Computer Engineering and Applications Vol. 6, No. 3, 2017

Arabic TranslationWeb Services: An Implementation Survey Towards Arabic Language Grid

Arif Bramantoro*

Faculty of Computing and Information Technology, King Abdulaziz University, Rabigh 21911, Saudi Arabia
**asoegihad@kau.edu.sa*

ABSTRACT

This research proposes the development of Arabic language service. It is a servletbased Web service which provides a translation from English into Arabic, using techniques to develop Web services such as Restful and API in Java language. This API is part of Language Grid, a project in Japan to collect, share and combine as many language resources as possible by wrapping the language resources as web services, which is also known as Everything as a Service (XaaS) technology. By having Arabic language services connected in the Language Grid, there will be a wider use of Arabic language resource in the world. An evaluation of running the service is provided to enhance the performance and reliability of the service.

Keywords: Machine Translation, Web Services, Arabic Language Resources.

1. INTRODUCTION

There are various available kinds of linguistic services on the Web, such as text translations and dictionaries. Furthermore, several natural language processing tools are currently available, where these tools can be transformed into web services. Web services can be defined as software components, which communicate with the use of various open persistent, standards dependent web techniques. The use of these standards makes web services independent on any programming language, hardware or operating system. In other words, applications written by using various programming languages and executed on several platforms are able to transmit data easily over the Internet with the use of these services. Thus, the main purpose of designing these services is to be accessed and used via other applications, where this allows them to be explored, located and published irrespective of fundamental architectures.

Web services assist organizations perform their businesses in open, easy and flexible manners over the Internet. Using web services solves the problem of the need to make organizations' applications work with the applications of their providers and customers. Web services depend on using universal languages to transmit instructions and information with translation, where they offer incorporated applications that are able to automatically perform businesses without the need of human intervention. Other benefits of web services are modifying the main program logic without critical impact on interfaces, supporting system architectures, hiding complexity of systems with the use of standard interfaces, providing platform

neutral applications and improving legacy systems without the need of changing the underlying code [1, 2].

Various community dependent language resources that target specific application fields have been developed. In practice, the customization of composite linguistic services according to users' requirements can be performed based on establishing specific infrastructures that allows combining elemental linguistic services with incorporated language resources. One of these infrastructures is the language grid, which permits users to generate new language services based on their multilingual or intercultural activities [3].

This work aims to develop an English-Arabic translation web-service to be used in the Language Grid with a purpose to facilitate accessing and using Arabic language resources. It can be considered as an enhancement for the traditional Google translation service. Google translation is a web dependent and statistically dependent Google machine translation service, which allows translating texts from a specific language into another based on offering a machine translation among languages. Web based translation infrequently offers perfect translations, but it assist readers to have rough ideas concerning the text basic contents.

Google translation API allows programs and websites incorporating with the translation service programmatically. It mainly has two API versions, a paid version, where its free trial version has limited number of words and a free beta version, which is mainly used by developers. In this work, beta is deployed since it is a free version. The main limitations of using Google translation service are that it is a process based service, which leads to a delay in the response time as well as users must accept all Google terms and conditions to use its API.

In practice, the increasing dependency on services and information that offered by current web servers led to increase the need for more improvements concerning the performance and reliability of these services. The most effective issue on both the performance and reliability is the web server architecture, which is mainly characterized by the processing model that demonstrates the kind of utilized threading or process scheme to support operations of web servers. The most used processing model in web applications is the process based one. However, it suffers from various problems, such as complexity, the need for more overhead and its own address space, the communication among process is expensive and the delay in the response time to the need to access the server by each process. One of the main process based services is the Google translation. Thus, this work offers an enhancement for the traditional Google translation service to offer a solution for all these problems based on developing a thread based translation web-service to be included as an enhanced language English-Arabic translation resource in the future Arabic language grid.

2. ARABIC MACHINE TRANSLATION

Arabic language is rich of suffixes where its derivational and inflectional productions produce numerous word forms. Arabic pronouns, prepositions and articles can be affixed to verbs, particles and nouns, where this causes challenges to translate words from/to Arabic [4].

In general, Arabic words are considered as ambiguous ones because of the word three-letter root method. This method let the Arabic language covers a large range of

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

words' meanings. Some derivations results in dropping root letters, where this causes the ambiguity.

An Arabic sentence with structural representation can be considered as ambiguous. In [5], the authors developed the initial Arabic parser to describe the basic syntactic Arabic structures with the use of Xerox linguistics environment that permits writing several notations and grammatical rules. However, the main limitation of this work is that it was evaluated using short sentences.

