Journal of Computer Science and Cybernetics, V.34, N.4 (2018), 283–294
DOI 10.15625/1813-9663/34/4/13161

VLSP SHARED TASK: NAMED ENTITY RECOGNITION

NGUYEN THI MINH HUYEN[1,*], NGO THE QUYEN[1], VU XUAN LUONG[2], TRAN MAI VU[3], NGUYEN THI THU HIEN[4]

[1] *VNU University of Science;* [2] *Vietlex*

[3] *VNU University of Engineering and Technology*

[4] *Thai Nguyen University of Education*

[*] *huyenntm@hus.edu.vn*

Abstract. Named Entities (NE) are phrases that contain the names of persons, organizations, locations, times and quantities, monetary values, percentages, etc. Named Entity Recognition (NER) is the task of recognizing named entities in documents. NER is an important subtask of Information Extraction, which has attracted researchers all over the world since 1990s. For Vietnamese language, although there exist some research projects and publications on NER task before 2016, no systematic comparison of the performance of NER systems has been done. In 2016, the organizing committee of the VLSP workshop decided to launch the first NER shared task, in order to get an objective evaluation of Vietnamese NER systems and to promote the development of high quality systems. As a result, the first dataset with morpho-syntactic and NE annotations has been released for benchmarking NER systems. At VLSP 2018, the NER shared task has been organized for the second time, providing a bigger dataset containing texts from various domains, but without morpho-syntactic annotation. These resources are available for research purpose via the VLSP website `vlsp.org.vn/resources`. In this paper, we describe the datasets as well as the evaluation results obtained from these two campaigns.

Keywords. CoNLL format; Evaluation; Named entity; Named entity recognition; Shared task, Vietnamese, VLSP workshop.

1. INTRODUCTION

Named entities (NE) are phrases that contain the names of persons, organizations, locations, times and quantities, monetary values, percentages, etc. Named Entity Recognition (NER) is the task of recognizing named entities in documents. NER is an important subtask of Information Extraction, which has attracted researchers all over the world since 1990s.

From 1995, the 6th Message Understanding Conference (MUC) has started evaluating NER systems for English [14]. Besides NER systems for English, NER systems for Dutch and Turkish were also evaluated in CoNLL 2002 [16] and CoNLL 2003 [16] shared tasks. In these evaluation tasks, four named entities were considered, consisting of names of persons, organizations, locations, and names of miscellaneous entities that do not belong to the previous three types. Recently, there have been some competitions about NER organized, for example the GermEval 2014 NER Shared Task[1].

[1] `https://sites.google.com/site/germeval2014ner/home`

For Vietnamese language, although there exist several research projects and publications on NER task before 2016, as in [6, 7, 9, 11, 12, 15], none of these works has resulted in a free/open-source software.

In 2016, the organizing committee of the VLSP workshop decided to launch the first evaluation campaign for Vietnamese NER systems, together with the shared task on Vietnamese sentiment analysis. These shared tasks are important to reach an objective evaluation of natural language processing tools, and to promote the development of high quality systems. As a result, the first dataset with morpho-syntactic and NE annotations has been released for benchmarking NER systems at VLSP 2016, using CoNLL 2003 compatible data format [13]. Three types of entities have been considered for evaluation: person, organization and location. The dataset also contains entities at nested levels. Training data consist of two datasets. In the first dataset in CoNLL format, data contain the information of word segmentation. The information of part-of-speech (POS) and phrase chunk was added by utilizing available tools. The second dataset contains only NE tags in XML format.

At VLSP 2018, the NER shared task has been organized for the second time, providing a bigger dataset containing texts from various domains. The corpus is annotated in XML format, containing only NE tags. The data preprocessing tasks are left to the participant systems.

All the resources built at VLSP 2016 and VLSP 2018 are available for research purpose via the VLSP website `vlsp.org.vn/resources`. In this paper, we describe the datasets as well as the evaluation results obtained from these two campaigns.

The rest of the paper is structured as follows. First, we define the shared tasks, the building of the gold data and the evaluation measures. Then we summarize the methods and discuss about the results of the participating systems. Finally we conclude the paper and propose some future works for Vietnamese NER.

