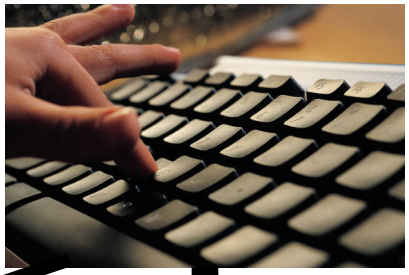


UNICODE

DoBeS Training Course

Paul Trilsbeek

Character input/output relations

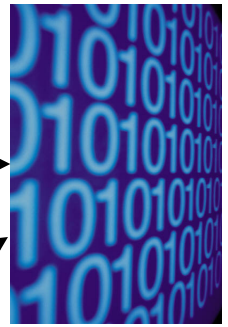


keyboard

Input method

mapping

Character set



file

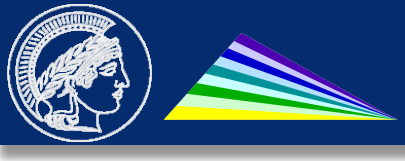
Font

printer



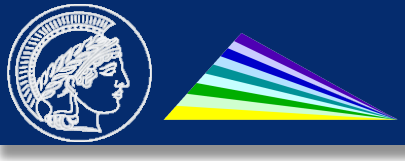


display



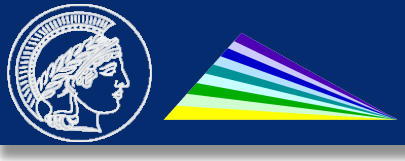
Character encoding

- 7-bit: $2^7 = 128$ different possibilities
  - ASCII
- 8-bit (1 byte): $2^8 = 256$ different possibilities
  - Extended ASCII, ISO-8859 “codepages” (e.g. ISO-latin-1)
- 16-bit (2 bytes): $2^{16} = 65536$ different possibilities
  - UNICODE
- 32-bit (4 bytes): $2^{32} = 4,294,967,296$ different possibilities
  - Extended UNICODE



UNICODE

- *"Unicode provides a unique number for every character, no matter what the platform, no matter what the program, no matter what the language." (www.unicode.org)*
- Attempt to create one large universal character set
- Continually expanding -> UNICODE consortium decides whether proposed characters are added or not

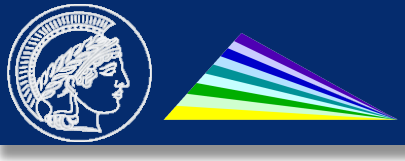


UNICODE

- First efforts started around 1989
- Version 1.0 appeared in 1992
- Slow adaptation, one reason was the lack of fonts (is still an issue)
- Last couple of years support has grown rapidly: operating systems, word processors, web browsers, email programs, etc. mostly support UNICODE

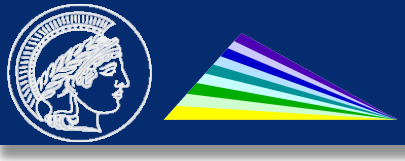
UNICODE

- Current version 6.0 has 109,449 characters
- Characters are organized in “blocks”, e.g. the “IPA Extensions” block
- Currently there are a few large UNICODE fonts, such as MS Arial UNICODE which contains over 50,000 glyphs, and many smaller UNICODE fonts for a specific language or set of characters.



UNICODE

- Problems:
  - “Your character” is not in the UNICODE set. You can propose the character to be added and hope that the committee will accept it.
  - There is no UNICODE font that contains your character(s). There might be people working on a font that does contain it, but if not, you would need to create one yourself.

