On Demand Web Page Translation -BEES in action-

B. Hettige[1], A. S. Karunananda[2]
[1]Department of Statistics and Computer Science, Faculty of Applied Science,
University of Sri Jayewardenepura, Sri Lanka.
[2]Faculty of Information Technology, University of Moratuwa, Sri Lanka.
budditha@yahoo.com[1] , asoka@itfac.mrt.ac.lk[2]

Abstract **–Web-enabled technologies including www, email are widely use and have become popular communication media in the modern world. However, many of these services are available only through the English language. This is a problem faced by millions of internet users who are not fluent in English. Therefore, many countries address this issue by using Machine Translation technologies to translate these English based web resources into their local languages. This paper reports the design and implementation of the English to Sinhala Machine Translation system (BEES) that has been developed to translate an English web page in to Sinhala through the concept of Varanagema (conjugation) in Sinhala Language. In addition, it uses a context-based approach to semantic handling. The design, implementation and major translation issues have been presented in the paper.**

Introduction

World Wide Web (www) is the most widely used and popular communication media in the modern world. From a technical viewpoint, it is a system of interlinked hypertext documents accessed via the Internet [19]. There are so many services and facilities available on the internet such as web, email, chat, forums, Facebook etc. It should be noted that, many of these services are available only in the English language. This is a problem for millions of internet users who are not fluent in English. The obvious solution for this issue is the use of modern computing technologies to translate English to local languages. This is call machine translation (MT). The machine translation is a sub field of Natural Language Processing (NLP), which is one of the most achieved areas in Artificial Intelligence (AI).

In Sri Lanka, Sinhala language is spoken by about 16 million people. Sinhala is one of the constitutionally-recognized official languages in Sri Lanka, along with Tamil. However, 80 % of Sinhala speaking people do not have the ability to read and write in English well. Therefore, the development of a English to Sinhala Machine translation system is a highly valuable product for all Sinhala speakers who are not fluent in English language.

Nowadays, thousands of Machine translation systems have been developed for different languages. Among others, Apertium [30], Google Translate [28], Babel Fish [25] and SYSTRAN [22] are well-known machine translation systems in the world. In the region, Anusaaraka [2], AnhalaHindi [4], ManTra [6], AngalaBaratha [5], English to Urdu machine translation system [35] belong to the Indo-Aryan family [19] of machine translation systems. On the other hand, perhaps, EDR [36], the machine translation system by the Japanese is the most completed system so far.

These translation systems use various approaches to machine translation, including, Human-Assisted translation, Rule based translation, Statistical translation and Example-based translation. However, due to various reasons associated with the complexity of languages, Machine Translation has been identified as one of the least achieved areas in computing over the last sixty years. Most of these issues are associated with semantic handling in the machine translation systems.

We have been working on a project to develop an English to Sinhala Machine translation system namely BEES. The BEES is acronym for Bilingual Expert for English to Sinhala. It has been powered by theory of Varanagema (conjugation) in Sinhala language. In this project we have already developed a Sinhala parser [7], intermediate-editor [11], Sinhala morphological analyzer [8], three lexical dictionaries [9] and Transliteration module [10]. Each of these modules and their prototype integrations have been tested through several real world applications namely Human-Assisted machine translation system [11], web-based selected text translation system [12] [13] and context-based machine translation system [14].

This paper reports a new version of the BEES that can translate a given web page into Sinhala. This system works based on the concepts of Varanagema and handles the semantics of the sentence through the context-based approach.

The rest of this paper is organized as follows. Section 2 gives an overview of some existing machine translation systems. Section 3 reports a brief description about previous developments of the BEES. Then section 4 gives the design of the

translation system. Section 5 discusses current issues in the English to Sinhala Machine translation and section 6 shows how the system works for the given web page. Finally, Section 7 concludes the paper with the conclusion and a note on further work.

