Vietnam Journal of Science and Technology ***58*** *(3) (2020) 344-354*

doi:10.15625/2525-2518/58/3/14744

PROPOSED MODEL OF HANDLING LANGUAGE FOR SMART HOME SYSTEM CONTROLLED BY VOICE

Phat Nguyen Huu*, Khanh Tong Van

School of Electronics and Telecommunications, Hanoi University of Science and Technology
No. 1, Dai Co Viet road, Hai Ba Trung, Ha Noi, Viet Nam

*Email: *phat.nguyenhuu@hust.edu.vn*

Received: 29 December 2019; Accepted for publication: 24 February 2020

Abstract. Voice interaction control is a useful solution for smart homes. Now it helps to bring the house closer to people. In recent years, many smart home-based voice control solutions have been introduced (for example: Google Assistant, Alexa Amazon etc.). However, most of these solutions do not really serve Vietnamese people. In this paper, we study and develop Vietnamese language processing model to apply it to smart home system. Specifically, we propose language processing methods and create databases for smart homes. Our main contribution of the paper is the Vietnamese language processing database for smart home system.

Keywords: VNLP – Vietnamese Natural Language Processing, smart home, signal processing, Google Assistant.

Classification numbers: 4.2.3; 4.5.3; 4.7.4.

1. INTRODUCTION

Language processing is a category in information processing with linguistic data input. In other words, it is text or voice. These data are becoming the main data types of people, and saved electronically. Their common characteristics are non-structured or semi-structured that cannot be saved as tables. Therefore, we need to deal with them to be able to transform from an unknown form into an understandable form. Some applications of natural language processing are such as: Voice recognition, Automatic translation, searching information, extracting information etc. Application of Vietnamese language processing into smart homes is a new field. For a model to handle well and accurately, the system requires the amount of data training to be of quality and realistic.

Nowadays, human needs are increasingly advanced when electronic technology develops. The trend of smart home is becoming popular as the demand for modern and thus comfortable and energy-saving houses gradually becomes a standard. There are many researches and solutions for smart home control by voice [1 - 5]. The authors [1] have come up with solution that combines the language processing on smartphone and IoTs to create a remote control system for voice devices of house. The authors [2] have come up with a solution to use Google Home to recognize and process voice. It sends commands to Raspberry Pi and Raspberry Pi transmits signals to Bluetooth devices to control devices. In [3], the authors used the Support Vector Machine (SVM) classification algorithm to classify monophonic sounds in speech and extracted features to control devices without having processing languages. In [4], the authors proposed several basic concepts of SVM, different function, and parameters selection of SVM. In [5], the authors presented Naïve Bayes (NB) algorithm and concluded that it was able to classify the quality of journals. However, their accuracy is not optimal. Therefore, journal classification using the Naive Bayes Classifier algorithm needs to be optimized with other algorithms.

The goal of integrating technology into home appliances is to easily control, connect via the internet, and automatically do the pre-programmed jobs to create a friendly modern home for a civilized life. Smart home solution that can interact by voice is no longer a strange concept for today's technology era. It really is a useful solution for smart home now and become closer to people, not simple as a machine. Therefore, we propose the construction of an interactive voice smart home system in this paper.

The goal of the paper is to build a smart home system that can control devices such as lights, fans, air conditioners, electric cookers, etc. remotely from the user's voice via the website. Our main contribution in this paper is to build a reference data set (including literal and figurative meanings) for Vietnamese language processing models and programs to support the control of remote devices in smart home. The system has the ability to predict human thoughts based on any command.

2. RELATED WORKS

There are many research works on Vietnamese language processing such as word segmentation studies [6 - 8], and [9]. In the study [7], a combination of dictionary and ngram were used, in which the "ngram model" was trained using Vietnamese treebank (70,000 sentences were separated from). Separating words are an indispensable stage in the preprocessing stage and separating words in Vietnamese is a fairly complicated step. We will give an example of Vietnamese "*Ông già đi nhanh quá*". For this sentence, it can be understood by two meanings: "*Ông già(subject)/đi(verb)/nhanh quá (adverb)*" or "*Ông(subject)/già đi(verb)/nhanh quá (adverb)*". This can lead to ambiguous semantics, and greatly affect the process of teaching machine to understand human language.