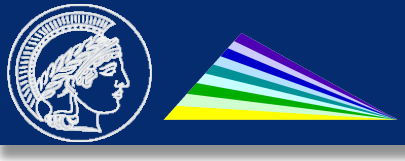


UNICODE

- Problems:
  - Sometimes UNICODE handling in certain software applications can be problematic, e.g. when using multiple diacritics on a single character

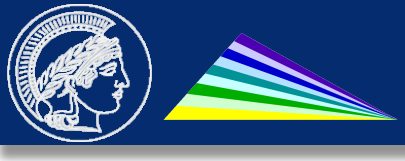
Keyman

- Utility for Windows to map keys on the keyboard to characters in a character set.
- Commercial program from Tavultesoft, http://www.tavultesoft.com
- 3 versions:
  - “light” version: supports 2 keyboard layouts max., € 14,99
  - “Pro” version: supports unlimited layouts and offers some more features, € 29,99
  - “Developer” version: needed in order to create your own keyboard mappings, € 119,99 (educational prices)



Keyman

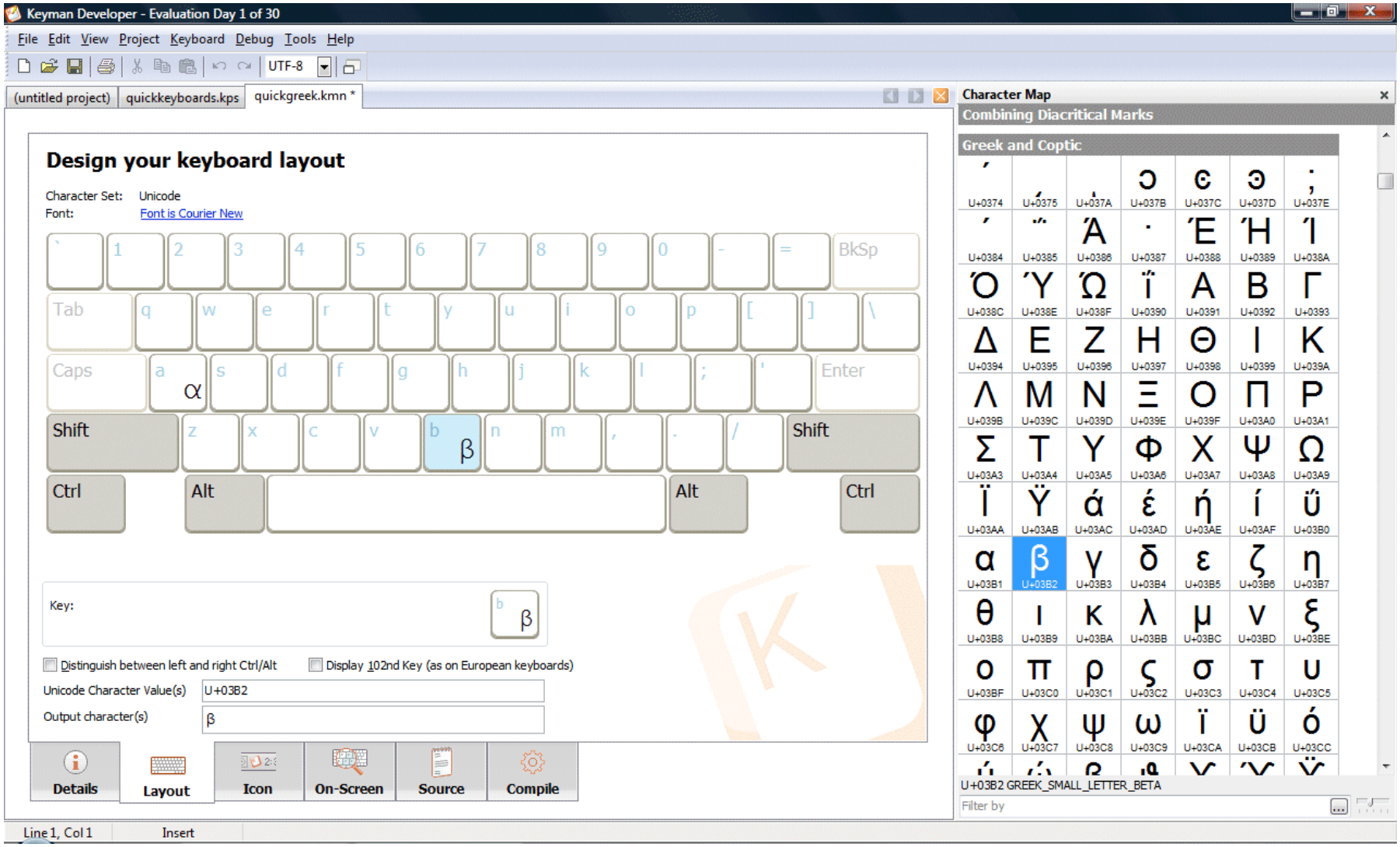
- One key can map to a sequence of characters, e.g. a character including one ore more diacritics.
- When mapping only a diacritic to a key, it will behave as a “modifier” key which you can use to add diacritics to any character.