2. TASK DESCRIPTION

2.1. NER-VLSP2016

2.1.1. Task definition

The scope of this first campaign on NER task is to evaluate the ability of recognizing NEs in three types, i.e. names of persons (PER), organizations (ORG), and locations (LOC), given an annotated sentence with manual word segmentation and automatic generated labels in POS tagging and phrase chunking. The nested NEs are taken in account. The dataset should be annotated following the CoNLL 2003 compatible data format [13] with morpho-syntatic information or XML format with only NE tags. Examples are given in Section 2.1.3.

2.1.2. Data collection

Data are collected from electronic news papers published on the web. Three types of NEs compatible with their descriptions in the CoNLL Shared Task 2003 [13] are considered.

Locations

- roads (streets, motorways)

- trajectories
- regions (villages, towns, cities, provinces, countries, continents,dioceses, parishes)
- structures (bridges, ports, dams)
- natural locations (mountains, mountain ranges, woods,rivers, wells, fields, valleys, gardens,nature reserves, allotments, beaches,national parks)
- public places (squares, opera houses, museums, schools,markets, airports, stations, swimming pools,hospitals, sports facilities, youth centers,parks, town halls, theaters, cinemas, galleries,camping grounds, NASA launch pads, clubhouses, universities, libraries, churches,medical centers, parking lots, playgrounds,cemeteries)
- commercial places (chemists, pubs, restaurants, depots,hostels, hotels, industrial parks,nightclubs, music venues)
- assorted buildings (houses, monasteries, creches, mills,army barracks, castles, retirement, homes, towers, halls, rooms, vicarages,courtyards)
- abstract "places"' (e.g. *the free world*)

Organizations

- companies (press agencies, studios, banks, stockmarkets, manufacturers, cooperatives)
- subdivisions of companies (newsrooms)
- brands
- political movements (political parties, terrorist, organizations,
- government bodies (ministries, councils, courts, political unions of countries (e.g. the *U.N.*))
- publications (magazines, newspapers, journals)
- musical companies (bands, choirs, opera companies, orchestras
- public organizations (schools, universities, charities
- other collections of people (sports clubs, sports teams, associations, theaters companies,religious orders, youth organizations

Persons

- first, middle and last names of people, animals and fictional characters, aliases

Here are some NE examples:

- Locations: Thành phố Hồ Chí Minh, Núi Bà Đen, Sông Bạch Đằng.

- Organization: Công ty Formosa, Nhà máy thủy điện Hòa Bình.

- Persons: proper name in "ông Lân", "bà Hà".

An entity can contain another entity, e.g. "Uỷ ban nhân dân Thành phố Hà Nội" is an organization, in which contains a location of "thành phố Hà Nội".

Training data consist of two datasets. In the first dataset, data contain the information of word segmentation. The information of POS and phrase chunks were also added by utilizing available tools. The second dataset is in XML format, containing only NE tags.

2.1.3. Data format

Dataset 1. Data have been preprocessed with word segmentation, POS tagging and phrase chunking, in CoNLL format. The data are structured in five columns, in which two columns are separated by a single space.

- The first column is the word;
- The second column is its POS tag;
- The third column is its chunking tag;
- The fourth column is its NE label;
- The fifth column is its nested NE label.

Each word has been put on a separate line and there is an empty line after each sentence.

NE labels are annotated using the IOB notation as in the CoNLL Shared Tasks. There are 7 labels: B-PER and I-PER are used for persons, B-ORG and I-ORG are used for organizations, B-LOC and I-LOC are used for locations, and O is used for other elements. More concretely, B-XXX is used for the first word of an NE in type XXX, and I-XXX is used for the other words of that NE. The O label is used for words which do not belong to any NE.

One thing to note is that POS tags and phrase chunk tags are determined automatically by publicly available tools, they may contain mistakes.

Dataset 2. Data contain only NE information in XML format.

Example. For example, given the following sentence for input:

Anh Thanh là cán bộ Uỷ ban nhân dân Thành phố Hà Nội.

Then the output could be in CoNLL format or in XML format.