In [6], the authors studied current standard Arabic with a focus on the concepts of Arabic phonology, morphology, orthography, scripts, semantics and machine translation. However, [7] presented major lexicons and rules based system, which demonstrates the morphological analysis and production of online Arabic language words that signified in the typical orthography. The authors demonstrated that patterns, roots and all affixes incorporate with characteristic tags represent voices, numbers, moods and persons. Thus, various efforts have been conducted concerning the development of Arabic machine translations.

Various efforts have been conducted in the processing of Arabic language in various applications, mainly in machine translation. In [8], the authors presented the development of a technique to enhance the statistical machine translation performance with the use of a syntactic parser to rearrange verb-subject sentence constructions into subject-verb. However, this did not offer a complete solution for the problem since several verb re-orderings were missed.

In [9], the authors presented a method to study Arabic-English translation depending on supervised alignment data, where the performance of this method was distinguished from that of human annotation. On the other hand, the authors in [10] implemented an Arabic-English machine translation system, which known as Npae-Rbmt based on a transfer-based technique. Based on evaluating the method using 88 computer science thesis titles and journals, it achieved 94.6% accuracy rate. In [11], the authors developed another Arabic-English machine translation system, which called UniArab, where it depended on a role and reference grammar scheme. However, the used restricted lexicon limits the ability of the system to translate several words since their structure is not available in the system.

[12] introduced a comparison among two freely available machine translation systems; Babylon and Google Translate in terms of the translation from Arabic to English. The conducted work depends on using a corpus that includes more than thousand Arabic sentences with more than four thousands Arabic words and applied for each machine translation system. After defining the corpus, each Arabic sentence is translated with the use of both systems; Babylon and Google Translate and two human translations. A preprocessing stage is applied on these sentences to segment the text into several n-gram sizes, where the precision for both systems is then computed for all gram sizes. The result demonstrates that the Google translate outperforms the Babylon one in terms of Arabic-English translation precision. Hence, the use of Google translate is proposed in this paper.

Machine translation developers must consider all orthographical differences among Arabic and English languages. In English language, italics are utilized to

point out emphasis to a specific word, while in Arabic language; there must be a change in the word order or the use of emphatic words.

English to Arabic machine translation system was developed by the authors in [13] with the use of the transfer approach. The author focused on analyzing the English language as a language, translation problems from English into Arabic and Arabic production as a target language. However, the conducted work was restricted to electronic texts, which represent texts written in a machine understandable layout.

In [14], the authors developed a method to assess Arabic language machine translation systems based on applying it in four developed commercial machine translation systems to translate from English to Arabic. The obtained evaluation results of those systems were demonstrated for Arabization and Internet domains.

It can be concluded that English-Arabic machine translation systems depend on using methods and approaches that to some extent differ from those used in ArabicEnglish machine translation ones. The majority of the conducted researches coped with the agreement and considered it as a main characteristic that seriously influences on the English-Arabic machine translation systems' outputs and offered solutions to enhance word reordering.

3. THE DEVELOPMENT OF ARABIC TRANSLATION SERVICE

The importance of information technologies is stronger day by day, in a manner that it opens a real competition between organizations. The translation and general lingual services are playing major roles in human life, especially in commercial, political, and educational purposes. This research is proposed to establish the role of Saudi Arabia in the language resource infrastructure community. By having this service, Saudi Arabia is one of 17 countries which are participating in the Language Grid to support the Arabic language data base for web applications associated with lingual translation and other uses.

Arabic language resources include Arabic parallel texts in other languages, bilingual dictionaries machine translators, dependency parsers, text to speech, speech recognizers, named entity detectors and others. Some of the Arabic language resources are currently online and some are not. To date, none of Arabic language resources are connected and shared. To enhance university's network in utilizing its resources of Arabic language resources establishing a new operation center of the Language Grid in the Kingdom is inevitable. This grid should handle the quality of service (QoS) of all distributed service and data utilizing Arabic language resources. QoS includes response time, scalability, accuracy, and others.

This research is conducted to enhance Arabic language performance for web services. It is a step toward building a grid for Arabic characters to be used in translation services. This research works on the responding time factor; hence the associated code will minimize the responding time to translate from English to Arabic via a proper algorithm. This algorithm will make an additive advantage to the translation process that is held by Google translator, hence that it will be proved that the outcomes are faster than before. Our work could be useful for web developers, organizations that are interested web services, and organizations which are interested in Arabic linguistic issues.

