

질문응답체계에서 패턴정합법에 의한 비사실적질문류형분류의 한가지 방법

리 청 한

위대한 령도자 김정일동지께서는 다음과 같이 교시하시였다.

《현시대는 과학과 기술의 시대이며 이르는 곳마다에서 요구하는것은 기술입니다. 기술을 몰라가지고서는 경제조직사업과 생산지휘를 바로할수 없으며 사회주의건설에 적극 이바지할수 없습니다.》(《김정일전집》 제2권 499~500페이지)

질문처리는 질문응답체계의 첫 공정으로서 질문응답체계전반에서 매우 중요한 부분의 하나이다. 왜냐하면 질문처리를 통하여 얻어진 결과들이 질문류사도평가, 문서검색, 패째지검색, 대담추출모듈에서 리용되기때문이다.

질문처리단계에서 가장 중요하고 또 가장 어려운 문제의 하나가 다양한 패턴으로 주어지는 질문들에 대한 질문류형분류이다.

질문류형은 사용자의 질문이 미리 정해놓은 분류체계가운데서 어디에 속하는가 하는 것이다.

질문응답체계[1]가 인터넷상에서 사용될 경우 사용자의 질문은 다양한 형태로 나타나게 된다.

따라서 질문류형분류는 문장의 형태나 단어의 쓰임에는 관계없이 같은 의도를 가진 질문을 같은 유형으로 분류해낼수 있어야 한다.

최근에 개발된 대부분의 질문응답체계[2]들은 사용자의 질문의도를 파악하기 위한 분석과정 즉 질문류형분류를 반드시 포함하고있다. 그러나 대체로 사실적질문에 대한 질문류형분류에 대한 연구는 진행되고있으나 비사실적질문에 대한 류형분류방법에 대한 연구는 진행되지 못하고있다.

일부 선행연구[3]들에서는 비사실적질문을 사실적질문으로 넘기여 질문류형분류를 진행하고있다.

이로부터 논문에서는 패턴정합법에 의한 비사실적질문에 대한 질문류형분류방법에 대하여 제안하였다.

1. 패턴정합법에 의한 비사실적질문의 류형분류

비사실적질문응답체계를 합리적으로 구축하는데서 가장 중요한 첫 공정이 바로 질문류형분류(question type classification)이다. 왜냐하면 학습자의 자연언어질문이 제기되면 그것이 어떤 류형인가를 식별하여야(질문류형이 결정되어야 질문류사도계산을 위한 검색공간이 줄어들게 되고 같은 류형의 질문들과만 류사성을 계산할수 있으므로) 검색속도의 측면에서 매우 효과적이기때문이다.

결국 질문류형분류는 질문류사도계산에서 중요한 열쇠로 된다.

비사실적질문이란 학습자의 질문에 대한 응답요구가 특정한 사실이 아니라 어떤 개념의 정의라든가 사건이 진행되게 된 리유 등과 같이 설명을 요구로 하는 질문 즉 간단

히 말하면 응답으로서 문장이나 단락을 요구하는 질문을 말한다.

실례로 《자료구조와 알고리즘》학과목에서 《2진나무란 무엇인가?》라는 질문은 개념의 정의를 응답으로서 요구하는 비사실적질문이다.

질문응답체계를 학과목자체학습에 리용할 때 학생들속에서 학과목자체학습수행기간에 제기되는 질문들을 통계적으로 조사하면 비사실적질문들의 비율이 사실적질문들보다 더 높다는것을 알수 있다.

론문에서는 xx학부 xx학과 과정안에 반영된 42개의 모든 학과목들에 대하여 자체학습과제서에 제시된 질문문제들과 학기말 및 학년말시험문제들을 통계적으로 조사하여 비사실적질문들의 유형을 12가지로 설정하고 사용자들로부터 제기되는 질문을 패턴정합법에 의하여 질문유형을 결정하였다.

론문에서는 질문문장들의 특성에 따르는 질문유형을 결정하기 위하여 선행연구[2]에 소개된 조선어질문의 품사적구성에 대하여 분석하였는데 그 분석결과는 다음과 같이 5가지 형식으로 된다.

불완전명사+물음종결토
물음대명사+물음종결토
자동사+물음종결토
타동사+물음종결토
형용사+물음종결토

우리는 xx학부 xx학과 1학년에서 취급하는 《프로그램작성법》학과목에 대한 자체학습진행시 학생들과의 질문응답과정에 수집한 610개의 질문문장들과 이 학과목의 5개장의 298개의 연습문제, 300개의 시험문제들에 대응한 문장들을 분석하였다.

이 분석에 리용된 물음대명사, 물음종결토, 시킴을 나타내는 종결토들에는 《무엇때문에, 왜, 언제, 어디서, 어느것이, 어떻게, 얼마나, ㄴ니까(습니까), ㄴ시요》 등 여러가지가 있다.

또한 질문의 기본형식은 《...단어 ... (해당한)물음토》로 되어있는데 크게 두가지 표현형식을 가진다.

첫째 형식은 문장안에 무엇, 누구, 어디, 몇, 언제, 얼마 등의 물음대명사와 그뒤에 물음종결토가 붙은 단어결합형식이고 둘째 형식은 무슨, 왜, 어찌, 웬, 어떻게 등의 물음의 뜻을 나타내는 관형사, 부사, 형용사, 동사단어와 그뒤에 역시 물음종결토가 붙은 단어결합형식이다. 여기서 《...알고리즘을 작성하여라.》와 같은 시킴문(연습문제실례, 시험문제실례의 기본형식의 하나)은 《...알고리즘은 어떻게 작성하는가?》와 같이 둘째 형식의 질문으로 바꾼다.

