Bharatavani Project - Reviving Linguistic Diversity and Cultural Heritage in India: A Case Study

Narayan Choudhary, LR Premkumar, Chandan Singh, Shubhanan Mondal, Shivangi Priya, Beluru Sudarshan, P. Perumal Samy, Shailendra Mohan

Central Institute of Indian Languages Mysuru

Karnataka, India

n.choudhary@gov.in, lrprem90@gmail.com, chandansingh.ciil@gmail.com, subhanan.ciil@gmail.com, priyashivangi6@gmail.com, beluru.sudarshana@gmail.com, director-ciil@gov.in

ABSTRACT

The Bharatavani project, launched in 2016 initiated by the Government of India, addresses the crucial need to preserve and promote indigenous languages and cultures. The paper presents an overview of the project, which focuses on recording socio-cultural and linguistic information about 121 Indian languages and making it accessible to a broader audience. The project leverages technological advancements to document significantly smaller and lesser-known languages and mother tongues in India, to raise awareness and maintain and promote the country's rich linguistic diversity. The Bharatavani project aims to bridge the digital divide and ensure equal access to knowledge and information by emphasising the importance of incorporating these languages into the digital sphere. Through the creation of e-content, the project offers multimedia resources, including text, audio, video, and images, through the online portal www.bharatavani.in and the Bharatavani Android App. This research highlights the significance of content generation, software development, and web portal creation for selected languages in the first phase, with subsequent plans for translation, online teaching-learning, and language teacher training in the second phase. By embracing the potential of technology, the Bharatavani project aspires to create a Knowledge Society in the digital era, enabling individuals across India to explore and celebrate their linguistic heritage.

CCS CONCEPTS

• Applied computing → Computers in other domains → Digital libraries and archives • General and reference → Document types → Surveys and overviews • Social and professional topics → Computing / technology policy → Government technology policy • Human-centered computing → Accessibility→ Accessibility technologies.

KEYWORDS

Digital Archives, Bharatavani, Ministry of Education, E-content, Indigenous Languages

1 INTRODUCTION

As a nation, India is known for its linguistic diversity and pluralism. Languages of India primarily belong to five linguistic families: Indo-European, Dravidian, Austro-Asiatic, Tibeto-Burman, and Samito-Hamitic [1]. According to the Census of 2011, there are 121 languages and 270 mother tongues with speakers' strength of 10 000 and above at the national level [3]. Twenty-two of these languages are included in the VIII Schedule of the Constitution of India, and there are 100 non-scheduled languages spoken by more than 10,000 speakers each. Moreover, there are several languages/mother tongues that are spoken by fewer than 10,000 persons each. The scheduled and non-scheduled languages and some other languages/mother tongues have writing systems.

In contrast, hundreds of languages/mother tongues remain oral. However, in the face of globalisation and the dominance of a few major regional languages, many regional and indigenous languages have been marginalised, risking the cultural heritage they embody. In response to this challenge, the Government of India launched the Bharatavani Project to preserve, promote, and propagate linguistic diversity by leveraging digital technologies. This research article provides an overview of the Bharatavani Project, its objectives, implementation strategies, and its impact on language documentation, promotion, and revitalisation. Additionally, it explores the project's significance in the broader context of linguistic heritage preservation and the challenges that need to be addressed for its sustained success.

Electronic content came rather late into Indian languages. In order to preserve knowledge about indigenous languages and

cultures, it is crucial to record socio-cultural and linguistic information about these languages and make it accessible to a wider audience. It is essential to utilise technological advancements for documenting vernaculars, mainly smaller and lesser-known languages/mother tongues in India. This project can effectively raise awareness about these languages/mother tongues, thus contributing to maintaining India's rich linguistic diversity.

In India, the exclusion of mother tongues from formal education is closely tied to the perception of inferiority and reduced vitality [2] attributed to minor, minority, and tribal languages compared to dominant majority languages like English. Furthermore, the preference for English-medium education has marginalised other major regional and constitutional languages, significantly weakening them across all aspects of Indian society. Globalising information through the internet primarily occurs in English and a few scheduled languages, leaving hundreds of languages/mother tongues absent in the digital content realm. Moreover, content development in these languages/mother tongues, including their cultural components (textual, auditory, or visual), progresses slowly. As a result, a significant portion of the population needs help to utilise the information available in cyberspace to enhance their knowledge and foster social and economic growth. The need for technology to incorporate these languages/mother tongues into digital networks or provide translation services further compounds the issue.

To overcome this gap, the Ministry of Education, Government of India, launched the Bharatavani Project to address these challenges and safeguard India's linguistic heritage. The Bharatavani Project is implemented by the Central Institute of Indian Languages, Mysore, in the form of a web portal and mobile application where registered users can access books dealing with encyclopaedia, language learning materials, dictionaries & glossaries, textbooks, grammar and more in 121 Indian languages for free. Till March 2021, Bharatavani had accomplished the task of hosting more than 5500 resources in 92 Indian languages. Besides books, Bharatavani hosts multimedia content describing various literary and cultural aspects of Indic linguistic communities. Already the biggest single knowledge portal in the world, Bharatavani mobile application is now operational with 200+ digital dictionaries in multiple languages and subject combinations, the first of its kind in the world. Bharatavani has become the world's largest single-point hub of important indigenous content. The objective is to make indigenous knowledge resources in their respective languages available through a robust digital platform.

