

# 온톨로지내 효율적 정보 탐색을 위한 CQET 시스템 구현

공현장\*, 황명권\*\*, 김원필\*, 김관구\*\*

\*조선대학교 전자계산학과

\*\*조선대학교 컴퓨터공학과

e-mail:kisofire@chosun.ac.kr

## A Study on the CQET(Controlled Query Engine for Triple) System

Hyun-Jang Kong\*, Myung-Gwon Hwang\*\*, Won-Pil Kim\*, Pan-Koo Kim\*\*

\*Dept of Computer Science, Chosun University

\*\*Dept of Computer Engineering, Chosun University

### 요 약

온톨로지에 대한 연구가 활발해지면서, 온톨로지를 실제 응용하고자 하는 사용자들의 요구가 커지게 되었다. 그렇지만 온톨로지의 응용을 위해서는 구축된 온톨로지를 효율적으로 다루고, 평가할 수 있는 시스템이 필요하다. 본 논문에서는 기존 온톨로지 질의 언어들의 한계점 중 구문 작성시 복잡성을 해결하기 위해 단순한 텍스트 기반의 질의 처리 인터페이스를 구현하여 온톨로지 사용의 범용성을 높이고, 사용자는 쉬운 질의 처리 환경에서 온톨로지에 대한 평가를 용이하게 수행함으로써, 온톨로지의 높은 활용성 또한 기대할 수 있다. 결론적으로, 본 연구의 온톨로지 질의 처리의 용이성과 트리플 사실 기반의 추론을 통하여 사용자는 온톨로지 구축의 완전성 및 응용 가능성을 쉽게 평가할 수 있다.

### 1. 서론

온톨로지에 대한 연구와 이를 응용하고자 하는 분야들 사이에는 실제 온톨로지 활용을 위해 해결해야 할 많은 문제들이 존재하고 있다. 특히, 온톨로지내의 효율적 정보 탐색을 위한 질의 처리 부분은 온톨로지 응용의 측면에서 반드시 해결되어야 할 부분이다. 기존의 온톨로지에 대한 질의 처리 시스템은 온톨로지에 표현된 사실만을 분석하여, 단순 구조적 사실 정보만을 사용자에게 제공함으로써 데이터베이스 시스템 및 텍스트 기반 정보처리 시스템과 크게 다른 점이 없었다. 사실상, 기존의 온톨로지 질의 시스템들에서 사용하고 있는 질의 언어의 구조나 형식이 데이터베이스 시스템의 질의 언어(SQL)에 모태를 두고 있으므로, 온톨로지와 데이터베이스의 질의 처리 및 그 결과는 거의 동일하다. 이러한 질의 처리 환경은 온톨로지 사용의 효율성을 낮추는 요인이 된다. 이에 본 연구에서는 온톨로지 사용의 효율성을 극대화하기 위해 온톨로지에 대한 단순 질의 처리 뿐만 아니라, 추론 규칙에 기반한 추론된 사실들

을 모두 처리할 수 있는 온톨로지 추론 및 질의에 대한 통합 시스템 개발의 필요성을 인식하고, 온톨로지에 대한 최적의 질의 처리를 위한 CQET 시스템을 구현하였다.

본 시스템에서는 기존 온톨로지 질의 언어(RQL, RDQL, SPARQL)들이 가지고 있는 질의 구문 작성시 복잡성을 해결하기 위해 단순한 텍스트 기반의 질의 처리가 가능한 질의 인터페이스를 구현하여 온톨로지 사용의 범용성을 높이고자 한다. CQET를 사용하여 사용자는 텍스트 입력 부분에 자연어와 비슷한 질의를 작성하여 온톨로지에 대한 질의 처리가 가능하다. 또한, CQET는 온톨로지 언어(OWL, RDF(S))의 어휘들의 특성을 기반으로 정의한 추론 규칙을 사용하여 온톨로지를 트리플 사실들의 집합으로 재구성하여, 추론된 새로운 사실들을 질의 처리함으로써, 온톨로지 활용성 또한 높일 수 있다. 끝으로, 질의 처리의 용이성과 트리플 사실 기반의 추론을 통하여 사용자는 온톨로지 구축의 완전성 및 응용 가능성을 쉽게 평가할 수 있다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 기존의 온톨로지 질의 시스템들과 질의 언어들에 대한 관련 연구를 기술하고, 3장에서는 CQET 시스템의 내용에 대하여 자세하게 기술한다. 그리고 4장에서는 본 연구에 대한 평가를 기술하고, 끝으로 5장에서는 결론을 맺는다.

