

A New Query Expansion Method for Statements Query Retrieval Based on Event Structure

Bing Li

School of Foreign Languages
Qilu University of Technology
Jinan, China
e-mail:edward791110@163.com

Abstract—This paper borrows the idea from the linguistic theories of lexicalism and makes use of event structure to explore how to understand the statement query, so as to eliminate ambiguity and refine the query. It proposes a new query expansion method based on event structure which helps to improve query expansion by combining with the Knowledge based Query Expansion Method. First, use Event Structure based Query Expansion Agent (ESQEA) to identify queries conforming to verbs' argument structures, and then realize the query refinement according to the knowledge structure. The simulation result shows that for the same recall, the precision ratio is improved by the new approach.

Keywords-Lexicalism, Event Structure, Knowledge Structure, Domain Ontology

I. INTRODUCTION

The World Wide Web [1] has become the victim of its own success in many respects. The rapid growth in the volume of available information has made it increasingly difficult for users to quickly locate pertinent information. The traditional keyword-based information retrieval technique is to assist users to look for a specific piece of information buried within a few documents in a large collection. In spite of the giant contributions of this approach, this knowledge based method is far from perfect. First of all, it can only perceive the sense of the separate query item, not syntactic structure. Once a grammatical structure is input, the current system may fail to comprehend the exact meaning.

Example 1

- (a) The drive loaded the hay onto the truck.
- (b) The drive loaded the truck with the hay.

Traditional information retrieval technology based on knowledge structure cannot tell the difference of the two statements in Example 1, because it does not take the function words *the* and *onto* into query. Nevertheless, the well-known semantic distinction between the two statements is that (b) implies that the truck is completely loaded with hay while (a) does not.

This paper borrows the idea from lexicalism and proposes a new query expansion method based on event structure to make amendments to the knowledge based approach.

Two agents are used to realize the statement query expansion: the Event Structure based Query Expansion Agent and the Knowledge based Query Expansion Agent. The former is used to understand syntactic structure of the initial query; the latter is to extract the feature words from the initial query by performing Part of Speech (POS), Stop Words Removal, and Words Stemming, and to determine which special domain the initial query belongs to. The expanded query will be realized in the special domain on the basis of the knowledge structure.

The innovation of this paper is that for the first time it applies the lexical approach, i.e., event structure, to the information retrieval system, which can assist the comprehension of the syntactical meaning and expand the lexical query instead of the initial input query, which can reduce redundancy expansion for some meaningless query items.

The paper is structured as follows: Section 2 outlines the related work. Section 3 describes the entire system, the realization of the event structure based query expansion agent and knowledge based query expansion agent. Section 4 gives the simulation based on the proposed method and experimental results. The last section contains concluding remarks and future directions.

II. RELATED STUDY

A. Research on semantic meaning understanding

I proposed a constructional approach to improve precision based on Construction Grammar in another paper [2]. According to Construction Grammar, argument structure is determined by the integration of the verb and the construction [3]. Therefore, differences in meaning of the same verb in different constructions are attributed directly to the particular constructions. The benefit of this approach is quite obvious as shown in the following:

Example 2

- (a) He kicked the ball. (transitive)
- (b) He kicked the football into the gate. (caused motion)
- (c) He kicked Tom the football. (ditransitive)
- (d) He kicked his way out of the classroom. (way construction)
- (e) He kicked Tom black and blue. (resultative)

Base on constructional approach, *kick* means “to move one’s legs with very quick, small and forceful movements” across the five syntactic structures. Argument structure construction instead of verbs provides the direct link between surface form and general aspects of the interpretation, such as something acting on something else (a), something causing something else to move (b), someone intending to cause someone to receive something (c), someone moving somewhere despite obstacles (d), someone causing something to change state (e).

However, sometimes, the meaning is rooted in verbs instead of constructions. Consider the following example:

Example 3

(a)#He loaned some money to me, but I never got it.

(b)I threw Tom the ball, but he didn’t catch it.

According to the constructional approach, double object construction denotes caused possession event schema while dative construction does not. Nevertheless, obviously possession is entailed in (a) (dative construction) that is supposed to express caused motion according to Construction Grammar, while possession is not entailed in (b) (double object construction) that is supposed to infer possession according to Construction Grammar.

