Journal of Engineering and Applied Sciences 13 (6): 1442-1450, 2018
ISSN: 1816-949X

The Role of Natural Language Processing on Technology Enhancement

[1]Ema Utami and [2]Suwanto Raharjo
[1]Department of Informatics Engineering, Universitas Amikom, Jl. Ring Road Utara, Condong Catur, Depok, Sleman, Daerah Istimewa, 55281 Yogyakarta, Indonesia
[2]Department of Informatics Engineering, Faculty of Industrial Technology IST AKPRIND, Jl. Kalisahak 28 Kompleks Balapan, Gondokusuman, Yogyakarta, Daerah Istimewa, 55222 Yogyakarta, Indonesia

Abstract: The grew of technological developments has made a lot of changes, adjustments and improvements in many areas included the field of computer science. Internet as one of the icons of the technological enhancement give impacts on the amount of data resulted by the internet services. Natural Language Processing (NLP) as one of the computer science and linguistics studies also grown in the presence of such changes. The development of machine translation also triggered by increasing number of the internet usage. Even not as machine translation progress but development of transliteration machines are also affected by internet. The amount of existing data will necessarily require fast processing but with accurate results. The presence of technology and the development of computer science can help data processing such make big data into smart data. Research on NLP field is still very wide open in the field of literacy, for example, especially, for countries having their own script. The integration of electronic sensors with the use of natural language will be easier for users to communicate with the machine are also big opportunities for research.

Key words: NLP, translator, transliterator, internet, smart data, opportunities, translation

INTRODUCTION

The current technological developments grew extremely fast and made a lot of changes, adjustments and improvements in many areas included the field of Computer Science. Until now, technological development still follow Moore's Law, simply said the technology is growing exponentially at a more affordable price. One example of these progresses is the greater capacity of data storage media (hardisk) for personal need indicated by the use of terabyte unit. The other example is the number of cores in a processor used in laptops and personal computers that are more and more increasing in number and speed. We can easily find the other sophisticated technologies with affordable prices.

The users of the internet as one icon of technology development are increasing in number and more rapidly increasing in the future. Nowadays, when we meet the users of mobile smart phone, their devices are very likely provided with internet connections. One statistics site reported that the users of internet in 2016 were estimated about 3.5 billions of people (http://www.internetlivestats.com/internet-users/, accessed Aug 26, 2016). The number of internet users are so large it would have a multiplier effect in many ways. Some of the impacts that we can already experience at this time is a change in social interaction, cultural, buying and selling and so forth.

In computer science, one of crucial effects is the huge size of data resulted from the services existing in the internet. Such huge size of data is recently often termed as the bigdata. The term bigdata is generally used to refer the huge set of data in number, type as well as in rate of increase (Chen *et al*., 2014). The term bigdata firstly came up on 1998 in a slide of presentation and title of book but noted in a study in year of 2000 (Fan and Bifet, 2013).

NLP (Natural Language Processing) also grown in the presence of such changes. NLP is one of the computer science and linguistics that studies about the interaction between computer and human natural language (Kumar, 2011). Natural language is a language that we use everyday which are growing and evolving to meet the needs of human communication. Natural language is not designed as a language to be used in the process of computing, so, it does not concern to the possible trouble that might arise in the processing. As the result, natural language processing is more difficult than artificial languages. The main characteristics that make difficulty for natural language processing are (Utami and Hartati, 2007a, b):

Corresponding Author: Ema Utami, Department of Informatics Engineering, Universitas Amikom, Jl. Ring Road Utara, Condong Catur, Depok, Sleman, Daerah Istimewa, 55281 Yogyakarta, Indonesia

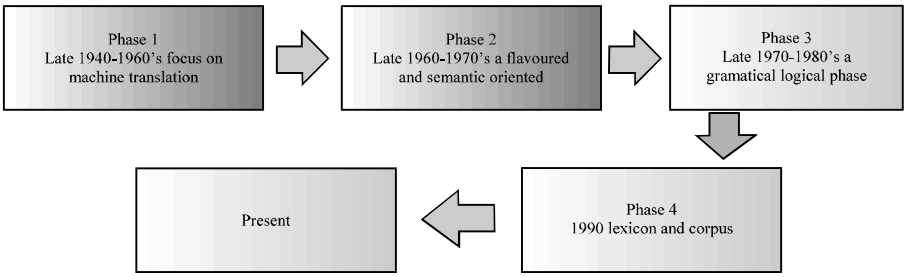

Fig. 1: History phases of NLP

- Ambiguities often appear in natural language
- The huge number of vocabulary in natural language that continuously developing from time to time language is very large and the expand from time to time

