Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 3, March 2014, pg.623 – 631

RESEARCH ARTICLE

A HYBRID APPROACH OF QUERY EXPANSION FOR VIETNAMESE QUERY

Tuan, Tran Anh
Faculty of Informatics & Foreign Language
DakLak College of Pedagogy, Vietnam
tuanta@dlc.edu.vn

Abstract - The semantics of a user's query plays an important role in supporting an Information Retrieval (IR), which returns results closer to the user's query. However, most of user's queries sometimes do not fully reflect the semantics. Therefore, it is necessary to add semantics to the user's query. The paper presents how to add semantics into a query in Vietnamese language using a hybrid method, which combines ontology and local analysis technique, in order to expand the user's queries. In the hybrid method, ontology – based query expansion technique analyzes semantics relationships in order to determine similar noun phrases, and local analysis technique is to get the most relevant documents which are to identify the context of user's query.

Keywords - Vietnamese language query expansion; hybrid model of query expansion; query expansion; local analysis; ontology

I. INTRODUCTION

Nowadays, Information Retrieval (IR) and Search Engine (SE) become one of the most important tools for our life. Most of users use IR or SE when they search new information on digital library or the Internet. However, the returned results from these systems always include some documents which are not relevant with the user's query. One of the reasons leading to the returned results which are not highly precise is a short query. To achieve more accurate results, a query will be added more keywords which related to its query semantics. This is a problem of query expansion.

There are two techniques of query expansion [20]: search result – based technique, and knowledge structure – based technique. The search result – based technique uses terms which are chosen from the retrieved documents in the relevance feedback process. The knowledge structure – based technique can either depend on corpus or be independent of it.

The collection – dependent knowledge structure expands user's query by adding keywords from a knowledge model as WordNet, ontology. The semantics of user's query is determined in this technique. In contrast to the collection – dependent knowledge structure, collection – independent knowledge structure is based on statistical analysis of the feedback from the corpus such as global analysis, local analysis, etc. User's query is expanded from returned documents which have relationship with user's query. One of the first techniques is global analysis technique which improves returned results. This technique analyzes the entire document corpus to determine word relationships. Otherwise, local analysis technique which uses top-ranked retrieved documents for query expansion is to identify the context of user's query and to achieve high efficiency in retrieving the specific domain as medicine, computer science, etc.

Accordingly, based on the advantages' techniques above, a query expansion model for Vietnamese query is proposed. This model uses a hybrid method which combines ontology and local analysis technique in order to expand the user's queries.

In the hybrid method, ontology – based query expansion technique analyzes semantics relationships in order to determine similar noun phrases, and local analysis technique is to get the most relevant documents which are to identify the context of user's query.

In the following sections, an overview of research works is described, and then present a hybrid model of query expansion for Vietnamese language query. Section 4 presents ontology and Vietnamese thesaurus development. The remaining section 5 and 6 will show how to expand a query based on ontology and local analysis technique and evaluate experiment. And the last is conclusion section, future works.

II. RELATED WORKS

There are lots of query expansion techniques investigated by many researchers.

A. *Query expansion technique based on ontology*

In this technique, the semantics of user's query is determined. There are many researchers in this technique. In the work of R.Navigli and P.Velardi [10], the query was processed and the ambiguity was reduced before expanding a query. They proposed an approach of word sense disambiguation via creating and intersecting semantic networks, then assigning scores to the configurations to find the best selections. They created a new model of semantic Network based on some relations, which were extracted from WordNet such as *hyperonymy*, *hyponmy*, *meronymy*, *holonymy*, *attribute*, *similarity* and some newly defined relations such as *gloss*, *topic*, and *domain*. A.Agissilaos [1] did the research on ontology for query expansion, which was built according to the approach using the concepts and semantic relations available in WordNet. He also developed an approach of handling word sense disambiguation to the keywords in the query. However, he used the page rank approach of Banerjee and Pedersen, which was different from the one proposed by R.Navigli and P.Velardi. Chen Gang in the biomedicine field [3] used query expansion based on WordNet, MeSH and UMLS SPECIALIST. For WordNet, he used it as a general ontology to expand the common keywords. As for the MeSH and UMLS SPECIALIST, he used them as the specific ontologies in the field of biomedicine to expand the terminology. And the synonym and hyponym relations of WordNet and UMLS were used to expand queries.

