The 2020 12th International Conference on Knowledge and Systems Engineering (KSE)

November 12 -14, 2020
Can Tho University, Viet Nam

Proceedings

IEEE CATALOG NUMBERS:
Part Number: CFP2003I-ART
ISBN: 978-1-7281-4510-5

Table of Contents

The 2020 12th International Conference on Knowledge and Systems Engineering (KSE)

Knowledge Integration Technologies and Applications

Data Science for Business

Bioinformatics and Computational Biology

Build a search engine for the knowledge of the course about Introduction to Programming based on ontology Rela-model

Xuan-Thien Pham[1,6], Tuan-Vi Tran[2,6], Van-Thanh Nguyen-Le[3], Vuong T. Pham[4], Hien D. Nguyen[5,6,*]
[1,2,5]*University of Information Technology, Ho Chi Minh City, Vietnam*
[3]*HCMC University of Techonology and Education, Ho Chi Minh City, Vietnam*
[4]*Sai Gon University, Ho Chi Minh City, Vietnam*
[6]*Vietnam National University, Ho Chi Minh City, Vietnam*
Email: {[1]18520158,[2]18520245}@gm.uit.edu.vn, [3]thanhnlv@hcmute.edu.vn, [4]vuong.pham@sgu.edu.vn, [5]hiennd@uit.edu.vn
[*]Corresponding author: Hien D. Nguyen

***Abstract*—STEM education is the modern educational method. The knowledge of programming is very important for studying in STEM education, especially in information technology. In this paper, an intelligent search engine for the knowledge of the course about Introduction to Programming is constructed. The knowledge model is organized based on the improved ontology of Rela-model, which represents concepts of the course, relations between those concepts and some inference rules to query the knowledge of programming. Based on this ontology, the problems for querying the knowledge of the course are proposed based on their semantic. The search system has been tested in the real-world by students studying the course. The experimental results are positive for the effectiveness of the system.**

***Index Terms*—search engine; knowledge of programming; knowledge base; ontology.**

I. Introduction

In the Fourth Industrial Revolution, the systems of manufacturing cooperate with each other in a flexible way at the global level [1]. The skill of programming is one of the key knowledge in education. Besides, STEM education is also the modern educational method, it helps students to develop their skills in science and technology [2], [3]. The knowledge of programming is very important for studying in STEM education, especially in information technology (IT).

Nowadays, there are many intelligent educational systems supporting the learning of students. The study in [4] uses a general Neural Cognitive Diagnosis (NeuralCD) framework, which incorporates neural networks to learn the complex interactions between student's and exercise's factor vectors. This method can diagnosis the students' proficiency level on specific knowledge concepts through results of their exercises. This method got the good experimental results in mathematics. In [5], a method for finding similar exercises in large-scale online education systems is developed based on Multimodal Attention-based Neural Network (MANN) framework by learning a unified semantic representation from the heterogeneous data. However, those systems have not yet supported the querying and searching the knowledge of programming.

Ontology is a data model that represents a field and is used to infer objects in that field and the relationship between them. Ontology provides a general vocabulary set of concepts, essential attributes, and definitions of these concepts and attributes [6], [7]. Ontology provides constraints, which are sometimes referred to as fundamental assumptions about the desirable meaning of the vocabulary set. It is used in a domain that can be communicated between people and other heterogeneous distributed application systems [8].

In the IT curriculum at university, the knowledge of the course about Introduction to Programming is very important. It is the foundation knowledge of IT students. In this paper, a method for building an intelligent search system on the knowledge of programming in this course is constructed. The structure of ontology Rela-model [9], [10] is applied to organize the knowledge model of this system. This ontology includes components representing concepts of the course, relations between those concepts and some inference rules to query the knowledge of programming. Based on this ontology, the problems for querying the knowledge of the course are proposed. These problems solve the matching the queries with the represented knowledge of this course based on their semantic.

II. Related Work

There are many search engines built based on ontology. However, they have not yet met some requirements of the application for supporting of learning in courses.

A model of Classed Keyphrase based Ontology (CK-ONTO) describes the domain knowledge, a database of document repository, semantic representations for documents [11]. This model is used to design advanced search techniques based on measuring semantic similarity. It is applied to build the Job-posting retrieval system. However, this method does not solve problems with the query as natural language.