**2. 관련 연구**

온톨로지 질의를 위해서 많은 연구가 진행되어져 왔으며, 그 결과 RQL[2], RDQL[4], SeRQL[1]과 같은 질의 언어들이 개발되었다. 그리고 이러한 언어들의 구조는 데이터베이스 질의 언어인 SQL[3]에 모태를 두고 있다. RQL, RDQL, SeRQL은 RDF 그래프 모델에 기반하여 사용자 질의를 처리함으로써, OWL 파일을 처리하는데 한계를 가지고 있으며, 질의를 수행하기 위해 복잡한 질의 구조를 사용하여 질의를 작성해야 하는 어려움이 있다. 그리고 SQL은 데이터베이스의 일반화로 질의 언어의 사용이 보편화 및 구조화되어 있지만, 역시 데이터베이스를 처리하는데 적합한 언어로써 온톨로지를 처리하는데는 한계가 있다. 그리하여 최근의 연구에서는 온톨로지를 데이터베이스화 함으로써 데이터베이스로 변형된 내용에 대하여 SQL을 사용하여 질의를 처리하는 연구도 있었다. <표 1>은 기존 온톨로지 질의 언어와 본 질의 처리 방식의 특징을 비교하여 보여 주고 있다.

<표 1> 온톨로지 질의 언어 비교

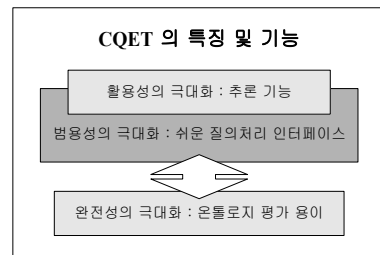
|      | RQL, RDQL, SeRQL                             | SQL   | CQET                                     |
|------|--|---|--|
| 처리영역 | RDF  | Database                                    | RDF, OWL                                 |
| 추론기능 | 없음   | 없음  | 있음                                       |
| 처리시간 | 느림   | 빠름  | 보통                                       |
| 처리량  | 보통   | 많음  | 보통                                       |
| 기본구조 | SELECT, FROM, WHERE                          |   |  |
| 특징   | RDF 그래프 모델에 기반한 텍스트 매칭의 온톨로지내의 간단한 구조적 정보 검색 | 온톨로지 검색을 위해서는 온톨로지 파일을 데이터베이스화하는 사전 작업이 요구됨 | 사용자중심의 텍스트기반 질의 인터페이스 및 추론과 질의의 통합시스템 제공 |

그리고, 이러한 온톨로지 언어들에 기반하여 온톨로지를 처리하기 위한 추론 엔진(Sesame, Jena, Jess, JTP)들이 개발되었다. 먼저 Sesame[8]은 RQL에 기반하고 있으며, 추론을 위한 추가의 기능을 포함하고 있지는 않다. 두 번째로 Jena는 RDQL에 기반한 추론 엔진이며, RDF(S), OWL, DAML+OIL로

작성되어진 모든 파일들에 대한 처리가 가능하다. 또, Jess[6]는 Rete 알고리즘에 기반하여 자바로 구현된 규칙 기반의 추론 엔진이다. 끝으로, JTP[5]는 객체 지향적 추론 시스템이며, 다른 시스템들에 비해 매우 간단하고 일반적인 추론 구조로 되어있다.

**3. CQET 시스템의 특징 및 구조**

서론에서 간단히 언급했듯이, CQET는 온톨로지에 대한 효율적 질의 처리를 지원하고 있으며, 세가지 특징을 가지고 있다. 첫 번째로 질의와 추론을 모두 처리할 수 있는 통합시스템적 특징을 가지고 있으며, 두 번째는 온톨로지 사용의 범용성을 높이기 위한 텍스트 기반의 질의 처리 인터페이스를 제공하고 있다. 세 번째 특징은 구축 또는 재사용하고자 하는 온톨로지에 대한 평가가 용이하다는 것이다.(그림 1)은 CQET의 특징을 설명하고 있다. 3.1절부터는 CQET의 구조 및 세가지 특징에 대하여 자세하게 소개한다.