Brief review of the Mchine Translation

The Machine translation approaches can be classified into three categories, namely, statistical approach, example based approach and rule-based approach [19]. The Statistical approach uses some statistics such as mean, variance on bilingual text corpora to find the most appropriate translation. The Example-based approach is often characterized by its use of a bilingual corpus with parallel texts as its main knowledge base. The rule based approach requires extensive lexicons with morphological, syntactic, and semantic information, and large sets of rules. Therefore, any rule-based machine translation system contains a source language morphological analyzer, a source language parser, translator, target language morphological analyzer, target language parser and several lexicon dictionaries. Further, in relation to English to Sinhala machine translation, the system needs an English dictionary, an English-Sinhala bilingual dictionary and a Sinhala dictionary.

A large number of machine translation systems have been developed under the above three broader headings.

For instance, Apertium [30] is a rule-based MT system that translates related languages. This is an open –source system that can be used to translate any related two languages. This MT engine follows a shallow transfer approach and consists of eight pipelined modules, such as de-formatter, morphological analyzer, part-of-speech (PoS) tagger, lexical transfer module, structural transfer module, morphological generator, post-generator, and re-formatter.

Google Translator [28] translates a section of a text, or a webpage, into another language. It does not always deliver accurate translations and does not apply grammatical rules, since its algorithms are based on statistical analysis rather than traditional rule-based analysis.

Babel Fish [25] is a web-based application developed by AltaVista, which translates text or web pages from one or several languages into another. The translation technology for Babel Fish is provided by systran [22], whose technology also powers the translator at Google and a number of other sites. It can translate among English, simplified Chinese, traditional Chinese, Dutch, French, German, Greek, Italian, Japanese, Korean, Portuguese, Russian, and Spanish. A number of sites have sprung up that use the Babel Fish service to translate back and forth between one or more languages.

The Anusaaraka [2] is a popular machine-aided translation system for Indian languages that makes text in one Indian language accessible to another Indian language. Further, this system uses Paninian Grammar model [1] to its language analysis. The Anusaaraka project has been developed to translate Punjabi, Bengali, Telugu, Kannada and Marathi languages into Hindi. The approach and lexicon is general, but the system has mainly been applied for children's stories.

Angalabharti [5][6] is also a human-aided machine translation system used in India. Since India has many languages, there are a variety of machine translation systems. For example, Angalahindi[5] translates English to Hindi using machine-aided translation methodology. Human-aided machine translation approach is a common feature of most Indian machine translation systems. In addition, these systems also use the concepts of both pre-editing and post-editing as the means of human intervention in the machine translation system.

Electronic Dictionary Research (EDR) [36], by the Japanese is the most successful machine translation system. This system has taken a knowledge-based approach in which the translation process is supported by several dictionaries and a huge corpus. While using the knowledge-based approach, EDR is governed by a process of statistical machine translation. When compared with other machine translation systems, EDR is more than a mere translation system and hence provides lots of related information.

Table 1 shows a comparison of some existing machine translation systems.

System	Language pair	Approach & Type
Anusaaraka	Among Indian languages	Human-assisted, application
Angalabarath	English to Indian languages	Human-assisted, rule-based, application
AngalaHindi	English to Hindi	Machine-aid, rule-based/ example-based, web-based
ManTra	English to Hindi	Human-aided, web based
English to Urdu MT	English to Urdu	Example based, application
Matra	English to Hindi	Human-aided, transfer-based application
Google TR	Several languages	Statistical, web-based
Bable fish	Several	Systran technology,

	languages	web based
Yahoo TR	Several languages	Statistical, web-based
Aprtium	Related languages	Rule-based, application
EDR	English/Japanese	Knowledge based, application

Table 1: Comparison of the MT systems

At present there are many Sinhala language resources available; including Sinhala Unicode [26], some bilingual dictionaries [20][21], Sinhala corpus[29], some transliteration and OCR systems. However, only few researches have been done on machine translation. Vitanage's English to Sinhala translator for weather forecasting domain [17] and Silva and others' Sinhala to English language translator [16] are some prototype projects. In addition, there some attempts have been taken to develop Sinhala to Tamil machine translation [18] and Japanese to Sinhala machine translation [15].

