

# Index

- Abstraction
  - data, 81
  - standards for, 4, 6–7, 10
- Acceleration, 117, 188
- ACID (atomicity, consistency, isolation, durability) tests, 103–105
- Addressable data, 98
- Adobe PostScript language
  - as de facto standard, 26
  - for printing, 73
- Advanced searches, 148–149
- Aircraft manuals, 56, 74
- American Iron and Steel Institute, 17
- American National Standards Institute (ANSI), 11
- Analysis of legacy information, 173–174
- ANSI (American National Standards Institute), 11
- APIs, 22
- Application specifications, 24–25
- Archiving data
  - benefits of, 80
  - in XML projects, 34, 36
- Artificial intelligence in search process, 162
- Associating Style Sheets with XML Documents*, 20
- Asynchronous networking, 119, 126–127
- Atomicity in ACID test, 103
- Attributes
  - in markup, 64
  - in relationships, 87
  - repetition of, 188
- Auditing data, 80
- Authors, in-line markup for, 154
- Automated conversions with legacy information, 181
- Automation, 77
- B2B (business-to-business)
  - automation for, 77
  - types of, 134–135
- Backward compatibility, 44
- Base64 encoding, 96
- Batch requests, 116
- Batch searching, 37
- Berners-Lee, Tim, 13, 162
- Binary data
  - external references, 96
  - internal, 96–97
  - representations of, 189
  - space efficiency in, 186
- Binary Large Objects (BLOBs), 34, 36
- Binding
  - data, 94, 109–110
  - programming languages, 22
  - in WSDL, 128–129
- BIS (Bureau of Indian Standards), 11
- Bloat in legacy information, 175–176
- BLOBs (Binary Large Objects), 34, 36
- Bloomberg service, 38
- Boolean searches, 148
- Bourret, Ronald, 110
- Bray, Tim, 147–148
- Broadcast networking, 119

- Budgets in XML projects, 46–47
- Bureau of Indian Standards (BIS), 11
- Business documents, 134
- Business-to-business (B2B)
  - automation for, 77
  - types of, 134–135
- bzip2 program, 199
  
- CALS Table Model, 21
- Canonical XML, 19
- Capitalization, 58–59
- Cascading style sheets (CSS), 12
  - as filters, 23
  - XML support in, 69
- CDATA sections, 52
- Change markup, 62–63
  - custom publishing issues in, 65–66
  - problems from, 63–65
- Character large objects (CLOBs), 36
- Character size, 197–198
- Client-side rendering, 38
- Client-side XML, 69
- CLOBs (character large objects), 36
- Cluster computing vs. grid computing, 131
- Code reduction in legacy information, 172
- Collaboration protocol profiles, 135
- Collections, 98
- Common data styles, 81–82
  - graph, 85–90
  - hierarchical, 90–92
  - tabular, 82–85
- Common Object Request Broker
  - Architecture (CORBA), 121
- Compatibility
  - networking for, 115–116
  - in standards, 27–29
  - in XML projects, 44–46
- Compressed XML, 107
- Compression, 199
  
- CompuServe, 33
- Computer games, external metadata for, 158
- Concurrent version system (CVS), 36, 74
- Consistency
  - in ACID test, 103
  - in standards, 27–29
- Content
  - management of, 74–75
  - non-XML, 95–97
  - vs. protocols, 136–137
- Context in searches, 143–145
- Converting
  - currency, 147
  - legacy information, 177–182
  - XML to HTML, 38
- CORBA (Common Object Request Broker
  - Architecture), 121
- Core specifications, 18
- Correlation in searches, 145–147
- Country identification codes, 11, 100
- Cover, Robin, 8, 17, 25
- Cover Pages, 8, 17, 25
- Creation stage in XML projects, 34–35
- Cross-conversions with legacy information, 177
- Currency
  - codes for, 11, 100
  - converting, 147
- Custom coding, transformation for, 37
- Custom libraries, reading XML with, 110
- Custom programming for queries, 192
- Custom publishing, 56, 65–66
- CVS (concurrent version system), 36, 74
  
- Data, 77
  - abstraction, 81
  - advantages of, 78
  - archiving and auditing, 80
  - binding, 94, 109–110