이상의 내용을 종합하여 론문에서는 질문형식과 물음종결토에 따라 질문문장들을 표 1과 같이 모두 12개의 유형으로 무리화한다.

표 1. 질문유형에 따르는 패턴정합표

No.	유형이름	기호	유형의 기본의미	물음종결토실례	대응하는 물음패턴
11	What is	Q1	무엇인가	...란 무엇인가?, ...은 무엇이라고 하는가?, ...란 어떤 것인가?	A+무엇입니까?
22	Why	Q2	무엇때문에	왜, 무엇때문에, 어떻게 되어, 무슨 리유로, 어찌	A+왜+...ㄴ니까?

표제 속

No.	류형이름	기호	류형의 기본의미	물음종결로실례	대응하는 물음패턴
33	When	Q3	언제	언제, 어느 시각에, 어느때	...때 + A + B니까?
44	Where	Q4	어디에서	어디에, 어디에서, 어디로, 어떤 위치에서, 어느곳에서	A + 어디에 + B + ... B니까?
55	Who	Q5	누구	누구, 어떤 사람, 어떤 단위, 어느것이, 어느	A + 어느것인가?
66	How To	Q6	어떻게	어떻게, 어찌하여, 어떤 상 태(모양)로	...이면 어떻게 됩니까?
77	How Many	Q7	얼마나	얼마나, 얼마만큼, 얼마인가	...은 얼마로 되는가, ...계산하 면 ...얼마입니까?
88	Yes No	Q8	인가 아닌가	인가 아닌가, 있을가 없을 가, 그런가 그렇지 않은가	...계산할수 있는가
99	When How	Q9	언제 어떻게	언제 어떻게, 어느때 어떤 상태로	...때에 ...이 어떻게 되는가
110	Where How	Q10	어디서...어떻게	어디서...어떻게	...에서 ...이 어떻게
111	When How Many	Q11	언제...얼마나	언제...얼마나	...일 때...이 얼마인가
112	When Where	Q12	언제...어디서	언제...어디서	...일 때...어디서

이로부터 논문에서는 질문자의 질문문장을 표 1에서 준 질문들에 쓰이는 대표적인 종결로, 관형사, 불완전대명사단어가 질문문장에 포함되는가에 따라 12가지 류형 Qi, i = 1, ..., 12의 하나로 판단하여 분류를 진행한다.

만일 한 질문문장에 서로 다른 류형을 의미하는 2개이상의 관형사, 불완전대명사, 물음도가 들어있는 경우에는 문장의 앞부분에서부터 처음으로 출현하는 품사의 의미를 가지는 류형에 속하는것으로 한다.

실례 1

질문문장 : C++에서 참조형변수란 무엇이며 원변수와 어떤 관계가 있습니까?

물음대명사 : 무엇이며, 어떤

정합된 질문류형 : Q1

실례 2

질문문장 : 클래스의 파괴자함수를 다중정의할수 있는가?

종결로 : ...할수 있는가

정합된 질문류형 : Q8

실례 3

질문문장 : 구축자함수란 무엇이며 왜 필요한가?

물음대명사 : 무엇, 왜

정합된 질문류형 : Q1

2. 실험결과 및 분석

론문에서는 비사실적질문에 대한 질문류형의 평가를 진행하기 위하여 대상자료로서 xx학부 xx학과 1학년에서 취급하는 《프로그램작성법》학과목에 대한 자체학습진행시 학생들과의 질문응답과정에 수집한 610개의 질문문장들과 이 학과목의 5개 장의 298개의 연습문제, 300개의 시험문제들에 기초하여 만든 1208개의 표준질문들을 준비하였다.

실험을 위한 대비방법들은 다음과 같다.(표 2)

방법 1: 유한자동체를 리용한 질문류형분류방법

방법 2: 호상정보량을 리용한 질문류형분류방법

표 2. 질문류형분류방법의 성능비교

질문개수	방법 1/%	방법 2/%	제안한 방법/%
1208	77.1	77.6	89.3

맺는 말

xx학부 xx학과 과정안에 반영된 42개의 모든 학과목들에 대하여 자체학습과제서에 제시된 질문문제들과 학기말 및 학년말시험문제들을 통계적으로 조사하여 비사실적질문들의 유형을 12가지로 설정하고 사용자들로부터 제기되는 질문을 패턴정합법에 의하여 질문류형을 결정하였다.

이 방법은 비사실적질문응답체계의 질문류형분류에 효과적으로 리용할수 있다.

참고 문헌

- [1] Zijng Hui et al.; Proceedings of the 5th International Joint Conference on Natural Language Processing, 938, 2011.
- [2] Zhicheng Zheng et al.; Proceedings of the 5th International Joint Conference on Natural Language Processing, 947, 2011.
- [3] Stefan Riezler et al.; Proceedings of the 22nd International Conference on Computational Linguistics, 737, 2008.

주체108(2019)년 11월 5일 원고접수

A Method of Non-Factoid Question Type Classification Based on Pattern Matching in Question Answering System

Ri Chong Han

In this paper, we consider the question type classification method for non-factoid question based on pattern matching.

Keywords: pattern matching, question answering, question type classification