

질문응답체계를 위한 지식기반온톨로지 자료통합의 한가지 방법

최명옥

현재 온톨로지기반질의응답체계의 하나인 AQUA(Automated QUESion Answering)는 닫힌
영역질의응답체계[1]이다. 따라서 다중화된 온톨로지를 조종하기 위하여 구성방식을 확장
할 필요성이 나서고있다.[2]

본문에서는 메타온톨로지와 류사도알고리즘에 기초한 중개자(mediator)를 리용하여
질문들과 온톨로지공간사이를 사영함으로써 온톨로지자료통합을 진행하는 방법을 제안
하였다.

1. 자료통합원리

온톨로지자료통합에서는 질문의 용어들과 온톨로지적인 관계사이의 사영을 진행할수
있는 중개자가 필요하다. 여기서는 중개자를 리용하여 AQUA에 통합체계를 결합하는 문
제를 논의하였다.

온톨로지자료통합문제는 류사도알고리즘과 결합되는 메타-온톨로지를 리용하는것으
로 해결한다. 메타-온톨로지는 매개 자원의 관계들에 대한 정보를 포함한다.

새 자원에 따르는 새로운 메타정보가 점차적으로 추가될수 있기때문에 체계는 제한
을 받지 않게 된다.

질문응답처리는 3개의 단계들로 나누어 진행한다.

① 질문작성

1계술어론리(FOL)로 씌여진 질문을 받고 그것을 부분질문들로 나눈다.

② 선택수속

질문에 적절한 온톨로지들의 부분모임을 제공한다.

③ 질문충족알고리즘

AQUA에서 이미 실현된 질의응답의 표준기술들을 리용하여 응답한다.

2. 자료통합방법

자료통합체계의 목적은 여러가지 자료원천들사이에 유일한 대면부를 제공하는것이다.
자료통합체계의 설계는 여러가지 측면들을 이루는 매우 복잡한 과제이다.

중개자의 선택절차는 각이한 자원들을 리용하여 주어진 질문에 적합한 토막들을 찾
는 방식으로 진행된다. 그것은 온톨로지적인 탐색공간을 횡단하는 방법으로 볼수 있다.

온톨로지들에 대하여 동작하는 관련온톨로지생성방법을 보기로 하자.

관련온톨로지생성방법은 각이한 원천(메타-온톨로지)들의 메타서술이 있다는 사실에

기초한다. 관련되는 원천들이 질문에 대한 답을 분간하는데 도움이 된다면 선택된 원천을 리용하여 수행될수 있다. 메타-온톨로지는 개념들과 개념들사이의 관계에 대한 형식적서술을 포함하며 우점은 이러한 서술들이 체계에 의존하지 않는다는것이다.

메타-온톨로지 Meta-O는 매 온톨로지의 매 개별적서술의 결합으로 이루어져있다.

$$\text{Meta-O} = O_1 \cup \dots \cup O_n$$

여기서 Meta-O는 전체적인 스키마(schema)이고 O_i 는 온톨로지 i 의 서술이다.

$$O_i = \{R_{1i}, \dots, R_{ni}\}$$

여기서 R_{ji} 는 온톨로지 i 에서 관계 j 를 의미한다.

매 온톨로지서술은 형과 값을 가진 FOL(1계술어론리)로 씌여진 관계들의 모임으로 구성된다. 이 서술들은 FOL 혹은 DL(서술론리)인 형식언어로 정의될수 있다.

관련온톨로지생성알고리즘을 다음과 같이 정의한다.

Procedure generate-relevant-ontologies(O, Q)

/* O 는 온톨로지들의 모임이고 Q 는 더하기표준형의 질문*/

걸음 1. 질문계획작성자는 질문을 부분질문으로 분해한다.

$$Q(X_1, \dots, X_n) = Q_1(Z_1) \cup Q_2(Z_2) \cup \dots \cup Q_n(Z_n)$$

여기서 매 Q_i 는 술어로서 정의될수 있다.

$$Q_1(W_1, \dots, W_n) = (X_1, \dots, X_n), \exists X_1, \dots, X_n \omega$$

걸음 2. $S = \theta; i = 1;$

걸음 3. For each Q_j do

걸음 4. For each $\eta_i(Y_1, \dots, Y_n) \in \text{Meta-O}$ do i

Begin

온톨로지기반류사도알고리즘을 리용하여 $\text{Sim}(\omega(X_1, \dots, X_n), \eta(Y_1, \dots, Y_n))$ 을 계산한다.

