

온톨로지 구조를 활용한 키워드 기반 질의 변환 (Keyword-Based Query Translation using Ontology Structure)

송 현 제 [†] 노 태 길 ^{**}
(Hyun-Je Song) (Tae-Gil Noh)

박 성 배 ^{***} 박 세 영 ^{***}
(Seong-Bae Park) (Se-Young Park)

요 약 질의 처리는 사용자가 입력한 질의를 분석하여 시스템이 이해할 수 있는 질의로 변환하는 작업으로, 온톨로지 및 시맨틱 웹 검색을 위한 질의 처리가 최근 주목을 받고 있다. 시맨틱 웹에서 사용하는 형식 질의 언어는 찾는 대상을 정확히 표현할 수 있지만, 키워드 기반과 달리 온톨로지의 구조 파악이 선행되어야 한다. 본 논문에서는 일반 사용자에게 익숙한 키워드로 작성된 질의문을 시맨틱 웹 환경에서 적합한 형식 질의문으로 변환하기 위한 키워드 기반 질의 변환 시스템을 제안한다. 제안한 방법은 온톨로지 구조 지식을 활용하여 개별 키워드가 어떤 제약 조건 또는 제약 대상과 결합하는지를 명시함으로써, 질의 키워드로부터 등가의 조건에 해당하는 형식 질의문을 생성할 수 있다.

키워드 : 질의 변환, 시맨틱 웹, 온톨로지, 구조 정보

Abstract This paper proposes a keyword-based query translation system for the semantic web. With the

· 본 연구는 한국과학재단의 "지역대학우수과학자지원사업(KRF-2005-202-D00465)"에 의하여 지원받았음

· 이 논문은 2009 한국컴퓨터종합학술대회에서 '온톨로지 구조 지식을 활용한 키워드 기반 질의 변환 시스템'의 제목으로 발표된 논문을 확장한 것임

[†] 학생회원 : 경북대학교 전자전기컴퓨터공학
hjsong@sejong.knu.ac.kr

^{**} 비 회 원 : 경북대학교 컴퓨터공학
tgnoh@sejong.knu.ac.kr

^{***} 증신회원 : 경북대학교 컴퓨터공학 교수
sbpark@sejong.knu.ac.kr
(Corresponding author)
syпарк@sejong.knu.ac.kr

논문접수 : 2009년 8월 14일

심사완료 : 2009년 10월 14일

Copyright©2009 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 컴퓨팅의 실제 및 래터 제15권 제12호(2009.12)

relationship between keywords and ontology structure information, the system converts keyword based queries into queries written by formal query language which is appropriate for the semantic web. As a result, casual web users could not only express queries easily but also obtain the better result.

Key words : Query translation, Semantic web, Ontology, Structure information

1. 서 론

시맨틱 웹은 온톨로지라는 지식 표현을 웹 단위로 공유함으로써 이루어지는 데이터의 웹이다. 최근 시맨틱 웹 기술의 발전과 함께 시맨틱 웹을 위한 질의 처리가 새로이 관심을 받고 있다. 기존 키워드 기반 질의 처리는 시맨틱 웹 환경에서 질의할 수 있는 다양한 의미적 질의들을 분석하기에 한계를 가진다. 예를 들어, "1980년대 이후에 태어난 블로거들이 주로 듣는 음악 그룹을 알려주세요,"라는 질의문의 경우, 단순히 키워드 매칭 방법으로 질의문이 내포하고 있는 복잡한 의미적 조건을 정확히 표현할 수 없다.

의미적 제약 조건을 가지는 질의를 처리하기 위해서, 시맨틱 웹 연구자들은 다양한 질의 입력 방식을 연구해 왔다. 시맨틱 웹의 표준 질의 언어는 SPARQL로서, 형식화된 시맨틱 질의어를 정의하고 있으며, 형식 언어를 사용하여 질의하는 방법은 다수의 시맨틱 웹 어플리케이션[1,2]들이 표준적으로 제공하고 있다. 이러한 형식 질의 언어는 찾고자 하는 정보를 형식화하여 정확하게 표현할 수 있지만, 형식 언어로 질의하기 위해서 사용자는 형식 언어의 문법을 미리 알고 있어야 할 뿐 아니라, 지식 기반인 온톨로지의 구조를 함께 파악하고 있어야 한다는 제약이 따른다. 따라서 다수의 일반 웹 사용자들이 형식 언어를 사용하여 질의하는 것을 기대하기는 어렵다.

