

A Review on Google Translation project in Tamil Wikipedia

Role of voluntarism, free and
organically evolved, in ensuring
quality of Wikipedia

- Ravishankar Ayyakkannu

(User name: ravidreams)

<http://ravidreams.com>

Coordination

Section 1

History

- Mid 2009 - Project started in Tamil Wikipedia. We couldn't identify who ran this project.
- Feb 2010 - Google informed Indian Wikipedians about this project
- Google - Tamil Wikipedia coordination started in Feb 2010. This was seen as a test case / pilot for coordinating other indic languages translation projects.
- Google running this also in Hindi, Telugu, Kannada, Arabic, Swahili Wikipedias. Bangla Wikipedia has blocked this project.

Coordination

- Channels - email, online chat, conference calls, online shared documents, spreadsheets, Wikipedia talk pages, face to face tutorial sessions for translators.
- Coordinators - on all three sides: Google, Wiki community and translators.
- Decisions - important policy decisions by usual wiki method in talk pages based on inputs from coordinators.
- Consultation - with Indian languages Wikipedians community and members of Wikimedia Foundation

Issues

- Technical
- Translation
- Philosophical

Technical Issues

- Too many unwanted red links - Red links in existing articles were initially manually removed. Now, we have a bot removing red links. Google translation kit's upcoming version may take care of not creating red links.
- Templates not imported - Google adding missing templates manually now.
- Images not imported - Images not in Wikimedia Commons may need to be removed.
- Google translation kit - Translators don't feel user friendly and reducing their productivity.

Operational issues

- Articles to be translated - Initially based on most searched terms from India. But many are not relevant to Tamil. For example, too many american pop stars and Hindi movies which Tamils may not need as a priority.
- Solution 1 - Most searched Tamil terms (coming from across the Globe not just India. Tamil is also an official language in SriLanka, Singapore and has a strong diaspora all over the world) - yet to be done.
- Solution 2 - Google translating needed articles list prepared by Tamil Wikipedians - yet to be done.
- Solution 3 - Tamil Wikipedians validate articles needing to be translated - done now.

Operational issues

- Existing articles getting replaced. Google thought overwriting stubs is OK. But it is not. Now we check the article to be translated does not exist already.
- Many translators operated through a few wiki accounts operated by team leaders. Now we insisted on unique account and total control, responsibility for each translator.
- Many new manual of style policy decisions needed as sudden increase in content and its variety.

Operational Issues

- Feedback mechanism - Translators never checked talk pages, edit changes. Now they have been instructed to check. Separate category for these talk pages. Unresponding users were stripped of their ability to create new pages. A new user group called nocreate was added for this purpose.
- Cleaning up existing articles - Issues with already existing articles were not cleaned up. Now, a target date has been set to complete corrections.
- Categorization - Added Google translated articles and contributors in exclusive categories. Useful to track the project.

Operational Issues

- Response time - Professional translators have strict deadlines and expect faster response from Wiki community. Also, non-wiki channel based support expected. But Wiki community can only work best on-wiki and at-its-own pace. But we try our best to give faster responses.
- Delay in Coordination - Due to delay in identifying Google and starting coordination, the issues have needlessly been duplicated in 100s of articles. Coordination from beginning of project must

Operational issues

- Google's understanding of community participation - Google initially thought the drawbacks in translation will be taken care by the community. But community can't correct the same mistakes again and again.
- Better coordination, translation kit updates and tutoring the translators is the key. We held 4 hours long direct face to face tutorial session for translators. Around 15 translators from all translation companies participated. Many more rounds of telephone conferences were held with Google.
- Inhouse translators working as a team is better. Freelancers may not be available for correction and may be exploited by vendors.

Translation issues

- Word for word and sentence by sentence translation giving a mechanical feel. Even if 95% + manual translation is involved.
- Errors, bias, non-updated info in English Wikipedia articles carried through.
- Lack of comprehension
- Too much use of transliterated English words as such against community's manual of style - lack of some Tamil words in new fields, inefficiency / carelessness of the translators.

Translation issues

- Amateur and literal translation by some translators
- Untranslated templates
- Ungrammatical sentences
- Wrong translations
- Poor copy editing, spelling mistakes, improper punctuations, no proof reading. Upload and disappear and no one to take care of article. Now we insisted on fixing these issues.
- Unconventional transliteration / translation

Section 2

Philosophical Issues

Philosophical Issues

- Paid editing - Against Wikipedia principles or spirit? Pay per se not wrong. But cause for other issues. Conflict of interest, motivation levels based on pay etc.,
- Google's motivation for the project? Community feels Google uses Wikipedia to test and develop their translation kit. Google says it wants to contribute to Indian languages. Can Wikipedia be a testing ground for such projects?

Philosophical Issues

- Regular Wikipedians Vs Paid translators: user - > user interaction changed to user-Wiki coordinator - Google coordinator - translators' coordinator - translators. Change in community dynamics? can the users act as a group? What if religion / ethic / any organization based users demand such group privileges citing this precedence?

Philosophical Issues

Do Paid translators contribute out of free will? choice of articles, work load, translation style, tool for translation?
Can this affect translation quality? How important is voluntarism and a free community to ensure the growth, quality and integrity of a project?

Content Vs Community.

Philosophical Issues

- Use of translation kit not wrong per se. But reason for many other issues. Should we translate by hand or using a Kit is OK?
- Local Wikipedias need not / should not be a copy of English Wikipedia. English Wiki articles have a western / Europe-centric view. Local view is needed.
- Community not motivated to donate its valuable time to correct mistakes left by paid translators.
- Lack of moral ownership and responsibility to upgrade quality of a translated article

Philosophical issues

- Style of English and style of Indian languages distinct. Machine aided translation resulting in a dry and artificial style for Indian languages. Given the influence of Wikipedia and Internet in general, this can wrongly influence the language style in long run (compare print media influence)
- Google and Wikipedia's influence on one another? Both being powerful entities on web.

Review

Section 3

Pros

- May be a good model to increase content in small Wikipedias quickly
- Diverse topics added. Not leaning towards few interest areas of few contributors.
- Development of Machine aided translation for many more languages
- Job opportunity for translators

Cons

- Lack of quality translation may degrade the overall positive image of work done by other contributors
- Change in community dynamics. Content vs Community?
- Quantity Vs Quality?
- Should Wikipedia become a testing field? Microsoft is also reportedly involved in a similar project in 30 Wikipedias.
- Lack of active community to coordinate this in some Wikis. Such Wikis may be irreversibly influenced by this.

Suggestions

- Local Wikipedia community should be involved from the beginning including pre-planning. More control with the Wikipedians the better the quality.
- Local Wikipedia community's decision should be final (it already is)
- Involve Universities / regular Wikipedians through grants and scholarships rather than professional translators. Try different payment models.
- Scale up only after all issues are resolved.