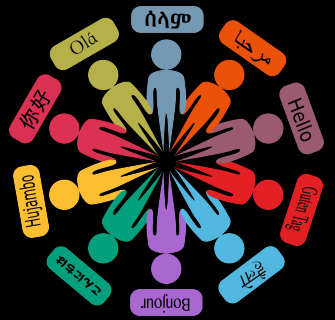
WIKIMEDIA FOUNDATION

Getting Started with Wikimedia's Language Tech Volunteering

Srishti Sethi, Senior Developer Advocate at Wikimedia Foundation



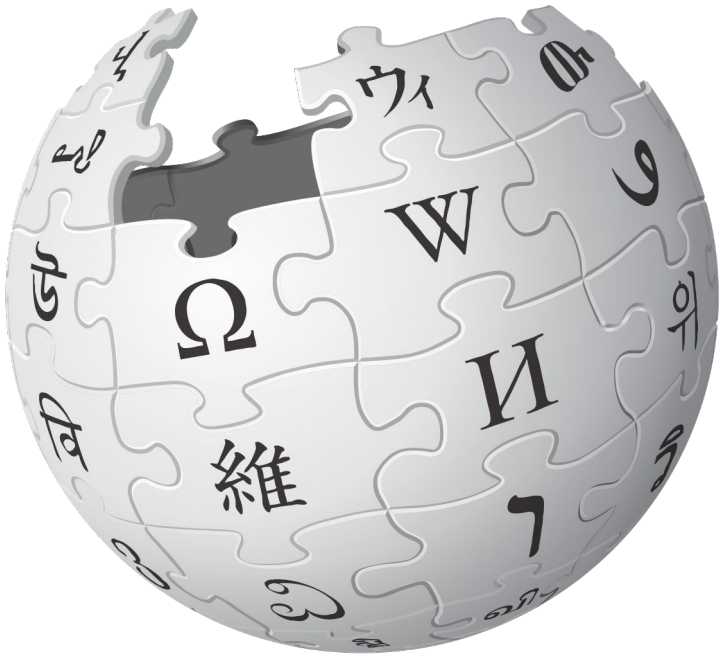
🤦 Which **modern-day platforms** do you see adapting to **diverse languages** and **cultures**, and how do they do it?

Did you know that **Wikipedia** is available in **339 languages**?

Including **27 Indian languages**!

This makes it one of the **most multilingual platforms** in the world.

Explore Wikipedia in your language today!



Did you know **Wikimedia** hosts over **14 free knowledge** projects, besides **Wikipedia**?

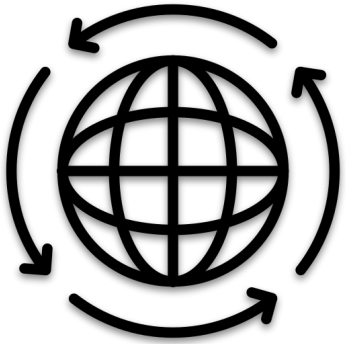
Most of Wikimedia's major **content projects** are multilingual.

🤦 Do you know if there is a **Wikipedia** available in your **local Indian language**?

And if it exists, have you **read** it or **contributed edits** to it?

Language Tech Overview

Wikimedia ensures its projects are accessible in many languages by supporting **internationalization** and **localization** efforts.



