DOI 10.15622/ia.23.5.2

N.V. Hung, N. Tan, N.T.T. Nga, L.T.H. Trang, T.T.T. Hang

USING ONTOLOGY TO ANALYZE ENGLISH COMMENTS ON SOCIAL NETWORKS

Nguyen Viet Hung, Nguyen Tan, Nguyen Thi Thuy Nga, Le Thi Huyen Trang, Tran Thi Thuy Hang. **Using Ontology to Analyze English Comments on Social Networks.**

Abstract. Chatbots have become interesting for many users as technology becomes more and more advanced. The need for information exchange among people through computer systems is increasing daily, raising the preference for using chatbots in most countries. Since Vietnam is such a developing country with a variety of ethnic groups, it requires much attention to the proliferation of social networks and the expansion of the cooperative economy. Regarding social networks, the inappropriate use of words in everyday life has become a significant issue. There are mixed reviews of praise and criticism on social networks; and we try to reduce the negative language use and improve the quality of using social networks language. We aim to meet users' needs on social networks, promote economic development, and address social issues more effectively. To achieve these goals, in this paper we propose a deep learning technique using ontology knowledge mining to collect and process comments on social networks. This approach aims to enhance the user experience and facilitate the exchange of information among people by mining opinions in comments. Experimental results demonstrate that our method outperforms the conventional approach.

Keywords: chatbot, ontology, deep learning, machine learning, social network, Vietnam.

1. Introduction. Recently, the development of social networks has promoted the development of the economy as data analysis has partly contributed to improving economic efficiency, especially in improving marketing strategy [1 – 4]. There have been more and more social networking sites used, including Facebook and TikTok, for business purposes. For many businesses, creating a social network website has helped them interact with customers to collect their opinions, helping to get customers' essential needs. Besides, for businesses to be closer to their customers, they desire to receive as many mixed opinions as possible. Chatbots have been no longer strange on social networks recently. It has become more familiar to a great number of users, where consultants or assistants work continuously to give answers for many problems in specific fields. A chatbot system can understand and communicate to people and do a specific task. Chatbots have become automated tools that are able to interact with users online in natural language processing. Besides, chatbots have also been co-operated with other platforms, such as E-learning, to answer all users' questions [5, 6]. This has made online classes more attractive, especially in virtual classrooms during the COVID-19 pandemic time [7]. Despite different requirements and purposes, it is necessary for Chatbots to provide all users with guidance, documents, answers, and even entertaining stories.

Various chatbots [8, 9] are categorized based on actual and domain, goals, input processing, response generation methods, provided services, business support, people support, and methods construction. For example, to help businesses like [4], the author analyzed, classified comments, and evaluated product quality through user comments. As it is a powerful tool, based on the knowledge domain, the authors [10] used an access-based model in which the domain ontology can be built to retrieve information. Besides, several chatbots based on ontology have been built to support autoresponders from different fields, for example, educational and professional orientation [11], shopping in e-commerce [12], and consulting drug information in health [13]. Chatbots have helped a lot in many fields, including helping businesses avoid overloading, and consulting customers in many different areas without being inaccessible to professionals.

On the other hand, feedback generation and input processing methods have helped general models to be closer to users. These models can be referred to as machine learning or deep learning models, such as Long short-term memory [14 – 16], Recurrent neural network [14, 17], Fake Information Recognition [18, 19], Sequence-to-sequence [20, 21], Hierarchical Recurrent Encoder-decoder [22, 23], and other proposed methods. Some of the built popular models are better than teacher consulting services [6]. However, current chatbots are mainly based on a generic model that often gives short answers and needs to train a large data set. This also shows that building a chatbot has become more complicated when the training data is not large enough from the beginning.

In this paper, we propose a method to analyze social network comments for users through a chatbot system. Our method is based on Ontology to collect and analyze online user comments on social networks, thereby providing comments on these to support businesses partly in developing products for sale on social networks and hoping to have positive user reviews. The contribution of this proposed method can be summarized as follows:

– when analyzing social media comments, ontology can help you understand the underlying topics, identify relevant concepts, and establish relationships between them. This can help improve social media data categorization, sentiment analysis, and information retrieval. Furthermore, we employ ontology-specific techniques such as OWL, algorithms, and applications developed specifically for social media comment analysis. Besides, we provide the following. First, a structured and formalized method of representing knowledge is needed. Second, it facilitates knowledge sharing and integration and enables more practical information navigation and organization;

– the proposed method can replace traditional manual methods because it evaluates and provides direct results without requiring manual operations such as selection, classification, etc. The empirical model also performs significantly better than the reference models, up to 80.52%.

The remaining paper is as follows. Section 2 gives an overview of the related work. Next, the proposed model is presented in Section 3. Section 4 provides an evaluation. Finally, the paper is concluded in Section 5.

2. Related work. In this section, we will present the related techniques and some chatbot models being used by ontology.

2.1. Chatbots. There are many built-in definitions of chatbot [9]. However, it is a computer-to-computer communication program simulating human conversation through voice commands, text chat, or both.

There have been a lot of recent chatbot developments, including [17, 24, 25] chatbot models based on the use of neural networks using deep learning technology. The neural network is trained on data sets to generate grammatically responsive and relevant responses, such as sentences and words in the response texts. However, the system's limited analysis, including careful analysis of the data for the question, leads to the fact that the answers are so far mechanical and unfounded.

With the proposed method, our system extracts and analyzes information with a technical approach and ontology features. The system will analyze and evaluate topical content based on essential keywords in the sentence. Besides, our chatbot system also expands the synonyms and related conjunctions we analyze in the file system. This is also important when the system is analyzed more closely regarding comments. Experiments show that the results of the proposed method are superior.

2.2. Ontology. According to research [26 – 29], Ontology is a Semantic Web tool that defines the real-world semantics of terms and describes data and data relationships. These knowledge representations are very effective because they are applied in many different fields. This is also why many studies use ontologies to represent knowledge bases or retrieve information. Besides, OWL [30] is the primary Web ontology language that meets the requirements for building domain ontology, including its semantic syntax, expression support, and convenience. Therefore, we use the ontology effectively in the chatbot's search engine.

2.3. Existing chatbot models based on ontologies. In this section, we will present some existing models using ontology. Since Ontology is a solution to the desire to understand what one user is expressing, chatbot models based on ontology were born. Ontology-based models are domain knowledge-oriented so that domain-oriented conversations can be generated, in which ontology is used to store and navigate domain knowledge. Based on the domain knowledge condition, the knowledge base relies on it to build an ontology that can provide information and generate answers to conversations [31]. Of all the benefits of this approach is that the history of mastery is fully preserved during the conversation. Besides, chatbots can be applied for closed domains providing more detailed answers.

Many question-answering systems called chatbots are emerging [32, 33]. Most usual question-answering systems perform three main tasks: analyzing the questions, searching the documents and databases containing the answers, and extracting the answers from them. The system aims to answer users' queries in their natural language, including structured and unstructured free-text databases. The response system will rely on the ontology to turn it into a semantic information inference language in that knowledge domain based on the ontology. Based on this, it does not require training but still naturally asks and answers questions in a particular case.

On the other hand, for better performance than conventional systems like Apache Lucene, it uses a keyword-based text search engine. An educational counseling system [34], which is for admissions counseling, has been proposed. It helps students to identify their future major. Since enrollment is always a big challenge for its time and effort consumption, building this system is very important. There has not been a reasonable response system for this work at universities in Vietnam yet, especially in the context of rapidly developing information technology.

In general, ontology-based models are constrained by a particular domain of knowledge but can be built from many other data sources. Therefore, the research on using ontology to support answering and solving communication problems is significant, especially in university admission counseling.

3. Proposed method

3.1. Research methodology. In this section, we propose a chatbot framework for opinion analysis, following these steps: First, the data is prepared. Next, we proceed to build structure, process, and argument by ontology in Figure 1. The stages are analyzed as follows:

– **step 01**: the user sends comments to the system; here, the system will use ontology to create semantic knowledge;

– **step 02**: the system is based on the stored knowledge bases, from which to analyze the requirements and put the needs into the system;

– **step 03**: the system conducts processing and analysis as required;

– **step 04**: the system understands the user's request;

– **step 05**: the system argues based on the requirements that the system has previously understood;

– **step 06**: the system returns opinions inside comments.

