Development of a Language Translator from English to Waray

Khenilyn P. Lewis
Cavite State University
Technological University of the Philippines-Manila, Philippines

ABSTRACT

The Philippines is a home of different languages. Languages are vital weapons of communicating with others and expressing points in the society more importantly in Education. Based from Department of Education, there are 12 major languages as part of K+12 Education System. Among these is Waray, language spoken by people from Samar, Northern Samar, Eastern Samar, Biliran, and Leyte. According to the Department of Education, the implementation of K+12 Education System in the country is the used of mother tongue as medium of instructions from Kindergarten to Grade 3. The development of the software is to translate instructional materials written in English to Waray language. The software could accept word, sentence, phrase, and document file with .doc and .docx extensions written in English to be translated in Waray language. In addition, the capabilities to add new words and rules, update of dictionary, and search for word and rule are part of the translator. The development of the translator used Rule-Based Algorithm that includes parsing, searching, and matching of words inside the database. The translator can translate document file and words from English to Waray Language. In addition, existing dictionary and rule can be modified and added inside the database. Test and evaluation results validated that the prototype could be an alternative way of translating instructional materials written in English to Waray language.

General Terms

Natural Language Processing

Keywords

Rule-Based Algorithm, Language Translator, Waray

1. INTRODUCTION

Education is the key to success. Students at an early age should be able to understand the process of learning that serves as their foundation of knowledge. The early stage of education also helps students to enhance their emotional, physical, social and intellectual capacities. Language is one of the important ingredients to make the learning process more comprehensible. The use of books and learning materials are being utilized in teaching students. In the Philippines, textbooks are written in languages such as Tagalog and English. Also, based on DepEd, the types of learning materials available are story books in big and small book formats; flash/activity cards using letters and numbers; basic sight words or grade level words and picture dictionary; and picture chart for oral literacy.

According from the Official Gazette of the Department of Education, "The K +12 Basic Education Program", introduced twelve (12) Mother Tongue (MT) languages for SY 2012-2013: Bahasa Sug (Tausug), Bikol, Cebuano, Chabacano, Hiligaynon, Iloko, Kapampangan, Maguindanaoan, Meranao, Pangasinense, Tagalog, and Waray. These local languages spoken in different regions of the country are used as medium of instruction from kindergarten to Grade 3. This crucial shift, known as "Mother Tongue-Based Multi-Lingual Education" (MTB-MLE), is part of the K+12 basic education reform program and other local languages will be added in succeeding school years.

The Philippines is home of more or less 175 languages, but some of them have just few speakers left while others are already considered extinct because there are no more people at all who use in them. According to Agence (2012), up to 50 of the country's minor languages could be lost within 20 years.

The DepEd also stated that the "Local and international studies have shown that using the language used at home (mother tongue) inside the classroom during the learners' early years of schooling produces better and faster learners who can easily adapt to learn a second (Filipino) and third (English) language." DepEd national coordinator for MTB-MLE, Rosalina Villaneza stressed that studies show that using the child's mother tongue as language of literacy "bolsters comprehension and critical thinking skills of children and facilitates acquisition of a second language such as English or Filipino.

DepEd directive aimed at enhancing language, cognitive and academic development of the child even as it also "seeks to create social and cultural awareness among the children by promoting their local heritage, language and culture." The use of mother tongue as the medium of instruction will require competent educators especially those teaching Science and Math subjects.

Furthermore, there are available software translators in the Internet. One of the most popular is the Google Translate that allow users to translate the source/ detect language to the target language. This translator comprises different languages all over the globe. In addition, dictionary for Native Philippine Languages called Waray English Dictionary is also accessible in the Web. This online dictionary accepts either Waray or English text and translates it word for word. Most language translators though translate languages does not focuses in Native Language and mostly connected in the Internet. In addition, software translator from English to Native Language nonetheless available has limited capacities.

The study is to developed prototype that translates English language to Waray and includes dictionary, sentence translation, and accept document file.

