, 140–1, 143–9, 151–8
  completeness, 47–8, 52–4, 58–61
  connected component, 173
consumerism, digital, 241
content analysis, 161, 164–5, 170, 188
.co.uk, 29–34, 42
coverage, of an individual website, 47–8, 52–4, 58–61
  coverage, web archives, 26
Cran, Rona, 222–3, 231, 235
crawl profile, 192
credit crunch, 242
croatia, 7
cyberbalkinization, 164
data loss, 244–5
data protection, 247–8
data, open, 245, 246
data, portability of, 246
deduplication, policies on, 192, 195
deleted content, 53, 55, 60
delicious, 58
denmark, 7, 8, 10, 63–80
density, network, 173
deswarte, richard, 223–4, 231, 232, 234–6
development of the web, 62–80
diaspora, 205
digital dark age, 244
digital humanities, 238, 239, 242, 246
diplomatic, 240
distribution, 55
dk registry, 67, 72–6
domain name registry, 67, 72–6
ephemeralty, of data, 245–6
european union, 223
event crawl, 66
everyday life, 4
evidence, survival of, 244

facebook
commercial service provider, 243
links to, 94
news, 98
screen shots of, 9
size, 3
social media, 160, 162–3, 174, 188, 245
taquwacore 205, 210, 215, 218
farage, nigel, 223
flickr, 162
france, 3, 7, 118–23, 126–33
france télécom, 126
gatekeeping, 101
gockeytes, 137–58, 175
geographic distance, 38–42
germany, 3
google, 1, 58, 163, 165–6, 168–71, 186–9, 235
.gov.uk, 29–34
gtld (generic top-level domain), 65
guestbooks, role in community, 155, 156
guldí, jo, 241

harvard university library web archive
collection service, 8
health 165, 180–2
heretrix, 7
history manifesto, the, 241
historical method, 195–7, 202–3
history of the web, 23
hitchcock, tim, 241
homophily, 164, 173
hope, david, 198 (see also archbishop of york)
html (hyper/text markup language), 86, 141, 142, 152, 153
huc-hepher, saskia, 224–5, 231

human rights web archive @ columbia
university, 9, 10
hyperlink, 24, 25, 28, 34
hyperlink network, 159, 161–5, 170–1, 174–5, 187–9

iceland, 7
iiipc (international internet preservation consortium), 7
image analysis, 149, 150, 151
inclusiveness, network, 173
indegree, network 173–7
india, 3
information highway, 123–4
information public good, 163
information studies, 2
instagram, 243
institut national de l'audiovisuel, 8
institute of historical research, university of london, 220, 244, 245
institutional history, digital, 241
internet archive, 1, 10, 26, 84, 107, 140, 239, 244–5
archive-it, 9
biases of, 45–6
danish web, 62–4, 66, 72–5, 78–80
establishing of, 6, 51
geoCities, 140, 158
uk web domain, 27–8, 191, 195, 197–8, 201, 220, 231
internet memory research, 9
issuecrawler, 187
issy-les-moulineaux, 125–6

japan, 3, 7
javascript, 51, 60
jisc (joint information systems committee), 52, 220
jisc uk web domain dataset, 191
host link graph, 192, 193, 195, 196
john johnson collection of political ephemera, 243

kay, alison, 225–6, 231, 233, 236
/kernel density, 55 (figure 2.5), 56 (figure 2.6), 57 (figure 2.7)
ketelaar, eric, 240
knight, michael muhammed, 204, 205, 207, 209, 211–2, 215–6
kominas, the, 215–6, 218
korea, 7