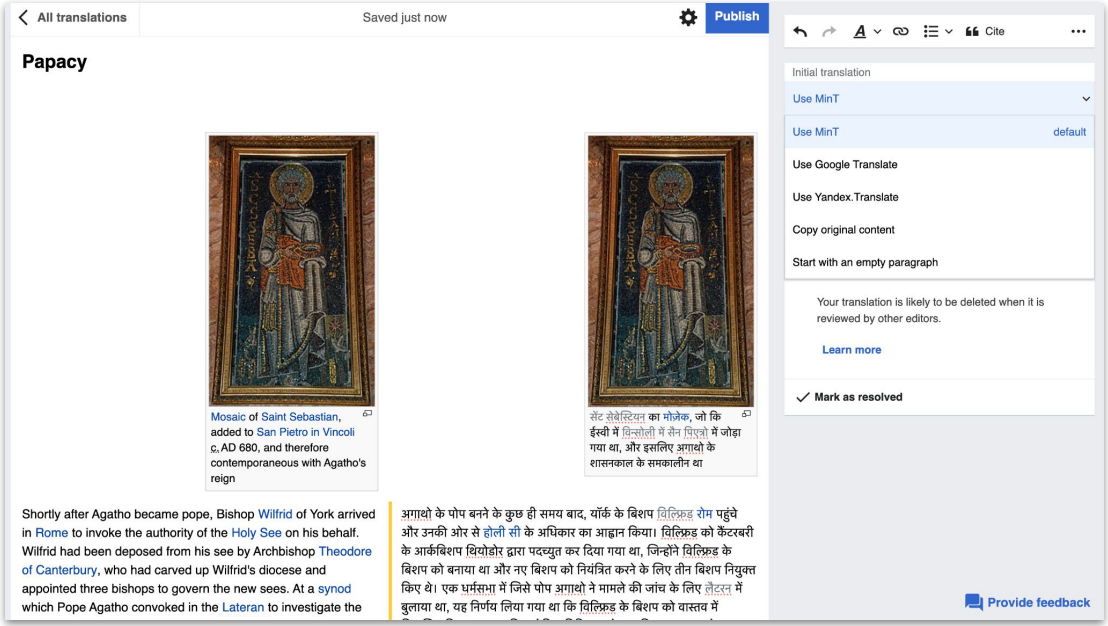
Internationalization is designing Wikimedia platforms, like Wikipedia to be used internationally and support localization efforts.

Localization is adapting Wikimedia software and content to meet the needs of local communities.

Internationalization Example

Wikipedia can be read and **translated** into **multiple languages.**

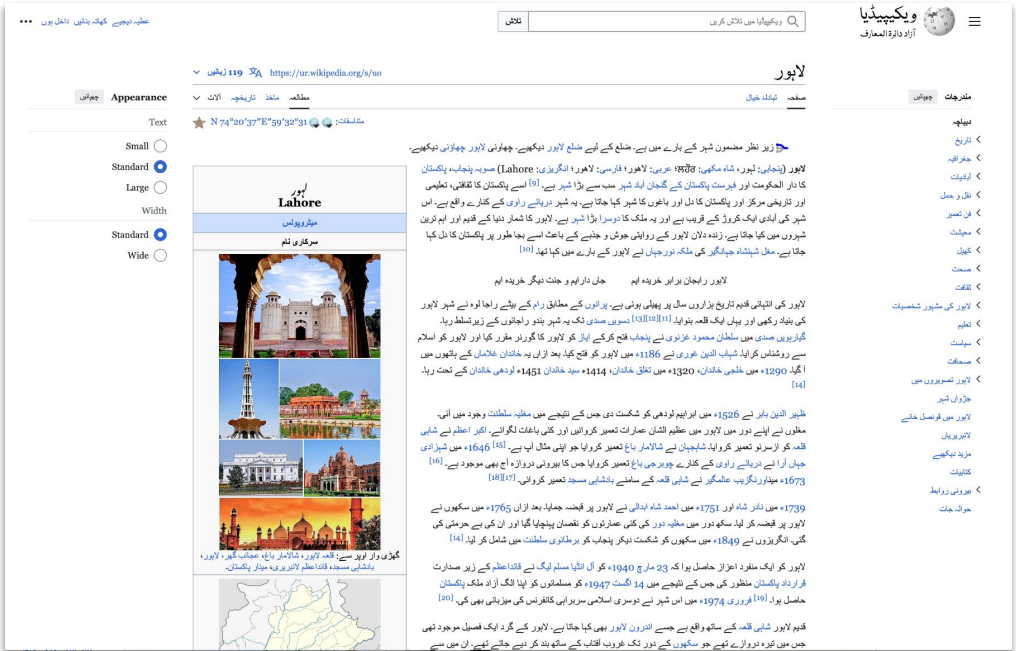
- Using **Content Translation** tool.
- A part of Content Translation, **Section Translation** tool provides experimental support for translating on **mobile devices.**
- These tools use automatic **machine translation services** (e.g., Google Translate, Yandex, etc.)