- common styles, 81–82
  - graph, 85–90
  - hierarchical, 90–92
  - tabular, 82–85
- data typing, 105–108
- identification of, 98–99
  - existing schemes for, 100
  - future possibilities, 101–103
  - in searches, 146
- interoperability of, 80–81
- markup issues in, 93
  - non-XML content, 95–97
  - whitespace handling, 93–95
- platform and storage independence, 78–79
- resources for, 81
- reusability of, 79
- self-documentation, 79
- summary, 111
- transactions, 103–105
- verification, 79–80
- Data interchange format (DIF), 78–79
- Data-oriented XML, 93
- Database metadata in searches, 37
- DataPower, 188
- DCMI (Dublin Core Metadata Initiative), 15, 149–150
- DCOM (Distributed Component Object Model), 121
- De facto* vs. *de jure* standards, 26–27
- Denormalized documents, 71
- Department of Defense (DoD) networking model, 136
- Diagrams, UML, 25
- Dictionaries, 74–75
- DIF (data interchange format), 78–79
- Direct mapping to programming interfaces, 110
- Disruptive technologies, 33, 40
- Distributed Component Object Model (DCOM), 121
- Distributed computing, 131
- DocBook specification, 13
  - interoperability with, 151
  - for Linux, 25
  - for master documents, 54
  - words in, 145
- Document Object Model (DOM)
  - for parsing, 190–191
  - programming bindings for, 22
  - tree-based, 180, 185, 192
  - W3C working group for, 13
- Document type definitions (DTDs), 23
  - costs of, 46
  - external, referencing, 189–190
  - for searching, 159–161
  - top-levels of, 75
  - validation, 27–28
- Documents, XML, 51–53
  - change markup in, 62–66
  - client-side XML, 69
  - content management in, 74–75
  - custom publishing, 56
  - hybrid data publishing, 55–56
  - idioms, 72–74
  - knowledge preservation in, 58–59
  - layout control for, 61–62
  - looseleaf publishing, 66–67
  - multiple text flows in, 67–69
  - people limitations in, 60–61
  - reusability of, 57, 70–72
  - single-source publishing, 53–54
  - summary, 76
- DoD (Department of Defense) networking model, 136
- DOM (Document Object Model)
  - for parsing, 190–191
  - programming bindings for, 22

- DOM (Document Object Model) (*cont.*)
  - tree-based, 180, 185, 192
  - W3C working group for, 13
- DOM-like interfaces, 190–191
  - pipelines in, 197
  - for XSLT, 194
- Domain names, 102–103
- Down conversions with legacy information, 177
- DTDs. *See* Document type definitions (DTDs)
- Dublin Core
  - for consistency and compatibility, 29
  - for RDF data, 156
  - with Semantic Web, 164
- Dublin Core Metadata Initiative (DCMI), 15, 149–150
- Duplicated parsing, 196–197
- Durability in ACID test, 104
- ebXML
  - as B2B initiative, 135
  - development of, 14
- EDI (electronic data interchange), 14, 134–135
- 8-bit ASCII encoding, 197
- Element content, 95
- Element names, repetition of, 188
- Emacs LISP language, 70
- Embeddable specifications, 19–21
- Emergency phone service, 5
- Empty tags, 63
- Encoding, 189, 197–198
- End users in search process, 161
- Enterprise resource planning (ERP), 176
- Entities in relationships, 87
- Enumerated values, 106
- ERP (enterprise resource planning), 176
- Event-based interfaces, 190–191
- Everything-is-a-string philosophy, 98
- Excel spreadsheet
  - as standard, 78
  - tabular style for, 82
  - XML format for, 24
- Expressions, XPath, 193
- Extensibility
  - in networking, 114–115
  - in XML projects, 44–46
- Extensible Business Reporting Language (XBRL) format, 80
- Extensible Hypertext Markup Language (XHTML), 25
  - consistency in, 29
  - interoperability with, 151
- External metadata in searches, 154–159
- External references, 189–190
- Facades for legacy interfaces, 168–170
- Fatal errors, 115
- Fear, uncertainty, and doubt (FUD)
  - standards, 7–8
- Feed readers and aggregators in RSS, 134
- Fielding, Roy, 119
- Financial-planning software, 133
- Firewalls, 117
- Flexibility in switching, 139
- Footnotes, text flow in, 68
- Format level, interoperability in, 5
- Formats vs. protocols, 113
- Formatting
  - change markup, 62–66
  - as information, 58–59
  - vs. meaning, 54
- Forward compatibility, 44–46
- FoxPro, 78
- Frustration problem, 41
- FTP, plaintext with, 185, 187