이에 사용자에게 친숙한 자연어를 시맨틱 질의 입력에 활용하고자 하는 연구들이 최근 진행되고 있다. Tablan[3]는 통제 자연어(Controlled Natural Language)를 이용한 질의문 작성 시스템을 제안하였다. 통제 자연어는 자연어의 형태를 가지고 있어 읽기는 쉬우나 논리 언어의 엄밀한 규칙을 가지고 표현해야 하는 언어이므로, 이를 사용하기 위해서는 논리 언어를 배우는 것과 같은 정도의 노력이 요구된다는 단점이 있다.

일반 사용자들이 쉽게 사용하는 키워드에 기반을 둔 온톨로지 검색방법으로 Ding[4]은 온톨로지로부터 추출된 키워드들을 사용하여 검색을 수행하였다. 하지만 질의문에 사용되는 키워드들은 단순히 클래스의 레이블 또는 이름에서 추출된 것이며, 온톨로지의 구조를 반영하지 않아 정확한 결과를 얻기가 힘들다. Zhou[5]는 Ding[4]과 비슷한 방법이지만 사용자로부터 입력받은 키워드

들 사이에 존재할 수 있는 구조를 온톨로지 내부에서 찾은 후, 사용자로 하여금 그 중에 가장 적절한 것을 선택하도록 하는 방법을 제안하였다.

본 논문에서는 온톨로지에 대한 키워드 기반 질의 변환 시스템을 제안한다. 온톨로지의 구조를 활용하여 사용자로부터 자연어 키워드와 제약사항을 입력받아, 이를 형식 질의 언어로 작성된 시맨틱 질의문으로 생성한다. 사용자들이 입력할 수 있는 어휘들은 키워드-형식어휘 사전에 등재되어 있으며, 이 사전과 온톨로지의 구조를 활용하여 키워드들을 형식 질의 언어로 변환하는 방법이다. 따라서 키워드 기반에 익숙한 일반 사용자들은 시맨틱 질의어의 문법과 온톨로지 구조를 모르더라도 제안된 방법을 통해 보다 쉽게 시맨틱 웹 환경에서 질의문을 작성할 수 있다.

2. 온톨로지의 구조를 활용한 키워드 제약

본 논문에서 실험 대상으로 삼고 있는 온톨로지는 이벤트 온톨로지이다(자세한 구조는 3장 참조). 이벤트 온톨로지에 표현된 다양한 이벤트들은 일반적으로 유하원칙에 의해 표현이 가능하다. 시간(언제), 주체(누가), 대상(무엇을) 등 각 요소를 지칭하는 자연어 키워드는 온톨로지 구조상에서 다른 요소와 중복이 발생하지 않는다. 예를 들어, 주체로 파악된 키워드 '20대 남성'은 유하원칙의 다른 요소인 시간 또는 대상에서 사용되는 질의 키워드와 중복이 발생하지 않고, 온톨로지 구조에서 대상으로 인지된 '2008년에 출판된 책'은 시간이나 주체 요소로 파악되지 않는다.

이벤트 온톨로지에서 사용하는 이벤트 구조는 기본적으로 3가지 요소(주체-이벤트-객체)로 이루어진다. 이벤트를 구성하는 단어들은 일부 제약만으로 이벤트 구조상에서 어느 위치에 들어가는지를 찾을 수 있다. 이와 같이 온톨로지의 특징적인 구조를 활용하면, 자연어 키워드가 제약할 대상이나 의미가 미리 제약되므로, 일정한 단어의 조합으로 적합한 조건을 표현하는 것이 가능하다.

본 논문에서는 사용자가 입력한 질의 키워드를 시맨틱 질의문으로 변환할 때, 각각의 질의 키워드들은 온톨로지 구조상에서 어떤 한 부분을 제약하는 제약 사항으로 식별된다. 즉, 개별 키워드는 이벤트 구조의 3가지 카테고리 중 하나에 대한 제약사항을 가진다. 예를 들어, '20대', '학생' 등 특정 경험의 주체가 되는 단어들은 이벤트 구조에서 주체에 대한 제약사항과 함께 사용된다. '도서관' 또는 '한국소설' 등 경험의 대상이 되는 단어들은 이벤트 구조에서 객체로 사용된다. 이 외에도 이벤트에 대한 제약사항으로서 사용되는 단어에는 '작년', '지난여름' 등이 있다. 따라서 이러한 키워드들의 제약 사항을 통해 입력된 키워드는 온톨로지 구조와 매칭될 수 있으며, 이를 바탕으로 형식 질의어의 변환을 수행한다.