978-1-7281-4510-5/20/$31.00 ©2020 IEEE

Ontology-L is used to represent to present the knowledge domain about the law law of public investment for cost estimation in the information technology field [7]. It includes concepts about expenses in a project, rules for computing expenses, law documents. This ontology is applied to build a support system to plan an IT hardware project based on law of public investment. Nonetheless, ontology-L needs to improve for representing the knowledge of programming courses in education to construct the intelligent search systems.

In [12], authors proposed the knowledge model to representing the knowledge of algorithms. This knowledge model is the integrating between ontology and frames [13]. Those methods are utilized to construct the the intelligent systems for learning courses about algorithms. Those systems can illustrate algorithms' process and the working of data structures. However, the function of querying the content of lessons in the course still has some limitations. It does not query related knowledge with the current results.

Besides, in [14], the author proposed two approaches to discover natural definitions of knowledge and wisdom using graph theory: The metrics approach is to produce graphs that force an increase in various graph metrics, and the dimensions approach is based on the observation that the graphical representation of aggelia in the DIKW (Data, Information, Knowledge and Wisdom) hierarchy seems to increase in dimension with each step up the hierarchy. Moreover, the ontology in [15] includes basic concepts in the agricultural domain. It also contains geographical, IoT, and other knowledge extracted from various datasets. Using this ontology, the agricultural data collected from many different data resources can be linked between them. It is a framework of knowledge for agriculture which can be used for the purpose of smart agriculture systems. However, those approaches are not effective to build the intelligent system supporting the searching and querying on the real-world knowledge, especially in education.

III. Knowledge model for the course about Introduction to Programming

Rela-model is an ontology including concepts, relations between concepts and inference rules of knowledge domain [9]. It is effective for representing the knowledge of relations to build intelligent educational systems [3], [16]. In this section, based on the foundation of ontology Rela-model, a knowledge model for representing the knowledge domain about programming in the course about Introduction to Programming is proposed. The structure of concepts in this model is innovated suitably for building an intelligent search system. Thus, the structure of relations between concepts needs to be improved too. The improved model is applied to design a knowledge base of the knowledge base of the search system.

Definition 1 [9]: Rela-model is a knowledge model of relations. It is a tube including three components as follows:

$$K = \textbf{(C, R, Rules)}$$

In which:

- **C** is a set of concepts. They are concepts about the knowledge of programming in the course.
- **R** is the set of relations between concepts in **C**.
- **Rules** is a set of inference rules in the knowledge domain.

A. *C - set of concepts*

Each concept c $\in$ **C** is a class of objects. It has the structure as follows:

(Attr, Inner-relation, Properties, Key)

In which:

- *Attr* is a set of attributes of a concept
- *Inner-relation* is the set of relations between attributes in c_i belong to **C**.
- *Properties* is the set of properties of the corresponding concept.
- *Key* is a set of keywords related to the corresponding concept.

Example 3.1:

- ALGORITHM - the concept about algorithms
  Attr = {Name, Chapter, Content}
  - Name: The name of Algorithm.
  - Chapter : The chapter which algorithm belong to.
  - Content: The definition and content of the algorithm.

  Inner-relation = {Belong to}
  - *Name* Belong to *Chapter*

  Properties = {Operations}
  Key = set of keywords of the algorithm
- POINTER - concept about Pointer
  Attr = {Name, Chapter, Content, Expression, Usage}
  - Name: The name of Pointer
  - Chapter : The chapter which Pointer belong to.
  - Content : The definition and content of the pointer.
  - Expression : The set of pointers in the lesson.
  - Usage : How to use of pointer.

  Inner-relation = {Belong to}
  - *Name* Belong to *Chapter*

  Properties = { Static memory allocation, Dynamic memory allocation, new, delete}

B. *R - set of relation*

There are two kinds of relations between concepts in **C**:

$$\mathbf{R} = R_{hirerachy} \cup R_{related\_to}$$

In which:

- $R_{hirerachy}$ is a set of hierarchical relations between concepts and can be considered as the inheritance relations. They are relations as "is_a".

  c_i is_a c_j: It means c_i is sub-concept of concept c_j (c_j, $c_j \in$ **C**).

