

# Domain-Independent Intent Extraction from Online Texts

Thai-Le Luong<sup>1,2</sup>, Nhu-Thuat Tran<sup>1</sup>, Tien-Son Dang<sup>1</sup>, Quoc-Long Tran<sup>1</sup>, Xuan-Hieu Phan<sup>1</sup>

<sup>1</sup> VNU University of Engineering and Technology,  
Faculty of Information Technology,  
Vietnam

<sup>2</sup> University of Transport and Communications,  
Faculty of Information Technology,  
Vietnam

luongthaile80@utc.edu.vn, hieupx@vnu.edu.vn

**Abstract.** Identifying user's intents from texts on online channels has a wide range of applications from entrepreneurship, banking to e-commerce. However, intent identification is not a simple task due to the intent and its attributes are various and strongly depend on the domain of data. If the number of intent domains increases, the number of intent's attributes will get bigger. As a result, the complexity of intent extraction task grows up significantly. Additionally, when a new domain comes, it involves considerable physical efforts to define specific labels for intent and attributes for that domain. Hence, it would be much better to come up with a new method for extracting user's intents which is not dependent on a specific domain. In our research, we study the problem of domain-independent intent identification from posts and comments crawled from social networks and discussion forums. We present ten general labels, i.e. labels do not depend on a specific domain, and utilize them when extracting intent and its related information. We also propose a map between general labels and domain-specific labels. We extensively conduct experiments to explore the efficiency of using general labels compared to specific labels in extracting user's intents when the number of intent domains increases. Our study is conducted on a medium-sized dataset from three selected domains: *Tourism*, *Real Estate* and *Transportation*. In term of accuracy, when the number of domains grows, our proposal achieves significantly better results than domain-specific method in identifying user's intent.

**Keywords.** Information extraction, intent identification, intent mining, domain-independent.

## 1 Introduction

With the explosion of Internet, there have been more and more people having accounts on social networks and discussion forums. These accounts daily generate a huge amount of valuable data. For example, a user posted on webtretho.com: “*Our family is going to Da Nang from 14/6 to 18/6, we have 5 adults and 1 child (1-year-old), could you recommend us the hotel, the best places to visit there and the total cost is about 20 million dong. Tks. Phone number: 0913 456 233*”. If a travel agent could analyze and extract user's intent, *going*, together with its related information: *destination* (Da Nang), *agenda* (from 14/6 to 18/6), *number of people* (5 adults and 1-year-old child), etc. from this post, they would give the inline advertising strategy to this user. Clearly, this advertising is very effective because it attends to the specific needs of users.

According to Luong et al.(2016) [19], the process of fully understanding user's intent includes three major stages: *user filtering*, *intent domain identification* and *intent parsing and extraction*. However, this approach faces some challenges. Firstly, user's intent and its attributes greatly depend on domain they come from. For instance, in *Transportation* domain, it is more likely for users to share a post containing *brand*, *price*, *color*, *model*, etc. when they are going to buy a car. In *Real Estate* domain, *number of floors*, *number of bedrooms*, *direction of façade* and *location* are

the most frequently mentioned attributes when people want to buy a house. It can be easily realized that the number of intent's attributes will increase dramatically when the number of intent domains increases. Thus, the complexity of intent identification grow sharply because of the growth of intent's attributes. Additionally, the stage of *intent parsing and extraction* is also hard to scale to other domains because it is domain-specific. When a new domain comes, it involves exhaustive work to define a new set of labels for intent and attributes for that domain.

To the best of our knowledge, there have been no researches that attempt to overcome the variety of intent's structure to achieve better result in extracting user's intent. "Intent extraction from social media texts using sequential segmentation and deep learning models" by Luong et al. (2017) [21] is probably the most related study to our work. However, in this study, authors attempted to extract user's intent and related information from texts generated on social networks but this method is domain-specific and does not generalize to other domains.

In this paper, we aim at building an intent extraction method which can deal with the variety of intent's attributes and also scale up to data coming from various domains. The underlying idea of this work is that for intent extraction task, we define a set of ten general labels, i.e. labels do not depend a specific domain. This idea comes from the fact that some attributes, also called labels, such as *intent, brand, contact, price*, etc. frequently appear in posts from almost domains, and, therefore, we take them as *general labels*.

For other attributes which are only mentioned in posts from a specific domain, for example, *color* in *Transportation* domain, *number of floors* in *Real Estate* domain, *time period* in *Tourism* domain we group them as *description* label.

In our work, we will intensively conduct experiments to explore the efficiency of proposed general labels compared to specific labels in intent extraction task when the number of intent domains increases. We extract the labels using various machine learning methods. Notable methods among them are the state-of-the-art Conditional Random Fields (2003) [16], Bidirectional

Long Short-term Memory and Bidirectional Long Short-term Memory – Conditional Random Fields (2016) [17].

Although we attempted to make our models be flexible, we still deal with some challenges. One of the most difficulty is the ambiguity of natural language. Such a challenge is presented in the post: "*If anyone want to liquidate your own LX motorbike, please call me. My phone number is 0983256999*".

The intent key word of this post is hidden and need some effort to deduce. While user needs to buy an old LX motorbike, the prediction model may extract the intent keyword is *liquidate*. Therefore, in the scope of this work, we only focus on extracting user's intent from posts containing explicit intents as described in [19]. Overall, the main contributions of our study are:

- We proposed a domain-independent method for intent extraction task based on the set of general labels. We also proposed the map between general labels and specific labels for data from *Tourism, Transportation* and *Real Estate* domains;
- We built an annotated medium-sized dataset containing more than 8500 Vietnamese posts from social networks and discussion forums. These data can be used for later researches in Vietnamese intent identification;
- We conducted a comparative and efficient experiments with multiple powerful machine learning models to verify the efficiency of proposed general labels.

The remainder of our paper is organized as following. Section 2 reviews some of the most related studies to our work. In section 3, we present our domain-independent approach for intent extraction task. Section 4 shows the experimental results and some discussions. Finally, section 5 concludes some main points in our work.

## 2 Related Work

To the best of our knowledge, there have been no domain-independent approaches attempted to solve the problem of intent extraction. The most related study to ours is Luong et al. (2017) [21]. In this research, authors approached the intent extraction task but in a domain-specific way, i.e. build the set of specific labels for two domains *Real Estate* and *Cosmetics & Beauty* and utilized these specific labels to extract user's intent. Another related study is the one using domain-adaptation approach by Xiao Ding (2015) [4].