language development, 242
latent content, 165, 177
latvia, 7
law, religious, sharia law, 194, 196, 200, 202
legal deposit legislation, uk, 246
legal frameworks, 246
legislation, health 160
library of congress, 7, 8, 10, 11
link density, 29, 36
links, interpretation of, 193, 195–6
lippman, walter, 101
local news media, 93, 97
london, 46, 49
longitudinal analysis, 45–6, 58, 60
macro-historical research, 241
macro-scope, historical, 241
manifest content, 165, 177
marriage equality, 188
media studies, 2
Memento API, 47
Memento protocol, 240
meta words, 170, 177, 180–5
Meyer, Eric T., 247
micro-historical research, 241
migration online, 106
Millward, Gareth, 226–7, 232, 236
Mind (UK charity), 226
Ministry of Defence, 228–9, 233
Minitel, 118–121, 123–4, 126, 133
mobile web, 98
modelling, 60
Mosaic, 86
music catalogue, image of, 243
Musso, Martha, 227–8
MySpace, 205, 215
n-grams, 242
National Archives of the UK, the, 239
national web archives, 24, 25
national web, 62–80, 192
neologisms, 242
Ness of Brodgar, 229
Netarkivet, 10, 11, 62–4, 66–7, 72–5, 78–80
Netherlands, 8, 10
network analysis, 242
network centrality, 38
New Zealand, 7
newsgroup, 162, 163
newspapers, 83
Non-Print Legal Deposit, 192
Norway, 7, 10
one-sample t-test, 55, 56
online religion, object of study, 191
Open Director Project (DMOZ), 47, 58
.org.uk, 29–34
outdegree, network, 174, 177, 179
Oxford English Dictionary, 242
Oxford internet Institute, University of Oxford, 220
PageRank, 187
PANDORA, 7, 10, 11
Parliament (United Kingdom), House of Lords, 190, 193, 198
Parthenay, 125
personalization, 60
petitioning, online, 246
political party, 168, 188
politician, 168, 188
politics, 159, 160
Portuguese Web Archive, the, 10, 11
power law, 187
probability sample, 47, 58–9
provenance, 240
publication, date of, 240
Python, 51
Raffal, Harry, 228–9
Reddit, 216–7
regular expressions, 51
religion, 168, 180, 182–3
religious leaders, relationship with news media, 193, 194, 195, 198, 202
reviews, 48
Rhizome's ArtBase, 8
Richardson, Lorna, 229, 231, 233
Robot Wisdom, 87
Rogers, Richard, 193
Royal National Institute of the Blind, 226–7
Russia, 3
sample, 47, 58–9, 165
Savile, Jimmy, 221–2, 233
scholarly editing, 240
Scope (UK charity), 226
search engine 165–6, 177, 181, 184, 187, 189
search methodologies, 243–4
Second World War, 225
seed URLs, 192
selective crawl, 66
Sentamu, John, 198–9 (See also Archbishop of York)
SHINE (web archive search interface), 192, 200, 242, 243
sitemap, 51
SixDegrees.com, 88
social issue, 163, 185, 187–8
social media, 160, 162, 174, 188–9, 241–2
social network analysis, 159, 161–2
social networking sites, 88, 98
South America, 7
spam, 168, 186
Spanish, 3
Stanford University Libraries, 8
Stonehenge, 229
Sweden, 7
t-test, one-sample, 55, 56
Taylor, Helen, 229–30, 233
text analysis, 148, 163, 177, 242–3
text content, 159, 161, 164–6, 168, 170–1, 182–3, 187–8
Thirty Year Rule, 239
top-level domains, 24–5
topic drift, 171
topic modelling, 148, 149
tourism, 48
title, 48
TripAdvisor, 45–61
Twitter, 160, 162–3, 174, 188
UCLA Library, 8
UK Conservative Party, 1
UK Government Web Archive, 8
UK Web Archive, 1, 8, 10, 26, 191, 192, 200, 246
UK web domain, 24–5, 27, 28, 29–33
UK, 7–8
Ukraine, 1–2
United Kingdom Independence Party, 223
United Kingdom, 7–8
universities, 34–42, 58
online, 247
Russell Group, 35, 36, 38
URL (Uniform resource locator), 85
USA, 3
user-generated content, 48–49

Vallely, Paul, 195
visualization, 164, 173, 180
VOSON crawler, 170–1
Voyant, 246

Warburg Institute Iconographic Database, 243
WARC, 7
Warcbase, 137
Wayback Machine, 11, 58, 84, 140, 157
Web 1.0, 161–2, 165, 175, 177, 188–9
Web 2.0, 162
web archives, computational analysis of, 26
history of, 6–9
web archiving strategies, 10, 66

web archiving, 26, 27
web crawler, 51, 165–6, 170–1, 187
web crawls, 25
web, Danish, 62–80
web, development of the, 62–80
web, history of the, 23
web, use of, 2–6
Wikipedia, 4
Williams, Rowan, 194–7, 200–2 (See also Archbishop of Canterbury)
Word cloud, 180–3
word of the year, 242
World Wide Web Consortium, 226

Yahoo!, 137, 139, 140, 142, 145
YouTube, 174, 175, 243

zines, 204