BIG DATA IN TAMIL: OPPORTUNITIES, BENEFITS AND CHALLENGES

R.S. Vignesh Raj[1], Babak Khazaei[2] and Ashik Ali[3]

[1, 2]*Sheffield Hallam University, United Kingdom*

E-mail: [1]classicalraj@gmail.com, [2]b.khazaei@shu.ac.uk

[3]*FireflyApps Ltd, United Kingdom*

E-mail: ashikali.hn@gmail.com

Abstract

This paper gives an overall introduction on big data and has tried to introduce Big Data in Tamil. It discusses the potential opportunities, benefits and likely challenges from a very Tamil and Tamil Nadu perspective. The paper has also made original contribution by proposing the 'big data's' terminology in Tamil. The paper further suggests a few areas to explore using big data Tamil on the lines of the Tamil Nadu Government 'vision 2023'. Whilst, big data has something to offer everyone, it is argued that Tamil language proficiency and favourable policy decision is key to the success of Big data in Tamil and more specifically in Tamil Nadu.

Keywords

Tamil Computing, Tamil Nadu, Big Data, English Language, Technology, Ilanji Tharavu

1. INTRODUCTION

Computing in Tamil is certainly one of the missions that the IT Department of Tamil Nadu is striving to achieve for over 70 million native Tamil speakers [25]. Therefore proficiency in Tamil in future could play a major role [3]. According to 2013 estimates, 4 zettabytes of data were accumulated. One zettabyte is 1,000,000,000,000,000,000,000 bytes [12] and that is 'Big Data'. It is estimated that by 2020, it would reach to 40 zettabytes. Language plays an important role in technology. Although there are quite a few advancements and technology in the field of Tamil Computing, the important question that needs to be answered is: 'Are they well received and put to use?' This paper introduces the potential opportunities of reinventing 'Big Data' in Tamil, the benefits it has to offer and the challenges it presents in order to achieve them.

Big data not just helps in understanding and predicting customer behaviours but also provides some useful insights on some of the key issues that almost all countries face ranging from unemployment to anti corruption and intelligent traffic system [11]. This paper aims to list some of the key areas that may be of some interest from the Tamil Nadu perspective along with the limitations identified.

According to Localisation Industry Standards Association (LISA), localisation is the process of adopting a product to meet the language, cultural and other requirements of a specific market (She-Sen Guo, 2003). It may not necessarily be confined to products. This paper further proposes the Tamil term for 'Big Data' on the lines of localisation to take it closer not just to the potential IT professionals who are native Tamil speakers but also to the not so tech savvy Tamil people who may not be into Computing since it is felt that 'Big Data' is a generic term that has something for everyone.

2. BIG DATA IN THE CONTEXT OF TAMIL LANGUAGE AND RELATED WORK

According to Oxford dictionary, data may be defined as 'facts and statistics collected together for analysis'. Literature review suggests that 'data' and reference to 'Big data' is not language specific which means that the data obtained can be in English, Tamil, Mandarin or Arabic. The word 'data' in this paper's context shall refer to Tamil. Although there have been efforts to translate the term 'Big Data' into Tamil, it is felt that the term is more of a literal translation of the English equivalent. There is very little work in done in 'Big data in Tamil' therefore it has been challenging in discussing related work with specific reference to Tamil.

3. RATIONALE BEHIND COINING A TAMIL TERM

Big data- as the name suggests refer to complex data in huge numbers - 4 zettabytes. Although the equivalent translation of 'Peruntharavu' exists in popular usage, it is felt that a more meaningful term is required to represent the volume of data that is being referred to. Along with translation on the lines of localisation, comes a social benefit. It is a pleasure for people to work with computers or technology in their native tongue. Recognition of language within technology is extremely important from the 'belongingness' point of view as the technology and computers have traditionally evolved around the English language to cater to the English speaking community. [18]. The non native English speakers may not interact with the technologies in the same manner as the native English speakers do. Many people quite strongly identify with the culture that they were brought up with and when their language is not accepted in technology, they may feel personally rejected and second rate and possibly they may never approach IT or even attempt to approach IT in their native tongue when an opportunity arises. But in a sense, the potential of these people are unknown until an opportunity is presented to them [18].

This paper proposes the Tamil terminology for Big Data as **இளஞ்சி தரவு** ('iLanji tharavu'). The proposal of this terminology is based on the volume of the data - the 'big data' represents. The Tamil term 'Ilanji' is equivalent to one zettabyte which is 1018. Therefore when the term 'iLanji tharavu' is used, it not just refers to 'Big Data' but also to the volume of 'Big data' and makes it meaningful for the Tamil developers and common people to visualise the amount of data that is being referred to.