The authors used datasets from specific domains to identify the consumption intention. Then they attempted to transfer the CNN mid-level sentence representation learned from one domain to another domain by adding an adaptation layer. They also proposed method to extract intention words from sentences with consumption intent. Intention word refers to the word that can best indicate users' needs. In our work, in addition to intent keyword extraction, we also extract necessary information related to the intent.

Prior to recent intent extraction researches, most of studies on detecting user's intention are based on classification. In the early 2000s, researchers have tended to classify the user's intention into three pre-defined classes: *navigation*, *information and transaction* [26, 13, 14]. Those classes do not seem to be clear enough to reveal someone's intent. Besides, authors just only focused on analyzing the queries from search engines to understand the users' intentions. In 2013, Zhiyuan Chen et al. (2013) [3] claimed that their solution was the first one that try to identify user intents in posts, i.e. the context of text documents from discussion forums.

After that time, researchers have drawn attention to online texts, and they also tried to make the intent identification more clearly. V.Gupta et al. (2014) [6] attempted to identify only purchase intent from social posts by categorizing the posts into two classes namely *PI* (purchase intent) and *non-PI*. This has been done by extracting features at two different levels of text granularity - word and phrase based features and grammatical dependency based features.

Purohit et al. (2015) [24] attributed intentions to every day behaviors, from a user query issued a search engine to buying a laptop to a user participating in a conversation. The authors define the problem of mining "relevant social intent from an ambiguous, unconstrained natural language short-text document" as a classification task with 3 classes: *seeking*, *offering*, and *none*. Luong et al. (2016) [20] followed this approach to identify intent domain.

The authors utilized two classification models, Maximum Entropy Classifier (MEC) and Support Vector Machine (SVM), to classify the intent posts into 12 major domains such as *electronic device*, *fashion & accessory*, *finance*, *food service*, *furnishing & grocery*, *travel & hotel*, *property*, *job & education*, *transportation*, *health & beauty*, *sport & entertainment*, and *pet & tree*.

Recently years, supervised learning has shown a critical drawback that it requires vast amount of manually annotated training data. To overcome this drawback, recent studies focus on domain-adaptation, transfer learning and semi-supervised learning approaches. Such approaches have been successfully applied in user's intent detection. Z. Chen et al. (2013) [3] leveraged labeled data from other domains to train a classifier for the target domain. They proposed a new transfer learning method to classify the posts into two classes: *intent posts* (positive class) and *non-intent posts* (negative class).

J. Wang et al. (2015) [27] proposed a graph-based semi-supervised approach to sort tweets into six intent categories, namely *food & drink*, *travel*, *career & education*, *goods & services*, *event & activities* and *trifle*. With effective information propagation via graph regularization, only a small set of tweets with category labels is needed as the supervised information. Ngo et al. (2017) [22] proposed a new method for intention detection, which leveraged labeled data in multi-source domains to improve performance of classification in the target domain. Specifically, they used stochastic gradient descent (SGD) to optimize the aggregation process of source and target data in a Naive Bayesian framework.

The method has shown positive improvement in intention detection task on the same benchmark dataset that used in [3].

### 3 Domain-Independent Proposal for Intent Identification

#### 3.1 Domain-Independent Intent Extraction

In the research presented by Luong et al. [19], authors defined the user explicit intent as a 5-tuple:

$$\mathbf{I}_u^e = \langle u, \mathbf{c}, d, w, \mathbf{p} \rangle, \quad (1)$$

in which:

- $u$  is the user identifier on social media services;
- $\mathbf{c}$  is the current context or condition around this intent. For example, a user may currently be pregnant, sick, or having baby. Context  $\mathbf{c}$  also includes the time at which the intent was expressed or posted on online;
- $d$  is the intent domain such as *Real-Estate*, *Finance*, *Tourism*, etc;
- $w$  is the intent name, i.e., a keyword or phrase representing the intent. It may be the name of a thing or an action of interest. For example,  $w$  can be *rent* (house), *borrow* (loan), or *book* (tour), etc;
- $\mathbf{p}$  is a list of properties or constraints associated with an intent. It is a list of property–value pairs related to the intent. For example,  $\mathbf{p}$  can be {location=*near Yen Pho industrial zone*, acreage=*90-120m2*, ...}.

And according to Luong et al. (2017) [21], authors proposed a domain-specific intent extraction model, where they tried to extract  $w$  and  $\mathbf{p}$  with the assumption that  $d$  had been identified. It means that the intent information can be extracted only if the domain of the intent have been known. However in real applications, the user's intents are very diverse, they may be "want to buy a house", or "are going to travel", and even "need to borrow some money" etc.

And that leads to a large amount of intent domain types, such as *Tourism*, *Transportation*, *Finance*, *Real Estate*, *Education*... And as we mentioned above, the more the number of the intent domains are, the more complicated it is to extract the intent information. Firstly, one needs to identify the intent domain  $d$ . And secondly, for each domain  $d$ , one has to build a set of specific labels to identify necessary attributes. For example, one could arrive at 15 specific labels for *Tourism*, 18 specific labels for *Real Estate* and 17 specific labels for *Transportation* domain respectively (see tables 1, 2, 3). Finally, after combining these three sets of specific labels, one could have the set of 33 specific labels (see table 9).

Therefore, if the number of intent domains get bigger, the number of labels will grow sharply. As a result, the bigger the number of labels is, the more difficult the predict model has to face. In this paper, we proposed a new method to identify the user intention that does not depend on the domain  $d$ . We still formulate our work as a sequential labeling problem, but the main improvement is the idea of generalizing a new set of domain-independent labels that we will describe more clearly in the next subsection. Instead of building a set of specific labels for each intent domain, we try to build the most suitable set of labels for all intent domains.

Then we built a carefully experimental strategy to verify our assumption that the set of general labels is more effective than the set of specific labels for intent extraction problem when intent domains are scaled up. This assumption lead us to the novel approach to identify users intents that be called domain-independent method. This approach allows us to extract  $w$  and  $\mathbf{p}$  without having to identify  $d$ .

This will help us to bypass one difficulty. Moreover, we are free from the worry of the number of labels increasing when a new domain comes, because we only need a few general labels as will be described in the next subsection for every domain. In the next subsection, we will present three sets of specific labels that we proposed for three domains *Tourism*, *Transportation* and *Real Estate*. Especially, we will explain more carefully about the way that we map from the set of general labels to these three sets of specific labels.

### 3.2 Domain-Independent Labels Versus Domain-Specific Labels

With three domains that we chose to crawl the data for model training (*Tourism*, *Transportation*, *Real Estate*), we built three specific sets of labels. We have 15 labels for *Tourism*, 18 labels for *Real Estate*, and 17 labels for *Transportation*, and they are described in detail in table 1, table 2 and table 3 respectively.