Initially, we should know how does Google translation tool works, where it is commonly known that Google translator considered as "statistical machine translation", which means it is a word-based translator, hence its technique gathers

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

as much as possible parallel words together, to match meaning between two languages, this technique aims to extend word pairs shared between two languages. On the other hand; this methodology of work can not apply grammatical rules to detect the meaning, which sometimes considered as a limitation.

Google translator does not translate from a language to another directly, but there is an intermediate language between them. Google initially translates to English, then translate again from English to the desired language. Unfortunately this is another source of error probability. Hence, English is not rich enough to match all lingual derivations words. It is noteworthy that the technique of ”statistical machine translation’ helped to establish a good number of word pairs that can be directly translated from a language to another without an intermediate language. When we search for the word meaning in Google translate, it looks for a hundreds of millions document, and hence Google translator is already supported by the algorithm of intelligent guess, it can show the proper meaning and also another alternative word for the searched word.

Google works in grids technology for classifying the involved languages characters. Consequently, there is a requirement for us to illustrate more deeply what the grid concept is, and where our work will help in Arabic language grid building. Google offers information about grids and the countries that are participating in building and enhancing the grids of their native languages. First, we define the Language Grid as an online web service for multilingual purposes, which provides easy registration and language information sharing in web online services like: machine translators and online dictionaries. In language grid, there are 144 organizations come from 17 countries participating.

The Language Grid is a service grid that aims at sharing language services around the world. It connects language resource providers and users using Web service technologies. Service Grid aims at connecting various types of Service Grids that are operated by different organizations, and building an infrastructure of Service Grid by federated operation among international organizations across domains.

There is a difference between the translation by the Language Grid machine translation and the system of conventional translation available on the Internet. The system user can improve the performance quality of translation by utilizing the existed the Language Grid facilities, and users can utilize the registered parallel texts for translating. The example is parallel texts provided by Kawasaki City Board, which is already registered in the Language Grid. The user requests a translation for a sentence, then some examples which have the same meanings to the requested sentence will be shown automatically. If the user find a suitable sentence that conveys the meaning, the correct meaning is obtained by applying the parallel texts results. If the needed expression is not available, a dictionary registered by users can also help to enhance the translation quality. For example, when the school dictionary terms is registered, it gives a significant improvement to the translation result.

English is normally the hub language for the machine translation. For example, the technique for translation from Japanese language to Portuguese language is done as the following: first the translation from Japanese to English is held, and then from English to Portuguese. The Language Grid can include and help in performing composite online services, such as the multi-hop translation illustrated above. We

can realize these when using the Language Grid, where it is not available when using conventional translation.

The codes that have been built in this research aim to add improvements on Google translate, it minimizes the translation time in each trial made for a certain word. The code was written in Java language. Google Company provided a free API, which can be used by developers to add their ideas in order to boost Google translator performance.

The main enhancement on Google translation when using this web service, is reducing the respond time needed, hence the system in our tool does not need to reach

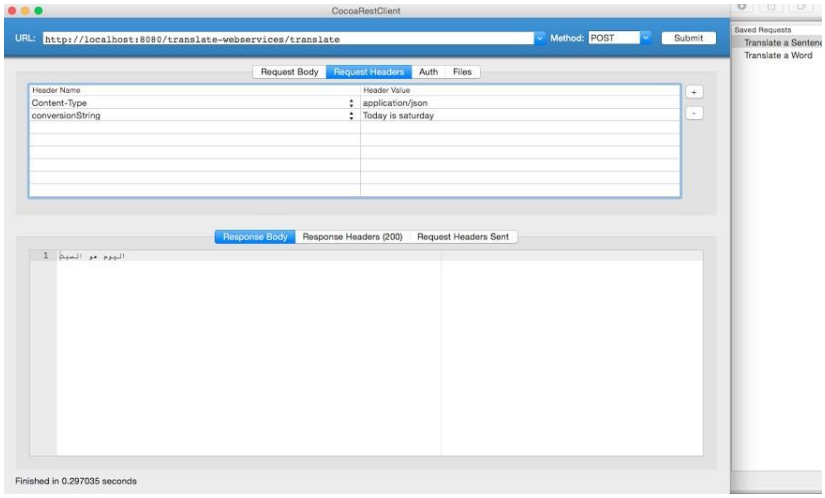


FIGURE 1. The interface when translating a sentence ”Today is Sunday”

the server, as the case in Google translator. This advantage was established by using java servlet (a tool that can be used to create a web application), also we used the beta version that is provided for free in Google translator API. The code is written in java language. Java has a property that helped in minimizing the time. Java forms threads instead of making full process, thus the threads are reaching the servers much faster, then the time response is decreed. The code actually depends on the java servlet to accomplish the work.

