Need of Separate metrics for Tamil Software Development

R. SETHURAMAN

Research Scholar, Department of Banking Technology, Pondicherry, India

sethuvandan@yahoo.co.in

V. PRASANNA VENKATESAN

Associate Professor, Department of Banking Technology, Pondicherry, India

prasanna_v@yahoo.com

ABSTRACT

Modern world introduces many electronic gadgets including computers with multilingual features to reach the international markets. Multilingual or Language other than English based Special software is required to the user to work on their own language comfortably. Tamil software development is progressive both in India and Abroad. Tamil character representation in computer application has more complex than English character. Software metrics are used to measure the software. The present metrics does not measure the language based features available in the Tamil software or multilingual software. To know the importance of special metrics to measure the Tamil software (multilingual) an experiment is conducted with sample data in Tamil and English.

Keywords

Internationalization, Multilingual software, Metrics

1. INTRODUCTION

The software in the world market is almost in a particular language like English. The efforts are taken to develop software to different users with different languages. Internationalization of the software is to develop the software with multilingual features to meet more than one locale. Almost all the leading software company are developing their software with multilingual features. The software with multilingual features has the provision to select the languages by the user dynamically. The software development in India has been carried out primarily through the English language There are many works going on to implement Tamil in software both in India and Abroad. The software in language like Tamil has to be measured. The Software metrics are used to measure the software. The problem in applying the existing metrics is experimentally presented here.

2. INTERNATIONALIZATION OF SOFTWARE

The software developed to international market must have special features for internationalization. Internationalization of a software products deal with developing the products to meet more than one locales.[7] The metrics used to measure software product with internationalization also focuses on process, project and product issues. The best method of reaching the locale is to develop the software to support multilingual features. The new coding system Unicode is introduced to maintain an international standard and to define characters from an international perspective. The Unicode system helps to develop software to reach international environment with capability of allowing more than one language user interface. All these efforts are to reach the locales by developing the software in their respective languages.

3. TAMIL LANGUAGE IN SOFTWARE DEVELOPMENT

In early stages software developments are only possible to those who know knowledge of English. The user of the software must be familiar with English. The developers are interested to make their products to reach all users in the world. But some software developer developed the software to select the language, which is listed by the developer, at the time of installation. The user operating environment is either one user with many languages or one application with many users on their own languages. This may be possible by developing the software with multilingual features. Multilingual software is software which allows the users to select their languages dynamically as and when they required. Almost all the leading software company are developing multilingual software. The multilingual feature in software development is one of the features expected by the user to use the computers or any electronic gadgets more comfortably. Developer of the multilingual software will follow one of the models, wrapper model, monolithic model and multilingual library model to develop software.[6] Multilingual software are categorised as multilingual content creation, multilingual processing like spell check and multilingual software development tools. The software development in India has been carried out primarily through the English language. Now there are few software development tools available to develop an application that incorporates little or no English in its user interface. There are many works going on to implement Tamil in software industries. Few of them are listed here as follows.

The department of Computer Science, Pondicherry University introduced the PONN project which implemented Tamil in computer in the year 1995.[5] It has lot of enhancement like WINKURAL -In 1997 a Tamil computing aimed in porting the DOS based kural to windows, PONN – A multilingual system was developed to provide a working environment for people to work in the language of their choice, PONN ANJAL – A multilingual mailing system, IPONN - The system is aimed to add networking features to PONN system and a multilingual SMS. The Indian Institute of Technology, situated in Chennai, undertakes various projects in multilingual systems. It provides a separate site named acharya.gen.in and works on various multilingual applications which also support Tamil language[9]. Some of the software developed in this institute are Multilingual text and document preparation packages, conversion utilities to support Indian languages, command processor with Indian language and web application supporting online databases in Indian languages. One of the leading software developments organization, Microsoft is now focusing on developing their applications to work in different regional languages including Tamil.[4] The software companies mainly like Yahoo, Google, etc are developing their products in multilingual environment for user interface.