2. METHODOLOGY
2.1.1 Project Design

The translator offers a way of converting English as source language to Waray dialect as target language.
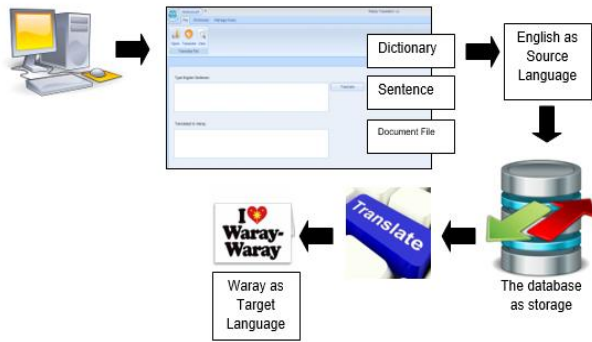
Fig 1: Block diagram of the language translator

Figure 1 shows the block diagram of the language translator. Microsoft SQL Server and executable file is installed in a computer. The prototype contains translation for dictionary, sentence(s), and document file. Document file to be translated should have file extensions of .doc or .docx only. The source language which is English will be parsed and matched. Inside the database are the fields for words, part of speech, linking verb, and rules that are ready for retrieval. The source language will be translated to Waray as target language.
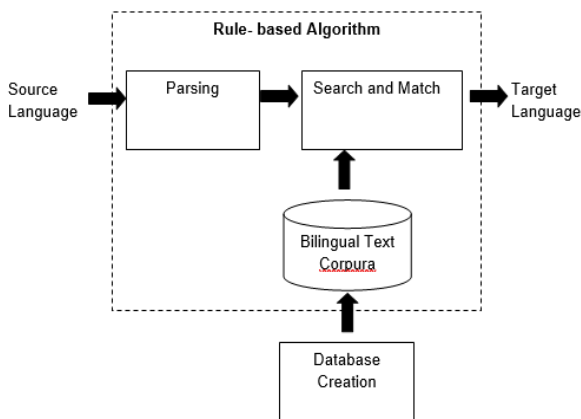

Fig 2: Conceptual design of the language translator

As shown in Figure 2, database creation used to provide bilingual text corpura. Source language parsed and identified match after search. Matched found translated based on rules whereas direct translation take place if not. The bilingual text corpura composed of source and target language, rules and linking verbs.

2.1.2 Database Design

Table 1. List of Tables for WarayDB Database

Table Name	Description
LinkingVerbs	This table list linking English linking verbs.
tblWords	This table composed of English and Waray words with Part_of_Speech, RuleValue, Meaning, and Marker
PartofSpeech	This table list of parts of speech
Rules	This table list of rules
Users	This table list of users

Table 1 shows the list of tables inside the WarayDB database. These are LinkingVerbs, tblWords, PartofSpeech, Rules, and Users. Further discussions are shown in the tables below.

Table 2. The Users Table of WarayDB Database

Field Name	Data Type	Description
UserID	int	Shows user ID and cannot be null
LoginID	varchar(20)	Accepts log in ID
LoginPW	varchar(20)	Accepts log in password
AcctType	varchar(50)	Account Type according to level of access
FullName	varchar(80)	Account full name
AcctStatus	varchar(1)	Account Status
CreateDate	datetime	Create date of account
CreatedBy	varchar(20)	Account that created other account
LastUpdate	datetime	Last update of the account
UpdateBy	varchar(20)	Holds who update the account

Table 2 shows the users table of the WarayDB database composed of UserID, LoginID, LoginPW, AcctType, FullName, AcctStatus, CreateDate, CreatedBy, LastUpdate and UpdateBy fields. The UserID field is int in data type, auto increment and cannot be null. The LoginID and LoginPW fields with data type varchar(20) holds the log-in information of the account. The FullName field with varchar (80) indicates the full name of the account user. For AcctStatus with varchar(1) data type specifies if the account is active or not. The fields CreateDate, LastUpdate, CreatedBy, and UpdateBy are for update of an account.

