

온톨로지에 기초한 질문응답체계실현의 한가지 방법

최명옥, 동승철

경애하는 최고령도자 김정은동지께서는 다음과 같이 말씀하시였다.

《과학기술과 경제의 일체화를 다그치고 나라의 경제를 현대화, 정보화하는데서 과학 기술부문이 주도적인 역할을 하도록 하여야 합니다.》

질문응답체계들은 전통적으로 질문과 가능한 응답들사이의 표면적인 차이들을 없애기 위해 다량의 어휘자원들에 의거하여 개발되였다.

선행연구[1]에서는 문법구조정합을 패세기(passage)검색과 응답추출에 적용하고 복합 질문에 대한 응답에서 의미정보의 역할을 강조하였으며 프레임망(FrameNet)에 의미와 관련한 정보를 추가한 《술어-인수》구조에 대하여 논의하였다.

선행연구[2, 3]에서는 후보단어들을 정합하기 위한 의미적관계를 리용하고 반결합과 확장된 어휘집합체계를 적용하였다. 그리고 온톨로지정보를 문장론적인 구문분석과 결합하는 방법에 대하여 논의하였다.

현재 세계적으로뿐만아니라 우리 나라에서도 질문응답의 자동화를 실현하기 위한 연구가 활발히 진행되고있지만 지금까지 응답의 정확률이 높지 않다.

본문에서는 의미정보를 리용하여 추론을 진행하는 온톨로지에 기초한 조선어질문응답체계의 실현방법을 제안하였다.

1. 류사도알고리즘

RDF(Resource Description Framework)나 RDFS(RDF-Schema)로 작성되는 의미론적인 서술문들은 웹메타자료를 표현하기 위한 기본프레임구조를 제공한다. 의미웹브에서는 프레임구조에 의한 검색을 지원하며 이것을 질문응답에 적용할수 있다.

AQUA(Automatic QUEStion Answering)는 온톨로지를 리용하는것을 전제로 한다.

AQUA는 의미론적인 서술문들을 리용하여 추론을 진행하며 질문에 대한 가능한 응답의 수를 감소시킨다. 온톨로지를 리용하여 전형적인 검색기구들에 의하여 얻어지는 색인단어를 탐색한다.

추론기구는 모든 용어가 형태를 가지고있으며 모든 술어가 령역과 관련되어있는 다중분류된 논리구조내에서 동작한다.

분류정보는 사용자들이 응답을 예측할수 있는 응답종류에 대한 정보를 제공한다. 그러므로 사용자들은 체계가 생성할수 있는 응답의 형태를 예측할수 있다.

AQUA체계에서 구현되는 류사도알고리즘은 다음과 같다.

1. 질문 Q 를 제공한다.
2. 질문의 문법구성요소들에서 물음을 구문분석한다.
3. 조선어질문을 질의로 전환한다.

4. 온톨로지에 대하여 질의를 실행한다. 만일 성공하면 응답을 제공하고 5단계에로 이행한다.

실패하는 경우 다음의 형태로 질문을 분류한다.

무엇—객체명세서, 동작정의

누구—사람명세서

언제—날자

어느—객체명세서, 속성

어디—지리적위치

새로운 질문 Q' 로 질문 Q 를 변환한다.

새로운 질문 Q' 로 검색기구를 실행시킨다.

질문 Q' 를 만족시키는 검색된 문서들을 분석한다.

응답추출을 진행한다.

응답선택을 진행한다.

5. 중지

질문평가는 사용자질문에서 리용되는 관계들의 이름과 지식기지에서 리용되는 관계들의 이름의 사영에 의존한다.

AQUA는 관계들의 대행체(agent)이름을 제공하는 류사도알고리즘을 포함한다. 류사도알고리즘은 Dice결수와 WordNet를 리용한다. 또한 온톨로지의 부분모임을 포함하고있는 그래프와 질문으로부터 얻어진 그래프를 리용한다.

출력은 개념/관계들과 객체관계이름사이의 류사도척도이다. 류사도값이 주어진 턱값 아래에 있는 경우 AQUA는 WordNet로부터 합성모임을 제공하고 사용자는 제공되는 합성모임들중에서 하나의 의미를 선택하여야 한다.

2. AQUA처리모형

AQUA처리모형은 자연언어처리, 논리학, 온톨로지들과 정보검색을 통합하는 단일한 프레임워크를 제공한다.

처리모형에는 사용자호상작용, 질문처리, 문서처리와 응답처리를 포함하는 4가지 단계가 있다.

① 사용자호상작용

사용자는 질문을 입력하고 응답이 정확한가, 아닌가를 확인한다.

