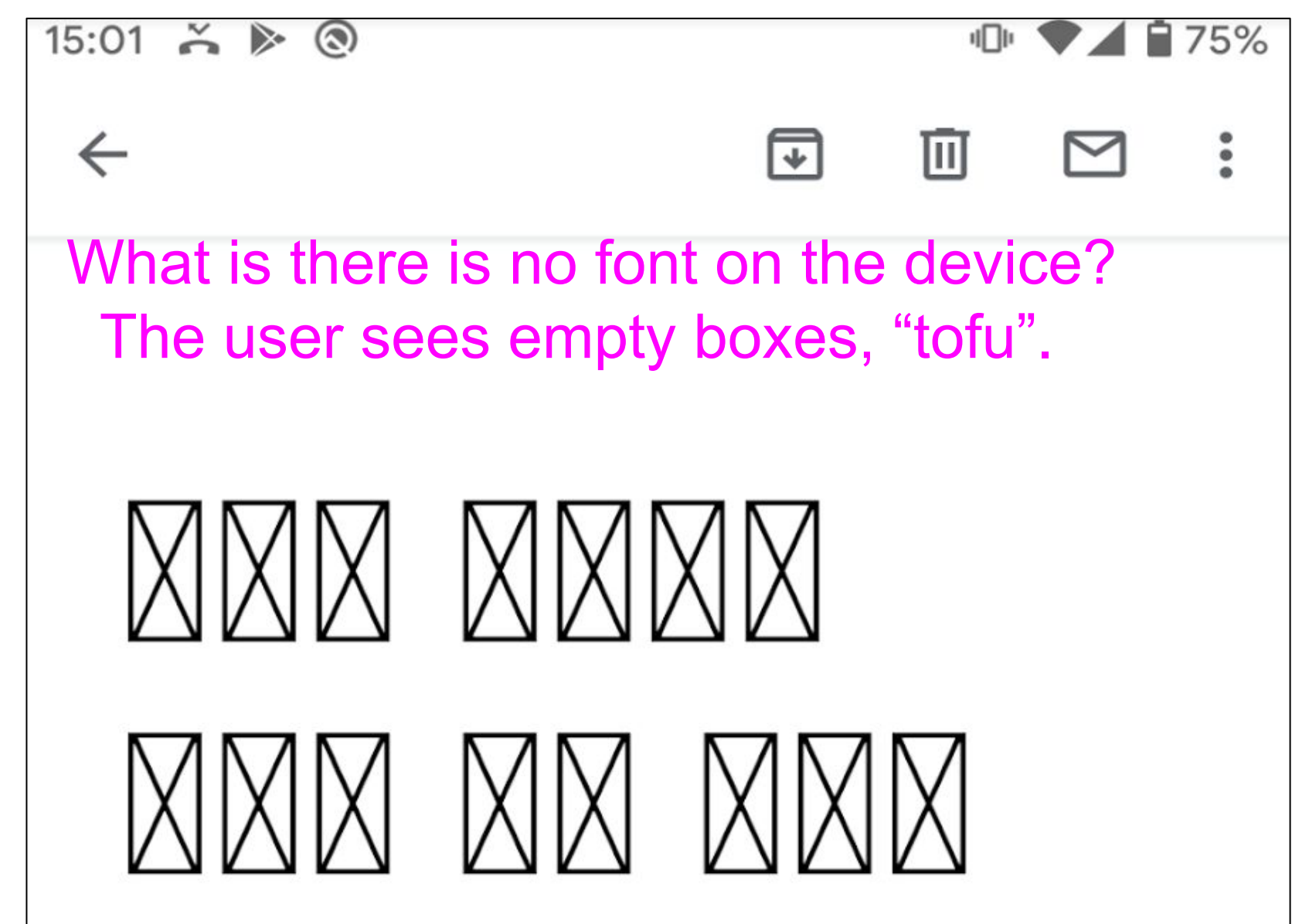