The research on eliminating stopwords is mentioned in [10]. Stopwords are words that appear in a sentence or text but do not carry much meaning of that sentence.

Studies on word and sentence classification in Vietnamese are mentioned in [11, 12]. In the study [11] the author used two models, NB and SVM to training data. As a result, the SVM model is higher than NB model with the same amount of data.

3. METHODOLOGY

3.1. Overview

The common language processing process will be as Fig. 1 [13].

Figure 1. Process of common language processing [13].

The raw data are initially pre-processed (cleaned, standardized, etc.) and then extracted. Depending on the purpose, it will extract different characteristics. Then the system will put data into the model for training. It will then perform the evaluation process and give the final result. More details can be seen in [13].

Based on [13], we propose a process for processing Vietnamese language shown in Figure 2. In this model, we use Google's service to convert voice data into text. This service makes language processing process convenient and permit to attain the highest accuracy when building speech recognition model. The function of this block is to convert user voice data into text. Details of the steps taken for the following blocks will be presented in the next section.

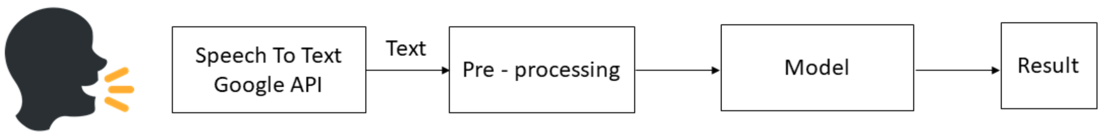

Figure 2. Proposed Vietnamese language processing diagram.

