

```

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
X This report, or parts of it, may be copied for non-commercial      X
X purposes, but only if the source, that is, title, author and version, X
X is stated explicitly as part of the copy.                          X
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

```

(C) 1990 J. W. van Wingen, The Netherlands

VERSION 1.5

1990-12-05
rev. 1992-02-20
rev. 1993-07-01
rev. 1994-04-05

```

+-----+
| A SURVEY OF LEXICAL ORDERS IN VARIOUS LANGUAGES USING THE LATIN SCRIPT |
+-----+

```

A language using the latin script may show the following phenomena that matter with lexical ordering (needed at alphabetic sorting).

- a. a simple alphabet of 26 small letters,
- b. the corresponding capitals,
- c. certain letters, that act as a single letter normally, but as two at sorting, called ligatures, even if they do not look like that,
- d. extra letters added or inserted to the alphabet, creating an extended order,
- e. digraphs of two letters which are allocated a place in the alphabet as a single unit,
- f. combinations of a letter with a diacritical mark that are considered: either as an extra letter (same as in d.), or as being not distinguished with respect to sorting from the letter without one.

In general, it is required that with every language in this class there is established:

- a first level ordering sequence, showing the extended alphabetic order of letters and digraphs,
- a second level ordering sequence, showing letters considered to have equal status to a certain letter of the first level, with respect to ordering, but having a precedence defined between each other,
- the expansion of ligatures to single letters,
- the precedence of small to capital letters or the reverse.

In the schemes attached these rules are specified, as far as these are known. For some languages more than one scheme may be in use, like one for a lexicon and another for the telephone directory.

At the first level there is specified the right order of letters, accented letters and digraphs, that are taken into account for constituting the alphabet. For the ligatures the expansion is given by "letters=ligature". For the second level order, strings of letters are specified, starting with a letter from the first level followed by all these that are equalized to it, separated by a hyphen. The precedence of accents may be or may not be according to the order given. Only for French it is known with certainty that it is. For other languages no exact information was available. Precedence of capital to small letters, as given here, is mostly based on assumption, not authorized by official statements.

1. General

In the tables some letters are omitted that do not normally occur in the language. It is assumed that where these appear, for example in quoting foreign names of persons or places, they are sorted at the usual place in the order.

In general, the customary order of A to Z is kept, but not with Estonian. Mostly, a variant of a letter, like one carrying a diacritic mark, is ordered after the letter itself, but not with Maltese.

If a letter sequence is compared with one resulting from expansion of a single letter, it precedes that. Thus Goethe precedes Göthe.

The lexical order for a certain language is more often than not dependent on unwritten rules rather than on official guidelines. Thus it should not cause surprise when more than one ruling is found, varying per field of application, or between groups of users or business interests. Some differences discovered are summarized in the following paragraphs, ordered per language.

Typical application categories are:

Dictionaries

Encyclopedias

Lists of geographic names (railway time tables)

Telephone directories

Dutch

The Van Dale Dictionary sorts 'ij' as two letters, but railway and telephone as equalised with 'y'. Then 'ij' precedes 'y'.

German

The order in Duden, Rechtschreibung, deviates from the customary order used in encyclopedias, catalogs and name lists (which is followed in the table). Accordingly, in dictionaries unlauded letters are not expanded but equalised with the letter. The ß precedes 'ss' in Duden.

Hungarian

The Railway Timetable does not equalise á to a and é to e, like is ruled by the Academy of Sciences

Czech / Slovak

The list as given serves Czech and Slovak combined. However, some letters occur only in Czech, others only in Slovak.

Czech only

E WITH CARON

R WITH CARON

U WITH RING ABOVE

Slovak only

A WITH DIAERESIS

O WITH CIRCUMFLEX

DZ

DZ WITH CARON ON Z

L WITH CARON

L WITH ACUTE

R WITH ACUTE

2. Letters

A justification for the selection of accented letters such as are required in each language is given in the Complete Repertoire of Letters included in the International Standard ISO 6937, (see below). The criterium for including letters used there distinguishes letters required for the language proper, those occurring in borrowed words, next to letters occurring in personal names of foreign origin. This last category is not included in the survey presented here.

3. References

Complete Repertoire of Letters included in the International Standard ISO 6937, as contributed to ISO / IEC JTC1 / SC2 / WG3 by J. W. van Wingen, January 1991. (An updated version is in preparation.)

Règles du classement alphabétique en langue française et procédure informatisée pour le tri, Alain LaBonté, Québec, Canada, Aug. 1988

Sort order schemes in different languages, ISO/IEC JTC1/SC2 N 2111, Johan W. van Wingen, Jan. 1989

Keys to sort and search for culturally expected results, Denis Garneau, 1990, IBM International Technical Support Centre, Roanoke, Texas, USA, document number GG24-3516

Nordic Cultural Requirements on Information Technology, Report TS3, Icelandic Council for Standardization, Reykjavík, 1992

4. Acknowledgements

Besides the people mentioned above, the author received valuable suggestions by e-mail from the following people.