- CoNLL format:

Anh	N	B-NP	O	O
Thanh	NPP	I-NP	B-PER	O
là	V	B-VP	O	O
cán_bộ	N	B-NP	O	O
Uỷ_ban	N	B-NP	B-ORG	O
nhân_dân	N	I-NP	I-ORG	O
Thành_phố	N	I-NP	I-ORG	B-LOC
Hà_Nội	NPP	I-NP	I-ORG	I-LOC
.	.	O	O	O

Table 1. Statistic of NEs in the VLSP2016 corpus

NE Type	Training Data		Test Data	
	First level	**Nested level**	**First level**	**Nested level**
PER	6230	480	1294	7
LOC	1210	1	1377	100
ORG	7478	7	274	0
Total	**14918**	**488**	**2945**	**107**

- XML format:
  Anh ⟨ENAMEX TYPE="PERSON"⟩ Thanh ⟨/ENAMEX ⟩ là cán bộ ⟨ENAMEX TYPE="ORGANIZATION" ⟩ Uỷ ban nhân dân ⟨ENAMEX TYPE="LOCATION"⟩ thành phố Hà Nội ⟨/ENAMEX⟩ ⟨/ENAMEX⟩ .

2.1.4. Annotation procedure

In the framework of this shared task, we choose to make use of the POS tagged dataset published by the VLSP project . Two annotators have worked on the NE labeling with double check.

The initial corpus is separated randomly in a training set and a test set.

The quantities of NEs (first level and nested level) in each set are reported in Table 1. Due to the relatively short time for the corpus annotation, we couldn't ensure a similar distribution of NE types in the training and the test set, as the training set was distributed before the annotation of the test set.

2.1.5. Evaluation measures

The performance of NER systems is evaluated by the F1 score

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{1}$$

where Precision and Recall are determined as follows

$$\text{Precision} = \frac{\text{NE-true}}{\text{NE-sys}}, \tag{2}$$

$$\text{Recall} = \frac{\text{NE-true}}{\text{NE-ref}}, \tag{3}$$

where,

NE-ref: The number of NEs in gold data;

NE-sys: The number of NEs extracted by the system;

NE-true: The number of NEs which is correctly recognized by the system. The results of systems will be evaluated at both levels of NE labels.

Table 2. VLSP2018 NER dataset

Category	Train				Dev				Test			
	PER	**ORG**	**LOC**	**DOC**	**PER**	**ORG**	**LOC**	**DOC**	**PER**	**ORG**	**LOC**	**DOC**
Giáo dục	636	459	596	75	209	163	214	25	84	57	57	9
Giải trí	1086	169	259	75	319	49	95	25	802	167	166	29
KH-CN	204	502	465	75	81	96	184	25	139	245	169	39
Kinh tế	416	1049	896	75	106	376	302	25	298	427	488	51
Nhà đất	-	-	-	-	-	-	-	-	3	24	9	1
Pháp luật	1071	493	822	75	438	248	254	25	342	187	250	15
Thế giới	602	609	1987	75	113	273	726	25	256	76	328	9
Thể thao	1089	878	859	76	426	347	346	25	801	598	281	26
Văn hóa	502	217	1614	90	252	99	468	30	409	63	517	17
Xã hội	392	754	1190	90	158	229	410	30	268	315	218	27
Đời sống	429	59	150	75	66	27	47	25	117	36	45	18
Total	6427	5189	8838	781	2168	1907	3046	260	3519	2195	2528	241

2.2. NER-VLSP 2018

Similarly to the first campaign, the second evaluation campaign for the task of Vietnamese Named Entity Recognition deals with recognizing NEs in three types, i.e. names of persons, organizations, and locations. The annotation procedure and the evaluation measure are equally similar. However, here are some different points:

- No linguistic information is given: the data contain only NE information in XML format (as the dataset 2 in Section 2.1.3;
- The datasets contain documents classified in various domains;
- For each domain, data were divided into three datasets: training, development, and test. Training and development datasets were used to train participating systems. Test dataset was used for the final evaluation purpose;
- The distribution of three NE types in the training, development and test data is comparable;
- A more important quantity of nested NEs is present in the corpus.

Table 2 shows the number of NEs in each dataset.