(그림 1) CQET의 세가지 특징

**3.1 질의 및 추론의 통합 시스템**

기존 대부분의 온톨로지 질의 처리 시스템에서는 단순한 온톨로지 질의 언어에 기반하여 온톨로지내의 구조적 사실들에 대한 질의 결과를 사용자에게 제공하고 있다. 그렇지만, 온톨로지의 사용 목적인 추론에 대한 결과를 제공하지 못하여, 사용자에게 온톨로지 사용에 대해 어필하지 못하고 있다. 이에 반해, CQET는 트리플 사실들에 기반하여 질의 처리가 이루어지고 있으며, CQET에 사용되는 트리플 사실들은 온톨로지 언어들의 어휘들에 기반하여 추론 규칙을 사전에 정의하여 새롭게 추론된 의미적 사실들의 집합이다. 그리하여 트리플 사실들은 구조적 사실들과 의미적(추론된) 사실들로 구성되어진다. CQET는 먼저 온톨로지를 사전에 정의된 추론 규칙에 기반하여 트리플 사실들의 집합으로 재구성한 후, 트리플 처리에 적합한 질의 처리 인터페이스를 설계하여 두 가지의 주요한 기능으로 구성되어져 있다. CQET에서는 온톨로지 언어의 어휘들에 대한

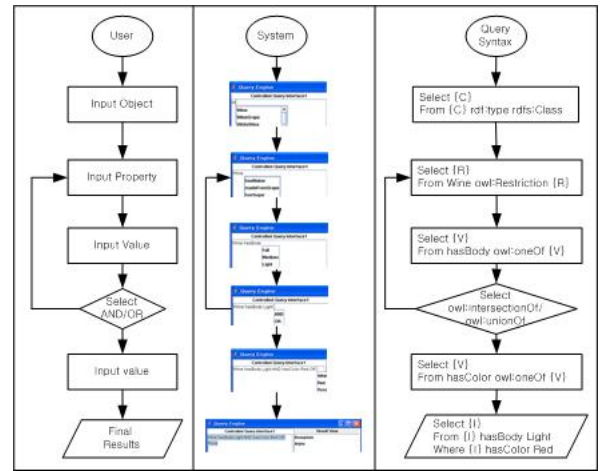
기본 그래프 모델 처리 규칙 30개와 어휘들에 대한 의미적 추론 규칙 20개를 각각 정의하고, 정의된 규칙들에 기반하여 온톨로지를 트리플 사실들로 변환하기 위한 온톨로지 파서 및 변환 모듈을 포함하고 있다.

이렇게 구현된 온톨로지 파서 및 변환기에 의해 새롭게 생성된 트리플 사실들의 집합을 사용하여 사용자가 쉽게 온톨로지에 대한 질의 처리를 할 수 있는 텍스트 기반의 질의 처리 인터페이스를 구현하였다. 3.2절에서 자세하게 소개한다.

### 3.2 텍스트 기반 온톨로지 질의 처리 인터페이스

CQET의 가장 핵심이 되는 목적은 온톨로지의 범용성을 높이는 것이다. 기존의 온톨로지 질의 처리를 위한 시스템 및 질의 언어들의 사용은 대부분 구조적 질의 형식(SELECT, FROM, WHERE)에 의존하여 사실상, 질의를 수행하는 사용자는 질의 언어의 구조적 사용 방법 및 질의 수행을 위해 온톨로지 내용을 사전에 파악하고 있어야 온톨로지에 대한 질의 처리가 가능하였다. 이러한 질의 처리 환경에서 온톨로지를 연구하는 대부분의 연구자들은 쉽게 이러한 질의 시스템을 다룰 수 있겠지만, 시맨틱 웹이 차세대 웹으로 자리매김하기 위해서는 이러한 연구자뿐만 아니라, 일반 사용자에게 쉽게 온톨로지가 사용될 수 있는 범용의 질의 처리 인터페이스가 존재해야 한다. 이에, 본 연구에서는 일반 사용자가 쉽게 온톨로지에 대한 질의를 수행할 수 있는 텍스트 기반의 질의 처리 인터페이스를 구현하고, 이를 활용하여 온톨로지 사용의 범용성을 높이고자 한다. CQET의 질의 처리 인터페이스는 3.1절에서 생성된 트리플 사실 집합에 기반하고 있으며, 사용자는 복잡한 질의 구문이나 온톨로지의 내용을 알지 못하더라도 쉽게 온톨로지 내용에 대한 질의 수행이 가능하다. CQET에서 사용자는 텍스트 창에 마치 자연어를 입력하는 것과 같은 질의 처리가 가능하며, 그 과정을 살펴보면 (그림 2)와 같다.