Internationalization Example

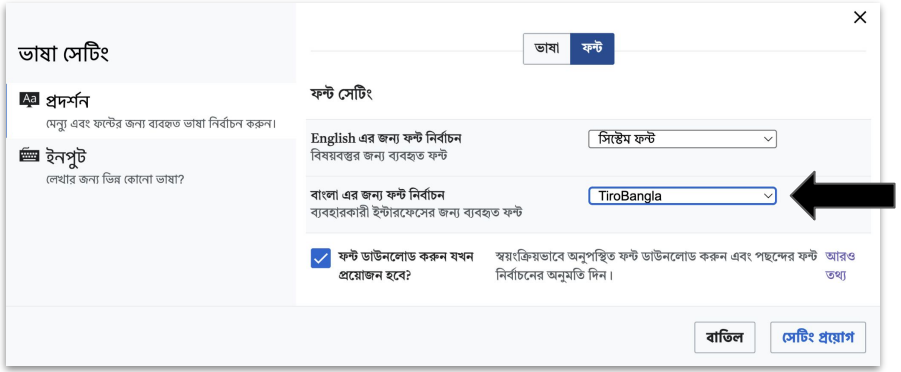
MediaWiki, the software behind Wikipedia is designed to support display of different **writing systems** including **right-to-left (RTL)** languages.

- Wikipedia supports **36 writing scripts**, including Latin, Cyrillic, Arabic, Devanagari.
- A few common scripts write characters **right-to-left**.
- Example in the screenshot shows **Urdu Wikipedia** RTL Interface.

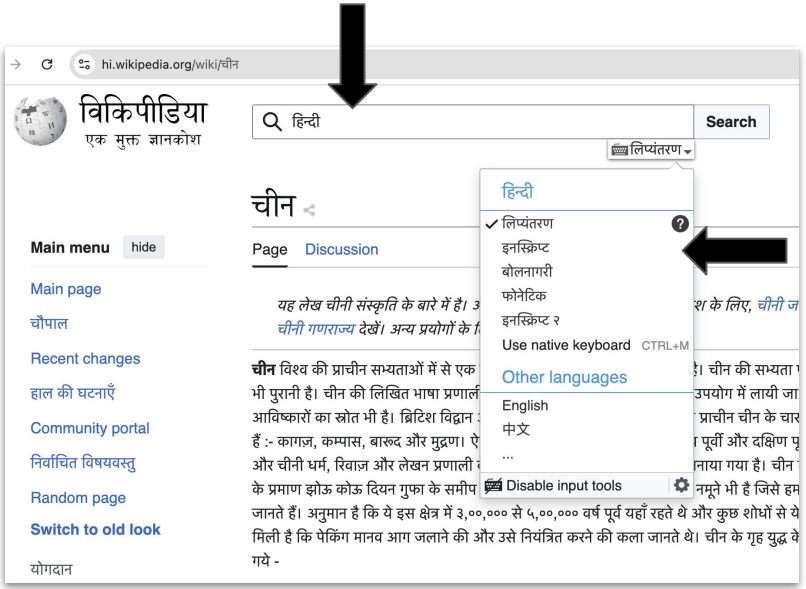


Localization Example

Wikimedia platforms support various **input methods, keyboards, and fonts** (e.g., **Tiro Bangla font** for **Bengali** and **Lipyantaran keyboard** for **Hindi**).



টিরো বাংলা



अ	आ		इ	ई	उ	ऊ		ए	ऐ	ऍ	ओ	औ	ऑ	अं	अः	ऋ
a	A	aa	i	I	ii	u	U	uu	e	ai	E	o	au	O	aM	aH	R

Opportunities

It's a great place for **newcomers** to learn, grow, and connect with others while contributing to the **open knowledge movement**!

- Develop an understanding of **internationalization principles, tools**, and **best practices for creating multilingual software** like **MediaWiki** that can be adapted to various regions and cultures.
- Gain hands-on experience with **open-source contributions**, including making changes, submitting patches, and collaborating with other contributors on platforms like Gerrit and GitHub.
- **Translate content** and making it accessible in various languages.
- **Navigate and manage tasks** across multiple project boards on collaboration platforms