질문대면부에서 사용자는 단순한 대화창문인 사용자대면부를 리용하여 조선어로 질문을 입력한다. 사용자는 응답이 만족하지 않으면 질문을 수정할수 있다. 응답들은 정렬된 모임으로 사용자에게 현시된다.

응답확인에서 사용자는 응답에 동의 혹은 반대를 지적함으로써 AQUA에 반결합한다.

② 질문처리

질문처리에서는 사용자에게 의하여 제시되는 질문을 리해한다. 여기서는 질문분석, 질문표현과 분류와 같은 여러 단계들을 요구한다.

NLP구문분석기는 문장을 주어, 동사, 전치사구, 형용사와 보어들로 분할한다. 이 모듈의 출력은 질문의 논리적표현이다.

해석기는 통합과 분해알고리즘들을 리용하여 지식기지에서 질문의 논리적증명을 발견한다.

WordNet류의어목록은 AQUA의 사전자원이다.

온톨로지는 사람, 조직, 연구분야, 계획, 발행, 기술과 사건을 포함한다.

실패-분석체계는 주어진 질문의 실패를 분석하고 질문이 왜 실패했는가에 대한 설명을 준다. 그다음 사용자는 이를 위하여 새로운 정보를 제공할수 있다. 이러한 처리는 필요할 때까지 반복될수 있다.

질문분류 및 재형성에서는 질문을 AQUA에서 지원하는 어떠한 형태에 속하는것으로 분류한다.(무엇, 누구, 언제, 어느것, 왜, 어디) 이 분류는 다만 증명이 실패했을 때 진행된다. 그리고 AQUA에서는 정보검색방법을 리용한다. 이것은 AQUA가 문서처리와 응답추출을 수행해야 한다는것을 의미한다.

③ 문서처리

문서모임이 선택되고 단락모임이 추출되며 2개의 부분품들로 구성된다.

탐색질문형식화에서는 초기의 질문을 변환한다. 변환규칙을 리용하여 Q 를 새 질문 Q' 로 변환한다. 동의어들이 리용되고 금지어들이 제거되며 어간추출이 진행된다.

탐색기구에서는 실마리어모임을 리용하여 문서모임을 위한 웹탐색을 진행한다.

④ 응답처리

응답처리에서는 단락으로부터 응답이 처리되고 응답에 관한 점수를 주며 2개의 부분품들을 리용한다.

단락선택에서는 응답을 가지는 적합한 문서모임으로부터 단락들을 추출한다.

응답선택에서는 응답들을 클러스터화하고 투표방법으로 응답에 점수를 주며 후보목표표를 얻는다.

3. 효과성평가

온톨로지에 기초한 질문응답체계를 구성하고 질문응답체계의 효과성을 평가하였다.

학습자 100명을 대상으로 《프로그램작성법》과목에 대하여 질문응답을 진행하여 고전적인 문장구조분석에 의한 질문응답체계와 온톨로지에 기초한 질문응답체계를 비교하였다. 실험에서는 《체계가 제출한 응답이 정확한가》에 대한 대답을 놓고 평가를 진행하였다. 제안한 질문응답체계의 효과성평가를 다음의 표에 보여주었다.

표. 질문응답체계의 효과성평가

구분		틀리다	정확하다	애매하다
선행방법	대답수/명	20	62	18
	비율/%	20	62	18
제안한 방법	대답수/명	4	75	21
	비율/%	4	75	21

표에서 보여준것처럼 문장구조분석에 기초한 질문응답체계에 비하여 온톨로지에 기초한 질문응답체계의 정확률이 훨씬 개선되었다는것을 알수 있다.

맺 는 말

의미정보를 리용하여 추론을 진행하는 온톨로지에 기초한 조선어질문응답체계의 실현방법을 제안하고 질문응답체계의 효과성을 평가하였다.

참 고 문 헌

- [1] Anjali Saini, P. K.Yadav; International Journal of Engineering and Computer Science, 6, 3, 20453, 2017.
- [2] Ajitkumar M. Pundge et al.; International Journal of Computer Applications, 141, 3, 34, 2016.
- [3] P. M. Athira et al.; International Journal of Web & Semantic Technology(IJWeST), 4, 4, 31, 2013.

주체108(2019)년 5월 5일 원고접수

A Method of Implementing an Ontology-based Question Answering System

Choe Myong Ok, Tong Sung Chol

In this paper we suggested a method of implementing automatic ontology-based question answering system(AQUA). AQUA uses the semantic annotations to performing inferences.

Key words: ontology, question answering, semantic retrieval