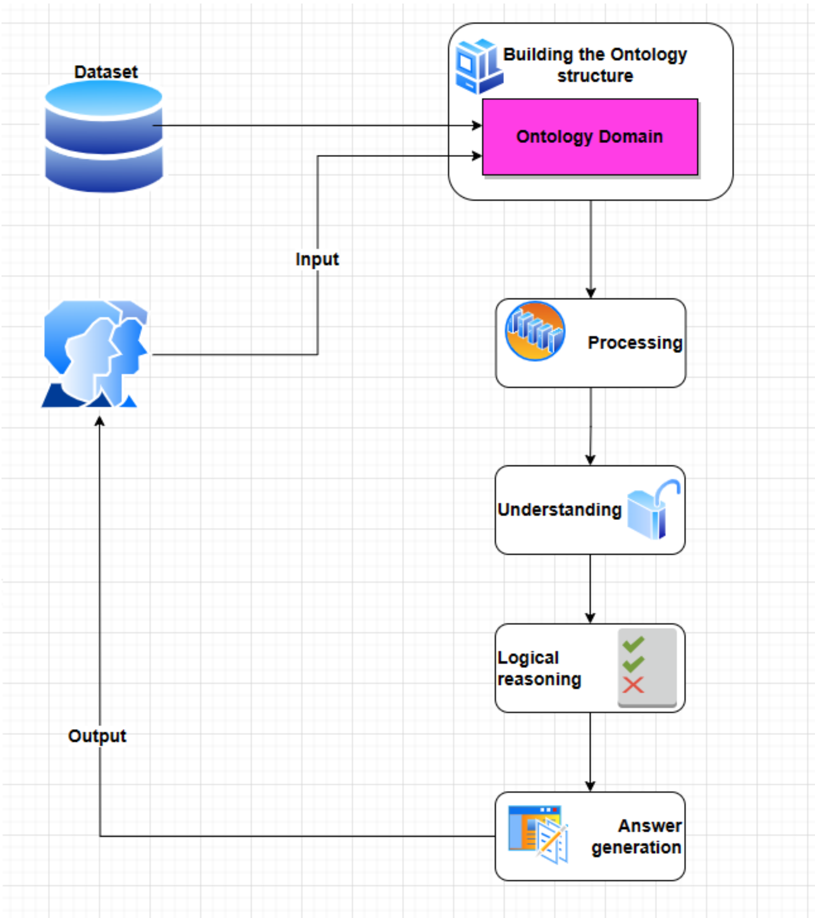

Fig. 1. The proposed communication analysis framework

Overall, this system is built to analyze a simple requirement, such as admissions advice.

3.2. Architecture of our system. This section details the architecture of the chatbot system we built in Figure 2.

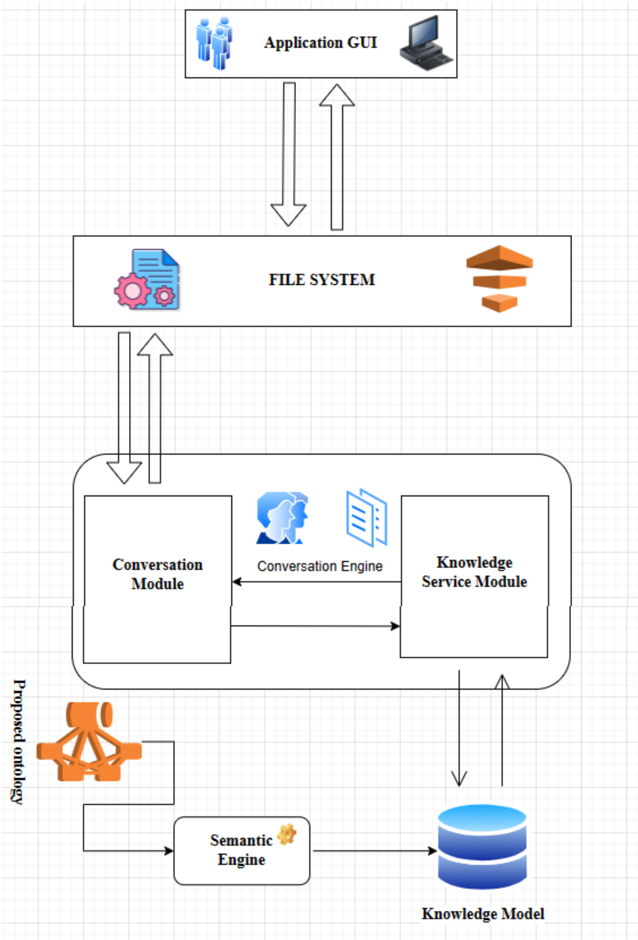

Fig. 2. System architecture diagram

This architecture not only allows interaction with the user but also provides interaction information between the ontology structure and the user's application based on the following criteria.

– **First**: a chatbot will be implemented to exchange information through a User Interface (GUI); the exchanged data is stored and processed through the file system. The file system is responsible for keeping the communication between the user and the system, processing and transmitting information to the conversation modules inside the system so that the conversation is always processed and ready to respond to the users.

– **Second**: the conversation takes place through two modules that are exchanged based on the dialogue patterns stored in the incoming file system. Based on the user-suggested question, the chatbot can access lead conversation templates that match the user's intentions. We use a knowledge mining service to get matching answer patterns and send feedback back to achieve this.

– **Third**: this step, which provides and contains the knowledge database, is crucial to get a suitable conversational sample to respond to the user. This knowledge model has both the dialogue patterns that may be relevant to the users and the problems arising from the conversations. The system will analyze and extract these derivatives to make the dialogue natural and reliable.

Furthermore, a further problem is giving users the feeling that the interaction is among humans rather than a computer system. For users, humans are preferred over technological computer systems because of their habits and reliability. A more significant challenge, therefore, is how to get the technology to understand the intent of the sentences and implications in human conversation and represent the authentic answers sent in. Consequently, we designed and exported a chatbot system based on the Neural Network model, called the Sequential model, with 3 Dense layers including:

– the first layer has 128 neurons;

– the second layer has 64 neurons;

– the last layer has the same number of neurons as the number of intents to predict the output.

On the other hand, using the Sequential model allows us to build an End-to-End model from input to creation efficiently. Keras Library supports model building and training thanks to its simplicity, ease of use, and high performance. Dense layers are connected consecutively, in which each layer is connected to all the nodes of the next layer, allowing the model to learn complex features from the data. The Neural Network deep learning model will help chatbots have better learning and prediction capabilities.

3.3. Ontology design our method. In this section, we build and develop the ontology to support the Chatbot system – a computer program is capable of communicating with users in natural language through processing techniques in Figure 3. Specifically, the ontology is designed to model critical concepts and relationships related to natural language's vocabulary, grammar, and semantics helping our Chatbot system parse the syntax, extract information, classify sentences, and finally understand the meaning of users' questions or comments on social networks and then give appropriate responses.

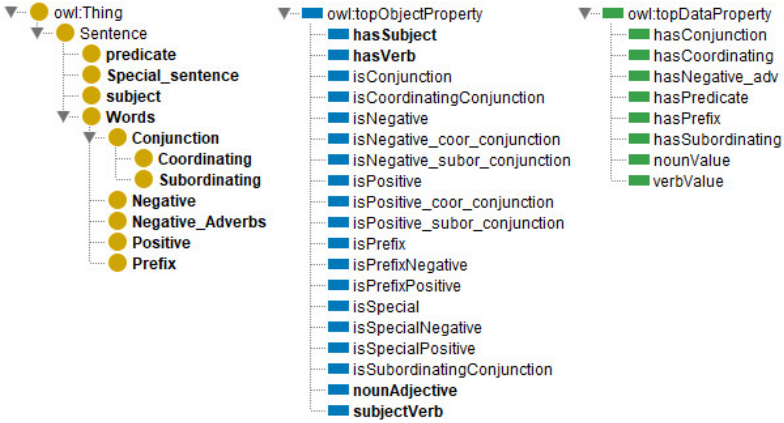

Fig. 3. Flowchart designed using our ontology

Furthermore, in artificial intelligence, ontologies play an essential role in providing general knowledge to computers, helping computers reason and solve problems. In other words, for Chatbot systems, ontology helps provide linguistic knowledge so Chatbots can understand and process natural human language. Therefore, we analyze and build a new model using ontology to evaluate the quality of comments better using the following tricks.