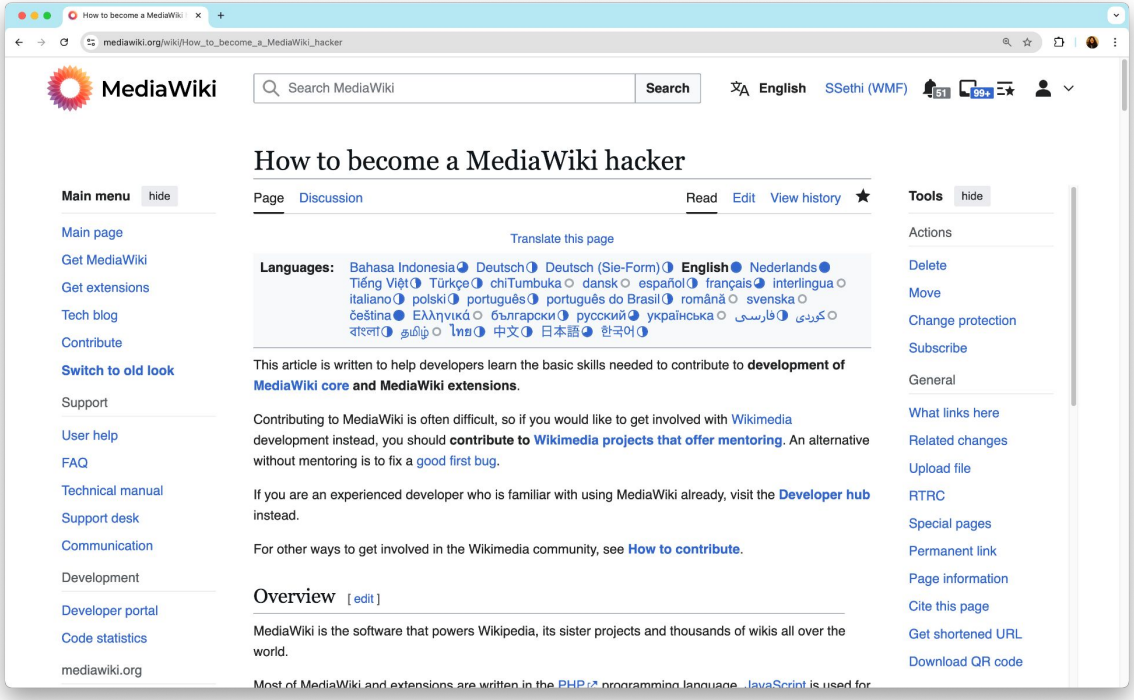
Opportunities

Some example of hands-on **coding** and **documentation** areas:

- Small edits to **improve source messages** or **translations** in:
  - MediaWiki core (the software behind Wikipedia)
  - Various apps and extensions
  - Edit related to spelling, style, grammar, clarity, gender and plural forms.
- Simple **bug fixes** or updates to language tools:
  - Content Translation
  - Adding font and keyboard support for languages in projects (e.g., in Hindi Wikipedia).
- Translating **interface messages** in your local messages
  - Using translatewiki.net

