Index

Abbreviations, 14, 37, 52, 68, 71, 129, 145, see also Acronym keys, Okapi '86 Go/see list
Access points, 9, 10-11, 149
Acronym keys, see Derived keys
Acronyms, 34, 52, 129, 145, see also Abbreviations, Okapi '86 Go/see list
ADAMSON, G W, 45
Affixes, see Stemming
ALBERGA, C N, 47

Algorithms,
Okapi '86 search processing and merging, 77-83, 141
spelling correction, 33-37, 73-74, 151, 152
stemming and conflation, 21-28, 29-30, 61-62, 64-66
term extraction and lookup, 77-78, 146
word similarity measures, 46-48, 49-52, 75, 152
American spelling, 34-35, 138
see also Spelling standardisation
Anagram technique for word matching, 75, 152
shortcomings of, 144
Analytical entries, 83, 149
see also Subject enrichment of bibliographic entries
AND (boolean), 11, 15, 16, 49, 51, 57, 60, 62, 63, 76, 94, 100, 115, 121, 139, 146
see also Combinatorial searching, OR (boolean)
Apple Ile, 3, 91-92
Articles (definite and indefinite), 38
Author/title search, 3, 118, see also Specific item searching
Authority control, 37, 52
CITE, 36-37
online catalogues, 52-53
phrase (or pre-coordinate) online catalogues, 37
see also Cross-references
Authority files, 14, 36, 37, 145
Automatic stemming, see Stemming

BATES, Marcia, 11
BELL, C L M, 13, 15, 25-26, 37
Beth Israel Hospital, see Paperchase
Bibliographic Access and Control System (BACS)
spelling correction, 15, 51
Bibliographic file(s), 3, 14
Okapi, 83-84, 85, 108
rare words in, 140
size, 83-84
as source of spelling dictionary, 73
subject description in, 147, 149
see also Indexes/Indexing

BLAIR, C R, 47
BOOKS project, 149 (reference [3])
Boolean searching, see AND (boolean), Combinatorial searching, OR (boolean)
BOURNE, C P, 41, 48
Brief record display, see Record displays
Browsing (scanning), 9, 10
subjects, 9, 11-12
VDU limitations (effect of), 11, 15
BUTCHER, Peter, 13

Cambridge University (online catalogue),
use of MARC cross-reference fields, 14
CARLSON, G, 45
Case, treatment of, 34, 92-93, 145
CATALOG, 13, 28
Catalogues,
  hard copy compared with online, 15, 147
  see also Online catalogues
CEREONE, N, 26
CITE, 15, 16, 36-37, 58, 60, 146
  automatic cross-referencing, 36-37
  combinatorial search, 16
  stemming, 13, 16, 29
City University, 13
Classification schemes,
  Dewey Decimal Classification Online Project, 11, 124
  language of, 11, 145
  for linking related records, 11
  for recall improvement, 10
  see also Subject enrichment of bibliographic records
CLR, see Council on Library Resources
Coloured keys, see Okapi coloured keys
Combinatorial searching, 3-4, 7-8, 15, 57, 60, 77, 100
  in CITE, 16
  cut-off rules (Okapi '86), 83
  design, 77
  interaction with stemming, 138-139
  merge procedure in Okapi '86, 82-83
  and spelling mistakes, 15
  term weights for, 79, 80-82
  used in CITE, LIBERTAS and Okapi, 57
  see also Ranked output, Term weighting, AND (boolean), OR (boolean)
Compound words, treatment of, 15, 27, 35, 37, 70
  MORPH5, 37
  Okapi '86, 38, 70, 93
Conflation, Chapter 3, 60-62, 138-139, 140-142
  see also Stemming, Truncation
CORNEW, R W, 43
Council on Library Resources,
  Online Public Access Catalog Study,
  cross-reference recommendations, 7
  difficulty with truncation, 12
subject search success rates, 7

Cross-references,
categories,
in LCSH, 9-10, 11
in Cambridge University online catalogue, 14
see also Okapi '86 go/see list, Tables and dictionaries,
Authority control CII, see Okapi '86, Control system

DAMERAU, F J, 48
DATTOLA, R T, 25
DAVIDSON, L, 47, 48
DAWSON, J, 28
Derived keys, 3, 15
Dewey,

Dewey Decimal Classification Online Project, 11, 124
in Okapi, 11, 85
see also Subject enrichment of bibliographic records

DILOG, 1

Dictionaries,
see Okapi '86 go/see list, Tables and dictionaries
Digrams, see N-grams
Disk storage and access, 85, 100, 149
Display, see screen displays
DOSZKOC, T E, 16

Ease of use, usability, 3

Equivalence Tables,
see Okapi '86 go/see list, Tables and dictionaries
Evaluation, Chapter 8
data analysis,

interview responses, 119-121
log analysis, 121-122
search repetition, 124-129
system performance, 129-135
data collation, 112
data collection, 108, 109-110
interviews, 111-112
logs, 112, Appendix 3
methodology, 109
objectives, 107
pre-test, 110-111
user searches, 113-115

number of terms in searches, 114-115
relationship between searches, 114
search classification, 114
sessions and searches, 113
SREC5S file, 113-116
user sessions, 113
boundaries, 113
see also Logs

