Modeling Web Knowledge for Answering Event-based Questions

ABSTRACT

For the TREC-style questions, the query terms we get from the original questions are either too brief or often do not contain much relevant information in the corpus. It will be very difficult to find an exact answer in a large corpus because of the surface string mismatch. In order to solve this problem, we present a question answering system QUALIFIER, which employs a novel approach to structurally model the external knowledge from the Web and other resource for Event-based question answering. The results obtained on TREC-11 QA corpus demonstrate the effectiveness of our approach.

1. INTRODUCTION

Open Domain Question Answering (QA) is an information retrieval (IR) paradigm. Modern QA systems [1,2] combine the strengths of traditional IR, natural language processing and information extraction to provide an appropriate way to retrieve concise answers to open-domain natural language questions against the QA corpus. In TREC-11 [4], we employed an innovative approach to model the lexical and world knowledge from the Web and WordNet to support effective QA. This paper investigates the integration and structured use of both world and linguistic knowledge for QA. In particular, we describe a high performance question answering system called **QUALIFIER (QUestion Answering by LexIcal FabrIc and External Re-sources)** and analyze its effectiveness by using the TREC-11 benchmark.

2. QUESTION ANSWERING EVENTS

We propose a novel way to investigate the QA problem and find the solution, which we called ***Event-based Question Answering***. The world consists of two basic types of things: ***entities*** ("*anything having existence (living or nonliving)*") and ***events*** ("*something that happens at a given place and time*") and people often ask questions about them. If we apply this taxonomy to QA task, questions can be considered as "*enquiries about either entities or events*". Generally, questions often show great interests in several aspects or elements of QA events, namely *Location*, *Time*, *Subject*, *Object*, *Quantity*, *Description* and *Action*, etc. Table 1 shows the correspondences of the most common WH-question classes and the QA event elements.

Table 1: Correspondence of WH-Questions & Event Elements

WH-Question	QA Event Elements
Who/Whose/Whom	Subject, Object
Where	Location
When	Time
What	Subject, Object, Description, Action
Which	Subject, Object,
How	Quantity, Description

Our main observation is that a QA event shows great cohesive affinity to all its elements and the elements are likely to be closely coupled by this event. Based on this observation, we derive the following Event-based QA hypothesis:

1. ***Equivalency***: *if all_elements(E_i) = all_elements(E_j), then $E_i = E_j$, and vice versa;*
2. ***Generality***: *if all_elements(E_i) is a subset of all_elements(E_j), then E_i is more general than E_j;*
3. ***Cohesiveness***: *if elements a, b both belong to an event E_i, and a, c do not belong to a known event, then co-occurrence(a,b) is greater than co-occurrence(a,c);*
4. ***Predictability***: *if elements a, b both belong to event E_i, then a => b and b => a.*

(Here " =>" means "induces".)

Normally, the question itself provides some known elements and asks for the unknown element(s). However, for most of the cases, it is difficult to find a correct answer, i.e., the correct unknown element(s). To solve the problems of insufficient known elements and inexact known elements, we model the Web and linguistic knowledge to perform effective QA.

3. EMPLOYING WEB KNOWLEDGE

As the Web is the most rapidly growing and complete knowledge resource in the world, QUALIFIER uses it as an external knowledge resource to solve the problem of insufficient known elements. The terms in the relevant web documents are likely to be similar to or even the same as those in the QA corpus since they both contain same information about the natural facts *(QA Entity)* or the factual events in the history *(QA Event)*.

QUALIFIER stores the original content words in $\underline{q}^{(0)}$ to retrieve the top N_w documents in the Web search engine (E.g. Google) and then extract the terms in those documents that are highly correlated with the original query terms. That is, for $\forall q_i^{(0)} \in \underline{q}^{(0)}$, it extracts the list of nearby non-trivial terms, $\underline{t}_i$, that are in the same sentence or snippet as $q_i^{(0)}$. We compute the weights for all terms $t_{ik} \in \underline{t}_i$ as:

$$weight(t_{ik}) = \frac{d_s(t_{ik} \wedge q_i^{(0)})}{d_s(t_{ik} \vee q_i^{(0)})} \qquad (1)$$

where $d_s(t_{ik} \wedge q_i^{(0)})$ gives the number of web snippets or sentences that contain t_{ik} and $q_i^{(0)}$; and $d_s(t_{ik} \vee q_i^{(0)})$ gives the number that contain either t_{ik} or $q_i^{(0)}$.

Finally, QUALIFIER merges all $\underline{t}_i$ to form $\underline{C}_q$ for $\underline{q}^{(0)}$. It then uses WordNet as a filter to adjust the term weights. The final weight of each term is normalized and the top m terms above the cut-off threshold s are selected to expand the original query:

$$\underline{q}^{(1)} = \underline{q}^{(0)} + \{\text{top } m \text{ terms} \in \underline{C}_q \text{ with weights=s}\} \qquad (2)$$

where m is initially set to 20 in our experiments.

The expanded query should contain more known elements of the QA event. If we classify the terms in $\underline{q}^{(1)}$, they are actually corresponding to one or more of the QA event elements we discussed in Section 2. We explore the use of semantic grouping to structurally utilize the external knowledge extracted from the Web. Given any two distinct terms t_i, t_j, we compute:

WWW 2003, May 20-24, 2003, Budapest, Hungary.
ACM xxx.