(그림 2)의 왼쪽 부분은 데이터의 흐름을 순서대로 표현하고 있으며, 중간 부분은 실제 CQET에서 데이터 흐름에 대한 실행 화면을 보여주고 있다. 그리고 오른쪽 부분은 각각의 질의 수행 절차에 대한 시스템 내부적으로 처리되는 모듈을 순서적으로 보여주고 있다.



(그림 2) CQET의 질의 처리 방식

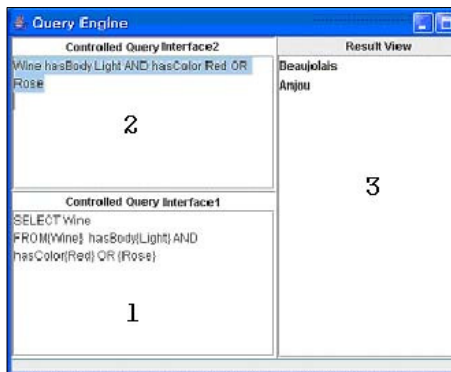
### 3.3 온톨로지 평가의 용이성

3.1절의 추론 기능과 3.2절의 텍스트 기반의 질의 처리 인터페이스에 기반하여 CQET 시스템의 사용에 의해 부수적으로 획득할 수 있는 효과중의 하나는 온톨로지 평가의 용이성이다. 현재 온톨로지에 관심을 가지고 있는 많은 사용자들이 온톨로지를 구축 또는 재사용하고자 하며, 온톨로지를 자신의 분야에 활용하고자 한다. 그렇지만 자신이 구축한 온톨로지가 완벽하게 구축되어 있는지, 자신의 온톨로지를 사용하여 어떠한 결과 값을 얻을 수 있으며, 자신이 온톨로지를 활용하여 획득하고자 했던 결과를 얻을 수 있는지에 대한 평가를 위해 많은 노력이 요구되어진다. 특히, 기존의 질의 처리 시스템들에서는 구축된 온톨로지에 대해 추론 과정을 거치고 이렇게 추론된 내용을 분석하고, 이를 처리할 수 있는 질의 언어를 사용하여 복잡한 질의 표현을 통해 온톨로지 평가를 수행해야 하는 상당히 어렵고 번거로운 작업이었다. 그리하여 사용자는 자신의 온톨로지에 대한 평가를 수행하지 않고 단순하게 구축의 완전성 및 유효성만을 검사하는 수준에 그치게 되어 온톨로지를 실제 응용하는 데는 그만큼의 한계가 존재하고 있다. 이에, CQET에서는 추론의 기능에서 질의의 수행까지 통합하여 처리할 수 있는 통합시스템을 제공함으로써 사용자는 CQET를 통하여 쉽게 자신의 온톨로지에 대한 평가가 가능하며, 특히 텍스트 기반의 쉬운 질의 처리 인터페이스는 온톨로지에 대한 사용자가 원하는 많은 질의를 단순하게 수행할 수 있어, 사용자는 온톨로지에 대한 다양한 적격 질의를 통한 질의 결과를 기반으로 온톨로지의 내용 수정이 가능하며 완벽한 온톨로지의 구축을 기대할 수 있고, 또한 실제 온톨로지를 응용하여 얻을

수 있는 결과값을 미리 추측할 수 있으므로 온톨로지의 산업 응용 측면에 크게 기여할 것으로 기대된다.