The development of NLP itself is a long time history, starting from the end of 1940's and generally can be grouped into 4 phases (Fig. 1). The first phase (from the end of 1940's to the end of 1960's) was focused on Translation Machine (MT), the second phase (from the end of 1960 an to the end of 1970's) was furnished with Artificial Intelligence (AI) with a great emphasis on knowledge and its role in semantic oriented development and manipulation, the third phase (from the end of 1970's to the end of 1980's) focused on computational grammar, and third phase (starting from 1990's) focused on the lexical and corpus studies (Jones, 1994).

NLP in each four phases have been progressing rapidly in line with the technological advances that occurred. One factor that has a major influence is amount of output data from the power of the internet as I have mentioned earlier. Nowadays, NLP is no longer just dealing with formal speech or writing as in the official documents, news, journals, speeches, conversations and other traditional communication. NLP now have to deal with new forms of communication, especially in the era of social media is much more diverse.

Social media such as Facebook, Twitter, Path, Line, and many others constitute software that can be used for communication recently. Such social media including their types of data such as text, graphics and sounds give contribution in creating bigdata. Researches in NLP in terms of text in social media constitute a new field that needs to adapt the existing traditional NLP methods in such a way that it can adjust to these types of text or even creating new methods more suitable in order to be able to extract the whole information or others (Farzindar and Inkpen, 2015).

MATERIALS AND METHODS

Machine translator: In computer sciences, translation machine constitutes one part of linguistics computation having points of study related to the problem of translating one language into another from both text and oral sources. The implementation of technology in the written and oral language communication has many differences. The oral language system to the larger extent, deals with speech recognition, development of text-to-speech synthesis system, dialogue modeling and so on whereas written language system to the greater extent, deals with written symbol processing. These written symbols were then modeled by Chomsky (1956) to represent a language. Several studies, conducted by the researcher has been linked with the discussion of a written language like this, namely the manufacture of machine-readable code (Utami and Hartati, 2008), the examination of writing (Utami *et al*., 2009), splitting string Latin (Utami *et al*., 2009), grammar checking (Utami and Hartati, 2007a, b), transliteration machine (Utami *et al*., 2014), non-verbal conversation machine (Utami and Hartati, 2007a, b) and machine translation (Utami and Hartati, 2007a, b).

The recent translation machines have shown some progression and have played some significant roles in providing supports in many fields. Computer history notes that the translation machine development has also been influenced by politics and economics fields of study, especially in terms of selecting the languages being translated. United States (US) for example, started to concentrate to use machines to translate Russian language into English in 1950's because during that era, US government had paid serious attention to the science and technology development in Soviet Union (Hutchins, 1982). Georgetown University in 1952 with the help of the United States government began the development of a machine translator to do the translation into English of the Russian language, called the Georgetown Automatic Translation (GAT) (Slocum, 1985). The history of modern translation machine has started from this development of the translation machine.

The quality of translation output by the translation machine where the semantic meaning in the target language is concurrent with the original meaning in the source language, constitutes the main objective of utmost

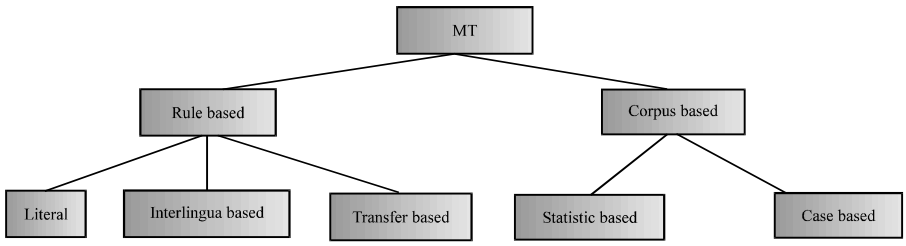

Fig. 2: Translation methods

importance. Many factors influence a translation result with good accuracy, one of which is the method that used to perform the translation process. In general, the methods used in translation machines can be grouped into 2 models, namely rule-based method and corpus-based method (Fig. 2).