- FUD (fear, uncertainty, and doubt) standards, 7–8
- Full-text searches, 37, 143
- Fully normalized data, 86–90
  
- Generalized Markup Language (GML), 51
- GET headers and requests
  - in HTTP, 137–138
  - in SOAP, 123
- Global identifiers, 100–102
- Globus Toolkit, 132
- GML (Generalized Markup Language), 51
- Gnumeric spreadsheet, XML format for, 24
- Goldfarb, Charles, 51
- Google search engine, 143–144
- Granularity
  - in in-line markup, 153
  - in legacy information mismatches, 174–175
- Graph style, 85–90
- Graphic Communications Association, 14
- Grid computing, 131–132
- Growth in legacy information, 176
  
- Hardware acceleration, 117, 188
- Hardware support in performance, 187–188
- Headers in HTTP, 137–138
- Health Level Seven (HL7), 17
- Heuristics in legacy information
  - conversions, 177
- Hierarchical style, 90–92
- Holman, Ken, 23, 195
- Homographs in searches, 144
- HTML (Hypertext Markup Language)
  - converting XML to, 38
  - development of, 12
  - extensibility of, 115
  - forms and scripts with, 9
  - meta elements in, 149–150
  - for online documentation, 53
  - SGML inspiration for, 52
  - vs. XML, 78
- HTTP (Hypertext Transfer Protocol)
  - extensibility of, 115
  - GET headers in, 137–138
  - library problems with, 10
  - plaintext with, 185, 188
  - specifications for, 12
  - support for, 113
  - for transport, 34
- Human intelligence in search process, 161
- Human-readable format, 170–172
- Hybrid data publishing, 55–56
- Hypertext Markup Language. *See* HTML (Hypertext Markup Language)
- Hypertext Transfer Protocol. *See* HTTP (Hypertext Transfer Protocol)
- HyTime specification, 157
  
- IANA (Internet Assigned Numbers Authority), 101
- IBM databases, XML support
  - for, 36
- ICANN (International Corporation for Assigned Names and Numbers), 101
- ICAO (International Civil Aviation Organization) airport identifiers, 101
- ICE (Information and Content Exchange), 15, 38, 134
- IDEAlliance, 14–15
- Identification of data, 98–99
  - existing schemes for, 100
  - future possibilities, 101–103
  - in searches, 146
- Idioms, 72–74
- IETF (Internet Engineering Task Force), 12
- Ignorable whitespace, 95
- In-line markup, 151–154

- Independence, platform and storage, 78–79
- Indexes
  - in in-line markup, 152
  - in structural searches, 37
- Influence in XML projects, 42–44
- Information and Content Exchange (ICE), 15, 38, 134
- Information exchange, 91
- Information publishing, 91
- Inheritance relationships in searches, 163–164
- Intelligence in search process, 161–162
- Interface designers in search process, 161
- Interfaces, legacy, 168–170
- Internal text entities, 71
- Internalization and size, 198–199
- International Civil Aviation Organization (ICAO) airport identifiers, 101
- International Corporation for Assigned Names and Numbers (ICANN), 101
- International Organization for Standardization (ISO), 11–12
- International Press Telecommunication Council (IPTC), 17
- International Standard Book Number (ISBN), 100
- Internationalization in networking, 114
- Internet, interoperability in, 5
- Internet Assigned Numbers Authority (IANA), 101
- Internet Engineering Task Force (IETF), 12
- Internet Explorer, XSLT support in, 22
- Internet Protocol (IP), 115
- Internet worm, 9
- Interoperability
  - of data, 80–81
  - in searches, 150–151
  - standards for, 4–6
- Interprocess communications (IPC), 121
- Inventory information, 109–110
- Invited experts, 13
- IP (Internet Protocol), 115
- IPC (interprocess communications), 121
- IPTC (International Press Telecommunication Council), 17
- ISBN (International Standard Book Number), 100
- ISO (International Organization for Standardization), 11–12
- ISO-8859 encoding, 198
- ISO Latin 1 alphabet, 197
- Isolation in ACID test, 103–104
- Java language
  - memory leaks with, 10
  - for server-side development, 9
- JavaDoc system, 71
- Jazz standards, 3
- JSPs (JavaServer Pages), 9
- Just-in-time rendering, 38
- Keys
  - for identification, 99
  - for relational databases, 85
- keywords meta element, 150
- Knowledge preservation in documents, 58–59
- Knuth, Donald, 71
- KOffice spreadsheet, XML format for, 24
- Koine (Greek dialect), 167–168
- Language codes, 11, 100
- LaTeX markup, 51–52
- Layer-8 protocols, 138
- Layering in networking, 136–138
- Layout control for documents, 61–62
- Layout designers, costs of, 47
- Legacy information, 167
  - advantages of XML for, 168
  - analysis of, 173–174