It is evident from the discussion that we have developed a English to Sinhala machine translation system (BEES). This system has also taken the approach of human-assisted translation and it works on the concepts of Varanagema in Sinhala language. This system has been tested through several standard desktop applications and a web application. Following section reports previous development of the BEES.

Previous Development of the BEES

Our English to Sinhala machine translation system has been primarily implemented with the use of SWI-Prolog [23], Java and Prolog Server Pages PSP [24]. The core of our MT system has seven modules, namely; English morphological analyzer, English parser, word level translator, Sinhala morphological analyzer, Sinhala parser, transliteration module and lexical dictionaries. Our project has introduced the first ever parser [7] and morphological analyzer [8] for Sinhala language. Figure 1 shows the basic interface of our stand-alone machine translation system. This first version of the BEES can translate only simple present tense sentences. It can handle only simple subject and object forms with adjectives, adverbs and articles. Further, to handle out-of-vocabulary issues, it can transliterate English terms into Sinhala. However, this version does not handle semantic issues.

To improve this basic system we have developed three types of systems namely; Human assisted machine translation system [11], web-based English to Sinhala translation system [12][13][27] and context based English to Sinhala machine translation system[14]. The web-based English to Sinhala translation system is a web-enabled version of the stranded English to Sinhala machine translation system. A brief description of the other two developments is given bellow.

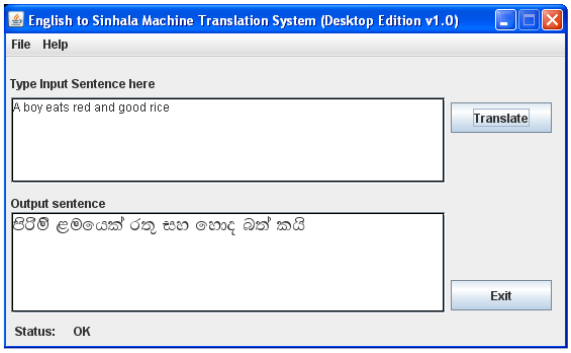

Figure 1: stand-alone Machine translation system

Human-Assisted machine translation system

Human-assisted machine translation system has been developed to solve out-of-vocabulary and semantic issues in the English to Sinhala machine translation. This application has been developed as a java based application and it runs on Linux or Windows based systems. This system provides user interface (Intermediate editor) to semantic handling. Figure 2 shows the user interface of the intermediate editor.

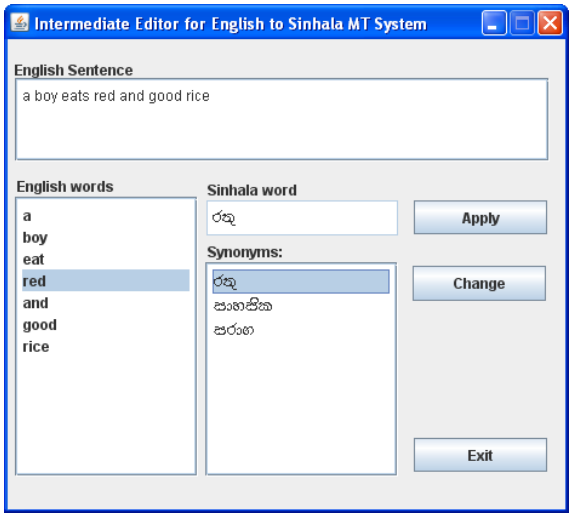

Figure 2: The intermediate editor

This editor provides facilities such as display of synonyms and antonyms and related words. The intermediate-editor is linked with both English and Sinhala dictionaries in the MT system. The process of intermediate-editing, before composing a Sinhala sentence drastically reduces computational costs of running a Sinhala morphological analyzer and parser. In addition, the requirement for post-editing can be reduced by the process of intermediate editing. On the other hand, intermediate-editing can be used as a means of continuous capturing of human expertise for machine translation. This knowledge can be

reused for subsequent translations. With the above ideas we have developed a context-based, English to Sinhala machine translation system to use the human knowledge through the concept dictionary.