3. SUBMISSIONS AND RESULTS

3.1. Submissions in NER-VLSP2016

This first NER shared task attracted 10 registered teams. Finally, we had only five teams submitting their results, one of them submitted two systems. Each team provided us with their full report, excepting one just sent us their short description. No team worked on the second dataset (XML format, NE annotation only).

3.1.1. Methods and features

Table 3 gives an overview of the methods and features applied by the submitted systems for detecting the NEs at first level.

Table 3. Methods and Features

Team	Methods	Features
ner1 [2]	Token regular expression + Bidirectional Inference	Basic features (word, pos tag, chunk tag, 2 previous NE tags) Word shapes Basic joint features Regular expression types
ner2 [3]	CRF	word, wordCombination, firstSyllable, lastSyllable, ngrams, initUpcaseWord, allCapWord, letterAndDigitWord, isSpecialCharacter, firstSentenceWord, lastSentenceWord and pos
ner3-1 [10]	Bidirectional Long short term memory (LSTM) – CRF	Head word, pos, chunk tag
ner3-2 [10]	Stack LSTM	
ner4 [8]	CRF/MEM+BS	Current word, pos, word form, context words, is syllable, is in dictionary, regular expression for dates, numbers
ner5	CRF	previous word, current word, next word, pos tag, previous pos tag, next pos tag, chunking tag, previous chunking tag, next chunking tag

For the nested level, only two teams ner4 and ner5 tried to tackle the problem.

3.1.2. Results

As we mentioned above, among six submitted systems only two systems extracted NE at the nested level. However, as the number of entities at this second level is relatively small in the training data as well as in the test set, it is the system performance at the first level that decides its final performance. It is worth mentioning that the result at the nested level of both systems ner4 and ner5 is very poor - it makes decreasing the general performance of these systems.

The F1-score at first level of these systems varies from 78.4% to 88.78%. The results in details of each system are shown in tables 4, 5, 6, 7, 8 and 9.

The comparison of the results of all the systems are reported in Table 10, where systems are ranked by their general F1 score.

In general, all the systems get the best result for the personal names (PER type), then for the locations (LOC type). The result for ORG type is much poorer for all the six systems.

If we look at the results for each NE type as well as for the whole system, the precision score is better than the recall in most of the cases.

Table 4. Result of ner1 system

NE Type	P	R	F1
PER	91.52	94.2	92.84
LOC	86.5	93.54	89.88
ORG	78.95	43.8	56.34
Total	88.36	89.2	88.78

Table 5. Result of ner2 system

NE Type	P	R	F1
PER	92.52	74.57	82.58
LOC	85.79	75.38	80.25
ORG	61.69	34.67	44.39
Total	87.16	71.24	78.4

Table 6. Result of ner3-1 system

NE Type	P	R	F1
PER	94.06	81.99	87.61
LOC	86.52	84.39	85.44
ORG	54.85	47.45	50.88
Total	86.89	79.9	83.25

Table 7. Result of ner3-2 system

NE Type	P	R	F1
PER	90.06	88.95	89.5
LOC	84.82	84.82	84.82
ORG	55.39	41.24	47.28
Total	85.06	82.58	83.8

Table 8. Result of ner4 system

NE Type	P	R	F1
PER	91.74	89.19	90.45
LOC	86.3	81.35	83.75
ORG	61.86	43.8	51.28
Total	87.06	81.3	84.08

3.2. Submissions in NER-VLSP2018

At VLSP 2018, 11 teams have registered and got the training and development datasets for the NER shared task. Finally only 4 teams submitted their results. Among them, three teams submitted their detailed technical reports and the other one sent a short description.

Table 9. Result of ner5 system

NE Type	P	R	F1
PER	88.19	89.41	88.8
LOC	83.01	92.23	87.38
ORG	96.64	52.55	68.09
Total	85.96	87.3	86.62

Table 10. Comparison of F1 score between 6 systems

NE Type	ner1	ner5	ner4	ner3-2	ner3-1	ner2
PER	92.84	88.8	90.45	89.5	87.61	82.58
LOC	89.88	87.38	83.75	84.82	85.44	80.25
ORG	56.34	68.09	51.28	47.28	50.88	44.39
Total	**88.78**	**86.62**	**84.08**	**83.8**	**83.25**	**78.4**

3.2.1. Methods

Table 11 summarizes learning algorithms and features used by the participating systems: NER1 [1], NER2 [4], NER3 [5] and NER4.