EXP system, see Okapi '86, Experimental system

False drops, 58, 67
false coordination, 60, 69-70, 144, Appendix 4
through go/see list, 135, 145
through stemming, 12-13, 30, 145
FENICHEL, R R, 47
FIELD, B J, 24-25
FIRST, 25
Foreign language records, 73
  effect on spelling correction, 73, 129
FRAKES, W B, 13, 26
FREUND, G E, 42-43, 45-46, 50
Fuzzy matching, 2, Chapter 5
  see also N-grams, Soundex and soundex-type codes, Stemming,
  Truncation

GALLI, E J, 34
Geac (online catalogue),
  explicit truncation (Hull), 12
  subject access (PSB), 11-12
GREENFIELD, R H, 48

HALL, P A V, 41
Hardware, 85
  adequacy of conventional, 2, 142, 143-148
  Okapi, 3
  limitations of, 73
  special hardware for spelling correction, 43
HERSCHMAN, J, 10
Highlighting of terms in record displays, 103-104
HILDRETH, Charles R, iii
Homographs, 37, 135, 145, 146
Hull University,
  truncation in Geac catalogue, 12
Hyphens, treatment of, 34, 37, 64, 84, 93

ILS (online catalogue), 16
Indexes/Indexing,
  analytical, 148
  effect of stemming, 72
  keyword (or postcoordinate),
    adjacency information in, 60
    Dewey index in, 11
    links, 37
    and stemming, 12
    storage requirements, 30
    and synonyms, 37-38
    title words for subject access, 11
    and truncation, 12-13
    updating, 148-149
  see also Okapi '86 indexes
language of, 118, 145-146
Okapi, see Okapi '86 indexes
phrase (or precoordinate), 9, 10, 14
  see also bibliographic files
Indexic, 118
Information Retrieval systems, see CATALOG, CITE, LEXICON,
  Massachusetts General Hospital, MORPHS
Initialisms, treatment of, 85, 92
IN5PEC, 24-25, 29
Interaction, see User(s) interaction
Intermediaries, see Search intermediaries
Interviewing, see Evaluation
INTREX, 23

Jargon,
   classification and indexing, 118
   library, iii
JOHNS, Nicky, iv, 2, 5
JONES, K P, 13, 15, 25-26, 37
JONES, Richard M, iv, 2, 4, 125
Keyword (or post-coordinate) online catalogues, see Online
catalogues
Known item searching, see Specific item searching

LAN(s), 2-3
LANDAUER, C, 29
Languages (programming) used for project, see Programs
LCSH, see Library of Congress Subject Headings
LENNON, M, 24, 29, 45
LEXICON, 52
   spelling correction, 35, 36, 52
LIBERTAS (online catalogue), see SWALCAP
Library of Congress Subject Headings, 138
   cross-references, 10
   language, 9-10
   pre-processing, 11
   see also Subject indexing
Local area networks, 2-3
Logs, 15, 49, 58, 61, 112, 135, 153-159, see also Evaluation
Lookup tables, see Okapi '86 go/see list, Tables and
dictionaries
LOVINS, J B, 23, 28, 29
LOWE, T C, 23

Malaysian Rubber Producers’ Association, see MORPHS
MANDEL, Carol A, 10
MARC, 4, 11, 14, 73, 83, 85, see also Subject descriptors
MARKEY, Karen, iii, 7, 11, 12, 109, 124
MARS, 30, 36-37
Massachusetts General Hospital (online catalogue),
   spelling correction, 15-16, 51-52
   stoplists, 51
MAW (Minimum Acceptable Weight), see Term weighting
Maximum possible weight, see Term weighting
Medical Subject Headings, see MeSH
MEVYVL, 14, 37
MeSH, 15, 16, 36-37
MGW (Minimum Good Weight), see Term weighting
Minimum acceptable weight, see Term weighting
Minimum good weight, see Term weighting
Miskeyings and misspellings, 2, 7, 15-16, 33-35, 41-55, 57, 59,
74, 125-131, 132-134, 138, 142-144, 151-159
   see also Spelling correction
MITEV, Nathalie Nadia, iv, 2, 60
Monograms, 41
MOORE, F J, 48
MORPH5, 13, 15, 37
  homographs, 146
  stemming, 13, 25-26
  tables, 15, 37
MORRIS, R, 43
MPW (Maximum Possible Weight), see Term weighting

N-grams,
  applications,
    detection of spelling errors, 43-45, 50
    improvement of recall, 42-43
  definitions, 41-42
  digrams, 41, 43
  disadvantages, 42
  effectiveness, 42-43, 45-46
  monograms, 41
  in online catalogues and IR systems, 50
  syllabic, 44, 46, 50
  techniques,
    frequency tables, 43
    invalid trigram detection, 44
    inverted file, 42-43
    length, 45
    positional analysis, 44-45
    syllable comparison, 43
  tetragrams, 41
  trigrams, 41-42, 44-46
  see also fuzzy matching
Natural language processing, 16, 26, 36
Nestar, 3
NOREAULT, T T, 42
NUSSBAUM, R, 43