Keyman

- If “normal” keys of the keyboard are also still needed, one needs to define key combinations or look for very rarely used keys

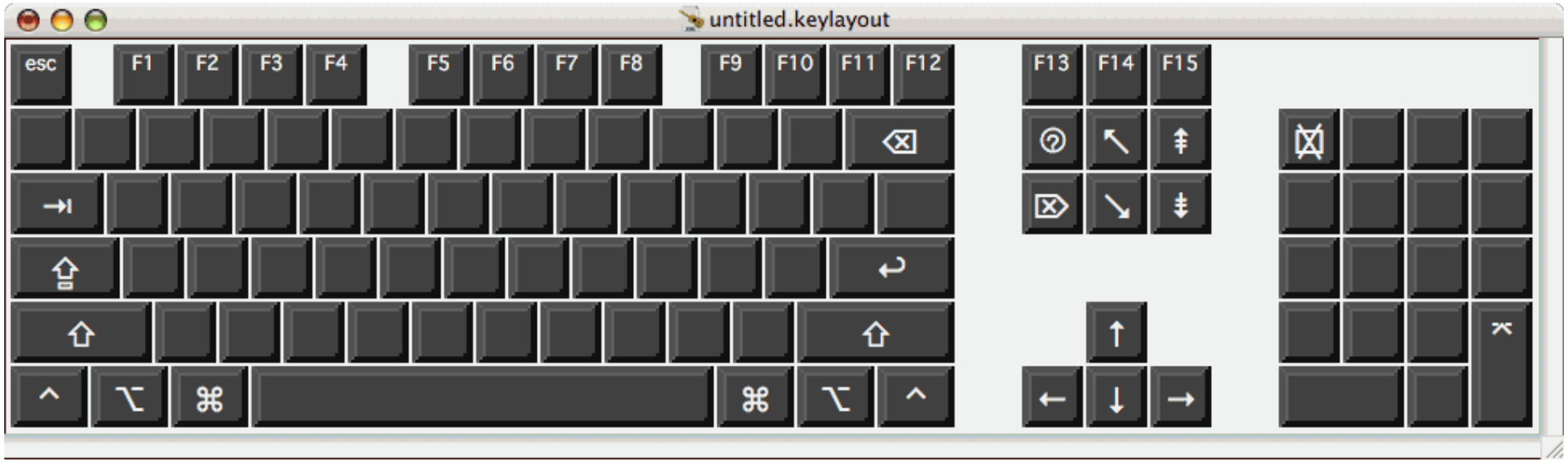
Keyman


Keyman Developer - Evaluation Day 1 of 30
File Edit View Project Keyboard Debug Tools Help
UTF-8
(untitled project)
quickkeyboards.kps
quickgreek.kmn *
Design your keyboard layout
Character Set: Unicode
Font: Font is Courier New
Key:
Distinguish between left and right Ctrl/Alt
Display 102nd Key (as on European keyboards)
Unicode Character Value(s) U+03B2
Output character(s) β
Details
Layout
Icon
On-Screen
Source
Compile
Line 1, Col 1
Insert
Character Map
Combining Diacritical Marks
Greek and Coptic
U+03B2 GREEK_SMALL_LETTER_BETA
Filter by

Ukelele

- Keyboard mapping tool for Mac, developed by SIL.
- Creates XML keyboard layout files for use with MacOS X.
- Free download from sil.org

Ukelele


untitled.keylayout
esc
F1
F2
F3
F4
F5
F6
F7
F8
F9
F10
F11
F12
F13
F14
F15