Context based handling system

Development of the fully automated, perfectly correct translation system is very difficult for any language pairs. However, we are researching to develop a fully automated machine translation system, using the captured human knowledge throughout the result of the intermediate editing. The result of the intermediate editing is stored in a dictionary named concept dictionary. This information can be used to handle the semantics in the Machine Translation. By using this context-based information, we have developed a context-based machine translation system that translates English paragraphs in to Sinhala. This system has the following features;

- Handling multiple sentences.
- Ability to handle semantics through concept dictionary.
- Ability to handle simple and complex sentences
- Ability to translate all tenses with active and passive

However we have noted that, English to Sinhala web page translation is more useful for many people who use web resources. Therefore, we have developed a new version of English to Sinhala machine translation system that can translate a given English web page into Sinhala. Design of the system is given bellow.

Design of the BEES

The translation system is designed to translate a given English web page into Sinhala. This system contains two modules namely translation module and the HTML parser. Figure 3 shows the overview of the web page translation system.

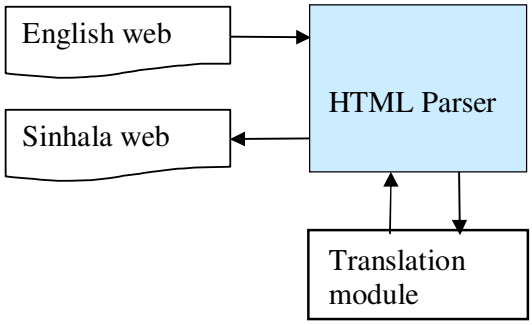

Figure 3: Over view of the translation system

The input of the system is an English web page and the output is a translated Sinhala web page. Brief description of each module is given below.

HTML Parser

The HTML parser is the controlling module of the system. As the first step, the parser analyzes the input HTML document and decodes the text and tags. Then the HTML parser sends the text into the Translation module and gets the Sinhala translated text. Finally, the system composes the web page using these text and tags. The HTML parser has been developed using JAVA.

Translation module (BEES)

We have designed the BEES with seven modules, namely; English morphological analyzer, English parser, word level translator, Sinhala morphological analyzer, Sinhala parser, transliteration module and four lexical dictionaries namely English dictionary, Sinhala dictionary, English to Sinhala bilingual dictionary and concept dictionary. Figure 4 shows the design of the BEES. Note that this new design of the BEES does not contain Inter-mediate editor. This is because this system uses concept dictionary for semantic handling. The concept dictionary is updated through the previous development of the BEES. Brief descriptions of each module are given bellow.

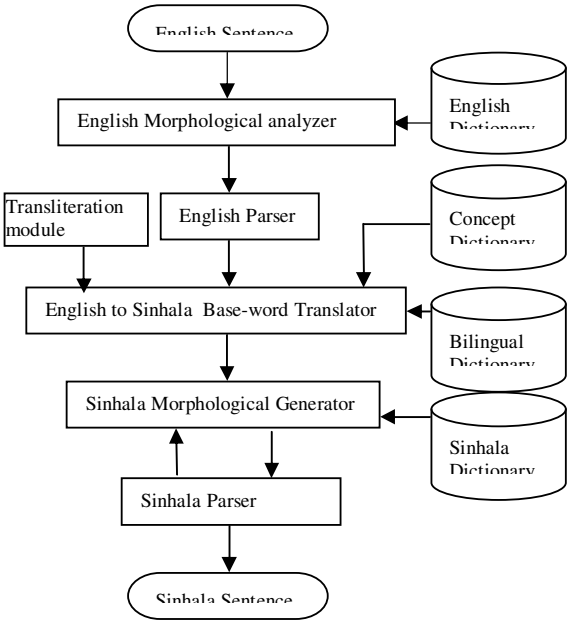

Figure 4: Design of the BEES

English Morphological analyzer reads a given English sentence, word by word and identifies morphological information for each word. The morphological analyzer in our MT system has linked up with an English dictionary to get grammatical information of the words in the input sentence. Using SWI-PROLOG, we have developed a rule based English morphological analyzer for our purpose.

The English parser receives source English sentences and tokens from English Morphological analyzer. This parser works as a syntax analyzer. Since there are many English parsers, we have customized an existing parser for our purpose. The current version of the parser can handle simple and complex sentences including active and passive tenses. The parser has also been implemented using SWI-PROLOG.