After conducting surveys carefully all the crawled data from three domains (*Tourism*, *Transportation* and *Real Estate*) and also some online data from other domains, we decided to proposed a set of 10 general labels. The table 4 presents these ten general labels and especially shows the mapping from the set of general labels to three sets of domain-specific labels. Some information/properties exists in almost sort of intent domains, such as *intent*, *object*, *price*..., then they are treated as general labels. Some other properties are just specific for each intent domain, for examplpe *time period* in *Tourism* domain, *acreage* in *Real Estate* domain or *color* in *Transportation* domain will be aggregated to make the label *description* in the set of general labels.

### 3.3 Intent Extraction Models

Given a post contains a user's intent which belongs to any intent domain, our model desires to extract the intent keyword (such as buy, sell, hire...) and all the necessary information that relates to the user's intention. So we proposed to use two advanced machine learning models to build our models, namely CRFs and LSTM.

#### 3.3.1 Conditional Random Fields

Conditional random fields [16] are probabilistic models has shown a great success in segmenting and labeling sequence data.

Given  $o = \{o_1, o_2, \dots, o_T\}$  as input observation sequence data, CRFs identifies  $s = \{s_1, s_2, \dots, s_T\}$ , which is a finite set of state associated with a set of labels  $y_i (y_i \in L = \{y_1, y_2, \dots, y_M\})$ , by a probability function:

$$p_{\theta}(s|o) = \frac{1}{Z_{\theta}(o)} \exp\left(\sum_{t=1}^T F(s, o, t)\right), \quad (2)$$

where  $Z_{\theta}(o) = \sum_{s'} \exp \sum_{t=1}^T F(s', o, t)$  is the normalizing factor to ensure that  $p_{\theta}(s|o)$  is a probabilistic distribution, and  $F(s, o, t) = \sum_i (\lambda_i f_i(s, o, t))$  is the sum of CRFs feature  $f_i$  with the feature weight  $\lambda_i$  correspondingly. CRFs is trained by searching the set of weights  $\theta^* = (\lambda_1^*, \lambda_2^*, \dots, \lambda_n^*)$  to maximize the log likelihood function. When the labels make the state sequence unambiguous, the likelihood function in exponential models such as CRFs is convex, thus searching the global optimum is guaranteed.

It has been shown that quasi-Newton methods, such as L-BFGS [18], is the most efficient for this issue. In our work, we utilized pycrfsuite (<https://python-crfsuite.readthedocs.io/en/>), which is a fast implementation of Conditional Random Fields on Python. We chose linear chain CRFs architecture because of faster training time. State features used in our model were as following:

- N-gram feature: we used unigram, bigram and trigram to capture the context of word in the posts;
- Part-of-speech (POS) tag of word was utilized to enrich linguistics features of word, i.e. user's intent is a verb or location is a noun. We used each single word (separated by spaces) as a word segmentaion unit. Then, we used pyvi which is a python based implementation for VN POS tagging <https://pypi.org/project/pyvi/>. After manually inspecting this POS tagging tool on social texts, we found that this tool is appropriate;
- Some of entities in our data have special forms so we used word format feature to improve the accuracy in recognizing them. For example, word contains digit tend to be a point of time or price, word is initialized by a capital character tend to be a location;
- We built a dictionary to improve the learning task and using dictionary looking-up feature for unigram, bigram and trigram. In this dictionary, we built lists of unigram, bigram or

**Table 1.** The 15 specific labels of *Tourism* domain

<b>Tourism Label</b>	<b>Abbreviation</b>	<b>Description</b>
Intent	<b>int</b>	User intent (travel, look for (hotel), book...)
Brand	<b>brd</b>	The object's brand (Vietnam Airlines, VietTran...)
Contact	<b>ctt</b>	User 's email or phone number
Context	<b>ctx</b>	User's condition that affects his/her intent (pregnant, with baby along...)
Description of Object	<b>obj-des</b>	More about object's characteristic (sea view)..
Destination	<b>dest</b>	The place where user is going to
Name of Accommodation	<b>accom-name</b>	Name of hotel, resort (Sealink, Sunwah, Ana Mandara)
Number of Objects	<b>obj-num</b>	The quantity of mentioned object
Number of People	<b>ppl-num</b>	The number of people in the journey
Object	<b>obj</b>	The object which user mentions
Point of Departure	<b>dpt</b>	The place where user's journey starts
Point of Time	<b>time-pnt</b>	When user's journey starts or finishes
Price	<b>prc</b>	The price of mentioned object
Time Period	<b>time-prd</b>	How long user's journey takes
Transport	<b>trp</b>	Means of transportation

**Table 2.** The 18 specific labels of *Real Estate* domain

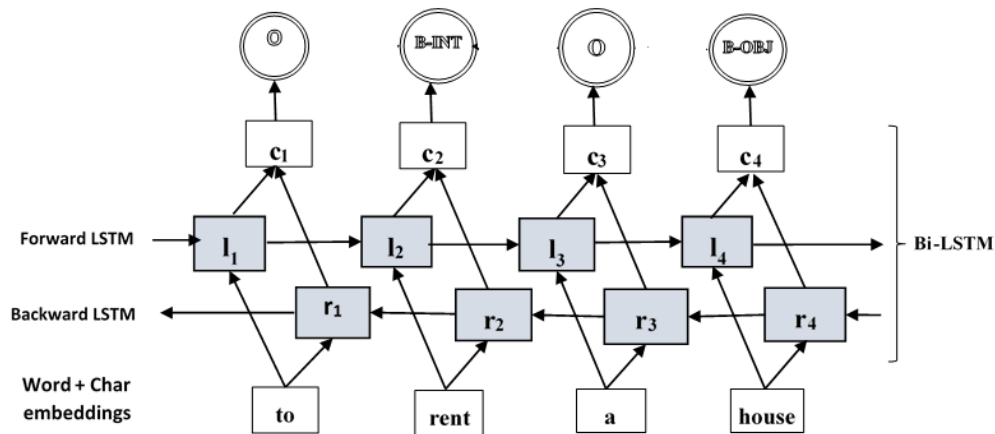
<b>Real Estate Label</b>	<b>Abbreviation</b>	<b>Description</b>
Intent	<b>int</b>	User intent (sell, buy, for rent..)
Acreage	<b>acr</b>	Object's acreage
Brand	<b>brd</b>	The object's brand
Contact	<b>ctt</b>	User 's email, phone number
Context	<b>ctx</b>	User's condition that affects his/her intent
Description of Object	<b>obj-des</b>	More about object's characteristic ( residential land, agriculture land)
Equipment	<b>eqm</b>	The equipments in house, flat
Facade Direction	<b>face-dir</b>	The direction of facade
Facade Size	<b>face-size</b>	Facade's width
Location	<b>loc</b>	Object's location or user's location
Number of Bedrooms	<b>bed-num</b>	The number of bedrooms
Number of Bathrooms	<b>bath-num</b>	The number of bathrooms
Number of Facades	<b>face-num</b>	The number of object's facades
Number of Floors	<b>fnum</b>	The number of floors
Number of Objects	<b>obj-num</b>	The quantity of mentioned object
Object	<b>obj</b>	The object which user mentions
Owner	<b>own</b>	The seller is the head-owner of object or not
Price	<b>prc</b>	The price of mentioned object