Furthermore, the rule-based method can be broken down into 3 models, namely literal translation, interlingua-based and transfer-based models while the corpus-based method can be broken down into 2 models, namely statistics-based and case-based models (Li, 2013). Research example about machine translation of English into Indonesian that use rule-based methods is done by Utami and Hartati (2007a, b) while the statistical methods is done by Adji *et al*. (2011), Manurung and Tanuwijaya (2012).

The increasing number of the internet usage is also, a factor that triggers the development and changes in machine translation. The easy of doing translation, the greater number of languages easily to be translated and capability of automatic translation from many types of documents are some examples of the changes. The phenomenal use of internet has changed many things related to the development of translation machine that has been unprecedented before. The development of translation machine which has affordable price and can be accessed on-line has made translation machine into a product that can be used daily by means of personal computer (Vasconcellos and Miller, 1996).

Big companies involved in internet business have captured these changes as opportunities and started to compete in developing translation machines by adjusting to the prevailing conditions such as through the changes in methods being used. For example, Alta Vista, one of internet search engines of early generation has started to develop an internet network-based translation machine. In cooperation with SYSTRAN Software, Alta Vista has developed translation machine babel fish in 1997 and for the first time offered concept of on-line translation for free (Yang and Lange, 1998). SYSTRAN itself is a translation machine using rule based method and is considered as the pioneer in developing modern translation machine.

Yahoo that has taken over Alta Vista including Babel Fish within in 2003, started to make changes in methods being used in translation machine in 2012. The statistics-based translation machine developed by Microsoft, i.e., Bing Translator (https://www.bing.com/translator) was adopted to replace the existing translation machine. Similarly, Google with Google Translate (GT) (https://translate.google.com) which up to 2016 has been able to translate 103 languages also make adjustments to the method used in the machine translator (https://translate.googleblog.com/2016/04/ten-years-of-google-translate.html).

In addition to using SYSTRAN system in the early development of translation machine, Google has also adopted the concept of on-line and free translation. In 2006, Google started to develop its own statistics-based translation machine to replace the SYSTRAN-based system (https://research.googleblog.com/2006/04/statistical-machine-translation-live.html). In 2013 a machine translator developed by Google has been used by about 200 million users per day (http://www.cnet.com/news/google-translate-now-serves-200-millionpeople-daily/). In addition to those developed by search engines, some other translation machines that can be found on-line and free of charge are among other Wordlingo, SDL free translation and ProMT.

A translation system that is able to transform one natural language into other natural language as the target language in both text and oral forms will of course give some contribution in various fields. Translating texts from the source language into target language will follow some processes conducted by lexical analysis (scanner), syntax analysis (parser), semantic analysis (translator) and pragmatic analysis (evaluator) before gaining the final outputs in the forms of texts in target language. Lexical analysis is to examine the textual forms and to group them into lists of token which are then forwarded to the next processes done by syntax analysis (Utami and Hartati, 2007a, b).

Syntax analysis is to trace the tokens for comparison to the available lists of token and then adjusted or matched to the existing grammatical rules. The process of syntax analysis is to analyze sentence syntactical structures by taking grammatical context into account. The next process is done when the grammatical rules are fulfilled. The next process is semantic analysis to depict the meaning of the texts. The process of semantic analysis

make use of knowledge on semantic meaning and linguistics structure such as the roles of noun or transitive verbs. This process is to test the semantic consistency as well. The output of semantic analysis will be forwarded to pragmatic analysis to display the translation output taking form of target language texts (Utami and Hartati, 2007a, b).

In terms of translation machine, the problem of doing parsing is a very crucial issue. When there are many grammatical rules, then the parsing process takes a lot of time. Time saving can be done by using main grammatical rules and assisted by other methods such as using statistics to give a rapid translation output. On the other hand, this may result in sentence output of translation machine unfitting with the true meaning.

Transliteration machine: In addition to semantic (meaning) issues from one language to another, the differences in alphabetic scripts or symbols used by source and target languages of translation process may result in distinctive problems. The process of alphabetic adjustment shall be done in order to solve this problem, and this process is called as transliteration. Transliteration is defened in the terminological dictionary of philology (1977) as the replacement of alphabetic script one by one from onealphabet to another without considering the phonetic sound of the alphabetic symbol or called as alphabetic conversion.