3. 이벤트 온톨로지

본 논문에서 대상으로 하는 이벤트 온톨로지의 구조는 그림 1과 같다.

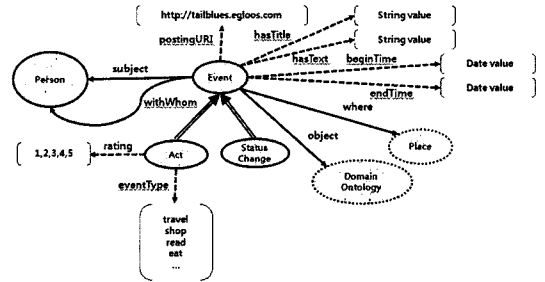


그림 1 이벤트 온톨로지의 구조

이벤트 온톨로지는 도서, 여행, 쇼핑 등 웹 상에 등장하는 실제계의 다양한 경험들을 수집, 기록할 수 있도록 디자인되어 있다. 경험의 주체가 되는 사람은 FOAF 온톨로지[6]를 활용하여 구축하였다. 현재, 경험의 대상을 표현하는 도메인 온톨로지는 총 3개의 분야(책, IT, 여행지)를 다루고 있으며, 각 분야를 하나의 온톨로지로 구축하였다. 책 온톨로지는 C. Bizer[7]의 온톨로지를 간략히 수정하여 사용하였으며, IT 온톨로지의 경우, [8]을 사용하였다. 또한 여행지 온톨로지는 GeoNames[9] 온톨로지 에서 일부 속성을 추가 및 수정을 하여 사용하였다.

본 논문에서 사용한 도메인 온톨로지는 복잡한 구조를 가지지 않는다. 예를 들어, '삼성전자 제품에 대한 사람들의 만족도' 또는 '여름에는 해운대에 사람이 많다.' 등의 대상에 대한 선호도 및 경향과 같은 복잡한 지식을 가지고 있지는 않다. 이들은 개별 인스턴스에 대한 최소한의 속성만을 담고 있으며, 경험의 대상을 식별하는 것이 주된 목적이다.

4. 키워드 기반 질의 변환 시스템

시맨틱 웹은 기존의 웹과 달리 온톨로지라는 지식 기반을 바탕으로 대상들간 다양한 의미 관계를 정의 및 표현하고 있다. 이에 질의문을 작성하기 위해서는 표준으로 정의된 형식 질의 언어를 사용한다. 예로, 아래는 시맨틱 웹 어플리케이션인 DBPedia¹⁾에서 "1970년 이후에 서울에서 태어난 축구선수와 그 선수가 현재 소속된 클럽을 찾아주세요"에 관하여 작성한 질의문이다.

위의 질의문을 작성하기 위해서는 사용자는 형식 질의 언어의 문법과 대상 온톨로지의 구조를 사전에 알고 있어야 한다. 따라서 키워드 기반 질의 입력에 익숙한

1) <http://dbpedia.org/snorql/>

```
PREFIX dbo: <http://dbpedia.org/ontology/>
SELECT ?person ?club WHERE {
  ?person dbpedia2:birthPlace <http://dbpedia.org/resource/Seoul>.
  ?person dbo:birthdate ?birth
  FILTER (?birth > "1970-01-01"^^xsd:date) .
  ?person rdf:type <http://dbpedia.org/ontology/FootballPlayer> .
  ?person dbpedia2:currentclub ?club . }
```

다수의 일반 사용자들은 형식 질의 언어로 질의문을 작성하기 어렵다. 본 논문에서는 시맨틱 웹 환경에서 일반 사용자들에게 익숙한 키워드로 작성된 질의문을 형식 질의 언어의 질의문으로 변환하려 한다.

제안하는 방법은 온톨로지의 레이블이나 Literal과 별개로, 사용자가 입력하는 자연어 키워드와 온톨로지의 관계 및 인스턴스를 연결할 사전의 역할을 하는 언어자료를 기반으로 동작한다. 이를 위해 온톨로지로부터 키워드로 사용할 수 있는 대상을 찾아 리소스로 구축하는 단계가 선행되어야 하며, 그러한 기반 위에서 사용자가 입력한 키워드와 제약 표기를 기반으로 지정된 표기법으로 입력 후, 시맨틱 질의문으로 변환한다.