trigram that belong to some labels. For the label *Brand* as an example, we created the list contains the words or phrases such as Hon da, Vietnam Airline, VinGroup... Then if the unigrams, bigrams or trigrams appear

in those lists, the correspondingly features of the current single word will be updated. For example, if  $w_0 w_1$  in *list\_brand* return predicate  $w_0 : w_1 : in\_dictionary=brand$

**Table 3.** The 17 specific labels of *Transportation* domain

Transportation Label	Abbreviation	Description
Intent	<b>int</b>	User intent (sell, buy, hire..)
Brand	<b>brd</b>	The object's brand (Honda, Yamaha, Toyota...)
Color	<b>clr</b>	Object's color
Contact	<b>ctt</b>	User 's email, phone number
Context	<b>ctx</b>	User's condition that affects his/her intent
Description of object	<b>obj-des</b>	More about object's characteristic
Location	<b>loc</b>	Object's location or user's location
License Plate	<b>lpe</b>	The license plate of object
Model	<b>mdl</b>	The model of the object (corola 1.6, wave rsx)
Number of Objects	<b>obj-num</b>	The quantity of mentioned object
Object	<b>obj</b>	The object which user mentions
Origin	<b>orig</b>	The place where object is manufactured
Owner	<b>own</b>	The seller is the head-owner of object or not
Price	<b>prc</b>	The price of mentioned object
Registration	<b>reg</b>	The object has legal documents or not
Registration Year	<b>reg-year</b>	When object is registered
State	<b>stt</b>	Object is old or new



**Fig. 1.** The Bi-LSTM Model

### 3.3.2 Bidirectional Long Short-Term Memory (Bi-LSTM)

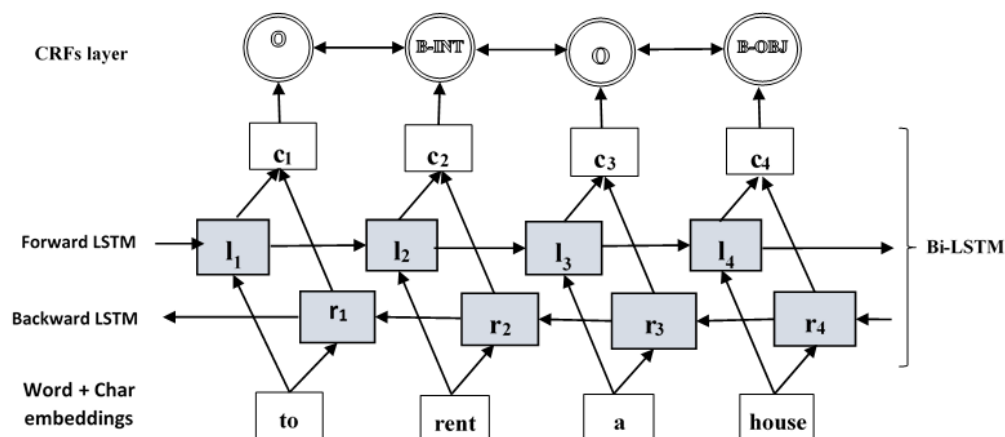
LSTM was developed based on recurrent neural network (RNN) architecture by Hochreiter and Schmidhuber (1997) [9] and it is known to be the most effective deep learning model in natural language processing problem. Given the input  $(x_1, x_2, \dots, x_n)$ , we have LSTM model computes

the state sequence  $(h_1, h_2, \dots, h_n)$  by iteratively applying the following updates:

$$\begin{aligned}
 i_t &= \sigma(W_x^{(i)}x_t + W_h^{(i)}h_{t-1} + W_c^{(i)}c_{t-1} + b^{(i)}), \\
 c_t &= (1 - i_t) \odot c_{t-1} + i_t \odot \tanh(W_x^{(c)}x_t + W_h^{(c)}h_{t-1} + b^{(c)}), \\
 o_t &= \sigma(W_x^{(o)}x_t + W_h^{(o)}h_{t-1} + W_c^{(o)}c_t + b^{(o)}), \\
 h_t &= o_t \odot \tanh(c_t),
 \end{aligned}$$

**Table 4.** The domain-independent labels

Domain-Independent Label	Abbreviation	Tourism Specific Label	Real Estate Specific Label	Transportation Specific Label
Intent	<b>int</b>	Intent	Intent	Intent
Brand	<b>brd</b>	Brand	Brand	Brand
Contact	<b>ctt</b>	Contact	Contact	Contact
Context	<b>ctx</b>	Context	Context	Context
Description	<b>des</b>	- Description of Object - Point of Time - Time Period	- Acreage - Description of Object - Equipment - Facade Direction - Facade Size - Number of Bathrooms - Number of Bedrooms - Number of Facades - Number of Floors	- Color - Description of Object - License Plate - Model - Origin - Registration Year - State
Location	<b>loc</b>	- Destination - Point of Departure	Location	Location
Number of Objects	<b>obj-num</b>	Number of Objects	Number of Objects	Number of Objects
Object	<b>obj</b>	Object	Object	Object
Other	<b>oth</b>	- Name of Accommodation - Number of People - Transport	- Owner	- Owner - Registration
Price	<b>prc</b>	Price	Price	Price



**Fig. 2.** The Bi-LSTM-CRFs Model



where  $\sigma$  is the element-wise sigmoid function and  $\odot$  is the element-wise product. LSTM have been designed to combat RNN issue by incorporating a memory-cell and have been shown to capture long-range dependencies. They do so using several gates that control the proportion of the input to give to the memory cell  $c_t$ , and the proportion  $i_t$  to forget from the previous state.

In sequence tagging task, we have to take care of both past and future input features for a given time, so we chose Bi-LSTM network [5] to do our second experiment. The figure 1 illustrates the structure of Bi-LSTM model that we used.

With this model,  $l_i$  represents the word  $i$  and its left context and  $r_i$  represents the word  $i$  and its right context. Then these two vectors will be combined to create the result vector represents the word  $i$  in its context,  $c_i$ .