Class. Ontology generally refers to exploring what entities exist, their relation, and the categories and properties that can describe and classify them. Therefore, in our ontology design, a class represents a concept, object type, or a collection of objects with common properties and behavior. A class defines a collection of things with common characteristics, typical properties, and ordinary relationships. For example, when looking at comment sentences, it might be that "positive" represents the positive comments and "Negative" represents the negative class. However, a class can have sub-classes, in which sub-classes are within the same super-class, defining more specific attributes and relationships. For example, the positive category can include simple, compound, and comparative sentences. By defining classes and relationships between them, ontology creates a structure as well as a model representing knowledge about a specific field. This helps clarify concepts and relationships between them, supporting searching, querying, and processing information in that field. In Figure 3, the primary classes of ontology are designed as follows:

Words. Words are an abstract class representing all words in a natural language. The Words class defines the general properties of a word by:

– part-of-speech: shows the word type (noun, verb, adjective, etc.);

– polarity: positivity/negativity of the word;
– sentiment: the emotion the word brings (happy, sad, angry, etc.).

Sentence. Class representing the sentence and its attributes, such as:

– sentiment: general feeling of the whole sentence;
– polarity: the degree of positivity/negativity of the sentence;
– complexity: the complexity of the sentence structure.

Conjunction. This class describes conjunctions, divided into:

– coordinating: conjunction connecting independent clauses (and, or, etc.);
– subordinating: conjunction connecting dependent clauses (if, because, although, etc.).

Positive/Negative. Based on word type to divide into positive or negative classes; however, we need to pay attention to other factors, such as negative or interrogative, etc.:

– class contains words with positive or negative meanings;
– identify words expressing emotions and subjective assessments.

Special sentence. Usually, special sentences are difficult to classify, so for this type of sentence, the system needs more analysis, which can be based on the two criteria below:

– represents sentences whose meaning depends on the context;
– identify sentences that need special treatment based on context rather than literal meaning.

Instances. Instances are particular objects or real-world patterns belonging to a specific ontology class. Each instance is an instance of a class containing particular values for the properties and relationships defined in that class. For example, the attributes can be words in sentences on social comments. Sentences are filled with information through words, which helps to understand the structures and the meaning of sentences. From there, map them to reality and support analysis in different comments.

Properties. Properties in an ontology are used to describe the properties or relationships of a class or instance. Attributes help to identify specific information about objects and their relationships in the ontology. In this design, we divide it into two main categories:

– Class Property: class properties describe the common properties of a class. It is applied for all instances of that class. For example, in the class "Positive", the class attribute could be "positive words" to describe the positive comments in the class.

– Object Property: object properties describe the relationships between instances of different classes. It defines the relationship or interaction between instances. For example, in the class "Positive" and class "Negative", the object attribute could be "Positive or negative", describing the relationship between affirmative or negative sentences, such as "good" or "not good".

In general, both of these property types can have constraints and values. Regulations define the rules or conditions that the properties must obey.

Relationship. We use it intending to connect layers, such as the input layer receiving input data and passing it to the following layers for processing. Next, we use an intermediate relationship that receives data from the previous layer and passes it to the next layer to perform calculations and data transformations. Finally, we use the relationship between the last layer in the model and the output. The last layer receives data from the previous layer and performs the final calculation to produce the prediction or final result.

3.4. Proposed model-UOAEN. In this section, we will detail our analysis and evaluation process. The steps are evaluated and analyzed according to our previous ideas and development models. We have changed and developed new methods that use more ontology technology. The steps tested are as follows:

Step 01: Data preprocessing. In this step, we analyze and remove punctuation such as $(, *, , @$, etc., and eliminate spaces between words in sentences along with "stopwords" from a comment entered into the system or text data file. The process is shown in Figure 4. The algorithm processing process will return an array containing the sentence's words for future comparison and evaluation.

Step 02: Data vectorization algorithm. In this step, we convert the data from non-vector to vector format to apply the natural language processing method in Figure 5. After the data is preprocessed in the above step, vectorizing the data and extracting the features, information, or documents will be encoded into digital vectors that machine algorithms can process and learn. This process is performed as follows.

First, we use word embedding [18], TF-IDF (Term Frequency-Inverse Document Frequency), and word inversion rate to accelerate the weight analysis. We perform the TF calculation by counting the keywords in the input data; if the key is too large, we use the ontology to determine the correlation. Similarly, when calculating IDF, we consider the parameter TF. It reflects whether the keyword is used too frequently or not, such as affecting the cumulative value extreme because it may need to measure the importance of a phrase, not just its frequency in terms of how many times it is used, but also its prepositions, pronouns, conjunctions, etc. A function calculates the reverse because it more accurately reflects the value when the phrase has a higher IDF score. After all, the linear IDF function overestimates the document's score. At the same time, phrases with high IDF scores could include uncommon words and misspelled terms. Below are the two formulas we use to calculate TF and IDF.

$$TF = \frac{1 + log(\text{Keyword number})}{log(\text{Word number})}, \tag{1}$$

$$IDF = log(1 + \frac{\text{Total data}}{\text{Data contains keyword}}). \tag{2}$$

Generally, $TF - IDF$ is a comprehensive measure instead of the keyword density measure, which only reflects the degree of "stuffing" a specific keyword into text. Furthermore, it reduces the prominence of meaningless words and phrases while increasing the significance of meaningful and uncommon terms.

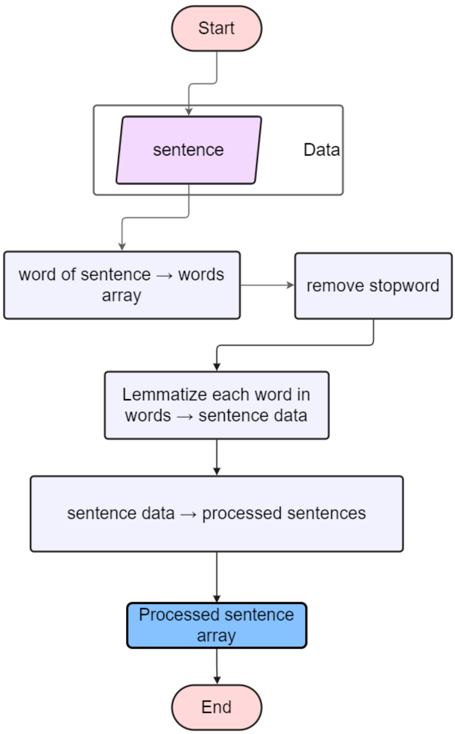

Fig. 4. Algorithmic diagram of the data preprocessing process

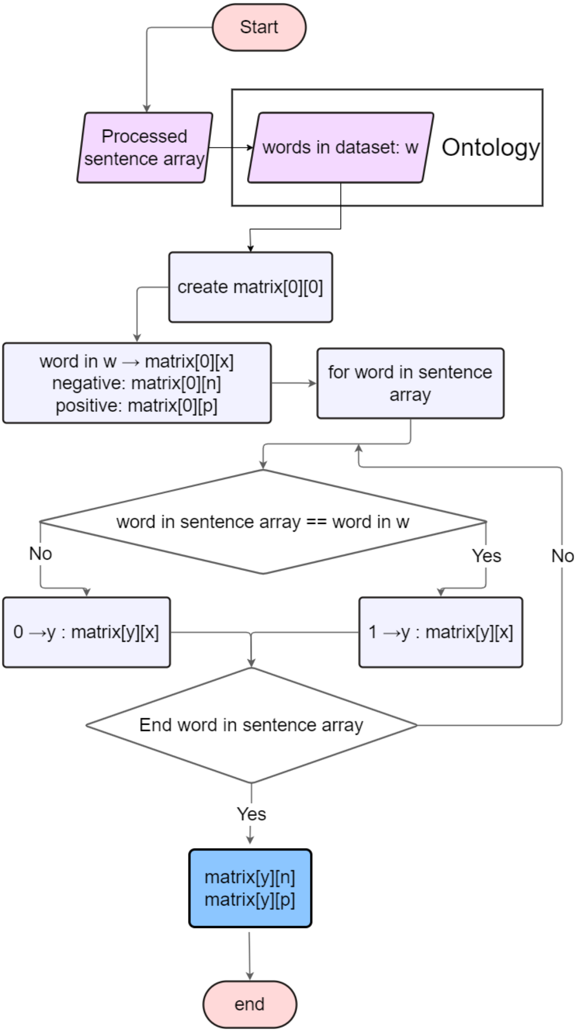

Fig. 5. Data vectorization algorithm

– **Second**, after the data is represented as a feature vector, the data will represent a characteristic or attribute, allowing the algorithm to analyze the data effectively. It helps convert unstructured or textual data into a numerical form to train models and extract information from the data in the future.