- bloat in, 175–176
- code reduction in, 172
- conversion strategies for, 177–182
- granularity mismatches in, 174–175
- growth and versioning in, 176
- interfaces, 168–170
- metadata with, 182–184
- network effect in, 172–173
- summary, 184
- transparency for, 170–172
- virtual XML for, 180–182
- LEPs (lists of effective pages), 66–67
- Libraries
  - HTML, 10
  - reading XML with, 110
  - for reuse, 6
- Life cycles of XML projects, 33–35
- LISP language, 70
- Lists of effective pages (LEPs), 66–67
- Literate programming, 71
- Logical format for data, 79
- Looseleaf publishing, 66–67
- Loosely-coupled data, 79
- Lorie, Ray, 51
- MAC (medium access control) addresses, 101
- Magna Carta, 5
- Markup budgets in XML projects, 46–47
- Markup issues
  - change markup, 62–63
    - custom publishing issues in, 65–66
    - problems from, 63–65
  - non-XML content, 95–97
  - in searches, 151
    - external metadata, 154–159
    - in-line markup, 151–154
    - whitespace handling, 93–95
- Marshaling, 116
- Mathematical equations, support for, 96
- Mathematical Markup Language (MathML), 19–20
- Meaning vs. formatting, 54
- Measurement units, identification codes for, 100
- Medium access control (MAC) addresses, 101
- Megginson, David, 15–16
- Memory leaks, 10
- meta element, 149–150
- Metadata
  - as commodities, 158
  - external, 154–159
  - with legacy information, 182–184
  - in searches, 37, 154–159
- Microsoft Excel
  - as standard, 78
  - tabular style for, 82
  - XML format for, 24
- Microsoft Networks (MSN), 5
- Microsoft Visual SourceSafe, 36, 74
- MIF for printing, 73
- Mixed content, 52, 95
- Mixed conversions with legacy information, 177
- Monocultures from standards, 8–9
- Morris Internet worm, 9
- Mosher, Ed, 51
- Mozilla Project, 17
  - XML support in, 69
  - XSLT support in, 22
- MSN (Microsoft Networks), 5
- Multiple text flows in documents, 67–69
- Namespaces, 18
  - in compatibility, 46
  - consistency in, 28
  - for embedded specifications, 19–20
  - for global identifiers, 102

