
사용자 질의 의미 해석을 위한 온톨로지 지식베이스 스키마 구축

도하나*, 이무훈**, 정훈***, 최의인****

Ontology Knowledge Base Scheme for User Query Semantic Interpretation

Hana Doh*, Moo-Hun Lee**, Hoon Jeong***, Eui-In Choi****

요약 최근의 검색 형태는 키워드 기반 검색에서 보다 더 정확한 결과를 제공하기 위한 시맨틱 검색 방법으로 변화하고 있다. 하지만 일반 사용자들은 여전히 기존의 키워드 기반 검색에 익숙하기 때문에 시맨틱 검색을 위한 형식화된 구조적 질의어를 구성할 수 없다. 따라서 본 논문에서는 이러한 사용자들의 키워드 해석을 위한 온톨로지 지식베이스 스키마를 제안한다. 제안된 스키마는 서술 논리 추론을 위해 OWL-DL을 기반으로 설계되었으며 최소한의 관계 정의와 추론 규칙 언어인 SWRL(Semantic Web Rule Language)으로 추론함으로써 보다 풍부한 개체 관계의 표현을 제공한다. 마지막으로 논문에서 제안한 사용자 질의 의미 해석 기법의 검증을 위해 키워드 유사도 실험 결과를 설명한다.

주제어 : 의미 기반, 시맨틱 검색, 온톨로지, 지식베이스, 질의 의미 해석

Abstract The method of recent information retrieval passes into an semantic search to provide more accurate results than keyword-based search. But in common user case, they are still accustomed to using existing keyword-based search. Hence they are hard to create a typed structured query language. In this paper, we propose to ontology knowledge-base scheme for query interpretation of these user. The proposed scheme was designed based on the OWL-DL for description logic reasoning, it can provide a richer representation of the relationship between the object by using SWRL(Semantic Web Rule Language). Finally, we are describe the experimental results of the similarity measurement for verification of a user query semantic interpretation.

Key Words : Semantic, Semantic Search, Ontology, Knowledge Base, Query Interpretation,

1. 서론

인류는 오랜 역사 속에서 지식을 표현, 공유하기 위해 많은 노력을 기울였다. 데이터와 정보의 생성 유통이 폭발적으로 증가됨에 따라, 인간이 수작업으로 그 안에 숨겨진 지식을 모두 처리하기에는 점점 더 불가능해지고 있다. 정보사회의 발달로 웹이 인류 지식의 보고가 됨에 따라, 웹에 있는 지식을 컴퓨터에 의해 생산적으로 분석, 발견, 공유, 활용하기 위한 방법을 찾아내는 것은 더욱 중

요해지고 있다. 따라서 사람만이 이해하는 웹이 아닌 컴퓨터가 읽고 처리할 수 있는 웹을 구현하기 위한 지식 표현 체계가 필요한 것이다. 특히, 정보 검색에서 사용자가 원하는 검색 결과를 제공하기 위해서는 사용자의 검색 의도를 정확히 파악해야 한다. 시맨틱 검색에서는 사용자가 입력한 모호한 키워드로부터 각 키워드의 개념적 의미와 키워드 간의 의미적 상관관계를 분석해야만 정확한 검색 결과를 사용자에게 제공할 수 있다[1][2][4]. 이

*한남대학교 컴퓨터공학과 석사과정

**한국전자통신연구원

***한남대학교 컴퓨터공학과 석사과정

****한남대학교 컴퓨터공학과 교수(교신저자)

논문접수: 2013년 1월 18일, 1차 수정을 거쳐, 심사완료: 2013년 2월 13일, 확정일: 2013년 3월 20일

를 위해 사용자가 입력한 검색 키워드, 사용자 선호 정보, 사용자의 검색 환경에 대한 상황정보 등의 정보를 활용한다면 사용자의 검색 의도를 보다 정확하게 파악할 수 있을 것이다[3][6][7]. 따라서 이러한 시맨틱 검색에서 가장 전제되어야 할 것은 잘 구축된 지식베이스이다. 지식베이스는 사용자 검색문 의미해석을 위한 메타데이터 간의 의미적 네트워크로 구성된 데이터베이스이다. 지식베이스 구축에 있어 모든 도메인에 대한 온톨로지 지식베이스를 구축하는 것은 거의 불가능한 일이지만, 특정 도메인의 온톨로지 구축은 지식 표현 언어 및 추론 기능의 발달로 최근 그 구축이 용이해졌다. 특히, 스키마 설계 단계에서 온톨로지 모델링 이론을 바탕으로 해당 도메인을 분석하여 스키마를 설계하고, 서술 논리 추론 기능을 적용함으로써 포함관계, 일관성, 동등관계, 멤버관계 등을 표현하는 온톨로지 지식베이스를 구축할 수 있다.