Requirements

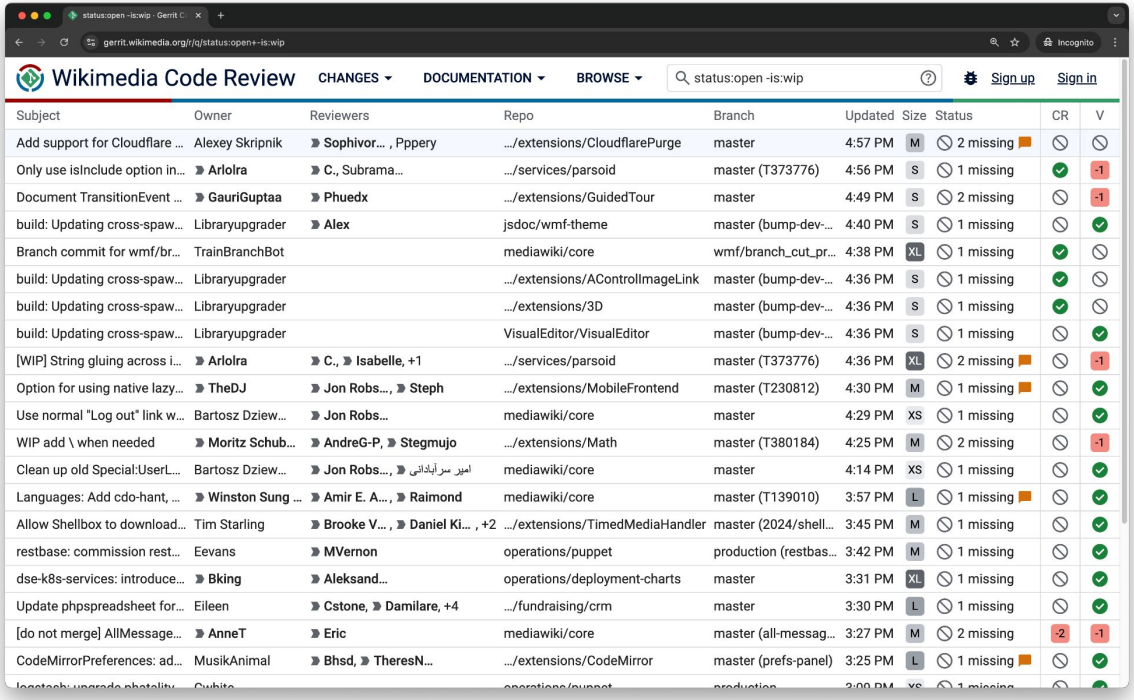
Experience with MediaWiki development



https://www.mediawiki.org/wiki/How_to_become_a_MediaWiki_hacker

Requirements

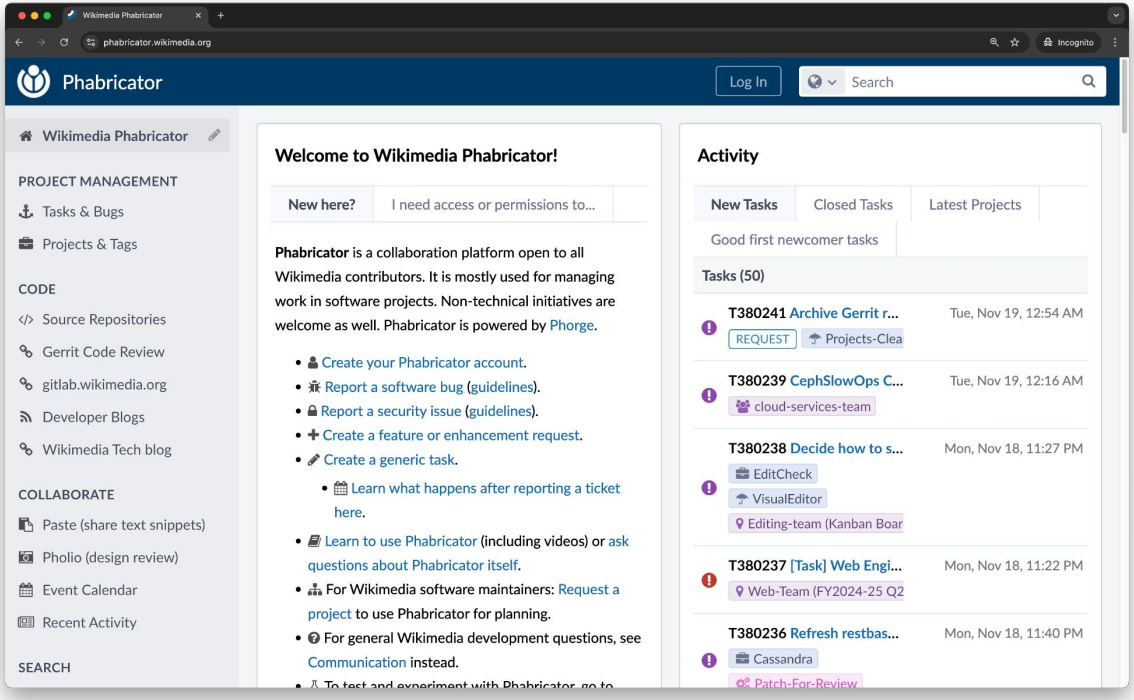
Gerrit Setup



https://www.mediawiki.org/wiki/Gerrit/Tutorial

Requirements

Phabricator Setup



https://www.mediawiki.org/wiki/Phabricator/Help

WIKIMEDIA FOUNDATION

Task Categories and Examples

Beginner-friendly tasks from actively maintained repositories, perfect for newcomers to start and get feedback easily.

