BUILDING A VIETNAMESE DIALOG MECHANISM FOR V-DLG~TABL SYSTEM

An Hoai Vo and Dang Tuan Nguyen

Faculty of Computer Science, University of Information Technology, Vietnam National University – Ho Chi Minh City

ABSTRACT

This paper introduces a Vietnamese automatic dialog mechanism which allows the V-DLG~TABL system to automatically communicate with clients. This dialog mechanism is based on the Question – Answering engine of V-DLG~TABL system, and composes the following supplementary mechanisms: 1) the mechanism of choosing the suggested question; 2) the mechanism of managing the conversations and suggesting information; 3) the mechanism of resolving the questions having anaphora; 4) the scenarios of dialog (in this stage, there is only one simple scenario defined for the system).

KEYWORDS

Semantics, Automatic Dialog Mechanism, Question – Answering Engine, Vietnamese Language Processing.

1. INTRODUCTION

In this paper, we focus on introducing the Vietnamese automatic dialog mechanism of our V-DLG~TABL system. This automatic dialog mechanism allows the system to communicate effectively with clients, helps them know the needed information about the tablets which they expect to find.

The V-DLG~TABL system's automatic dialog architecture composes the following supplementary mechanisms: 1) the mechanism of choosing the suggested question; 2) the mechanism of managing the conversations and suggesting information; 3) the mechanism for resolving the questions having anaphora; 4) the scenarios of dialog (in this first time, we only developed one simple scenario for the system). This automatic dialog mechanism is used to build the component "V-DLG~TABL Dialog" of the system.

In addition, the automatic dialog mechanism of our system bases on an important engine of answering Vietnamese questions about tablets. This Vietnamese language based Question – Answering engine is built by applying the theoretic methods and logic programming techniques of computational semantics proposed by Patrick Blackburn and Johan Bos [1], and by reusing and referring to some basic implementations (including technical solutions and program source codes) of Nguyễn Thành and Phạm Minh Tiến [2].

2. USING SIMPLE LAMBDA EXPRESSION FOR REPRESENTING THE SEMANTICS OF VIETNAMESE QUESTIONS

To represent and analyze the semantics of Vietnamese questions, we apply the approaches, methods and implementation techniques which are proposed in [1], [2]. According to [1], the

DOI : 10.5121/ijnlc.2014.3104

semantics of sentences, including questions, can be represented by lambda expressions (cf. [1], [2]):

- Following [1], each unidentified part of a lambda expression is symbolized by a variable going after an operator λ, and the operator @ is used to assign an argument to a variable in a lambda expression (cf. [1], [2]).
- Following [1], the β-transformation uses operator @ to combine the lambda expressions of syntactic elements with each others. We use the algorithm proposed in [1], which is based on β transformation, to combine the semantic components of the Vietnamese question (cf. [1], [2]).

Based on [1], we propose a semantic analysis algorithm for Vietnamese questions in V-DLG~TABL system as follows:

ALGORITHM:

1) Step 1: Remove the supplementary syntactic elements in the syntactic tree of the question. The principal syntactic elements are retained: n_tablet, pn_tablet_name, n_component_<name>, n_property_<name>, and "literal".
2) Step 2: Identify a lambda expression corresponding to each syntactic element which is retained after removing the supplementary syntactic elements.

   Based on the proposed methods in [1] and their applications in [2], the lambda expressions corresponding to the retained syntactic elements are defined as follows (cf. [1], [2]):

Syntactic element	**Lambda expression**
pn_tablet_name	λx.x@<tablet_name>
n_component_<name>	λx.component_name(x)
n_property_<name>	λx.λy.property_name(x, y)
literal	λx.x@<literal>

3) Step 3: Apply β-transformation [1] to combine the lambda expressions of syntactic elements with each other (cf. [1], [2]). The complete FOL (First-Order Logic) expression can represent the semantics of Vietnamese question.

