Appendix 1

The soundex algorithm

If word length < 5 or > 18 or word contains a digit, exit.

Copy first letter to output.

While the next letter is the same as the first, skip a letter.

Assert 'duplicates are to be ignored'.

For each remaining letter except the last

   if letter is in (aeiouywh) skip it but assert 'duplicates not
   to be ignored'

   else if letter is in (cgjkqsxz) hold it as 'c'

   else if letter is in (bfpv) hold it as 'b'

   else if letter is in (dt) hold it as 'd'

   else if letter is in (mn) hold it as 'm'

   else (letter is 'l' or 'r') hold it unchanged.

   Unless previous output character is the same as this one
   and duplicates are to be ignored, copy held letter to
   output.

   Assert 'duplicates are to be ignored'.

If final letter is in (aiouy) copy 'y' to output, otherwise
treat it as previous letters.

End.

Examples:

```
economics )
economic  ) --> ecmmc   (repeated "m" because of intervening
*ecomonic )                vowel)

*econmic     --> ecmc

rabbit )
rabid  )
rapid  )  --> rbd
repeat )

sociology  )
*socialogy ) --> sclcy
*sociolgy  )
```

Appendix 2

The word-matching algorithm

Set best score to zero.

For each word in the candidate list

   Set score to zero.

   For each letter (except the first)

      If the letter also occurs anywhere in the user's word
      AND this letter has not already been counted, increase
      score by one and mark the letter (in the user's word)
      as 'already counted'.

      If score is better than best so far, or score is equal
      to best so far but this word is shorter, save this word
      as 'best so far'.

If any of the following four conditions hold, return
failure and exit:

   best score is less than (length of user's word - 2)

   user's word has five letters and best score is less than
   four

   user's word has fewer than nine letters and best match is
   more than one letter shorter or longer

   user's word has more than eight letters and best match is
   more than two letters shorter or longer

otherwise return success (best match will be offered to user).

End.

Appendix 3

Extracts from a log file

Key

Lines from the log are in roman, comments and annotations in italic.

The first character in each line identifies the type of data:

    I Terminal and program identification
    T Date and time to nearest second
    C Comment
    S Screen identification
    U User searches and commands
    E Elapsed time to nearest second since last "E" or "T" line.
    L Lookup. Contains enough information to reconstruct what was displayed on the screen.
    O Only used here to show number of items with maximum possible weight, minimum good weight and minimum acceptable weight. These figures can be used to deduce how the result of a search was presented to the user.
    X Suggested spelling correction.

Use of the coloured function keys is logged as :R, :Y, :B, :G, :X for red, yellow, blue, green, black respectively. The yellow key had no effect, but is still logged. A timeout is logged as :T.

The log

  I $OC 30-Oct-86 195
Terminal and program identification. The date is a program revision date. The version code 195 shows that this is the EXP system.
  T 861104092641
Date and time
  S SCR1
Introductory screen (Fig 7.1)
  U :G
User presses green key

C ** Start
T 861104093400
S SCR7.1
Search input screen (Fig 7.2)
U relationship and social classes¦G
User enters a search and keys green -
E +34
- this took 34 seconds.
L relationship (228), social (6257), classes (565),
Lookup. The numbers are the number of postings as shown to the user.
E +4
Lookup took 4 seconds. System displays 'Looking for books ..'
E +30
and performs the merge, which took 30 seconds
O NMPW=1 NGW=4 NPW=512
*Merge found 1 book with max. poss. weight, 3 more books with min. good weight and
more than 500 books altogether.*
U ¦G¦G¦G¦G¦G¦G¦G¦G¦G¦R
User presed green key 10 times and saw the first 10 records, then the red key -
E +191
- spending 191 seconds looking at records.
C BR: (0), FR: (10),
No brief records, 10 full records (these systems don't use a brief display).
S SCR7.2
*Last key user pressed was red (¦R). This led to second or subsequent input screen
(Fig 7.13).*
U social relationship and classes in great britain¦G
E +51
Another related (narrower) search by the same user.
L social (6257), relationship (228), classes (565), great britain (11019),
GREAT BRITAIN is a phrase in the 'go/see list'.
E +3
E +62
*This merge of about 30,000 postings (strong stems are not shown in the log) took
over a minute. This is nearing the limits of acceptability. Luckily, not many
searches are as heavy as this one.*
O NMPW=0 NGW=11 NPW=128
U ¦G¦G¦G¦G¦G¦G¦G¦G¦G¦G¦G¦B¦G¦G¦G¦G¦G¦G¦G¦G¦G¦G¦G¦G¦G¦G¦G¦G¦G¦G¦G¦
   G¦G¦G¦G¦G¦G¦G¦G¦G¦G¦R
E +322
C BR: (0), FR: (49),
S SCR7.2
U relation s¦R
User starts to type another search, but abandons it by pressing the red key.
E +19
T 861104094556
C ** Finish - duration 716 sec(s)
*The duration shown is the time from the timestamp after the 'Start' comment to the
timestamp before the 'Finish' comment.*
S SCR1