B. Research on query expansion techniques

Query expansion is the process of supplementing the original query with additional terms, which can be viewed as a method for improving retrieval performance [4].

The query expansion research has shifted the eye to delivering tailored, adapted and personalized information presentations due to the growing user information needs and queries in various contexts with different intentions. A challenging task of the query expansion is how to choose the proper terms to match users’ reading preference.

Query expansion techniques can be broadly categorized into two groups: those based on query results and those based on knowledge structure. The former is dependent on the search process and makes use of relevance feedback in the earlier iterative search as the source to locate the query expansion terms [5]. The latter does not rely on the search process and additional query terms are obtained by traversing a semantic network built on the basis of knowledge structure. Knowledge structures used by this group of techniques can either be a general-purpose ontology (or thesaurus) [6], an ontology built for a specific domain [7], or an ontology constructed from document collection based on term clustering [8]. However, neither of them can perceive the grammatical relationship of the searched items, except the independent verbal meaning. Therefore, if users input a logical expression, this approach cannot understand the correct meaning.

III. THE PROPOSED METHOD

A. Framework of the Event Structure based Query Expansion Method for Statements Query

Two agents are designed in the whole system, which can communicate with each other, while having their own responsibilities, shown in Fig. 1.

Firstly, use the lexical resource VerbNet to detail semantic classes of verb. Secondly, in the Knowledge based Query Expansion Agent, extract the feature words from the lexical query expanded by ESQEA with Feature Words Extraction Subagent, and then determine which domain ontology the lexical query belongs to. Finally, expand the query with WordNet based on the Knowledge-Based Query Expansion Subagent.

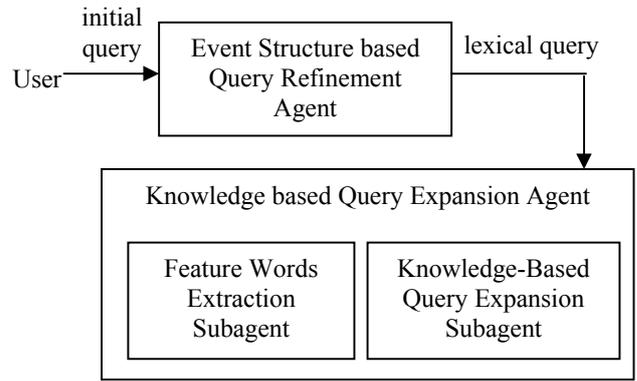


Figure 1. Framework of the Event Structure based Query Expansion Method

B. Event Structure based Query Expansion Agent

Lexicalism speculates that the syntactic subcategorization frames of a verb may be uniquely predictable from the verb’s lexical semantics.

	Event Structure	Meaning
<i>give</i> -type	[[X ACT] CAUSE [BECOME [Y HAVE<POSS-TYPE> Z]]]	X cause Y to have Z
<i>throw</i> -type	[[X ACT<THROW>] CAUSE [Y GO [PATH Z]]]	X cause Z to be at Y

Figure 2. Event Structure of Give-type and Throw-type verbs

As shown in Fig.2, the root of *give*-type verbs such as *loan* is encoded in the caused possession event type and the root of *throw*-type verbs is associated with the caused motion event type. This approach can solve the problem left by constructional approach as in Example 3. Since argument structure is determined by verbs instead of constructions, *loan* as a *give*-type verb entails possession while *throw* entails caused motion.

The lexical resource VerbNet details semantic classes of verbs, where a class comprises verbs that have similar argument structures, following work by Levin [9]. Verbs are classified by their syntactic realization or frames, and each frame is associated with a meaning. For example, verbs such as *push* and *pull* are grouped together in class 12 with roughly the meaning of “exerting force”.