Example 1: *"Máy tính bảng Nexus 7 có kích thước màn hình là bao nhiêu?"*

(English: *"How much Nexus 7 tablet screen size is"*)

The system resolves this question in the following steps:

- Step 1: The syntactic tree of the question in example 1 is presented in Figure 1.

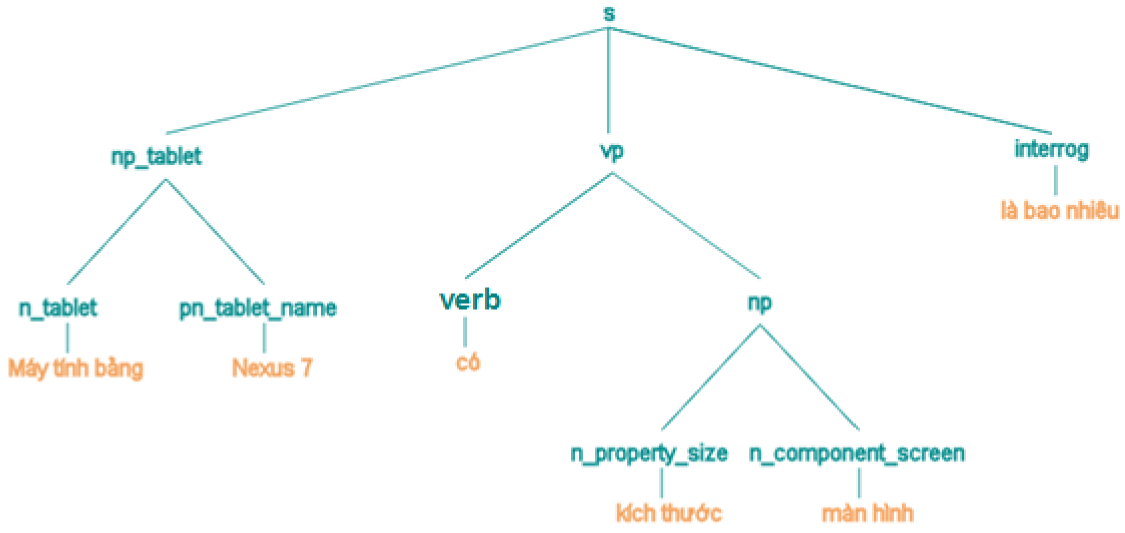

Figure 1: The syntactic tree of the Vietnamese question in example 1

- Step 2 (a): Remove the supplementary syntactic elements in the syntactic tree. The syntactic tree becomes as in Figure 2.

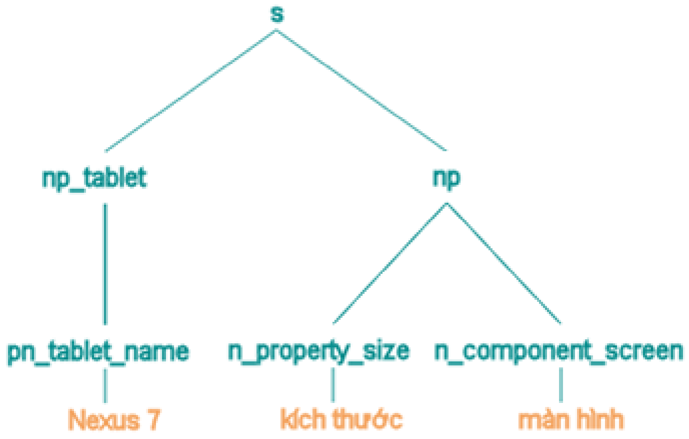

Figure 2: The syntactic tree of the Vietnamese question in example 1 (after removing the supplementary syntactic elements)

- Step 2 (b): Determine the lambda expressions corresponding to each retained syntactic element:

  + The lambda expression for the syntactic element n_property_size: $\lambda x.\lambda y.$kích_thước(x, y).

  + The lambda expression for the syntactic element n_component_screen: $\lambda x.$màn_hình(x).