4. SOFTWARE METRICS

The entities are measured by quantifying their attributes or characters. The software also measured in terms of software attributes. Software metrics are intended to measure software quality characteristics quantitatively. According to the IEEE Standard Glossary of Software Engineering Terms, metric is defined as "a quantitative measure of the degree to which a system, component, or process possesses a given attribute."[2] The LOC (Line Of Codes) is used to measure the production ratio of the programmers at the first time by Wolverton in 1974 and improved metrics as person/month. Later, in the requirement specification, Albrecht introduced function points in the year 1979 to measure the software. In 1994, Chidamber Kemerer gave C&K metrics particularly for Object-oriented program. In 2001, metrics for software components was proposed by Victor and Daily, which is to estimate the development time of the modules. Vector Size Measure (VSM) was used for metrics on software scale, software classification which was proposed by Hastings and Sajeev in 2003.[8][3] The software metrics are classified as process metrics, project metrics

and product metrics.[1] Process metrics is meant for the improvement of the process and the prediction for the future process. This metrics includes maturity, management, life cycle, product ratio, defect ratio, etc... Project metrics is to avoid the risks and to optimize the development plans. It includes cost, workload, status, production power, risk, the degree of satisfaction from clients, etc.. Product metrics is to calculate the quality of the product and it includes the reliability, maintainability, product scale, software complexity, portability, documents, etc.

The software development in Tamil requires additional efforts to develop software. The existing metrics can be used to measure the Tamil software. The language based features available in the Tamil software or multilingual software are not considered in the existing metrics. Language like Tamil characters representation is different from English characters. The processing of Tamil characters requires additional efforts to software developer. In this regard, an experiment is conducted to process data in Tamil words and English words for sorting and details of the experiment is given in next section..

5. EXPERIMENTATION

To study the requirements of special metrics for Tamil software two sets of data, one set with list of Tamil words and another set with list of English words are considered for process. But the representation of data on the above two cases are not same. In English each character is represented easily by assigning unique code and in Tamil each character is represented with little complexity by more than one Unicode[9]. The processing of the above two types of data are not same. A common task is considered to operate on these two sets of data. Here sort task is the one selected and operated on the two sets of data and their respective execution time is obtained for multiple runs. Each runs the length of the string that is the number of characters and the quantity of the strings are same in both the sample sets. The strings having length of ten, twenty, thirty, forty and fifty are considered to study the performance of the implementation of sort. The quantity of the strings in each data set is in thousands and multiples of thousands up to nine thousands. The execution time of the Tamil strings and English strings are obtained and tabulated in Table 1.

Table 1- Execution time of sort process in multiple runs

width (no. of char)X →	10	20	30	40	50	10	20	30	40	50
Quantity of strings↓	Execution time of sort process for English characters in milliseconds (Y)					Execution time of sort process for Tamil characters in milliseconds(Z)				
1000	15	16	20	24	31	4424	8315	12462	16372	20411
2000	44	53	74	90	98	19697	33337	49509	65501	83226
3000	93	121	160	206	231	40102	75457	112614	150375	188320
4000	167	216	280	361	406	71309	134223	214065	265246	343209
5000	251	332	440	563	652	112535	211902	314096	428190	509245
6000	365	485	629	863	936	161670	304509	456079	597692	755789
7000	505	673	858	1163	1300	220552	417101	611217	835805	1029700
8000	654	919	1126	1529	1674	289292	548656	796520	1056295	1328444
9000	834	1172	1420	1930	2169	364790	693650	1088146	1383699	1711501

The data labelled X in table 1 represents the length of characters and Y and Z are to represent the execution time for sort process in milliseconds for English characters and Tamil characters respectively. The relationships between X and Y is $Y = a1 * X + b1$ and between X and Z is $Z = a2 * X + b2$. The values of parameters a1 and a2 are calculated with help of experiment data and they are listed in the table 2.

