

Digital Divide of Dogri Language

Shubhnandan S. Jamwal

jamwalsnj@gmail.com

Department of Computer Science and IT, University of Jammu, Jammu

Abstract

Dogri language is the most prominent language of the J&K and now has also become the official language of the J&K. In this paper, the challenges before the Dogri language have been discussed. The challenges such as development of the text processing tools such as Machine translation (MT), text processing tools, spell checkers, stemmers, morph analyzers etc., has been discussed. The initiatives taken by the govt. of India for managing the digital divide of the Dogri language has also been discussed.

Keywords: Digital Divide, Dogri, internet, languages

Introductions

The term "Digital Divide" refers to the disparity that exists in most countries between those who have quick access to information and communication technology tools and the knowledge that they supply and those who do not. Sometimes, a significant symmetry in the distribution and efficient use of information and communication resources between two or more populations is also referred to as the digital divide. With the emergence of the omnipresent society, the world we live in has been fast changing, bringing with it incredible benefits and opportunities, as well as new challenges. The ability to generate and use data plays a crucial role. Within the global community, a new divide in technology adoption between the industrialized and developing worlds is visible, and it may be even more substantial. The digital divide arising from the lack of reliable information, infrastructure and digital literacy is a major cause of social and economic backwardness. The development of the regional languages in the modern digital world remains an efficient tool for the development of the social and economic conditions of the society. The roll of the digital technology and the language is very much evident in the field of the digital transactions, news, gaming, ICT practices, and in many other services offered by the government. Therefore, the development of the regional language becomes very important for getting the benefits of the modern digital world.

Literature Review

K. V. Ketan, A. Ponnada and P. Yammiyavar[1] have highlighted the potential of several existing technologies in enhancing the gaming experiences of rural and semi-urban populations of India and emphasized on the importance of cultural differences in designing the delivery content for the digital games. We argue that the widely discussed digital divide can be mitigated by considering both technological aspects and cross-cultural aspects of the target users. K. Mahmood, Z. Nayyar and H. Mushtaq [2] published a paper for filling the gaps between E-Government and digital divide by utilizing the multi-channel services governed by transformational leaders. R. Handayani and G. Afrizal[3] in Indonesia carried out a study where it has been observed that nowadays the price of smartphone is getting affordable, yet smart phone cannot replace the computer completely and to overcome the digital divide, the society need to experience in using computer and the skill of the computer



can be more effective and efficient. N. Wedasinghe and R. Wicramaarchchi[4] proposed a model to eliminate Disability Digital Divide in Sri Lanka. This model includes guidelines to eliminate categories under disability digital divide in Web services, Mobile devices, Computer devices. This paper further proposes to consider the areas in ICT to be considered and the responsible authorities and stakeholders who need to take the initiation to eliminate the disability digital divide in Sri Lanka.

The problem of the digital divide in African-American students are also studied[5] and it is observed that digital divide is not merely about Internet access but it involves access to the social networks that ease the path to success in high-tech careers. Muhammed Azam [6] observed that ICT is relatively less expensive technology and so would solve many of the problems of the poor regions; especially in rural villages. Its expansion may have far-reaching effects on reducing poverty and Bridging the Digital Divide. They have explored the role of the telecommunication sector for bridging the digital divide. It is also believed that connecting people across the Digital Divide is as much a social effort as a technological one [7] and a community-centered approach is required. Preliminary trials have yielded interfaces that deal with poor quality by adapting Instant Messaging techniques for multiple modalities, providing improved semi-synchronous communication.

The researchers using data sets of the Dogri language in the digitization of the Dogri language have worked on models for generation of the verbs of the language [10] and the automatic generation of the Noun using Machine learning [11]. English-Dogri parallel corpus has also been prepared for machine translation from one language into another language and Moses, which is a statistical machine translation system has been used to train translation models for any language pair[12]. A rule based transliteration system has also been developed for detection of proper nouns for Dogri to English[13].