One of the project's key objectives is to develop e-content in different languages/mother tongues and showcase India's linguistic diversity in cyberspace. It is an integral part of the broader mission of creating a Knowledge Society in the Digital India. The project ran in two phases. In the first year, 18 Scheduled languages (Assamese, Bengali, Bodo, Dogri, Gujarati, Kannada, Kashmiri, Konkani, Malayalam, Manipuri, Maithili, Marathi, Nepali, Oriya, Punjabi, Santali, Tamil, Telugu) and 32 non-scheduled languages were be covered. In the second year, 68 non-scheduled (excluding Arabic, Afghani, English and Persian) Indian languages/mother tongues were covered.

The project revolves around an online portal called www.bharatavani.in and a corresponding Bharatavani Android App. These platforms deliver knowledge in multimedia formats, including text, audio, video, and images, for all languages in India. The project's initial focus is on content aggregation and developing a corner on the web portal for selected languages. The project aims to encompass translation, online teaching-learning, and online language teacher training in the second phase. Overall, the Bharatavani Project aims to leverage digital platforms to facilitate access to knowledge and information in and about all languages in India. By doing so, it seeks to embrace the country's linguistic diversity and contribute to advancing a Knowledge Society in the digital age.

2 OBJECTIVES OF THE BHARATAVANI PROJECT

The primary objectives of the Bharatavani Project include:

1. Creating a digital repository of linguistic resources: The project focuses on collecting, digitising, and archiving language-related materials, such as dictionaries, grammar, texts, and audio-visual content, to build a comprehensive repository.
2. Development of linguistic tools and technologies: The project emphasises the development of language learning tools, machine translation systems, text-to-speech synthesisers, and other language technologies to facilitate linguistic research and communication.
3. Dissemination of linguistic resources: Bharatavani aims to make linguistic resources accessible to a wide range of users through online platforms, mobile applications, and offline modes to promote language learning and research.

3 IMPLEMENTATION STRATEGIES

The Bharatavani Project employs a multi-faceted approach to achieve its objectives:

1. Language documentation and digitisation: Linguistic scholars and experts collaborate to document and digitise linguistic resources in various Indian languages, ensuring their preservation and accessibility.
2. Technology development: The project invests in research and development of language technologies to facilitate language learning, content generation, and translation services.
3. Language promotion and awareness: Bharatavani conducts workshops, seminars, and awareness campaigns to promote linguistic diversity and foster pride in regional languages among communities.

4   IMPACT AND ACHIEVEMENTS

Since its inception, the Bharatavani Project has made significant strides in language preservation and revitalisation. Some of the notable impacts and achievements include.

1. Preservation of endangered languages: By digitising and archiving endangered languages, the project has helped prevent the loss of linguistic knowledge and cultural heritage associated with these languages.
2. Language learning and dissemination: The project has provided digital platforms and mobile applications that enable users to learn Indian languages, facilitating cross-cultural understanding and communication.
3. Academic and research support: The availability of linguistic resources and technologies has enhanced linguistic research, enabling scholars to delve into the grammatical structures, dialects, and socio-cultural aspects of different languages.

4   CONCLUSIONS

Despite its successes, the Bharatavani Project faces several challenges that need to be addressed for its continued effectiveness. Indian language scripts are complex and challenging to connect with digitally generating software programmes. In contrast to English, where consonants and vowels have different representations in word creation, the alphabet, which consists of both, is written as a single unit for Indian languages. Indian scripts, sometimes known as abugida scripts due to this distinguishing trait, are challenging to create and incorporate into programmes. Also, making the Indic scripts compatible across devices is a major task for developing language technologies-based applications. In addition, for error-free performances, the unique glyphs used by Indian languages that cause issues with various devices must be fixed.

The Bharatavani Project is a significant initiative to protect India's linguistic diversity and preserve the cultural heritage embedded within these languages. The project has made commendable progress in language preservation, dissemination, and research by leveraging digital technologies and fostering community involvement. However, to ensure its sustained success, continued support, collaboration, and innovative strategies are necessary to overcome the challenges and propel the project towards a more inclusive and linguistically vibrant future for India.

REFERENCES

[1]   Abbi, Anvita (2018). A sixth language family of India: Great Andamanese, its historical status and salient present-day features. In The Dynamics of Language: Plenary and focus lectures from the 20th International Congress of Linguists (p. 134). Juta and Company (Pty) Ltd

[2]   Giles, H. (1977). Towards a theory of language in ethnic group relations. In H. Giles (Eds.). Language, ethnicity and intergroup relations, (pp.307-348). London: Academic Press

[3]   Office of the Registrar General of India. 2018. The Census of India 2011): Paper 1 of 2018- Language. Office of the Registrar General of India, New Delhi.

[4]   Perumal P. Samy, Narayan Choudhary, L.R. Premkumar, Chandan Singh, Subhanan Mandal. 2021. Indian Languages in Digital Space: Sharing Knowledge through Web-portal and Mobile App. Central Institute of Indian languages, Mysore.