THESIS INFORMATION

Title:	**Object Coreference Resolution for Sentiment Analysis**
Major:	Computer Science
Major Code:	62.48.01.01
PhD. Student:	Le Thi Thuy
Advisors:	Prof. Phan Thi Tuoi
University:	Ho Chi Minh City University of Technology, VNU-HCMC

1. ABSTRACT

Coreference resolution and aspect-based sentiment analysis are common problems in natural language processing (NLP) and are going under research by the NLP community with different approaches. In order for the sentiment analysis problem to have complete and detailed results that bring practical benefits, the thesis proposes to solve the object coreference resolution problem for sentiment analysis. This is the aim of the thesis. Results of the problem are triplets of object-aspect-sentiment, which belong to special products or services on a domain. The limit of the thesis is applied to English sentiment texts.

To build object coreference resolution in sentiment analysis, the thesis proposes a popular knowledge-based approach, sentiment ontology, combining natural language processing algorithms, machine learning, and deep learning on contextual sentiment corpus. Based on the characteristics of the sentiment text and the goal of the problem, the thesis needs to solve the following problems. Firstly, the thesis proposes the problem of object coreference resolution for sentiment analysis with one object. The second is sentiment ontology enrichment that supports object co-referencing resolution. The third is to define an implicit aspect that supports object co-referencing resolution. The fourth is object coreference resolution for multi-object sentiment analysis. The experiments of the thesis are carried out on the corpus of sentiment texts commenting on products and services on the Amazon website and the YouNetMedia Company. The experimental results reflect the feasibility of the proposed algorithms and models in the thesis.

The thesis has 6 published articles, among which 2 articles are in the list of journals (1 article in a domestic journal, 1 article in SCIE), and 4 articles in the proceedings of the international scientific conferences.

2. MAIN CONTRIBUTIONS

- ***Building the sentiment ontology (SO) and the coreference graph (CRG) to support the problem of object coreference resolution in sentiment analysis***. SO identifies the triples of object, aspect, and sentiment; can determine implicit aspects for sentiments; solves named entity coreference resolution; identifies the relationships between components of the triples of object, aspect and sentiment. The second tool, CRG, is to generate the triples which is the output of the thesis.
- ***Proposing the model of object coreference resolution on texts with one object***. From the first contribution, the thesis builds an object coreference resolution model for sentiment analysis for single-object text.
- ***Proposing the sentiment ontology enrichment model***. The thesis's sentiment ontology enrichment focuses on adding instances to classes, such as words, and phrases indicating objects, aspects and sentiments and the relationships between them.
- ***Proposing the implicit aspect identification model in sentiment texts***. To identify the implicit aspect in sentiment texts, the thesis proposes a deep learning method based on a multi-layer artificial neural network and the context of sentiment words referring to the aspect in a corpus.
- ***Proposing the object coreference resolution model in sentiment analysis with more than one object***. For texts with many objects, the thesis uses texts with one object as the basis for determining the relationship between objects, aspects and sentiments. The approach of the model is applied according to the current new method of machine learning and deep learning based on the context of text and sentiment ontology.

3. ISSUES FOR FUTURE RESEARCH

Although the results are quite good and have been evaluated experimentally, the thesis still has some small issues that need attention and development in the future.

- Building and enriching sentiment ontology by the semi-automatic method has some limitations, so improvement and use of the automatic method are necessary.
- For the problem of coreference resolution for sentiment analysis with many objects is affected by the pre-training step. If there is a new aspect that does not exist in the original lexicon, then the pre-training step must be performed again, or not change the lexicon, and thus the results obtained will be affected.
- Developing the model of the thesis on Vietnamese sentiment text will have high scientific and practical significance in the field of natural language processing in Vietnam.

ADVISORS	PHD STUDENT
Prof. Phan Thi Tuoi	Le Thi Thuy