Furthermore, Thanh.N.C and Tuoi.P.T [13][14][15] had created a new approach called *object-oriented* to expand a query based on OMP ontology. This ontology which was based on the relational characteristics and the concept of ontology was the noun phrase extracted from the TREC corpus. They also proposed algorithms to support query expansion using OMP ontology: *Similar Noun Phrase Expansion (SNPE), Verification Complete Noun Phrase (CNPV), etc*.

B. *Query expansion technique based on Local analysis*

The top-ranked retrieved documents are used for query expansion in local analysis in order to identify the context of user's query and to achieve high efficiency in retrieving the specific domain as medicine, computer science, etc.

In the research of Attar and Fraenkel [2], the top ranked documents for a query were used to build an automatic thesaurus which is to cluster and treat as quasi-synonyms. An automatic query expansion via adhoc feedback was improved by Mandar Mitra, Amit Singhal and Chris Buckley [9]. In their approach, term co-occurrence was used to estimate word correlation and to identify independent concepts in a user's query.

In addition, a local context analysis method, which combines both local analysis and global analysis, was proposed by Xu and Croft [19]. Expansion terms were not based on frequencies in the top-ranked documents, but based on co-occurrences with the query terms within the top-ranked documents.

C. *A hybrid technique of query expansion*

In the work of Thanh.N.C and Tuoi.P.T [16], a hybrid ontology-based solution was proposed to expand user's queries by authors. They proposed for mechanisms not only to look for relative result in the OMP ontology to complete and expand user's entered query/noun phrase, but also to expand the search progress by linking the OMP ontology to indexes of information retrieval system. Ounas Asfari et al [11] presented a context-based hybrid method for query expansion that automatically generates context-related terms. The model considered the context as the actual state of the task that the user was undertaking when the information retrieval process takes place. They also introduced a new concept of SRQ (State Reformulated Queries), which was used to reformulate queries according to the user task context and the ontological user profile.

Furthermore, the HQE method, which combines ontology - based collaborative filtering with neural networks to improve query expansion, was proposed by Lixin Han and Guihai Chen [8]. In the HQE method, collaborative filtering was used to analyze semantics relationships that are acquired from the constructed ontologies in order to find the similar users, and the radial basis function (RBF) networks were used to acquire the most relevant web documents and their corresponding terms from these similar users' queries.

III. A PROPOSED HYBRID MODEL OF QUERY EXPANSION

A hybrid model of query expansion is proposed for Vietnamese language query as illustration in Figure 1 below.

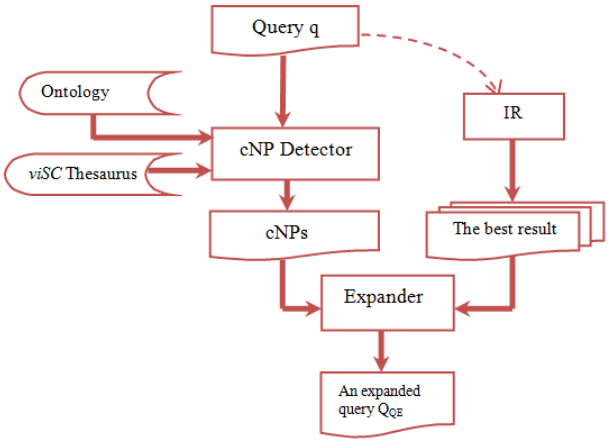

Fig 1. A Hybrid Model of Query expansion for Vietnamese query

In the model, The cNP Detector will select candidate noun phrases (cNPs) which have semantics relations with the user's query q from the ontology and viSC dictionary. Before selecting cNPs, The cNP Detector also checks the user's query if it is complete or not *(A query is complete if it consists description of the object, key member, key property and they are in R^m and R^p relations [13][14])*. If the query is not complete, the detector will add components from the ontology in order to complete the query in a form of noun phrase.

The cNP Detector is built by the *improved Similar Noun Phrase Expansion (iSNPE)* algorithm based on OOMP ontology. The result of this detector can have a lot of noun phrases corresponding with query.

Otherwise, the user's query q is used to retrieve documents which have context with query from corpus and are ranked by IR.