When the Language Grid tool is open, it informs us that we can use any API involved with their grids. This massage is available in the home page of the Language Grid tool web site which says ”Multilingual Studio is a set of APIs for using the multi-language functions provided by the Language Grid, such as back translation, text-to-speech, multi-hop translation and so on”.

This research is part of a big work to improve and support the Arabic grid on which the kingdom will be the participant in it. Hence, we focus on translation from English to Arabic. One problem can face in this work is that Arabic does not have a distinct API, so we must choose an intermediate language that has a distinct API, Romanian language has been chosen randomly, we have also provided the system with an interface as illustrated in Figure 1.

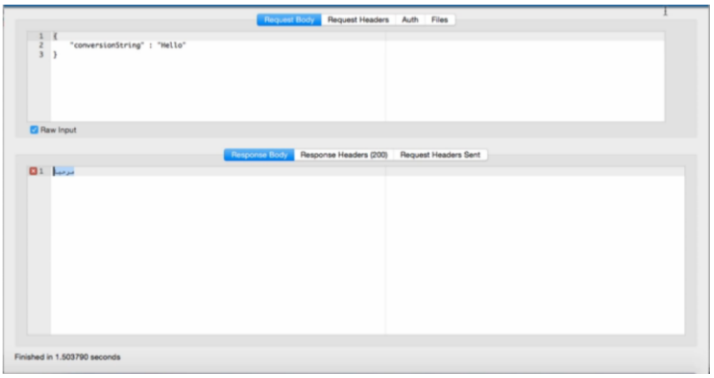
ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

4. DISCUSSION

The application was tested repeated times to check the accuracy and precession of the output, we have applied this improvement by writing a code using java servlet, in order to develop a grid for Arabic language. Knowing that this research is a part of a large work aims to develop a special grid for the Arabic language, which can support the translation from and to Arabic.

The system is composed of a code and a proper interface, which allows translating word or sentences, saves the previous requests, and shows the responding time associated with each translation request. The Romanian language was used as an intermediate language between Arabic and English; hence the Arabic language does not own a distinct grid until now.

The main idea of this research aims to reduce the respond time of translation, to



be less that that needed by Google translator. In addition the algorithm is designed in a manner that causes a reduction in respond time with each submitting trial. For example, Figure 2 shows the first trial to submit the word ”hello”, at the first trial the respond time took about 1.5 seconds, the second submit took about 0.67 seconds, the third took about 0.58 seconds and finally, the fourth trial took about 0.12 second, which means a 91% of the first trial is eliminated. The system works into two topologies in translation process, one is for translating single words, and the other is for translating multi-words sentences. Table 1 shows the recorded time for translating different number of words, done for four trials for each. All the outcomes were typically match the required time reduction. The expected outcomes pay attention to that the reduction in time became larger as we increase the number of trials. This can give us a good indication, about success in building the required grid for Arabic language in the near future.

Even the previous work added more advantages to the translation process; it still needs more work for building a distinctive Arabic grid. Hence the reduction in response is a small step in the whole work needed. The faster the data base for Arabic vocabulary is formed; the easier the grid is built. After forming a complete data base, translation between Arabic and English will not need an intermediate language any more. Instead, Arabic then could be used as an intermediate language between two languages, and this will take it a step for being a global language.

Furthermore, we made other trials but with using the German language as intermediate, hence the German language also has an API. The results were also satisfying. However, the Romanian is the chosen language to go to the next step in developing this algorithm.

After the interface had been ready to use, we surveyed a group of people from a deferent educational level to try the system, after explain the way to use it and the aim of this research. The number of participants was 53. Most of them welcomed the idea from the beginning, the main reason is they tend to support their mother language at the national levels. 43 people expressed that the system is satisfying enough and easy to use, where five people is see that Google translator can be trusted more. However, all of the users do not prefer the method that the system is operated, like writing the required word in the URL. This will be solved in the next steps of the future development.

Nevertheless, executing the program several times did not show significant errors, and the translation in each time was as expected, thus the most suitable meaning has been chosen by the tool, with a reduction in time in each trial.