본 논문에서는 온톨로지 스키마 설계 단계에서 서술 논리 추론을 위해 OWL의 하위 언어 중 OWL-DL을 기반으로 한 스키마 설계를 제안한다. 또한, 최소한의 관계 정의와 추론 규칙 언어인 SWRL(Semantic Web Rule Language)을 정의하여 추론함으로써 보다 풍부한 개체 관계를 표현하도록 하였다.

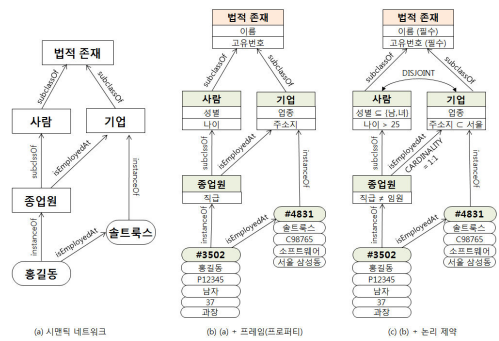
본 논문의 구성은 다음과 같다. 2장에서는 관련연구에 대해 소개하고, 3장에서는 제안한 사용자 질의 의미 해석 시스템과 이에 필요한 온톨로지 지식베이스 스키마를 설명한다. 4장에서는 제안 시스템의 성능 평가 및 분석에 대해 설명하고, 마지막으로 5장에서 결론을 맺는다.

2. 관련 연구

온톨로지 지식베이스는 검색의 대상인 멀티미디어 콘텐츠에 대한 메타데이터를 표현한 저장소이면서 사용자 검색문의 의미를 해석하기 위한 사전과 같은 논리적 지식표현 체계로서 논리, 규칙, 프레임 등의 개념을 모두 통합했다고 할 수 있다. 온톨로지에서 원하는 정보를 검색하기 위해서는 형식화된 구조 질의 언어인 SPARQL과 같은 형태의 질의를 수행하여야 하지만 일반 사용자들은 이러한 질의어를 구성할 수 없을 뿐만 아니라 온톨로지 자체에 대한 이해가 부족하다. 따라서 기존의 익숙한 검색에서 사용하는 질의어를 입력하고 온톨로지를 기반으로 키워드를 해석하여 구조적 질의어로 자동 변환해줌

으로써 시맨틱 검색을 손쉽게 이용할 수 있게 된다.

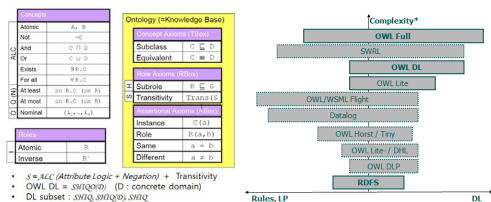
[그림 1] (a)의 시맨틱 네트워크는 지식을 구성하는 개념을 관계로 연결하고 있다. 관계는 두 개체를 연결하는 것이다. 여기에 프로퍼티를 추가하면 (b)와 같이 된다. 프로퍼티는 두 개체만을 연결한 경우이다. 마지막으로, 논리적 제약조건을 추가해서 (c)와 같이 만들면, 완전한 온톨로지 모양을 갖추게 되고, 여기에 규칙이 결합될 수 있다. 따라서 가능한 의미해석 결과는 다양한 형태로 존재할 수 있으며, 이러한 의미해석 결과를 탐색하는 곳이 온톨로지 지식베이스이다.



[그림 1] 온톨로지에 포함된 지식 표현들

이러한 온톨로지 개념을 컴퓨터가 처리하려면, 기계가 읽고 처리할 수 있는 언어로 기술되어야만 할 것이다. 온톨로지 표현 언어는 매우 다양하며 대부분 1차 논리에 기반하고 있다. 추론시 결정 가능성을 보장하는 서술 논리 기반의 언어들이 대세를 이루고 있다. 특히, OWL은 시맨틱 웹의 지식표현에 이용되고 있으며 OWL DL은 계산학적 완전성과 결정 가능성이 유지되는 최대의 표현력을 제공한다. 완전성은 모든 결론이 계산될 수 있다는 것이고, 결정 가능성은 모든 계산이 유한한 시간 내에 끝날 수 있다는 것을 의미한다[5].