The Expander calculates the score of the cNPs based on the top ranked documents. After being calculated the score, candidate noun phrases have high score will be selected and added into the user's query.

The expanded query has semantics and context with the user's query.

IV. ONTOLOGY AND VIETNAMESE DICTIONARY DEVELOPMENT

A. *Ontology Development*

Ontology which is called OOMP is used to expand user's query. The OOMP has hierarchy of semantics concepts based on relationships (R^m, R^p, R^m_φ, R^p_φ) [13][14][15]. Figure 2 presents an OOMP ontolgy example.

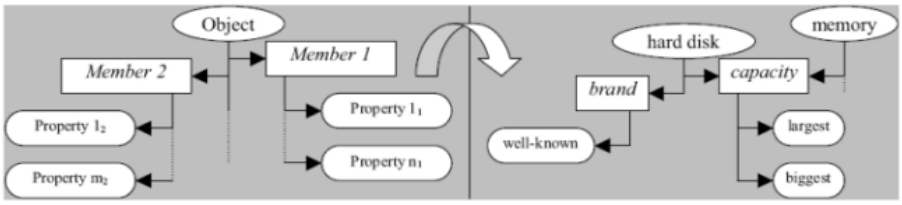

Fig 2. Example of OOMP ontology [13][14][15]

The OOMP ontology structure was proposed by Thanh.N.C and Tuoi.P.T and was constructed in relational data model [13][14][15].

In detail of figure 3, Candidate noun phrases which are extracted from corpus in Vietnamese documents are determined the components O, M, and P basing on the pattern of OMP components [18]. After being determined, semantics concepts and relations (Rm, Rp, R^m_φ, and R^p_φ) are built from sets of components {O}, {M}, {P}. The best concepts and relations for ontology are selected by expert.

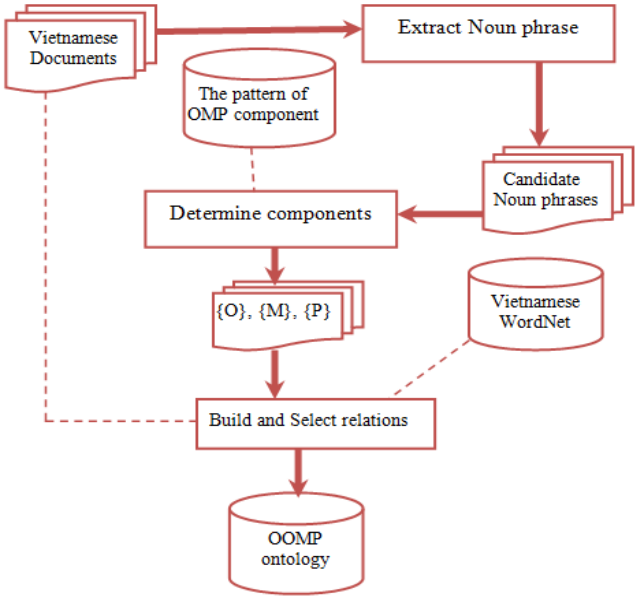

Fig 3. Training the OOMP ontology

B. Vietnamese Thesaurus Dictionary Development

Vietnamese thesaurus dictionary called *viSC* is built by expert with the help of the Transact-MySQL and stored in tables in MySQL. This dictionary has words extracted from the following sources: WordNet 3.0 [29]; Vietnamese dictionary for natural language processing complied by Luong.V.X, Bao.H.T, and Huyen.N.T.M [28]. The steps to build this viSC are as follows:

- *Step 1*: *Extract words (Lemma) in the computer field from WordNet, and then remove overlapping words.*
- *Step 2: Determine synonyms with the words in step 1 from WordNet, and then select and remove duplicate words that are not related to selected areas.*
- *Step 3: Use the online Bambo dictionary [21] and LacViet dictionary to translate words in step 2 into Vietnamese.*
- *Step 4: Select and add the synonyms in the field of computer from the Vietnamese dictionary [28] to viSC dictionary.*
- *Step 5: Add abbreviations which correspond to the words in the dictionary viSC if any.*

V. A HYBRID MECHANISM OF QUERY EXPANSION FOR VIETNAMESE LANGUAGE

This mechanism determines and selects keywords, which have semantics and context with the user's query, do complete and expand Vietnamese language query. These steps of a hybrid mechanism are as follows:

- *Step 1: Determine keywords have semantics with the user's query from ontology*

Keywords in form of noun phrase, which will be called *vnNP* (Vietnamese Noun Phrase), are extracted from ontology of OOMP. To get better results with the Vietnamese language, the Similar Noun Phrase Expansion (SNPE) algorithm [14] is improved by using viSC dictionary, then to produce *iSNPE* algorithm (improved Similar Noun Phrase Expansion).