  + The lambda expression for the syntactic element pn_tablet: $\lambda x.x$@Nexus_7.

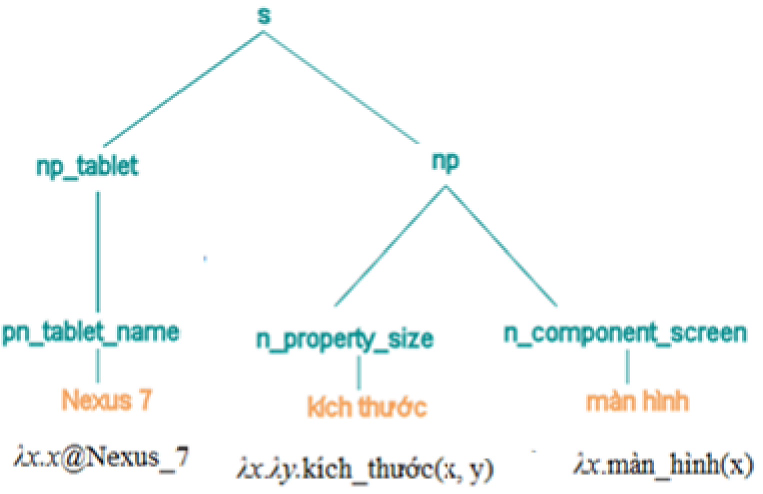

Figure 3: The lambda expressions corresponding to retained syntactic elements of the Vietnamese question in example 1

- Step 3: The lambda expressions representing retained syntactic elements will be combined by using β transformation, in the following steps:

- Combine ingredients lambda expressions to create a new lambda expression:

  [λt.t@Nexus_7]@[λx.λy.kích_thước(x, y)]@[λz.màn_hình(z)]

- Apply β transformation for ingredients lambda expressions to create a completed lambda expression:

  [λt.t@Nexus_7]@([λx.λy.kích_thước(x, y)]@[λz.màn_hình(z)])

  = [λx.x@Nexus_7]@[λy.kích_thước(λz.màn_hình(z), y)]

  = [λx.x@Nexus_7]@[λz.λy.kích_thước(màn_hình(z) , y)]

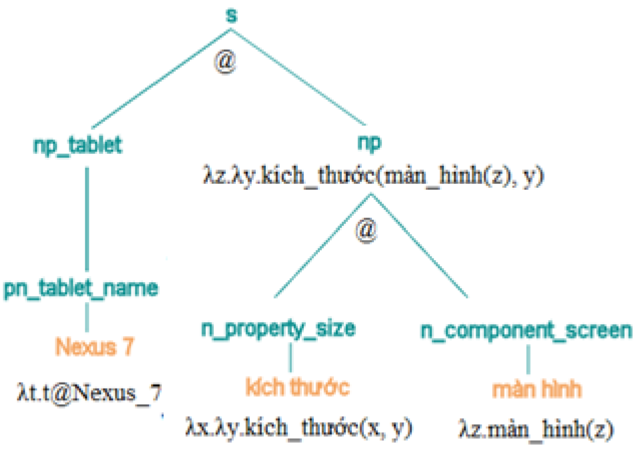

Figure 4: Each step for applying β transformation to create a completed lambda expression of the Vietnamese question in example 1

- Continue to apply β transformation for the other side of the lambda expression:

= [λz.λy.kích_thước(màn_hình(z), y)]@Nexus_7

= [λy.kích_thước(màn_hình(Nexus_7), y)]

= λy.kích_thước(màn_hình(Nexus_7), y)