[그림 2]는 OWL과 OWL DL의 표현력을 나타낸다.



[그림 2] OWL 및 OWL DL의 표현력

2.1 온톨로지 모델링 방법

온톨로지 모델링은 온톨로지 구축 생산성과 관리 가능한 품질 달성을 위해, 각 도메인에 적합한 표준 방법론을 정립하고 이에 기반 한 온톨로지 모델링을 수행하는 것이다. 실제 구축 방법론은 각 단위 프로세스를 반복적으로 사용한다.

온톨로지 구축의 첫 작업은 적용 범위와 응용 사례를 결정하는 것이다. 온톨로지가 표현할 지식과 도메인이 무엇이며, 그 응용 소프트웨어는 무엇이 될지, 온톨로지가 어떤 질문에 대한 답을 주어야할지를 결정한다. 온톨로지 모델링 시 활용 어플리케이션과 시나리오를 초기에 명확하게 만들수록 더 좋은 온톨로지를 더 적은 비용으로 만들 수 있다. 온톨로지 구축에 있어 CQ단계는 온톨로지의 역할, 범위 뿐 아니라 핵심 개념과 관계를 추출할 수 있는 가장 중요한 도구이다.

2.2 온톨로지 추론

시맨틱 웹 관점에서의 추론은 기존에 알고 있던 명제로부터 새로운 사실이나 명제를 논리적으로 유도하는 과정으로 정의할 수 있다. 즉, 1차 논리, 특히 서술 논리와 2차 논리 규칙에 의한 추론만을 그 대상으로 할 수 있다.

OWL DL 추론 시스템은 다음과 같은 문제들을 취급한다.

- 포함관계(subsumption) 추론: 지식이 올바른가?
- 등가(equivalence) 추론: 중복된 지식은 없는가?
- 일관성(consistency) 추론: 일관적인 지식베이스인가?
- 유효성(satisfiability) 추론: 지식이 의미가 있는가? (인스턴스를 가질 수 있는가?)
- 지식 질의(querying) 추론: 인스턴스 i가 클래스 C의 인스턴스인가?

포함관계 추론은 하나의 클래스가 다른 클래스의 하위 개념인지 아닌지를 검사하는 것이다. 이러한 과정을 통해 추론기는 온톨로지 클래스의 계층체계 오류를 발견하고, 올바른 구조를 추론해낼 수 있다. 일관성 추론은 두 클래스에서 동시에 상속받은 클래스가 존재한다는 것 자체가 모순이 된다는 것을 제약조건 검사를 통해 알게 된다.

이러한 온톨로지 지식베이스 기반의 시맨틱 검색은 사용자의 키워드를 해석하여 구조적 질의어로 자동 변환해

줌으로써 시맨틱 검색이 가능하도록 해준다.

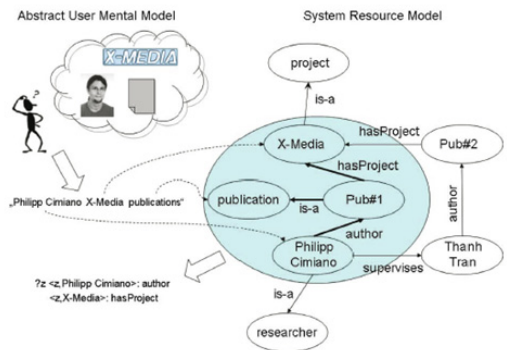
3. 온톨로지 지식베이스 설계 및 구축

앞 절에서 언급한 바와 같이 모든 도메인 지식을 온톨로지 구축하기에는 어려움이 따른다. 본 논문에서는 온톨로지 스키마 설계 단계에서 대상 도메인을 멀티미디어 콘텐츠로 한정하였으며, 더 세부적으로는 영화, 드라마, 예능으로 도메인을 한정하였다. 그리고 각 멀티미디어 콘텐츠의 메타데이터를 분석하고 영화, 드라마, 예능을 잘 표현할 수 있는 온톨로지 스키마를 설계하였다.

3.1 온톨로지 기반 키워드 해석

온톨로지 기반 키워드 해석을 위해 사용자가 입력하는 키워드 질의는 텀(term)으로 구분하고 각 텀에 매핑되는 온톨로지 개체를 식별한다. 식별된 개체를 기반으로 온톨로지 탐색을 통해 개체 간 관계를 찾고, 개체들 간의 관계로부터 서술 논리 결합적 질의(DL conjunctive query)를 유도함으로써 정확한 검색 결과를 제공한다[8].