The word level translator is used to translate English base-word into Sinhala base-words with the help of the bilingual dictionary and the concept dictionary.

The Sinhala morphological analyzer [7] works as a morphological generator. This morphological analyzer reads the words from the translator word by word. For each word, the morphological analyzer generates the appropriate word with full grammatical information such as nama (nouns), kriya (verb) and nipatha (preposition) in the Sinhala language [31][32]. This analyzer is based on Akshar's and others Morphological Analysis Shell[3] and uses rule based approach for concepts of "Varanagema". It works with the help of two dictionaries, namely, Sinhala rule dictionary and Sinhala word dictionary. All these dictionaries and the Sinhala morphological analyzer have been implemented using Prolog.

The Sinhala parser [6] works as a sentence composer. It receives tokenized words from the Sinhala morphological analyzer and composes grammatically correct Sinhala sentences. In general, a Sinhala sentence contains 5 components, namely Ukktha vishashana (adjunct of subject), Ukkthya (Subject), karma vishashanaya (attributive adjunct of object), karmaya (object) and akkyanaya [33][31]. These five components of a Sinhala sentence are the building blocks for the design and implementation of a Sinhala parser. The parser is also one of the key modules of this English to Sinhala Machine Translation System and it has also been implemented using SWI-PROLOG.

Translation system uses four dictionaries such as English dictionary, English-Sinhala bilingual dictionary, Sinhala dictionary and concept dictionary. The English word dictionary contains English words and the lexical information. English to Sinhala bilingual dictionary is used to identify appropriate Sinhala base word for a given English word and it contains the relation between English and Sinhala words. Sinhala dictionary contains two sub dictionaries namely; Sinhala word dictionary and Sinhala rule dictionary. The Sinhala word dictionary stores Sinhala regular base words and lexical information. The Sinhala rule dictionary stores rules required to generate various word forms. These are the inflection rules for formation of various forms of verbs and nouns from their base words. The rule dictionary also stores vowels, consonants, upasarga (prefix) and vibakthi (postfix). The concept dictionary contains three sub dictionaries namely; English concept dictionary, Sinhala concepts dictionary and bilingual concept dictionary. The English concept dictionary contains synonyms, anti-synonyms and general knowledge about English words. Similar to the English dictionary, Sinhala concept dictionary stores Symantec information. The bilingual concepts dictionary stores bilingual semantic information which are update by humans through the intermediate editing.

Transliteration module is used to solve out-of-vocabulary problems and to translate technical terms. Transliteration is the practice of transcribing a word or text written in one writing system into another writing system [10]. In other words, machine transliteration is a method for automatic conversion of words in one language into phonetically equivalent ones in another language. At present, we have developed two types of transliteration models. One of these models transliterates original English texts into Sinhala Transliteration and the other transliterates Sinhala words that are written in English, which transliterate into Sinhala. Finite state transducers are used to develop these two modules.

The following section reports some translation issues that are handled by the system.

Translation issues

The English to Sinhala web page translation is a critical process considering the large and complex type of sentences. This section describes some common issues that are addressed by the BEES.

Text manipulation issues

An html document contains a lot of tags and text. The text on the web document is not completely sentences. These texts are available in several formats such as;

- Complete sentences
- Noun phrases
- URLs
- Equations
- Numbers etc.

The web translation system needs to handle these texts for target language generation. Identification of the complete sentence is one of the critical problems in the context based machine translation. Any sentence in English ends with a dot sign (.) and after the dot sign the space appears. Using these two character combinations, the system identifies the sentence. However there is a problem in understanding names (Example: A. B. Fernando) Note that, the "A." is not a sentence ending therefore HTML parser uses internal mechanism to remove this issue. Also Noun phrase identification is another issue in the translation. As an example consider the following

phrase "A Computer Science Subject", is translated as a "mrs.Kl jsoHd jsIhla". Note that there are grammatical differences between English and Sinhala languages; therefore, word level translation cannot be used. This is because there is a difference between Sinhala nouns in the noun form and adjective form ("mrs.Klh" is a noun form and mrs.Kl is an adjective form.) [33] Also in Sinhala, article comes with a Sinhala noun. According to the above reasons, we have developed a translation module to translate noun phrases. However, URLs, Numbers and equations cannot be translated.