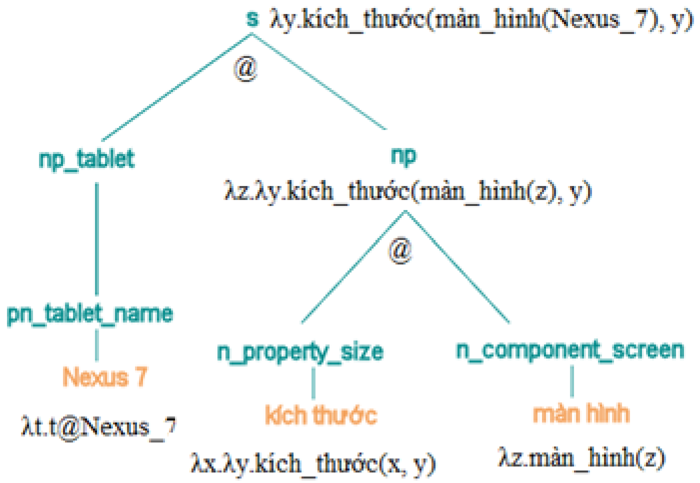

Figure 5: The completed lambda expression of the Vietnamese question in example 1 (after applying β transformation)

- Finally, the completed lambda expression of the question in example 1 is: λy.kích_thước(màn_hình(Nexus_7), y). In this lambda expression, y is the unidentified variable, and represents the value of the screen size asked in the question.

3. ANSWERING VIETNAMESE QUESTIONS ABOUT TABLETS

To answer Vietnamese questions in V-DLG~TABL system, we need to define the data to describe the information about the tablets. Based on [1], we apply the methods and techniques of Nguyễn Thành and Phạm Minh Tiến [2] to define descriptive facts and organize the modules in Prolog.

When a Vietnamese input question is entered into the system, the component "V-DLG~TABL Syntactic Parsing and Processing" will analyze it to return its syntactic tree. The retained syntactic elements of the syntactic tree, which are determined by our algorithm mentioned above, will correspond to the components of the "information structure model" proposed in [2]. These retained syntactic elements are then processed by the component "V-DLG~TABL Semantic Analyzing" to determine the lambda expression representing the semantics of the Vietnamese question. To create a lambda expression for each Vietnamese question, we use the methods and implementation techniques proposed in [1] (cf. [2]).

> For example, the lambda expression of the question in example 1 is λy.kích_thước (màn_hình(nexus_7), y). In this lambda expression, variable y indicates the unidentified element in the question.

Basing on the methods and techniques of implementation proposed in [1], and referring to their applications presented in [2], the system uses the β-transformation to reduce the lambda expression of the Vietnamese question, and acquires the complete FOL expression; then, this FOL expression is used to query the database of facts in Prolog and returns the found result. Following the techniques proposed in [1] (these techniques have been applied in [2], [3], [4], [5]), the querying mechanism is implemented by using the unification between the FOL expression of the question and the facts of the system (cf. [1], [2]).

> Example 2: Based on [1], [2] the lambda expression λy.kích_thước(màn_hình(nexus_7), y) of the Vietnamese question in example 1 will be unified with the fact kich_thuoc(man_hinh(nexus_7), '7 inch')) in the database of facts. The unification result of y is "7 inch". The value of y is the screen size of the tablet nexus_7.

To create Vietnamese answer in V-DLG~TABL system, we apply the method which is developed in [2] (the original idea of this approach was proposed in [7], [9]): based on the unidentified node in the syntactic tree of the question, the system queries the database of facts to find the data, and creates the answer.

> Example 3: The system performs the following steps to resolve Vietnamese question in example 1:

- Step 1: The syntactic structure of the question is presented in Figure 6.