The transliteration is a systematic methods to convert (transform) character of a font or speech sound (phonetic sound) into additional fonts (Arms, 2000). Similarly, Haizhou *et al*. (2004) described that transliteration is a process that uses language source code as input and generates code the target language as output.

The transliteration has been used in a translation machine system, since, the very beginning of translation machine development (Deselaers *et al*., 2009). The research by Knight and Graehl (1997) with the aim to transliterate Latin text of proper names and terms in English to Japanese characters (Katakana) use statistical-based methods, this study is an early study transliteration machine problems (Deselaers *et al*., 2009). There are many methods in literatures that can be used to do the transliteration process from rules-based model to statistics based model (Kumaran and Kellner, 2007). Transliteration models can be categorized into three models graphme based where transliteration is done without pay attention to the pronounciation of the language that used in part or whole documents that will be transliterated phoneme based where transliteration is done by paying attention to the pronounciation of the language used in part or whole documents to be transliterated combination of both.

Recently, the development of transliteration machine has not been as fast as that of translation machine most likely because the use of transliteration machine is not as much as that of translation machine. Transliteration process itself has it own complexity that shall be resolved requiring the knowledge on the target alphabet having the alphabetic scripts different from the source alphabet. Other issues potentially existing during the transliteration processes are dependent upon the natural language texts being used, among other: hyphenation, punctuation rules, capital rules, compound words, diacritics and special scripts (Lagally, 2004).

Transliteration of Latin into Javanese alphabets, for example, requires two main processes, namely, checking and breaking down Latin strings hyphenation patterns of other Latin strings as well as the conversion of the Latin string hyphenation patterns into certain Javanese alphabetic characters (Utami and Hartati, 2008). Javanese alphabetic characters or scripts have their own uniqueness making the transliteration process not that simple. One example is the scripting of consonant in a word will be different when it is written in two words or phrases: an example is the word 'tempe', letter 't' at the or character 't' at the initial position will be different from that in word 'tumbas tempe' (Utami and Hartati, 2008). There are other problems making the transliteration process from Latin alphabet into Javanese characters more complicated, among other, not all Javanese characters are written to the right direction but some of them are added in front or at the top or below other characters, according to it syntax and scripting rules.

JawaTEX (http://www.jawatex.org) constitutes a contribution by the researcher to the field of transliteration machine. JawaTEX has a main function of mapping the Latin characters to Javanese characters. This transliteration machine is built using three methods. The context free recursive sescent parser to conduct parsing in order to produce the hyphenation patterns of Latin strings that have been made (177 models of Latin string hyphenation patterns) to make the mapping processes into Javanese characters easy that have been adjusted to the syntax and scripting rules of Javanese alphabet where the Latin string hyphenation is to accommodate spatial solving to avoid the problem of ambiguities, considering that Javanese alphabet do not accommodate spatial placement to hyphenate words

because the Javanese alphabet do not adopt spatial rules; rule-based model to hyphenate Latin strings pattern matching to structurally map the Latin string hyphenation patterns into the JawaTEX format in order to obtain the relevant syntax and scripting patterns of Javanese alphabet/characters.

Several other studies that related to machine transliteration is Wan and Verspoor (1998) who examined the text transliteration Latin name in English to a Chinese character (Al-Onaizan and Knight, 2002) who examine the names in Arabic script transliteration Latin text into English using finite state machines, AbdulJaleel and Larkey (2003) who studied Latin text transliteration English to Arabic script-based statistics, Oh and Choi (2002) who studied text transliteration in English to the Korean alphabet and Kumaran and Kellner (2007) who studied manufacture generic framework transliteration machine.

Google also has made the project of making machines for transliteration (https://www.google.com/transliterate) but it was discontinued in 2011 and was replaced with software building-up project called Google input. This software function as a virtual keyboard to write different forms of characters that can be used online (https://www.google.com/inputtools/try) and offline by downloading the application. Beside being a keyboard, this software complete with the ability to transliterate. The transliteration process carried out by Latin text entered by the user and then Google will provide an alternative script of the expected language based on knowledge of its character dictionary. Then, users are expected to choose a character goal well, it shows that knowledge of script purpose is required to perform the transliteration method. The number of characters in the list Google Software as many as 118 but not all can be used including Javanese alphabet.

RESULTS AND DISCUSSION

Smart data: Data generated by internet media is enormous, one of which is data in text form. The amount of existing data will necessarily requires rapid processing but with accurate results. The presence of technology and the development of computer science can help transforming bigdata into data that having more valuable and usable such as to support a decision system. A process of adding more value into data usually called smart data.