On the other hand, to do this, we calculate by matric by:

– First, create matrix[0][0].

– Second, create two negative and positive matrices of the form [0][1] according to the words in the dataset.

– Third, match each word in the words dataset. If the word in the sentence matches the word, assign 1 to position y in the form [y][x] $>$ [1][1]; otherwise, transfer 0.

– Finally, once the sentence is finished, it will stop and return two matrices.

Step 03: Our Training Model. In this step, we build a model to train the data. This is an important step to evaluate the quality of the proposed model compared to previously researched models. To implement this model, we first perform steps 1 and 2 above. After initial data preprocessing, the next steps are described in Figure 6 as follows:

– First, we use features and labels to train a model and predict based on the corresponding components.

– Second, we initialize the RNN/CNN model through two main libraries, TensorFlow and PyTorch, to learn and create predictions as well as extract the necessary information in order to classify "positive" or "negative" for the training model. Then, they practice with features and labels as objects.

– Third, we use the gradient descent algorithm to find out the standard weight value in order to determine an objective function that needs to be optimized. To do this, we initialize the model parameters with initial values. First, we use the gradient descent algorithm to repeat the following steps until the stopping condition is achieved. We calculate the objective function's gradient at the parameters' current points. The gradient represents the fastest increasing direction of the objective function. We update the model parameters based on the calculated gradient by moving in the opposite direction. Finally, we repeat the process of calculating the gradient and updating the parameters until a stopping condition is reached, such as reaching a sufficient number of iterations (epoch) or convergence of the objective function. Finally, we finish and return to the trained model.

Step 04: Prediction model. To build a model based on prediction and evaluate the results as shown in Figure 7, in this section, we conduct the experimental evaluation with the proposed model as follows:

– Sentence ← preprocess sentence.

– Vec ← vectorize sentence.

– Pred ← predict vec using model.

– Label ← get the label with the highest probability from pred.

– End the function and return the results.

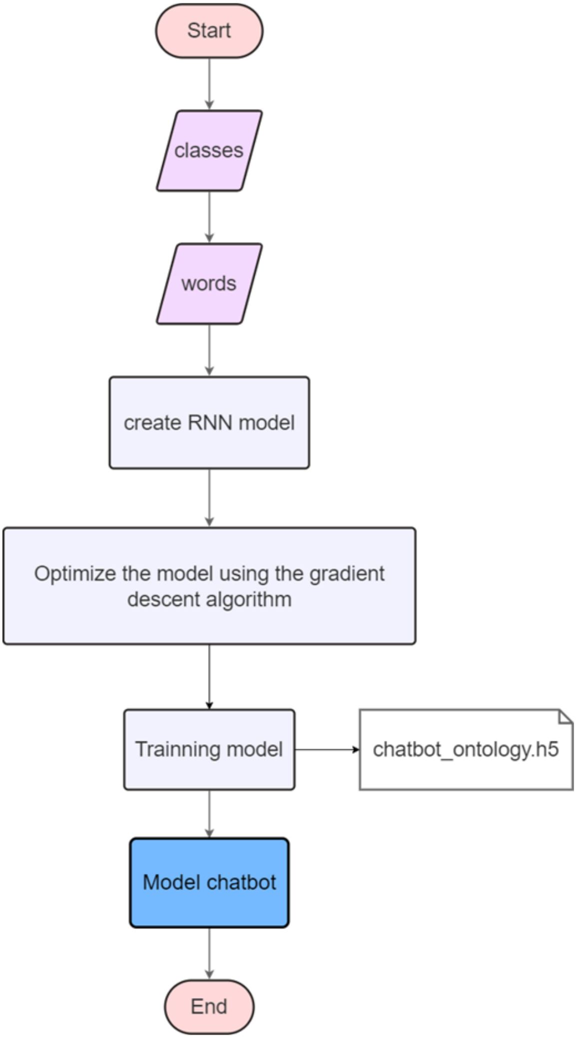

Fig. 6. Domain ontology training flowchart

Overall, in the prediction model building section, we have described it in detail in Figure 7. Based on this model, our results are better than the referenced methods.

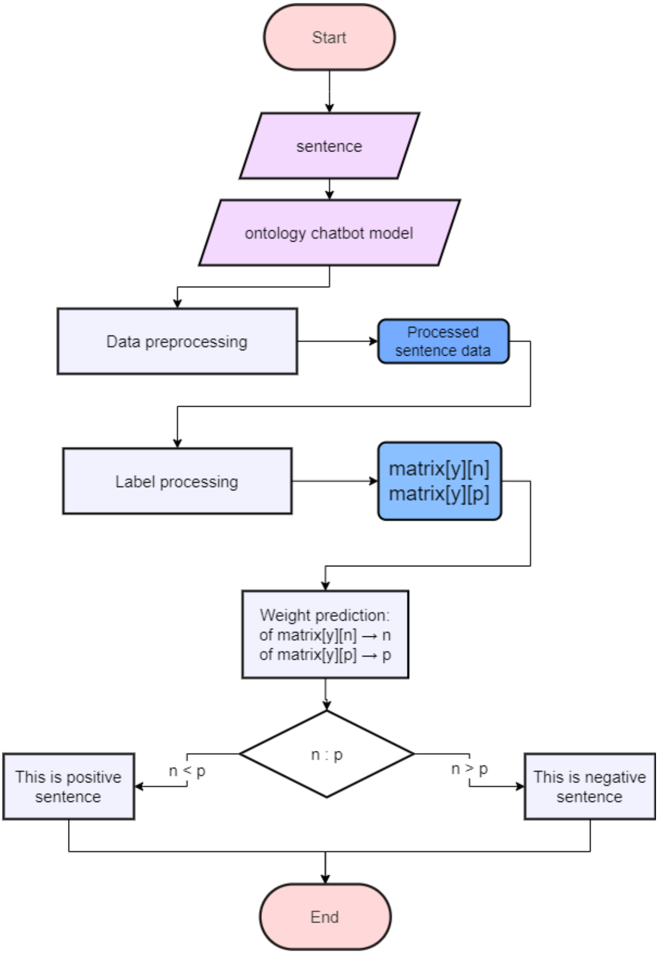

Fig. 7. Prediction model flowchart

4. Evaluation

4.1. Experimental settings. In this section, we will install and test on Python language and run on a 64-bit Windows 11 Pro computer with the following settings: Core i5-6300U (i5-6200U) / 16GB RAM / 512GB SSD / 12.5 inch HD 1366x768 screen. Furthermore, we use negative and positive comments in the dataset from [4].

On the one hand, to test the accuracy of the proposed method with human evaluation, we have created some accurate comments, as shown in Table 1.