  And we have, $c_j.Attr \subseteq c_i.Attr$.

  Example 3.2:
  - "High level programming language" is_a "Programming language"

- "Multidimensional array" is_a "Array", "Multidimensional array" has all of attributes that "Array" has like Name, Usage, etc. .

- $R_{related\_to}$ is a set of related knowledge with the main knowledge. It helps to deduce the related knowledge when the machine operate the proof layer.

  c_i related_to c_j: It means c_i have the relationship with c_j (c_i, $c_j \in$ **C**) and it is also a 2-way relationship.

  Example 3.3:
  - "Program" related_to "Programming language"
  - "Statement" related_to {"DataTypes", "Key word", "Typedef"}. It means:
    * "Statement" related_to "DataTypes"
    * "Statement" related_to "Key word"
    * "Statement" related_to "Typedef"

C. Rules - set of rules

Each rule $r \in$ **Rules** is an inference rule, it has the form: $u(r) \rightarrow v(r)$, which $u(r), v(r)$ are set of facts. Facts are concrete statements about "properties of relations", "relations between concepts", "relations between attributes of concepts".

Example 3.4: $\forall c_1, c_2 \in$ **C**:

- $r_1 : \{c_1 \text{ related\_to } c_2\} \rightarrow \{c_2 \text{ related\_to } c_1\}$
- $r_2 : \{c_1 \text{ is\_a } c_2\} \rightarrow \{c_2.Attr \subseteq c_1.Attr\}$

IV. Some problems for querying the knowledge of programming

When a query for retrieve the knowledge of programming is inputted, it will be processed to extract some main keywords. After that, those keywords are compared to match with the content in the knowledge base of the search system based on their semantic. The matching is worked by using the comparison with the structure of the knowledge model's components, especially the concepts and relations. The inference rules of the knowledge model helps to deduce some more relationships related to the content of the query. There are some main problems for querying the knowledge of programming in the search system:

(1) Problem 1: *Processing of the query*. From the inputted query as Vietnamese text, the problem will extract the main keywords of this query. Those words are used to retrieve the suitable knowledge from the knowledge base.

(2) Problem 2: *Matching the keywords with the content of the knowledge*. Based on the set of keywords extracted from the query, the method to compare the similarity between the meaning of keywords and the content in knowledge base is proposed. This method will determine the required knowledge's content for the inputted query. Besides, this problem also solves the determination of the knowledge related to the knowledge of query.

A. Classification of the query

In the search system on the knowledge domain, the classification of a query helps to reduce the searching space, and increase the accuracy of the searching. In this paper, the query is divided into two kinds:

- "What is" query is the query to determine the definitions, properties, and relations of concepts or attributes of concepts.
- "Compare" query is the query to find the difference and the similarity between two concepts and attributes of concepts.

Example 4.1:
- *"What is Algorithm?"* or *"Algorithm"*: The result of this query is the definition of Algorithm, the chapter which Algorithm belongs to and the other related knowledge, such as some popular algorithms.
- "What's the difference between integer array and string ?": The result of this query is the definitions of integer array and string, the chapter which integer array and string belong to, and the other related knowledge, such as some kinds of exercises for each concept. It shows the results as a comparison table.

B. Processing of the inputted query

The first stage of the search system is to process the inputted query to extract the main keywords. The performance of processing queries plays an essential role in the subsequent stages. Because it contributes directly to limit the search space, make the acceleration of search-processing and increasing accuracy. The queries-processing is illustrated in detail in Fig 1. This process consists of three main parts: word segmentation, chunking and select candidates.

Word segmentation is one of the most important tasks which has a considerable impact on language processing. In this system, we use Underthesea - Vietnamese NLP Toolkit for handling not only queries segmentation but also chunking stage [1].

Chunking is a process of meaningful extracting phrases from unstructured sentences. Chunking is very useful when we want to extract constituents from sentence [17]. The Underthesea is also used for the chunking process. The core idea is the choosing of the phrases, which have chunking tags: NP (Noun phrase), VP (Verb phrase) and QP (Question phrase), as candidate keywords. While NP and QP contain the key information of queries, QP determines the type of queries for query classification that will be presented in subsequent section.