Olle Järnefors, Stockholm, S
Erik Naggum, Oslo, N
Keld Simonsen, Copenhagen, DK
Kevin P. Donnelly, Edinburgh, UK
Andrew Hawke, Aberystwyth, UK
Michael Everson, Dublin, IR
Dimitri Vulis, New York, USA

LEXICAL ORDER FOR ALPHABETS (LATIN SCRIPT)

J. W. van Wingen

1990-09-13 Version 1.5

rev. 1991-01-25, 1992-03-02, 1993-09-30, 1994-04-05

Included are official languages (in Europe) or minority or regional languages for which an official authority has been established:

Changes have been made to Icelandic, Frisian, Catalan, Croat (1.3).

Changes have been made to ROMANIAN, CROAT (1.3 -> 1.5), SWEDISH.

Not included yet (in Europe):

GREENLANDIC, RUMANTSCH (RHAETIAN), LETZEBURGESCH, FAROESE, SAMI, SORBIAN, BASQUE, GALICIAN

```

+-----+
|LITHUANIAN:
|First level order:
|  a b c č d e f g h i j k l m
|  n o p q r s š t u v w x y z ž
|Ligatures:
|  none
|Second level order:
|  a-ą e-ė-ę i-į-y u-ū-ų
|Small letters preceding capitals
+-----+
    
```

```

+-----+
|LATVIAN:
|First level order:
|  a ā b c č ch d dz dž e ē f g ģ h i ī ie j k ķ l ļ m
|  n o o ō p r r s š t u ū v z ž
|Ligatures:
|  none
|Second level order:
|  none
|Small letters preceding capitals
+-----+
    
```

```

+-----+
|ESTONIAN:
|First level order:
|  a b c d e f g h i j k l m
|  n o p q r s š z ž t u v õ ä ö ü x y
|Ligatures:
|  none
|Second level order:
|  õ-õ š-sh ž-zh
|Capitals preceding small letters
+-----+
    
```

```

+-----+
|SWEDISH / FINNISH:
|First level order:
|  a b c d e f g h i j k l m
|  n o p q r s t u v x y z å ä ö
|Ligatures:
|  none
|Second level order:
|  a-á-à e-é-ë y-ü ö-ø v-w
|Capitals preceding small letters
+-----+
    
```

```
+-----+
|DANISH / NORWEGIAN:
|First level order:
|  a b c d e f g h i j k l m
|  n o p q r s t u v w x y z æ ø å
|Ligatures:
|  none
|Second level order:
|  e-é y-ü æ-ä ø-ö å-aa
|Capitals preceding small letters
+-----+
```

```
+-----+
|ICELANDIC:
|First level order:
|  a á b d ð e é f g h i í j k l m
|  n o ó p r s t u ú v x y ý z þ æ ö
|Ligatures:
|  none
|Second level order:
|  ö-ø v-w
|Capitals preceding small letters
+-----+
```

```
+-----+
|GAELIC:
|First level order:
|  a b c d e f g h i j k l m
|  n o p q r s t u v w x y z
|Ligatures:
|  none
|Second level order:
|  a-á-à e-é-è i-í-ì o-ò u-ù
|Small letters preceding capitals
+-----+
```

```
+-----+
|IRISH:
|First level order:
|  a b c d e f g h i j k l m
|  n o p q r s t u v w x y z
|Ligatures:
|  none
|Second level order:
|  a-á e-é i-í o-ó u-ú
|Small letters preceding capitals
+-----+
```

```
+-----+
|WELSH:
|First level order:
|  a b c ch d dd e f ff g ng h i j l ll m
|  n o p ph r rh s t th u w y          (k,q,v,x,z as usual if needed)
|Ligatures:
|  none
|Second level order:
|  a-á-à-â-ä e-é-è-ê-ë i-í-ï-î-ï o-ó-ò-ô-ö u-ú-ù-ü-ü
|  w-ŵ-ŵ-ŵ-ŵ y-ý-ÿ-ÿ-ÿ
|Small letters preceding capitals
+-----+
```

```
+-----+
|BRETON:
|First level order:
|  a b c h c'h d e f g h i j k l m
|  n o p r s t u v y z
|Ligatures:
|  none
|Second level order:
|  e-ê n-ñ u-ù-ü
|Small letters preceding capitals
+-----+
```

```
+-----+
|DUTCH:
|First level order:
|  a b c d e f g h i j k l m
|  n o p q r s t u v w x y z
|Ligatures:
|  ij=ij (the first ij is understood to be a single letter)
|Second level order:
|  a-á-à-ä e-é-è-ê-ë i-í-ï o-ó-ö u-ú-ü y-ij
|Capitals preceding small letters
+-----+
```

```
+-----+
|FRISIAN:
|First level order:
|  a b c d e f g h i j k l m
|  n o p q r s t u v w x y z
|Ligatures:
|  none
|Second level order:
|  a-à-â-ä e-é-ê-ë i-ï o-ô-ö u-û-ü-Û y-ij
|Capitals preceding small letters
+-----+
```

```
+-----+
|GERMAN:
|First level order:
|  a b c d e f g h i j k l m
|  n o p q r s t u v w x y z
|Ligatures:
|  ae=ä oe=ö ue=ü ss=ß
|Second level order:
|  a-â e-é-ê-ë o-ô
|Small letters preceding capitals
+-----+
```

```
+-----+
|FRENCH:
|First level order:
|  a b c d e f g h i j k l m
|  n o p q r s t u v w x y z
|Ligatures:
|  ae=æ oe=e
|Second level order (applied in reverse):
|  a-à-â c-ç e-é-ê-ë i-ï o-ô u-û-ü-Û y-ÿ
|Small letters preceding capitals
+-----+
```