Table 3. The LinkingVerb table of WarayDB database

Field Name	Data Type	Description
LvID	int	Accepts linking verb ID`
LvName	varchar(150)	Accepts linking verb name

Table 3 shows the LinkingVerb table of WarayDB database. This includes LvID with data type Int and LvName with data type varchar (150). The table list all English linking verbs ready for retrieval.

Table 4. The tblWords table of WarayDB database

Field Name	Data Type	Description
ID	int	Shows ID number and cannot be null
Waray	nvarchar(255)	Holds Waray words
English	nvarchar(255)	Holds English words
Part_of_Speech	nvarchar(255)	Holds part of speech
Rulevalue	nvarchar(255)	Holds Rule Value
Meaning	nvarchar(255)	Holds the meaning of English and Waray word
Markers_RuleNum	nvarchar(255)	Holds marker value

Table 4 shows the tblWords table from WarayDB composed of ID, Waray, English, Part_of_Speech, RuleValue, Meaning, and Markers_RuleNum. ID field has data type int and cannot be null. Waray field holds word of target language with nvarchar (255) data type. English field hold word as source language with nvarchar (255). Part_of_Speech, Rulevalue and Markers_RuleNum fields hold the equivalent value of the source and target language with data type of nvarchar (255). Moreover, meaning field with data type nvarchar (255) of the

table holds the description of the English and Waray words. This table serves as the dictionary of the database

Table 5. The PartofSpeech Table of WarayDB Database

Field Name	Data Type	Description
ID	int	Shows ID number and cannot be null
PartofSpeech	varchar(50)	Holds identified part of speech
RuleValue	varchar(50)	Holds rule value

Table 5 shows the PartofSpeech table of WarayDB database composed of ID, PartofSpeech, and RuleValue. The PartofSpeech and RuleValue with data type varchar(50) holds marker for translating based on rules. While ID with data type int serves as the control number and cannot be null.

2.1.3 Link Grammar Algorithm

Each word with corresponding part of speech assigned a rule value with the following equivalent:

Table 6. Part of speech and Rule value

Part of Speech	Rule Value
Noun	N
Pronoun	PRO
Adjective	ADJ
Verb	V
Preposition	PREP
Adverb	ADV
Conjunction	CONJ
Article	A
Contraction	CONT
Interjection	INTJ
Markers	M

After assigning the rule value:
1. Provide noun markers that include an, it, ha, hi, hin, han, and hit.
2. The program randomly selects from noun markers.
3. Words with more than one equivalent get the first value found in the database.

2.1.4 Re-writing Document File as Output

The program includes:
1. Saving word and rule matched in a data table in the memory.
2. Automatically save the translated file in an output folder (C:\WarayProj\WarayDoc\Output\) without deleting the original file. File will have the same name as the original file.

Project Development

The conceptual and database design guided the development process and program flow chart of the language translator as shown in the Figures 3 to 5.
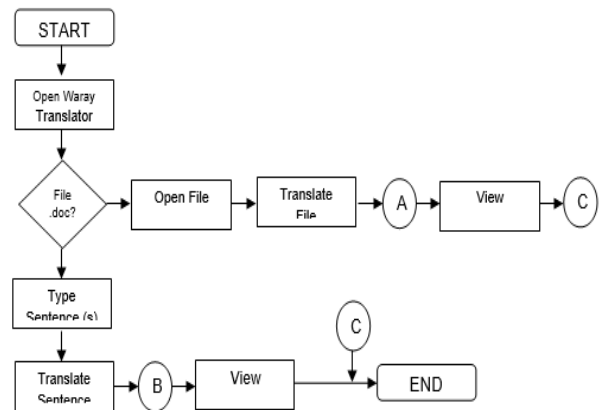