3.2. Pre-processing process

3.2.1. Preprocessing language steps

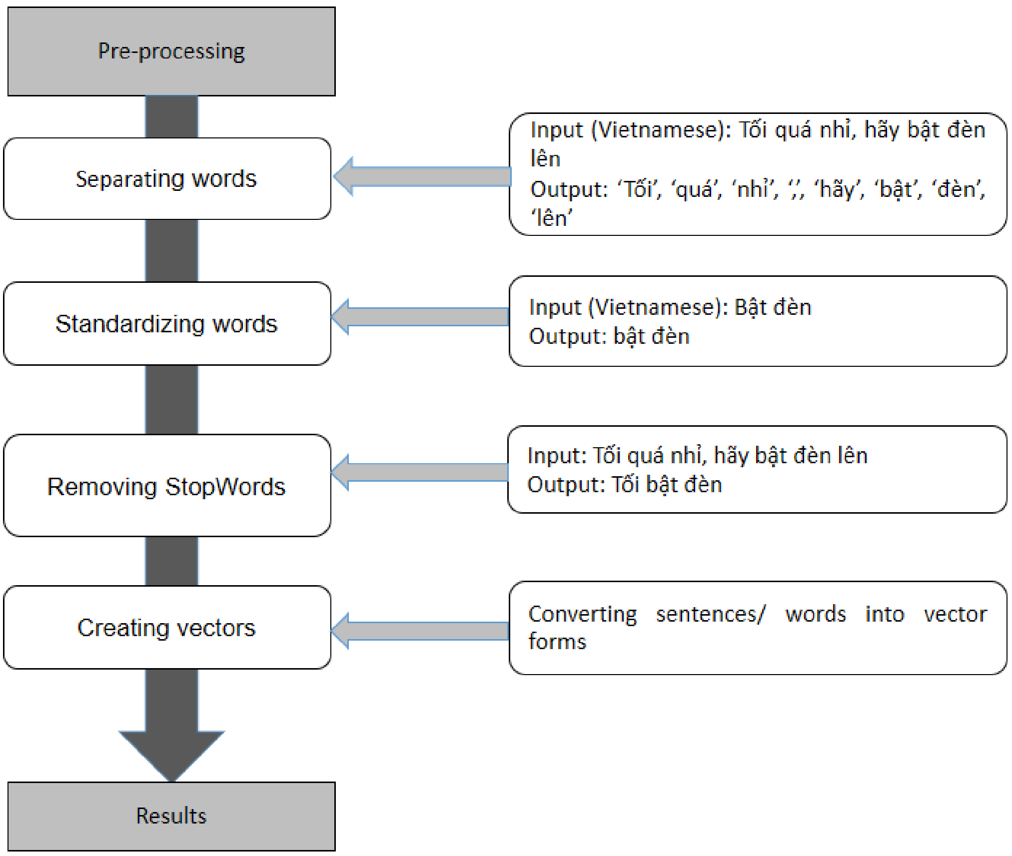

Figure 3. Proposing steps in language preprocessing.

Language preprocessing is an indispensable step in natural language processing. The text is inherently listed without structure. If we keep the original text, the processing is very difficult. Therefore, we will propose preprocessing steps in Vietnamese language processing as shown in Figure 3.

Word segment

Separating word plays an important role to improve accuracy in language processing. A word can have one, two or more ways of dividing syllables into words. Therefore, it causes semantic ambiguity. In this study, we use Vitokenizer () [7] to separate words. For example, we have sentence as " Ôi sao phòng tối thế" and output is then as "Ôi", "sao", "phòng" "tối", "thế".

3.2.2. Removing stopWords

In order to eliminate stopWords effectively for the model, we must prepare a stop-word dataset that is realistic for the purpose of training. Within this paper, we propose a solution to build stop-word data using IF-IDF [14].

The term frequency inverse document frequency (TF-IDF) is a feature extraction technique used in text mining and information retrieval is calculated as follows:

$$idf(t,d) = \log\left(\frac{\text{how many times the term } t \text{ appears}}{\textit{number} \text{ of documents containing the term } t}\right) \quad (1)$$

Based on the calculation of the idf for each word in a sentence, the machine can know which words are less important (small idf) and important (large idf). Therefore, we will remove words with IDF <= threshold.

After building stopwords, we proceed to delete stopwords. For example, if the input is ("ôi", "sao", "phòng" "tối", "thế") then the output is ("phòng", "tối"). Therefore, three words ("ôi", "sao", thế") are stopwords that are removed.

To verify this step, we compared the data set with the algorithm in [15]. The result is shown in Table 1.

Table 1. Table comparing the Vietnamese stop-word data sets with other data sets.

Command	**Expected**	**Our stopwords**		**Others stopwords** Error! Reference source not found.	
		Time	**Actual**	**Time**	**Actual**
Ôi sao phòng tối thế	Phòng tối	0.0022	Phòng tối	0.0210	Phòng tối thế
Hôm nay nóng quá đi	Nóng	0.0027	Nóng	0.0029	Nóng quá đi
Chán quá có phim gì hay không	Phim	0.0020	Phim	0.002	Chán có phim gì

3.2.3. Creating vectors

To create vectors for words, we use the "One-Hot" method [16]. The process of vector formation is as follows:

For example, the following sentence: "Ôi sao phòng nóng thế" (Oh, why is it so hot), the vector of words would be as

"Ôi" [1,0,0,0,0], "sao"[0,1,0,0,0], "phòng"[0,0,1,0,0], "tối"[0,0,0,1,0], "thế"[0,0,0,0,1].

Therefore, the position of the word in a sentence will be 1 and the rest will be 0.

3.2.4. Collecting additional data

For more diverse data, we surveyed nearly 200 figurative sense commands to control the device, including (Commands to turning on / off the light, commands to turning on / off the fan, commands to turning on / off the television) in Fig. 4.