The iSNPE algorithm steps are as belows:

- *Step 1.1*: The user's query is verified if it is complete or not. If the query is not complete, the query will be added components from the ontology in order to complete the query in a form of noun phrase. This step is done by CNPV (Complete Noun Phrase Verification), NPC (Noun Phrase Completion) algorithm [13][14].
- *Step 1.2:* The completed query in step 1 will be analyzed into components *(Object – O, Member – M, and Property – P).* After being analyzed, synonyms corresponded the components *(O, M, P)* will be selected from Vietnamese thesaurus dictionary to add into sets of synonyms *(O_c, M_c, P_c).* The VnSS *(Vietnamese Synonym Selection)* algorithm [18] does select synonym from thesaurus dictionary.
- *Step 1.3*: The similar noun phrases are generated from the sets of synonyms in step 2 by CNPG (*Complete Noun Phrase Generation*) algorithm [13][14].

- *Step 2: Select the best documents from corpus*

In this step, Documents retrieved and ranked by Lucene have context with the user's query. The top N documents, which are selected from the retrieved and ranked documents, are the best results. The N argument plays an important role in calculating the score of the keywords.

- *Step 3: Select the highest scored keywords*
  - Calculate the score of the keywords in step 1 based on the best documents. The score is based on conditional statistics as equation belows:

$$score(vnNP) = score(t_1, t_2, \dots, t_h) = \prod_{j=2..h} p(t_j|t_{j-1}) \quad (5.1)$$

Where: conditional probability of t1 with t2

- t_j is a word in the vnNP.
- $p(t_j|t_{j-1})$ is conditional statistics between two words

  - Choose the M keywords have the highest score. The M argument also plays an important role in supporting IR, which returns results closer to the user's query.
  - Add the M keywords into the user's query.

For example:

- With initial query "*bộ nhớ máy tính*" (computer memory).
- Step 1: Determine keywords have semantics with the user's query from ontology
  - query is tagged and resulted *"bộ nhớ /Nc máy tính/Nc"*. The CNPV algorithm is called and the retrieved result is {TH_5}.
  - With {TH_5}, NPC is called to complete noun phrase and define o_i = "*máy tính*", m_j = "*bộ nhớ*" and p_k = "*mới*".
  - Basing on the viSC, we obtain O_c=*{"máy tính", "máy vi tính"}*, M_c={*"bộ nhớ", "bộ lưu trữ", "memory"*}, P_c={*"mới", "thấp"*}
  - Basing on ontology, we retrieve the results: *{"bộ lưu trữ của máy tính mới", "bộ lưu trữ máy tính mới", "bộ lưu trữ máy vi tính mới", "bộ nhớ của máy tính mới", "bộ nhớ của máy tính thấp", "bộ nhớ máy tính mới", "bộ nhớ máy tính thấp", "bộ nhớ máy vi tính mới", "memory của máy tính mới", "memory của máy tính thấp", "memory của máy vi tính mới", "memory máy tính mới", "memory máy tính thấp", "memory máy vi tính mới", "memory máy vi tính thấp"}*
- *Step 2: Select the best documents from corpus*
  - Documents retrieved and ranked by Lucene have context with the user's query. Choose N = 10 and N = 20 in order to evaluate.
- Step 3: *Select the highest scored keywords*
  - With N = 10 and N = 20, the score of the keywords in step 1 is presented in the following table 1:

TABLE I

THE SCORE OF THE KEYWORDS

No	Candidate noun phrase	Score	
		N = 10	N = 20
1	bộ lưu trữ của máy tính mới (new storage of computer)	0.047	0.056
2	bộ lưu trữ máy tính mới (new computer storage)	0.104	0.113
3	bộ lưu trữ máy vi tính mới (new computer storage)	0.043	0.041
4	bộ nhớ của máy tính mới (new memory of computer)	0.154	0.146
5	bộ nhớ của máy tính thấp (small memory of computer)	0.064	0.047
6	bộ nhớ máy tính mới (new computer memory)	0.261	0.290
7	bộ nhớ máy tính thấp (small computer memory)	0.109	0.093
8	bộ nhớ máy vi tính mới (new computer memory)	0.087	0.056
9	memory của máy tính mới (new memory of computer)	0.250	0.301
10	memory của máy tính thấp (small memory of computer)	0.104	0.097
11	memory của máy vi tính mới (new memory of computer)	0.104	0.110
12	memory máy tính mới (new computer memory)	0.778	0.729
13	memory máy tính thấp (small computer memory)	0.324	0.234
14	memory máy vi tính mới (new computer memory)	0.324	0.267
15	memory máy vi tính thấp (small computer memory)	0.135	0.086

  - Choose the 10 keywords have the highest score. The result is presented in the following table 2:

TABLE II

THE 10 KEYWORDS HAVE THE HIGHEST SCORE

N = 10; M=10		N = 20; M=10	
Candidate noun phrase	**Score**	**Candidate noun phrase**	**Score**
memory máy tính mới	0.778	memory máy tính mới	0.729
memory máy tính thấp	0.324	memory của máy tính mới	0.301
memory máy vi tính mới	0.324	bộ nhớ máy tính mới	0.290
bộ nhớ máy tính mới	0.261	memory máy vi tính mới	0.267
memory của máy tính mới	0.250	memory máy tính thấp	0.234
bộ nhớ của máy tính mới	0.154	bộ nhớ của máy tính mới	0.146
memory máy vi tính thấp	0.135	bộ lưu trữ máy tính mới	0.113
bộ nhớ máy tính thấp	0.109	memory của máy vi tính mới	0.110
bộ lưu trữ máy tính mới	0.104	memory của máy tính thấp	0.097
memory của máy tính thấp	0.104	bộ nhớ máy tính thấp	0.093

VI. EXPERIMENT AND EVALUATION

A. *Data source*

For Vietnamese language, it currently does not have a standard corpus like TREC. So that, a corpus in computer field must be built. A corpus is extracted from websites: *Network administration (D1) 1,571 files – 111MB[24], PC World Magazine (D2) 1,070 files – 80.8MB[27], Vietnamese Wikipedia (D3) 92 files – 5.8MB[26], Science and Technology (D4) 156 files – 6.63MB[25], Information technology (D5) 660 files – 46MB[22], IT News (D6) 609 files – 34.2MB[23]*. It was extracted from web (.html, .htm) to text (.txt) and normalized to Unicode. A corpus is divided into two parts: one is used to training ontology (D1, D2) and the other is used to indexing (D3, D4, D5).

B. *Experiment of a Hybrid Mechanism*

To evaluate experiment, Lucene is used, however, it does not support for Vietnamese language. So that, a VietAnalyzer package is built to support pre-processing Vietnamese text for an information retrieval in indexing and searching. The fourteen queries in English or abbreviation are used to evaluate a reliability of VietAnalyzer package. The twenty five queries are also selected from http://www.pcworld.com.vn/hoi-dap/phan-cung/ to experiment. And the result is introduced as below Table 3.

Two arguments playing an important role in a hybrid model are N and M. The N argument is number of the best ranked documents to calculate the score of the keywords and the M argument is number of the highest scored noun phrases to add into the user's query. According to the research of Claudio and et al [4], the N argument has the best result is 4 to 10; with N=20, returned results are not good and decreasing; and the M argument is 10. So that, the experiment of query expansion has 3 cases: the first is a hybrid model of query expansion with N=10, M=10; the second is a hybrid model with N=20, M=10; and the last is ontology – based model. The experimental result is presented in Table 3.