4.1 키워드 테이블

시맨틱 질의에서는 단순 일치 방법을 통해 질의를 처리하는 것은 불가능하다. 예를 들어, “2000년 이후 출생한”과 같은 질의는 단순히 특정 스트링에 매치되는 것이 아니라, 질의의 조건에 맞는, 즉 “생년월일”에 해당하는 속성이 2000년 이후의 값을 가지는 모든 대상을 찾아 주어야 한다.

본 시스템에서는 이러한 질의문을 처리하기 위해서, 주어진 온톨로지서 처리 가능한 키워드들을 먼저 추출하여 그 키워드들과 시맨틱 형식 질의어로 동일한 의미가 되는 형식 질의 어구들을 미리 작성하여 테이블로 구축하며, 추후 이 테이블에 따라 입력된 질의를 변환하는 방법을 택하고 있다. 이 테이블을 키워드 테이블이라 부르며, 표 1은 키워드 테이블에서 추출한 일부 항목을 보여준다.

키워드 테이블은 키워드, 질의문, 역할표지의 총 3부분으로 구성된다. 키워드 부분은 사용자가 질의로 사용하는 일반적인 키워드 스트링으로, 이후 질의문에서 매치할 대상이기도 하다. 기본적으로 질의 키워드의 선정은 시스템 구축 과정에서 이루어진다. 키워드들은 해당 온톨로지의 인스턴스가 가지는 값, 속성, 클래스의 레이

표 1 키워드 테이블

키워드	질의문	역할표지
30대	?X foaf:birthday ?Y FILTER { 'current-date' - ?Y > 30 && 'current-date' - ?Y < 40 }	S
베스트셀러	?X book:label 'BestSeller'	O
2008년 이후	?X event:eventTime ?Y FILTER { ?Y > 2008-01-01 }	E
남성	?X foaf:title 'Mr'	S
핑크	?X IT:a_Color '핑크'	O
인기 있는	?X event:rating ?rating FILTER { ?rating >=4 }	E

블, 클래스 이름에서 추출되며, 그 중 조사 및 특수기호 등은 제거한다. 뿐만 아니라 일반 사용자들이 시스템에 질의할 때 사용할 수 있는 새로운 키워드들을 수집, 선정하는 과정을 거쳐 함께 키워드 테이블로 구축한다. 본 논문에서 실험 대상이 되는 이벤트 온톨로지로부터 약 500개의 질의 키워드를 선정하였으며, 표 2는 각 역할표지에 따른 키워드 목록의 일부분이다.

질의문 부분은 키워드에 해당하는 형식 질의 어구를 의미하며 형식 질의 언어인 SPARQL로 작성한다. 즉, 시스템은 사용자로부터 질의 키워드를 입력받아, 키워드 테이블의 질의문 부분을 참고하여 최종 시맨틱 질의문을 생성한다. 자세한 방법은 4.2절에서 설명하도록 한다. 각 키워드에 대응하는 형식 질의 어구는 키워드 추출과 함께 미리 작성한다.

키워드 ‘20대’, ‘남성’ 등 특정 경험의 주체가 되는 단어들은 앞서 언급된 이벤트 구조에서 주체에 대한 제약 사항으로 식별될 수 있다. 객체에 대한 단어, ‘베스트셀러’, ‘도서관’ 등은 이벤트 구조에서 객체 부분에 사용된다. 이처럼 이벤트 온톨로지의 구조에 기반하여 키워드들의 역할은 제약될 수 있으며, 본 논문에서는 이러한 제약사항을 주체는 ‘S’, 객체는 ‘O’, 이벤트는 ‘E’로 나타내어, 키워드 테이블의 역할표지 부분에 함께 표시한다. 따라서 사용자 질의를 형식 질의문으로 변환 시, 각 키워드의 역할표지 부분에 적힌 제약을 살펴봄으로써 질의 의미에 적합한 제약 조건을 표현할 수 있다.