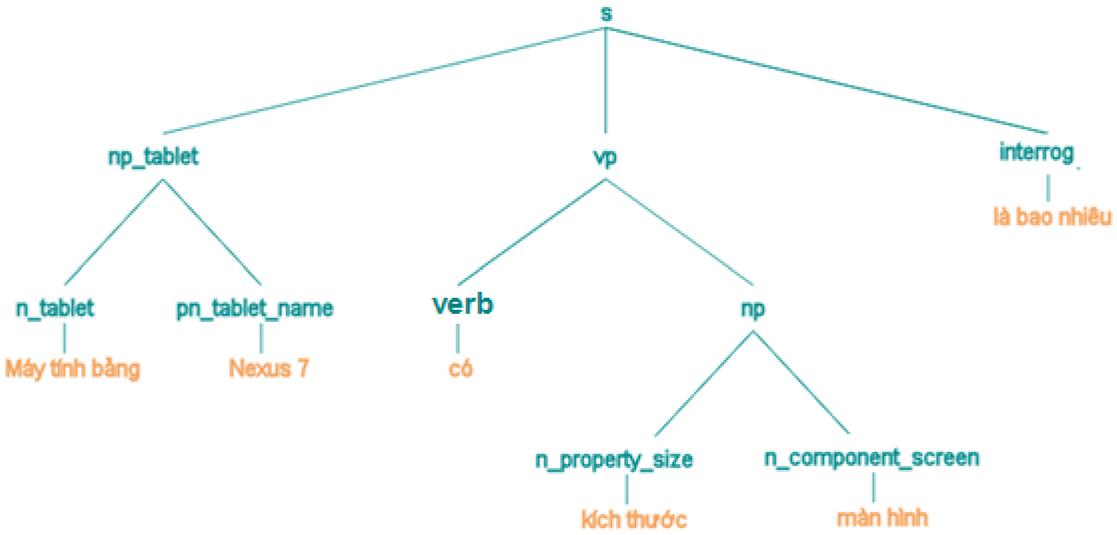

Figure 6: The syntactic tree of the Vietnamese question in example 1

- Step 2: Query and take the result: "7 inch".
- Step 3: Find the value which answers the interrogative node in the syntactic tree: see Figure 7.

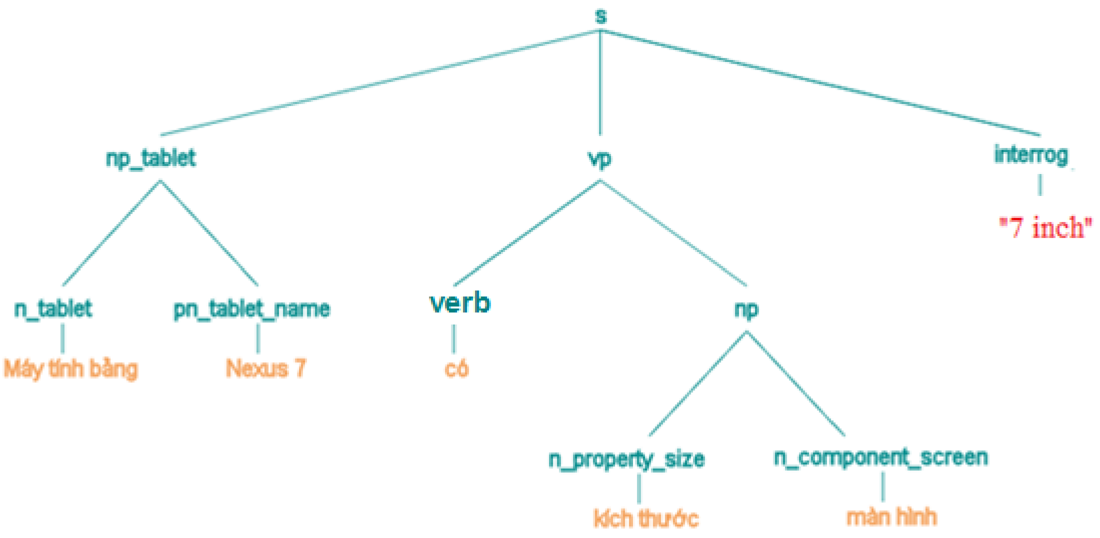

Figure 7: The interrogative information is answered in the syntactic tree of the Vietnamese question in example 1

- Step 4: Create the answer based on combining the information components in the syntactic tree, from left to right: *"Máy tính bảng Nexus 7 có kích thước màn hình 7 inch"*.