Following the Bi-LSTM architecture in [17], we trained our Bi-LSTM model with the following set up:

- Because our data contains both words in formal and informal convention so it is very hard to use pre-trained word embeddings as input to Bi-LSTM model. Instead, we utilized the embeddings learned through our network.
- We combined both word embedding feature and char embedding feature as input to Bi-LSTM to reduce the affection of words which are not in vocabulary.

Specifically, the size of char embedding and the number of char long short-term memory unit in our model are both 25. These ones for the size of word embedding and the number of word long short-term memory unit are both 100.

We also used dropout technique to reduce the over-fit phenomenon. Our optimization method was Adam with learning rate, learning rate decay and clip gradients initialized by 0.001, 0.9, 5.0 respectively. All of these hyper-parameters would be tuned during training phase.

### 3.3.3 Bidirectional Long Short-Term Memory - CRFs (Bi-LSTM-CRFs)

Instead of making tagging independently, a CRFs layer is added at the end of the tagging process of a Bi-LSTM model. The output of Bi-LSTM layer had been considered as the input of CRFs layer and the output of CRFs layer will be the final tags. Based on the model described in [17], we utilized Bi-LSTM-CRFs model described in figure 2 for our problem. The initialization of this model was same as the one described in Bi-LSTM model above.

## 4 Experimental Evaluation

### 4.1 Experimental Data

In our work, we used the data from online forums, social media network and other websites. We collected data for *Tourism* domain from two main sources: <https://www.webtretho.com/forum/f110/> and <https://dulich.vnexpress.net/>. In *Real Estate* domain, data was mostly crawled from <https://batdongsan.com.vn/>. Some Facebook public groups, such as: <https://www.facebook.com/groups/xemay-cuhamoi>, were used for collecting data for out last selected domain, *Transportation*. We only used the posts which have length from 30 characters up to 800 characters in order to reduce noisy data come from advertisement posts.

Overall, our built dataset contains about 3000 posts for each domain. After that, we had a group of 5 students to tag the data with the labels that we had built. We carefully did the cross-check among of these students works to choose the most suitable annotation. Finally, for each post, we have two sets of labels to tag. The first one is the set of specific domain labels of a post and the second one is the set of general labels. Figure 3 presents an example post be tagged with domain-specific labels and domain-independent labels in turn.

After that, we carried out the experiments with both tagged types, the results and discussion will be presented bellow. We used 60% of data to train our model, 20% of data to tune the hyper-parameters. Finally, to evaluate our model we used the remaining 20% of our collected data.

<b>Example post</b>	Nhà tớ đi Đà Nẵng ngày 14/6 đến 18/6, nhà có 5 người lớn và 1 trẻ em (1 tuổi), các bác thông thái tư vấn cho tớ chọn khách sạn và đi tham quan những đâu là hợp lý nhất mà trọn gói chỉ tầm 20 triệu thôi ạ, thanks. <a href="#">sđt 0913 456 233</a> . (Our family are going to go to Da Nang from 14/6 to 18/6, we have 5 adults and 1 child (1 year old), could you recommend us the hotel, the best places to visit there and the fee is only about 20 million dong, thanks. <a href="#">sđt 0913 456 233</a> )
<b>Tagging with specific labels</b>	Nhà tớ đi <dest>Đà Nẵng</dest> ngày <time_pnt>14/6 đến 18/6</time_pnt>, nhà có <ppl_num>5 người lớn và 1 trẻ em</ppl_num> (1 tuổi), các bác thông thái tư vấn cho tớ chọn khách sạn và đi tham quan những đâu là hợp lý nhất mà trọn gói chỉ tầm <prc>20 triệu</prc> thôi ạ, thanks. <a href="#">sđt</a> <ctt>0913 456 233<ctt>.
<b>Tagging with general labels</b>	Nhà tớ đi <loc>Đà Nẵng</loc> ngày <des>14/6 đến 18/6</des>, nhà có <oth>5 người lớn và 1 trẻ em</oth> (1 tuổi), các bác thông thái tư vấn cho tớ chọn khách sạn và đi tham quan những đâu là hợp lý nhất mà trọn gói chỉ tầm <prc>20 triệu</prc> thôi ạ, thanks. <a href="#">sđt</a> <ctt>0913 456 233<ctt>.

Fig. 3. Tagging posts with specific labels and general labels

#### 4.1.1 Results on a Specific Domain

#### 4.2 Evaluation Measures

For all experiments, *precision*, *recall* and *F1-score* at the segment (or chunk-based) level are used as the evaluation measures. Specifically, assume that the true segment sequence of an instance is  $\mathbf{s} = (s_1, s_2, \dots, s_N)$  and the decoded segment sequence is  $\mathbf{s}' = (s'_1, s'_2, \dots, s'_K)$ . Then,  $s'_k$  is called a true positive if  $s'_k \in \mathbf{s}$ . The precision and recall are the fractions of the total number of true positives among the total number of decoded and true segments respectively. We report the F1-score which is computed as  $2 \cdot \text{precision} \cdot \text{recall} / (\text{precision} + \text{recall})$ . Besides, we have the *support* as the number of the true segment corresponding to each label in the testing set.

#### 4.3 Experimental Results and Discussions

We conducted the experiments with three models as we described above, namely CRFs, Bi-LSTM,

and Bi-LSTM-CRFs. With an attempt to prove our assumption that using general labels is more effect than specific ones if the number of intent domains increases, we carefully conducted totally 42 experiments, including:

- For each individual intent domain (*Tourism*, *Real Estate*, *Transportation*), extract intent with both set of specific labels and general ones;
- For each combination of 2 domains (*Tourism* vs. *Real Estate*, *Tourism* vs. *Transportation*, *Real Estate* vs. *Transportation*), extract intent with both specific labels and general ones;
- For the combination of all 3 intent domains, extract intent with both specific labels and general ones.

In the next sub section, we will present some of the most interesting results and their discussions.