From here on, many lines of the log have been removed.

U :G

C ** Start
T 861104120248
S SCR7.1

Start of a long sequence of related searches.

U german media:G
E +7
L german (1665), media (439),
E +2
E +4
O NMPW=6 NGW=6 NPW=6
U :G:G:G:G:G:G:R
E +139
C BR: (0), FR: (6),
S SCR7.2
U german press:G
E +5
This user can type. Screen display time, thinking time and 13 keystrokes took only 5 seconds.
L german (1665), press (475),
E +2
E +6
O NMPW=21 NGW=21 NPW=21
U :G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:R
E +260
C BR: (0), FR: (21),
S SCR7.2
U german jouralism:G
E +8
L german (1665), jouralism [F,
U :R
User aborts search instead of correcting. May not have noticed misspelling.
S SCR7.2
U german broadcasting:G
 . . . . .
U german radio:G
 . . . . .
U german television:G
 . . . . .
U soviet media:G
 . . . . .
U soviet press:G
 . . . . .
U soviet journalism:G
 . . . . .
U german journalism:G
 . . . . .

```
T 861104122137
C ** Finish - duration 1129 sec(s)

S SCR1
U :G

C ** Start
T 861104122251
```
Although only 74 seconds have elapsed this is almost certainly a different user.
```
S SCR7.1
U social organizations of prison:G
E +45
L social (6257), organizations (1732), prison (102),
```
The spelling standardisation ensures that 'organisation(s)' is also included.
```
E +1
E +37
O NMPW=1 NGW=13 NRW=143
U :G:G:G:X
E +40
C BR:  (0), FR:  (3),
T 861104122454
C ** Finish - duration 123 sec(s)

S SCR1
U :G

C ** Start
T 861104122848
```
Different user.
```
S SCR7.1
U rape:G
E +15
```
*Probably unfamiliar with the catalogue. Few people take 15 seconds to make 6
keystrokes, so user was probably reading the screen.*
```
L rape (16),
E +0
E +1
O NMPW=16 NGW=16 NRW=16
U :G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:X
E +482
```
Eight minutes looking at 16 records.
```
C BR:  (0), FR:  (16),
T 861104123706
C ** Finish - duration 498 sec(s)

S SCR1
U :G
```

C ** Start
T 861104123736
Cannot assume different user - only 29 seconds time lapse. The fact that previous
search ended with black key is not a reliable indicator.
S SCR7.1
U maternal deprivation:G

....

O NMPW=5 NGW=5 NPW=67

....

E +430
C BR: (0), FR: (54),
T 861104124528
C ** Finish - duration 473 sec(s)

S SCR1
U :G

C ** Start
T 861104124551
S SCR7.1
U role of women as mothers and feminist views:G

.....

O NMPW=0 NGW=0 NPW=4
This search is too precise to work well. Two of the four books found might be
judged relevant, but there is nothing in the catalogue indexed under WOMAN/WOMEN and
MOTHER(S) and FEMINIST/FEMINISM.
U :G:G:G:G:G:X
E +308
User spent more than five minutes looking at and/or making a note of these four
records.
C BR: (0), FR: (4),
T 861104125235
C ** Finish - duration 404 sec(s)

.....