Edit	Copy	Delete	32	nóng thế nhỉ	Bật điều hòa
Edit	Copy	Delete	33	Lạnh quá đi mất	Tắt quạt
Edit	Copy	Delete	34	Ôi sướng thế	Bật quạt
Edit	Copy	Delete	35	nóng quá đi mất thôi	Bật quạt
Edit	Copy	Delete	36	Ôi trời phòng khách tối thế	Bật đèn phòng khách
Edit	Copy	Delete	37	ôi sao hôm nay nóng thế	Bật quạt
Edit	Copy	Delete	38	sao phòng tối vậy	Bật đèn phòng khách
Edit	Copy	Delete	39	chán quá nhỉ	Bật Tivi
Edit	Copy	Delete	40	chán thế	Bật Tivi
Edit	Copy	Delete	41	Sao trong này tối om vậy	Bật đèn phòng khách
Edit	Copy	Delete	42	Sáng quá	Tắt đèn phòng khách
Edit	Copy	Delete	43	Trời ơi chả nhìn thấy gì cả	Bật đèn phòng khách
Edit	Copy	Delete	44	Tivi bây giờ có gì hay không nhỉ	Bật Tivi
Edit	Copy	Delete	45	Nóng thế nhỉ	Bật quạt
Edit	Copy	Delete	46	Trời ơi lạnh quá	Tắt quạt
Edit	Copy	Delete	47	Trời ơi nóng quá	Bật quạt
Edit	Copy	Delete	48	Tối thui vậy	Bật đèn phòng khách
Edit	Copy	Delete	49	Tối om thế này	Bật đèn phòng khách
Edit	Copy	Delete	50	Tối om thế này không nhìn thấy gì	Bật đèn phòng khách
Edit	Copy	Delete	51	Tối quá không nhìn thấy gì	Bật đèn phòng khách
Edit	Copy	Delete	52	Tối quá đèn đâu rồi	Bật đèn phòng khách
Edit	Copy	Delete	53	Tối quá công tắc đèn đâu nhỉ	Bật đèn phòng khách
Console	Copy	Delete	54	Bật đèn lên xem nào	Bật đèn phòng khách

Figure 4. Result of collecting additional data.

3.3. Training

With training data for 6 Vietnamese actions as "Bật đèn phòng khách", "Tắt đèn phòng khách", "Bật quạt", "Tắt quạt", "Bật tivi", "Tắt tivi", we get the results as in Table 2.

Discussion: With the results received, we see two models to predict the intent of sentence. However, the SVM model is more accurate. Besides, accuracy also depends on a lot of data training. In the future, we will try to improve the data training to achieve the highest accuracy.

Due to the small amount of data but many features, we chose the SVM model [4] to train the data. In this article, we train for 6 actions, namely "Bật đèn phòng khách" (Turn on the living room lights), "Tắt đèn phòng khách" (Turn off the living room lights), "Bật quạt" (Turn on the fan), "Tắt quạt" (Turn off the fan), "Bật tivi" (Turn on the TV), "Tắt tivi" (Turn off the TV). Details of the assessed results are shown in the following section.

Table 2. Result of SVM and NB models.

Command	SVM Model		NB Model	
	Accuracy	Target	Accuracy	Target
Hãy bật đèn phòng khách lên	0.8954	Turn on the living room lights	0.8125	Turn on the living room lights
Tắt đèn phòng khách đi nào	0.8896	Turn off the living room lights	0.7956	Turn off the living room lights
Bật quạt lên đi nào	0.8973	Turn on fan	0.8354	Turn on fan
Tắt quạt đi nào	0.8795	Turn off fan	0.8025	Turn off fan
Bật tivi lên xem phim nào	0.8965	Turn on TV	0.8276	Turn on TV
Hãy tắt tivi đi	0.8868	Turn off TV	0.8375	Turn off TV

4. RESULTS AND DISCUSSION

To test the language processing algorithm, we performed with 2 sets of Vietnamese and English dictionaries. The results shown are based on the evaluation of criteria such as execution time and accuracy.

4.1. Preprocessing process results

4.1.1. Result of word separation

In the word separation algorithm, we use data from Vitokenizer.tokenize () [17]. The results are shown in Table 3.

Table 3. Table of results of Vietnamese word separation.