- Namespaces (*cont.*)
  - for naming collisions, 115
  - in searches, 159–164
- Namespaces Recommendation, 18
- Network effect
  - in legacy information, 172–173
  - in transportation standards, 4–5
- Networking, 113–114
  - advantages of, 114
  - for compatibility, 115–116
  - disadvantages of, 117
  - extensibility in, 114–115
  - internationalization in, 114
  - layering in, 136–138
  - performance of, 117
  - resources for, 116
  - security for, 117–118
  - states in, 118
  - styles, 118–120
    - asynchronous, 119, 126–127
    - grid computing, 131–132
    - miscellaneous, 134–135
    - RPC, 121–126
    - syndication, 132–134
    - Web Services, 127–131
  - summary, 139–140
  - switching and routing in, 138–139
  - transparency in, 114
- New Testament, 167
- News Industry Text Format (NITF), 144–146
- NewsML
  - identification codes in, 99
  - interoperability of data in, 80
  - for referencing external objects, 97
  - standardized metaformat in, 157
  - for syndication, 134
  - for transport, 38
- Newspapers, text flow in, 68–69
- 911 phone service, 5
- NITF (News Industry Text Format), 144–146
- Non-XML content, 95–97
- Normalized data, 86–90
- OASIS (Organization for Advancement of Structured Information Standards), 13–14
  - ad hoc groups, 15–16
  - IDEAlliance, 14–15
  - RosettaNet, 15
  - specialist groups, 16–17
  - WS-I, 14
- Object-oriented programming, 71
- OGSA (Open Grid Services Architecture), 131–132
- OGSI (Open Grid Services Infrastructure), 131–132
- One-way conversions with legacy information, 178
- Opaque names, searching documents with, 160–161
- Open Grid Services Architecture (OGSA), 131–132
- Open Grid Services Infrastructure (OGSI), 131–132
- Open Office spreadsheet, XML format for, 24
- Open Systems Interconnection (OSI) model, 136–137
- Oracle databases, XML support for, 36
- Organization for Advancement of Structured Information Standards. *See* OASIS (Organization for Advancement of Structured Information Standards)
- OSI (Open Systems Interconnection) model, 136–137
- Overwork problem, 41
- OWL (Web Ontology Language), 164
- Oxford English Dictionary*, 74

- Packets, HTTP, 138
- Parsers and parsing
  - duplicated, 196–197
  - external references with, 190–191
  - in in-line markup, 152
  - performance of, 188, 190–191
  - SAX for, 180–181
  - security of, 117–118
  - Unicode support by, 197
- Pattern matching in legacy information
  - conversions, 177
- PDF format for printing, 73
- Pecking order in XML projects, 42–43
- People limitations, 60–61
- Performance, 185
  - encoding in, 189
  - external references in, 189–190
  - of networking, 117
  - parsing interfaces in, 190–191
  - of pipelines, 196–197
  - of queries, 191–194
  - repetition in, 188–189
  - size in, 197–199
  - software and hardware support in, 187–188
  - space efficiency in, 186
  - summary, 199–200
  - with transformations, 194–196
- Photographs, external metadata for, 157–158
- Physical data format, 79
- Phytophthora infestans, 9
- Pipeline performance, 196–197
- Plaintext, 187–188
  - performance with, 185
  - for transparency, 170–172
- Planning XML projects, 33
  - components in, 33–35
  - creation stage, 35
  - forward compatibility and extensibility in, 44–46
  - markup budgets in, 46–47
  - pitfalls in, 38
  - rendering stage, 38
  - resistance to change in, 40–41
  - search and retrieval, 36–37
  - social side of, 42–44
  - storage and archiving stage, 36
  - summary, 47
  - technology in, 44–47
  - transformation stage, 37–38
  - transport stage, 38
  - unrealistic expectations in, 39–40
  - unspoken expectations in, 39
- Platform independence, 78–79
- Polling in RSS, 134
- POP3, plaintext with, 185
- Port types in WSDL, 128
- Postel, John, 115
- PostScript language
  - as de facto standard, 26
  - for printing, 73
- Potato blight, 9
- Primary keys for identification, 99
- Printing, 73
- PRISM (Publishing Requirements for Industry Standard Metadata), 15, 29
- Processing instructions, 63–64
- Processing pipelines, 196–197
- Profiles, WS-I, 14
- Programmers in search process, 161
- Programming interfaces, direct mapping to, 110
- Property inheritance in RDF, 163–164
- Protocol level, interoperability in, 5
- Protocols
  - vs. content, 136–137
  - vs. formats, 113
- Proximity searches, 148