As mentioned above, we did the experiments for each of three specific domains: *Tourism*,

**Table 5.** The best chunk-based result with the set of specific labels for a specific domain - *Tourism* domain

<b>Specific Label</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-score</b>	<b>Support</b>
Intent	86.65	86.38	86.52	661
Brand	0.00	0.00	0.00	14
Contact	89.91	92.45	91.16	106
Context	64.71	51.76	57.52	85
Description of Object	39.47	40.91	40.18	110
Destination	86.46	85.32	85.89	756
Name of Accommodation	51.09	54.65	52.81	86
Number of Objects	93.33	86.42	89.74	81
Number of People	89.23	82.39	85.67	352
Object	81.48	76.92	79.14	143
Point of Departure	72.84	72.84	72.84	81
Point of Time	86.04	89.29	87.64	794
Price	74.12	76.83	75.45	164
Time Period	84.88	85.71	85.29	203
Transport	56.14	58.18	57.12	55
avg/total	82.29	81.82	82.01	3691

**Table 6.** The best chunk-based result with the set of general labels for a specific domain - *Tourism* domain

<b>General Label</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-score</b>	<b>Support</b>
Intent	91.43	83.96	87.54	661
Brand	50.00	14.290	22.22	14
Contact	95.16	92.45	93.78	106
Context	72.06	57.65	64.05	85
Description	83.72	85.00	84.36	1107
Location	91.98	79.45	85.26	837
Number of Objects	95.77	83.95	89.47	81
Object	85.04	75.52	80.00	143
Other	82.03	76.88	79.37	493
Price	69.94	69.51	69.72	164
avg/total	86.38	80.71	83.33	3691

**Table 7.** The average F1-score for each specific domain with general labels and specific labels

<b>General label</b>	Tourism	Transportation	Real estate
CRFs	80.08	79.69	71.24
Bi-LSTM	81.71	77.43	72.51
Bi-LSTM-CRFs	<b>83.33</b>	<b>79.75</b>	<b>74.21</b>
<b>Specific label</b>	Tourism	Transportation	Real estate
CRFs	79.34	<b>79.78</b>	71.29
Bi-LSTM	80.89	78.00	71.70
Bi-LSTM-CRFs	<b>82.01</b>	79.76	<b>74.85</b>

**Table 8.** The best chunk-based F1-score result with the set of specific labels for Transportation domain and Real Estate domain

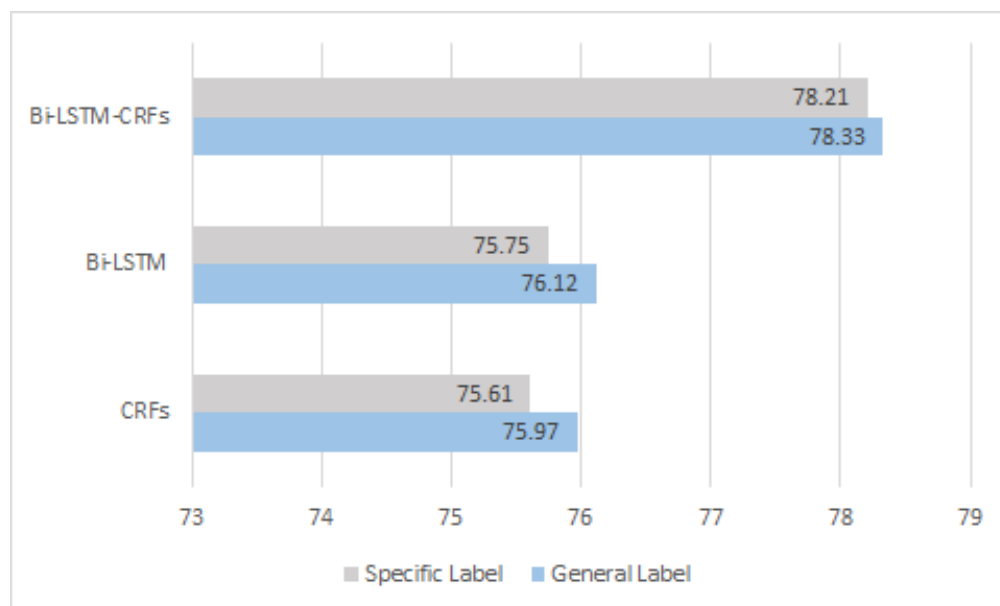
<b>Transportation Label</b>	<b>F1-score</b>	<b>Support</b>	<b>Real Estate Label</b>	<b>F1-score</b>	<b>Support</b>
Intent	90.03	661	Intent	93.37	569
Brand	87.26	192	Brand	25.00	10
Contact	94.63	458	Contact	93.23	402
Context	52.75	57	Context	40.32	51
Color	63.27	109	Facade Direction	60.91	96
Description of object	60.78	239	Acreage	83.56	575
License Plate	71.90	124	Description of Object	50.00	131
Location	78.76	403	Location	56.83	1052
Model	74.23	663	Number of Bathroom	93.33	70
Number of Objects	53.61	54	Number of Objects	51.28	39
Object	76.13	426	Object	76.80	553
Origin	81.55	111	Number of Floor	72.22	139
Owner	84.09	135	Facade Size	57.68	137
Price	88.16	501	Price	92.44	452
Registration	71.58	106	Equipment	58.17	85
Registration Year	86.90	90	Number of Bedroom	88.21	104
State	53.88	148	Number of Facade	41.18	32
			Owner	60.10	182
avg/total	79.78	4477	avg/total	74.85	4679

**Table 9.** The best chunk-based result with the set of specific labels for the combination of 3 domains

Specific Label (32)	Precision	Recall	F1-score	Support
Intent	90.94	89.69	90.31	1891
Object	75.80	79.86	77.78	1122
Acreage	83.64	80.00	81.78	575
Brand	74.66	76.39	75.51	216
Color	81.00	74.31	77.51	109
Contact	94.14	94.72	94.43	966
Context	58.22	44.04	50.15	193
Description	67.13	40.00	50.13	480
Destination	83.70	84.92	84.31	756
Equipment	77.97	54.12	63.89	85
Facade Direction	58.82	62.50	60.61	96
Facade Size	61.11	56.20	58.56	137
License Plate	75.00	75.00	75.00	124
Location	61.82	62.54	62.18	1455
Model	71.30	74.21	72.73	663
Name of Accommodation	45.95	59.30	51.78	68
Number of Bathroom	95.45	90.00	92.65	70
Number of Bedrooms	92.08	89.42	90.73	104
Number of Facades	50.00	50.00	50.00	32
Number of Floors	69.23	64.75	66.91	139
Number of Objects	75.30	71.84	73.53	174
Number of People	82.04	86.93	84.41	352
Time Period	91.01	84.73	87.76	203
Price	86.10	83.71	84.88	1117
Origin	76.32	78.38	77.33	111
Owner	72.58	68.45	70.45	317
Point of Departure	72.00	66.67	69.23	81
Point of Time	86.08	88.04	87.05	794
Registration	83.15	69.81	75.90	106
Registration Year	94.67	78.89	86.06	90
State	60.87	47.30	53.23	148
Transport	58.93	60.00	59.46	55
<b>avg/total</b>	<b>79.26</b>	<b>77.57</b>	<b>78.21</b>	<b>12847</b>