Challenges before Digitization in India

Having a smart phone is not a proof of being digital because the smart phone without internet is of no use and unless a person knows how to get relevant and timely information on the internet he/she cannot be called as 'digital savvy'. According to the data available with National Sample Survey of India, only 27 percent of the households in India, where only one member has internet facility. Therefore, this limitation is the big problem in the digital world and only 2.5 percent of the students in India have internet access. When we know that 67 percent of India's total population lives in rural areas and 3G & 4G technologies are hardly available in remote areas of J&K, the dream of the digital J&K is quite distant. It is also said by many experts that without educating a mother we can educate a family. When the gender gap in India translates into ownership of mobile use, this gender gap widens and only 16 percent of women in India are connected to mobile internet.

Challenges for the Dogri language

The challenges which are faced for the development of the Dogri language are not because of the less advances for the digital technology but the main reason is the lack of fusion of the technocrats and linguists of the Dogri language. The knowledge of the characters is the literacy of the old era, which seems to be invalid in today's world. If the ignorance of the characters set is considered as illiteracy in the earlier world then the having

no knowledge of the digital world is considered as digital illiteracy. The Digital India program is a new initiative of the Government of India towards making India prosperous. Its main objective is to make the country new records in the field of science and technology. Through this, the only goal is to empower the country digitally. In the present era, today the same country is ahead which has made science and technology the medium of progress of its country. This campaign is to revolutionize the country through the Internet, as well as strengthen the technological side of India by empowering the Internet. Due to digitization, now we can book train, plane, and bus tickets sitting at home. No need to stand in long queues anymore. Now everything is possible online. Any information needed, everything is available on the Internet. No time, and have to shop, no problem, shop online, sitting at home. E-commerce platforms have given livelihood to many. Now the researchers and digital game publishers are also thinking to reduce the digital divide between urban and rural digital gamers in India.

1. The problem of the low resourced language remains another challenge for the Dogri. The non availability of the NLP tools, high-quality corpora, morphological engines, spellcheckers, stemmers, POS taggers, Text-to-speech etc., makes the language processing hard.

2. The other challenge that is faced by the Dogri language is the non standardization. The various dialects of the Dogri language have different words for the common things has not been standardized among the different dialects.

3. Another problem in the design and development of the language is the orthography. There are some consonants such as घ (gh), झ (jh), ढ (dh), ध (dh), भ (bh) in Devanāgarī script which exist only in Dogri orthography. And they are used phonetically used for representing tonal क (k), च (c), ट(t) , त (t), प (p) at initial stage of a word e.g. घर (ghar/house), झड़ना (jhaṛanā/ shed), ढक्कन (ḍhakkān/ lid) , धमकी (dhamakī/ threat), भारी (bharī/ heavy) etc.

4. Dogri language is some time also called as tonal language and tone is the Major feature of the Dogri language. In Dogri language we have three types of tones i.e., level tone, low or low-rising tone, high or high falling tone. In the digitalization of the Dogri language for the development of the spoken text recognition, the representation of the tones remains a big challenge.

Initiatives taken to remove the digital divide

Numbers of initiatives have been taken by the government for the development of the language resources like the development of the glossary of commonly used free and open source software terms for the Mozilla firefox in Dogri-English language in xls format, Unicode Typing Tool, Dogri SakalBharati Unicode font [8]. The web site also contains English Dogri parallel corpus links of the terms, which has been imported programmatically from online glossary tool available at tdil-dc portal. This parallel corpus contains the glossary of the terms of Mozilla Firefox but this browser is having only 3.37% market share. Therefore, the relevance of the developing such a system becomes impractical when only few people are using any technology in the world. The web site also contains some utilities which are either not download able or not installable because of the change in the architecture of the computers in the forms of bits.

Text corpora of the Dogri has been developed by the Linguistic Data Consortium for Indian Languages (LDC-IL) which is the scheme of the Department of Higher Education, Ministry of Human Resource and Development, Government of India and the concerned development is going at the Central Institute of Indian Languages, Mysore[9]. The concerned agency is also instrumental in the development of the other language resources like text and speech corpus of other Indian languages but the data for the development of the text to speech recognition for the Dogri language is still not available.