Table 1. Some representative review comments between ITEAI (called ITE.), ACCLE (called ACC.), BCSAO (called BCS.), UOAEN (called UOA.) methods, and Human

No.	Some English comments on social networks	Evaluates				
		ITE.	ACC.	BCS.	UOA.	Human
1.	The replacement unit had a problem too: the little switch on the bottom wasn't working properly, making it impossible to select among city forecasts in my area.	t	f	t	t	Negative
2.	Acer has been awful to deal with, they treat their customers like rubbish, and have a miserable repair facility to match their products.	t	f	t	t	Negative
3.	However, since I am in an office, I normally keep them at a pretty low volume, and at low volumes, they sound a little weak.	t	f	f	t	Negative
4.	I am so tired of Netgear 's product quality & customer service that I would NEVER recommend Netgear products to anyone I know.	t	f	f	t	Negative
5.	The Netbook is great and rates 5 stars EXCEPT – Customer service is useless.	t	t	t	t	Negative
6.	I wished that they could make the volume a little bit louder.	t	t	t	t	Negative
7.	Amazing speed and is easy to set up.	f	t	t	t	Positive
8.	So far, I am super happy with this small yet powerful enough laptop.	f	t	t	t	Positive
9.	They are very sturdy, and have no trouble balance-wise, or handling oddly shaped speakers.	f	t	t	t	Positive
10.	This Pinnacle speaker bar is well worth the money.	f	t	t	t	Positive
11.	It is a wonderful little router.	f	t	t	t	Positive
12.	As a final comment, I am happy with the D-Link DIR-655 so far and my home wireless connections are smooth whether I'm using my ASUS netbook or Apple iPad.	f	t	t	t	Positive

We identify that t is true for humans, and f is false for humans. For example, comment line number 1 in Table 1 is "The replacement unit had a problem too: the little switch on the bottom wasn't working properly, making it impossible to select among city forecasts in my area". Humans evaluate this sentence as "Negative", meaning a negative sentence. ITEAI, BCSAO, and UOAEN evaluate the following evaluation methods as "Negative" (t), while the ACCLE method is evaluated as "Positive" (f). Experiments show that our method is successful. The results show that the accuracy of the system we evaluated by manual evaluation has reached 100%, while the remaining methods achieved accuracy rates of ITEAI (67%), ACCLE (50%), and BCSAO (83%), respectively.

On the other hand, to evaluate the proposed method and other methods, in this section, we use three existing methods to assess our approach; the referenced methods include: a chatbot system building to analyze opinions of English opinions (called **UOAEN**), a Chatbot for Changing Lifestyles in Education (called **ACCLE**), and an interactive Transport Inquiry with AI Chatbot (called **ITEAI**). The following is a summary of those reference methods:

– **BCSAO** [4]: This method is similar to our proposed method. However, the data processing is still done primarily through programming techniques, as opposed to our process, which goes deeper into the use of programming techniques. Creating an ontology is to classify sentences and evaluate in greater depth for each separate topic, then automatically separate based on individual models and optimize comment sentences.

– **ACCLE** [35]: The author proposes a Chatbot system to help teachers and students work together. The system is implemented by having students ask text-based questions to the Chatbot, processing the data using natural language processing and deep learning technology, and then responding to the student. However, this only serves schools without analyzing the respondents' emotions.

– **ITEAI** [36]: Like the ACCLE method, this method creates a Chatbot system that confirms the user's current location and final destination by asking questions. This method design examines the user's query and extracts the relevant entries from the database. It provides the receiver with complete information about the name and the number of the bus. By this, individuals can safely move to the desired location.

On the other hand, to evaluate the performance of the proposed method with the evaluation methods, we use the following formulas to calculate Accuracy and F1-score from [37].

Accuracy. In this research area, we use accuracy measurement to evaluate the performance of a prediction model. It calculates the ratio between the number of correct predictions and the total amount of predicted data in formula 3. The formula for calculating accuracy is expressed as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}. \tag{3}$$

Besides, we realize that accuracy is sometimes considered as a good measurement for all situations. Especially in data imbalance problems, when one class of data accounts for the majority and another class accounts for only a tiny part, the accuracy can be limited. Therefore, we additionally calculate other measurements, including precision (called Pre) in formula 4, recall (called Rec) in formula 5, and F1-score in formula 6 using the parameters shown in Figure 8, with the following formula:

$$Pre = \frac{TP}{TP + FP}, \tag{4}$$

$$Rec = \frac{TP}{TP + FN}, \tag{5}$$

$$F1\text{-}score = 2 * \frac{Pre * Rec}{Pre + Rec}, \tag{6}$$

where:

- TP: the model predicts 1 while actually, it is 1;
- TN: the model predicts 0 while actually, it is 0;
- FN: the model predicts 0, but the truth is 1;
- FP: the model predicts 1, but the truth is 0.

4.2. Experimental results. Our system outperforms the three methods listed in Table 2. When considering 736 sentences, the results in Table 2 show that the proposed method always accounts for a higher percentage than the other methods. The ACCLE method yields low average results of 50.27%, while the ITEAI and BCSAO methods yield 76.22% and 78.53%, respectively. At 80.52%, the proposed method (UOAEN) achieved the lowest level.

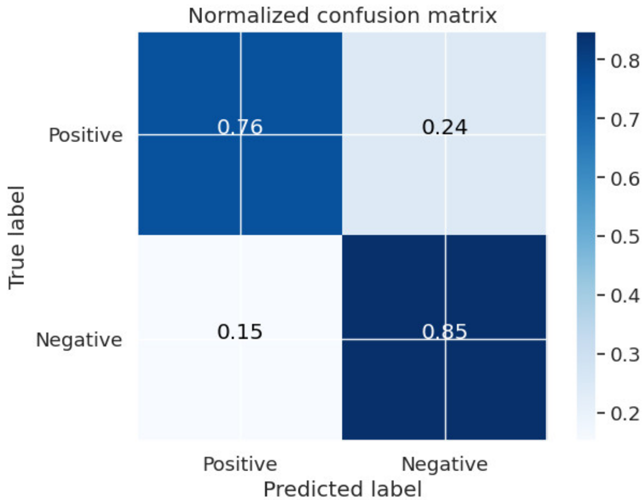

Fig. 8. Confusion matrix

Table 2. Performance of UOAEN and existing methods

Values	ACCLE	ITEAI	BCSAO	UOAEN
Accuracy	50.27%	76.22%	78.53%	80,52%
F1-score	0,82%	73.81%	76.02%	76,72%

On the one hand, we conducted experiments to measure the F1-score value of the algorithms, and the experiments showed that the proposed method achieved better results. Meanwhile, the ACCLE method has an F1-score value of at most 1% since this only focuses on responses based on available data without analyzing in-depth questions.

On the other hand, we create a system like Figure 9. Our program is analyzed through ontology to evaluate comments on social networks to evaluate positive and negative comments. The chatbot system will automatically analyze and automatically respond to each comment.

Overall, the method for exporting has better results than the referenced methods, with deep learning techniques and ontology-based analysis in Figures 10 and 11.