4. Building Dialog Mechanism of System

The dialog mechanism of V-DLG~TABL system is built basing on the following mechanisms:

- The mechanism for choosing the suggested question.
- The mechanism for managing the dialog and suggesting the information.
- The mechanism for resolving the questions having anaphora.
- The sample scenarios.

4.1. The mechanism for choosing the suggested question

The mechanism of choosing the suggested question is based on some predefined samples of suggested questions. Depending on the communication scenario between client and the system, the system will create the appropriate suggested questions.

- Type 1: These are the suggested questions about other information of the tablet which the client has just been interested in.

  For example, as soon as the client asks the information about screen, size, and weight of a tablet, the system suggests a question about its color.

- Type 2: These are the suggested questions about the tablets which similar to the one which client has just been interested in.

  For example, after the client asks about tablet Nexus 7 having screen size '7 inch', the system can suggests some more different tablets having the same screen size.

- Type 3: These are the suggested question about other information when the above cases do not occur. These questions can be: *"Bạn có muốn biết thêm thông tin nào không?"* (English: "Would you like to know any information more?"), or *"Bạn có muốn biết thêm* *thông tin về máy tính bảng nào khác không?"* (English: "Would you like to know any information more about another tablet?"), ...

Table 1 presents the samples of suggested questions which are defined in V-DLG~TABL system.

Table 1: The samples of suggested questions

Suggested question types	Elementary questions
Type 1	Question 1.1: <component> Question 1.2: <component> <tablet> Question 1.3: <property>, <component>, <tablet> Question 1.4: <component>, <tablet >
Type 2	Question 2.1: <tablet> Question 2.2: <same_tablet>
Type 3	Question 3.1: <what_tablet> Question 3.2: <what_component> Question 3.3: <what_info>

Table 2 presents some examples for the suggested question samples.

Table 2: Some examples for the suggested question samples

Suggested question types	Elementary questions
Type 1	Question 1.1: Bạn có muốn biết thông tin <thành phần> không? Question 1.2: Bạn có muốn biết thành phần <thành phần> của máy tính bảng trên không? Question 1.3: Bạn cần biết thông tin <thuôc tính> của <thành phần> máy tỉnh bảng <X> không? Question 1.4: Bạn có biết thông tin về <thành phần> máy tỉnh bảng <X> không?
Type 2	Question 2.1: Bạn có muốn biết thêm thông tin máy tính bảng <X> hay không? Question 2.2: Bạn có muốn biết thêm thông tin các máy tính bảng cùng loại không?
Type 3	Question 3.1: Bạn có muốn biết thêm thông tin máy tính bảng nào không? Question 3.2: Bạn có muốn biết thêm thông tin các thành phần nào khác không? Question 3.3: Bạn có muốn biết thêm thông tin nào không?

4.2. Managing the conversations and suggested information

The mechanism of managing the conversations and suggested information is integrated into the Question – Answering engine, and based on the following tasks:

- Management of time: operate based on managing downtime between each question and answer, and between user and system.
- Auto generation of suggested question: after identifying the suggested question type, the suggested question is created based on the suggested question samples.

- The mechanism for managing the dialog: This mechanism is integrated into the Question – Answering engine, implemented some basis tasks:
  + Saving to communication information between the client and the system.
  + Managing time with the component "Manage time" and giving the suggestion with the component "Auto-generate the suggested question".

4.3. Revolving the questions having anaphora

In a series of Vietnamese questions and answers following in a dialog, usually there are the existences of anaphora. According to [8] and [9], the anaphora can be omitted, or can refer to a queried object in previous questions.

Based on [8] and [9], we propose a method for treating the anaphora problems existing in Vietnamese questions about tablets.