TABLE 1.
The recorded time for translating different number of words

Number of words	1st trial (sec)	2nd trial (sec)	3rd trial (sec)	4th trial (sec)
1 word	0.305582	0.237148	0.205296	0.155085
2 words	0.267320	0.167424	0.133814	0.209118
4 words	0.299808	0.219695	0.217045	0.162021
8 words	0.296764	0.271510	0.201099	0.171853
16 words	0.280653	0.276265	0.238175	0.165363

5. CONCLUSION

Digital data and web services are becoming more and more important a day by a day, translation services are mostly from the common and daily used services in web. Each country should be responsible for supporting the translation services associated with their mother language, hence the language is a major part in the concept of the ”local wisdom”, and Kingdom of Saudi Arabia took part in enhancing the rating of Arabic importance among the web services, especially in translation. Trends are going on now a day to build a proper grid for Arabic language. This research is considered as a part of this total work. On this work, we made an enhancement on the performance of Google translator, in a manner that decreases the time of word translation. The code also works in an accumulative performance in response time reduction; hence the time is decreased as the number of translation trials are increased; for the same required phrases. The result showed a reduction reaches 91% after four requesting trials, which is a very satisfying percentage for this aim.

We are expecting more developed data base and more support for Arabic language, by the kingdom or other countries, and more efforts to support the using of this global translator by people who Arabic is not their mother language, hence the associated services for this purpose is still not strong enough. In addition, we expect more enhancement in the quality of translation, especially in terms of

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

accuracy, hence Arabic language can have many meanings to single word, differs according its place in the sentence, also there should be more work to simplify finding the related derivations for a certain word.

In the future, our work is the base of developing the Arabic Language Grid. Hence, there will be more languages can be translated from/to Arabic by connecting to other language grids in Japan, Thailand, and Indonesia.

ACKNOWLEDGEMENT

This work was supported by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, Saudi Arabia. The authors, therefore, gratefully acknowledge the DSR technical and financial support. The author also thanks Fatemah Ahmed Al-Jeraisy for the implementation in Saudi Arabia.

REFERENCES

[1] E. Cavanaugh, "Web services: Benefits, challenges, and a unique, visual development solution. a white paper published by altova," Inc. USA, 2006.

[2] P. Miller, S. K. Sharma, and F. L. Kitchens, "A flexible services architecture based translator web services." in M-Services, 2002.

[3] T. Ishida, "Language grid: An infrastructure for intercultural collaboration," in Applications and the Internet, 2006. SAINT 2006. International Symposium on. IEEE, 2006, pp. 5 –pp.

[4] I. T. Khemakhem, S. Jamoussi, and A. B. Hamadou, "The miracl arabic-english statistical machine translation system for iwslt 2010." in IWSLT, 2010, pp. 119 – 125.

[5] M. A. Attia, "Handling arabic morphological and syntactic ambiguity within the lfg framework with a view to machine translation," Ph.D. dissertation, University of Manchester, 2008.

[6] N. Y. Habash, "Introduction to arabic natural language processing," Synthesis Lectures on Human Language Technologies, vol. 3, no. 1, pp. 1–187, 2010.

[7] L. S. Larkey, L. Ballesteros, and M. E. Connell, "Improving stemming for arabic information retrieval: light stemming and co-occurrence analysis," in Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2002, pp. 275–282.

[8] M. Carpuat, Y. Marton, and N. Habash, "Improved arabic-to-english statistical machine translation by reordering post-verbal subjects for word alignment," Machine Translation, vol. 26, no. 1-2, pp. 105–120, 2012.

[9] A. Ittycheriah and S. Roukos, "A maximum entropy word aligner for arabic-english machine translation," in HLT/EMNLP 2005, Human Language

Technology Conference and Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, 6-8 October 2005, Vancouver, British Columbia, Canada. The Association for Computational Linguistics, 2005, pp. 89–96. [Online]. Available: http://aclweb.org/anthology/H/H05/H051012_.pdf

[10] O. Shirko, N. Omar, H. Arshad, and M. Albared, "Machine translation of noun phrases from arabic to english using transfer-based approach," Journal of Computer Science, vol. 6, no. 3, pp. 350–356, 2010. [Online].
Available: https://ukm.pure.elsevier.com/en/publications/machine-translationof-noun-phrases-from-arabic-to-english-using-

[11] Y. Salem, A. Hensman, and B. Nolan, "Implementing arabic-to-english machine translation using the role and reference grammar linguistic model," Dublin Institute of Technology, 2008.

[12] L. S. Hadla, T. M. Hailat, and M. N. Al-Kabi, "Evaluating arabic to english machine translation," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 5, no. 11, pp. 68–73, 2014.

[13] M. A.-A. Attia, "Implications of the agreement features in machine translation," Ph.D. dissertation, Faculty of Languages and Translation, Al-Azhar University, 2002.

[14] A. Guessoum and R. Zantout, "A methodology for evaluating arabic machine translation systems," Machine Translation, vol. 18, no. 4, pp. 299–335, 2004.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)