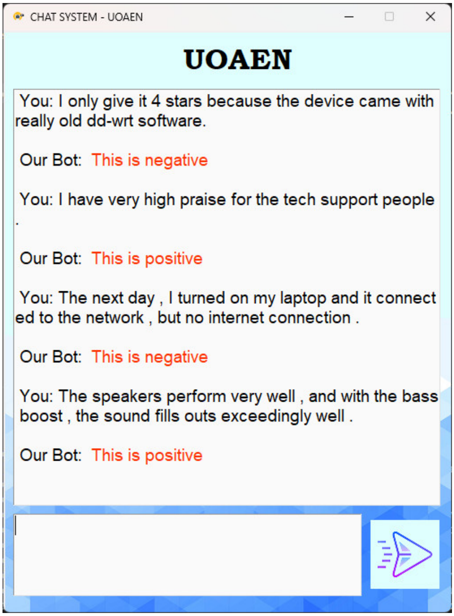

Fig. 9. GUI - UOAEN

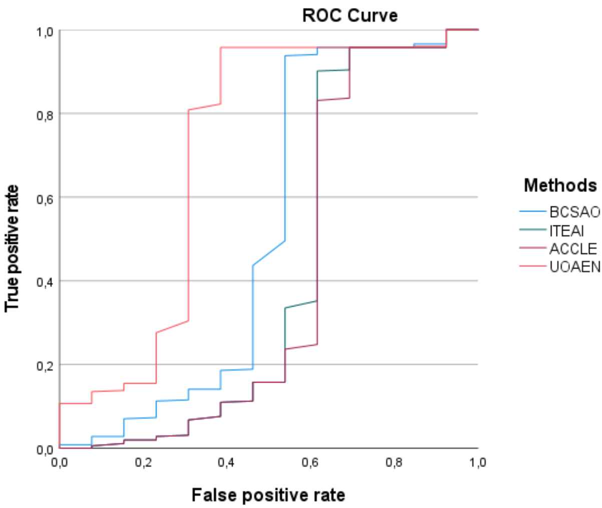

Fig. 10. Accuracy of UOAEN vs. the reference

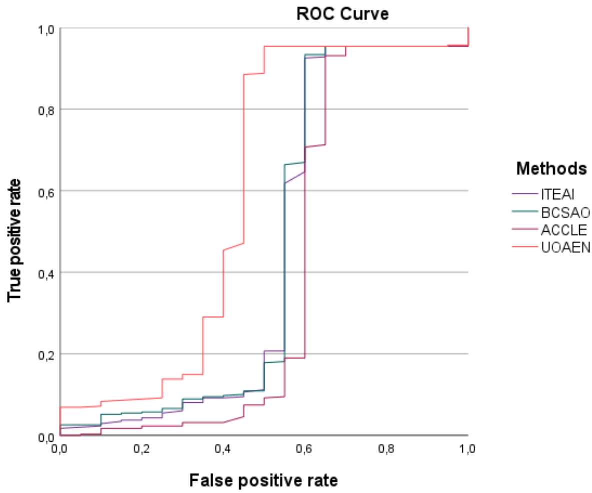

Fig. 11. F1-score of UOAEN vs. the reference

5. Conclusions. In this paper, we have built a model for evaluating comments on social networks using ontology techniques to determine sentence types such as simple sentences, compound sentences, and complex sentences.

The article has expanded on exploiting user opinions on social networks through the Chatbot system. This technique interests many businesses, especially in Vietnam, when they deal with many customer comments while the number of people serving the company is limited. Furthermore, this new research contributes to significant performance gains compared to the referenced methods. The mentioned improvement shows that the efficiency is up to 80.52%. However, we also realize some existing problems in our studies, including:

– First, testing data is still a significant factor that needs to be expanded not only in English but in other languages also.

– Second, the country's culture also dramatically affects language as well as comment sentence use.

Therefore, we will overcome and supplement the above issues in the future.

References

1. Nasir M., Rehmat U., Ahmad I. Social media analysis of customer emotions in pizza industry. The Computer Journal. 2023. vol. 66. no. 7. pp. 1777–1783.
2. Rehmat U., Javed A., Nasir M., Bashir M. Sentimental analysis of beauty brands on social media. Proc. GS Int. Conf. on Computer Science on Engineering 2020 (GSICCSE 2020).

2020. no. 37.
3. Chen Y., Wang L. Misleading political advertising fuels incivility online: A social network analysis of 2020 us presidential election campaign video comments on YouTube. Computers in Human Behavior. 2022. vol. 131(37). DOI: 10.1016/j.chb.2022.107202.
4. Nguyen H., Tan N., Quan N., Huong T., Phat H. Building a chat-bot system to analyze opinions of english comments. Informatics and Automation. 2023. vol. 22. no. 2. pp. 289–315.
5. Fidan M., Gencel N. Supporting the instructional videos with chat-bot and peer feedback mechanisms in online learning: The effects on learning performance and intrinsic motivation. Journal of Educational Computing Research. 2022. vol. 60. no. 7. pp. 1716–1741.
6. Wu E., Lin C.-H., Ou Y.-Y., Liu C.-Z., Wang W.-K., Chao C.-Y. Advantages and constraints of a hybrid model k-12 e-learning assistant chatbot. Ieee Access. 2020. vol. 8. pp. 77788–77801.
7. Surani D., Hamidah H. Students perceptions in online class learning during the covid-19 pandemic. International Journal on Advanced Science, Education, and Religion. 2020. vol. 3. no. 3. pp. 83–95.
8. Suta P., Lan X., Wu B., Mongkolnam P., Chan J. An overview of machine learning in chatbots. International Journal of Mechanical Engineering and Robotics Research. 2020. vol. 9. no. 4. pp. 502–510.
9. Adamopoulou E., Moussiades L. An overview of chatbot technology. IFIP international conference on artificial intelligence applications and innovations. Springer, 2020. pp. 373–383.
10. Nguyen T., Ho D., Nguyen N. An ontology-based question answering system for university admissions advising. Intelligent Automation and Soft Computing. 2023. vol. 36. no. 1. pp. 601–616. DOI: 10.32604/iasc.2023.032080.
11. Zahour O., Eddaoui A., Ouchra H., Hourrane O., Benlahmar E. A system for educational and vocational guidance in morocco: Chatbot e-orientation. Procedia Computer Science. 2020. vol. 175. pp. 554–559. DOI: j.procs.2020.07.079.
12. Hallili A. Toward an ontology-based chatbot endowed with natural language processing and generation. Processing of the 26th european summer school in logic, language and information. 2014. 7 p.
13. Avila C., Calixto A., Rolim T., Franco W., Venceslau A., Vidal V., Pequeno V., Moura F. Medibot: an ontology-based chatbot for Portuguese speakers drug's users. International Conference on Enterprise Information Systems. 2019. vol. 1. pp. 25–36.
14. Muangkammuen P., Intiruk N., Saikaew K. Automated thai-faq chatbot using rnn-lstm. in 2018 22nd International Computer Science and Engineering Conference (ICSEC). IEEE, 2018. pp. 1–4.
15. Blanc C., Bailly A., Francis E., Guillotin T., Jamal F., Wakim B., Roy P. Flaubert vs. camembert: Understanding patient's answers by a french medical chatbot. Artificial Intelligence in Medicine. 2022. vol. 127.
16. Hung N., Loi T., Binh N., Nga N., Huong T., Luu D. Building an online learning model through a dance recognition video based on deep learning. Informatics and Automation. 2024. vol. 23. no. 1. pp. 101–128.
17. Dhyani M., Kumar R. An intelligent chatbot using deep learning with bidirectional rnn and attention model. Materials today: proceedings. 2021. vol. 34. pp. 817–824.
18. Hung N., Loi T., Huong N., Hang T., Huong T. Aafndl-an accurate fake information recognition model using deep learning for the vietnamese language. Informatics and Automation. 2023. vol. 22. no. 4. pp. 795–825.