Command	**Expectation**	**Actual**	**Unittest**
Đi ngủ nào bật đèn ngủ lên	"Đi" "ngủ" "nào", "bật", "đèn", "ngủ" "lên"	"Đi" "ngủ" "nào", "bật", "đèn", "ngủ" "lên"	OK (0.001s)
Bật đèn phòng khách lênh nào em ơi	"Bật", "đèn", "phòng" "khách", "lênh", "nào", "em", "ơi"	"Bật", "đèn", "phòng" "khách", "lênh", "nào", "em", "ơi"	OK(0.001s)
Nóng quá bật quạt lên nào	"Nóng", "quá", "bật", "quạt", "lên", "nào"	"Nóng", "quá", "bật", "quạt", "lên", "nào"	OK(0.001s)
The room so hot man	"The", "room", "so", "hot", "man"	"The", "room", "so", "hot", "man"	OK(0.001s)
Evaluation			100%

4.1.2. Stop-word removal results

Results of stop-word removal are shown in Table 4.

Table 4. Results table of Vietnamese stop-words removal.

Command	**Expectation**	**Actual**	**Unittest**
“Đi” “ngủ” “nào”, “bật”, “đèn”, “ngủ” “lên”	“bật”, “đèn”, “ngủ”	“bật”, “đèn”, “ngủ”	OK(0.001s)
“Bật”,“đèn”,“phòn”“khách”,“lênh”,“nào”, “em”,“ơi”	“Bật”,“đèn”,“phòng”, “khách”	“Bật”,“đèn”,“phòng” ,“khách”,	OK(0.001s)
“Nóng”,“quá”,“bật”, “quạt”,“lên”, “nào”	“Nóng”,“quá”,“bật”,“quạt”	“Nóng”,“quá”,“bật”, “quạt”	OK(0.001s)
Evaluation			100 %

Discussion: The above results are evaluated in an objective manner by Unittest [18] as shown in Fig. 5. Although the above assessment is not entirely accurate because of the small amount of input test data, it is sufficient to conclude that using Vitokenizer () to separate words and stop-word sets for smart home is effective. It will help train the model to achieve the best results.

4.1.3. Training results using SVM

We continue to experiment with two sets of English and Vietnamese data for different emotions. Judging by 6 corresponding emotions for the above 6 actions, we obtained the following results:

For the English data set, we have the following results as shown in Tabs. 5 and 6.

Table 5. Results of testing 10 different statements related to hot emotions by English.

No.	**Command**	**Predict rate**	**Target**
1	Oh, so hot man	0.8253	Turn on the fan
2	Too hot	0.8252	Turn on the fan
3	The weather so hot	0.8256	Turn on the fan
4	Oh my god how too hot	0.8254	Turn on the fan
5	Hot sweating	0.8251	Turn on the fan
6	Too hot turn the fan on please	0.7327	Turn on the fan
7	Oh my god the room so hot	0.8251	Turn on the fan
8	Hot like a sexy girl	0.8251	Turn on the fan
9	I feel hot like standing outside	0.8256	Turn on the fan
10	Turn on the fan please	0.8279	Turn on the fan
Average			0.8163

Table 6. Results of testing 10 different statements related to dark emotions by English.

No.	Command	Predict rate	Target
1	Too dark	0.8211	Turn on the living room lights
2	The living room so dark	0.8581	Turn on the living room lights
3	So dark turn on the light please	0.8918	Turn on the living room lights
4	Oh my god so dark	0.8214	Turn on the living room lights
5	so dark I can't see anything	0.8213	Turn on the living room lights
6	Turn on the living light please	0.8242	Turn on the living room lights
7	It's seem like too dark	0.8217	Turn on the living room lights
8	Why the living room so dark	0.8585	Turn on the living room lights
9	How the living room dark	0.8585	Turn on the living room lights
10	Why don't you turn the living light on	0.8232	Turn on the living room lights
Average			0.8399

For the Vietnamese dataset, the results are shown in the following Tabs. 7, 8, 9, 10, 11, and 12.

Table 7. Table of training results related to hot emotions by Vietnamese.