C ** Start
T 861104145533
S SCR7.1
U economic development in the third world:G
E +55
L economic (5545), development (3364), third world (482),
THIRD WORLD is a phrase, equivalenced to DEVELOPING COUNTRIES, UNDERDEVELOPED
COUNTRIES etc
E +2
E +34
O NMPW=188 NGW=431 NPW=482
U :G:G:G:G:G:G:G:G:G:G:G:G:G:R
E +60
C BR: (0), FR: (14),

S SCR7.2
U transforming traditional agriculture¦G
This is a title.

    .....
O NMPW=1 NGW=2 NPW=13

    .....
C BR: (0), FR: (1),

S SCR7.2
U agarian reform¦G [2 dels]
Use of the white (delete) key is logged.
E +21
L agarian (2), reform (412),
AGARIAN finds misspellings ('Essays in agarian history') in the catalogue. User
doesn't notice -

    .....
O NMPW=0 NGW=0 NPW=2
U ¦G¦G¦G¦G¦B¦R
- and even looks at one of the records twice.
E +74
C BR: (0), FR: (3),
S SCR7.2

    .....
Many more related searches, some probably for specific titles.

    .....
T 861104151212
C ** Finish - duration 999 sec(s)

S SCR1
U ¦G

C ** Start
T 861104151934
S SCR7.1
U conciliation in industrial relations¦G
'Industrial relations' is a phrase in the 'go/see' list.

    .....
O NMPW=12 NGW=12 NPW=29
U ¦G¦G¦G¦G¦G¦G¦G¦G¦G¦G¦R
E +30
C BR: (0), FR: (11),
S SCR7.2
U industrial relations¦G
Broader search.

    .....
O NMPW=475 NGW=475 NPW=475
U ¦G¦G¦G¦G¦G¦G¦G¦G¦G¦G¦R
E +33
C BR: (0), FR: (10),
S SCR7.2
U ¦T
Leaves terminal to time out from the input screen.

```
E +80
T 861104152236
C ** Finish - duration 182 sec(s)
S SCR1
U :G
```

```
C ** Start
T 861104153342
S SCR7.1
U economics ddictionary:G
```

*The duplicated 'd' may be due to the 'auto-repeat' function. Online catalogue
terminals should have this function disabled.*

```
E +15
L economics (5545), ddictionary OF,
X dictionary
```

System suggests a correction -

```
U :G
```

- which the user accepts.

```
L dictionary (1326),
  .....
O NMPW=45 NGW=45 NPW=45
```

*(Weak) stemming works well here. Without stemming this search retrieves only the 28
books which have 'economics' and 'dictionary' in the title. The Library of Congress
Subject Heading is 'Economics - dictionaries'.*

```
U :G:G:G:G:B:B:B:G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:
   G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:G:X
```

Blue key used to look at some of the records again.

```
E +238
C BR:  (0), FR:  (51),
T 861104153825
C ** Finish - duration 283 sec(s)
```

```
S SCR1
```

Appendix 4

Notes on some failed sessions

These are the experimenter's notes taken while repeating searches which the user reported as failures. These sessions are discussed in 8.4.2.

1   13.11.86 CONTROL [system] 12.41

HMSO employment statistics 1986
Retrieves 0 MPW (max. poss weight), 0 GW (good weight), 1 (acceptable weight)
Looks at this record and presses G[reen] K[ey] another 3 times.
Obviously not sure what to do.

Record not useful: 'employment' and 'hmso'
employment statistics finds 30 of mpw.
employment statistics 1986 is not in the file. (There are 6 records under '1986'.)

Describes subject as: 'Government statistical data'

2   3.11.86 CONTROL 13.03

Unemployment
111 mpw
Looks at 38 before deciding that it is too general.

Employment structure
2 mpw
Looks at each record twice: false co-ordination.
'North East Lancashire Structure Plan survey report. Employment and industry'
The use of time-span of discretion to measure the level of work in employment
roles and to arrange an equitable payment structure.'

Employment structures in britain since 1945
86 naw
Looks at first 2
First 10 are on 'britain' and '1945' although some are on economics.