19. Suganthi S., Ayoobkhan M., Kumar K., Bacanin N., Venkatachalam K., Hubalovsky S., Trojovsky P. Deep learning model for deep fake face recognition and detection. PeerJ Computer Science. 2022. vol. 8. DOI: 10.7717/peerj-cs.881.
20. Rakib A., Rumky E., Ashraf A., Hillas M., Rahman M. Mental healthcare chatbot using sequence-to-sequence learning and bilstm. International Conference on Brain Informatics. Springer, 2021. pp. 378–387.
21. Palasundram K., Sharef N., Nasharuddin N., Kasmiran K., Azman A. Sequence to sequence model performance for education chatbot. International journal of emerging Technologies in Learning (iJET). 2019. vol. 14. no. 24. pp. 56–68.
22. Piccini R., Spanakis G. Exploring the context of recurrent neural network based conversational agents. arXiv preprint arXiv:1901.11462. 2019.
23. Sun R., Chen B., Zhou Q., Li Y., Cao Y., Zheng H.-T. A non-hierarchical attention network with modality dropout for textual response generation in multimodal dialogue systems. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022. pp. 6582–6586.
24. Karna M., Juliet D., Joy R. Deep learning-based text emotion recognition for chatbot applications. 4th International Conference on Trends in Electronics and Informatics (ICOEI) (48184). IEEE, 2020. pp. 988–993.
25. Sperli G. A cultural heritage framework using a deep learning-based chatbot for supporting tourist journey. Expert Systems with Applications. 2021. vol. 183. DOI: 10.1016/j.eswa.2021.115277.
26. Hitzler P., Krotzsch M., Parsia B., Patel-Schneider P., Rudolph S., et al. Owl 2 web ontology language primer. W3C recommendation. 2009. vol. 27. no. 1.
27. Antoniou G., Groth P., Van Harmelen F., Hoekstra R. A semantic web primer 3rd ed. The MIT Press, 2012. 288 p.
28. Casillo M., De Santo M., Mosca R., Santaniello D. An ontology-based chatbot to enhance experiential learning in a cultural heritage scenario. Frontiers in Artificial Intelligence. 2022. vol. 5.
29. Chi Y.-L., Sung H.-Y. Building the knowledge base of folk beliefs based on semantic web technology. International Conference on Human-Computer Interaction. Springer Nature Switzerland, 2023. pp. 172–182.
30. Owl web ontology language. Available: https://www.w3.org/TR/owl-features/ (accessed: 2023-08-15).
31. Aceta C., Fernandez I., Soroa A. Todo: A core ontology for task-oriented dialogue systems in industry 4.0. Further with Knowledge Graphs. IOS Press, 2021. pp. 1–15.
32. Bouziane A., Bouchiha D., Doumi N., Malki M. Question answering systems: survey and trends. Procedia Computer Science. 2015. vol. 73. pp. 366–375.
33. Budler L.C., Gosak L., Stiglic G. Review of artificial intelligence-based question-answering systems in healthcare. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2023. vol. 13. no. 2.
34. Majid M., Hayat M., Khan F., Ahmad M., Jhanjhi N., Bhuiyan M., Masud M., AlZain M. Ontology-based system for educational program counseling. Intelligent Automation and Soft Computing. 2021. vol. 30. no. 1. pp. 373–386.
35. Kasthuri E., Balaji S. A chatbot for changing lifestyle in education. Proceedings of the Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV). IEEE, 2021. pp. 1317–1322.
36. Dharani M., Jyostna J., Sucharitha E., Likitha R., Manne S. Interactive transport enquiry with ai chatbot. Proceedings of the 4th International Conference on Intelligent Computing and Control Systems (ICICCS). 2020. pp. 1271–1276.

37. Nguyen H., Dao T., Pham N., Dang T., Nguyen T., Truong T. An accurate viewport estimation method for 360 video streaming using deep learning. EAI Endorsed Transactions on Industrial Networks and Intelligent Systems. 2022. vol. 9. no. 4.

Nguyen Viet Hung — Ph.D., Lecturer, Faculty of information technology, East Asia University of Technology. Research interests: multimedia communications, network security, artificial intelligence, traffic engineering in next-generation networks, QoE/QoS guarantee for network services, green networking, applications. The number of publications — 18. hungnv@eaut.edu.vn; Ky Phu – Ky Anh, Ha Tinh, Viet Nam; office phone: +84(098)911-2079.

Nguyen Tan — Research assistant, East Asia University of Technology. Research interests: applications, data analysis. The number of publications — 2. tan25102000@gmail.com; Trung Dung – Tien Lu, Hung Yen, Viet Nam; office phone: +84(035)919-0216.

Nguyen Thi Thuy Nga — Lecturer, Faculty of information technology, East Asia University of Technology. Research interests: multimedia communications, database management systems, artificial intelligence, applications. The number of publications — 2. ngantt@eaut.edu.vn; Ta Thanh Oai – Thanh Tri, Hanoi, Viet Nam; office phone: +84(097)529-4770.

Le Thi Huyen Trang — Lecturer, Faculty of information technology, East Asia University of Technology. Research interests: multimedia communications, database management systems, artificial intelligence, applications. The number of publications — 1. tranglth@eaut.edu.vn; Phuong Tri – Thi Tran Phung – Dan Phuong, Hanoi, Viet Nam; office phone: +84(032)8899-334.

Tran Thi Thuy Hang — Lecturer, Faculty of information technology, East Asia University of Technology. Research interests: multimedia communications, database management systems, artificial intelligence, applications. The number of publications — 2. hang42c@gmail.com; Mao Khe - Dong Trieu, Quang Ninh, Viet Nam; office phone: +84(090)496-8545.

УДК 006.72 **DOI** 10.15622/ia.23.5.2

Н.В. Хунг, Н. Тан , Н.Т.Т. Нга , Л.Х. Транг , Т.Т. Ханг

ИСПОЛЬЗОВАНИЕ ОНТОЛОГИИ ДЛЯ АНАЛИЗА АНГЛИЙСКИХ КОММЕНТАРИЕВ В СОЦИАЛЬНЫХ СЕТЯХ

Вьет Хунг Н., Тан Н., Тхи Туй Нга Н., Хуен Транг Л., Туй Ханг Т. **Использование онтологии для анализа английских комментариев в социальных сетях.**

Аннотация. Чат-боты заинтересовывыют многих пользователей по мере того, как технологии становятся все более продвинутыми. Потребность в обмене информацией между людьми через компьютерные системы увеличивается с каждым днем, в результате чего в большинстве стран растет предпочтение использовать чат-боты. Поскольку Вьетнам является развивающейся страной с множеством этнических групп, требуется усиленное внимание к распространению социальных сетей и расширению кооперативной экономики. Серьезной проблемой стало неуместное использование слов в повседневной жизни. В социальных сетях встречаются неоднозначные отзывы с похвалой и критикой о том, что мы пытаемся уменьшить использование негативной лексики и улучшить качество использования языка в социальных сетях. Мы стремимся удовлетворить потребности пользователей в социальных сетях, способствовать экономическому развитию и более эффективно решать социальные проблемы. Для достижения этих целей предлагается метод глубокого обучения, использующий интеллектуальный анализ онтологических знаний для сбора и обработки комментариев в социальных сетях. Этот подход направлен на улучшение пользовательского опыта и облегчение обмена информацией между людьми путем анализа мнений в комментариях. Результаты экспериментов показывают, что наш метод превосходит традиционный подход.

Ключевые слова: чат-бот, онтология, глубокое обучение, машинное обучение, социальная сеть, Вьетнам.

Литература

1. Nasir M., Rehmat U., Ahmad I. Social media analysis of customer emotions in pizza industry. The Computer Journal. 2023. vol. 66. no. 7. pp. 1777–1783.
2. Rehmat U., Javed A., Nasir M., Bashir M. Sentimental analysis of beauty brands on social media. Proc. GS Int. Conf. on Computer Science on Engineering 2020 (GSICCSE 2020). 2020. no. 37.
3. Chen Y., Wang L. Misleading political advertising fuels incivility online: A social network analysis of 2020 us presidential election campaign video comments on YouTube. Computers in Human Behavior. 2022. vol. 131(37). DOI: 10.1016/j.chb.2022.107202.
4. Nguyen H., Tan N., Quan N., Huong T., Phat H. Building a chat-bot system to analyze opinions of english comments. Informatics and Automation. 2023. vol. 22. no. 2. pp. 289–315.
5. Fidan M., Gencel N. Supporting the instructional videos with chat-bot and peer feedback mechanisms in online learning: The effects on learning performance and intrinsic motivation. Journal of Educational Computing Research. 2022. vol. 60. no. 7. pp. 1716–1741.
6. Wu E., Lin C.-H., Ou Y.-Y., Liu C.-Z., Wang W.-K., Chao C.-Y. Advantages and constraints of a hybrid model k-12 e-learning assistant chatbot. Ieee Access. 2020. vol. 8. pp. 77788–77801.

