n Documentation n

Jörg Knappen

Barbarossaring 43

55118 Mainz

email: `knappen@vkpmzd.kph.uni-mainz.de`

1mJUNm1996

1 Introduction

1.1 The dc fonts

In 1990 at the TUG meeting at Cork, Ireland, the european TEX user groups agreed on a 256 character incoding supporting many european languages with latin writing. This encoding is both an *internal encoding* for TEX and a *font encoding*. This double nature is a consequence of the fact, that both kind of encodings cannot be entirely separated within TEX.

The design goals of the Cork encoding are to allow as many languages as possible to be hyphenated correctly and to guarantee correct kerning for those languages. Therefore it includes many ready-made accented letters.

It also includes some innovative features, which have not become very popular yet, though they deserve to become so. First to mention is a special, zero width invisible character, the compound word mark (cwm). Its design purpose is to assist correct hyphenation at morpheme boundaries, which can occur in minimal pairs (e. g. german *Wachs-tube* vs. *Wach-stube*), if the previous example is typeset in *fraktur* different styles of the letter isj are needed.

In the new version dc 1.3 I have given the cwm a height, namely x-height, in order be function as a carrier of accents which are placed between letters like in the german abbreviature *-buj027g. (-burg)*.

The second innovative feature is the separation of the two characters <hyphen> and <hyphenchar>. The hyphenchar is designed as a hanging hyphen, giving a smother text boundary.

The final version of the Cork encoded fonts will be called ec (European Computer Modern or Extended Computer Modern) fonts. The current version, dc 1.3, is the last intermediate step towards the final version. Note, that in the cause of bug fixes and improvements, the metrics may change. After the renaming to ec the metrics will remain stable, as the metrics of the cm fonts do.

1.2 The tc-fonts

The need for a text companion font was first articulated in the discussion of new 256 character mathematical fonts in 1993. In order to achieve a better orthogonality

The twelve dex if the math, font list as symbols listed in the inalnln fonfsphoufct be moyed te cube text coltpánidnfouty pub/tex/mathfont.. The text companion fonts are also the ideal place to store some new characters, like currency symbols.

There are now 108 characters in the tc fonts, including special uppercase versions of the accents, oldstyle digits, genealogical symbols, footnote symbols, currency symbols and other custom characters. The selection is unique and superior to most commercially available expert sets.

2 Supported languages

The following languages are supported by the Cork encoding: Afrikaans, Albanian, Breton, Croat, Czech, Danish, Dutch, English, Estonian, Faroese, Finnish, French, Frisian, Gaelic, Galician, German, Greenlandic, Hungarian, Icelandic, Irish (modern orthography), Italian, Letzeburgish, Lusatian (Sorbian), Norwegian, Polish, Portuguese, Rhaetian (Rumantsch), Rumanian, Slovak, Slovene, Spanish, Swedish, Turkish. Many non-european languages using the standard latin alphabet (e. g. Bahasa Indonesia, Suaheli) are also supported.

In europe, the following languages arenjt supported: Azeri, Basque, Catalan, Esperanto, Irish (old orthography), Latvian, Lithuanian, Maltese, Sami, Welsh. Of course, Greek and all languages with cyrillic writing are outside the scope of the Cork encoding.

3 Standard Control Sequences

The following standard control sequences are assigned with LaTEXjs `T1` encoding for the dc fonts:

° Ring accent (°u gives u)
Ogonek (e gives e)
, Icelandic letter edh (,)
, Letter d with stroke (,)
, Letter eng (,)
, Icelandic letter thorn (,).
To load the tc fonts, use the `textcomp`-package, which provides control sequences for all the included symbols.