Candidates selection is the final step of the query processing. There are a lot of NP and VP in a chunking tree that need to process to select the most meaningful features. Based on the built ontology, the heuristic approach is used for selecting keywords. Three sets of keywords are sorted by increasing level of internal node, in which the limitation is three nodes for each set.

C. Matching the keywords with the content of the knowledge

After extracting the main keywords from the inputted query, the search system will compare the similarities between those

[1] We use the latest version 1.1.17 downloaded from https://pypi.org/project/underthesea

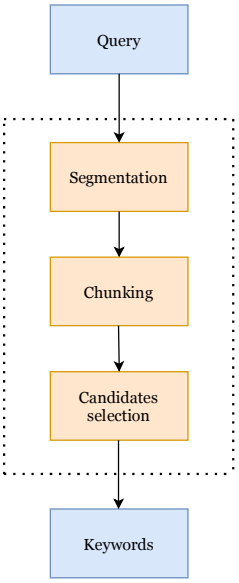

Fig. 1. *The processing of query*

keywords and the knowledge's content from the knowledge base as Rela-model represented in section III. The technique for matching to design the search engine is described as Fig 2.

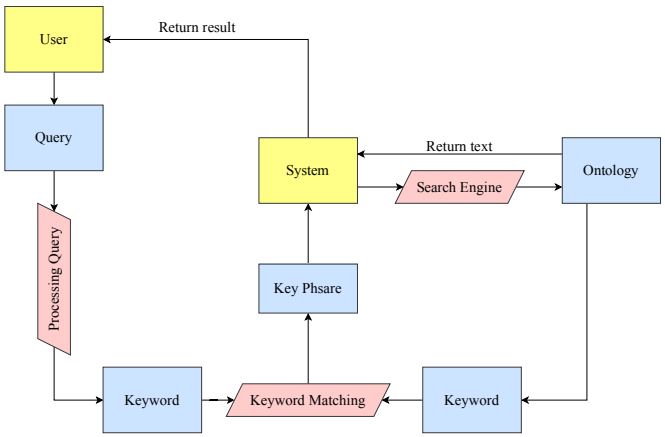

Fig. 2. *The technique for matching of the search engine*

The dictionary is the set of keywords in the course about Introduction to Programming. At the first step, the dictionary is established from the knowledge base of the search system. It also contains question words that help to classify the kinds of queries.

Example 4.2: Some words in the dictionary:

- Some individuals in the dictionary: *Array, Algorithm, Define, Variable, Usage*.
- The synonyms of a keyword in dictionary: *"How to use"* equivalent *"Usage"*, or *"What is"* equivalent *"Define"*.

At the second step, the extracted keywords of the inputted query is compared with the dictionary to create a set of key phrases. The search engine uses those key phrases to retrieve the knowledge of programming from the ontology. Moreover, from the retrieval results, the system also search the related knowledge based on the relations of those results and the inference rules of the knowledge base. Finally, all results will be returned as a document in the form for each kind of the query.

V. Experimental Results

A. The architecture of the search system on knowledge of programming

The system for searching of the knowledge of programming have a knowledge base which is organized based on the ontology as section III and a search engine with the described problems as section IV. The architecture of the search system is shown in Fig 3