4. SCOPE OF BIG DATA

The English Oxford dictionary defines a world language as “A language known or spoken in many countries”. With over 70 million native Tamil speakers and official in two countries and two dependent territories, Tamil can certainly be classified as a ‘world language’. It is important to understand how the native Tamil speakers ‘perceive’ Tamil language and to what extent is the language being used in daily lives because much of big data, its analysis and complexities of analysis is quite reliant on the language used. Eagle and Lazer (2009) have demonstrated that it is possible to accurately infer 95% of friendships based on observational data alone. The issues of behavioral and social science from the big data perspective have already begun. It is possible to use large scale mobile data as input possibly from social networking sites like Facebook and Twitter to characterize and understand individual trait, behaviour pattern, and human mobility to name a few [16]. One of the critical aspects that is suggested be taken into account while dealing with Big data in Tamil is the script used on social networking sites like Facebook - Is it Tamil in native script (**வணக்கம்**) or is it Tamil in Roman script (Vanakkam). It is in this context, the challenge that one may have to deal with is with the issue of code mixing especially in when using the Roman script to write Tamil language [20]. Within this context, the language and its proper use by the community plays a vital role and is more likely to influence the technology. Although, one may claim to use Tamil on Facebook, previous research experiments suggest that the reference in most cases refers to Tamil being written in the Roman script.

4.1 BIG DATA AND POTENTIAL BENEFITS FOR TAMIL NADU GOVERNMENT

Under the Tamil Nadu IT initiatives, Computing in Tamil and e-Governance has gained prominence [25]. The author in the context of data security suggests that the Government has vast amount of data online. The larger question from the big data perspective is ‘What language is the data in?’ Although, Tamil is the only official language of Tamil Nadu, over ninety percent of the Government websites (statistics based on Tamil Nadu Corporations, Tamil Nadu police) are only in English [20]. It suggests that the divide between the people and Government is seemingly apparent since the ‘Official language’ is yet to be fully implemented in the state.

Businesses, governments and the research community can all derive value from ‘big data’. The main goal of a government is to maintain domestic tranquility, achieve sustainable development, secure citizens’ basic rights and promote the general welfare and economic growth [15]. Some of the developed countries have shown the way in adopting the concept of big data in order to develop useful applications that takes priority. Some of the examples include Japan’s Intelligent Traffic System, Korea’s Employment Position statistics, Singapore’s Risk Assessment and Horizon Scanning, UK’s Horizon Scanning Center.

The first step could be to identify issues within Tamil Nadu that needs immediate addressing. The expectation of a government out of big data typically revolves around economy, health care, job creation, natural disaster and terrorism [15]. According to Economic Survey 2012-13, Tamil Nadu ranks 7th in unemployment. Owing to lack of clarity and data, it is at this point assumed that unemployment in Tamil Nadu refers to the native Tamil speakers and the migrant non Tamil speakers who are permanent residents in Tamil Nadu state.

Big data is a multidimensional concept that embraces technology, decision making and public policy. It is quite possible that the socio-technological pose challenges and problems [9], [10]. The policy decisions to some extent influences the data and its analysis. One of the likely challenges in dealing with the data sets from Tamil Nadu is its language policy. Although the state’s only official language is Tamil, anecdotal evidence suggests a disconnect in its implementation and the members of the public adhering to the law of the state. As a result, the data is more likely to be in either English or Tamil or both which might add to the complexity of analysis. Therefore the first step suggested towards Big data in Tamil is to translate non Tamil data which may include but is not limited to data in Romanised Tamil into Tamil and in Tamil script. This in itself might sound like a ‘big’ task but it is argued that it could help in analysing and interpreting the data and reduce the complexities at that level.

4.2 APPLICATIONS OF BIG DATA

Big Data could play a pivotal role in some of the advanced technologies like data mining, speech recognition, cross lingual information retrieval. Shriram and Sugumaran (2009), in their research paper demonstrate on how cross lingual information retrieval can be achieved using data mining techniques. Some of the examples cited by them relate to English- Tamil- English but dealing with Tamil script- Romanised Tamil- Tamil script can particularly be challenging. It is further felt that while dealing with cross lingual information retrieval, it is important to take into account the native script rather than Romanised Tamil. Muni Kumar and Manjula (2014), in their paper have argued that ‘Big data’ could help in ‘big’ ways to achieve a more reliable and an efficient healthcare system. It is envisaged that the health care data is more likely to be in English than in Tamil which could possibly add to the complexity of data fusion, analysis, classification, inference and interpretation. It is suggested that any big data project in Tamil from the analysis, data fusion and interpretation point of view must take into account the following:

1. Tamil script
2. Regions of potential interest where Tamil is official or is a recognised minority language such as Singapore, Sri Lanka, Malaysia, Mauritius and the Reunion.