4 Ligatures

In the proportional fonts, the following ligatures are implemented:

```
--                                              m (en dash)
---                                             n (em dash)
``                                              k (english opening quotes, german closing quotes)
''                                              l (english and polish closing quotes)
,,                                              „ (german and polish opening quotes)
<<                                              « (french opening quotes)
>>                                              » (french closing quotes)
!`                                              ¡ (spanish opening exclamation mark)
?`                                              ¿ (spanish opening question mark)
fi                                              fi
ff                                              ff
fl                                              fl
ffi                                             ffi
ffl                                             ffl
```

In the typewriter fonts, the following ligatures are implemented:
```
--                                              m  (en dash or number range dash)
---                                             n  (em dash)
```

```
``          k  (english opening quotes, german closing quotes)
''          l  (english and polish closing quotes)
,,          „  (german and polish opening quotes)
<<          «  (french opening quotes)
>>          »  (french closing quotes)
!`          ¡   (spanish opening exclamation mark)
?`          ¿   (spanish opening question mark)
```

The convention on the dashes suites british usage for number range dashes best and does not interfer with any other known usage. In verbatim mode, all ligatures are switched off.

5 Hints on usage

The dc fonts are intended for text usage in european languages. The Cork font encoding is selected with the command `\usepackage[T1]fontenc` in LaTEX2e.

The tc fonts are a multi-purpose font. Suggested usages include verbatim setting of latin-1 and latin-2 listings, avoiding the so-called khidden mathl in text mode (thatjs the reason, why there are footnote symbols in), providing building blocks for virtual fonts (oldstyle digits are included for this reason), or just providing otherwise unavailable symbols (like the permille sign).

Some characters are primarily intended for verbatim listings, in plain text they may be replaced with macros. These characters include the raised digits, the fractions, the trademark sign, and the ordinal indicators.

For text fractions, the following macro is suggested (from the TEXbook, exercise 11.6):

```
\newcommand\nicefrac[2]\leavevmode\kern.1em
\raise.5ex\hbox\the\scriptfont0 #1\kern-.1em
/\kern-.15em\lower.25ex\hbox\the\scriptfont0 #2
```

It can produce arbitrary fractions and is not restricted to some simple cases, the output looks .5ex0 1 /0 2, .5ex0 5 /0 4, .5ex0 17 /0 42.

For the ordinal indicators ($1ex_o$ and $1ex_a$), the following macros are suggested (from spanish.ldf, babel bundle):

```
\newcommand\ordmale\raise1ex\hbox\underbar\scriptsize o
\newcommand\ordfemale\raise1ex\hbox\underbar\scriptsize a
```

6 Naming of the font files

Currently, the extended computer modern font have the prefix dc. This prefix will changew to ec with the final release after another round of bug fixing. I hope to make the transition from dc to ec in about one year. The text companion fonts have the prefix tc, which is not subject to change. However, later releases may included more characters and therefore have different checksums. No characters shall be removed from the tc fonts.

Most of the dc fonts can be generated at any size one want in the range from 5pt to 100pt. For each size, a unique name is needed.

With the release 1.2 of the dc fonts, a new, more precise naming scheme is in effect. Since there are widely used operating sytems limiting the file name to 8 character (plus an extension of 3 characters) the following scheme is used:

The first two letters (either dc or tc denote the encoding and the general design of the font.

The one or two following letters denotes the family, shape, and series attributes of the font. E. g. r for roman, bx for bold extended, ti for text italic, or bi for bold extended italic. A complete overview is given at the end of this section.

The following four digits give the design size in TEXjs points multiplied with 100. E. g 1000 denotes tex point, 1440 denotes magstep 2, i. e. 14.4 point, and 0500 denotes five point.

Here are the implemented styles:

Roman family: r roman, b bold, bx bold extended, sl slanted, bl bold extended slanted, cc caps and small caps, xc bold extended caps and small caps, sc slanted caps and small caps, oc oblique (bold extended slanted) caps and small caps, ti (text) italic, bi bold extended italic, u unslanted italic, ci classical serif italic (new design).

Sans serif family: ss sans serif, si sans serif inclined (slanted), sx sans serif bold extended, so sans serif bold extended oblique (slanted).

Typewriter family: tt typewriter, tc typewriter caps and small caps, st slanted typewriter, it italic typewriter.

Variable width typewriter family vt variable width typewriter, vi variable width italic typewriter.

Various other fonts: bm variant bold roman,