

TAKING A SUSTAINABLE AND CLASSIFICATION-BASED APPROACH TOWARDS PROTECTION OF KNOWLEDGE RESOURCES IN INDIAN LANGUAGES VIA BHARATAVANI MULTISITE PORTAL

Manikanta V

Central Institute of Indian Languages, Mysore
cii.manikanta@gmail.com

Abstract: The already available digital knowledge platforms are dedicated towards hosting of books without any classification and dedicated sections based on the content genre. Bharatavani stands out from the rest by not only providing important and detailed classification of each of the knowledge contents in and about Indian languages but also helps the users to modify their search according to their needs. A web portal dedicated to hosting of resources in Indian languages is the first of its kind and thus Bharatavani has already gained wide-scale appreciation for such an approach towards knowledge management. The management is inclusive of a single language-based content as well as multilingual content wherein the same is broadcasted to concerned language domains for usage and viewing by the registered users.

Keywords: Bharatavani Project, Dictionaries, Jnanakosha, Bhashakosha

1. Introduction

The Bharatavani Website provides the users the opportunity to view the ethnic, traditional, cultural, knowledge-based diversity associated with language communities of the entire nation through a single integrated knowledge portal. This facilitates the academicians, researchers and scholars to refer to a one-point solution in form of Bharatavani multisite portal for all the knowledge contents in and about Indian languages.

Indian languages are essentially classified into 5 family categories.

a) Indo-european

The family belongs to regions of Southwest and South Asia with Europe spoken languages.

- i. INDO-ARYAN From 1800-1500 BCE, ethnolinguistic South asia group people are diverse speak to the Indo-aryan languages. There are 21 INDO-ARYAN languages [8] 1. Assamese (S), 2. Bengali(S), 3. Bhili/Bhilodi, 4. Bishnupuriya, 5. Dogri(S) 6. Gujarati(S), 7. Halabi, 8. Hindi(S), 9. Kashmiri(S), 10. Khandeshi, 11. Konkani(S), 12. Lahnda. 13. Maithili(S), 14. Marathi(S), 15. Nepali(S), 16. Odia(S), 17. Punjabi(S), 18. Sanskrit(S), 19. Shina, 20. Sindhi(S), 21. Urdu(S).
- ii. IRANIAN Community of ethno-religious Zoroastrians migrated in the 19th and 20th centuries from Iran to British [8] 1. Afghani/Kabuli/Pashto

- iii. GERMANIC Northern Germany and southern Scandinavia of Germanic tribes. 1. English.

b) Dravidian

The Dravidian languages are divided into literary language and nonliterary language. The literary languages are Kannada, Tamil, Malayalam and Telugu. Nonliterary languages are completely Proto-South Dravidian languages, South-Central Dravidian languages, Central Dravidian Languages and North Dravidian languages. Proto-South Dravidian is divided into ProtoTamil-Kannada and Tulu language. Tulu is the Karnataka of Dakshin kannada district and the Kerala of Cannanore district. The Proto-Tamil-Kannada is divided into ProtoTamil-Toda and Pre-Kannada type of languages. PreKannada type of languages belong to Badaga and Kannada language. Proto-Tamil-Toda is divided into Proto-Tamil-Kodagu and Proto-kota-Toda. The ProtoKota-Toda languages belong to Kota and Toda. Proto-Tamil-Kodagu is divided into Proto-TamilMalayalm and Kodagu language. Kodagu language belongs to Coorg district. Proto-Tamil-Malayalam is divided into Pre-Tamil and Malayalam language. PreTamil languages belong to Tamil and Irula languages. South-Central Dravian languages are divided into ProtoTelugu and Proto-Gondi-Kui. The Proto-Telugu languages belong to Savara and Telugu languages. Proto-Gondi-Kui languages are divided into Proto-Gondi, Proto-Kui-Kuvi languages, Pengo and Manda languages. Proto-Gondi languages belong to Gondi and Konda languages. Proto-Kui-Kuvi languages belong to Kui and Kuvi languages. The Proto-Central Dravidians are divided into Proto-Kolami-Naki and Proto-Gadaba-Parji. Proto-Kol Ami-Naki languages belong to Kolami andNaki. Proto-Gadaba-Parji languages belong to Gadaba and Parji languages. North Dravidian languages are divided into Proto-Kurukh-Malto and Proto-Brahui. Proto-Kurukh-Malto languages belong to Kurukh and Malto Language. Proto-Brahui belongs to the Brahui language.