Grammatical issues

There are several issues that have been addressed by the present system. Due to having different language structures in English and Sinhala languages, the translation of English to Sinhala is a difficult process. English is a West Germanic language that originated in Anglo-Saxon England. Sinhala belongs to the Indo-Aryan branch of the Indo-European languages [29]. Following list shows some grammatical issues in both languages.

- The literary language and the spoken language differ from each other in Sinhala.
- Sinhala uses SOV (Subject Object Verb) word order and English uses SVO (Subject Verb Object) word order.
- Sinhala nouns have five types of inflections, namely, gender, number, person, case and artical (difinite/indifinite). English nouns have four types of inflections, namly; gender, number, person and case.
- Sinhala has nine cases and these differ from English.
- There is a difference between noun and the adjective form of the noun in Sinhla but no such difference is found in English.
- Sinhala language contains only three tenses while English has 12 tenses.
- Sinhala sentences contain 5 components, namely Ukktha vishashana (adjunct of subject), Ukkthya (Subject), karma vishashanaya (attributive adjunct of object), karmaya (object) and akkyanaya. However, this structure is different from the English sentence structure.

This English to Sinhala machine translation system uses the concept dictionary to its semantic handling. The following section shows how the system works for a given input text.

How system work

This section describes how the system translates a given English web page into Sinhala. Figure 5 shows the user interface of the system.

Figure 5: User interface of the BEES

To start the translation, you need to select a web page and click the translation button. After the translation, the system shows the output of the translation by using a web browser. Figure 6 shows the translated output of the Sinhala web page. Assuming that the system reads the following simple HTML document, as the first step HTML parser analyzes the document and identifies the tags and the text. Consider the following simple part of the html document.

```
<tr><td>
        The Rabbit
</td></tr>
<tr><td>
        <img src="trabsl1.jpg">
        The Rabbit is a small and herbivorous animal.
It lives in the jungle. Rabbit has long and powerful legs.
</td></tr>
```

This HTML source contains several HTML tags and text. "The rabbit" is a text identified by the HTML parser. Then the parser sends this text into the translation module. Translation module reads the above text and tries to translate. In the sentence analyzing stage, the English parser rejects the input text, because it is not a sentence. Therefore, the system tries to identify it as a noun phrase. At the moment, the English parser recognized the input text "The rabbit" as a noun phrase. Then the translation module uses the English to Sinhala word translator, Sinhala morphological analyzer and the Sinhala parser, and generates the appropriate Sinhala translation as "ydjd".

This is the time to show how a translation module works for a given complete sentence. Assume that the translation module reads the sentence "The Rabbit is a small and herbivorous

animal" as an input text. Then thw English morphological analyzer reads the input sentence and returns the following.

```
eng_detm([e1000002], dr, 'the').
eng_noun([e1000077], td, sg, ma, sb, 'rabbit').
eng_verb([e1000057], if, 'is').
eng_detm([e1000001], id, 'a').
eng_adjv([e1000074], p, 'small').
eng_conj([e1000020], 0, 'and').
eng_adjv([e1000076], p, 'herbivorous').
eng_noun([e1000059], td, sg, co, sb, 'animal').
```

eng_detm/3, eng_noun/6, eng_verb/3, eng_adjv/3 and eng_conj/3 are the prolog predicates to represent English words. Then English parser reserves above information and analyzes the English sentence. The English parser returns the following predicates.

```
eng_sentence_type(simple,if).
eng_sen_verb([e1000057]).
eng_sen_complement([e1000001, e1000074, …]).
eng_sen_subject([e1000002, e1000077]).
eng_sen_ekeys([e1000002, e1000077, …]).
```