4.3.1. Anaphora in questions about tablets/components

Example 4: Consider following dialog:

> User: *Nexus 7 sử dụng loại màn hình gì?*
> (English: What type of screen does Nexus 7 use?)
> System: *Nexus 7 sử dụng loại màn hình WXGA.*
> (English: Nexus 7 uses WXGA screen)
> User: *Máy tính bảng đó có kích thước màn hình bao nhiêu?*
> (English: What screen size of the tablet is?)

In the above example, "máy tính bảng đó" at the second question of the client refers to the same object with the "Nexus 7" at the previous question.

4.3.2. Anaphora referring to queried tablets/components in previous questions

Example 5: Consider following dialog:

> User: *Máy tính bảng nào có chức năng đàm thoại?*
> (English: Which tablets have conversation function?)
> System: *Nexus 7 có chức năng đàm thoại.*
> (English: Nexus 7 has conversation function)
> User: *Máy tính bảng đó có kích thước màn hình bao nhiêu?*
> (English: What screen size of the tablet is?)

In the above example, the information about the tablet at the second question of the client can be identified via the queried information in the first question: this anaphora can only be identified after answering the previous question.

4.3.3. Elliptical anaphora in questions about tablets/components

Example 6: Consider following dialog:

> User: *Nexus 7 sử dụng loại màn hình gì?*
> (English: What type of screen does Nexus 7 use?)

System: *Nexus 7 sử dụng loại màn hình WXGA.*
(English: Nexus 7 uses WXGA screen)
User: *Có kích thước màn hình bao nhiêu?*
(English: What screen size is?)

In the above example, the second question of the client does not contain any information about the tablet although this is mandatory information of the question. Therefore, this information is omitted. The system can identify it basing on the previous question.

4.3.4. Resolution of anaphora in V-DLG~TABL system

The anaphora appearing in syntactic tree of question is processed by an algorithm which we propose based on the information in Table 3.

Table 3: Anaphoric nodes in V-DLG~TABL system

Anaphora types	Anaphoric nodes	Descriptive words
Type 1	this_tablet, that_tablet	"máy tính bảng/tablet" + "này / đó / trên"
Type 2		
Type 3	na_tablet	Representing the omitted information in question

We propose an algorithm for resolving the anaphora existing in Vietnamese questions about tablets:

Algorithm:

Input: The syntactic structure of Vietnamese question.

Step 1: Searching anaphoric nodes based on the listed cases in Table 3.

Step 2: For each anaphoric node, searching a node corresponds to the anaphoric node in the syntactic structure of the question.

- Type 1: The corresponding node of an anaphoric node in the current question is the node of noun phrase describing tablet or components in the previous question.
- Type 2: The corresponding node of an anaphoric node in the current question is a queried node needing to be answered in the previous question.
- Type 3: The corresponding node of an omitted anaphoric node in the current question is node of noun phrase describing tablet or components in the previous question.

Step 3: Replacing an anaphoric node with corresponding node in syntactic structure of the question.

4.4. Building the scenario

To set up and build the dialog system, we build a scenario based on mechanism of choosing the suggested question, mechanism of managing the dialog, and mechanisms of resolving anaphora.

Example 7: A dialog follows the scenario:

User: *Nexus 7 sử dụng loại màn hình gì?*

(English: What type of screen does Nexus 7 use?)

System: *Nexus 7 sử dụng loại màn hình WXGA.*

(English: Nexus 7 uses WXGA screen)

User: *Có kích thước màn hình bao nhiêu?*

(English: What screen size is?)

System: *Có kích thước màn hình 7 inch*

(English: the screen size is 7 inches)

…20s

System: *Bạn có muốn biết thông tin bộ nhớ không?*

(English: Would you like to know the information of its memory)

User: *Có*

(English: Yes)

System: *Bộ nhớ trong là 115 MB.*

(English: Memory is 115 MB)

…20s

System: *Bạn có muốn biết thêm thông tin nào không?*

(English: Would you like to know any other information more?)