If $\exists \beta$ 가 $\omega(X_1, \dots, X_n)$ 과 $\eta(Y_1, \dots, Y_n)$ 사이에서 사영 then $\Pi = \Pi \cup \{(\eta(Y_1, \dots, Y_n), O_i)\}$ 이다. 즉 관계 $\eta(Y_1, \dots, Y_n)$ 을 온톨로지 O_i 우에 추가한다.

Else

End do i

걸음 5. $S = S \cup \Pi$

End do j

걸음 6. Return S

걸음 7. If S 가 θ 이면 《질문에 적합한 온톨로지모임이 없습니다.》를 출력한다.

else S 에서 온톨로지부분모임을 리용하여 질문을 평가한다.

걸음 8. End generate-relevant ontologies

3. 류사도계산

류사도는 언어학, 인공지능, 자연언어처리와 모호론리에서 중요하게 리용되며 그 측정범위는 단어의미매성으로부터 시작하여 본문요약, 정보추출과 검색, 질의응답, 자동색인과 코드의 자동교정에 놓인다.

분류법에서 유사도를 평가할 때 유사도에 접근하는 가장 최초의 방법은 비교되는 2개의 개념들사이거리를 평가하는것이다. 이 방법은 유사도를 측정함으로써 리용된다.

논문에서는 개념 유사도와 관계 유사도를 평가하기 위한 유사도알고리즘을 제안하고 실현하였다. 알고리즘은 사용자질문(질문에서 실체들과 온톨로지에서 정보적클래스들)으로부터 얻어진 확장그래프와 질문과 관련있는 온톨로지들의 부분모임을 표현하는 그래프들을 비교한다. 매개물단계로서 그것은 2개의 그래프들사이 마디상의 공통부분을 창조한다. 다시말하여 알고리즘은 온톨로지에서부터 가장 정보적인 클래스들과 Dice결수를 리용하여 개념 유사도와 관계 유사도를 평가한다.

알고리즘은 사용자질문에서 모든 기초적인 용어를 리용하며 온톨로지에서 개체들로서 그것들을 찾는다. 용어들이 발견되기만 하면 온톨로지 (G_2)의 부분이 린접한 클래스들을 포함하여 검토된다. 그다음 온톨로지에서의 지식을 리용하여 증대된 질문 G_1 과 G_2 와의 사림이 구조적 유사도를 평가하기 위하여 수행된다. 그것은 공통부분 G_3 에서 여러가지 관계들이 기초적인 인수들을 포함할수 있는 경우이다. 그다음 Dice결수를 리용하여 사용자질문의 원소들을 포함한 모든 관계들에 대한 유사도가 계산된다. 마지막으로 최대 Dice결수값을 가진 관계가 가장 유사한 관계로 선택된다.

관계에 대한 유사도알고리즘은 다음과 같다.

X_1 과 X_2 는 기초원소이다.

걸음 1. 질문을 1계술어론리로 번역한다. 실례로 술어_이름 (X_1, X_2)

걸음 2. \exists 관계 (C_1, C_2) $\wedge \exists C_1 \supset X_1 \wedge \exists C_2 \supset X_2$ 이면

관계 (C_1, C_2)는 C_1 과 C_2 사이의 온톨로지적관계

걸음 3. \exists 관계 (C_1, C_2) $\wedge \exists C_1 \supset X_1 \wedge \exists C_2 \supset X_2$ 이면

관계 (C_1, C_2) $\wedge \exists S_1 \supset (U_{11} \subset U_{12} \subset \dots \subset U_{1n}) \supset$

$C_1 \wedge \exists S_1 \wedge (U_{11} \subset U_{12} \subset \dots \subset U_{1n}) \supset C_2$

걸음 4. \exists 관계 (C_1, C_2) $\wedge \exists C_1 \supset X_1 \wedge \exists C_2 \supset X_2 \wedge S_1 \supset X_1 \wedge S_2 \supset X_2 \wedge$

$\exists (U_{11} \subset U_{12} \subset \dots \subset U_{1n}) \subset C_1 \wedge \exists (U_{21} \subset \dots \subset U_{2n}) \supset C_2$

여기서 U_{ij} 는 U_{ij+1} 의 하위클래스이다.

걸음 5. 마디표식에 기초하여 G_1 과 G_2 의 사림 G_3 을 구한다.