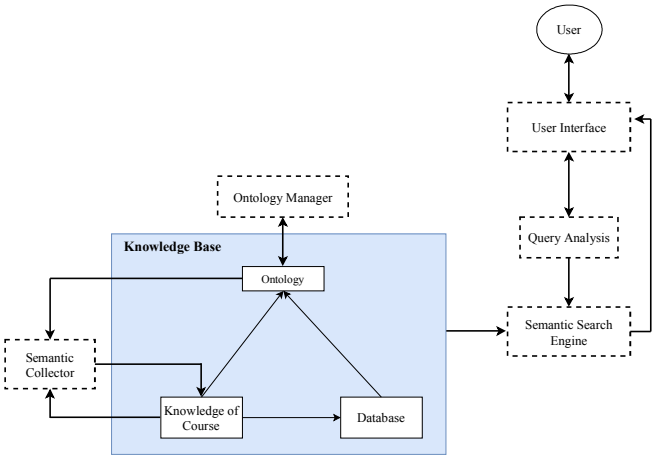

Fig. 3. *The architecture of the search system on the knowledge domain*

Fig 3: The knowledge of course is collected from the books [18], [19]. It consists of concepts, relations and rules in the course. This collecting knowledge can be classified in some ways such as chapters, topics, subjects, and problems, exercises in the course. Some examples of those knowledge are the database of the knowledge of course. Based on the collecting knowledge and database, the ontology representing the knowledge of course is represented by using the structure of Rela-model. This ontology is the knowledge base of the search system. The problems for searching and querying knowledge are worked on this knowledge base. The search engine will retrieve the results for the inputted query.

The most fundamental ontology languages include XML/XML Namespace/XML Schema, RDF, RDFSchema and OWL [20]. In reality, RDFS is a standard vocabulary just like RDF. The RDF has terms for creating instances, RDFS has terms for creating classes. Using both of them, the knowledge can be described more detailed. OWL terms are defined by using RDF and RDFS terms. In the system for searching the knowledge of the course about Introduction to programming, the ontology Rela-model of its knowledge base is built on RDF model and RDFS. This knowledge domain

is represented on the web and performed the semantics information in a way that the computer can understand.

B. Testing

Example 5.1: Query:*"Thuật toán là gì ?"* (It means *"What is Algorithm ?"*).

After the processing of query, we have a set of keywords S_1: {"Thuật toán", "là gì"} (it means {"Algorithm", "What is"}). The set S_1 is compared with the dictionary extracted from the knowledge base, the world "là gì" is the question word to classify the kind of the query. The system determines the query is a "What is" query. The main keyword is "Thuật toán" (Algorithm).

- Definition of Algorithm-concept: is a finite sequence of well-defined, computer-implementable instructions, typically to solve a class of problems or to perform a computation.
- Belong_to "Chapter 2": it means the lesson about "Algorithm" is in "Chapter 2"
- Besides, it also shows the other knowledge related to "Algorithm", such as "The complexity of algorithm", "Sorting algorithms", "searching algorithm".

The related knowledge are the concepts or attributes which have the direct or transitive relations with the main keywords.

Example 5.2: Query: *"So sánh mảng đa chiều và mảng 1 chiều số nguyên?"*(It means *"What is the difference between multi-dimensional array and one-dimensional array?"*.)

After the processing of query, we have a set of keywords S_2:{"Mảng 1 chiều", "so sánh", "Mảng đa chiều"} (it means {"one-way array", "What is the difference", "multi-dimensional array"}). The set S_2 is compared with the dictionary extracted from the knowledge base, the word "so sánh" is the question word to classify the kind of query. The system determines the query is a "Compare" query. Because this is a "Compare" query, the main keywords is the words representing compared objects. In this query, the main keywords are "one-dimensional array" and "multi-dimensional array".

- Definition of "one-dimensional array" is: a type of linear array. Accessing its elements involves a single subscript which can either represent a row or column index.
- Definition of "multi-dimensional array" is an array that has more than one dimension. It is an array of arrays; an array that has multiple levels. The simplest multi-dimensional array is the 2D-array, or two-dimensional array. It's technically an array of arrays, as you will see in the code. A 2D array is also called a matrix, or a table of rows and columns.
- Belong_to "Chapter 6": it means the lesson about "one-dimensional array" and "multi-dimensional array" is in "Chapter 6".
- Usage of "one-dimensional array" and "multi-dimensional array"
- How to Install "one-dimensional array" and "multi-dimensional array"
- Besides, it also shows the other knowledge related to arrays, such as "Data-types", "Identifier", "Operators on array".

C. Experimental results

The search system for the knowledge of the course about Introduction to Programming is tested by IT students studying this course at University of Information Technology and Sai Gon University (Ho Chi Minh city, Vietnam). The students input the queries for searching the knowledge of this course. Those queries are two kinds: "What is" and "Compare" queries. The results of the inputted queries have been double checked:

- If the students recognizes the results being suitable for them, they will check on the system.
- Besides, those results are noted. They will be evaluated by the lecturers who are teaching this course.
- If the results are checked by students and lecturers, they will be correct.