Table 2 - The values of a_1 and a_2

Sl.	Quantity of strings	Parameter for English characters(a1)	Parameter for Tamil characters(a2)
1	1000	0.40	400.31
2	2000	1.45	1592.22
3	3000	3.61	3713.54
4	4000	6.23	6748.23
5	5000	10.33	10097.08
6	6000	15.20	14814.19
7	7000	20.81	20369.97
8	8000	26.51	25859.47
9	9000	34.29	33834.63

6. ANALYSIS OF RESULT

The value of a1 and a2 implies that the execution time is not the same. The Fig.1 shows that the two parameters, a1 for English character sets and a2 for Tamil character sets increases directly with the volume of data. There is a steep increase in Tamil character set processing than the English character set processing. These two curves never get chance neither to meet nor to coincide with each other. So the metrics used for software using language like English having single code mapping to each character is not the same for software using the Tamil characters (which requires more than one code to represent a single character). So, it is necessary to have a special metric to measure the Tamil software.
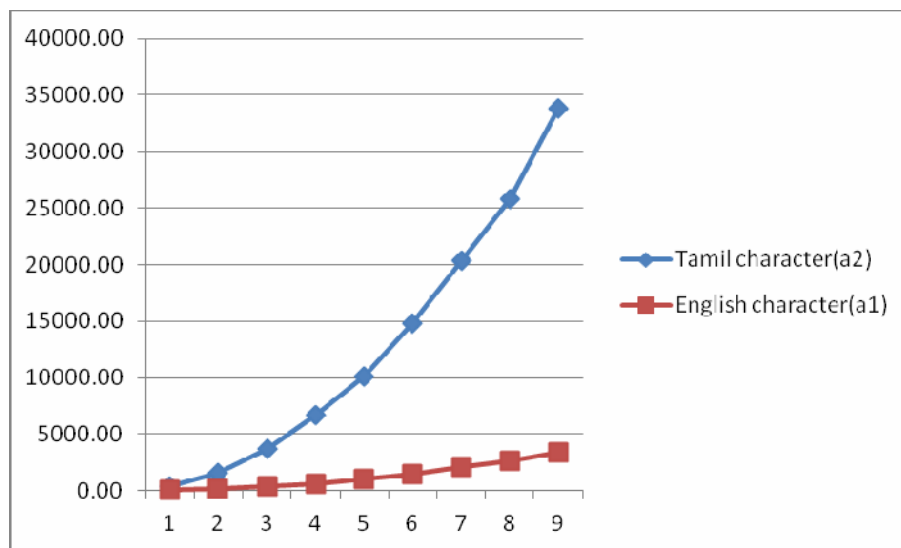

Figure 1 Curves representing two parameters a1 and a2

7. CONCLUSION

The need of software development in multilingual features is important to meet user requirements in the international market. The efforts to implement Tamil language in software development are listed. The processing of the Tamil characters and English characters are not same. These two sets of data (English words and Tamil words) are applied for same process (sort) for multiple runs. The result leads to work for future to design metrics for Tamil software development or multilingual software development.

References

[1] Gurdev Singh, Dilbag singh, Vikram Singh, A Study of Software Metrics, IJCEM International Journal of Computational Engineering & Management, Vol. 11, January 2011.

[2] IEEE Software Engineering Standards, Standard 610.12-1990, pp. 47–48.

[3] Norman E. Fenton , Martin Neil, Software metrics: successes, failures and new directions, The Journal of Systems and Software 47 (1999).

[4] Peter Constable, Understanding Multilingual Software on MS Windows, November 01, 2000 Rev:16, Copyright © 2000 Peter Constable and SIL International

[5] Prasanna Venkatesan V(1995)., PONN : A Multilingual Computer system, M.Tech Project report,   Department of Computer Science, Pondicherry University

[6] Prasanna Venkatesan V and S.Kuppuswami, Formal Models of Multilingual Software, International Journal o f Computer Science and System Analysis, Vol. 1(2), 65-76, 2007, ISSN 0973-7448.

[7] Sonia Dube, Yatrik Patel, T A V Murthy, Globalization of Software Applications Using UNICODE Based Multilingual Approach, 3rd International CALIBER - 2005, Cochin, 2-4 February, 2005, © INFLIBNET Centre, Ahmedabad

[8]. Tu Honglei, Sun Wei, Zhang Yanan , The Research on Software Metrics and Software Complexity Metrics, 2009 International Forum on Computer Science-Technology and Applications, IEEE Computer Society

[9] http://www.acharya.gen.in:8080/tamil/tamilcomp.php