Fig 3: Program Flowchart of the Language Translator

Figure 3 illustrates the program flowchart of language translator from English to Waray. It covers the translation of document file with file extensions .doc or .docx.
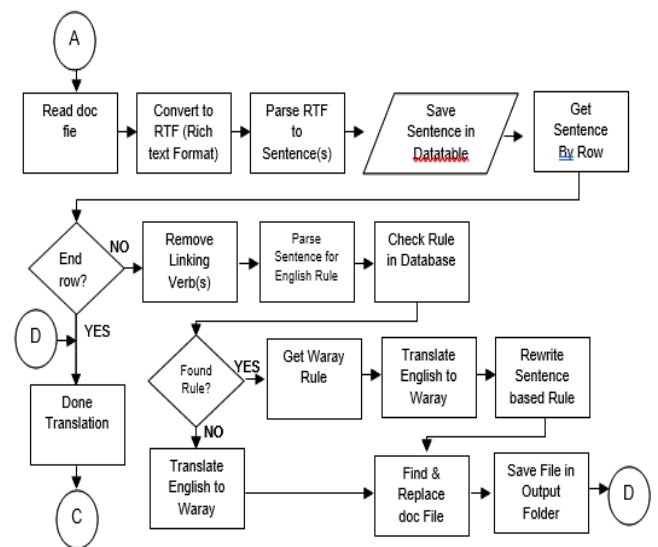

Fig 4: Translate File Program Flowchart

Figure 4 shows the program flow chart for translating file document. The opened file is read and converted to Rich Text Format (.rtf). Rich Text format allows text be read and figure/ images retained. The parsed sentence will be saved inside the database table. Since, Waray grammars do not have equivalent linking verb/s then it is removed in the parsed sentence. Parsed sentence will be divided according to English rules counterpart. If there are matched rules inside the database, it will be re-written and replace document file uploaded in the translator. If there are no matched found inside the database, word for word translation takes place. Replacing the document file after translation do not delete or re- write original file. Furthermore, translated file is saved inside the output folder of the computer's drive.
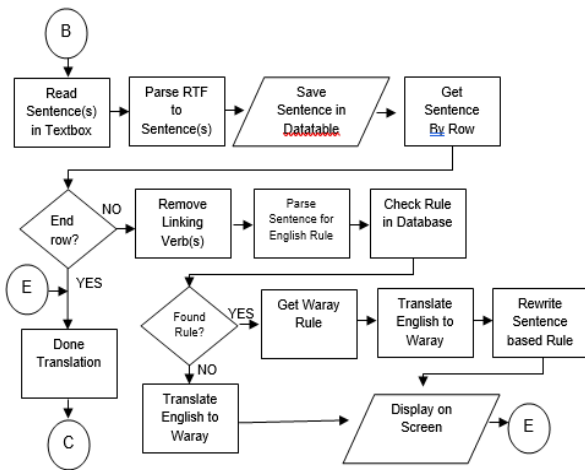
Fig 5. Program flowchart for sentence translation

Sentence translation in Figure 5 follows the same rule as the document file. However, there is no uploading of file and input and output display will be found in text fields. Finally, when word entered found matched, Waray word equivalent, meaning and part of speech will be displayed in the screen.

For the dictionary, entering English word that matched inside the database will provide the corresponding Waray word, meaning and part of speech.

2.1.5. Operation and Testing Procedure
2.1.5.1. Testing Procedure
The main purpose of this stage is to test software's functionality and accuracy. It was tested by five (5) information technology professionals and five (5) teachers from the Department of Education who speak Waray language.

The following were undertaken during the testing:

Table 6: Testing procedures

Module/ Component	Test Conducted
Translate Document	1. Opened the translator .exe file. 2. Clicked Open button under the file menu tab. 3. Selected file to open in any location in the computer. Learning materials from DepEd was first uploaded to the system. 4. Clicked Translate button under the file menu tab. 5. Clicked View button under the file menu tab. 6. Opened another document file to translate.
Dictionary	1. Clicked Dictionary menu tab. 2. Clicked Search button. 3. Entered English word and clicked Search button.
Sentence Translation	1. Entered sentence in the first text field. 2. Clicked Translate button. 3. Clicked Clear button to reset fields.
Account	1. Created account by entering log-in

	ID, password, account type and full name. 2. Clicked the Save button. 3. Clicked New Word button to dictionary menu tab. 4. Verified user name and password. 5. Supplied fields for Waray word, English word, part of speech and meaning. 6. Clicked the Save button and Clear button to reset. 7. Clicked find rules under Manage Rules menu tab. 8. Entered English rule, Waray Rule and Rule Description. 9. Clicked the Save button and Clear button to reset fields.