일반 키워드와 달리 베스트셀러, 일러어택터, SF매니아 등 추상적인 개념을 표현하는 특정 키워드들은 기본적으로 온톨로지 구조만으로 나타낼 수 없다. 그러나 다량의 이벤트가 축적되면, 이들로부터 추상적인 개념을 추론할 수 있게 된다. 가령 실험 대상인 이벤트 온톨로지에는 책 4만권이 대해서 5년간의 37만 건의 포스팅 데이터가 포함되어 있다. 이 온톨로지로부터 “2007년에 많은 사람들이 논했던 책 10선”이나, “지난 5년간 꾸준히 인기가 있었던 책 100권” 등을 선정할 수 있으며, 선정된 대상에 이들을 부가적인 속성으로 부착하는 것이 가능하다. 본 논문에서는 규칙 기반 추론으로 개별 인스턴스들에 대한 추상적인 개념들을 추가 속성(slabel)으로 부가하였다. 추상적인 개념을 사용함으로써, 사용자는 쉽게 온톨로지에 질의할 수 있으며, 의미 있는 결과를 얻을 수 있다.

표 2 키워드 목록 예

역할표지	키워드
S	남성, 여성, 애플매니아, 스티븐 킹 팬, 20대
O	MP3, 테마공원, 1월의 Best 여행지, 스테디셀러, 삼성전자기기, 제목:해리포터와 마법사의 돌, 제품명:iPhone3G, 한국소셜
E	지난여름, 올해, 인기있는, 선호하는, 사계절

4.2 질의 변환 방법 및 표기법

앞서 구축된 키워드 테이블을 바탕으로 사용자가 입력한 키워드들을 하나의 최종적인 시맨틱 질의문으로 만들기 위해서는 각 키워드에 해당하는 형식 질의 어구를 질의 의도에 맞게 조합하여야 한다. 하지만, 단순히 형식 질의 어구들을 조합할 경우, 사용자가 의도한 대상을 반영하지 못해, 정확한 시맨틱 질의문을 만들 수 없다. 또한, 입력한 질의문에서 키워드 간의 의존 관계가 존재하는 경우, 의존 관계를 반영하여 시맨틱 질의문을 작성하여야 한다. 따라서 본 장에서는 사용자가 입력한 키워드 기반 질의문을 시맨틱 질의문으로 변환하는 방법과 다수의 이벤트를 가진 질의 의도도 처리할 수 있는 표기법을 설명한다.

기본적으로 사용자가 입력한 키워드를 질의 의도에 부합하는 시맨틱 질의문으로 변환하기 위해 표 3의 표기법으로 표기된 질의문을 입력받아 질의 변환을 수행한다.

표 3 질의 작성 표기법

```

<query> - ( <event> ) <symbol>
          | <keywords> <symbol>
<event> - ( <event> ) <symbol> <keywords>
          | <keywords>
<keywords> - <keywords> <keyword>
              | <keyword>
<keyword> - [ keyword ]
<symbol> - S | E | O
  
```

질의문은 키워드로 구성되며, 각 키워드는 사용자로부터 종괄호로 표시된다. 키워드를 명시적으로 표시함으로써 띄어쓰기 문제를 쉽게 해결할 수 있다. 다음의 질의 의도를 살펴보자.

* “20대 남성이 놀러가는 여행지는?”

위 질의 의도는 20대, 남성, 놀러가는, 여행지 총 4개의 키워드로 표현되며, 찾는 대상은 여행지라는 것을 나타낸다. 사용자는 키워드와 찾는 대상을 명시하여 아래와 같이 질의문을 입력한다.

* [20대][남성][놀러가는][여행지]O

시스템에서는 입력된 질의문을 표 4의 함수를 통해 하나의 완성된 시맨틱 질의문으로 변환한다.

사용자로부터 괄호로 표시된 질의문과 찾는 대상을 입력받아, 키워드 테이블의 질의문 부분과 역할 부분을 참조하여 형식 질의 어구를 조합하고, 찾는 대상을 SELECT문에 명시하여 하나의 시맨틱 질의문을 생성한다.

사용자가 입력하는 질의 의도에는 단일 이벤트 구조로 표현할 수 있는 질의 의도와 다수의 이벤트로 구성된 질의 의도로 나눌 수 있다. 위 예제의 질의 의도 경우, 하나의 이벤트 구조인 주체(20대 남성)-객체(여행지)-이벤트(놀러가는)로 표현이 가능하다. 찾고자 하는

표 4 시맨틱 질의문 변환 함수

```

Convert (Keywords, Target )
Query = SELECT Target WHERE
K = Keywords
Until ki is not null, Do
    Q = GetSPARQL( ki )      - *1
    D = GetRole( ki )      - *2
    Query.D += Q
return Query
*1 GetSPARQL( ki ) - return ki's SPARQL
*2 GetRole( ki ) - return ki's role
  
```

질의 의도가 하나의 이벤트 또는 이벤트의 일부만으로 표현되는 단일 이벤트 구조의 경우, 입력된 키워드 간의 의존 관계 분석 없이 쉽게 질의 변환을 수행할 수 있다.