- Publish/subscribe networking style, 119–120, 132–134
- Publishing Requirements for Industry Standard Metadata (PRISM), 15, 29
- Punctuation, 58–59
- Queries
  - performance of, 191–194
  - WSDL, 128
- Railroad system standards, 4–5, 7
- Random access
  - for queries, 193
  - for references, 89
  - in SAX-like interfaces, 191
  - in XSLT, 194–195
- Raw tables, 83–84
- RDDL (Resource Directory Description Language), 15–16
- RDF (Resource Description Framework)
  - data representation in, 29
  - for data sharing, 78
  - embedded data in, 21
  - for normalized data, 89
  - property inheritance in, 163–164
  - schema languages for, 23
  - in searches, 156, 159
  - Semantic Web with, 162
  - specifications for, 10
  - standardized metaformat in, 157
  - whitespace handling in, 94
- Reactivity, 188
- Readability of data, 84–85, 89
- References, external
  - binary, 96
  - DTDs, 189–190
- Regions, identification codes for, 100
- Relational databases, 85
- Relationships, 87, 163
- RelaxNG schema language
  - development of, 14
  - external references with, 190
  - schema, 23
- Remote Method Invocation (RMI), 121
- Remote procedure calls (RPC), 121–122
  - with REST, 122–124
  - with SOAP client, 123–126
  - vs. syndication, 132–133
- Rendering stage in XML projects, 34, 38
- Repeatable information, 86
- Repetition in performance and size, 188–189
- Representational state transfer (REST)
  - networking
    - characteristics of, 119–120
    - RPC with, 122–124
- Request/response networking style, 119–120
- Resistance to change, 40–41
- Resource Description Framework. *See* RDF (Resource Description Framework)
- Resource Directory Description Language (RDDL), 15–16
- Resources
  - for data, 81
  - for networking, 116
- Responses in WSDL, 128
- REST (Representational state transfer)
  - networking
    - characteristics of, 119–120
    - RPC with, 122–124
- Reusability
  - data, 79
  - documents, 57, 70–72
  - standards for, 4, 6
- Reuters, 38
- Rival management in XML projects, 43–44
- RMI (Remote Method Invocation), 121
- Robustness Principle, 115
- Roman Empire, 167
- RosettaNet, 15, 135

- Round-trip conversions with legacy information, 178–179
- Routing in networking, 138–139
- RPC (remote procedure calls), 121–122
  - with REST, 122–124
  - with SOAP client, 124–126
  - vs. syndication, 132–133
- RSS, 15–16
  - consistency in, 29
  - and syndication, 134
  - for transport, 38
  - for web logs, 25
- RTF format for printing, 73
- SAI (Standards Australia International), 11
- Sarvega, 188
- SAX (Simple API for XML) and SAX-like interfaces, 16
  - as de facto standard, 26
  - development of, 15–16
  - for parsing, 180–181, 190–191
  - pipelines in, 197
- Scalable Vector Graphics (SVG)
  - development of, 19
  - support for, 96
- Schemas, 107–108
  - for data identification, 100
  - in searches, 159–164
  - as stylesheets, 23
- Screen-based formats, 73
- Search stage in XML projects, 34, 36–37
- Searching, 143
  - context in, 143–145
  - correlation in, 145–147
  - interoperability in, 150–151
  - markup considerations in, 151
    - external metadata, 154–159
    - in-line markup, 151–154
  - namespaces, schemas, and Semantic Web in, 159–164
  - summary, 164–165
  - trust in, 149–150
  - usability in, 147–149
- Security for networking, 117–118
- Self-documentation, 79
- Self-labeling data, 172
- Semantic Web project
  - with RDF, 21, 25, 28
  - in searching, 162–164
- Session states in networking, 118
- SGML (Standard Generalized Markup Language), 11–12, 51–53
- SGML Open, 13
- Sharing information, 45, 78
- Shift-JIS encoding, 197
- Simple API for XML (SAX) and SAX-like interfaces, 16
  - development of, 15–16
  - for parsing, 180–181, 190–191
  - programming bindings for, 22
- Simple Mail Transfer Protocol (SMTP), 126
  - attacks on, 9
  - plaintext with, 185, 187
- Simple Object Access Protocol (SOAP)
  - with REST, 123–124
  - RPC with, 124–126
- Single-source publishing
  - characteristics of, 53–54
  - vs. idioms, 73
- Size, 185
  - compression in, 199
  - external references in, 189–190
  - and internalization, 198–199
  - repetition in, 188–189
  - space efficiency in, 186
  - summary, 199–200
  - Unicode and character size, 197–198
- SMTP (Simple Mail Transfer Protocol), 126
  - attacks on, 9
  - plaintext with, 185, 187