The interesting thing is that all the teams make use of CRF models by formalizing the NER as a sequence labeling problem. Two teams combine CRF and LSTM models. The features of sentence segmentation, word segmentation, Brown and word embeddings are used by a majority of participating systems.

Table 11. Features and approaches. SS: sentence segmentation, WS: word segmentation, WE: word embeddings

Team	Model	SS	WS	POS	Subword	Gazetteers	Brown	WE	
NER1	Model 1	x	x	x		-	x		CRF
	Model 2	x	x	x	X	-	x	Glove	LSTM+CRF
	Model 3	x	x	x	X	-	x	Glove	Multi-LSTM
NER2	Model 3	x	x	-	X	-	-	Fastext	BiLSTM+CRF
	Model 4	x	x	-	X	-	-	Glove	BiLSTM+CRF
NER3	Model 1	x	x	-	-	-	x	Glove	CRF
	Model 2	-	x	-	-	-	x	Glove	CRF
	Model 3	x	x	-	-	-	x	Glove	CRF
	Model 4	-	x	-	-	-	x	Glove	CRF
	Model 5	x	x	-	-	-	x	Glove	CRF
	Model 6	-	x	-	-	-	x	Glove	CRF
NER4	Model 1	x	-	-	-	x	-	-	CRF
	Model 2	x	-	-	-	x	-	-	CRF

3.2.2. Results

Tables 12 and 13 summarize results of participating systems by domains and by NE types. The best score for each domain or NE type is colored in red.

In general, the best system comes from the NER3 team, who uses a small number of features and a simple CRF model.

Table 12. NER 2018 results by domains

		CN	GD	GT	KH	KT	ND	PL	TG	TT	VH	XH	DS
NER1	Model 1	54.25	70.84	66.00	60.98	62.48	47.27	71.78	55.40	47.61	49.31	67.95	63.13
	Model 2	45.07	64.64	66.44	53.13	60.91	31.88	69.60	59.12	46.15	50.11	59.60	70.14
	Model 3	55.00	75.68	71.79	67.33	71.82	54.55	75.80	65.34	49.65	59.43	74.15	70.00
	Model 4	50.22	69.27	64.71	61.54	62.85	43.48	68.09	59.38	42.40	51.05	67.74	64.13
NER2	Model 1	65.18	75.07	77.8	66.86	75.24	86.57	79.6	**73.28**	63.49	71.2	73.67	77.72
	Model 2	63.9	72.48	**79.46**	67.4	76.66	**88.24**	79.27	73.23	61.92	73.78	73.66	80.22
	Model 3	**68.72**	73.83	78.17	63.84	76.82	86.57	**79.69**	72.28	63.67	**71.55**	74.52	78.47
NER3	Model 1	65.19	83.5	77.62	74.69	78.85	67.74	76.5	71.14	73.15	67.15	74.3	84.16
	Model 2	65.6	**84.42**	78.27	76.16	78.57	60	76.06	70.75	73.27	67.37	**74.66**	83.68
	Model 3	66.93	83.92	77.68	76.01	**79.21**	68.75	77	71.5	72.23	66.67	74.25	**85.51**
	Model 4	66.41	83.29	78.34	**76.4**	**79.21**	69.7	76.76	71.84	**73.41**	66.88	74.51	84.43
	Model 5	65.02	83.21	77.58	74.92	78.63	67.74	76.42	70.99	73.06	67.15	73.35	84.46
	Model 6	65.43	83.84	78.24	**76.4**	78.14	56.14	75.89	70.6	73.21	67.41	73.72	83.68
NER4	Model 1	31.64	29.79	39.34	42.31	37.56	7.41	35.02	45.30	32.82	26.15	17.26	39.66
	Model 2	23.61	30.27	43.41	33.43	35.20	16.13	37.71	42.28	33.24	26.34	20.14	32.81