The government of the India has also taken the initiative for the development of the morphological engine of the Dogri language, under the project "Shallow Parser Tool for Indian Languages" with the University of the Hyderabad. The project was initiated in 2012 and only first phase of the project is completed.

Future of the digital Dogri

There are number of projects associated with other languages of the world, which are using Artificial Neural Networks systems based on the functioning like that of neurons in the human body. Models are created that are based on Artificial intelligence that learns by examples. The text Sequences are formed with words that needs to be changed as required by the as input to the LSTM-RNN. The development of such systems needs to be taken for the Dogri language also.

References

- [1] K. V. Ketan, A. Ponnada and P. Yammiyavar, "Reflections on user-centered and technology-centered issues in a digital game design process: An approach to bridge the digital divide," 2012 4th International Conference on Intelligent Human Computer Interaction (IHCI), 2012, pp. 1-3, doi: 10.1109/IHCI.2012.6481777.
- [2] K. Mahmood, Z. Nayyar and H. Mushtaq, "Role of transformational leadership on e-government switching: Multi-channel and digital divide," 2017 12th International Conference for Internet Technology and Secured Transactions (ICITST), 2017, pp. 475-478, doi: 10.23919/ICITST.2017.8356453.
- [3] R. Handayani and G. Afrizal, "Thin Clients as Memoryless Computer for Reducing Digital Divide in East Indonesia," 2018 12th International Conference on Telecommunication Systems, Services, and Applications (TSSA), 2018, pp. 1-5, doi: 10.1109/TSSA.2018.8708828.
- [4] N. Wedasinghe and R. Wicramaarchchi, "Web, Mobile and Computer related model to bridge the disability digital divide in Sri Lanka," 2014 International Conference on Electrical Engineering and Information & Communication Technology, 2014, pp. 1-6, doi: 10.1109/ICEEICT.2014.6919120.
- [5] Fay Cobb Payton. 2003. Rethinking the digital divide. *Commun. ACM* 46, 6 (June 2003), 89–91. DOI:<https://doi.org/10.1145/777313.777318>
- [6] Muhammed Azam. 2008. Bridging digital divide: role of telecommunication sector. In *Proceedings of the 2nd international conference on Theory and practice of electronic governance (ICEGOV '08)*. Association for Computing Machinery, New York, NY, USA, 487–488. DOI:<https://doi.org/10.1145/1509096.1509204>.
- [7] William D. Tucker. 2004. Connecting bridges across the digital divide. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems (CHI EA '04)*. Association for



Computing Machinery, New York, NY, USA, 1039–1040.

DOI:<https://doi.org/10.1145/985921.985968>

[8] <http://www.tdil-dc.in>

[9] <https://www.ldcil.org/resourcesTextCorp.aspx>

[10] Jamwal S.S., Gupta P., Sen V.S. (2021) Hybrid Model for Generation of Verbs of Dogri Language. In: Singh T.P., Tomar R., Choudhury T., Perumal T., Mahdi H.F. (eds) Data Driven Approach Towards Disruptive Technologies. Studies in Autonomic, Data-driven and Industrial Computing. Springer, Singapore. https://doi.org/10.1007/978-981-15-9873-9_39.

[11] Shubhnandan S. Jamwal, Named Entity Recognition for Dogri using ML, International Journal of IT & Knowledge Management, Jan-June 2017 Volume-10, Number-2 pp. 141-144.

[12] Singh, Avinash & Kour, Asmeet & Jamwal, Shubhnandan. (2016). English-Dogri Translation System using MOSES. Circulation in Computer Science. 1. 45-49. 10.22632/ccs-2016-251-25.

[13] Tejpaul S. Sasan, Dr. Shubhnandan S. Jamwal, Transliteration of Name Entities Using Rule Based Approach, International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume 6, Issue 6, June 2016.