['Occupational structure' gives 3 nmpw which are probably better.]
Describes subject as 'Employment structures'

Unsuccesful - nothing useful found.

3   3.11.86   CONTROL   13.08

Sterling
16 mpw, 6 looked at.
Unsuccessful - nothing useful found - 'Not really one on shares'
Describes subject as: 'Economics sterling shares and gold'.
Seems to want effect of sterling on shares.

4   3.11.86   CONTROL   17.05

Sociology of shopping
NBF ['no books match your search']

Shopping
149 MPW
50 looked at

Anthropology of shopping
284 acceptable weight (all either 'anthropology' or 'shop(ping)'
R[ed K[ey]

Unsuccessful - nothing useful found. 'Didn't expect any [useful books] - too
general'.
'consumer behaviour' finds 59 mpw
'Shopping behaviour' finds 2 MPW, one of which is good.

5   4.11.86   EXPERIMENTAL   13.36

aids
302 mpw
Looks at 12 - all false drops

acquired immunity deficiency syndrome
2 aw
Looked at. Both fds or marginally relevant.

aquired immune deficiency syndrome
NBF

Unsuccessful - nothing useful found.

6   4.11.86   EXPERIMENTAL 14.01

Generric
X Generic [i.e. system suggests 'generic']
User presses RK (aborts search)

Generic social work in Britain
1 GW
Looks at it

Unsuccessful - nothing useful.

7   5.11.86   CONTROL   15.16

Recent changes in Londons economy
13 AW
All looked at
All contain 'recent' and one other word

'changes in Londons economy' finds 14 gw, 29 aw, the 14 contain 'economy' and
'changes', 15 contain 'london' and one of the others.

'londons economy' finds 8 mpw, 1 of which is v good

8   6.11.86   EXPERIMENTAL   12.01

Economic trends
27 mpw - looks at 3
Obviously too general - many different countries

Employment trends post-war
52 aw - looks at first one

Social economics employment trends post-war
9 aw - looks at 4. Several look generally relevant but not specifically about
employment.

employment trends in england
3 gw, 234 aw

'work post-war' finds 7 mpw, 72 aw. One good.

Unsuccessful - but useful books found

These searches demonstrate the importance of hitting on the correct word which might
be odd. For example: 'employment patterns', 'work patterns' and 'labour trends' all
give good records.

9   6.11.86   EXPERIMENTAL   13.50

The work of social work
The work of social work today
The functions of contempory [sic] social work
Tasks of social workers today
A definition of social work
Meaning of social work
What is social work
Development of social work tasks
What social workers do
Social policy and the tasks of social work

Altruism
Selfless aciot [sic]
RED KEY

Selfless activity and social work tasks

Spelling correction suggests 'sleepless'
Presses B[lue] K 10 or so times
Replaces 'selfless' with 'unselfish action'
Can't Find 'unselfish'
Green Key (causes to ignore) and then Red Key (which aborts)

Selflessness and social work
RK   :

Perspectives in social work

Unsuccessful - nothing useful found.

Some of these searches seem to work well (e.g. 'tasks of social workers today') but
see below.

User commented that it only found basic books which she already knew about. She
complained that it couldn't find synonyms for unmatched words

10   6.11.86   EXPERIMENTAL   15.37

Job evaluation
Successful

Textile industry
20 mpw, 20 gw, 35 aw
Looks at all 35

Input output tables
2 mpw ,28 gw ,28 aw
Looks at 3

Unsuccessful - nothing useful.

11   6.11.86   EXPERIMENTAL   17.35

Truancy
CF, GK, NBF

Absenteeism
7
Looks at 7

School absenteeism
0 mpw, 0 gw, 7 aw
All absenteeism only - looks at 1st

Unsuccessful - 'nothing useful found '
['Truant' gives 1 and is a good record.]

Commented 'Didn't expect to find anything - just checking'

12   10.11.86   EXPERIMENTAL   14.26

Feuerbach
CF GK NBF

Philosophy
> 512
RK

Hegel
34
Looks at 34

K MARK/FEUERBACH
Black key

Unsuccessful - nothing useful found.