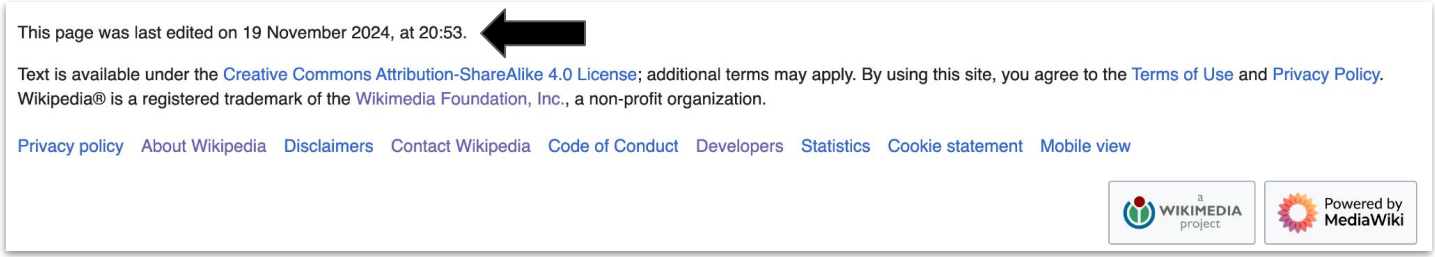
1. Plural Support for Messages

Overview

Languages have unique pluralization rules. **Adding plural support in MediaWiki** ensures messages handle these correctly. For example, English has singular ("1 page") and plural ("2 pages"), while languages like Russian and Arabic have **multiple plural forms**.

Task Example

- **Title:** Plural support required for messages `MediaWiki:Lastmodifiedat` and `MediaWiki:Lastmodifiedatby`
- **Phabricator Link:** T191687



1. Plural Support for Messages

Code Location

- Modify the messages `MediaWiki:Lastmodifiedat` and `MediaWiki:Lastmodifiedatby` in File: `languages/i18n/en.json` in repository: `MediaWiki Core`.

Hints for Implementation

```
"lastmodifiedat": "This page was last edited on $1, at {{PLURAL:$2|hour|hours}}. Note that $2 will require extraction and appropriate processing to determine the number of hours.
```

- $1: Represents the date when the page was last edited (e.g., "November 18, 2024").
- $2: Represents the hour of the day (0–23).
- If $2 = 1, the output will be: *"This page was last edited on November 18, 2024, at 1 hour."*
- If $2 = 0 or $2 >= 2, the output will be: *"This page was last edited on November 18, 2024, at 0 hours."* or *"at 23 hours."*

2. Capitalization for Karakalpak Wikipedia

Overview

Capitalization is the process of converting letters to **uppercase** or **lowercase**. In localization, it's important to handle capitalization correctly across languages. **Each language has unique rules**. In Turkish, i (lowercase) and İ (uppercase) are distinct.

Task Example

- **Title:** Page starting with letter i should not redirect to İ but to I in **Karakalpak** Wikipedia
- **Phabricator Link:** T335175

```php
/**
 * Fixes an issue with ucfirst for transforming 'i' to 'İ'
 *
 * @inheritDoc
 */
public function ucfirst( $str ) {
    if ( substr( $str, 0, 1 ) === 'i' ) {
        return 'İ' . substr( $str, 1 );
    }
    return parent::ucfirst( $str );
}

/**
 * Fixes an issue with lcfirst for transforming 'I' to 'ı'
 *
 * @inheritDoc
 */
public function lcfirst( $str ) {
    if ( substr( $str, 0, 1 ) === 'I' ) {
        return 'ı' . substr( $str, 1 );
    }
    return parent::lcfirst( $str );
}
```

2. Capitalization for Karakalpak Wikipedia

Code Location

- Modify the existing capitalization configuration in file: `includes/languages/LanguageKaa.php` in repository: `MediaWiki Core`.

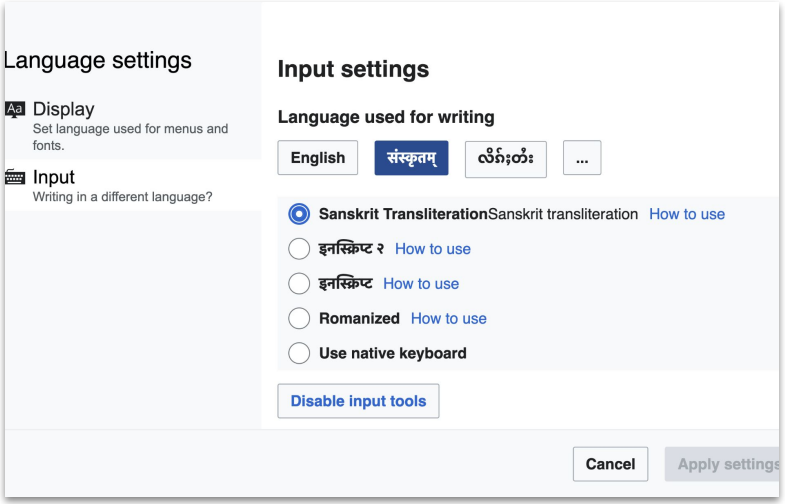
Hints for Implementation

- Read through the comments in Phabricator task!