This English parser indentifies the subject, verb and complement of the sentence. It stores these information using prolog predicates such as eng_sen_verb/1, eng_sen_complement/1 and eng_sen_subject/1. After successful syntax analysis, word translator translates the correspondent Sinhala root word for a given input root word. The word translator returns the following predicates.

```
estrwords(1001, e1000002, s1000000, dt).
estrwords(1002, e1000077, s1000078, na).
estrwords(1003, e1000057, s1000059, vb).
estrwords(1004, e1000001, s1000000, dt).
estrwords(1005, e1000074, s1000076, aj).
estrwords(1006, e1000020, s1000018, cn).
estrwords(1007, e1000076, s1000077, aj).
estrwords(1008, e1000059, s1000060, na).
```

The estrwords/4 prolog predicates represent bilingual information for each English root word. By using this information Sinhala morphological generator generates suitable Sinhala words for the corresponding English word with full grammatical information.

```
snoun([s1000078], td, sg, ma, li, dr, v1,'ydjd').
sin_fverb([s1000059], td, sg, pr,'h').
sin_adjv([s1000076],'l=vd').
sin_conj([s1000018],'iy').
sin_adjv([s1000077],'Ydl NlaIl').
snoun([s1000060], td, sg, co, li, id, v1,'isjqmdfjla').
```

Using all these information the Sinhala parser generates the appropriate Sinhala sentence as "ydjd l=vd iy Ydl NlaIl isjqmdfjlah'".

After the successful translation HTML parser reads these translated texts and composes a corresponding web page. Using this interface the user can see the original English web page and the translated Sinhala web page separately. Figure 6 shows the output web interface of the translator.

Figure 6: Translated output web page

Conclusion and Further Works

This paper has reported the design and implementation of the English to Sinhala machine translation system that can translate an English web page into Sinhala using the concept of "Varanagema". The "Varanagema" concept has reduced the workload of the Sinhala morphological generation and the number of word forms to be stored in dictionaries. Further the context based approach is used to semantic handling in the system. Therefore this system becomes a fully automated system.

However, we have identified that the identification of the context in the paragraphs or a sentence is a complex task and hence needs improvement. Updating the lexical resources and generating an algorithm to identify the context of the text are further work of this project.

References

[1] Akshar B., Chaitanya V., Sangal R., Natural Language Processing: A Paninian Perspective, Prentice Hall of India, New Delhi, India, 1995.

[2] Akshar Bharati, Vineet Chaitanya, Amba P. Kulkami, & Rajeev Sangal: Anusaaraka: machine translation in stages. Vivek: a

Quarterly in Artificial Intelligence, vol.10, no.3, July 1997, pp.22-25.

[3] Akshar B., Sangal R., Sharma D. M., Mamidi R., Generic Morphological Analysis Shell, In Proceedings of LREC 2004 -SALTMIL Workshop: First Steps in Language Documentation for Minority Languages. Lisbon, Portugal, 2004.

[4] Sinha R.M.K, Jain A., "AnglaHindi: an English to Hindi machine-aided translation system", MT Summit IX, New Orleans, USA, 23-27 September 2003; pp.494-497.

[5] Mahesh R, Sinha K., Integrating CAT and MT in AnglaBhart-II architecture, 10th EAMT conference "Practical applications of machine translation", May 2005, pp235-244.

[6] Hemant D., Computer-assisted translation system – an Indian perspective. Machine Translation Summit VII, 13th-17th September 1999, Kent Ridge Digital Labs, Singapore. Proceedings of MT Summit VII "MT in the Great Translation Era"; pp.80-85.

[7] Hettige B., Karunananda A. S., A Parser for Sinhala Language – First Step Towards English to Sihala Machine Translation, proceedings of International Conference on Industrial and Information Systems (ICIIS2006), IEEE, Sri Lanka, 2006, pp583-587.

[8] Hettige B., Karunananda A. S., A Morphological analyzer to enable English to Sinhala Machine Translation, Proceedings of the 2nd International Conference on Information and Automation (ICIA2006), Colombo, Sri Lanka, 2006, pp. 21-26.

