

DISSERTATION INFORMATION

Title:	Summarizing Vietnamese paragraphs based on generating approach
Major:	Computer Science
Major Code:	62.48.01.01
PhD Student:	Tran Trung
Advisors:	Assoc. Prof. Dr Nguyen Tuan Dang, Assoc. Prof. Dr Pham Huu Duc
University:	University of Information Technology, VNU-HCM

1. ABSTRACT:

The objective of the dissertation is to propose a solution to generate summarizing paragraphs to summarize the information content of a given Vietnamese paragraph. The proposed summarizing model in the dissertation consists of two components:

- Component “Analysing and Representing the semantics of Vietnamese text”. This component receives the input as Vietnamese paragraph. The output of this component is an abstract semantic representation structure of the source text.
- Component “Generating Vietnamese text”. The input of this component is an abstract semantic representation structure. The output of this component, also the final result of the model, is a complete summarizing paragraph.

The dissertation focuses on studying the abstract model for short Vietnamese paragraphs, in which the anaphoric factors may appear.

2. THE MAIN CONTRIBUTIONS OF THE DISSERTATION:

○ Scientific contributions:

- Proposing a solution to generate summarizing paragraphs.

- Proposing methods of creating abstract semantic representation structure Discourse Representation Structure – DRS (according to Discourse Representation Theory).
 - The method of using a surface phrase structure consists of two stages. In the first stage, the dissertation proposes a method to create a surface-level representation structure called a surface phrase structure. Also at this stage, the dissertation proposes a rule-based method to transform into paragraph composing simple structured Vietnamese sentences. In the second phase, the dissertation proposes a method to determine the relationship between each anaphoric factor and the corresponding antecedence and at the same time creates DRS of the result of the first phase.
 - The method using a tagged-phrase graph structure consists of two stages. In the first stage, the dissertation proposes a method to create a surface-level representation structure called a tagged-phrase graph structure. In this period, the dissertation also proposed a method to determine the relationship between each anaphoric factor and the corresponding antecedence. In the second stage, the dissertation proposes a method to transform the intermediate representation model into DRS.
- Proposing methods of determining the antecedences for different types of anaphoric factor in the source Vietnamese paragraph.
- Proposing methods of generating Vietnamese summarizing sentences paragraphs based on an abstract semantic representation structure.
- **Practical contributions:**
 - The dissertation's research results can be applied to the construction of high quality and effective text summarizing systems.
 - The dissertation's research results can be applied in other research directions of the Natural Language Processing field such as: semantic parsing, text mining, natural language generation, ...

- The dissertation's research results can be applied to the construction of human-machine interaction systems such as robot systems, virtual assistants, ...

The research and experimental results of the dissertation are published in 16 scientific articles: 02 international journal articles, 02 domestic journal articles, 08 articles at specialized international conferences with critical independent reviews that the conference proceedings were published by the prestigious publisher Springer, IEEE; 04 articles published in the proceedings of national specialized conferences.

Advisor 1

Advisor 2

Ph.D student

Assoc. Prof. Dr Nguyen Tuan Dang

Assoc. Prof. Dr Pham Huu Duc

Tran Trung