걸음 6. 유사도를 비교하는데 리용되는 특징들을 포함하는 벡토르들을 A 와 B 라고 하고 다음과 같이 계산한다.

Concept_similarity=0 공통개념이 없다.

Concept_similarity=1 개념모임일치, 아니면

$$\text{Concept_similarity} = \text{sim_dice}(A, B) = 2 \times \sum_{i=1}^n a_i b_i / \sum_{i=1}^n a_i^2 + \sum_{i=1}^n b_i^2$$

벡토르 A 와 B 는 각각 그래프 G_1 과 G_3 의 개념마디들의 수로 채워진다.

걸음 7. Relation_similarity=di=sim_dice(A, B)= $2 \times \sum_{i=1}^n a_i b_i / \sum_{i=1}^n a_i^2 + \sum_{i=1}^n b_i^2$ 을 계산

벡토르 A 와 B 는 각각 그래프 G_2 와 그래프 G_1 의 린접호들의 수로 채워진다.

결음 8. $\text{maximum}(\text{di}) \quad i=\overline{1, n}$

알고리즘은 사용자질문 G_1 로 그래프들을 만들고 관련있는 마디 G_2 를 포함하는 온톨로지에서 토막을 얻어 G_1 과 G_2 사이의 공통부분(G_3)을 구축한다.

결음 2는 질문을 위한 그래프 G_1 의 구축을 서술한다. 이것은 주어(X_1), 관계, 보어(X_2)와 X_1 과 X_2 의 가장 대표적인 클래스들을 리용하여 창조된다.

결음 3은 사용자질문으로부터 기초원소들과 개념들을 포함하는 하위계층을 발견하는 과정을 서술한다.(즉 기초원소들을 포함하는 린접)

결음 6은 Dice결수를 리용하여 G_1 과 G_3 사이의 유사도를 계산한다.

벡토르 A 와 B 는 각각 그래프 G_1 과 G_3 의 개념마디수로 채워진다.

결음 7은 Dice결수를 리용하여 G_1 과 G_2 사이의 유사도를 계산한다.

벡토르 A 와 B 는 각각 그래프 G_2 와 G_1 의 린접호들의 수로 채워진다.

SimilarityTop알고리즘에 대하여 보기로 하자. 이 알고리즘은 위에서 정의한 SimilarityBase알고리즘, WordNet 그리고 사용자반결합을 리용한다.

SimilarityTop알고리즘은 질문에서 관계, 개념과 선택된 온톨로지에서 관계, 개념 사이에 유사성이 있는가를 검사한다. 만일 유사성이 없다면 그것은 WordNet씨소라스에서 찾은 모든 정보를 사용자에게 제공한다. SimilarityTop알고리즘은 선택된 정보를 가지고 론리 규칙을 다시 리용한다.

기본단계는 다음과 같다.

SimilarityTop procedure:

- SimilarityBase알고리즘을 호출
- If ontological_relation $\neq \theta$ then

평가_질문(ontological_relation(β)(β_1, β_2))

- Else

WordNet씨소라스를 리용하여 관계_질문을 위한 동의어모임을 얻는다.

사용자에게 WordNet씨소라스가 제공하는 정보들을 선택하도록 요구한다.

평가_질문(선택된_정보(β_1, β_2))호출

AQUA상에서 클래스들의 린접을 리용하여 사용자질문과 온톨로지적인 관계사이의 사영을 수행하는 유사도알고리즘을 실현하였다.

맺는 말

온톨로지적인 공간에서 매 개별적인 온톨로지에 대한 서술들을 포함하는 메타-온톨로지와 유사도알고리즘들을 리용하여 자료통합을 자동적으로 실현하는 한가지 방법을 제안하였다.

참고 문헌

- [1] Anjali Saini, P. K.Y adav; International Journal of Engineering and Computer Science, 6, 3, 20453, 2017.

- [2] Marta Tatu et al.; A Semantic Question Answering Framework for Large Data Sets, Open Journal of Semantic Web, 3, 1, 16, 2016.

주제 107(2018)년 11월 5일 원고접수

A Method of Knowledge-Based Ontology Data Integration for QAS

Choe Myong Ok

In this paper we proposed a multiple ontology data integration system for the question-answering framework called AQUA. This method used Meta-Ontology and the similarity algorithm based on analysis of neighboring classes.

Key words: ontology, question answering, data integration