It is observed that in the relentless pursuit of achieving something out of big data, some of the very basic yet critical aspect such as language and the script used is quite often either ignored or missed. For Big Data to be reinvented in Tamil, it is recommended to obtain a reasonable proficiency in Tamil which includes, reading writing and speaking. Muni Kumar and Manjula (2014), in their paper have referred to terms like ‘data’ and ‘medical records’ and ‘medical history’. But it in the context of big data, is it assumed that they are all in English? We wish to reiterate that approaches to big data and chosen methodology to pursue with complex analysis and interpretation involving big data could have consequences. For instance, if the data, for some reason are captured in English in Tamil Nadu where Tamil is the only official language, the ethical question that needs considering is 'What if a person who is only literate in Tamil wishes to record his details?' There is a social implication for the technological decisions made at each level. One of the key dimensions that could help in bringing technology closer to the intended users is the 'cognitive dimension' which is usually applied to people who share common language. The common language facilitates their ability to gain access to people and their information [30]. This could then be further extended to countries like Sri Lanka, Singapore, Malaysia and broadly to the Tamil diaspora.

Speech recognition is typically monolingual and such technologies are largely user dependent. Therefore it is argued that Tamil language proficiency and the ability to pronounce the Tamil sounds correctly is vital. Owing to the nature of the language, it is perceived that mispronouncing could have an impact on speech recognition especially in sounds like **ழ**, **ள**, **ஞ**, **ண**. Code mixing Tamil and English could possibly be another challenge that needs to be dealt with. There are quite a few reasons for code switching. Some of them are:

- Limitation of vocabulary in the language
- When someone lives in a society where more than one language is spoken. . Some of the examples are: French- German in Switzerland, Cantonese- English in Hong Kong, English- Spanish in the US, Mandarin- English in Malaysia [29]

It is opined that none of the above is applicable to Tamil and Tamil Nadu. From the official language perspective, Tamil is the only official language of Tamil Nadu and the Official language Act 1976 of the Indian Union exempts Tamil Nadu from the Act. It is therefore argued that Tamil Nadu is largely a monolingual state. The percentage of migrants to the total population of Tamil Nadu state as of 2001 was 25.4 [27].

While there are ongoing researches on speech recognition in Tamil and other Indian languages, it is important that the Tamil speech recognition is categorised into (a) Native Tamil speakers and (b) Non Native Tamil speakers. One of the reason for this categorisation is that the accent non native Tamil speakers is more likely to differ from the native Tamil speaker and in that respect it is important to closely study the pronunciation pattern of native Tamils and non native Tamil speakers. [31], [28]. Some of the research experiments suggest that the native Tamil speakers themselves have difficulty in getting the Tamil pronunciation right [20]. Big data although to a great extent can be beneficial in such applications, it is suggested that the underlying emphasis ought to be given to the language skills. And as discussed in the previous sections on how some of the more advanced countries have used big data in order to address their priorities, and on the lines of the Tamil Nadu Government's 'Vision 2023' that seeks to enhance the state's economic and social performance and further generate employment, the following areas could be explored using 'Big Data' and may be of particular interest to Tamil Nadu:

- Employment scope for Tamil medium students in Tamil Nadu.
- Is there a relationship between unemployment and health issues?
- Ways of bridging the employment gap filled by migrants.
- Choice of employing migrants in Tamil Nadu: Is it a will issue or a skill issue?
- Prospects of Tamil as a language in the context of employment and commerce.