[그림 3]은 사용자 검색 키워드로부터 질의를 해석하는 절차를 설명하고 있다.



[그림 3] 질의 워크플로우

입력된 키워드의 각 텀을 온톨로지 기반으로 매핑하게 되면 그림과 같이 온톨로지 내의 개체로 식별된다. 매핑된 개체들 간의 연결 관계를 탐색하면 "X-Media"는 "project"의 인스턴스이고, "Philipp Cimiano"는 "researcher"의 인스턴스임을 알 수 있다. 또한 연결 관계 탐색을 통해 "Philipp Cimiano"가 "X-Media의

“researcher”였으며, 그가 작성한 어떠한 출판물이 있을 수 있다. 이러한 관계 탐색을 통해 아래와 같은 정형화된 질의어를 유도해 낼 수 있다.

“?z = < z, *Phlipp Cimiano* >
: author and < z, *X - Media* >
: hasProject“

〈표 1〉 인덱스 종류

인덱스 종류	인덱스 설명
Entity Index	지식 베이스 내의 Class, ObjectProperty, DatatypeProperty, Instance의 레이블 (user_term)에 대한 인덱스
Literal Index	지식 베이스 내의 class에서 가질 수 있는 리터럴에 대한 인덱스

〈표 1〉은 논문에서 사용한 인덱스로서, 텀 매핑 과정에서 개체 식별을 위해 두 가지 형태의 인덱스를 구성하였다. 이 인덱스로부터 사용자 검색 키워드의 개별 텀에 대한 지식베이스 내의 식별자인 URI를 제공한다. 제공된 URI를 통해 지식베이스에서 실제 개체를 찾고, 그 개체의 타입이나 특성을 분석하면 심볼 수준의 텍스트에서 개념 수준의 텍스트 정보를 획득할 수 있게 된다. 예를 들면, “KOREA”라는 키워드는 시스템 측면에서는 단순한 텍스트일 뿐이다. 하지만 이를 개체 식별하면 “KOREA”는 “나라”의 “이름”이며, 그 “의미”는 “고려”이며, “KOREA”의 위치는 “동아시아”라는 개념 수준의 의미를 찾을 수 있게 된다.

3.1.1 지식 개체 식별

사용자 검색 키워드에 해당되는 지식 개체를 찾는 텀 매핑 단계는 개체 식별, 지식 개체 조합 구성, 지식 개체 필터링의 3단계로 이루어진다.

- 1단계: NER(Named Entity Recognition)
지식베이스 인덱스로부터 사용자 검색 키워드를 포함하는 지식 개체 획득
- 2단계: NEC(Named Entity Combination)
획득된 지식 개체로부터 사용자 검색 키워드에 맞는 가능한 모든 NEC 생성

- 3단계: NEC Filtering
의미해석을 위한 연결 그래프를 생성할 수 없는 NEC 필터링

3.1.2 질의 의미 해석

개체 식별이 완료되면, 지식베이스로부터 식별된 개체를 모두 포함하는 부분 그래프를 찾는 질의 의미 해석을 수행한다. 부분 그래프는 다양하게 존재 가능하며, 사용자가 입력한 검색 키워드로부터 해석 가능한 다양한 의미 해석 후보가 될 수 있다. 이러한 의미 해석 후보를 지식베이스 기반으로 평가하여 가장 적합한 질의 의미 해석을 결정하고 질의문을 생성하게 된다.

일반적으로 구축된 온톨로지에서는 풍부한 메타데이터를 바탕으로 다양한 형태의 연결 관계를 뽑아 낼 수 있게 된다. 심지어 인스턴스 간의 관계 깊이가 깊어질수록 무수히 많은 부분 그래프를 뽑아낼 수 있는데, 본 논문에서는 개체 식별 결과로부터 가능한 모든 부분 그래프를 뽑는 것에 초점을 두었다.

논문에서 제안하는 질의 의미 해석 과정은 다음과 같이 NEC 분석, CG 구성, 질의 의미 해석 생성의 3단계로 이루어진다.