2.1.5.2. Evaluation Procedure
To determine the performance of the language translator, ISO 9126 evaluation criteria used to evaluate functionality, usability, reliability, efficiency, portability, and maintainability.

The following were undertaken during the evaluation:
1. The translator presented to the evaluator-respondents consisting of ten (10) information technology professionals and twenty (20) Waray speakers.

2. The evaluator-respondents explored, navigated and used the translator for evaluation;

3. Requested them to rate the translator using Likert Scale shown in Table 7 and the prepared evaluation instrument;

4. Collected and tabulated the data and computed the mean for each criterion and the overall mean of all the criteria; and

5. Results were interpreted for the equivalent descriptive rating using Table 8.

Table 7. Likert Scale

Numerical Scale	Interpretation
5	Excellent
4	Very Good
3	Good
2	Fair
1	Poor

Table 8. Range of Mean Values Scale and Interpretation

Numerical Scale	Interpretation
4.51 – 5.00	Excellent
3.51 – 4.50	Very Good
2.51 – 3.50	Good
1.51 – 2.50	Fair
1.00 – 1.50	Poor

3. RESULTS AND DISCUSSIONS
3.1. Project Description

The prototype is capable of translating word for dictionary, sentence, and document file with file extensions .doc or .docx from English to Waray. Document file translation uploads file inside the system. The uploaded file will be converted to .rtf format to segregate image and text. Text is parsed, and terminators served as the indicator. Matched words and rules will be used for translation. On the other hand, no rules and word matched will create word for word translation. Translated file saved in the output folder. Words inside the dictionary matched and displayed part of speech, Waray word equivalent, and meaning. Moreover, parsing and matching be applied for sentence translation. Additional feature of the translator is the capability of admin to add and manage word and rules inside the database. The translator was created using C# programming language, Microsoft SQL Server as database, and Telerik for interface design.

3.1.1. Project Capabilities and Limitations

The software could translate word, sentence, and documents file from English to Waray. It is capable of adding and editing word and rules inside the database. Further, the user account is used to manage update inside the database. The translator converts document file to Rich Text Format. Converted .doc to .rtf will be parsed and matched using the database. The prototype could re- write document file content to Waray language without deleting original file. Output is saved in an output folder in the computer's drive.

As limitations, the words and rules used in translator came from *Komisyon sa Wikang Filipino*- Manila and *Panrehiyong Sentro ng Wikang Filipino (Region 8)*. Hence, output of the translator is limited and varies only to those cited sources. It is not connected in the Internet or any dictionary except to the sources mentioned above. Document file that can be uploaded is with .doc and .docx file extensions only. The software is not capable of identifying wrong entries of the user.

3.2. Project Evaluation
3.2.1. Test Results

The results during the test procedure show that:
There are English rule that does not have match in the database and produced word for word translation. There are words not translated because it does not exist in the database. Special characters and text cases were not read. Also, the respondent did not able to translate more than one sentence in the field. In addition, to be able to add and edit word and rule, the respondent must have a valid account. The test result shows that because of limited sources, there were words and rules that were not translated. Since word for word translation occurred the context of the sentence also changed. The testing result was the reference for the software's improvement. In addition, translation time was considered during the testing as shown in the table below.

Table 9. Test conducted for Translation Time

Computer Specifications	Number of words in the file (from DepED)	Uploading Time in the translator	Translation Time
Windows 7 Starter 32 bit Intel Atom CPU @ 1.66 GHz 2.00GB RAM	842	1 minute 12 seconds	12 minutes 24 seconds
Windows 7 Ultimate 32 bit AMD E-450 APU @ 1.65 GHz 2.00GB RAM	842	24.57 seconds	7 minutes 24 seconds
Windows 7 Home Edition 64 bit Intel Core ™ i5-3210M CPU @ 2.50 GHz 4.00 GB	842	10.12 seconds	3 minutes 05 seconds

Table 9 shows the test conducted and result of translation time of the prototype. Three different computer specifications were used. The result of the table shows that translation time defends on the number of words in file and computer specifications. Furthermore, table 8 implies that the higher the computer specifications, the faster translation time is.