Objectives (of project), 1, 7, 13, 57
Observation, see Evaluation
OCLC LS/2000,
  tables, 14
  truncation, 12
Okapi '84, iii, 2-3, 15, 83, 85, 109-110, 137
  logs, 8-9, 58-59, 76, 109
Okapi '86,
  0STEM system, 4, 78, 85, 89, 126-128, 138-139
  bibliographic file, 83-84
  coloured keys, 91, 102, 153-159
  combinatorial search, 4, 8, 57, 60, 77-83, 110, 112, 138,
    140
  Control system (CTL), 4, 57, 62, 89, 93-97, Chapter 8,
    138-142
  design, 57-106
  evaluation, 107-136
  Experimental system (EXP), 4, 11, 38, 57, 62, 78, 89, 93-94,
    97-100, Chapter 8, 138, 140-142, 153
Pre-coordinate online catalogues, see Online catalogues
Prefixes, see Stemming
Programs,
   library automation, 148-149
   Okapi, see Okapi '86 programs
   online catalogues, 137
   stemming, 30

RAOCOL, 23-24, 29
Ranked output, 16, 42, 77, 86, 104, 141, 149
   see also Combinatorial searching, Term weighting
Recall
   comparison between systems, 124-129, 141-142
   evaluation, Chapter 8
   improvement devices, 1, 4, 41, 57
   fuzzy matching, 2, Chapter 5
   stemming, Chapter 3, 141-142
   tables and dictionaries, Chapter 4
Record displays, 85, 102-104
   brief and full, 102-103
   highlighting, 103-104
   sequencing of, 104
REINTJE5, J F, 23
Relevance feedback, 5, 102
Response times, 100, Appendix 3
RISEMAN, E M, 44, 46

Scanning, see Browsing (scanning)
Screen displays, Chapter 7
Search, searches, searching,
   analysis and classification, 113-115, 147-148
   failures, 4, 7, 9-12, 117-119, 121-123, 125-126, 133-134,
      Appendix 4
   initial, 126-129, 138, 140
   mistakes, see Miskeyings and misspellings, Spelling correction
   problems, Chapter 2, 58-59
   processing, 77, Chapter 7, 142
   repetition by experimenter, iv, 2, 4, 124-129, 137
   rephrasing, 141, Appendix 3, see also sessions
   sessions, 5, 58, 108, 111-113, 116-119, 124-129, 141,
      Appendix 3, Appendix 4
   statements, 9, 58-60
      language of, 58-59, 113, 114, 117-119, 138, 147, 149,
         Appendix 4
      number of terms in, 80-82, 114-115, 122-123
      parsing, 77-79, 86, 92-93
      sample of, 9
      stemming and reassembly of, 66-67
      success rate, 12, 51-52, 118-117, 125-129
      typing problems, 119
   see also User(s)
Search intermediaries, 2, 13, 21, 57-58
See Also references, see Cross-references
See references, see Cross-references
Sessions, 113, see also Searches
SIEGEL, E R, 16, 103
SMART, 25
Software, see Programs, Okapi '86 programs
Soundex and soundex-type codes, 46-48
see also Fuzzy matching
  applications,
    detecting spelling errors, 46-48
    improving recall, 46-48
  definitions, 45-47
  effectiveness, 47-48
in online catalogues, 51-53
  Bibliographic Access and Control System, 51
  LEXICON, 52
  Massachusetts General Hospital, 51-52
  Okapi, see Okapi '86, spelling correction
  Paperchase, 52
techniques,
  Blair, 47
  Bourne & Ford, 48
  Damerau, 48
  Davidson, 47, 48
Source file, see Bibliographic file
Specific item searching, 3-4, 12, 15, 49, 110, 111, 114, 117
see also author/title searching
SPEECOP Project, 33, 44, 46, 50
see also N-grams, Spelling correction
Spelling correction, 15-16, 49-50
  effectiveness, 33, 35,
  in Okapi '86, 74
  dictionary, 73
  processing, 75
  similarity algorithm, 75, 84, 152
  soundex algorithm, 84, 151
  user presentation, 86-87, 95-96, 98-99
using n-grams, 50
using soundex codes, 35, 51-52
using tables, 33-34
see also Miskeyings and mispellings
Spelling standardisation, 14, 34-35
  Okapi '86, 30, 61, 139-140, Appendix 3
  evaluation, 107, 130
  recommendations, 139-140, 142
  rules, 64
see also American spelling
Staffing of the project, 2, 5
Stem modification, see Stemming techniques
Stemming,
  algorithms,
    Cercone, 26
    Dawson, 28
    FIRST, 25
    Galli and Yamada, 34
    INSPEC, 24-25, 29,
    INTEX, 23


17. Developments in scholarly communication in the humanities in the USA by Paul Sturges. ISBN 0 7123 3119 0. 1987.