C. Knowledge based Query Expansion Agent

In the feature words extraction subagent, by performing Part of Speech (POS) [10], Stop Words Removal, and Words Stemming, the stemmed word is extracted from the lexical query. The traditional vector space model (VSM) is adopted to represent each stemmed words set $V(sw_1, sw_2, \dots, sw_n)$. For each stemmed words set, the ‘‘Similarity Matching Algorithm’’ [11] is used to search the most similar concept between the stemmed words set and the personalized ontology profile $C_j = [(t_1, w_1), (t_2, w_2), \dots, (t_n, w_n)]$ [12]. For one stemmed word, if it has no similar concept in personalized ontology profile, then discard it from the feature words set; otherwise select the most similar concept from the personalized ontology profile, and choose the top M related terms (t_1, t_2, \dots, t_M) and insert them into the feature words set $STW = V(CQW_{i1}, CQW_{i2}, \dots, CQW_{iM})$. The feature words extraction subagent is shown as Fig.3.

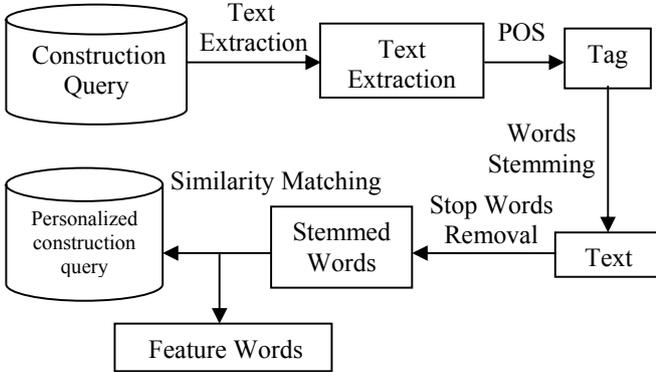


Figure.3. Framework of Feature Words Extraction Subagent

For the Knowledge-Based Query Expansion Subagent, first of all, use the ‘‘special domain ontology determination method’’ [13] to modify the lexical query set as $\{\alpha_1(STW, Df_1), \alpha_2(STW, Df_2), \dots, \alpha_s(STW, Df_s)\}$, where α_i is the weight of the query in a certain domain. More attention users receive, the bigger is the value of α_i , and thus the better α_i can effectively reflect the users’ needs. Through the modification process, modified query with a weight smaller than a threshold is abandoned. By this means, the system can sort out the most popular domain so as to determine the real domain of the lexical query in the form of (STW, Df_i) .

Second, add terms related to the semantics (hyponymy, hypernym, meronymy, synonymy and entailment) into the ontology profile according to WordNet. Semantic relevance is proportional to semantic similarity. This paper calculates the related degree of the terms and the corresponding concept with the method based on the WordNet. The function $STW(CQW_{ij}, C_k)$ is defined as function(1):

$$STW_{j,k} = -\log \frac{(\text{len}(CQW_{ij}, C_k) + 1)}{2 \times \text{depth}} \quad (1)$$

Here, $\text{len}(SW_{ij}, C_k)$ makes reference to the length of the shortest path between the concept k and the term j; depth is the height of the classifier tree. The synonyms are treated as $\text{len}(SW_{ij}, C_k) = 0$, $\text{depth} = 1$, $STW_{j,k} = 1$. The terms will be regarded as the related terms with the concept, and be added

into the refinement query if the related degree is larger than a threshold T .

IV. SIMULATION RESULT

In this study, the system environment is: CPU-- Inter Core2, 2G Memory. The proposed system is implemented in ASP.Net as web-based system using Visual Studio 201. Net Framework 4.

The system result is compared with that of the only knowledge based Query Refinement method. In the simulation, the experiment parameter is set $\alpha=0.5$. To test the approach proposed in this paper, the TREC 8 query sets were adopted as tools, and 8 Masters and 7 Bachelor students majoring in computer science and education participated in the experiments. The new approach is compared with the traditional only knowledge based query expansion method. The performance of the two methods is evaluated in terms of Precision Ratio, Recall Ratio, shown as Fig. 4 and Fig.5.