- SOAP (Simple Object Access Protocol)
  - with REST, 123–124
  - RPC with, 124–126
- Social side of XML projects, 42–44
- Software components for reuse, 6
- Software support in performance, 187–188
- Solicit/response method, 131
- Space efficiency, 186
- Specifications vs. standards, 3
- Spelling, 60–61
- Splitting markup, 64
- Spreadsheets
  - RPC model for, 133
  - standards for, 78
  - tabular style for, 82
  - XML format for, 24
- SQL for searches, 34
- Square peg problems for standards, 9
- Standard Generalized Markup Language (SGML), 11–12, 51–53
- Standard structures for metadata markup, 157
- Standards, 3
  - advantages of, 4
    - abstraction, 6–7
    - interoperability, 4–6
    - reuse, 6
  - consistency and compatibility in, 27–29
  - de facto vs. de jure, 26–27
  - disadvantages of, 7–8
    - abstraction, 10
    - FUD, 8
    - monocultures, 8–9
    - square pegs, 9
  - organizations for, 10
    - ISO, 11–12
    - OASIS, 13–17
    - W3C, 12–13
  - summary, 29–31
  - XML-related specifications, 17–18
    - application specifications, 24–25
    - core specifications, 18
    - embeddable, 19–21
    - utility specifications, 21–24
- Standards Australia International (SAI), 11
- Star Office, XML format for, 24
- Stateless protocols, 118
- States in networking, 118
- Stephenson, George, 4
- Stock quotes, syndication for, 132–133
- Storage and archiving stage in XML
  - projects, 34, 36
- Storage independence, 78–79
- Streaming APIs, 22
- Strings
  - efficiency of, 105
  - everything-is-a-string philosophy, 98
- Structural searching, 37
- Structural transformations, XSLT for, 194
- Structured programming, 71
- Style sheets, 12
  - costs of, 46
  - as filters, 23
  - linking, 20
  - XML support in, 69
- Styles, networking, 118–120
  - asynchronous, 119, 126–127
  - grid computing, 131–132
  - miscellaneous, 134–135
  - RPC, 121–126
  - syndication, 132–134
  - Web Services, 127–131
- Subsets in searches, 147
- Supersets in searches, 147
- SVG (Scalable Vector Graphics)
  - development of, 19
  - support for, 96
- Switching in networking, 138–139

- Synchronous networking, 119
- Syndication in networking, 132–134
- Synonyms in searches, 147
- System integraters in search process, 161
  
- Tabular style, 82–85
- TCP (Transmission Control Protocol), 115
- Technical committee (TC), 14
- Technology in XML projects, 44–47
- TEI. *See* Text Encoding Initiative (TEI)
- TELNET, plaintext with, 185, 187
- TeX language, 51
- TeXInfo, 54
- Text Encoding Initiative (TEI), 17, 25
  - consistency in, 29
  - interoperability with, 151
  - words in, 145
- Tracking changes in markup, 65
- Training, 47, 60–61
- Transactions, 103–105
- Transformation pipeline, 37
- Transformation stage in XML projects, 37–38
- Transformations
  - in hybrid data publishing, 55–56
  - performance with, 194–196
  - XSL. *See* XSL Transformations (XSLT)
- Translators for legacy interfaces, 168
- Transmission Control Protocol (TCP), 115
- Transparency of information, 107
  - for legacy information, 170–172
  - in networking, 114
- Transport stage in XML projects, 34, 38
- Transportation system standards, 4–5
- Tree-based APIs, 22
- Tree-based DOM, 180, 185, 192
- Tree-based interfaces, 190–191
- TROFF language, 51, 54
  
- Trust
  - in searches, 149–150
  - in Semantic Web, 164
- Two-phase commits, 105
- type field, 106
  
- U.S. Patent and Trademark Office (USPTO), 17
- UCS-4 encoding, 189
- UDDI (Universal Description, Discovery, and Integration of Web Services), 131
- UML (Unified Modeling Language), 25
- UN/EDIFACT, 135
- Unicast networking, 119
- Unicode encoding
  - for internationalization, 114
  - ISO specification, 11
  - size of, 197–198
  - UCS-4, 189
- Unified Modeling Language (UML), 25
- Uniform Code Council, 100
- Uniform resource identifiers (URIs), 101
- Uniform resource name (URNs), 102
- unit attribute, 147
- United Nations Centre for Trade Facilitation and Electronic Business (UN-CEFACT), 14
- Units of measure, identification codes for, 100
- Universal Description, Discovery, and Integration of Web Services (UDDI), 131
- Universal product codes (UPCs), 100
- Untyped XML data, 106–107
- Up conversions with legacy information, 177
- UPCs (universal product codes), 100
- URIs (uniform resource identifiers), 101