3. Sanskrit Language Selection User Interface

Overview

jQuery.IME is a library supporting several keyboard input methods that are used across Wikimedia projects. It is used by **Universal Language Selector**, a MediaWiki extension that allows users to select and modify language settings, including fonts, interface language, and input methods.



Task Example

- **Title:** Make Devanagari Transliteration as the default input method for sa.wikipedia (Sanskrit)
  Phabricator Link: T64145

3. Sanskrit Language Selection User Interface

Code Location

- Modify the Javascript file `src/jquery.ime.inputmethods.js` in repository: `wikimedia/jquery.ime` on GitHub.

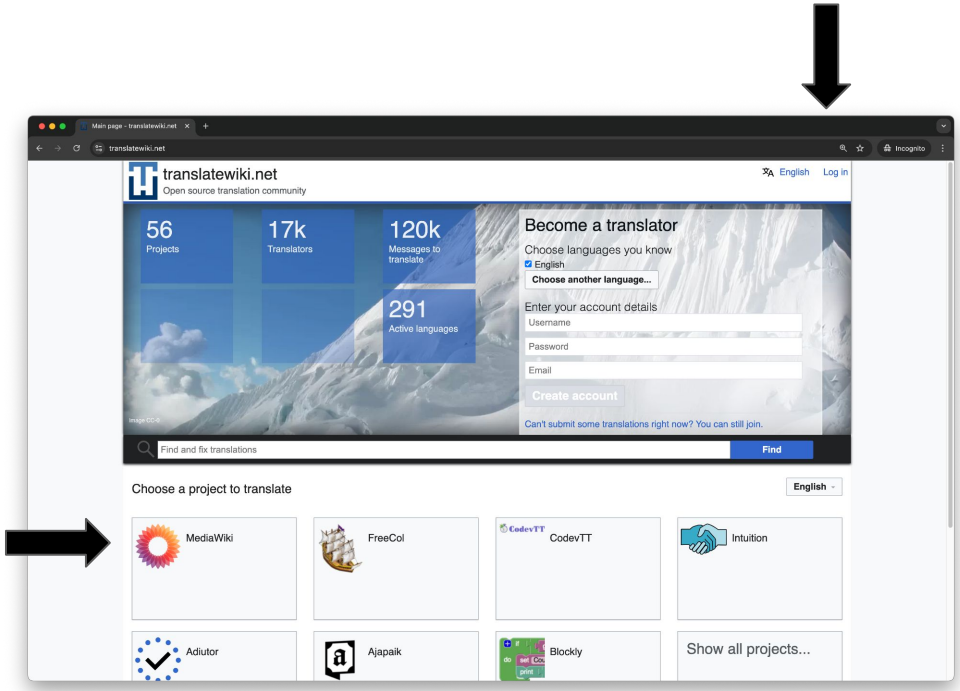
Hints for Implementation

- There is currently no way to set an input method as the default. Instead, input methods can be reordered, as was done for the Santali language. See example: T372146.