**Table 10.** The best chunk-based result with the set of general labels for the combination of 3 domains

General Label (10)	Precision	Recall	F1-score	Support
Intent	90.35	91.06	90.70	1819
Object	80.78	77.18	78.94	1122
Brand	85.96	70.83	77.66	216
Contact	94.17	95.34	94.75	966
Context	56.05	45.60	50.29	193
Description	76.58	70.10	73.20	3960
Location	69.69	71.12	70.40	2292
Number of Objects	72.84	67.82	70.24	174
Other	75.45	72.82	74.11	916
Price	87.38	86.12	86.74	1117
<b>avg/total</b>	<b>79.72</b>	<b>77.08</b>	<b>78.33</b>	<b>12847</b>



**Fig. 4.** The average F1-score for the combination of three domain datas with Bi-LSTM-CRFs, Bi-LSTM, CRFs models correspondingly

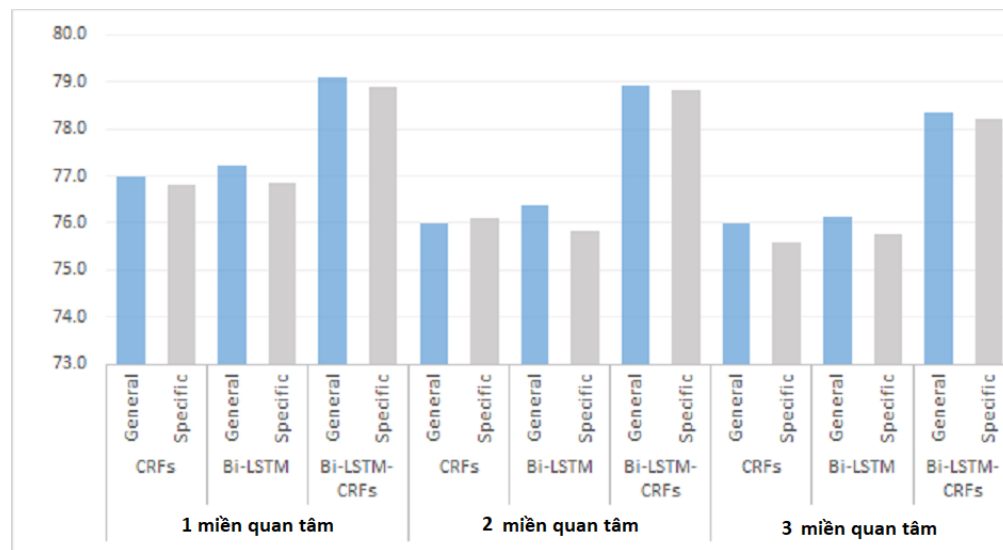
*Transportation* and *Real Estate* with both the set of general labels and the set of specific labels. Table 7 presents the overall results for these experiments with CRFs, Bi-LSTM, and Bi-LSTM-CRFs respectively. The highest F1-score we received when conducting experiments for each domain separately are the results of *Tourism* domain, they are 83.33% for extracting general labels and 82.01% for specific labels.

As described above, *Tourism* domain has least number of labels compared to *Real Estate* and *Transportation* (15, 18 and 17 labels respectively). Moreover, after carefully analyzing data from three domains, we found that *Tourism* domain contains less noisy data, such as improper abbreviation, emoticons than two remaining domains. Table 7 also shows that Bi-LSTM-CRFs achieves better results than CRFs and Bi-LSTM our experiments, although there isn't any hand-crafted feature was used in Bi-LSTM-CRFs. It proves that Bi-LSTM-CRFs is the most suitable model for our problem.

For more detail, table 6 shows the best chunk-based results when applying Bi-LSTM-CRFs model for the *Tourism* data using the set of general

labels. And table 5 shows the best chunk-based results when applying Bi-LSTM-CRFs model for the *Tourism* data using the set of specific labels. As we recognize, it is better to use the set of specific labels when extracting users intents in each individual domain. The reason is for a specific domain, the disparities in accuracy between using general labels and specific label are small, see table 7, while specific labels can describe the entities in greater detail.

So, we then present the best chunk-based results of *Transportation* domain and *Real Estate* domain when doing experiments with the set of specific labels in the table 8. We find that almost labels that benefit from high number and also their values have the recognizable form, such as *Intent*, *Price*, *Contact*,..., usually get high accuracy in all of three domains. However, some labels although have quite high number, such as *Location*, *Description*, *Equipment* and *Context*, they still get not really high accuracies. This can be explained by their complicated and barely recognizable value forms.



**Fig. 5.** The average F1-score when applying CRFs, Bi-LSTM and Bi-LSTM-CRFs models in experiments with 1 domain, 2 domains and 3 domains using general labels and specific labels correspondingly

#### 4.3.1 Using General Labels or Specific Labels when Scaling up Intention Domains

We would like to show our results and discussions to do the comparison between using the set of general labels and the set of specific labels. Figure 5 presents the average F1 score when we apply CRFs model in experiments with 1 domain using the set of domain-independent labels (general labels) and the set of domain-specific labels (specific labels) respectively, and the corresponding results when we increase the number of domains to 2 domains, 3 domains. Similarly it also shows the results when applying Bi-LSTM model and Bi-LSTM-CRFs model respectively.

We realize that it is usually gets better results when using the set of general labels rather than using the set of specific labels. So we could come to the conclusion that it would be better to use the set of general labels when identifying user's intent from collections of data combining from various domains. And as we mentioned above, one more reason for this conclusion is using the set of general labels help to get rid of rebuilding a new set of labels when a new intent domain comes.

#### 4.3.2 The Best Result for the Combination of Three Domains

Figure 4 shows the results when we apply CRFs model, Bi-LSTM model and Bi-LSTM-CRFs model to extract users intention for the combination of data from three selected domains. For each model, we conducted experiments with the set of general labels and the set of specific labels respectively. In this situation, Bi-LSTM-CRFs model still achieves higher average F1-score than the two remaining models and the experiments using the set of general labels always show higher results than the set of specific labels.

Table 9 and table 10 below show the best chunk-based results when we do the experiment with the set of 32 specific labels and the set of 10 general labels for the combination of data from three selected domains respectively. This is the result when we applied Bi-LSTM-CRFs method into our model. With the set of general labels we find that the accuracies for almost labels are quite stability. They are almost over 70%, except the label *Context*. This can be explained by the number of the *Context* labels are small, moreover the description of the *Context* labels in this problem is very diverse and complicated as can be seen in the

table 1, 2 and 3. Moreover in the experiment with the set of general labels, *Intent* and *Object* labels, which are the most important labels to identify users intents, always achieve higher F1-score than themselves in the experiment with the set of specific labels.

