

Language Support in Wikipedia

Siebrand Mazeland

Product Manager Localisation

Wikimedia Foundation



Wikipedia and MediaWiki

- MediaWiki is the software Wikipedias and sister projects run on (~800 wikis - 20,000,000+ pages)
- Hundreds of thousands installations elsewhere
- Usable in over 300 languages
 - ~7,500 Languages classified



WIKIPEDIA
The Free Encyclopedia



Powered By
MediaWiki

What is language support about

- ✦ User Interface - How does MediaWiki “talk” against its users
- ✦ Input - How can content be contributed in MediaWiki (scripts)
- ✦ Output - How is content being represented in MediaWiki (fonts, script



Who does it

- ✦ Volunteers
 - ✦ Translation
 - ✦ Describe language properties
 - ✦ Contribute to language related code
- ✦ Wikimedia Foundation
 - ✦ Localisation team with developers
 - ✦ Other developers



http://commons.wikimedia.org/wiki/File:Wikimedia_Hackathon_Berlin_2011_group_photo.jpg
CC-BY-SA 3.0 DE / Tobias Schumann

User Interface Internationalisation

- Internationalisation is making software ready for using it in multiple languages



[http://commons.wikimedia.org/wiki/File:Petersburger Platz.JPG](http://commons.wikimedia.org/wiki/File:Petersburger_Platz.JPG)
CC-BY-SA 3.0 / Achim Raschka

User Interface

Internationalisation

- Plural forms
- Gender
- Namespace names, magic words, URL of special pages
- Prefix and suffix for grammatical cases, number conversion for non-Latin scripts, toolbar buttons, script conversion (for example Latin<->Cyrillic for Serbian)
- Some Western European languages look a lot like English, but in India that's very far from the truth



User Interface

Localisation

- ✦ Localisation is the process of translating a product for use in a language
 - ✦ Text
 - ✦ Icons
 - ✦ Symbolism/culture



Logo translatewiki.net / © 2008 Danny B.

User Interface

Localisation

- ✦ Translate MediaWiki in as many languages as possible at translatewiki.net
- ✦ Checking English language messages
 - ✦ Style, patchwork messages, <the things developers do>...
 - ✦ Add support for gender and plural forms
 - ✦ Namespace names, magic words, special page URLs
 - ✦ Date and time notation and more...

Input

Not every language is the same

- ✦ Some languages are still in the digital Middle Ages
- ✦ No or developing Unicode standard
- ✦ No keyboards



Tibetan keyboard / MIT license / [[en:User:EdC]]

Input

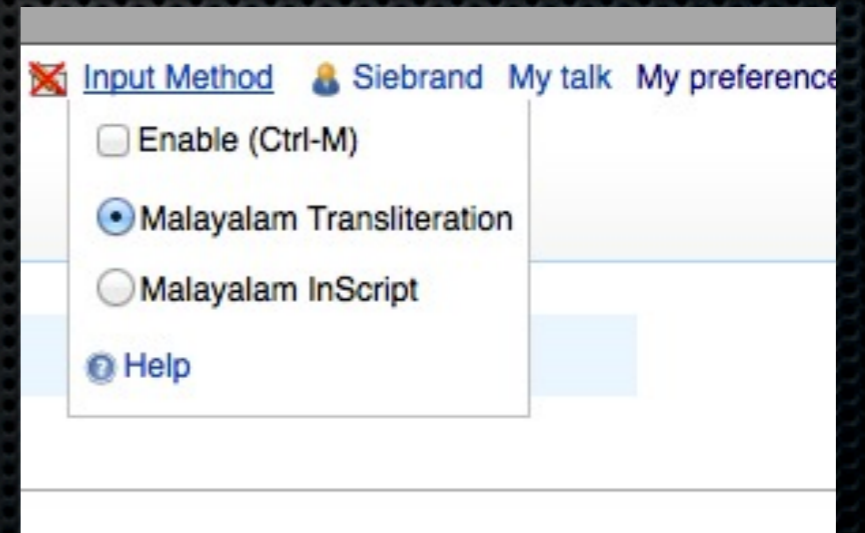
Not every user is the same

- ✦ Not always using the same computer
- ✦ Not always the same keyboard
- ✦ India, Africa
 - ✦ Bengali, Hindi, Tamil, Nepali, Assamese, Malayalam, Amharic, Marathi, Konkani, ...