There are 17 DRAVIDIAN languages [8] 1. Coorgi/Kodagu, 2.Gondi, 3. Jatapu, 4. Kannada(S), 5. Khond/Kondh,, 6. Kisan, 7. Kolami, 8. Konda, 9.Koya, 10. Kui, 11. Kurukh/Oraon, 12. Malayalam(S), 13. Malto, 14. Parji, 15. Tamil(S), 16. Telugu(S), 17.Tulu.

c) Austro-asiatic

Austro-Asiatic Languages are divided into Khasi-Khmuic, (Nuclear) Non-Khmer and Munda. Munda languages belong to Remo, Savara, Kharia-Juang, Korku.

There are 14 AUSTRO-ASIATIC languages [8] 1. Bhumij, 2. Gadaba, 3. Ho, 4. Juang, 5. Kharia, 6. Khasi, 7. Koda/Kora, 8. Korku, 9. Korwa, 10. Munda, 11.Mundari, 12. Nicobarese, 13. Santali(S) 14.Savara.

d) Tibeto-Burmese

The Indian languages consists Southeast Asia and other some parts of East Asia and South Asia which include three types of languages Tibetan, Chinese and Burmese. Tibeto-burmese is a Sino-Tibetan language family. There are 66 TIBETO-BURMESE languages [8] 1. Adi, 2. Anal, 3. Angami, 4. Ao, 5. Balti, 6. Bhotia, 7. Bodo (S), 8. Chakesang, 9.

Chakru/Chokri, 10. Chang, 11. Deori, 12. Dimasa, 13. Gangte, 14. Garo, 15. Halam, 16. Hmar, 17. Kabui, 18. Karbi/Mikir, 19. Khezha, 20. Khiemnungan, 21. Kinnauri, 22. Koch, 23. Kom, 24. Konyak, 25. Kuki, 26. Ladakhi, 27. Lahauli, 28. Lakher, 29. Lalung, 30. Lepcha, 31. Liangmei, 32. Limbu, 33. Lotha, 34. Lushai/Mizo, 35. Manipuri(S), 36. Mao, 37. Maram, 38. Maring, 39. Miri/Mishing, 40. Mishmi, 41. Mogh, 42. Monpa, 43. Nissi/Dafla, 44. Nocte, 45. Paite, 46. Pawi, 47. Phom, 48. Pochury, 49. Rabha, 50. Rai, 51. Rengma, 52. Sangtam, 53. Sema, 54. Sherpa, 55. Tamang, 56. Tangkhul, 57. Tangsa, 58. Thado, 59. Tibetan, 60. Tripuri, 61. Vaiphei, 62. Wancho, 63. Yimchungre, 64. Zeliang, 65. Zemi, 66. Zou.

e) Semito-Hamitic

The language family consists of around 300 languages which belong to North Africa, West Asia, Sahel, etc. 1. Arabic/Arbi.

Bharatavani is broadly divided into 121 language-domains which consist of the scheduled and non-scheduled languages of India. These language domains host a wide variety of knowledge contents representing various fields like Language Learning, Encyclopedia, Textbooks, Dictionaries and Glossaries. Bharatavani is the only unique portal to host such diverse resources in a single portal. These contents are received from various universities/Institutions/individual authors/publishers who partner with Bharatavani in this knowledge sharing initiative of the Government of India under the Digital India program. Besides providing meaning of a word that may be available across multiple dictionaries, Bharatavani stands out by providing the users the opportunity to transliterate generated meaning in the Dictionary domain in nine Indic scripts.

2. Application development methods

Bharatavani applications categorized contents as

- a) Jnana Kosha**
- b) Bhasha Kosha**
- c) Shabda Kosha**
- d) Paathyapustaka Kosha**
- e) Suchanaa Praudyogikii Kosha**
- f) Bahumaadhyama Kosha**

The Bhashakosha section is created for language learning books whereas Jnanakosha section is meant for encyclopedic knowledge contents in Indian languages. On the other hand, Paathyapustakakosha domain is created for hosting textbooks from various textbook production agencies across India. Shabdakosha section is dedicated towards hosting Dictionaries and Glossaries, Bahumaadhyamakosha for multimedia content and finally, Suchana Praudyogikki Kosha is meant for IT tools. In addition to the hosting of the resources, these sections provide descriptions about the content category, book type,

authors, content partners, publishers, media type, producer, etc. These types of additional features are not available in any part of the journal or book related websites.