No.	Commands	Predict rate	Target
1	Ôi sao nóng quá nhỉ	0.9238	Turn on the fan
2	Nóng quá đấy	0.9246	Turn on the fan
3	Trời sao nóng thế	0.9049	Turn on the fan
4	Nóng không chịu nổi	0.9056	Turn on the fan
5	Trời oi bức thể nhỉ	0.8765	Turn on the fan
6	Nóng toát mồ hôi	0.9042	Turn on the fan
7	Phòng nóng như cái lò	0.8455	Turn on the fan
8	Sao phòng nóng thế	0.8438	Turn on the fan
9	Phòng nóng thế này sao chịu được	0.8426	Turn on the fan
10	Nóng quá đi bật quạt lên nào	0.9716	Turn on the fan
Average		0.8943	

Table 8. Table of training results related to cold emotions by Vietnamese.

No.	Commands	Predict rate	Target
1	Ôi sao lạnh quá nhỉ	0.9164	Turn off the fan
2	Lạnh quá đấy	0.9162	Turn off the fan
3	Trời sao lạnh thế	0.8949	Turn off the fan
4	Lạnh không chịu nổi	0.8936	Turn off the fan
5	Trời lạnh thể nhỉ	0.8944	Turn off the fan
6	Lạnh run người	0.8939	Turn off the fan
7	Phòng lạnh thế	0.8210	Turn off the fan
8	Sao phòng lạnh thế	0.8209	Turn off the fan
9	Phòng lạnh thế này sao chịu được	0.8213	Turn off the fan
10	Lạnh quá đi tắt quạt lên nào	0.9663	Turn off the fan
Average		0.8389	

Table 9. Results of training action on lights.

No.	Commands	Predict rate	Target
1	Ôi sao tối quá nhỉ	0.9059	Turn on the light
2	Trời sao tối thế	0.8918	Turn on the light
3	Tối om thế này không nhìn thấy gì	0.8919	Turn on the light
4	Trời nay tối sớm thế	0.8919	Turn on the light
5	Tối quá em ơi	0.9058	Turn on the light
Average		0.8974	

Table 10. Results of training on turning off lights by Vietnamese.

No.	Commands	Predict rate	Target
1	Ôi sao sáng quá nhỉ	0.9012	Turn off the light
2	Trời sáng rồi	0.8983	Turn off the light
3	Sáng lắm rồi	0.9124	Turn off the light
4	Phòng sáng quá	0.8872	Turn off the light
5	Sáng rồi em ơi	0.8743	Turn off the light
Average		0.8946	

Table 11. Results of training action on television by Vietnamese.

No.	Commands	Predict rate	Target
1	Chán quá nhỉ có gì hay ho không	0.8406	Turn on the TV
2	Hôm nay tivi có chương trình gì không nhỉ	0.8401	Turn on the TV
3	Tivi bây giờ có gì hay không nhỉ	0.8404	Turn on the TV
4	Không biết có phim gì hay không ta	0.8379	Turn on the TV
5	Không có gì xem à	0.7680	Turn on the TV
Average		0.8254	

Table 12. Results of training action to turn off the TV by Vietnamese.

No.	Commands	Predict rate	Target
1	Hết thứ để xem rồi	0.7467	Turn off TV
2	Không xem tivi đâu	0.8326	Turn off TV
3	Tắt tivi đi nào	0.9403	Turn off TV
Average		0.8400	

5. CONCLUSIONS

In this paper primarily conducted a study of language processing to apply it to smart home system, we have achieved some results as follows:

- Proposed solutions to smart home control by voice through emotional commands,
- Completing the data processing language through emotions exclusively for smart home,
- Application of SVM algorithm in text classification for predictive results over 80%,
- Running experimental tests of control commands on Raspberry Pi 3 embedded computer successfully.

However, the remaining problem is that the proposed model does not recognize the non-control statements. Therefore, in the future, we will further improve the system structure and machine learning ability and expand more actions to control the device.

Acknowledgements. This research was supported by Hanoi University of Science and Technology and Ministry of Science and Technology under the project No. B2020-BKA-06, 103/QD-BGDT signed on 13/01/2020.