하지만, 아래의 예와 같이 다수의 이벤트 구조로 이루어진 경우에는, 단일 이벤트와 달리 이벤트간의 의존 관계를 고려하여 질의 변환을 수행하여야 한다.

* “나의 문화유산 답사기 책을 읽은 사람들이 놀러가는 여행지는?”

위 질의 의도는 ‘나의 문화유산 답사기’ 책을 읽은 사람과 사람들이 놀러 가는 여행지라는 복수의 이벤트를 가지며, 두 이벤트가 동일 주체를 공유하는 이벤트 간 의존 관계가 존재한다. 의존관계를 가진 다수의 이벤트로 구성된 질의 의도인 경우, 각 이벤트들을 독립된 단일 이벤트로 표현하면, 사용자 의도에 부합하는 시맨틱 질의문으로 질의 변환을 할 수가 없다.

본 시스템에서는 이러한 의존 관계를 표현하기 위해 표 3과 같이 괄호 표기법을 도입한다. 사용자는 질의문 입력 시, 이벤트 간의 의존관계를 괄호로 묶고 이벤트 간 공유하는 대상을 명시적으로 표시함으로써 이벤트 간의 의존 관계 분석으로 발생하는 모호성을 제거할 수 있다. 따라서 시스템은 사용자의 질의 의도가 정확히 반영된 질의문을 입력받으며, 공유 대상과 키워드 테이블의 역할표지와 함께 질의 변환을 수행한다. 앞선 예제의 질의 의도는 제안한 표기법에 따라 다음과 같이 작성된다.

* (((제목: 나의 문화유산 답사기) (읽은) (사람)) S [놀러 가는] (여행지)) O

괄호를 이용하여 이벤트 간 의존 관계를 나타내고, 이벤트 사이에 존재하는 역할표지 S를 통해 이벤트 들이 공유하는 대상이 주체임을 표현할 수 있다. 아래는 “나의 문화유산 답사기” 책을 읽은 사람들이 놀러가는 여행지는?”라는 질의 의도에 대해 시스템이 최종적으로 생성한 시맨틱 질의문의 예이다.

```

SELECT DISTINCT ?Object1 WHERE {
?Subject0 rdf:type 'People' .
?Object0 rdf:type 'Book' .
?Object0 book:title '나의 문화유산 답사기'
?Event0 rdf:type 'Event' .
?Event0 event:subject ?Subject0 .
?Event0 event:object ?Object0 .
  
```

```

?Event0 event:type 'Review' .
?Object1 rdf:type 'Travel' .
?Event1 rdf:type 'Event' .
?Event1 event:subject ?Subject0 .
?Event1 event:object ?Object1 .
?Event1 event:type 'Travel-review' .
    
```

5. 실험

본 논문에서 제안한 키워드 기반 질의 변환을 평가하기 위해 검색을 수행하여 성능을 비교하였다. 실험에 사용한 온톨로지는 상용 블로그의 포스팅에서 이벤트를 자동, 반자동으로 추출하여 구축된 이벤트 온톨로지이다. 해당 온톨로지는 책, 여행지와 IT기기의 쇼핑에 관련하여 394,392개의 이벤트 인스턴스와 이들 이벤트의 주체가 되는 48,208명의 사람 정보를 포함하고 있다.

실험을 위해 이벤트 온톨로지에 적합한 가상의 질의문을 만들어 평가에 사용하였다. 표 5는 실험에 사용한 질의문으로, '20대 여성'과 같이 온톨로지 구조 속성 정보를 찾는 검색 질의와 질의문 6, 7, 8번과 같이 의존성이 있는 다수의 이벤트 검색 질의까지 모두를 포함할 수 있도록 작성하였다.

표 5 실험에 사용한 질의 의도 및 질의문

번호	질의 의도
1	20대 여성은?
2	한국 소설인 책은?
3	최근 1년 사이 출시된 노트북은?
4	온다 리쿠가 쓴 책을 읽은 사람은?
5	여성이 구매한 핑크색 mp3는?
6	애플 제품을 좋아하는 사람들이 구매하는 삼성전자 기기는?
7	40대가 읽은 책을 읽은 20대 여성은?
8	나의 문화유산답사기 1을 읽은 사람이 놀러가는 여행지는?