3.2.2. Evaluation Results

The software was evaluated according to functionality, usability, reliability, efficiency, portability, and maintainability. It was evaluated by 30 respondents with 10 information technology professionals and 20 Waray speakers. Table below shows the evaluation result of each criteria and overall result.

The result of evaluation shows the overall mean rating for the language translator. It was rated 4.47 by the respondents and found very good. Furthermore, functionality is 4.22, reliability is 4.29, efficiency is 4.45, and portability is 4.33 and all found as very good in performance. On the other hand, usability is 4.66 and maintainability is 4.60 and both found excellent in performance. The result of evaluation discusses that usability got the highest mean and functionality got the lowest mean but still very good as construed. The result implies that the prototype could be an alternative way of converting English text in Waray language.

4. SUMMARY, CONCLUSION AND RECOMMENDATIONS
4.1 Summary of Findings

Based on tests and evaluation conducted on the performance capability of the language translator, the following were the findings of the study. The software is capable of translating English words as source language to Waray as target language. It can provide dictionary that displayed the meaning, English and Waray words, and part of speech. It could translate sentence and document file with .doc and .docx extensions. It was created using C# as programming language, Microsoft SQL Server as database, and Telerik as interface design. Rule based algorithm was used to get matches in the database. The software was tested by five (5) Department of Education teachers that speak Waray and five (5) information technology professionals for improvements. It was evaluated to be "Very Good" with an overall mean of 4.47 signifying that the prototype could be alternative way of translating English to Waray language.

4.2 Conclusions

In consideration of the objectives of the study and the results of testing and evaluation carried out, the following conclusions were derived. The language translator was developed such that it provided dictionary for English to Waray, could translate sentence and word document with .doc and .docx file extensions. The software was created using C# programming language, Microsoft SQL Server as database and Telerik to interface design. Test results validated that the

software could translate words, sentence and document file from English to Waray language. It could execute to Windows 7 platforms and the higher the computer specification, the faster the translation time. The performance of the software was rated "Very Good" by the respondents which proved that it can be an alternative way of converting English to Waray language.

4.3. Recommendations

Implementing the prototype highly recommended since it can be an alternative in translating English to Waray language. However, several recommendations to further enhancement of the project are hereby endorsed including the Upgrade of the translation algorithm, providing markers when adding rules and could be an online translator.

5. ACKNOWLEDGEMENT

6. REFERENCES

[1] Guidelines on the Implementation of the Mother Tongue-Based- Multilingual Education. 2012. Department of Education. Retrieved last August 5, 2012 from www.deped.gov.ph.

[2] Agence, France- Presse. 2012. Along fight begins to save Philippine languages. Retrieved last August 5, 2012 from http://www.rappler.com/nation/4537-a-long-fight-begins-to-save-philippine-languages

[3] Discussion Paper on the Enhanced K+12 Basic Education Program. 2010. DepEd Discussion Paper.

[4] Clarke, Derek. 1996. Software Evaluation Criteria- The Users. Retrieved last August 2, 2012 from http://content.alterra.wur.nl/Internet/webdocs/ilri-publicaties/special reports/Srep2-h5.pdf

[5] Schneidewind, Norman F. 2003. Software Reliability Measurement, or Life Cycle Core Knowledge Requirements for Software Reliability Measurement. Volume23 Number 2, June 2003. Reprinted from Reliability Review, The R & M Engineering Journal.

[6] Mathew, Tom V. 2005. Genetic Algorithm. Retrieved last August 3, 2012 from http://www.civil.iitb.ac.in/tvm/2701_dga/dga/dga.html

[7] Parikh, Prashant. 2001. The Use of Language. Distributed for Center for the Study of Language and Information. CSLI Publications.