Smart data is a process of conducting the data collection from various sources including bigdata then it is associated and analyzed to be used as bait in the next process or decision making (Iafrate, 2015). Vettor *et al.* (2014) formulates smart data as a function of I(G, Da, Db, ...,)≥Dmix with Dmix as a set of obtained smart data from the function I where G is a semantic graph of the data manipulate while Da, Db is a semantic data transformation extracted from the data source.

Data quality is influenced by many factors, there are 16 dimensions to assessment of data quality, one of them is free from error (Pipino *et al*., 2002). Data error can be caused by miss type, format or other that can lead miss information or decision. The quality of the inputted data is one factor influencing over all data quality.

For example, when we do searching using the Google search engine and we make mistake of spelling words or sentence we are going to search through Google search, then we are going to be offered with some input alternatives we likely intend to input. Through such suggested input revision, Google intends to obtain the most fitted and qualified data input to be used as feeder for next process for producing the best output of searching (Fig. 3).

Similarly, these also prevail in data used in a NLP processing. Before the data being processed in NLP, the data need to be cleansed, corrected and revised. Cleansing, correction and revision of the data are often called as a preprocessing step expected to obtain the most qualified input.

A method is also used to check the input data when the data are going to be stored in a Data Base Management System (DBMS). For example, in a Relational DBMS, validation of data being stored can be filtered by means of applying constraints. Applying the constraints can be done at a table level in such a way that it can be separated from the application being used to access the data. Applying the constraints in a table within a relational DBMS keeps the data valid because the data not fitted with the criteria are rejected (Utami, 2014). The wellstored and highly integrated data are a robust foundation for accurate and precise decision making (Utami and Raharjo, 2014).

Nevertheless, relational DBMS developers shall pay specificant attention to recent big data, considering that not all the data are well-structured, some of them are semi-structured in nature and even not structured at all. Traditional relation DBMS will not be able to handle the volume and heterogeneity of such a big size of data (Chen *et al*., 2014). The development of such data today requires much adjustments in both storing and processing. Method adjustment or new method breakthrough is now urgently needed to keep pace with the existing data expansion and development. One instance is textual data growth resulted from the

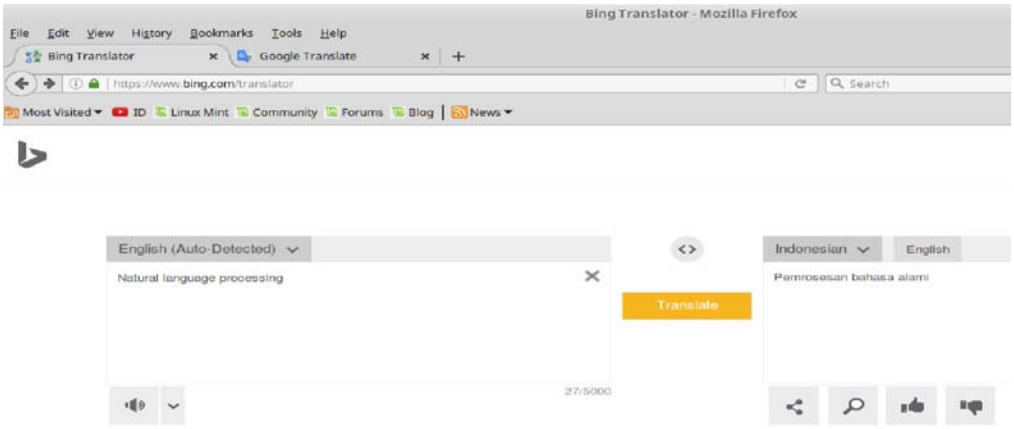

Fig. 3: Google search input suggestion

communication media in the early massive use of mobile phone, the communication through Short Message Service (SMS) is still used with its all limitation. The limited length of text being sent through SMS creates a distinctive linguistic "culture", namely SMS communication language full of abbreviation in order to fulfill the text length quota. Keeping pace with this demanding requirement, there exist some internet-based communication media or often called as Multimedia Internet Messaging Services (MIMs) such as WhatsApp, BlackBerry Messenger, Line, Twitter and so on. These changes in text communication media can be used as a research topics as that conducted by Church and Oliveira (2013) which investigated the differences of perception and purpose on using SMS and WhatsApp with interview and survey method.