For all these functionalities to be implemented, development of a multisite website is required to invoke the structural representation.

i. Install WordPress:

Many open source softwares are available for website creation. The WordPress software is an open source tool for multisite based portal development than others.

ii. Create Multisite:

a) Settings of wp-config.php

Change to site name, cache location file (define('WPCACHEHOME',
'/public_html/wp-content/plugins/wp-super-cache/');),
database name, username, password, etc.

b) Create .htaccess file

```
<IfModule mod_rewrite.c>  
RewriteEngine On  
RewriteBase / Options +FollowSymLinks  
RewriteCond %{HTTP_HOST} ^www\.bharatavani.in\.tld$ [NC]  
RewriteRule ^(.*)$ http://bharatavani.in.tld/$1 [R=301,L]  
RewriteRule ^index\.php$ - [L]  
RewriteRule ^wp-admin$ wp-admin/ [R=301,L]  
RewriteCond %{REQUEST_FILENAME} -f  
RewriteCond %{REQUEST_FILENAME} -d  
RewriteRule ^ - [L]  
RewriteRule ^(wp-(content|admin|includes).*) $1 [L]  
RewriteRule ^(\.\.|\.)$ $1 [L]  
RewriteRule \.index.php [L]
```

c) Enable a multisite network in WordPress for the location of wp-config.php file to add the following lines

```
define('WP_ALLOW_MULTISITE', true );  
define('WP_DEBUG', false);  
define('MULTISITE', true);
```

```
define('SUBDOMAIN_INSTALL', true);  
define('DOMAIN_CURRENT_SITE', 'bharatavani.in');  
define('PATH_CURRENT_SITE', '/');  
define('SITE_ID_CURRENT_SITE', 1);  
define('BLOG_ID_CURRENT_SITE', 1);  
define('WP_MEMORY_LIMIT', '5126M');
```

WordPress Dashboard of Network setup to make or configure multisite network features.