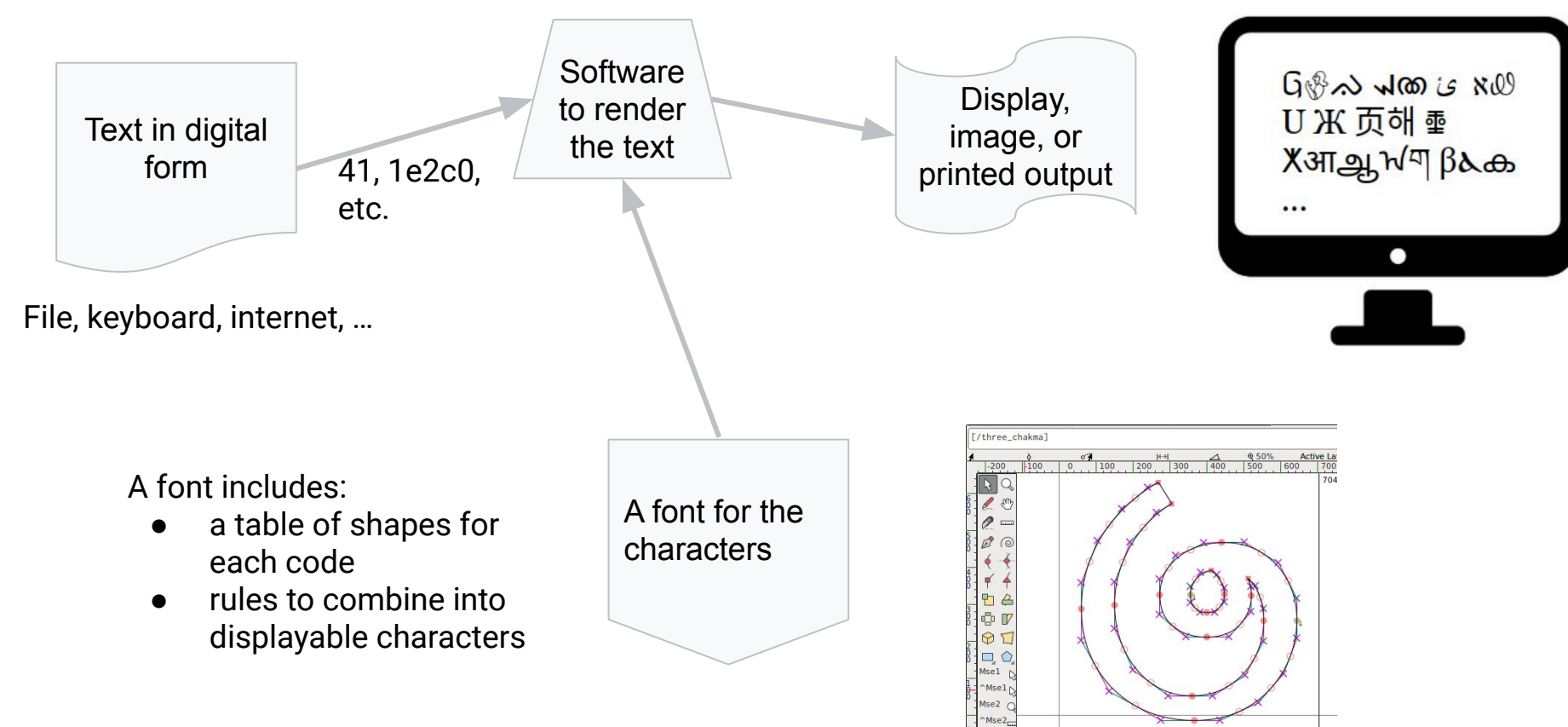