There are 652 queries has been inputted by students. The experimental results is shown in Table I:

TABLE I
THE RESULTS FOR TESTING OF QUERIES

Kind	Quantity	Correct	Rate
"What is" queries	542	337	62%
"Compare" queries	110	58	53%
Total	**652**	**395**	**61%**

In the kind of "What is" query, there are some words in dictionary cannot be separated, so they cannot match with the keywords in the query. Hence, some "What is" queries do not have correctly results. In the kind of "Compare" query, some queries are not able to parse to compared objects, so those queries do not return the correctly results of comparison.

However, the method of the search system can retrieve some complex queries including many requirements for many kinds of objects based on the knowledge of programming. If this system is integrated some improved techniques for Vietnamese processing [21], it will be increase the ability to match and extract the knowledge from the knowledge base.

VI. CONCLUSIONS AND FUTURE WORKS

E-learning is the modern method for studying in the era of information technology. Knowledge of programming is a piece of important knowledge for many people. In this paper, a method for building an intelligent search system on this knowledge domain is proposed. The knowledge base of this system is organized based on the improvement ontology of the Rela-model. It includes concepts of programming, relations between those concepts and some inference rules to query the knowledge of programming. Using this knowledge base, the techniques for querying the knowledge of the course are proposed and solved: the extracting of keywords from an

inputted query and the matching keywords with the content of the knowledge base. Moreover, the search system also can retrieve the knowledge related to the semantic of the query.

The current intelligent search system is a design based on the knowledge of course about Introduction to Programming in the IT curriculum at university. It is built based on RDF and RDFS models and runs on the web. The system can process the Vietnamese queries in two kinds "What is" and "Compare". It can suggest the knowledge which is related to the content of the query.

Besides, the data was built based on RDF and RDFS models. So we can apply for another course that can be built in the same structure. This will support a lot of stages in the future if you want to approach for another field.

Vietnamese is a language isolated [22]. The meaning of Vietnamese sentence belongs to the way for organizing of its predicates [23], especially the position of words. In the future, the technique for processing the Vietnamese query is studied base on its grammar structure [21]. This work will make the semantic analysis more accuracy. Moreover, the intelligent search system will be also constructed as an intelligent chatbot [24]. Through the chatbot, the interaction with the users is more easily and exhaustively. In that case, the intelligent chatbot helps to trace the behaviors and analysis the knowledge level of users. From that, the system tends to a personal search system satisfying the requirements of each user.

REFERENCES

[1] K. Schwab, *The Fourth Industrial Revolution*. Penguin UK, 2017.

[2] Y. Xie, M. Fang, and K. Shauman, "Stem education," *Annual Review of Sociology*, vol. 41, pp. 331–357, 2015.

[3] H. Nguyen, N. Do, N. Tran, X. Pham, and V. Pham, "Some criteria of the knowledge representation method for an intelligent problem solver in stem education," *Applied Computational Intelligence and Soft Computing*, vol. 2020, Article ID 9834218, 2020.

[4] F. Wang, Q. Liu, E. Chen, and Z. Huang, *Interpretable cognitive diagnosis with neural network for intelligent educational systems*, 2019. arXiv: 1908.08733.Prepint [cs.CL].

[5] Q. Liu, Z. Huang, Z. Huang, C. Liu, E. Chen, Y. Su, and G. Hu, "Finding similar exercises in online education systems," in *2018 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2018)*, 2018, pp. 1821–1830, London, UK, Aug. 2018.

[6] T. Berners-Lee, J. Hendler, and O. Lassila, "The semantic web," *Scientific American*, vol. 284, no. 5, pp. 34–43, 2001. [Online]. Available: http://www.jstor.org/stable/26059207.

[7] H. Do, N. Do, and H. Nguyen, "A consulting system for estimating costs of an information technology hardware project based on law of public investment," in *2019 International Conference on System Science and Engineering (ICSSE)*, 285–290, Vietnam, July 2019.