#### 4. 구현 및 평가

이러한 특징들을 포함하고 있는 CQET를 구현하여 테스트한 결과 온톨로지에 대한 질의 수행의 용이성을 확인할 수 있었다. 본 시스템의 평가를 위해 W3C에서 만든 와인 온톨로지(wine.rdf)를 실험 대상으로 삼고, 내용 평가를 수행하였다. 평가 내용을 살펴보면 다음과 같다. 먼저 본 시스템의 추론 규칙에 기반하여 총 1505개의 트리플 사실들이 생성되었으며, 1505개의 사실 트리플들은 온톨로지 언어의 구조적 특징에 기반한 구조적 사실 798개와 온톨로지 어휘의 규칙에 기반한 의미적 사실 707개로 구성되어 있다. 생성된 트리플들에 기반하여 텍스트 기반의 질의 처리 인터페이스에서 질의를 수행하는 실행화면은 (그림 3)과 같다.



(그림 3) CQET 실행 화면

CQET 실행 화면은 (그림 3)과 같다. (그림 3)의 내용은 평가를 위해 테스트한 질의 내용 “맛이 Light한 레드와인이나 로즈와인”에 대한 결과값을 보여주고 있다. CQET는 ②번과 같은 텍스트 형식의 자연어와 같은 질의 처리가 가능하며, ①번에서는 기존의 질의 처리 방식의 구조적 형식을 보여주고 있다. ①번과 ②번은 동일한 결과에 대한 질의 형식이다. 질의에 대하여 ③번 화면에서 결과값을 보여주며, 이에 기반하여 온톨로지의 평가가 용이하게 된다. 결론적으로, CQET를 통하여 온톨로지 처리의 통합화와 쉬운 질의 처리 환경을 제공함으로써 온톨로지 사용의 범용성을 높이고, 끝으로 온톨로지의 쉬운 평가를 통한 온톨로지에 대한 높은 활용성을 기대할 수 있다.

#### 5. 결론

본 연구는 쉬운 온톨로지의 운용을 위해 요구되는 몇가지 해결책에 대한 연구를 하였다. 먼저 기존 질의 언어들이 처리하지 못하는 추론 메커니즘을 질의와 통합하며 수행함으로써 온톨로지 구조를 파악하는 단순 정보검색 뿐만 아니라, 온톨로지에서 표현하고 있는 의미적 결과에 대해서도 질의 처리가 가능함을 보여주고 있다. 또한, 온톨로지 산업 및 시맨틱 웹 산업의 활성화에 발 맞추어 온톨로지 응용 및 범용성의 측면에서 살펴보면, 기존의 질의 언어들이 형식에 얽매인 구조적 어휘를 사용하여 질의를 수행하는 반면, CQET에서는 사용자 질의를 단순한 자연어 형식으로 작성하여 처리하는 것이 가능하여 온톨로지 사용의 범용성을 높일 것으로 기대된다. 끝으로, CQET를 사용하여 온톨로지를 평가하는 작업이 단순화됨으로써 사용자는 새롭게 구축된 혹은 재사용하고자 하는 온톨로지에 대하여 쉬운 질의 처리를 수행하고 질의 결과를 분석하여 온톨로지의 내용 및 응용 가능성의 평가가 용이하게 되며, 더욱 완벽한 온톨로지 구축을 기대할 수 있다.

#### 참고문헌

- [1] Broekstra, J. and Kampman, A. : The SeRQL Query Language User manual, Aduna. <http://sesame.sduna.biz/publications/SeRQLmanual.html>, 2003.
- [2] Broekstra, J., Kampman, A., and van Harmelen, F. : Sesame: A Generic Architecture for Storing and Querying RDF and RDF Schema. ISWC2002, page 54-68.
- [3] ISO : Information Technology-Database Language SQL. Standard No. ISO/IEC 9075, 1999.
- [4] Seaborne, A. : RDQL A Query Language for RDF. W3C member submission, <http://www.w3.org/Submission/2004/SUBM-RDQL-20040109>, 2004.
- [5] JTP, <http://www.ksl.stanford.edu/software/JTP>
- [6] Jess, <http://herzberg.ca.sandia.gov/jess>
- [7] Jena, <http://jena.sourceforge.net/index.html>
- [8] Sesame, <http://www.openrdf.org>