5. CHALLENGES

Raj, Babak and Ashik (2015), in their paper on 'Attitude of Tamil Nadu Tamils towards Tamil Computing' have identified that there exists a strong relationship between 'attitude' towards Tamil language and its acceptance and usage in various fields especially in technology. Based on UNESCO factors, the author's have attempted to classify Tamil in the context of Tamil Nadu and technology and have predicted it to be vulnerable since the language is largely restricted to certain domains. For the purpose of their research, the Tamil referred was without the effect of code mixing. It also depends on the source of data. For instance, if one decides to use limited amount of data from social networking sites such as Facebook and Twitter, then language and script could be a possible challenge. There is little guarantee that all the data will be in Tamil and in Tamil script. At the same time one needs to be familiar to deal with the issue of code mixing and more importantly without changing the meaning in that context. Related to the source data is ethical issue. One of the key factors that play a major role is ethics. It is felt that the data should not be exploited and privacy of individual must be protected. Exploitation of big data without regarding the legal issues, data quality and process quality quite often results in poor decisions and puts the individual to a greater risk. In most cases, the individual bear the consequences of an organization's low- quality decision making [4]. However, the perception of the word 'privacy' means different to different people. There is a huge difference in the way the Westerners perceive 'privacy' and the manner in which 'privacy' is perceived in India. To a statement 'Data security and privacy is not really a problem because I have nothing to hide', 89% of the US subjects disagree while only 21% of the Indian subjects disagree [22]. In this context, it might be interesting to find out the perception of 'privacy' from the native Tamil speakers who are residents of Tamil Nadu. Therefore, there seems to be a social science element involved even though the major focus is on 'big data' and technology. When a data is said to be protected, it would mean that an individual cannot be identified. In most of the western countries like the UK, data protection law exists and it is felt that there is a substantial awareness since almost all companies are required to adhere to the 'Data protection Act'. To a question on 'Identity theft' in a research, 21% of the Indian subjects were concerned against 82% of the US subjects [22]. Although that research study concludes that the perception of 'privacy' was very different in the two regions, it cannot be ignored on those grounds. Understanding ethical implications and ensuring ethical compliance while collecting data could be a challenge especially for native Tamil speakers. The reason why we call it as a 'challenge' is based on some of the findings in the previous research on 'Attitude of Tamil Nadu Tamils towards Computing in Tamil'. A few components that were taken into account and measured against certain parameters was 'attitude towards language', 'language skills' against their environment and domestic conditions that included but was not limited to policy decisions and its implementation. It is further important from an ethical perspective to reach out to the potential participants and make them fully aware of what data is being collected, why is it being collected, how will it be used, how is it going to affect or not affect the participants, potential benefits and if there are any consequences or risk in their participation. Therefore, even if the data is being collected from social media sites like Facebook, and even if the person is a close friend, it is felt that using their data on the Facebook without their consent is a breach of privacy. Although there are a few other challenges that are related to language and data analysis and data mining, from Tamil Nadu perspective, it is felt that collecting data ethically could perhaps be the first challenge. Surrounding the ethical issue is the policy decision and law that relates to data protection. The Indian constitution has provisions for privacy and data security, but the extent to which it is being implemented needs to be understood in the socio-technological context. Do people realise its importance? Are people aware of such laws? Do the people realise its consequences? Are some of the basic questions that need to be answered as a part of any big data project in Tamil Nadu.

Though many feel that Tamil as a language has little value in Tamil Nadu [20], contrary to their perception, it is felt that the attitude could deter progress of the language in technology especially in big data. One of the key things that needs to be borne in mind is the proficiency in language in order to analyse and interpret the results. Therefore, proficiency in Tamil is extremely important should the community aspire the language to grow in technology. Code mixing effect and writing Tamil in Roman script is yet another challenge that needs to be dealt with whilst analysing the data set. The findings from previous research experiments, suggest that quite a large number of native Tamil speakers for various reasons use the Roman script to write Tamil [20]. For quite a few of them, Romanised Tamil seem ‘proper Tamil’. It is therefore felt that the perception and definition of Tamil in itself is slowly changing depending on various other factors including but not limited to the commercial value of the language. There is no doubt that big data has enough to offer for businesses, government and community but as mentioned earlier, it is a multi- disciplinary approach and it is opined that change at any level may have consequences. It is inevitable to consider the social aspect, attitude, way of life and to certain extent the policy decisions that has an overall impact whilst answering some of the suggested questions relating to Tamil Nadu.

Like in research methodology, the data collection process, source and the methodology involved in collecting the data is critical since the analysis and the final result is arrived on the basis of the data obtained. Some of the things to consider is the heterogeneity of the data, bias in the data and data sampling. It is argued that strictly adhering to ethics is more likely to produce a quality and a more authentic data than otherwise. Big data although useful in predictive analysis to come up with innovative solutions, it must always be remembered that most of the times, it could be an individual’s data that might be used and carelessness in handling the data could expose the individual’s identity. Some of the actions may not be intentional but in the process of data fusion, at some point, it is important to think about the question ‘Will this identify the individual?’ Constant process improvement could help improve individual’s identity protection and at the same time, positively contribute to big data analysis.