Input

[[mw:Extension:Narayam]]

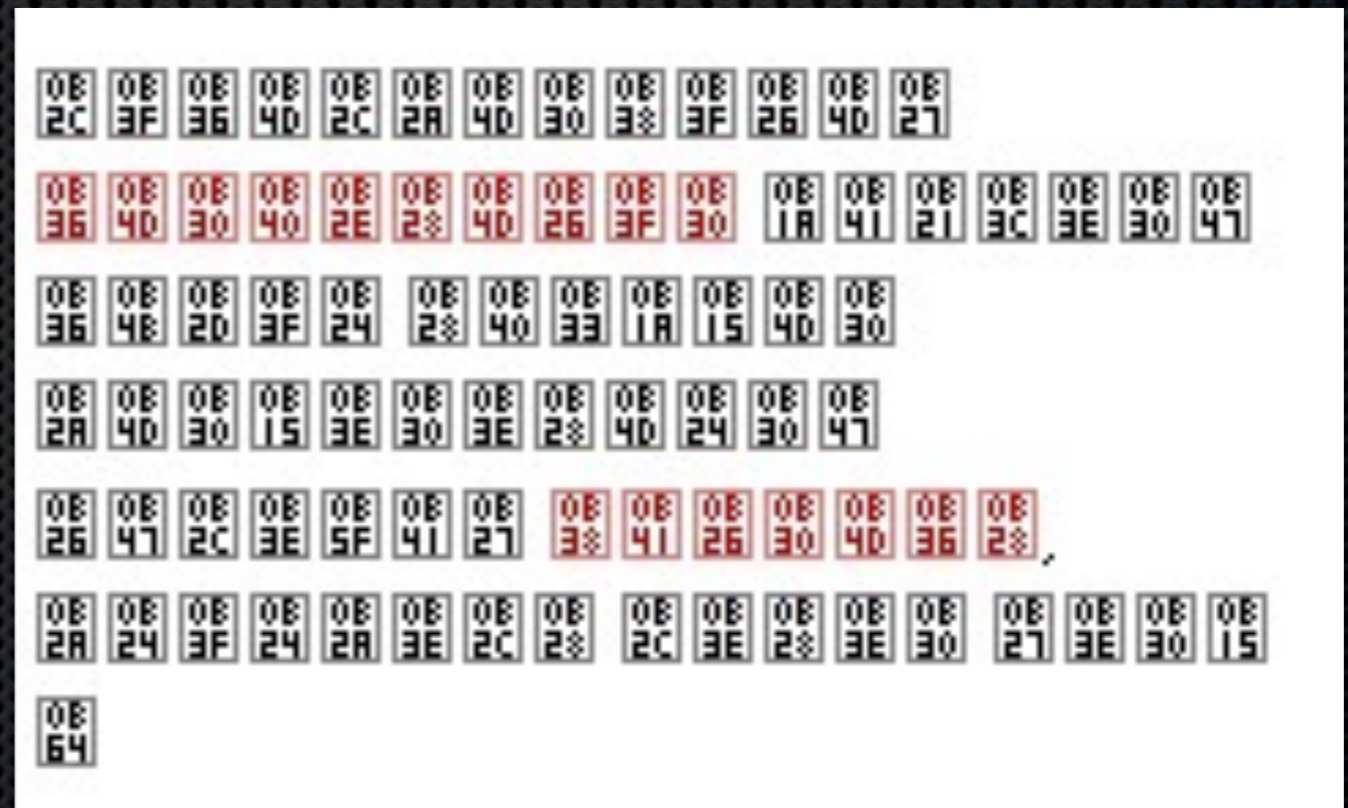
- ✦ Choose your keyboard layout in MediaWiki
- ✦ Future: on-screen keyboard for even better accessibility