7. Surani D., Hamidah H. Students perceptions in online class learning during the covid-19 pandemic. International Journal on Advanced Science, Education, and Religion. 2020. vol. 3. no. 3. pp. 83–95.
8. Suta P., Lan X., Wu B., Mongkolnam P., Chan J. An overview of machine learning in chatbots. International Journal of Mechanical Engineering and Robotics Research. 2020. vol. 9. no. 4. pp. 502–510.
9. Adamopoulou E., Moussiades L. An overview of chatbot technology. IFIP international conference on artificial intelligence applications and innovations. Springer, 2020. pp. 373–383.
10. Nguyen T., Ho D., Nguyen N. An ontology-based question answering system for university admissions advising. Intelligent Automation and Soft Computing. 2023. vol. 36. no. 1. pp. 601–616. DOI: 10.32604/iasc.2023.032080.
11. Zahour O., Eddaoui A., Ouchra H., Hourrane O., Benlahmar E. A system for educational and vocational guidance in morocco: Chatbot e-orientation. Procedia Computer Science. 2020. vol. 175. pp. 554–559. DOI: j.procs.2020.07.079.
12. Hallili A. Toward an ontology-based chatbot endowed with natural language processing and generation. Processing of the 26th european summer school in logic, language and information. 2014. 7 p.
13. Avila C., Calixto A., Rolim T., Franco W., Venceslau A., Vidal V., Pequeno V., Moura F. Medibot: an ontology-based chatbot for Portuguese speakers drug's users. International Conference on Enterprise Information Systems. 2019. vol. 1. pp. 25–36.
14. Muangkammuen P., Intiruk N., Saikaew K. Automated thai-faq chatbot using rnn-lstm. in 2018 22nd International Computer Science and Engineering Conference (ICSEC). IEEE, 2018. pp. 1–4.
15. Blanc C., Bailly A., Francis E., Guillotin T., Jamal F., Wakim B., Roy P. Flaubert vs. camembert: Understanding patient's answers by a french medical chatbot. Artificial Intelligence in Medicine. 2022. vol. 127.
16. Hung N., Loi T., Binh N., Nga N., Huong T., Luu D. Building an online learning model through a dance recognition video based on deep learning. Informatics and Automation. 2024. vol. 23. no. 1. pp. 101–128.
17. Dhyani M., Kumar R. An intelligent chatbot using deep learning with bidirectional rnn and attention model. Materials today: proceedings. 2021. vol. 34. pp. 817–824.
18. Hung N., Loi T., Huong N., Hang T., Huong T. Aafndl-an accurate fake information recognition model using deep learning for the vietnamese language. Informatics and Automation. 2023. vol. 22. no. 4. pp. 795–825.
19. Suganthi S., Ayoobkhan M., Kumar K., Bacanin N., Venkatachalam K., Hubalovsky S., Trojovsky P. Deep learning model for deep fake face recognition and detection. PeerJ Computer Science. 2022. vol. 8. DOI: 10.7717/peerj-cs.881.
20. Rakib A., Rumky E., Ashraf A., Hillas M., Rahman M. Mental healthcare chatbot using sequence-to-sequence learning and bilstm. International Conference on Brain Informatics. Springer, 2021. pp. 378–387.
21. Palasundram K., Sharef N., Nasharuddin N., Kasmiran K., Azman A. Sequence to sequence model performance for education chatbot. International journal of emerging Technologies in Learning (iJET). 2019. vol. 14. no. 24. pp. 56–68.
22. Piccini R., Spanakis G. Exploring the context of recurrent neural network based conversational agents. arXiv preprint arXiv:1901.11462. 2019.
23. Sun R., Chen B., Zhou Q., Li Y., Cao Y., Zheng H.-T. A non-hierarchical attention network with modality dropout for textual response generation in multimodal dialogue systems. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022. pp. 6582–6586.

24. Karna M., Juliet D., Joy R. Deep learning-based text emotion recognition for chatbot applications. 4th International Conference on Trends in Electronics and Informatics (ICOEI) (48184). IEEE, 2020. pp. 988–993.
25. Sperli G. A cultural heritage framework using a deep learning-based chatbot for supporting tourist journey. Expert Systems with Applications. 2021. vol. 183. DOI: 10.1016/j.eswa.2021.115277.
26. Hitzler P., Krotzsch M., Parsia B., Patel-Schneider P., Rudolph S., et al. Owl 2 web ontology language primer. W3C recommendation. 2009. vol. 27. no. 1.
27. Antoniou G., Groth P., Van Harmelen F., Hoekstra R. A semantic web primer 3rd ed. The MIT Press, 2012. 288 p.
28. Casillo M., De Santo M., Mosca R., Santaniello D. An ontology-based chatbot to enhance experiential learning in a cultural heritage scenario. Frontiers in Artificial Intelligence. 2022. vol. 5.
29. Chi Y.-L., Sung H.-Y. Building the knowledge base of folk beliefs based on semantic web technology. International Conference on Human-Computer Interaction. Springer Nature Switzerland, 2023. pp. 172–182.
30. Owl web ontology language. Available: https://www.w3.org/TR/owl-features/ (accessed: 2023-08-15).
31. Aceta C., Fernandez I., Soroa A. Todo: A core ontology for task-oriented dialogue systems in industry 4.0. Further with Knowledge Graphs. IOS Press, 2021. pp. 1–15.
32. Bouziane A., Bouchiha D., Doumi N., Malki M. Question answering systems: survey and trends. Procedia Computer Science. 2015. vol. 73. pp. 366–375.
33. Budler L.C., Gosak L., Stiglic G. Review of artificial intelligence-based question-answering systems in healthcare. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2023. vol. 13. no. 2.
34. Majid M., Hayat M., Khan F., Ahmad M., Jhanjhi N., Bhuiyan M., Masud M., AlZain M. Ontology-based system for educational program counseling. Intelligent Automation and Soft Computing. 2021. vol. 30. no. 1. pp. 373–386.
35. Kasthuri E., Balaji S. A chatbot for changing lifestyle in education. Proceedings of the Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV). IEEE, 2021. pp. 1317–1322.
36. Dharani M., Jyostna J., Sucharitha E., Likitha R., Manne S. Interactive transport enquiry with ai chatbot. Proceedings of the 4th International Conference on Intelligent Computing and Control Systems (ICICCS). 2020. pp. 1271–1276.
37. Nguyen H., Dao T., Pham N., Dang T., Nguyen T., Truong T. An accurate viewport estimation method for 360 video streaming using deep learning. EAI Endorsed Transactions on Industrial Networks and Intelligent Systems. 2022. vol. 9. no. 4.

Вьет Хунг Нгуен — Ph.D., преподаватель, факультет информационных технологий, Восточноазиатский технологический университет. Область научных интересов: мультимедийные коммуникации, сетевая безопасность, искусственный интеллект, управление трафиком в сетях нового поколения, гарантия QoE/QoS для сетевых сервисов, экологически чистые сети и приложения. Число научных публикаций — 18. hungnv@eaut.edu.vn; Ки Фу – Ки Ань, Хатинь, Вьетнам; р.т.: +84(098)911-2079.

Тан Нгуен — научный сотрудник, Восточноазиатский технологический университет. Область научных интересов: прикладные программы, анализ данных. Число научных публикаций — 2. tan25102000@gmail.com; Чунг Зунг – Тьен Лу, Хынгйен, Вьетнам; р.т.: +84(035)919-0216.

Тхи Туй Нга Нгуен — преподаватель, факультет информационных технологий, Восточноазиатский технологический университет. Область научных интересов:

мультимедийные коммуникации, системы управления базами данных, искусственный интеллект, приложения. Число научных публикаций — 2. ngantt@eaut.edu.vn; Та Тхань Оай – Тханьчи, Ханой, Вьетнам; р.т.: +84(097)529-4770.

Хуен Транг Ле Тхи — преподаватель, факультет информационных технологий, Восточноазиатский технологический университет. Область научных интересов: мультимедийные коммуникации, системы управления базами данных, искусственный интеллект, приложения. Число научных публикаций — 1. tranglth@eaut.edu.vn; Фуонг Три – Тхи Чан Пхунг – Дан Фуонг, Ханой, Вьетнам; р.т.: +84(032)8899-334.

Туй Ханг Тран Тхи — преподаватель, факультет информационных технологий, Восточноазиатский технологический университет. Область научных интересов: мультимедийные коммуникации, системы управления базами данных, искусственный интеллект, приложения. Число научных публикаций — 2. hang42c@gmail.com; Мао Кхе - Донг Чиу, Куанг Нин, Вьетнам; р.т.: +84(090)496-8545.