# Unicode for Indigenous Languages

## Standards and technology for getting online

Craig Cornelius, Senior Software Engineer, Google

### How text works on a computer / mobile device



### A pre-internet practice: use a font with modified shapes for the characters - a *font-encoding*.

Change the shapes (glyphs) for codes of an existing font to show desired characters. This works because a font simply defines the shapes for each of the digital text codes

#### Advantages:

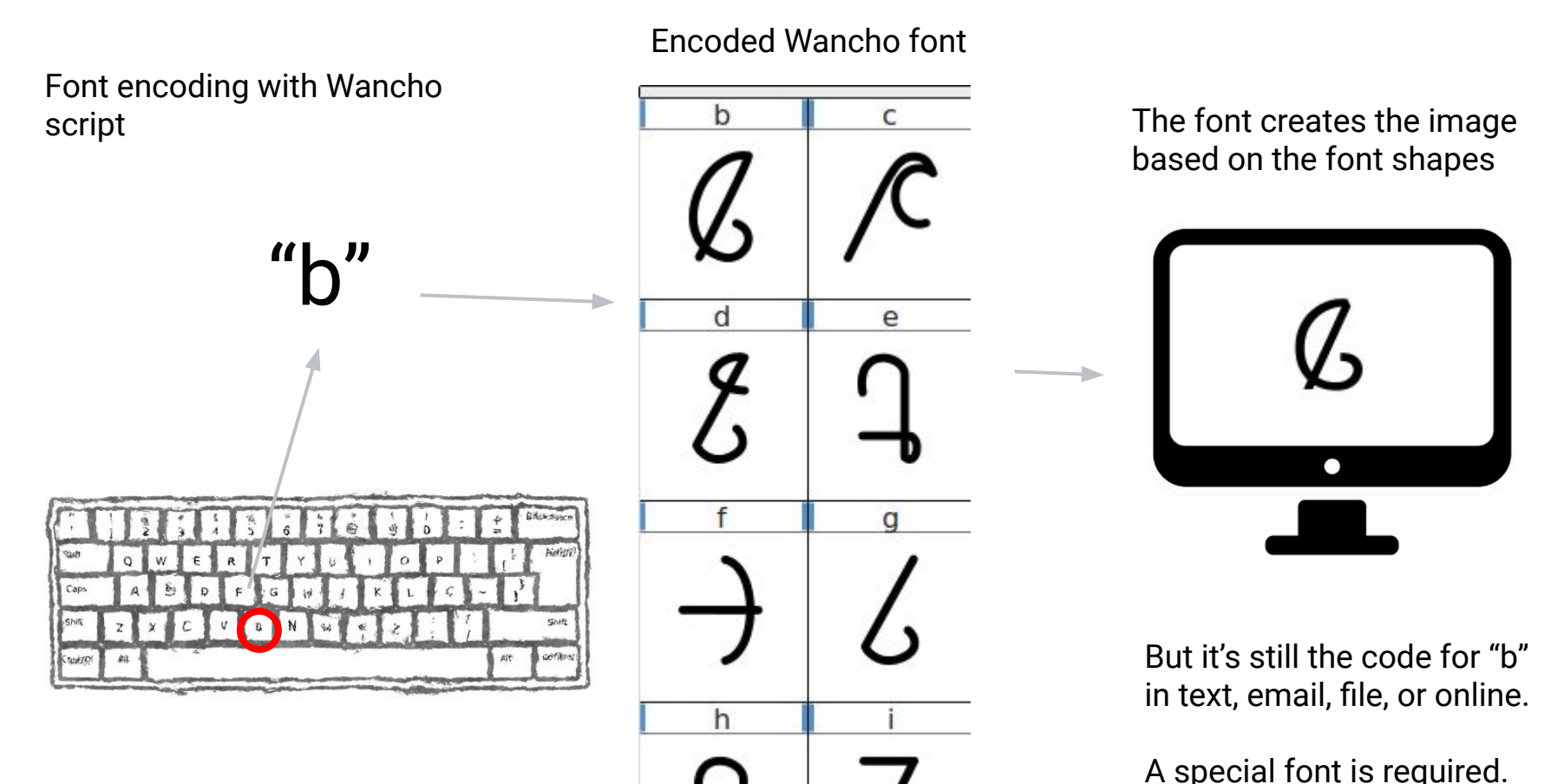
- Easy to create special fonts
- Simple to use in applications
- Font is easily shared with community
- Allows use in documents, newspapers, education, etc.
- Works for online websites when font is installed locally
  - Web fonts: embed font with the site

#### Problems:

- Fails if the font is not present on computer / device, especially on mobile.
- Since characters are redefined, no text processing works right. This includes casing, spell check, search, etc.
- Does not work for users without the installed font for websites, blogs, etc.

b	c
Q	Q
d	e
Q	Q
f	g
Q	Q
h	i
Q	Q

### Font encoding on a computer, *if font is present*



### The solution: Unicode - a standard for all scripts

Encoding for the writing systems of the world:

- Each character has a unique number, never reused
- Each code such as U+0416 includes:
  - **name**, e.g., CYRILLIC CAPITAL LETTER ZHE
  - **representative shape**, e.g., Ж
  - **properties**
    - Type (letter, digit, punctuation, space, combining mark, etc.)
    - Casing (upper or lower)
    - Sort order
    - Direction of text (RTL, LTR, vertical)
    - And more...



### More about Unicode

An international standard for a “**unique, unified, universal encoding**” Envisioned in 1987. First release in 1991 by the Unicode Consortium ([unicode.org](http://unicode.org))