Several other studies relating to WhatsApp in recent years grow larger with various topics from variety fields. The study by Johnston *et al*. (2015a, b) in field of medicine who evaluated the use of communication method from WhatsApp in emergency surgical team. Research from another study was conducted by Mahapatra *et al*. (2016) studying the use of MIMs such as WhatsApp to be integrated into Learning Management System (LMS). The discipline of computer security has also been keeping pace with the development as the research focusing on forensic analysis of WhatsApp in Android conducted by Anglano (2014).

Other researches related to MIMs in NLP field of study with the objective to find the extracted information to obtain data as smart data have also been carried out. One example is the use of data from Twitter to map critical areas affected by natural disaster (Middleton *et al*., 2014). The data taken from Twitter during the critical conditions can be used as the resources for decision making (Imran *et al*., 2015).

In addition to text processing for gaining the extracted information, there is also an effort to enable text data to provide sufficient information to be used as smart data. Adding information into a data or often referred as annotation is one of the efforts that can be done. In the text data annotations can provide additional explanation, so, text messages can be delivered more clearly. Giving annotations on the Twitter data relating to the state of crisis is believe will provide benefits, particularly the improvement in NLP automatically on messages related to a crisis situation (Imran *et al*., 2015).

Annotation adding into textual data has also currently been carried out at the state level, namely by collecting the textual data of certain countries language then added with additional information within the textual data. This kind of textual data set is often called as national corpus and there are some countries having established their national corpus such as Czechoslovakia, Hungary, Russia, Greek, Germany, Slovakia, Poland, China, Korea, Japan and some other countries (Xiao, 2007). Indonesia has started to plan the national corpus development but there are not much scientific publication on this. Selecting textual source storage method using a database system has many advantages and can be taken into account for developing Indonesian corpus.

Annotation adding can also be applied to other than textual data such us pictorial data such as in research who did the annotations based on metadata for simplify the classification and research by Glavatskih *et al*. (2015) who did the annotation to the voice corpus for sound recognition. Recently, there are some picture search engines such as Google images, TinEye or Pinterest that have been able to do searching using pictures or images as the feeders (input). This kind of picture searching is often called as reverse image search using the

Content-Based Image Retrieval (CBIR) method where the annotation into images or pictures can be used as one of methods in the process.

The presence of smart data in various forms of media is of course, very beneficial, especially in the decision making process. One of the benefits is the availability of image searching machine that can do searching using several imagery data as the input or feeder in order to trace information originality of the images often shared through social media.

CONCLUSION

Researches on NLP discipline are still widely open, for example, the discipline related to alphabet, not every country in the world has her own distinctive alphabet, Indonesia is one of the countries that has many alphabets. The existences of Javanese, Balinese, Sundanese, Batak alphabets and many other, provide a great opportunity for further researches, especially, related to NLP.

The technological development has also created many innovation and breakthrough related to NLP that provides the opportunity for further studies. For example, utilizing technology and NLP can be applied to give assistance for people with visual disorder that can be aided with speech to text and text to speech program for internet communication such as email writing, sending, and reading (Wan and Verspoor, 1998).

The intention of many cities and regencies in Indonesia to adopt the smart city/smart regency model provides some opportunity for further researches related to NLP. Integrating electronic sensors, data video (closed circuit television), global positioning system and natural language uses will make user communication with machine easier. For example, a question "whether the crossroad at the Jalan Kaliurang Ringroad is jammed or not" can be asked to the smart traffic management system for having information about traffic jam.

REFERENCES

AbdulJaleel, N. and L.S. Larkey, 2003. Statistical transliteration for English-Arabic cross language information retrieval. Proceedings of the 12th International Conference on Information and Knowledge Management, November 03-08, 2003, ACM, New Orleans, Louisiana, USA., ISBN:1-58113-723-0, pp: 139-146.

Adji, T.B., Y. Astuti and S.S. Kusumawardani, 2011. Statistical-based machine translation for prepositional phrase using link Grammar. Proceedings of the 2011 International Conference on Electrical Engineering and Informatics (ICEEI), July 17-19, 2011, IEEE, Bandung,Indonesia,ISBN:978-1-4577-0753-7,pp:1-6.