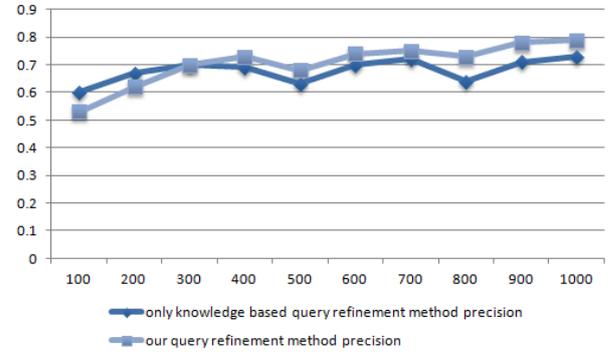


Figure 4. Precision ratio

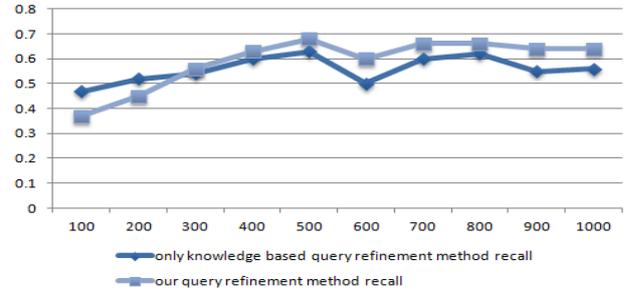


Figure 5. Recall ratio

The overall results show that our approach is better than the traditional only knowledge based query expansion method in terms of the Recall Ratio and Precision Ratio.

V. CONCLUSION

This paper attempts to apply the theory of event structure of lexicalism to information retrieval system to solve the problem of predicate query unsolved by knowledge based approach. The purpose of this paper is to combine the lexical approach and knowledge base approach to realize the expansion of the initial query.

The Event Structure based Query Expansion Agent is to identify the matching argument structures, and the

Knowledge based Query Expansion Agent is to determine the exact domain and to expand the construction query on the basis of the WordNet.

The simulation result shows that the proposed method is more effective than the traditional knowledge based query expansion approach. As the explorative study to apply event structure to query refinement, this paper only compares the new approach with the traditional knowledge based approach. In future studies, this method should be compared with other query improvement methods to test its performance.

REFERENCES

- [1] Berners-Lee, T., Cailliau, R., Graff, J., and Pollerman, B., "World Wide Web: The Information Universe," *Electronic Networking: Research, Applications and Policy*, vol.2, no.1, pp.52-58, 1992.
- [2] Bing Li, Multi-Agent New Query Refinement Approach for Statements Query Retrieval Based on Construction Grammar, *Proceedings of The 2013 2nd International Symposium on Computer, Communication, Control and Automation*, in press.
- [3] A.E. Goldberg. *Constructions at work: The nature of generalization in language*. New York: Oxford University Press, 2006.
- [4] Carola Carstens, "Ontology Based Query Expansion –Retrieval Support for the Domain of Educational Research," *Linguistics and Information Science of the University of Hildesheim*, 2011.
- [5] D. Cai, C. J. Rijsbergen and J. M. Jose, "Automatic Query Expansion based on Divergence," *Proceedings of the Tenth International Conference on Information and Knowledge Management (CIKM-01)*, pp.419–426, 2001.
- [6] E. M. Voorhees, "Query Expansion Using Lexical-Semantic Relations," *Proceedings of the 17th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, pp.61–69, 1994.
- [7] Gaihua Fu, Christopher B. Jones and Alia I. Abdelmoty, "Ontology-based Spatial Query Expansion in Information Retrieval," *OTM Confederated International Conferences*, pp.1466–1482, 2005.
- [8] J. Wang, Z. Li, J. Yao, Z. Sun, M. Li, and W. Ma, "Adaptive User Profile Model and Collaborative Filtering for Personalized News," *In Proceedings of the APWeb 2006*, pp. 474-485.
- [9] Levin, Beth. *English Verb Class and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press, 1993.
- [10] The Stanford Natural Language Processing Group. <http://nlp.stanford.edu/software/tagger.shtml>.
- [11] Jibrán Mustafa, Sharifullah Khan and Khalid Latif, "Ontology Based Semantic Information Retrieval," *Intelligent Systems, 4th International IEEE Conference*, vol.3, pp.14-19, 2008.
- [12] I.Young, "A Dynamic Ontology-based Multi-Agent Context-Awareness User Profile Construction Method for Personalized Information Retrieval," *International Journal of Fuzzy Logic and Intelligent Systems*, December, 2012, vol.12, No.4, pp.270-276, 1598-2645.
- [13] I.Young, A Multi-Agent Personalized Query Refinement Approach for Academic Paper Retrieval in Big Data Environment[J], *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol.16, no.7, pp.874-880, 2012