TABLE III

RESULT OF QUERY EXPANSION

(1)	(2)	(3)	(4)	(5)	(6)
Q1	card đồ họa mới	a new graphics card	*0.14*	*0.14*	*0.14*
Q2	bộ nhớ máy tính	computer memory	*1.19*	*1.19*	*0.20*
Q3	màn hình tinh thể lỏng	liquid crystal displays	*0.47*	*0.46*	*0.40*
Q4	card mở rộng	expansion card	*0.57*	*0.57*	*0.57*
Q5	màn hình LCD	LCD screen	*16.23*	*16.23*	*15.97*
Q6	máy tính Dell	Dell Computer	*0.20*	*0.20*	*0.20*
Q7	máy chủ	server	*47.80*	*47.80*	*47.18*
Q8	máy tính	computer	*99.52*	*99.52*	*97.28*
Q9	bộ xử lý	processor	*12.12*	*12.12*	*12.07*
Q10	bộ nhớ	memory	*62.70*	*62.70*	*36.21*
Q11	màn hình	monitor	*32.34*	*32.34*	*32.34*
Q12	CPU AMD	CPU AMD	*0.62*	*0.62*	*0.60*
Q13	HDD box	HDD box	*0.32*	*0.32*	*0.20*

Q14	mainboard Gigabyte	Gigabyte mainboard	*0.07*	*0.07*	*0.07*
Q15	AMD	AMD	*97.89*	*97.89*	*26.07*
Q16	Asus	Asus	*95.33*	*95.33*	*27.15*
Q17	Samsung	Samsung	*92.40*	*92.40*	*92.40*
Q18	chipset	chipset	*96.24*	*96.24*	*60.81*
Q19	Intel	Intel	*64.69*	*64.69*	*63.86*
Q20	LAN	LAN	*85.83*	*85.83*	*85.83*
Q21	mainboard	mainboard	*8.79*	*8.79*	*4.08*
Q22	RAM	RAM	*11.34*	*11.34*	*11.15*
Q23	SATA	SATA	*18.04*	*18.04*	*18.04*
Q24	Seagate	Seagate	*95.24*	*95.24*	*7.14*
Q25	CRT	CRT	*0.61*	*0.61*	*0.60*
		Average	**37.63**	**37.63**	**25.62**

Here:

(1): No., (2): Vietnamese query, (3): Vietnamese query in English

(4): Precision P (%) hybrid of query expansion with N = 20, M = 10

(5): Precision P (%) hybrid of query expansion with N = 10, M = 10

(6): Precision P (%) ontology – based query expansion

As the result is presented in table 3 and figure 4 and 5, the hybrid model of query expansion has precision higher than or equal to ontology – based query expansion. The first case, the hybrid model's precision is greater than the ontology model's: An expanded query is to increase a number of retrieved documents and a number of relevant retrieved documents. The second case, the hybrid model's precision is equal to the ontology model's: An expanded query and original query are the same. The reason explains why experiment has result above: number of relevant documents is not still large and limit. This result shows that a corpus also plays an important role in experimental evaluation. It is one of difficulties of the paper because a standard corpus for Vietnamese language does not introduce at present.

