

Q&A

{jheom, btzhang}@scai.snu.ac.kr

A Query Classification Method for Question Answering on a Large-Scale Text Data

Jae-Hong Eom Byoung-Tak Zhang
School of Computer Science and Engineering, Seoul National University

가 가 , 가 (가)
가 가 . 가
Conference) 가 NIST DARPA TREC(Text REtrieval
1999 Q&A trec
TF-IDF
(Tagging)
1999
TREC-8 (Question) (Answer) Question & Answer
1. IST (corpus)
[3].
(Information Overload) 가 [1].
가 가 FAQ
(Web) 가 ' FAQFINDER ' [8]
가 가 WordNet [9]
가 가 가
가 가 () , 가 가
가 가 () WordNet
가 가 가 [10] (knowledge-base)
가 가 가
가 가 가
[8][9][10].
가
[2] 'SUMMAR-

2. TREC Q&A

TREC Q&A

(Question)

가

2.1 TREC Q&A

NIST DARPA

TREC(Text REtrieval Conferenc-

e)

(CD 5) Q&A

4-5 CD

TREC

1

528155

(field)

SGML(Standard Generalized Markup Language)

(Tag)가 가

1.

DISK 4	FR 94	55631	Federal Register (1994).
	FT 91 ~ 94	210158	Financial Times (1991 ~ 1994)
DISK 5	FBIS	130471	Foreign Broadcast Information Service (1996)
	LATIMES	131896	Los Angeles Times (1989, 1990)

1.96GB , CD 4

CR93(Congressional Record of the 103rd Congress 1993) 가 가 Q&A

2.2 TREC Q&A

50 (byte) 250

가

5

1.

```
<top>
<num> Number: 1
<desc> Description:
Who is the author of the book, "The Iron Lady: A Biography of Margaret Thatcher"?
</top>
```

```
Answer of Question 1:
LA111289-0002 1 210677 ... THE IRON LADY: A BIOGRAPHY OF MARGARET THATCHER BY HUGO YOUNG (... FT943-8867 2 200395 ...dubbed "the Iron Lady" the book of HUGO YOUNG was a...
```

3. Q&A

3.1 (Question Classification)

16가

(Q-Tag)

2

2.

Q-Tag	
Q-NAME	" Who/What/Which ~"
Q-COST	" How much ~"
Q-LOCATION	" Where/What ~"
Q-TIME	" When ~"
Q-YES-NO	" Is it ~"
Q-DURATION	" How long ~"
Q-PERSON	" Who ~"
Q-DATE	" What date/When ~"
Q-AREA	" How big ~"
Q-LENGTH	" How long/high/etc ~"
Q-ORGANIZATION	" Who ~ /What ~"
Q-AGE	" How old ~"
Q-WEIGHT	" How heavy/big ~"
Q-NUMBER	" How many ~"
Q-QUANTITY	" How big/much ~"
Q-YEAR	" When/What year ~"

200

(parsing)

2

[4]

(term)

가 (weight)가

3.2

(Target Document Preprocessing)

2

(Tagging)

가

3.3

가 (Answer Evaluation)

$$Score(i) = \frac{1}{Rank(i)}$$

0.25, 0.20, 0.0

가

1.0, 0.5, 0.33,

가 NIST

TREC-8 Q&A

4.

TREC-8
 198 (200 2)
 528155
 가
 1
 3 250byte run
 TREC-8 Q&A

4.1 (98-199)
 50 가 250

3. tf-idf

	50 byte run	250 byte run
1	11	13
2	12	12
3	4	5
4	8	11
5	2	3
	161	154
	0.105	0.121

4.

	50 byte run	250 byte run
1	33	33
2	21	24
3	12	12
4	13	18
5	9	13
	110	98
	0.265	0.283

5.

Question Answering

TREC-8

3,4
 가
 (tf-idf)

(가

- [1] Pattie Maes, "Agents that reduce work and information overload," Communications of the ACM, Vol. 37, No 7, pp.30-31, 1994.
- [2] Inderjeet Mani, Mark T. Maybury, Advances In Automatic Text Summarization, 1999.
- [3] Eduard Hovy, Chin-Yew Lin, "Automated Text Summarization in SUMMARIST," in Proceedings of the Workshop of Intelligent Scalable Text Summarization, July 1997.
- [4] Chin-Yew Lin, Eduard Hovy, "Identify Topics by Position," in Proceedings of the 5th Conference on Applied Natural Language Processing, March 1997.
- [5] D.-H. Shin, B.-T. Zhang, "A Two-Stage Retrieval Model for the TREC-7 Ad Hoc Task," Draft of Text Retrieval Conference-7, pp. 501, November 1998.
- [6] J. Prager, D. Radev, E. Brown, A. Coden, "The Use of Predictive Annotation for Question Answering in TREC8," in Proceedings of Text Retrieval Conference-8, pp. 323-335, 1999.
- [7] D.-H. Shin, Y.-H. Kim, S. Kim, J.-H. Eom, H.-J. Shin, B-T Zhang, "SCAI TREC-8 Experiments," Draft of TREC-8, 1999.
- [8] K. Hammond, V. Kulyukin, S. Lytinen, N. Tomuro, and S. Schoenberg, "Question Answering from Frequently-Asked Question Files: Experiences with the FAQ Finder System," Univ. of Chicago, Department of Computer Science Technical Report TR-97-05, 1997.
- [9] Miller, G. A. 1995. WordNet: A Lexical Database for English. Communications of the ACM, 38(11).
- [10] Sanda Harabagiu and Dan Moldovan, "An Intelligent System for Question Answering," in the Proceedings of the 5th Conference on Intelligent Systems, June 1996, Reno NV, pages 71-75.