+-----+
|SPANISH:
|First level order:
| a b c ch d e f g h i j k l ll m
| n ñ o p q r s t u v w x y z
|Ligatures:
| none
|Second level order:
| a-á e-é i-í o-ó u-ú-ü
|Small letters preceding capitals
+-----+

+-----+
|CATALAN:
|First level order:
| a b c d e f g h i j k l m
| n o p q r s t u v w x y z
|Ligatures:
| none
|Second level order:
| a-à c-ç e-é-è i-í-ï o-ó-ò u-ú-ü
|Small letters preceding capitals
+-----+

+-----+
|PORTUGUESE:
|First level order:
| a b c d e f g h i j k l m
| n o p q r s t u v w x y z
|Ligatures:
| none
|Second level order:
| a-á-â-ã-ä c-ç e-é-ê-ë i-í o-ó-ô-õ u-ú-ü
|Small letters preceding capitals
+-----+

+-----+
|MALTESE:
|First level order:
| a b ċ d e f ġ g ħ h i j k l m
| n għ o p q r s t u v w x ž z
|Ligatures:
| none
|Second level order:
| għ-'
|Small letters preceding capitals
+-----+

+-----+
|ITALIAN:
|First level order:
| a b c d e f g h i j k l m
| n o p q r s t u v w x y z
|Ligatures:
| none
|Second level order:
| a-à e-é-è i-í-ì-ï o-ó-ò u-ú-ü
|Small letters preceding capitals
+-----+

```
+-----+
|ROMANIAN:
|First level order:
|  a â ã b c d e f g h i î j k l m
|  n o p q r s ș t ț u v w x y z
|Ligatures:
|  none
|Second level order:
|  none
|Small letters preceding capitals
+-----+
```

```
+-----+
|HUNGARIAN:
|First level order:
|  a á b c cs d dz dzs e é f g gy h i j k l ly m
|  n ny o ö p q r s sz t ty u ü v w x y z zs
|Ligatures:
|  none
|Second level order:
|  i-í o-ó ö-ő u-ú ü-ű
|Small letters preceding capitals
+-----+
```

```
+-----+
|ALBANIAN:
|First level order:
|  a b c ç d dh e ë f g gj h i j k l ll m
|  n nj o p q r rr s sh t th u v w x xh y z zh
|Ligatures:
|  none
|Second level order:
|  none
|Small letters preceding capitals
+-----+
```

```
+-----+
|TURKISH:
|First level order:
|  a b c ç d e f g ğ h ı i j k l m
|  n o ö p q r s ş t u ü v w x y z
|Ligatures:
|  none
|Second level order:
|  a-â i-î u-û
|Small letters preceding capitals
+-----+
```

```
+-----+
|POLISH:
|First level order:
|  a ą b c ć d e ę f g h i j k l ł m
|  n ń o ó p q r s ś t u v w x y z ź ż
|Ligatures:
|  none
|Second level order:
|  none
|Small letters preceding capitals
+-----+
```



```
+-----+
|SLOVENE:
|First level order:
|  a b c č d e f g h i j k l m
|  n o p q r s š t u v w x y z ž
|Ligatures:
|  none
|Second level order:
|  none
|Small letters preceding capitals
+-----+
```

```
+-----+
|CROAT:
|First level order:
|  a b c č ć d dž đ e f g h i j k l lj m
|  n nj o p q r s š t u v w x y z ž
|Ligatures:
|  none
|Second level order:
|  đ-dj
|Small letters preceding capitals
+-----+
```

```
+-----+
|CZECH / SLOVAK:
|First level order:
|  a ä b c č d dz dž e f g h ch i j k l m
|  n o ô p q r ř s š t u v w x y z ž
|Ligatures:
|  none
|Second level order:
|  a-á d-ď e-é-ě i-í l-ĺ-ļ
|  n-ň o-ó r-ř t-ť u-ú-ů y-ý
|Small letters preceding capitals
+-----+
```