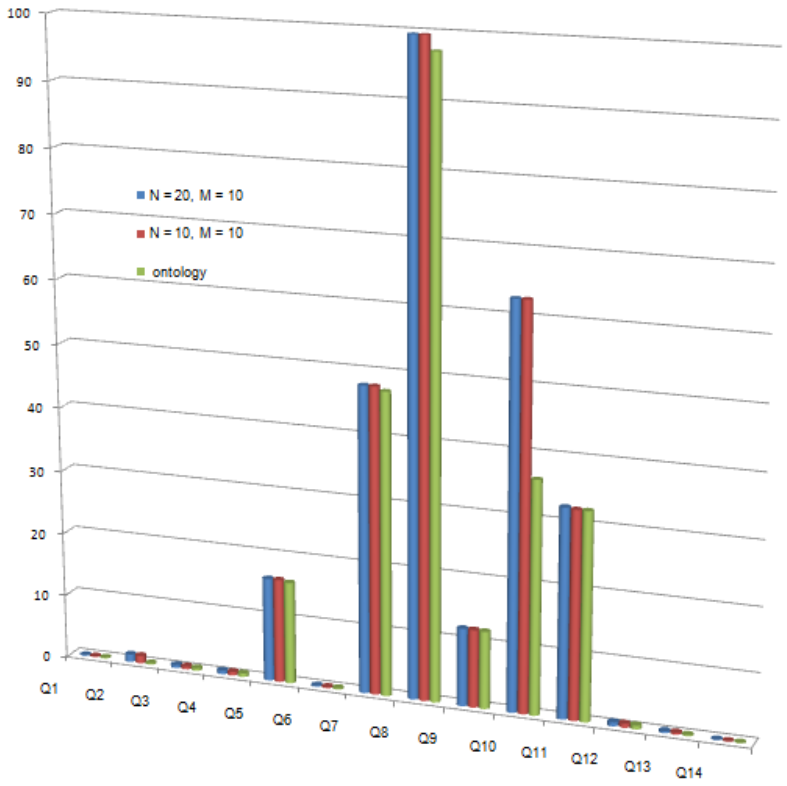

Fig 4. Statistic of Q1 to Q14 Precision

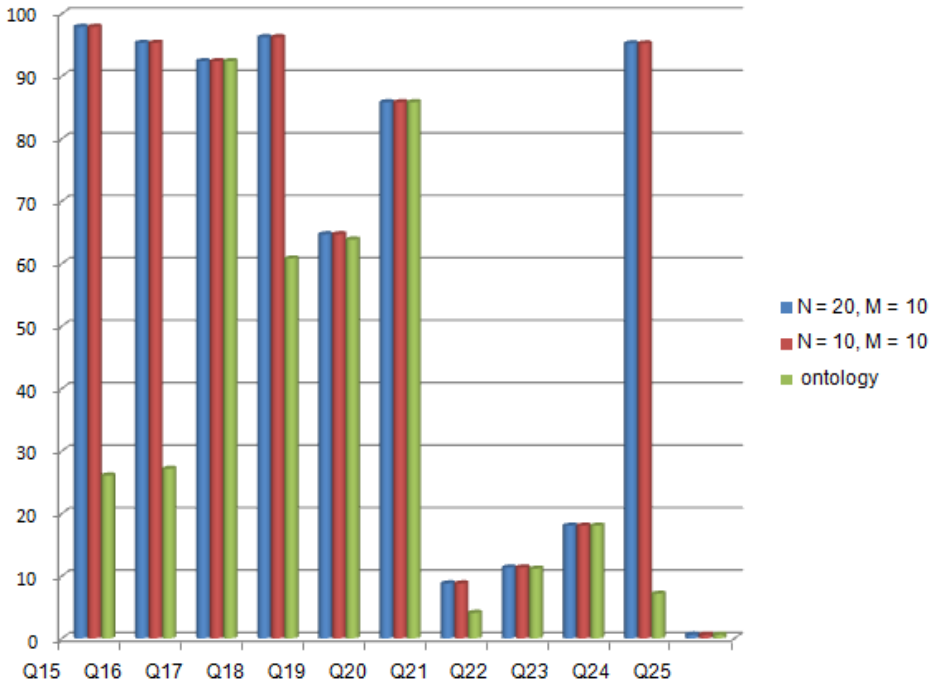

Fig 5. Statistic of Q15 to Q25 Precision

VII. CONCLUSION AND FUTURE WORK

This paper proposes how to expand Vietnamese language query based on a hybrid approach. To do a hybrid approach, a hybrid mechanism of query expansion is proposed. This mechanism has two phases: the first phase determines noun phrases have semantics with the user's query from OOMP ontology; the last phase selects the highest scored noun phrases based on the best documents from corpus. In the experimental result, an average precision of the hybrid model (37.63%) is greater than an average precision of the ontology model (25.62%). This result shows that the hybrid model is better than the ontology model is.

In future, paper does work as follows:

- Improve a hybrid mechanism of query expansion in order to increase precision;
- Build a semantics indexer; and
- Build a larger corpus to train ontology.