13   10.11.86   EXPERIMENTAL   14.48

How to beat unemployment
0 mpw, 0 gw, 118 aw
Looks at 7

Layard
CF BK RK

Macroeconomics and the British economy
5 13 110
Looks at 1

Unsuccessful - useful material found.

14   10.11.86   EXPERIMENTAL   17.10

House of commons
235 mpw
Looks at 18

Parliament and the Law
43 mpw
Looks at all, 1 twice

The passing of laws
4 mpw, 4 gw, 28 aw
Looks at 27. [3 of the mpw are false coordination as 'How to pass exams in law'.]

Passing of laws by the House of Commons
28 aw
Looks at all

Parliament and legislation
23 mpw
Looks at all

Unsuccessful - useful books found
He defines his search as: 'the mechanics of passing legislation'

'Parliament procedure' gives 9 mpw and they all look good.

15   11.11.86 CONTROL   11.36

Developement of amrica
Amrica CF
BK Britain [replaces 'Amrica' by 'Britain']
1 book found [under the file misspelling 'developement']
Looks at it

America as a developing country
0 7 16
Looks at all

Britain as a developing country
0 7 417
Looks at 50

Unsuccessful - useful books found.
Defines subject as: 'economic development of Britain in the 18th century'.
This oral definition gives good results.

16   11.11.86 CONTROL   13.56

Britain in indi
3 3 16 ['indi' is weak stem from 'indies']
8 records looked at

British rule in India
3 8 34
15 looked at

The advantage of india to britain in colonial rule
0 0 2
Looks at both

Unsuccessful - nothing useful found - 'It's a bit hard to communicate with it'

17   11.11.86 CONTROL   15.10

The police
201 mpw
Looks at 23

Unsuccessful - useful books found, said that he located too many books and couldn't be bothered to browse thru them all. First several were useful but were in Law Library and couldn't be bothered to go there.

Appendix 5

List of equivalence classes of terms used in the
experimental catalogue (go/see list)

```
1st, first, i, one, 1
2nd, second, ii, two, 2
3rd, third, iii, three, 3
4th, fourth, iv, four, 4
5th, fifth, v, five, 5
6th, sixth, vi, six, 6
7th, seventh, vii, seven, 7
8th, eighth, viii, eight, 8
9th, ninth, ix, nine, 9
10th, tenth, x, ten, 10
16th, sixteenth, sixteen, 16
17th, seventeenth, seventeen, 17
18th, eighteenth, eighteen, 18
19th, nineteenth, nineteen, 19
20th, twentieth, twenty, 20
advertising, advertisement, adverts
aeroplane, airplane
africa, african, afro
algeria, algerian
alps, alpes, alpine
aluminium, aluminum
america, american
arab, arabic, arabian
argentina, argentinian
aristotle, aristotelean, aristotelian
asia, asian
athens, athenian
australia, australian, australasia, australasian
austria, austrian, austro
bbc, british broadcasting corporation
bcs, british computer society
belgium, belgian
bfi, british film institute
bhs, british home stores
bl, british leyland
botb, british overseas trade board
bnfl, british nuclear fuels limited, british nuclear fuels ltd,
    british nuclear fuels
bolshevik, bolshevism
br, british rail
brazil, brazilian
brecht, brechtian
```

bt, british telecom
california, californian
cambodia, cambodian
canada, canadian
cba, cost benefit analysis
cbi, confederation of british industries
channel four, channel 4
child, children, childhood, childish
chile, chilean
china, chinese
cia, central intelligence agency
clp, city of london polytechnic
cnd, campaign for nuclear disarmament
communist party
co, company
conservative party, tory party
corsica, corsican
cpsa, civil and public servants association
cttee, committee
cuba, cuban
curriculum, curricula, curricular
czar, csar, tsar, tzar
czechoslovakia, czech, czechoslovak, czechoslovakian
denmark, danish, dane
dhss, department of health and social security
doi, department of industry
dti, department of trade and industry
eec, european economic community
eoc, equal opportunities commission
egypt, egyptian
england, english, anglo
ethiopia, ethiopian
europe, european
facts of life
fao, food and agriculture organisation
fauve, fauvism
finland, finnish
flanders, flemish
florence, florentine
france, french, francais
freud, freudian
gaol, jail
gdp, gross domestic product
general strike
germany, german, germanic, germano, deutsch, deutschland
glc, greater london council
gnp, gross national product
gov, govt, government
greece, greek, grecian
haiti, haitian
hebrides, hebridean
hegel, hegelian