6. CONCLUSION

Big Data in Tamil - is a unique approach contrary to the popular assumption in Tamil Nadu that anything to do with technology has to be in English [3]. It is believed that for Big Data Tamil initiative to reach the level of 1 zettabyte, the Tamil community ought to use Tamil language in Tamil script on the internet and handheld devices. Although for some of the techniques and computation, English may be used but it is opined that it may still be possible to innovate equivalent techniques in Tamil. By taking Cross Lingual Information Retrieval using the data mining approach to the next level, although an innovative approach, are we encouraging ‘code mixing’ amongst the native Tamil speakers? The intent of pursuit may be different but it could implicitly mean encouraging code mixing. It is opined that there are still quite a few unsolved challenges ahead such as - Tamil language skills, code mixing, attitude of native Tamil speakers towards Tamil and Computing in Tamil, writing in Romanised Tamil to name a few. Based on the previous studies and experiments, it is foreseen that these could perhaps be detrimental in taking Tamil Computing including the proposed Big Data in Tamil to the next level.

This paper has contributed the Tamil term for big data - ‘Ilanji Tharavu’ which is believed to be the first of its kind in the field of Tamil Computing with the logic and reasoning that any terminology and methodology in the target language Tamil in this case, must take into account the cultural aspect rather than a literal translation of the English equivalent. The researcher and the team hopes that the term would be well accepted and put to good use by the Tamil community.

The challenges discussed in this paper is extensively about ethics and language although there are a few others such as government policies towards Tamil development and Information Technology, sampling issues but it was felt that of all these, ethics and language was perhaps the most important challenge that one has to overcome in the near future even before moving on to the next stage in big data analysis. Some of the challenges may relate to the policy decisions of the government. It is further felt that laws introducing Tamil proficiency test for jobs in all sectors in Tamil Nadu, mandatory Tamil medium education could not just help improve Computing in Tamil but may also have a positive impact on the usefulness of Tamil as a language in Tamil Nadu which is vital in order to further in the field of technology [3].

7. LIMITATIONS AND FUTURE WORK

The researcher and the team recognise that there are quite a few avenues for improvement within this area. Although this paper could form the base for many more innovative approaches within ‘Ilanji Tharavu’, we believe that a pilot study in this field could perhaps be more informative and may reflect on the discussions presented in this paper.

REFERENCES

[1] Tim Harford, “Big data: A big mistake?”, *Significance*, Vol. 11, No. 5, pp. 14-19, 2014.

[2] Nicole Lazar, “The Big Picture: Big Data Computing”, *Chance*, Vol. 28, No. 2, pp. 39-42, 2013.

[3] R.S. Vignesh Raj, “Attitude of Tamil speakers towards Tamil Computing,” *Proceedings of International Tamil Internet Conference*, pp. 275-279, 2014.

[4] Marcus R WIGAN, Roger CLARKE, “Cover feature”, *IEEE Computer Society*, pp. 46-53, 2013.

[5] Kanthimathi Krishnaswamy, “Code mixing among Tamil-English bilingual children,” *International Journal of Social Science and Humanity*, Vol. 5, No. 9, pp. 788-792, (to be published in Sep. 2015).

[6] http://www.unesco.org/new/en/culture/themes/endangered-languages/atlas-of-languages-in-danger/ accessed on January 06, 2015.

[7] http://www.census.tn.nic.in/PCA_data_highlights/fig_glance_tn.pdf accessed on January 03, 2015

[8] Xindong Wu, Xingquan Zhu, Gong-Qing Wu and Wei Ding, “Data Mining with Big data”, *IEEE Transactions on Knowledge & Data Engineering*, Vol. 26. No. 1, pp. 97-107, 2014.

[9] Connie L. McNeely and Jong-on Hahm, “The Big (Data) Bang: Policy, Prospects and Challenges”, *Review of Policy Research*, Vol. 31, No. 4, pp. 304-309, 2014.

[10] Judith Bayard Cushing, “Beyond Big Data?”, Computing in Science and Engineering, Vol. 15, No. 5, pp. 4-5, 2013.

[11] Giuditta De Prato and Jean Paul Simon, “The next wave: big data?”, *Communications*, 1st Quarter, No. 97, pp. 15-39, 2015.