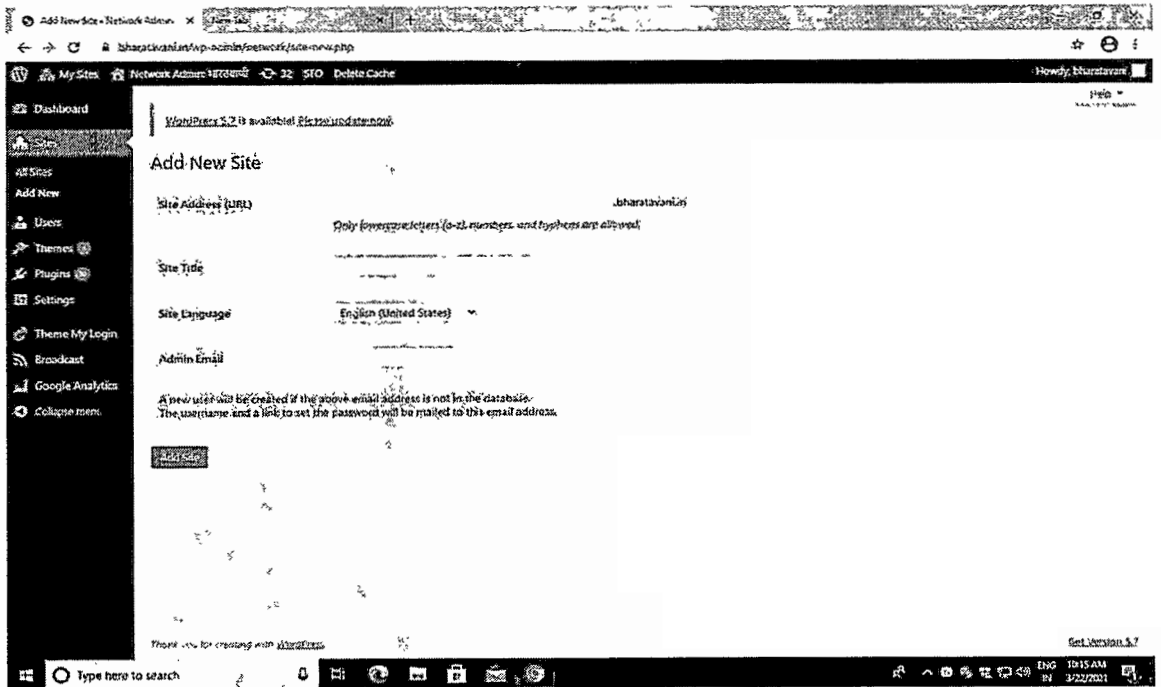


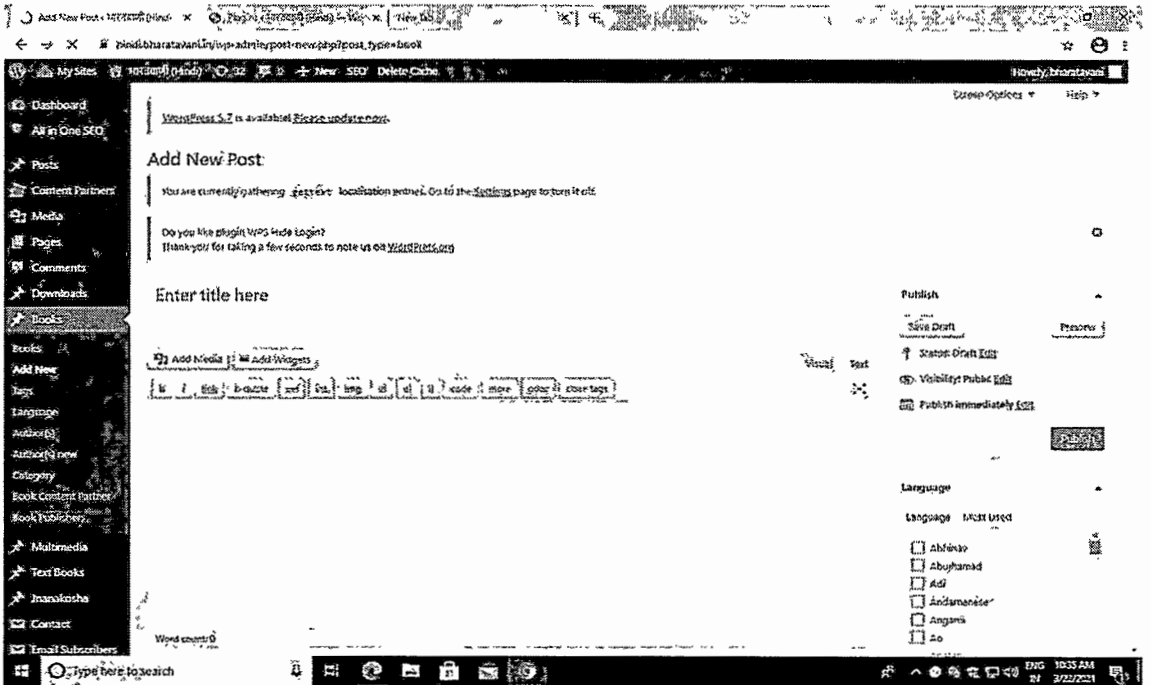
Figure 1 represents the Create new site for subdomain language option in Network Admin and create their structure based on theme for displaying the respective language domain. The result of the websites will be hindi.bharatavani.in, kannada.bharatavani.in, tamil.bharatavani.in, etc.

iii. Create custom post type

Bharatavani project resources are mainly available in the form of PDF files through scanning of books that is received after signing legal agreements with concerned content partners. Bharatavani also receives soft copies of audio and video files etc. Based on classification of contents, the same is hosted in different sections like bhashakosha,

Textbooks, Dictionaries, Multimedia, etc. and thereafter custom posts are created using WordPress Plugins.

Ex: In Hindi domain, for adding the new bhashakosha book post.



In Figure 2, the data will show the title, language fields are visible. For adding a book, the title of the book, language, author, category, publishers, content partner, broadcast language (if book is mapped multi language), featured image, pdf location of url, unique id, etc are required to be updated.

iv. Create Subdomains

Subdomains are created using network setup tools but the functionality view and outside view is different from their menu structures like pages, page setup, tag lines, etc which are required while inserting subdomain language information.

v. Create taxonomies related with custom post type

Classification of books are important to manage the website in a user-friendly manner that would help to search the books based on dictionaries, grammar, glossary, literature, Language learning, Terminology, Literature, etc. Similarly taxonomies are involved the publishers, content partners, multi languages, producer, classes, subject, book types, etc.

vi. Create a custom page for website requirements of visible information.