Table 13. NER 2018 results by NE types

Team	Model	PERSON			LOCATION			ORGANIZATION			OVERALL		
		P	R	F	P	R	F	P	R	F	P	R	F
NER1	Model 1	70.54	63.29	66.72	76.67	56.00	64.72	59.24	28.18	38.19	70.48	51.56	59.56
	Model 2	65.62	63.27	64.42	72.69	53.32	61.52	53.17	31.45	39.52	65.20	51.68	57.66
	Model 3	79.26	63.06	70.24	82.81	65.26	73.00	73.61	35.98	48.33	79.46	56.54	66.07
	Model 4	71.05	53.21	60.85	76.21	56.97	65.20	64.75	35.26	45.66	71.48	49.62	58.58
NER2	Model 1	77.40	82.84	80.03	85.98	58.94	69.94	71.05	52.21	60.19	78.05	67.35	72.31
	Model 2	77.33	84.31	80.67	80.44	63.92	71.24	73.07	49.20	58.81	77.32	68.71	72.76
	Model 3	78.77	82.89	80.78	82.96	61.43	70.57	71.00	52.21	60.17	78.11	68.14	72.78
NER3	Model 1	78.94	78.09	78.51	76.82	73.42	75.08	77.04	57.18	65.64	77.85	71.09	74.32
	Model 2	77.94	79.31	78.62	79.14	72.19	75.51	77.99	55.85	65.09	78.32	70.88	74.42
	Model 3	78.40	78.18	78.29	78.24	72.11	75.05	77.15	58.13	66.30	78.07	70.98	74.36
	Model 4	78.63	78.74	78.69	78.69	71.88	75.13	75.76	60.09	67.02	77.99	71.67	74.70
	Model 5	78.94	78.09	78.51	76.82	73.42	75.08	76.97	56.04	64.86	77.84	70.78	74.14
	Model 6	77.94	79.31	78.62	79.18	72.23	75.55	78.07	54.17	63.96	78.35	70.44	74.19
NER4	Model 1	40.56	38.82	39.67	69.12	23.73	35.36	62.41	8.24	14.57	47.44	26.05	33.63
	Model 2	29.24	47.80	36.29	66.27	24.32	35.59	40.90	13.62	20.48	35.03	31.50	33.17

4. CONCLUSION

In this paper, we have described the results of the shared tasks on named entity recognition, organized in the framework of two last editions of VLSP workshop series: VLSP 2016 and VLSP 2018. Together with the Sentiment Analysis shared task, these two evaluation campaigns have attracted an important number of research teams as well as the public attention.

These challenges have allowed the construction of Vietnamese datasets for benchmarking named entity recognizers, as well as an overview on performance of different machine learning approaches and features for Vietnamese Named Entity Recognition.

At VLSP 2018, only 4 among 11 teams registered to the shared task arrived to the step of final result submission. This can be explained by the fact that the task was more complicated as no preprocessing was provided: the participants had to do all the tasks of preprocessing (sentence segmentation, word segmentation, POS tagging etc.) by their own tools or other available tools.

In the next campaigns, we expect to build new datasets containing a richer set of named entity categories. We hope that these open datasets for research community contribute strongly to the improvement of Vietnamese language processing systems.

ACKNOWLEDGMENT

We would like to exress our special thanks to the sponsors of VLSP 2016 and VLSP 2018 shared tasks on Named Entity Recognition: Alt Vietnam, InfoRe, VCCorp, Viettel Cyberspace Center and Zalo Careers, as well as to all the research teams who have participated to these competitions.

REFERENCES

[1] N. T. Dong, "An investigation of Vietnamese nested entity recognition models," in *The Fifth International Workshop on Vietnamese Language and Speech Processing (VLSP 2018)*, 2018. [Online]. Available: http://vlsp.org.vn/archives

[2] P. L. Hong, "Vietnamese named entity recognition using token regular expressions and bidirectional inference," in *The Fourth International Workshop on Vietnamese Language and Speech Processing (VLSP 2016)*, 2016. [Online]. Available: http://vlsp.org.vn/archives

[3] T. H. Le, T. T. T. Nguyen, T. H. Do, and X. T. Nguyen, "Named entity recognition in Vietnamese text," in *The Fourth International Workshop on Vietnamese Language and Speech Processing (VLSP 2016)*, 2016. [Online]. Available: http://vlsp.org.vn/archives