REFERENCES

1. Chen Y. P. and Rung C. C. - Voice recognition by Google Home and Raspberry Pi for smart socket control, 10^{th} International Conf. on Advanced Computational Intelligence (ICACI), Xiamen, 2018, pp. 324-329.

2. Karan G. B., Kumar D., Pai K., and Manikandan J. Manikandan - Design of a phoneme based voice controlled home automation system, IEEE International Conf. on Consumer Electronics-Asia (ICCE-Asia), Bangalore, 2017, pp. 31-35.
3. Aml A. A. and Mohamed S. M. - Applying voice recognition technology for Smart home networks, International Conf. on Engineering & MIS (ICEMIS), Agadir, 2016, pp. 1-6.
4. Durgesh K. S. and Lekha B. - Data classification using support vector machine, Journal of Theoretical and Applied Infor. and Technol. **12** (2010) 1-7.
5. Wibawa A., Kurniawan A., Murti D., Adiperkasa R., Putra S., Kurniawan S., and Nugraha Y. - Naïve Bayes Classifier for Journal Quartile Classification, International J. of Recent Contributions from Engineering, Scie. & IT (iJES) **7** (2019) 91.
6. Dien D., Hoang K., and Toan N. V. - Vietnamese Word Segmentation, in Proc. of the Sixth Natural Language Proc. Pacific Rim Symp., Tokyo, Japan, 2001, pp. 749-756.
7. Phuong L. H., Huyen N. T. M., Azim R., Vinh H. T. - A Hybrid Approach to Word Segmentation of Vietnamese Texts, Lecture Notes in Computer Scie., Springer **5196** (2008) 240-249.
8. Trung T. V. - Python Vietnamese Toolkit, Version 1 [Online], viewed 20 July 2019 from: < https://libraries.io/github/trungtv/pyvi>.
9. Song N. D. C., Quoc H. N., and Rachsuda J. - State-of-the-Art Vietnamese Word Segmentation, 2nd International Conf. on Sci. in Infor. Technol. (ICSITech), 2019, pp. 119-124.
10. Al-Shalabi R., Kanaan G., Jaam J. M., Hasnah A., and Hilat E. - Stop-word removal algorithm for Arabic language, Proc. 2004 International Conf. on Infor. and Comm. Technol.: From Theory to Applications, Damascus, Syria, 2004, pp. 545-550.
11. Ha P. T. and Chi N. Q. - Automatic Classification for Vietnamese News), Advances in Computer Science: an International Journal **4** (4) (2015) 545-550.
12. Hoang V. C. D., Dinh D., Nguyen N. L., and Ngo H. Q. - A Comparative Study on Vietnamese Text Classification Methods, 2007 IEEE International Conf. on Research, Innovation and Vision for the Future, Hanoi, 2007, pp. 267-273.
13. Angermueller C., Parnamaa T., Parts L., and Stegle O. - Deep learning for computational biology, Molecular Syst. Biol. **12** (7) (2016) 1-16.
14. Wu H. C., Luk R. W. P., Wong K. F., and Kwok K. L. - Interpreting TF-IDF term weights as making relevance decisions, ACM Trans. on Infor. Syst. **26** (3) (2008) 13.1-13.35.
15. Duyet L. V. - Stopwords/Vietnamese-stopwords, Version 1.0, [Online] viewed 31 August 2019, from: <https://github.com/stopwords/vietnamese-stopwords/blob/master/vietnamese-stopwords.txt>.
16. Xilinx, HDL Synthesis for FPGAs Design Guide -Encoding State Machines, Appendix A: Accelerate FPGA Macros with One-Hot Approach, 1995.
17. Trung T. V. – Vietnamese language model for spacy, Version 2, [Online] viewed 19 October 2019 from: < https://libraries.io/github/trungtv/vi_spacyhttps://pypi.org/project/pyvi/>.
18. Hao N. (2014) – Unit Test, Version 1 [Online] 5 November 2018, from: <https://viblo.asia/p/unit-test-la-gi-maGK7m3Llj2>.