high school, secondary school
himalayas, himalayan
holland, dutch, netherlands
house of lords
hungary, hungarian
iba, independent broadcasting authority
iceland, icelandic
ici, imperial chemical industries
imf, international monetary fund
india, indian, indo
industrial relations
industrial revolution
iq, intelligence quotient
ira, irish republican army
iran, iranian
iraq, iraqi
ireland, irish
israel, israeli
italy, italian, italia, italiane, italiano
itn, independent television news
japan, japanese
jung, jungian
kenya, kenyan
keynes, keynesian
kibbutz, kibbutzim ·
klein, kleinian
korea, korean
labour party
lan, local area network
lbc, london broadcasting company
lcc, london county council
lebanon, lebanese
liberal party
library science, librarianship
lse, london school of economics
lt, lte, lptb, london transport
ltd, limited
machiavelli, machiavellian
maff, ministry of agriculture fisheries and food
malta, maltese
man, men
marcuse, marcusian
marx, marxian, marxist
maths, mathematics
matrix, matrices
mexico, mexican
micro computers, microcomputers, micro-computers
microelectronics, micro electronics, micro-electronics
middle ages, medieval
middle class
milan, milanese
mit, massachusetts institute of technology

movies, moving pictures
mp, member of parliament
napoleon, napoleonic
nato, north atlantic treaty organisation
ncb, national coal board
nccl, national council for civil liberties
nelp, north east london polytechnic
new york, ny
new zealand, nz
nf, national front
nft, national film theatre
nhs, national health service
nigeria, nigerian
norway, norwegian, norse
notting hill
nspcc, national society for the prevention of cruelty to children
o and m, organisation and methods
oecd, organisation for economic cooperation and development
official secrets act
opec, organisation of petroleum exporting countries
ou, open university
palestine, palestinian
paraguay, paraguayan
paris, parisian, parisien
pavlov, pavlovian
pcl, polytechnic of central london
persia, persian
peru, peruvian
phenomenon, phenomena
pirate radio
plo, palestine liberation organisation
pnl, polytechnic of north london
poland, polish
poly, polytechnic
porn, porno, pornography
portugal, portuguese
prussia, prussian
rspb, royal society for the protection of birds
rspca, royal society for the prevention of cruelty to animals
russia, russian, soviet, soviet union, ussr, union of soviet socialist republics
scandinavia, scandinavian
scotland, scottish
sdp, social democratic party
serbocroat, serbo-croat, serbo croat, serbocroatian, serbo-croatian, serbo croatian
soap opera
social science
south america, latin america, south american, latin american
south bank polytechnic, polytechnic of the south bank
spain, spanish, hispanic, espana
standard deviation
std, sexually transmitted diseases
sweden, swedish

switzerland, swiss
swp, socialist workers party
tanzania, tanzanian
tax, taxation, taxable, taxability
thatcher, thatcherite, thatcherism
third world, under developed countries, underdeveloped countries,
  under-developed countries, developing countries
tibet, tibetan
tuc, trade union congress
turkey, turkish
tv, television
ucca, universities central council on admissions
uruguay, uruguayan
uk, united kingdom, britain, great britain, british isles, gb
un, united nations
us, usa, united states, united states of america
utd, united
uv, ultra violet, ultraviolet
vat, value added tax
vd, venereal disease
vdu, vdt, visual display unit, visual display terminal
venice, venetian
vienna, viennese
vietnam, vietnamese
wales, welsh
welfare state
west indies, west indian
wife, wives
woman, women
working class
world war i, world war one, world war 1, first world war, 1st world war, 1914-18 war
world war ii, world war two, world war 2, second world war, 2nd world war,
  1939-45 war
yugoslavia, yugoslav, yugoslavian
z cars