4. Translating Software Interface Messages

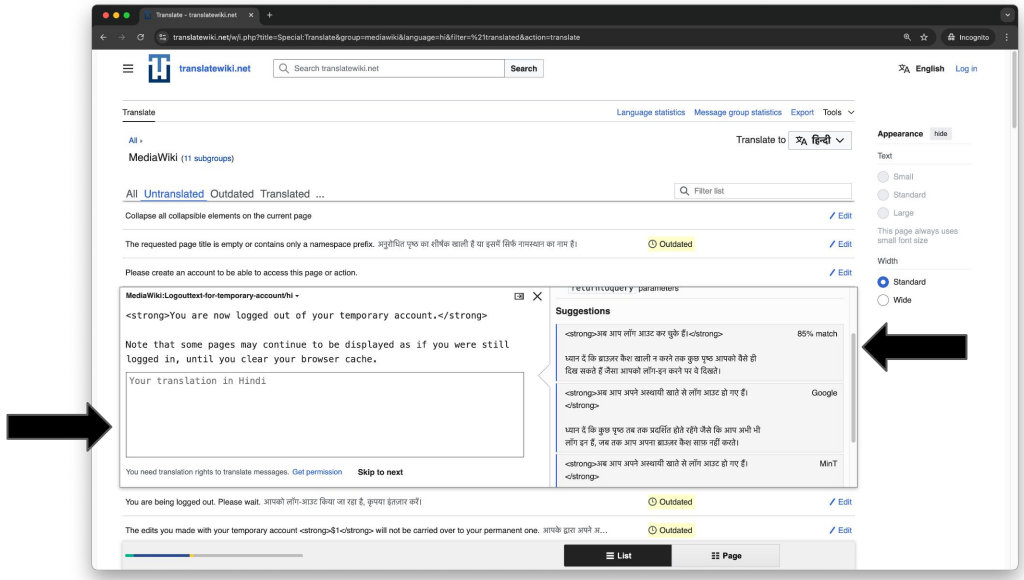
To help translate interface messages for various languages, follow these simple steps:

- Sign up on **translatewiki.net** and select your preferred languages, such as Indian languages.
- Navigate to the **MediaWiki project** to find **most used messages** that need translation (e.g., untranslated messages for Hindi language)



4. Translating Software Interface Messages

- Use the editor to enter translations for the messages.
- Save and submit your translations, then review any feedback provided.
- Make adjustments based on feedback to improve your translations.



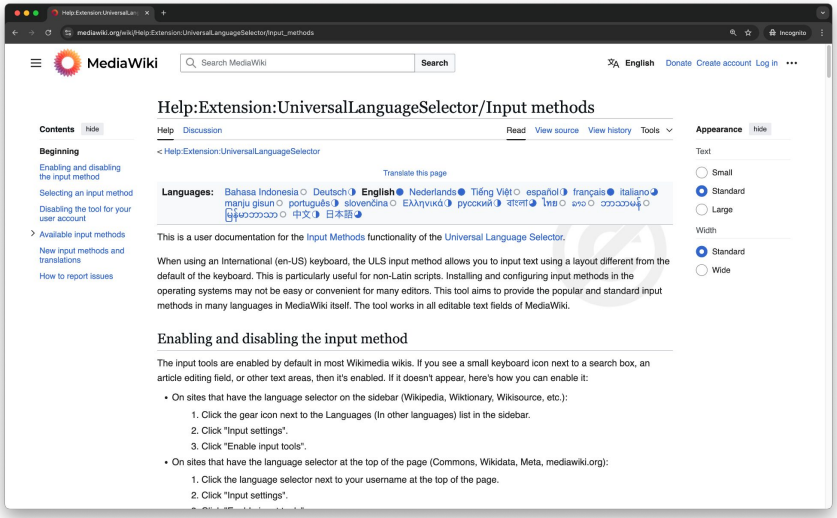
5. Documentation for Keyboard Inputs

Overview

In **jQuery.IME** library, there are certain keyboard input methods that lack corresponding documentation on **MediaWiki.org** for keyboard patterns.

Task Example

- **Title:** Add Missing Documentation for Keyboard Input Methods
- **Phabricator Link:** T381067



https://www.mediawiki.org/wiki/Help:Extension:UniversalLanguageSelector/Input_methods

5. Documentation for Keyboard Inputs

Documentation Location

- As an example, for Fur Tilde, the code file is `fvr-tilde.js` in the jQuery.ime repository, and the corresponding documentation on MediaWiki.org to be added at `Help:Extension/Input methods/fvr-tilde.`

Hints for Implementation

- Use the documentation for existing input methods here: https://www.mediawiki.org/wiki/Help:Extension:UniversalLanguageSelector/Input_methods to improve it.

More Tasks

Message/Translation Fixes

These tasks involve simple message or translation fixes that improve language support:

- Task T306326
- Task T159643
- Task T155993
- Task T335175
- Task T350575
- Task T294727 (Indian Language-Related)
- Task T331193

More Tasks

Bug Fixes/Feature Enhancements

These tasks focus on small bug fixes or feature improvements:

- Task T195229
- Task T379828

Code Organization

These tasks help improve code organization for better readability and maintainability:

- Task T115218
- Task T167762

🙋 Now it's your turn to ask questions!

Contact me:
Telegram: @srishakatux
Tag me on Wikimedia Collaboration Platforms (Preferred)