Creating a Custom page is required for all the necessary information to be visible in the website. It also interacts with many tools of plugins for any of the functionality.

vii. Theme options

Theme options are required in the WordPress environment for providing templates and developers can also create their own theme. The Appearance involves the Widgets, Menus, Header, Customizing the CSS, etc and chooses options for their subdomain websites.

viii. Installing supported plugins for multisite websites

The plugins are required to meet various parameters in many aspects of websites. The plugins are involved the design and coding is simple to manage the website especially.

ix. Creating own dictionary plugins or retrieving data through a database of source code.

x. All other minor functionalities of search, delete, add, update, etc.

3. Results

Bharatavani website is working properly throughout the multiple language environments and Google analytics dashboard expresses the user and activities. More than 27000 are registered and page views activities are in Figure 3.

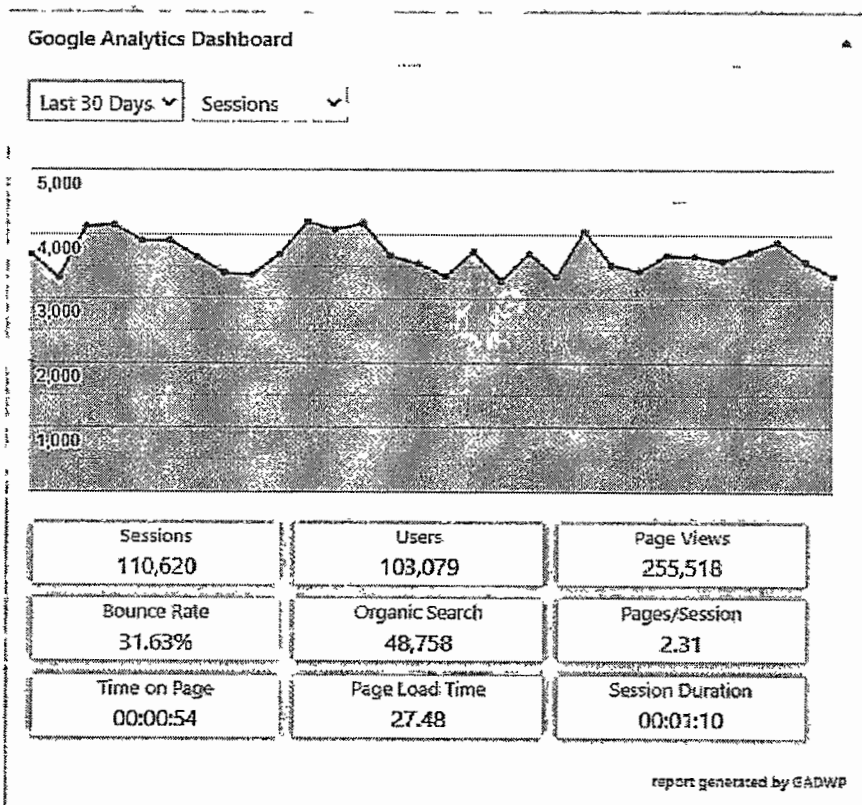


Figure 3

The Bharatavani project published over 220 text searchable dictionaries, more than 430 PDF dictionaries, 1000 plus bhashakosha resources, more than 2800 Jnanakosha contents, 900+ textbooks and above 1000 video files.

Conclusion

Bharatavani project is therefore a complete solution for viewing knowledge resources in and about Indian languages. It can be effectively concluded that the Bharatavani portal gives the opportunity to the users to not only learn various languages but also enables them to get access to the huge and diverse knowledge that are produced by Indian linguistic communities. Furthermore, the website structure is an example in itself about management of a multisite environment through the inclusion of specific domains based on Indian languages.

REFERENCE

Brain Messeniechner and Jason Coleman, Building Web Apps with WordPress

James Mallison, Mastering PHP 7.

Krishna Shasankar, Zend Framework 2.0 by Example.

Nathan George. (2018) WordPress FOR BEGINNERS: A Visual Guide to Building Your WordPress Site.

Arun K Pujari.(2001) Data Mining Techniques. Hyderabad, University press

Jaiwei Han, Michelinne Kamber. Data Mining : Concepts and Techniques.

C S R Prabhu (2002). Data Warehousing Concepts, Techniques and Applications 2nd Edition. New Delhi: Prentice Hall of India,

Online Sources

Census of India (2011) paper 1 of 2018 Language [Internet]. Available: http://censusindia.gov.in/2011Census/C16_25062018_NEW.pdf