

Languages of Knowledge Infrastructures

Learnings from Research on Indian Language Wikimedia Projects

Puthiya Purayil Sneha
Centre for Internet and Society
Bengaluru, India
sneha@cis-india.org

ABSTRACT

Indian Language Wikimedia projects have been growing at a steady pace in the last few years, with several new projects and emerging areas of work. At present, Wikipedia is available in 23 Indian languages [1], the latest being Santali Wikipedia which went live in 2018.[2] In addition to this, there are also active Indian language communities working on other projects such as Wikisource, Wikimedia Commons and Wikidata for example. While each project has its own specific goals and objectives in terms of growing content, and creating a diversity of use and access to the knowledge thus created, there are also challenges and knowledge gaps, some of which have remained persistent, evolving over a period of time. These range from technical issues related to access, resources and training, to more specific concerns related to certain languages, such as with those that may not have a fixed script. As a result, language communities have also been working to identify and address some of these concerns, both infrastructural, but also conceptual, in terms of their development in the context of the larger open source and access to knowledge movements and work on non-dominant languages in the global South.

The role of research in understanding and elucidating some of these challenges and opportunities, and supporting ongoing work on building Indian language Wikipedia projects is something to yet to be explored more widely in the Indian context. While there is definitely now a substantial body of research and writing,

particularly related to technical aspects of Indian language Wikipedias, outreach and awareness of such work among communities still requires focused efforts. Further, several important efforts are being undertaken in collaboration with colleagues in the larger open source movement. Similarly, ongoing research and work in areas such as computational sciences, digital humanities and cultural analytics, or arts and cultural heritage, design thinking and creative practice for example would also find many resonances with some of the challenges being discussed as part of the Indian language projects.

In order to understand and explore further the role of research in identifying and addressing some of these concerns, a small pilot initiative was undertaken by the Access to Knowledge Programme, located at the Centre for Internet and Society, India. [3] This comprised a set of six short-term research studies undertaken by team members, over a course of a year and a half, on key topics of interest within their own Indian language communities. The research studies are closely aligned with continuing areas of work, like the gender gap in Indian Wikimedia communities [4], creating multilingual and open educational platforms and resources, and efforts and challenges with content creation, access and outreach in specific language communities.[5, 6] Reports on some these studies have been published, while the rest are in progress and will be completed soon. This year, following up on learnings from some of this work, the team also undertook a needs assessment exercise to understand the awareness about research within

Indian language Wikimedia communities and identify existing projects if any, and b) gather community inputs on knowledge gaps and priority areas of focus, and the role of research in addressing the same.

This presentation will seek to present some early observations from both the above efforts. With the inputs and collaborations from the researchers/team members themselves, it will delve into some of the motivations and challenges of designing and operationalizing the research projects. It will also present some early takeaways from the research assessment exercise as well. The objective of this presentation would therefore be to present some early learnings from the research undertaken so far, to a diverse audience, and gather feedback that may help inform next steps, and further align these efforts with the work of the larger movement.

KEYWORDS

Wikimedia research, Indian languages, movement resources, multilingual content

ACKNOWLEDGMENTS

The author would like to thank the CIS-A2K team members/researchers for their efforts in undertaking the research, and Wikimedia contributors for their inputs and observations on the projects.

REFERENCES

- [1] Wikimedia Meta-Wiki. 2020. List of Indian Language Wiki Projects. Retrieved from https://meta.wikimedia.org/wiki/List_of_Indian_language_wiki_projects
- [2] Wikipedia.2021. Santali Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Santali_Wikipedia
- [3] Wikimedia Meta-Wiki. 2021. CIS-A2K. Retrieved from <https://meta.wikimedia.org/wiki/CIS-A2K>
- [4] Wikimedia Meta-Wiki. Research: Bridging the Gender gap: A Report on Indian Language Wikimedia Communities Retrieved from https://meta.wikimedia.org/wiki/Research:Bridging_the_Gender_gap:_A_report_on_Indian_Language_Wikimedia_Communities
- [5] Wikimedia Meta-Wiki. 2020. Research: Understanding the Data Gaps on Wikidata Concerning Heritage Structures of West Bengal. Retrieved from https://meta.wikimedia.org/wiki/Research:Understanding_the_data_gaps_on_Wikidata_concerning_heritage_structures_of_West_Bengal
- [6] Wikimedia Meta-Wiki. 2020. Research: Mapping GLAM in Maharashtra. Retrieved from https://meta.wikimedia.org/wiki/Research:Mapping_GLAM_in_Maharashtra,_India