Al-Onaizan, Y. and K. Knight, 2002. Machine transliteration of names in Arabic text. Proceedings of the ACL-02 Workshop on Computational Approaches to Semitic Languages, July 11, 2002, Association for Computational Linguistics, Philadelphia, Pennsylvania, pp: 1-13.

Anglano, C., 2014. Forensic analysis of WhatsApp messenger on android smartphones. Digital Invest., 11: 201-213.

Arms, W.Y., 2000. Digital Libraries. The MIT Press, Cambridge, Massachusetts,.

Chen, M., S. Mao and Y. Liu, 2014. Big data: A survey. Mobile Networks Applic., 19: 171-209.

Chomsky, N., 1956. Three models for the description of language. IRE Trans. Inform. Theor., 2: 113-124.

Church, K. and D.R. Oliveira, 2013. What's up with Whatsapp?: Comparing mobile instant messaging behaviors with traditional SMS. Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services, August 27-30, 2013, ACM, New York, USA., ISBN:978-1-4503-2273-7, pp: 352-361.

Deselaers, T., S. Hasan, O. Bender and H. Ney, 2009. A deep learning approach to machine transliteration. Proceedings of the 4th Workshop on Statistical Machine Translation, March 30-31, 2009, Association for Computational Linguistics, Stroudsburg, Pennsylvania, USA., pp: 233-241.

Fan, W. and A. Bifet, 2013. Mining big data: Current status and forecast to the future. ACM. SIGKDD. Explor. Newsl., 14: 1-5.

Farzindar, A. and D. Inkpen, 2015. Natural language processing for social media. Synth. Lect. Hum. Lang. Technol., 8: 1-166.

Glavatskih, I., T. Platonova, V. Rogozhina, A. Shirokova and A. Smolina *et al*., 2015. The Multi-Level Approach to Speech Corpora Annotation for Automatic Speech Recognition. In: Speech and Computer, Ronzhin, A., R. Potapova and N. Fakotakis (Eds.). Springer, Berlin, Germany, ISBN:978-3-319-23131-0, pp: 438-445.

Haizhou, L., Z. Min and S. Jian, 2004. A joint source-channel model for machine transliteration. Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, July 21-26, 2004, Association for Computational Linguistics, Stroudsburg, Pennsylvania, USA., pp: 159-166.

Hutchins, W.J., 1982. The Evolution of Machine Translation Systems. In: Practical Experience of Machine Translation, Lawson, V. (Ed.). North-Holland Publishing Company, Amsterdam, Oxford, pp: 21-37.

Iafrate, F., 2015. From Big Data to Smart Data. Vol. 1, John Wiley & Sons, Hoboken, New Jersey, USA., ISBN:978-1-84821-755-3,.

Imran, M., C. Castillo, F. Diaz and S. Vieweg, 2015. Processing social media messages in mass emergency: A survey. ACM. Comput. Surv. CSUR., 47: A1-A36.

Johnston, J., L. Ballan and L. Fei-Fei, 2015a. Love thy neighbors: Image annotation by exploiting image metadata. Proceedings of the IEEE International Conference on Computer Vision, August 30, 2015, IEEE, New York, USA., pp: 4624-4632.

Johnston, M.J., D. King, S. Arora, N. Behar and T. Athanasiou *et al*., 2015b. Smartphones let surgeons know WhatsApp: An analysis of communication in emergency surgical teams. Am. J. Surg., 209: 45-51.

Jones, K.S., 1994. Natural Language Processing: A Historical Review. In: Current Issues in Computational Linguistics: In Honour of Don Walker, Zampolli, A., C. Nicoletta and P. Martha (Eds.). Springer, Berlin, Germany, ISBN:978-0-7923-2998-5, pp: 3-16.

Knight, K. and J. Graehl, 1997. Machine transliteration. Proceedings of the 8th International Conference on European Chapter of the Association for Computational Linguistics, July 07-12, 1997, Association for Computational Linguistics, Stroudsburg, Pennsylvania, USA., pp: 128-135.

Kumar, E., 2011. Natural Language Processing. I.K. International Publishing House Pvt., New Delhi, India, Pages: 201.

Kumaran, A. and T. Kellner, 2007. A generic framework for machine transliteration. Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, July 23-27, 2007, ACM, New York, USA., ISBN:978-1-59593-597-7, pp: 721-722.