[8] Sugbo, Victor N. 2003. Language Policy and Literature in the Philippines. University of the Philippines in the Visayas.

[9] Rubrico, Jessie Grace U. 2006. The Languages of the Philippines. Language Links. Org. Retrieved last August 4, 2012 from http://www.languagelinks.org/onlinepapers/fillang.html

[10] The Official Website of Samar Province. Retrieved last August 8, 2012 form http://www.samar.lgu-ph.com.

[11] The Provincial Government Website of Leyte. Retrieved last August 8, 2012 from http://www.leyte.org.ph/about_leyte-gen_info.htm.

[12] Rouse, Margaret. 2005. Expert System. Retrieved last August 3, 2012 from http://searchcio-midmarket.techtarget.com/definition/expert-system

[13] Nueva, Frederico G. Development of Translation Software of Four Philippine Major Dialects. Master's thesis. Technological University of the Philippines, Manila. May 2004.

[14] Prasad, Rajkishore and Sinha, A.K. 2003. Using ESTA to Develop Expert System for the Natural Resource Management. 6th Annual International Conference and Exhibition, New Delhi, India.

[15] Basnette, Susan. 2002. Translation Studies- Third Edition. Routledge Taylor and Francis Group London and New York.

[16] Bosco, Gabriela. 2012. "Translation Techniques". Interpro translation Solutions. Retrieved last August 2, 2012 from http://www.interproinc.com/articles.asp?id=0303

[17] Herther, Nancy K. 2012. Language Translation in the Internet Age 'My Hovercraft Is Full of Eels'. Information Today, Inc. Vol. 20, No. 1.

[18] Top Ten Reviews. Retrieved last August 8, 2012 from http://translation-software-review.toptenreviews.com.

[19] Zogheib, Ali. 2007. Genetic Algorithm-based Multi-Word Automatic Language Translation. IT-Universitetet i Goteborg - Department of Applied Information Technology. ISBN 978-83-60434-59-8, pages 751–760.

[20] PhP: Hypertext Preprocessor. Retrieved last August 3, 2012 from http://www.webopedia.com/TERM/P/PHP.html

[21] Rouse, Margaret. 2007. Visual Basic .NET (VB.NET or VB .NET). Retrieved last August 3, 2012 from http://searchwindevelopment.techtarget.com/definition/Visual-Basic-NET

[22] Gaurav, Khanna. 2010. Benefits of VB .NET. Retrieved last August 3, 2012 from http://social.msdn.microsoft.com/Forums/en/netfxnetcom/thread/a5c456c8-e432-4b46-b963-3b4559e0a81d

[23] Hanumesh, V. J. and Seema Acharya. (2005). Relational Database Management System. Education and Research Department Inforsys Technologies Ltd. No. 350, Hebbal Electronics City Hootagalli, Mysore.

[24] The Tech Terms Computer Dictionary. Retrieved last August 5, 2012 from http://www.techterms.com

[25] Rouse, Margaret. (2006). MySQL. Retrieved last August 3, 2012 from http://searchenterpriselinux.techtarget.com/definition/MySQL

[26] Hubpages. How Google Translate Works. Retrieved last February 1, 2013 from http://animshare.hubpages.com/hub/How-Google-Translate-Works

[27] Microsoft Website. Retrieved last August 5, 2012 from http://office.microsoft.com/en-us/products

[28] ISO 9126 Software Quality Characteristics. Retrieved last August 3, 2012 from http://www.sqa.net/iso9126.html

[29] Uebersax, John S. (2006). Likert Scales: Dispelling the Confusion. Retrieved last August 8, 2012 from http://www.john-uebersax.com/stat/likert.htm.

[30] Education. US Library of Congress. Retrieved last August 3, 2012 from http://countrystudies.us/philippines/53.htm.

[31] The K to 12 Basic Education Program. Official Gazette of the Republic of the Philippines. Retrieved last August 2, 2012 from http://www.gov.ph/k-12.