[9] Hettige B., Karunananda A. S., Developing Lexicon Databases for English to Sinhala Machine Translation, proceedings of second International Conference on Industrial and Information Systems (ICIIS2007), IEEE, Sri Lanka, 2007, pp. 215-220.

[10] Hettige B., Karunananda A. S., Transliteration System for English to Sinhala Machine Translation, proceedings of second International Conference on Industrial and Information Systems (ICIIS2007), IEEE, Sri Lanka, 2007, pp.209-214.

[11] Hettige B., Karunananda A. S., Using Computer-Assisted Machine Translation to overcome language barrier in Sri Lanka, Proceedings of the 4th Annual Sessions of Sri Lanka Association for Artificial Intelligence (SLAAI), University of Moratuwa, 2007.

[12] Hettige B. Karunananda A.S., Web-based English-Sinhala translator in action, Proceedings of the 4th International conference on Information and Automation foe Sustainability (ICIAfS 08), IEEE, Sri Lanka 2008, pp 80-85.

[13] Hettige B. Karunananda A.S., Web-based English to Sinhala Selected Texts Translation system, Proceedings of the 5th Annual Sessions of Sri Lanka Association for Artificial Intelligence (SLAAI), The Open University of Sri Lanka, October 2008.

[14] Hettige B. Karunananda A.S., Context-Based approach to semantic handling in English to Sinhala Machine Translation, Poster presentation on 27th National Information Techonology conference (NITC09), Sri Lanka, September 2009.

[15] Herath. A , Hyodo. Y, Kawad Y, Ikeda T, A Practical Machine Translation, System from Japanese to Modern Sinhalese, The Logico-Linguistic Society of Japan ,1995.

[16] De Silva, D.; Alahakoon, A.; Udayangani, I.; Kumara, V.; Kolonnage, D.; Perera, H.; Thelijjagoda, S.,Sinhala to English Language Translator, Proceedings of the 4th International conference on Information and Automation foe Sustainability (ICIAfS 08), IEEE, Sri Lanka 2008, pp 419-424.

[17] Vithanage N. V. C. T., English to Sinhala Intelligent Translator for Weather forecasting domain, Thesis submitted BIT degree, University of Colombo, Sri Lanka, 2003.

[18] Weerasinghe R, A Statistical Machine Translation Approach to Sinhala-Tamil Language Translation, Department of Computation and Inteligent Systems, University of Colombo School of Computing, Sri Lanka.

[19] Wikipedia, http://www.wikipedia.org

[20] Madhura dictionary, http://www.maduraonline.com/

[21] Vidudaya dictionay, http://www.dscs.sjp.ac.lk/sinres/index.htm

[22] SYSTRAN : www.systransoft.com/

[23] SWI-PROLOG: http://www.swi-prolog.org

[24] Prolog server page: URL: www.benjaminjohnston.com.au

[25] Babel Fish, http://babelfish.yahoo.com/

[26] Sinhala Unicode, http://locallanguages.lk/

[27] BEES: http://www.dscs.sjp.ac.lk/psp/bees.htm

[28] Google translator, http://translate.google.com

[29] Language Technology Research Laboratory" URL:http://www.ucsc.cmb.ac.lk/ltrl/

[30] Apertium Machine translation system, http://www.apertium.org/

[31] Gunasekara A. M., A Comprehensive Grammar of the Sinhalese Language, Asian Educational Services, New Delhi, Madras, India, 1999.

[32] Karunathilaka W. S., Sinahala Basha Viyakaranaya, M.D. Gunaseena & Company, Clolombo 11, Sri Lanka, 2003.

[33] Karunarathna S., "Sinahala Viharanaya", Washana prakasakayo, Dankotuwa, Sri Lanka, 2004.

[34] Wren P.C., Martin H., High School English grammar and Composition, S. Chand and Company Ltd, Ram Nagar, New Delhi, India, 2005.

[35] Tafseer A., Sadaf A., English To Urdu Translation System, University of Karachi, 2002.
URL:
http://www.khazina.org/files/translator02.pdf

[36] Toshio Y, "The EDR electronic dictionary", Communications of the ACM, Volume 38, Issue 11, 1995, pp. 42 – 44.