All in all, it reconfirms that Bi-LSTM-CRFs and the set of general labels are suitable for identifying users intents in the combination of various intent domain.

## 5 Conclusion

In this work, we present a novel method to deal with the problem of intent parsing and extraction. We call it the domain-independent intent extraction model. In this model, we propose a set of 10 general labels that is generated mainly base on three domains *Tourism*, *Transportation*, *Real Estate* and some other domain data as well.

We carefully conduct more than 40 experiments to verify our assumption that the set of general labels is more effective than the set of specific labels in the user intent identification task especially when intent domains are scaled up. Finally, most of experimental results show that our proposed general labels achieve higher accuracy than specific labels in almost experiments. The average accuracies with the set of general labels are stability and almost be over 74%.

Although these accuracies are not quite high, but it reconfirms that our approach is sensible. We also realize that we should improve our models and also the data to achieve higher results.

## Acknowledgements

This work was supported by Ministry of Education and Training under the grant number B2019-SKH-01.

## References

1. **Ashkan, A., Clarke, C.L., Agichtein, E., & Guo, Q. (2009)**. Classifying and characterizing query intent. *Proceeding of ECIR'31*, pp. 578–586. DOI: 10.1007/978-3-642-00958-7\_3.
2. **Bratman, M. (1987)**. Intention, plans, and practical reason. *Harvard University Press*.
3. **Chen, Z., Liu, B., Hsu, M., Castellanos, M., & Ghosh, R. (2013)**. Identifying intention posts in discussion forums. *Proceeding of the HLT-NAACL*, pp. 1041–1050.
4. **Ding, X., Liu, T., Duan, J., & Nie, J.Y. (2015)**. Mining user consumption intention from social media using domain adaptive convolutional neural network. *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, pp. 2389–2395.
5. **Graves, A. & Jürgen, S. (2005)**. Framewise phoneme classification with bidirectional LSTM networks. *Proceeding of (IJCNN'05)*, Vol. 4.
6. **Gupta, V., Varshney, D., Jhamtani, H., Kedia, D., & Karwa, S. (2014)**. Identifying purchase intent from social posts. *Proceeding of (ICWSM)*.
7. **Haque, R., Hasanuzzaman, M., & Way, A. (2019)**. Mining Purchase Intent in Twitter. *Computación y Sistemas*, Vol. 23, No. 3.
8. **Hashemi, H.B., A.Siaee, A., & Kraft, R. (2016)**. Query intent detection using convolutional neural networks. *Proceeding of WSDM QRUMS Workshop*.
9. **Hochreiter, S. & Jrgen, S. (1997)**. Long short-term memory. *Neural computation*, pp. 1735–1780.
10. **Hollerit, B., Kroll, M., & Strohmaier, M. (2013)**. Towards linking buyers and sellers: detecting commercial intent on twitter. *Proceeding of WWW*.
11. **Hu, J., Wang, G., Lochovsky, F., Sun, J.T., & Chen, Z. (2009)**. Understanding user's query intent with wikipedia. *Proceeding of the WWW*.
12. **Hu, D.H., Shen, D., Sun, J.T., Yang, Q., & Chen, Z. (2009)**. Context-aware online commercial intention detection. *Proceeding of The ACML*.
13. **Jansen, B. J., Booth, D. L., & Spink, A. (2007)**. Determining the User Intent of Web Search Engine Queries. *Proceeding of The 16th ACM*, pp. 1149–1150.



14. **Kathuria, A., Jansen, B.J., Hafernik, C., & Spink, A. (2010).** Classifying the user intent of web queries using k-means clustering. *The Emerald Group Journal*, Vol. 20, No. 5, pp. 563–581.
15. **Kim, J.K., Tur, G., Celikyilmaz, A., Cao, B., & Wang, Y.Y. (2016).** Intent detection using semantically enriched word embeddings. *Proceeding of SLT Workshop, IEEE*.
16. **Lafferty, J., Andrew, M., & Fernando, P. (2001).** Conditional random fields: probabilistic models for segmenting and labeling sequence data. *Proceeding of ICML*.
17. **Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., & Dyer, C. (2016).** Neural architectures for named entity recognition. *arXiv:1603.01360*.
18. **Liu, V. & Nocedal, J. (1989).** On the limited memory BFGS method for large-scale optimization. *Mathematical Programming*, Vol. 45, pp. 503–528.
19. **Luong, Th.L., Tran, Th.H., Truong, Qu.T., Truong, Th.M.Ng., Phi, Th., & Phan, X.H. (2016).** Learning to filter user explicit intents in online vietnamese social media texts. *Proceeding of ACIID*, pp. 13–24.
20. **Luong, Th.L., Truong, Qu.T., Dang, H.Tr., & Phan, X.H. (2016).** Domain identification for intention posts on online social media. *Proceeding of SolCT*, pp. 52–57.
21. **Luong, Th.L., Cao, M.S., Le, D.T., & Phan, X.H. (2017).** Intent extraction from social media texts using sequential segmentation and deep learning models. *Proceeding of the 9th KSE*, pp. 215–220.
22. **Ngo, X.B., Le, C.L., & Tu, M.Ph. (2017).** Cross-Domain Intention Detection in Discussion Forums. *Proceeding of the Eighth SolCT*, pp. 173–180.
23. **Nobari, G.H. & Chua, T.S. (2014).** User intent identification from online discussions using a joint aspect-action topic model. *Proceeding of AAAI*.
24. **Purohit, H., Dong, G., Shalin, V., Thirunarayan, K., & Shethet, A. (2015).** Intent classification of short-text on social media. *IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*, pp. 222–228.
25. **Ren, X., Wang, Y., Yu, X., Yan, J., Chen, Z., & Han, J. (2014).** Heterogeneous graph-based intent learning with queries, web pages and wikipedia concepts. *Proceeding of ICWSDM*, pp. 23–32.
26. **Rose, D.E. & Levinson, D. (2004).** Understanding user goals in web search. *Proceeding of the 13th ACM*, pp. 13–19.
27. **Wang, J., Cong, G., Zhao, W.X., & Li, X. (2015).** Mining user intents in Twitter: a semi-supervised approach to inferring intent categories for tweets. *Proceeding of the 29th AAAI*.
28. **Xu, J., Zhang, Q., & Huang, X. (2013).** Understanding the semantic intent of domain-specific natural language query. *Proceeding of International Joint Conference on Natural Language Processing*, pp. 552–560.

*Article received on 25/12/2018; accepted on 04/03/2019.  
Corresponding author is Thai-Le Luong.*