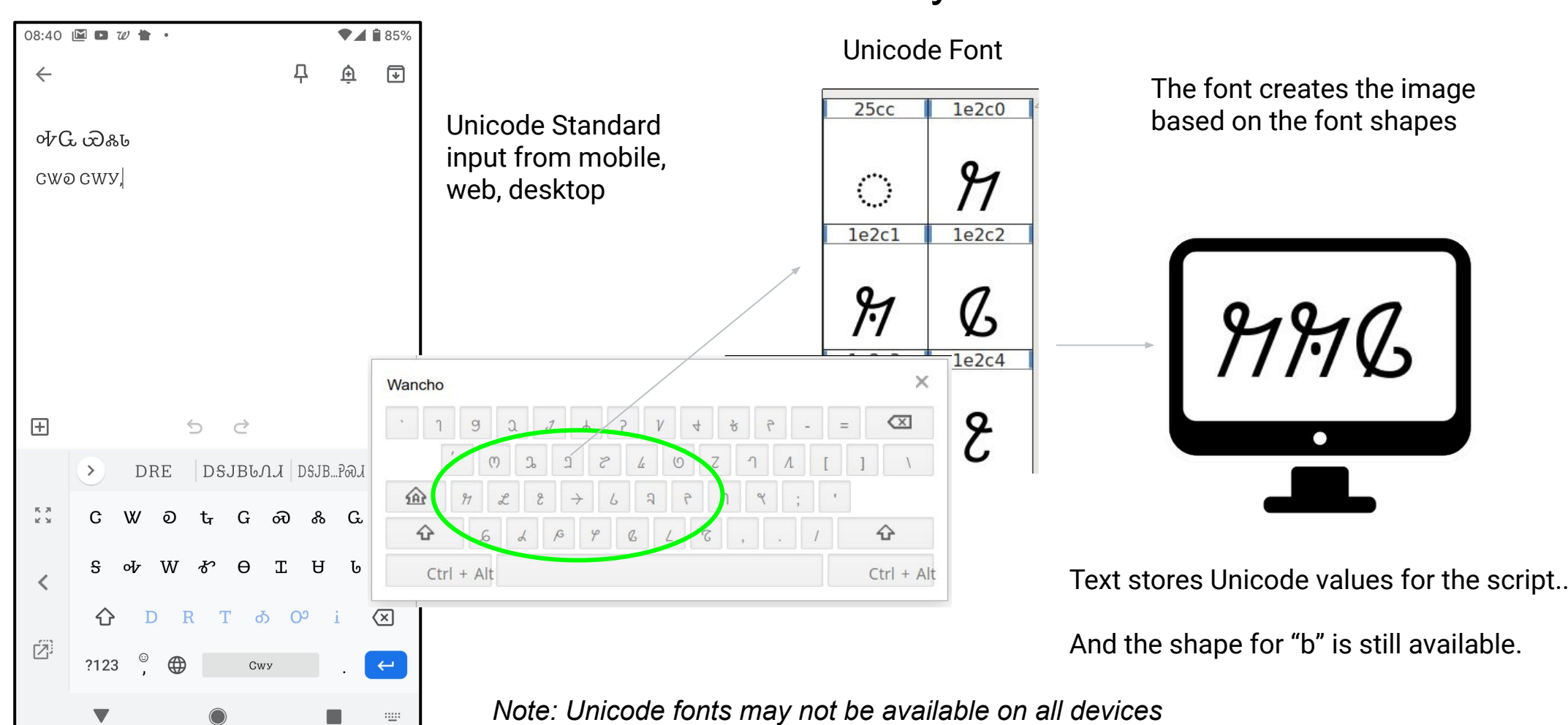
Version 12.0 (2019) has 150 scripts, 137,994 codes. New versions are released annually.

- Present on all modern computers and mobile
- Stable: codes never removed
- An open standard via Unicode proposal process

Supported by International Components for Unicode (ICU) software, free as open-source libraries



### Here's how Unicode works on any device



### Already in Unicode! Ready for your language!

Indigenous languages use Unicode

- Most living writing systems are already in Unicode. Every new device has it already!

To use your language:

0. Recruit community champions
1. Choose a writing system (or propose for standardization)
2. Find/create Unicode fonts
3. Find/create input methods in Unicode
4. Use the language online, texting, documents, social media, blogs, web sites



[Google Noto Fonts](#)  
Beautiful and free fonts  
for all languages

A list of some of the scripts in Unicode: Arabic, Gujarati, Lao, Armenian, Gurmukhi, Latin, Bengali, Han, Malayalam, Bopomofo, Hangul, Oriya, Cyrillic, Hebrew, Tamil, Devanagari, Hiragana, Telugu, Georgian, Kannada, Thai, Greek, Katakana, Mongolian, Syria, Canadian Syllabics, Myanmar, Thaana, Cherokee, Ogham, Yi, Ethiopic, Runic, Khmer, Singala, Deseret, Gothic, Buhid, Tagalog, Hanunóo, ... Elymaic, Nyaikeng Puachue Hmong, Nandinagari, Wancho as of Unicode 12.0