REFERENCES

[1] Andreou, A. (2005). Ontologies and Query Expansion, Master Thesis, School of Informatics University of Edinburgh.
[2] Attar, R., Fraenkel, A.S. (1997). Local Feedback in Full-Text Retrieval systems. Journal of the Association for Computing Machinery, 24(3), pp. 397 – 417.
[3] Chen, G. (2010). Ontology – based Query Expansion in biomedicine domain. Master Thesis, Roskilde University.
[4] Claudio Carpineto et al (2001). An Informatic – Theoretic Approach to Automatic Query Expansion. ACM Press, USA.Fensel, D. (2000). The Semantic Web and its Languages, IEEE Intelligent Systems, Vol 15, No. 6, November/December, pp. 67-73.
[5] Croft, W.B, Cook, R., and Wilder, D. (1995). Providing Government information on the Internet: Experiences with THOMAS. In Digital libraries Conference DL'95, pp. 19 – 24.
[6] Fu, G., Jones, C.B., Abdelmoty, A.I. (2005). Ontotology – based Spatial Query Expansion in Information Retrieval. International Conference on Ontologies, Databases and Applications of Semantics, Agia Napa, Cyprus.
[7] Imran, H., Sharan, A. (2009). Thesaurus and Query Expansion. International Journal of Computer Science and Information Technology (IJCSIT), Vol 1, No 2, Pp.89 – 97.
[8] Lixin Han, Guihai Chen. (2009). HQE: A hybrid method for query expansion. Expert Systems with Applications 36 (2009), pp.7985–7991.

[9] Mandar Mitra, Amit Singhal and Chris Buckley (1998). Improve Automation Query Expansion, ACM – 1998, University of Melbourne.
[10] Nagvili, R. and Verladi, P. (2003). An Analysis of Ontology based Query Expansion Strategies. 14th European Conference on Machine Learning (ECML 2003), September 22-26.
[11] Ounas Asfari, Bich-Lien Doan, Yolaine Bourda, Jean-Paul Sansonnet (2010). Context-based Hybrid Method for User Query Expansion. SEMAPRO 2010: The Fourth International Conference on Advances in Semantic Processing. ISBN: 978-1-61208-104-5. Pp. 69-74.
[12] Revuri, S., Upadhyay, R.S., and Kumar, P.S. (2006). Using Domain Ontologies for Efficient Information Retrieval. Proceedings of the 13th International Conference.
[13] Thanh, N.C. (2010). Query Expansion Model Construction in Text Information Retrieval. Doctoral Dissertation, HCMC University of Technology.
[14] Thanh, N.C, Tuoi, P.T. (2007). An Ontology - based approach of Query Expansion. Proceedings of the 9th iiWAS2007, Jakarta, Indonesia.
[15] Thanh, N.C and Tuoi, P.T. (2008). Information Retrieval: An Ontology Solution for Query Complete. Journal on ICT of Ministry Information and Communications, Pp 84 – 92, No 19.
[16] Thanh, N.C. and Tuoi, P.T. (2008). A hybrid solution of ontology-based query expansion. Journal of Web information systems, Vol 4 No 2, 2008, pp. 215 – 227.
[17] Thanh, N.C. and Tuoi, P.T. (2008). Information Retrieval: An Ontology Solution for Query Complete. Journal on ICT of Ministry Information and Communications, Pp 84 – 92, No 19.
[18] Tuan, T.A, Tuoi, P.T and Thanh, N.C. (2012). Query Expansion based on Ontology for Vietnamese Query. In Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services (iiWAS2012), ACM ISBN: 978-1-4503-1306-3, December 3-5, 2012, Bali, Indonesia, pp 332-335.
[19] Xu, J., & Croft, B. W. (2000). Improving the effectiveness of informational retrieval with local context analysis. ACM Transactions on Information Systems, 18(1), pp.79–112.
[20] Williams, Martha E., ed (1996). Query Expansion - Annual Review of Information Systems and Technology (ARIST), v31, pp 121-187, 1996.
[21] Bambo – dictionary online, http://tratu.soha.vn/dict.
[22] Information Technology, http://www.thongtincongnghe.com/maytinh.
[23] IT News, http://ictnews.vn/home/CNTT/Phan-cung/22/0/index.ict.
[24] Network Administrator, http://www.quantrimang.com.vn/phancung/index.aspx
[25] Science and Technology, http://www.khoahocphothong.com.vn/newspaper/lam-ban-voi-may-vi-tinh.
[26] Vi.Wikipedia – Computer Category, http://vi.wikipedia.org/wiki/Thể_loại:Máy_tính.
[27] Vietnamese Pcworld, http://www.pcworld.com.vn/articles/san-pham/.
[28] Vlsp – Vietnamese dictionary, http://vlsp.vietlp.org:8080.
[29] WordNet 3.0, http://wnsqlbuilder.sourceforge.net.