[12] Michael J. Pentecost, “Big Data”, *Washington watch*, Vol. 12, No. 2, 2015.

[13] Babak Falsafi and Boris Grot, “Big Data - Guest editor’s introduction”, *IEEE Micro*, Vol. 34, No. 04, pp. 4-5, 2014.

[14] Kathy Wren, “Big data, big questions”, *Science*, Vol. 344, No. 6187; pp. 982-983, 2014.

[15] Gang-Hoon Kim, Silvana Trimi and Ji-Hyong Chung, “Big-data applications in the government sector”, *Communications of ACM*, Vol. 57. No. 3, pp. 78-85, 2014.

[16] Juha K. Laurila , Jan Blom , Olivier Dousse , Daniel Gatica-perez , Olivier Bornet , Julien Eberle , Imad Aad and Markus Miettinen, “The mobile challenge: Big Data for mobile computing research”, *Proceedings of MDC Workshop*, 2012.

[17] Nathan Eagle, Alex (Sandy) Pentland and David Lazer, “Inferring friendship network structure by using mobile phone data”, *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 106, No. 36, 15274–15278, 2009.

[18] Dai Griffiths, Stephen Heppell, Richard Millwood and Greta Mladenova, “Translating software: What it means and what it costs for small cultures and large cultures”, *Computers and Education*, Vol. 22, No. 1/2, pp. 9-17, 1984.

[19] Alvin Yeo, “Software Internationalisation and localization”, *IEEE Proceedings of Sixth Australian Conference on Source*, 1996.

[20] R S Vignesh RAJ, Babak Khazaei, Ashik Ali, “Attitude of Tamil speakers in Tamil Nadu towards computing in Tamil”, *14th International Tamil Internet Conference*, 2015.

[21] R S Vignesh RAJ, Babak Khazaei, Ashik Ali, “A study of Tamil transliteration and the choice of Roman script for Tamil input”, *14th International Tamil Internet Conference*, 2015.

[22] http://ec.europa.eu/justice/policies/privacy/docs/studies/final_report_india_en.pdf accessed on 03, July 2015.

[23] http://mic.com/articles/93438/how-germany-is-using-big-data-to-win-the-world-cup accessed on 02 July 2015.

[24] http://www.tnidb.tn.gov.in/ accessed on 03rd July 2015

[25] Kumar Indrajeet, “IT initiatives in Tamil Nadu”, *egov*, 2013.

[26] R. Shriram and Vijayan Sugumaran, “Cross Lingual Information Retrieval Using Data Mining Methods”, *Americas Conference on Information Systems*, 2009.

[27] K. Jothy and S. Kalaiselvi, “Patterns of internal migration: An analysis using census data of Tamil Nadu” *International Journal of Current Research*; Vol. 3, No. 11; pp. 89-96, 2011.

[28] Jian Yang, Yuanyuan Pu, Hong Wei and Zhengpeng Zhao, “Acoustic models adaptions in large vocabulary continous Mandarin speech recognition non- native speakers”, *Proceedings of 7th International Conference on Signal Processing*, Vol. 1, pp. 687-690, 2004.

[29] Basem H. A. Ahmed and Tien-Ping Tan, “Automatic speech recognition of code switching using 1-Best rescoring”, *International Conference on Asian Language Processing*, pp- 137-140, 2012.

[30] Tzu-Chuan Chou, Jau-Rong Chen and Shan L Pan, “The impacts of social capital on information technology outsourcing decisions: A Case study of a Taiwanese high Tech firm”, *International Journal of Information Management*, Vol. 26, No. 3, pp. 249-256, 2006.

[31] Zhirong Wang, T. Schultz and Alex Waibel, “Comparison of acoustic model adaptation techniques on non- native speech”, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 1, pp. I-540-I-543, 2003.

[32] Dinesh Mavaluru and R. Shriram, “Telugu English cross language information retrieval: A case study”; *International Journal of research in Advanced technology in Engineering*, Vol. 1; Special issue, pp. 78-83, 2013.

[33] Muni Kumar and R. Manjula, “Role of Big data analytics in rural health care - A step towards svasth bharath”, *International Journal of Computer Science and Information Technologies*, Vol. 5, No. 6, pp. 7172-7178, 2014.

[34] http://rajbhasha.nic.in/UI/pagecontent.aspx?pc=Mzk= accessed on July 05, 2015.

[35] She-Sen Guo, “Learning from software localization”, *British Journal of Educational Technology*, Vol. 34, No. 3, pp. 372-374, 2003.