[8] K.Giri, "Role of ontology in semantic web," *DESIDOC Journal of Library Information Technology*, vol. 31, pp. 116–120, Mar. 2011.

[9] N. Do, H. Nguyen, and A. Selamat, "Knowledge-based model of expert systems using rela-model," *International Journal of Software Engineering and Knowledge Engineering (IJSEKE)*, vol. 28, no. 8, pp. 1047–1090, 2018.

[10] N. Do and H. Nguyen, "A knowledge model about relations and application," in *Proceedings of 4th International Conference on Data Mining and Intelligent Information Technology Applications (ICMIA 2012)*, Taiwan, Dec. 2012, pp. 701–704.

[11] T. Huynh, A. Pham, and N. Do, "A method for designing domain-specific document retrieval systems using semantic indexing," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 10, pp. 461–481, 2019.

[12] H. Nguyen, N. Do, T. Mai, and V. Pham, "A method for designing the intelligent system in learning of algorithms," in *Proceedings of 18th International Conference on Intelligent Software Methodologies, Tools, and Techniques (SOMET 2019)*, Kuching, Malaysia, Sep. 2019, Frontiers in Artificial Intelligence and Applications, vol. 318, pp. 658 –671.

[13] T. Le, S. Luu, H. Nguyen, and N. Do, "Knowledge representation method for designing an intelligent tutoring system in learning of courses about algorithms," in *Proceedings of 2019 25th Asia-Pacific Conference on Communications (APCC 2019)*, Ho Chi Minh, Vietnam, Nov. 2019, pp. 310–315.

[14] M. Atkins, "Two approaches toward graphical definitions of knowledge and wisdom," in *Natarajan Meghanathan et al. (Eds): SIPP, BIGML, DaKM, SOEN, AISC - 2020*, pp. 59–74, March 2020.

[15] H. Ngo, A. Le, and T. Kechadi, "Ontology based approach for precision agriculture," in *Proceedings of 2018 Multi-disciplinary Trends in Artificial Intelligence (MIWAI 2018)*, Springer, pp. 175–186, Hanoi, Vietnam, Nov. 2018.

[16] N. Do, H. Nguyen, and T. Mai, "A method of ontology integration for designing intelligent problem solvers," *Applied Sciences (Appl. Sci.)*, vol. 9, no. 18, p. 3793, 2018.

[17] M. Peters, W. Ammar, C. Bhagavatula, and R. Power, "Semi-supervised sequence tagging with bidirectional language models," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017)*, Vancouver, Canada, Jul. 2017, Vol. 1, pp. 1756–1765.

[18] T.Tran, P. Nguyen, T. Dinh, and T. Tran, *Introduction to Programming*. Publishes of Science and Technology. (Vietnamese), 2015.

[19] J. Hubbard, *Theory and Problems of Fundamentals of Computing with C++*. Schaum's Outlines Series, McGraw-Hill, 1998.

[20] S. Decker, S. Melnik, F. Harmelen, D. Fensel, M. Klein, M. Erdmann, and I. I. Horrocks, "The semantic web: The roles of xml and rdf," *IEEE Internet Computing*, vol. 4, Oct. 2000.

[21] H. Nguyen, T. Huynh, S. Hoang, V. Pham, and I. Zelinka, "Language-oriented sentiment analysis based on the grammar structure and improved self-attention network," in *Proceedings of 15th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE 2020)*, Prague, Czech Public, May 2020, pp. 339–346.

[22] B. Nguyen, H. Nguyen, and a. L. V. L. Romary, "Lexical descriptions for vietnamese language processing," *Language Resources and Evaluation*, vol. 40, no. 3/4, Asian Language Processing: State-of-the–Art Resources and Processing, pp. 291–309, 2006.

[23] M. Clark, *Passive ane Ergative in Vietnamese*. In: Nguyen Dang Liem (ed.), South-East Asian Linguistic Studies, 1974.

[24] L. Zhou, J. Gao, D. Li, and H. Shum, "The design and implementation of xiaoice, an empathetic social chatbot," *Computational Linguistics*, 2020. [Online]. Available: https://doi.org/10.1162/coli_a_00368.