Output

Fonts - The key to correct display

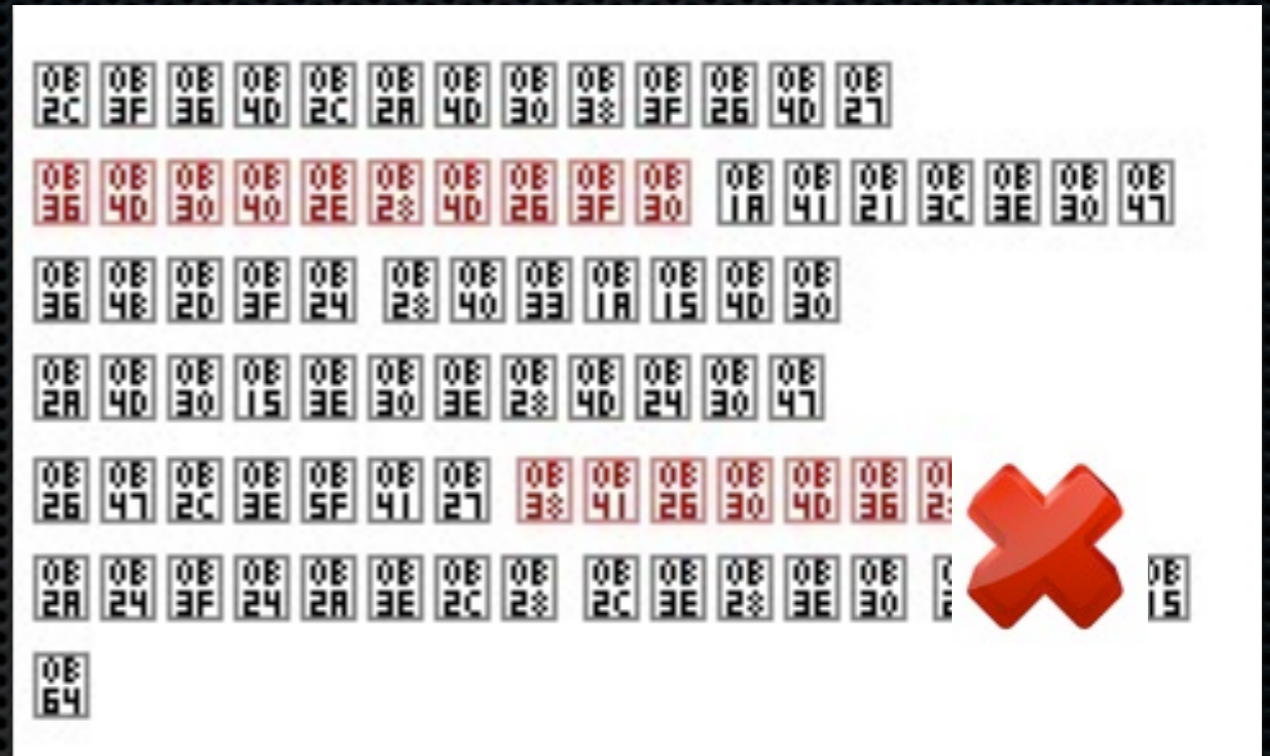
- ✦ Some scripts are still in the digital Middle Ages
- ✦ Users do not have the correct fonts installed



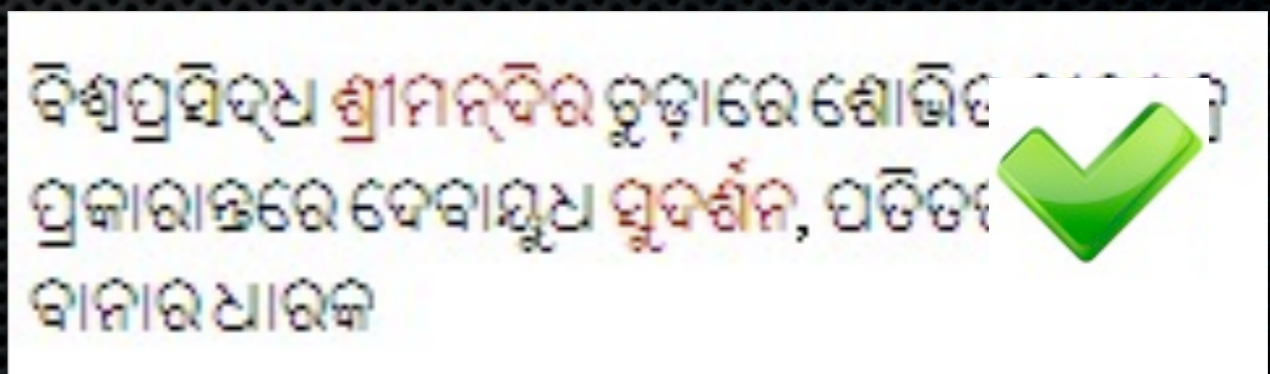
Output

[[mw:Extension:WebFonts]]

- ✦ Embed fonts
 - ✦ The website supplies fonts using
 - ✦ User language
 - ✦ Content language
 - ✦ Language tags



Text in Odiya script without a font installed



Text in Odiya script with WebFonts installed

What else?

Don't get me started!

- ✦ Make better what we have, and for more languages
- ✦ Language support teams for more information and testing
- ✦ Plural and gender in JavaScript
- ✦ Search (stemmers, scripts, ...)
- ✦ Structured translation of MediaWiki documentation
- ✦ Wiki page translation tools (Microsoft WikiBhasha, Google Translator Toolkit, ...)

Questions and answers

E-mail address: smazeland@wikimedia.org

- ✦ Have a look at mediawiki.org. We are looking for volunteers and paid developers!
<http://jobs.wikimedia.org>
- ✦ Help with translations?
<http://translatewiki.net>