이벤트 온톨로지를 잘 알고 있는 전문가가 각 질의문에 대한 시맨틱 질의어를 작성하여 얻은 결과를 정답으로 간주하였다. 베이스 라인은 온톨로지 항목들의 레이블 값과 키워드를 매치하는 방법으로 설정하였다. 그림 2는 각 질의에 대한 결과를 보여준다.

실험 결과, 베이스 라인 방법(방법1)은 질의문 1, 2와 같이 온톨로지 속성에서 검색할 수 있는 정보에 대해서는 약간의 결과를 출력하였지만 키워드 간의 의미 있는 관계를 포함한 질의 의도에 대해서는 검색 결과를 출력하지 못하였다.

본 논문에서 제안하는 방법(방법2)은 전문가가 작성한 SPARQL과 등가의 결과를 보여주었다. 이는 정의된 키워드 테이블이 온톨로지 구조를 적절히 반영하며, 이벤트 간의 의존관계를 사용자로부터 입력받아 질의 의도를 정확히 분석할 수 있었기 때문으로 보인다. 특히 질의문 6, 7, 8과 같이 다수 이벤트를 포함한 복잡한 질의에 대해서도 제안한 시스템은 유용한 결과를 보여주었다.

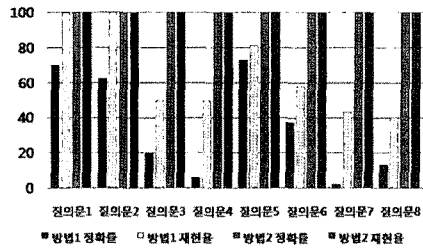


그림 2 실험 결과

6. 결론 및 향후 연구

본 논문에서는 시맨틱 웹 환경에서 일반 사용자들에게 익숙한 키워드 기반 질의 변환 시스템을 제안하였다. 본 시스템은 주체-이벤트-객체의 이벤트 구조를 가진 온톨로지를 활용하여 자연어 키워드 및 간단한 제약을 입력받아 그에 대응하는 시맨틱 질의문을 생성한다. 이때, 사용자가 질의어로 사용하는 자연어 기반 키워드의 후보들은 미리 온톨로지로부터 추출하여, 그에 해당하는 형식 질의 어구와 함께 키워드 테이블로 구축하였다. 식별된 질의 키워드들은 제약사항과 조합을 하여 하나의 완성된 시맨틱 질의문으로 생성된다. 제안한 질의 작성법을 통해 다수 이벤트 구조로 표현되는 질의 의도도 정확하게 시맨틱 질의문으로 변환할 수 있다.

참고 문헌

- [1] John G. Breslin, Andres Harth, Uldis Bojars and Stefan Decker, "Toward Semantically-Interlinked Online Communities," In *proceedings of The 2nd European Semantic Web Conference*, pp.500-514, 2005.
- [2] <http://wiki.dbpedia.org>
- [3] Valentin Tablan, Danica Damjanovic, and Kalina Bontcheva, "A Natural Language Query Interface to Structured Information," In *Proceedings of 5th European Semantic Web Conference*, pp.361-375, 2008.
- [4] Li Ding, Tim Finin, Anupam Joshi, Rong Pan, R. Scott Cost, Yun Peng, Pavan Reddivari, Vishal Doshi, Joel Sachs, "Swoogle: A Semantic Web Search and Metadata Engine," In *proceeding of the 13th ACM CIKM*, pp.652-659, 2004.
- [5] Qi Zhou, Cong Wang, Miao Xiong, Haofen Wang and Yong Yu, "SPARK: Adapting Keyword Query to Semantic Search," In *proceedings of 2nd Asian Semantic Web Conference*, pp.694-707, 2007.
- [6] <http://xmlns.com/foaf/spec/>
- [7] C. Bizer, R. Cyganiak and T. Gauss, "The rdf book mashup: from web APIs to a web of data," In *proceedings of the 3rd Workshop on Scripting for the Semantic Web, 4th ESWC*, pp.389-393, 2008.
- [8] <http://cscola.kaist.ac.kr/wiki/index.php>
- [9] <http://www.geonames.org/ontology>