- URLs
  - RESTful, 124
  - vs. URNs, 102
- URNs (uniform resource name), 102
- Usability in searches, 147–149
- User control in XML projects, 42
- User training, 47, 60–61
- USPTO (U.S. Patent and Trademark Office), 17
- UTF-8 encoding, 189, 197–198
- UTF-16 encoding, 189, 197–198
- Utility specifications, 21–24
- Validation
  - data, 79–80
  - DTDs, 27–28
- Vector graphics (SVG)
  - development of, 19
  - support for, 96
- Verification of data, 79–80
- version attribute, 46
- Versioning in legacy information, 176
- Video clips, metadata for, 183–184
- Virtual Reality Markup Language (VRML), 9
- Virtual XML, 180–182
- VisiCalc spreadsheet, 82
- Visual SourceSafe, 36, 74
- Vocabulary in searches, 148
- VRML (Virtual Reality Markup Language), 9
- W3C (World Wide Web Consortium), 8, 12–13
  - for namespace identifiers, 102
  - Semantic Web project, 25
- Web browser, XSLT for, 22
- web logs, 16, 25
- Web Ontology Language (OWL), 164
- Web pages in relationships, 87
- Web Services, 14
  - specifications for, 10
  - working with, 127–131
- Web Services Definition Language (WSDL), 128–131
- Web Services Interoperability Organization (WS-I), 14
- Web Services Reliable Messaging (WSRM), 14
- Web Services Security (WSS), 14
- Whitespace
  - in formatting, 58–59
  - handling, 93–95
- Workload management in XML projects, 43
- World Wide Web Consortium (W3C), 8, 12–13
  - for namespace identifiers, 102
  - Semantic Web project, 25
- Wrapping interfaces, 173, 184
- WS-I (Web Services Interoperability Organization), 14
- WSDL (Web Services Definition Language), 128–131
- WSRM (Web Services Reliable Messaging), 14
- WSS (Web Services Security), 14
- WYSIWYG systems, 40
- XBRL (Extensible Business Reporting Language) format, 80
- XBRL International, 17
- XForms, 21
- XHTML (Extensible Hypertext Markup Language), 25
  - consistency in, 29
  - interoperability with, 151
- XLink, 20
  - consistency in, 28

- for referencing external objects, 97
- W3C working group for, 13
- XML support in, 69
- XMI (XML Metadata Interchange), 25
- XML applications, 16
- XML Applications and Initiatives page, 25
- XML Encryption, 21, 24
- XML Entity Catalog specification, 24
- XML Information Set, 19
- xml:lang attribute, 19
- XML Linking Language. *See* XLink
- XML Metadata Interchange (XMI), 25
- XML Namespaces, 18
  - consistency in, 28
  - for embedded specifications, 19–20
  - for global identifiers, 102
- XML Pointer Language, 20–21
  - W3C working group for, 13
  - XML support in, 69
- XML Recommendation, 18
- XML Schema language, 14, 107–108
  - external references with, 190
  - specification for, 23
- XML Signature, 21, 24
- xml:space attribute, 19, 94–95
- XML Stylesheet Language (XSLT)
  - consistency in, 29
  - for rendering, 34
  - W3C working group for, 13
- XML Topic Maps (XTM), 25
  - for data sharing, 78
  - standardized metaformat in, 157
  - whitespace handling in, 94
- XPath language
  - for content switching, 139
  - for queries, 192–193
  - random access in, 194
  - structural searching in, 37
- XPointer language, 20–21
  - W3C working group for, 13
  - XML support in, 69
- XQuery
  - consistency in, 28
  - development of, 22–23
  - structural searching in, 37
- XSL Transformations (XSLT)
  - characteristics, 22
  - limitations of, 9
  - performance with, 185, 188, 194–195
  - XML support in, 69
- XSLT. *See* XML Stylesheet Language (XSLT); XSL Transformations (XSLT)
- XTM (XML Topic Maps), 25
  - for data sharing, 78
  - standardized metaformat in, 157
  - whitespace handling in, 94