[4] V. T. Luong and L. K. Pham, "Za-ner: Vietnamese named entity recognition at VLSP 2018 evaluation campaign," in *The Fifth International Workshop on Vietnamese Language and Speech Processing (VLSP 2018)*, 2018. [Online]. Available: http://vlsp.org.vn/archives

[5] P. Q. N. Minh, "A feature-based model for nested named-entity recognition at VLSP-2018 ner evaluation campaign," in *The Fifth International Workshop on Vietnamese Language and Speech Processing (VLSP 2018)*, 2018. [Online]. Available: http://vlsp.org.vn/archives

[6] D. B. Nguyen, S. H. Hoang, S. B. Pham, and T. P. Nguyen, "Named entity recognition for Vietnamese," in *Intelligent Information and Database Systems*, N. T. Nguyen, M. T. Le, and J. Swiatek, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 205–214.

[7] H. Nguyen and T. Cao, "Named entity disambiguation: A hybrid approach," *International Journal of Computational Intelligence Systems*, vol. 5, no. 6, pp. 1052–1067, 2012.

[8] T. C. V. Nguyen, T. S. Pham, T. H. Vuong, N. V. Nguyen, and M. V. Tran, "Dsktlab-ner: Nested named entity recognition in Vietnamese text," in *The Fourth International Workshop on Vietnamese Language and Speech Processing (VLSP 2016)*, 2016. [Online]. Available: http://vlsp.org.vn/archives

[9] T. T. V. Nguyen and H. T. Cao, "Vn-kim ie: Automatic extraction of Vietnamese named-entities on the Web," *Journal of New Generation Computing*, vol. 25, no. 3, pp. 277–292, 2007.

[10] T. S. Nguyen, L. M. Nguyen, and X. C. Tran, "Vietnamese named entity recognition @VLSP 2016 evaluation campaign," in *in The Fourth International Workshop on Vietnamese Language and Speech Processing (VLSP 2016)*, 2016. [Online]. Available: http://vlsp.org.vn/archives

[11] Q. H. Pham, M.-L. Nguyen, B. T. Nguyen, and N. V. Cuong, "Semi-supervised learning for Vietnamese named entity recognition using online conditional random fields," in *Proceedings of the Fifth Named Entity Workshop, joint with 53rd ACL and the 7th IJCNLP*, Beijing, China, July 2015, pp. 50–55.

[12] T. Pham, L. M. Nguyen, and Q. Ha, "Named entity recognition for Vietnamese documents using semi-supervised learning method of crfs with generalized expectation criteria," in *2012 International Conference on Asian Language Processing*, Nov 2012, pp. 85–88.

[13] E. F. T. K. Sang and F. D. Meulder, "Introduction to the conll-2003 shared task: Language-independent named entity recognition," in *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*, 2003. [Online]. Available: http://www.aclweb.org/anthology/W03-0419

[14] B. M. Sundheim, "Overview of results of the MUC-6 evaluation," in *Proceedings of the 6th Conference on Message Understanding*, ser. MUC6 '95. Stroudsburg, PA, USA: Association for Computational Linguistics, 1995, pp. 13–31. [Online]. Available: https://doi.org/10.3115/1072399.1072402

[15] P. T. X. Thao, T. Q. Tri, D. Dien, and N. Collier, "Named entity recognition in Vietnamese using classifier voting," *ACM Transactions on Asian Language Information Processing (TALIP)*, vol. 6, no. 4, pp. 3:1–3:18, Dec. 2007. [Online]. Available: http://doi.acm.org/10.1145/1316457.1316460

[16] E. F. Tjong Kim Sang, "Introduction to the conll-2002 shared task: Language-independent named entity recognition," in *Proceedings of the 6th Conference on Natural Language Learning - Volume 20*, ser. COLING-02. Stroudsburg, PA, USA: Association for Computational Linguistics, 2002, pp. 155–158. [Online]. Available: https://doi.org/10.3115/1118853.1118877

Received on October 03, 2018
Revised on December 28, 2018