User: *Không*

(English: No)

5. CONCLUSIONS

This paper presents the model of Vietnamese automatic dialog mechanism that we propose for V-DLG~TABL system. This dialog mechanism bases on supplemental mechanisms: choosing the suggested questions, managing the conversations and suggesting information, revolving the anaphora in conversion, and Question – Answering engine. Among these supplemental mechanisms, the engine of answering Vietnamese questions, which is built based on [1] and [2], is an important element of our dialog mechanism.

We have implemented this dialog mechanism to build V-DLG~TABL system. We have tested the system according to the defined scenario. Although at the current time, the system only operates according to one simple dialog scenario, but this scenario is quite general to cover some simple communication of the user with the system. However, we see that need to developer the other scenarios for the system.

ACKNOWLEDGEMENTS

This research is funded by University of Information Technology, Vietnam National University – Ho Chi Minh City (VNU-HCM) under grant number C2011CTTT-06.

REFERENCES

[1] Patrick Blackburn, Johan Bos, *Representation and Inference for Natural Language: A First Course in Computational Semantics*, September 3, 1999.

[2] Nguyễn Thành, Phạm Minh Tiến, "Xây dựng cơ chế hỏi đáp tiếng Việt cho hệ thống tìm kiếm sản phẩm máy tính bảng", B.Sc. Thesis in Computer Science, University of Information Technology, Vietnam National University – Ho Chi Minh City, 2012.

[3] Vương Đức Hiền, "Xây dựng công cụ truy vấn tiếng Việt về các phần mềm máy tính", B.Sc. Thesis in Computer Science, University of Information Technology, Vietnam National University – Ho Chi Minh City, 2013.

[4] Phạm Thế Sơn, Hồ Quốc Thịnh, "Mô hình ngữ nghĩa cho câu trần thuật và câu hỏi tiếng Việt trong hệ thống vấn đáp kiến thức lịch sử Việt Nam", B.Sc. Thesis in Computer Science, University of Information Technology, Vietnam National University – Ho Chi Minh City, 2012.

[5] Lâm Thanh Cường, Huỳnh Ngọc Khuê, "Biểu diễn và xử lý ngữ nghĩa dựa trên FOL (First Order Logic) cho các dạng câu đơn tiếng Việt trong hệ thống hỏi đáp kiến thức xã hội", B.Sc. Thesis in Computer Science, University of Information Technology, Vietnam National University – Ho Chi Minh City, 2012.

[6] Son The Pham and Dang Tuan Nguyen, "Processing Vietnamese News Titles to Answer Relative Questions in VNEWSQA/ICT System", *International Journal on Natural Language Computing (IJNLC)*, Vol. 2, No. 6, December 2013, pp. 39-51. ISSN: 2278 - 1307 [Online]; 2319 - 4111 [Print].

[7] Dang Tuan Nguyen, An Hoai Vo, Phuc Tri Nguyen, "A Semantic Approach to Answer Vietnamese Questions in OpenCourseWare Retrieval System", *Proceedings of the 2011 International Conference on Software Technology and Engineering (ICSTE 2011)*, August 12-13, 2011, Kuala Lumpur, Malaysia, pp. 314-320.

[8] Dang Tuan Nguyen, An Hoai Vo, Phuc Tri Nguyen, "Answering a Series of Vietnamese Questions in Library Retrieval System", *Proceedings of the 2011 2nd International Conference on Future Information Technology (ICFIT 2011)*, September 16-18, 2011, Singapore, pp. 445-449. ISBN: 978-981-08-9916-5. IPCSIT, Vol. 13, 2011.

[9] Võ Hoài An, Nguyễn Trí Phúc, "Mô hình xử lý cú pháp và ngữ nghĩa cho các chuỗi câu hỏi tiếng Việt trong hệ thống tìm kiếm thư viện", B.Sc. Thesis in Computer Science, University of Information Technology, Vietnam National University – Ho Chi Minh City, 2011.