1) Lexical correlation:

$$R_l(t_i, t_j) = \begin{cases} 1, \text{ if } t_i \text{ and } t_j \in \text{ same synset;} \\ 0, \text{ otherwise.} \end{cases} \quad (3)$$

2) Co-occurrence correlation:

$$R_{co}(t_i, t_j) = \frac{d_s(t_i \wedge t_j)}{d_s(t_i \vee t_j)} * \max\{0, p_{ij} - \frac{1}{k_j}\} \quad (4)$$

where $p_{ij} = \frac{d_s(t_i \wedge t_j)}{\sum_{k=1}^{k_j} d_s(t_k \wedge t_j)}$ (5)

where $d_s()$ is as defined in Eqn 1 and k_j gives the number of other terms in $\underline{K}_q$ that co-occur with t_j. Thus the max{} expression indicates that only those terms whose normalized co-occurrence probability is above $1/k_j$ (or average) will have a positive co-occurrence correlation value.

3) Distance correlation:

$$R_d(t_i, t_j) = \frac{1}{\frac{\sum |Pos(t_i) - Pos(t_j)|}{ds(t_i \wedge t_j)}} \quad (6)$$

where $Pos(t_i)$ (or $Pos(t_j)$) denotes the position of term t_i (or t_j) in a web snippet or sentence. $|Pos(t_i) - Pos(t_j)|$ gives the term distance.

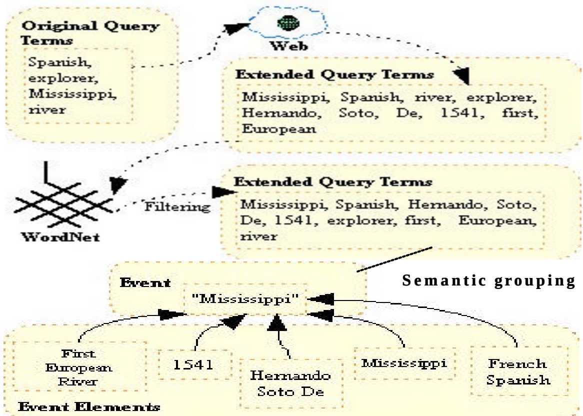

Figure 1: Example for Structured Query Formulation

Then we form the semantic groups for the terms using a modified version of the algorithm outlined in [3]. For question *"What Spanish explorer discovered the Mississippi River?"*, we perform the knowledge modeling and obtain the semantic groups as shown in Figure 1. One promising advantage of our approach is that we can answer any factual questions about the elements in this QA event. For instance, *"When Mississippi River discovered?" and "Which river were discovered by Hernando De Soto?"* etc.

4. EVALUATION

4.1 TREC-11

For the NIST human assessment on the 500 questions of TREC-11 QA track. QUALIFIER answers ***290*** questions correctly with confidence weighted scores ***0.610***, which places us among one of the top performing systems. Table 3 shows our system statistics:

Table 3: Performance over TREC-11 500 Questions

# right	290	Precision	0.580
# unsupported	18	Confidence-weight score	0.610
# inexact	17	Precision of recognizing NIL	0.241
# right	175	Recall of recognizing NIL	0.891

4.2 Effects of Web Search Strategies

For Web search, we adopt Google as the search engine and examine snippets instead of looking at full web pages as reported in [5,6]. We study the performance of QUALIFIER by varying the number of top ranked web pages returned N_w, and the cut-off threshold s (see Eqn 2) for selecting the terms in $\underline{C}_q$:

Table 2 summarizes the effects of these variations on the performance of TREC-11 questions by showing the precision, which is the ratio of correct answers returned by QUALIFIER. From the results, we can see that the best result is obtained when N_w = 75 and s = 0.2.

Table 2: The Precision Score of 25 Web Runs

s \ N_w	10	25	50	75	100
0.1	0.492	0.492	0.494	0.500	0.504
0.2	0.536	0.536	0.538	**0.548**	0.544
0.3	0.506	0.506	0.512	0.512	0.512
0.4	0.426	0.426	0.430	0.432	0.428
0.5	0.398	0.398	0.412	0.418	0.412

4.3 Event-based Query Formulation

We conducted several tests on modeling the knowledge for QA. For each run, we compute *P*, the precision, and *CWS*, the confidence-weighted score. Table 3 summarizes the test results.

Table 3: Results of Different Query Formulation Methods

Method	P	CWS
Baseline	0.438	0.640
Baseline + Web	0.548	0.754
Baseline + Web + WordNet	0.588	0.795
Baseline + Web + WordNet + structure analysis	**0.634**	**0.824**

Here we can draw the following observations.

- The Simple Web-based query formulation improves the baseline performance by 25.1% in Precision and 31.5% in CWS, which is more significant than WordNet's contribution, with only 7.3% increments in Precision.
- The best performance (P: 0.634, CWS: 0.824) is achieved by the structured modeling of the Web and WordNet knowledge as outlined in Section 3.

5. CONCLUSION

We have presented the techniques used in QUALIFIER system, which employs a novel approach to Event-based QA with the modeling of the Web knowledge. Using the structured query formulation, we can achieve an answer accuracy of 0.63 and CWS of 0.82, which showed the effectiveness of our approach.

6. REFERENCES

[1] E. Brill, J. Lin M. Banko, S. Dumais, and A. Ng, "Data-intensive question answering", (TREC'2001)

[2] C. Clarke, G. Cormack and T. Lynam, "Exploiting redundancy in question answering". (SIGIR'2001).

[3] J. Liu and T. S. Chua, "Building semantic perceptron net for topic spotting",(ACL'2001).

[4] E. M. Voorhees. "Overview of the TREC 2002 Question Answering Track." (TREC'2002)

[5] H. Yang and T. S. Chua, "The Integration of Lexical Knowledge and External Resources for Question Answering", (TREC'2002)

[6] H. Yang and T. S. Chua, "QUALIFIER: Question Answering by Lexical Fabric and External Resources", (EACL'2003). To appear.