INFORMATION OF DOCTORAL DISSERTATION

PhD student name: **NGUYỄN CHÁNH THÀNH**

Dissertation title: **QUERY EXPANSION MODEL CONSTRUCTION IN TEXT INFORMATION RETRIEVAL**

Major: **COMPUTER SCIENCE**

Major code: **62.48.01.01**

Scientific Advisor: **Associate Professor, Doctor. PHAN THỊ TƯƠI**

School: **HCMC University of Technology, Vietnam**

DISSERTATION SUMMARY:

In information retrieval, because of different reasons, users usually use simple queries in presenting their expected questions. By this, retrieved results neither have a good quality nor meet expectation of users. Therefore, query expansion is a necessary solution providing more context information to an information retrieval system then improving its retrieved result quality. The context information here can retrieve from relevant feedback, co-occurrence or knowledge base such as ontology.

The dissertation proposes an approach (methodology) of ontology-based query expansion. To achieve above goal, the dissertation solves two key problems as follows:

*** Problem 1–Constructing Ontology of Object-Member-Property (OOMP)**: researching and develop OOMP structure and training methodology to enrich OOMP based on TREC English documents and WordNet semantic data, then to serve problem 2 as below.

*** Problem 2–Constructing methodology of query completion and expansion**: based on OOMP of above problem's solving, researching and proposing a new approach for query expansion solution to provide new expanded queries in *complete query* format (presenting in kind of noun phrase in linguistics), then proposing new algorithms to solve this problem.

The experiment of the dissertation was done on TREC English documents. Its results highlighted the feasibility of proposed methodologies of the dissertation also shown promises of theoretic proposals of the dissertation in future.

NEW CONTRIBUTIONS / RESULTS OF THE DISSERTATION:

The dissertation proposed a theoretic model and methodology of ontology-based query expansion in information retrieval field based on constructing *complete noun phrase* (a kind of noun phrase which satisfies natural language's syntax also provides semantic information to help improving quality of information searching), then developing a theoretic base of query expansion model also algorithms for verifying, completing and expanding user's query also constructing a semantic index system. The results satisfied the goal of Problem 1 and 2 as above, also highlighted scientific and realistic significations based on new key contributions and results as follows:

*** Result 1: a proposed methodology which is used to be the theoretic base for all problems of the dissertation and theirs solving solutions.** It includes proposed concepts of semantic relations (R^m, R^p, $R^m_{\succ}$ and $R^p_{\succ}$), complete query, analysis tree and semantic graph

also theoretic models of query expansion system, MQE_{-IR}^{+OB}, MQE_{-IR}^{+OB+P}, MQE_{+IR}^{+OB} and MQE_{+IR}^{+OB+P}, which can be applied in the dissertation and other sub-fields of natural language processing. Also, object clustering will be done more effectively based constructing equivalence classes based on above relations.

*** Result 2: a proposed model of OOMP ontology and its training methodologies** such as CB-KBT, WB-KBT and A-KBT algorithm. OOMP can serve to solve the query expansion problem and other problems that focus on exploiting relational concepts; also it can be developed more conveniently for other natural languages (Vietnamese, French …).

*** Result 3: a proposed query processing model** (for query in kind of noun phrase) **including CNPV, NPC, SNPE and NPMR** algorithms thus we can apply to solve Problem 2 of the dissertation also other problems in natural language processing fields (information retrieval - extraction, text summarization).

*** Result 4: a proposed semantic index creation methodology.** The methodology found a new research orientation of semantic index in information retrieval and semantic web fields. It can support to develop smart information retrieval system which can provide a better result quality for user's searched result. It also laid the foundation for relevant researches in semantic index of information retrieval.

Above results were also presented in published papers of the author.

APPLICATIONS OF DISSERTATION'S RESULT IN PRACTICE:

The achieved results of the dissertation lay the foundation of further research of the author. They also support to develop applications in research projects of the author and other external information retrieval system in practice.

FURTHER RESEARCH OF THE DISSERTATION:

There are several matters for further research of the author as follows.

*** Problem 1:** Optimizing algorithms' implementation in dissertation's experiment and utilizing advantages of object-oriented database management systems to save algorithms run-time cost also improving their performance.

***Problem 2:** Updating some pre-processing steps of existing solutions in the dissertation to be configurable that can apply to Vietnamese or other natural languages.

*** Problem 3:** Developing continuously researches on semantic index and its relevant applications.

*** Problem 4:** Optimizing run-time cost of HS algorithm in the dissertation.

Scientific Advisor	PhD student
Prof. Dr. PHAN THỊ TƯƠI	NGUYỄN CHÁNH THÀNH