Lagally, K., 2004. ArabTeX: Typesetting Arabic and Hebrew. Master Thesis, University of Stuttgart, Stuttgart, Germany.

Li, P., 2013. A survey of machine translation methods. Indonesian J. Electr. Eng. Comput. Sci., 11: 7125-7130.

Mahapatra, J., S. Srivastava, K. Yadav, K. Shrivastava and O. Deshmukh, 2016. LMS weds WhatsApp: Bridging digital divide using MIMs. Proceedings of the 13th Web for All Conference, April 11-13, 2016, ACM, New York, USA., ISBN:978-1-4503-4138-7, pp: 42-42.

Manurung, H.M. and H. Tanuwijaya, 2012. [Translation English-Indonesian document using machine of statistical translation with word reordering and phrase reordering (In Indonesia)]. J. Comput. Sci. Inf., 2: 17-24.

Middleton, S.E., L. Middleton and S. Modafferi, 2014. Real-time crisis mapping of natural disasters using social media. IEEE. Intell. Syst., 29: 9-17.

Oh, J.H. and K.S. Choi, 2002. An english-korean transliteration model using pronunciation and contextual rules. Proceedings of the 19th International Conference on Computational Linguistics Vol. 1, August 24- September 1, 2002, Association for Computational Linguistics, Stroudsburg, Pennsylvania, USA., pp: 1-7.

Pipino, L.L., Y.W. Lee and R.Y. Wang, 2002. Data quality assessment. Commun. ACM., 45: 211-218.

Slocum, J., 1985. A survey of machine translation: Its history, current status and future prospects. Comput. Ling., 11: 1-17.

Utami, E. and S. Hartati, 2007a. Botqa application to improve the way of human and machine interaction (In Indonesia)]. Nat. Seminar Inf. Technol. Appl., 1: B1-B8.

Utami, E. and S. Hartati, 2007b. [The approach of rule-based method in translating English text to Indonesian text (In Indonesian)]. J. Inf., 8: 42-53.

Utami, E. and S. Hartati, 2008. [Software design for ASCII character recognition from bitmap images using artificial neural networks back propagation method (In Indonesia). J. Technol. Acad. ISTA., 12: 212-222.

Utami, E. and S. Raharjo, 2014. Database security model in the academic information system. Intl. J. Secur. Appl., 8: 163-174.

Utami, E., 2014. The advantages of using CHECK constraints in the academic database table. J. Software, 9: 382-388.

Utami, E., J.E. Istiyanto, S. Hartati and A. Ashari, 2009. Developing transliteration pattern of Latin character text document algorithm based on linguistics knowledge of writing Javanesse script. Proceedings of the 2009 International Conference on Instrumentation, Communications, Information Technology and Biomedical Engineering (ICICI-BME), November 23-25, 2009, IEEE, Bandung, Indonesia, ISBN:978-1-4244-4999-6, pp: 1-6.

Utami, E., J.E. Istiyanto, S. Hartati, M. Marsono and A. Ashari, 2014. [Applying the rule based on the build Transliterator jawatex (In Sundanese)]. Berkala Ilmiah MIPA., 23: 78-94.

Vasconcellos, M. and L.C. Miller, 1996. Recent trends in machine translation. Proceedings of the International Conference on Translating and the Computer, November 14-15, 1996, Association for Information Management, England, UK., pp: 1-10.

Vettor, D.P., M. Mrissa and D. Benslimane, 2014. Models and architecture for smart data management. Master Thesis, Claude Bernard University Lyon 1, Villeurbanne, France.

Wan, S. and C.M. Verspoor, 1998. Automatic English-Chinese name transliteration for development of multilingual resources. Proceedings of the 17th International Conference on Computational Linguistics Vol. 2, August 10-14, 1998, Association for Computational Linguistics, Montreal, Quebec, Canada, pp: 1352-1356.

Xiao, R.Z., 2007. Well-Known and Influential Corpora. In: Handbooks of Linguistics and Communication Science, Steger, H. (Ed.). Mouton de Gruyter, Berlin, Germany, ISBN: 9783110124217, pp: 1174-1174.

Yang, J. and E.D. Lange, 1998. Systran on AltaVista a user Study on Real-time Machine Translation on the Internet. In: Machine Translation in the Americas, Farwell, D., G. Laurie and H. Eduard (Eds.). Springer, Berlin, Germany, ISBN:978-3-540-65259-5, pp: 275-285.