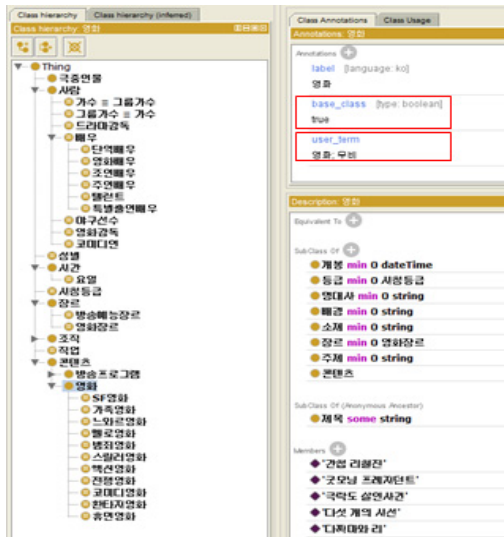
- 1단계: NEC 분석
텀 매핑 결과로 생성된 NEC 세그먼트의 구조를 분석하고, 개체 식별 타입에 따라 target class와 constraint로 구분
- 2단계: CG(Connection Graph) 구성
분석된 텀 매핑 세그먼트에 따라 지식 베이스를 탐색하여 부분 그래프인 CG를 구성
- 3단계: 질의 의미 해석 생성
CG로부터 질의 의미 해석 결과를 생성

3.2 온톨로지 지식베이스 스키마 설계

온톨로지를 구성하는 중요한 요소 중 하나는 인스턴스의 집합으로 정의될 수 있는 클래스이다. 클래스는 개념, 타입, 카테고리로서 구축된 멀티미디어 콘텐츠는 “콘텐츠”, “영화”, “드라마”, “예능”, “사람”, “영화배우”, “영화감독” 등 91개 클래스로 구성된다.

[그림 4]는 멀티미디어 콘텐츠를 표현하기 위한 온톨로지 스키마 구조를 설명하고 있다. 스키마 설계에 있어 특징적인 점은 Base class를 정의하고, 이로부터 인스턴

스를 생성하도록 설계하였다.



[그림 4] Base class 계층 구조 및 명세

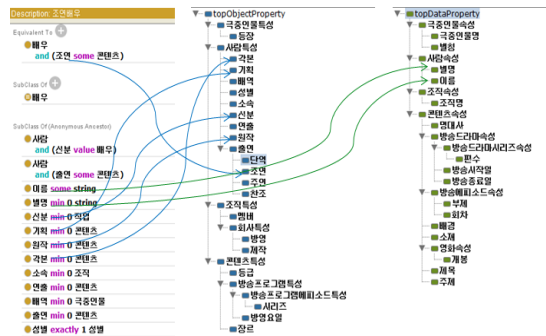
[그림 4]의 왼쪽 부분은 설계된 온톨로지 스키마의 각 클래스에 대한 계층 구조를 보여주고 있으며, 오른쪽 부분은 그 중 Base class인 영화를 정의하는 세부 명세를 보여주고 있다. Base class는 인스턴스를 생성할 수 있는 클래스로써 annotation을 활용하여 Base class임을 표현한다. 또한 Base class는 Primitive class로써 필요조건만으로 클래스를 명세하고 있다.

Defined class는 필요조건과 충분조건을 모두 충족시키는 클래스로서 지식베이스 기반으로 의미해석하기 위한 common sense로 사용된다. 한 예로 "주연배우" 클래스는 "배우" 클래스의 서브 클래스이며, "콘텐츠" 클래스의 멤버와 "출연", "연출", "각본", "원작", "기획" 등의 ObjectProperty 관계를 가져야 하며, "이름", "별명"의 DatatypeProperty로 리터럴 값을 가져야 한다는 필요조건을 명세하고 있다. 또한 "주연배우" 클래스는 "콘텐츠" 클래스의 멤버와 "주연"이라는 ObjectProperty 관계를 반드시 갖는 충분조건을 명세하고 있다.

이러한 ObjectProperty와 DatatypeProperty는 온톨로지에서 또 하나의 중요한 요소로써 관계를 표현한다. ObjectProperty는 인스턴스 간의 관계를 표현하며 프로퍼티(Property)라고 표현한다. DatatypeProperty는 인스턴스가 갖는 실제 값과의 관계를 표현하며 attribute라고 표현한다. 인스턴스가 가질 수 있는 값을 literal이라고 하

며, 정수, 문자열, 날짜 등을 값으로 가질 수 있다. 온톨로지에서는 보통 바이너리 관계를 표현하고 있으며, 다중 관계도 표현 가능하다.

[그림 5]는 온톨로지 스키마의 ObjectProperty와 DatatypeProperty의 계층 구조를 설명하고 있다. 가장 왼쪽의 클래스 명세에서 클래스가 가질 수 있는 ObjectProperty와 DatatypeProperty를 보여주고 있으며, 중앙은 인스턴스와 인스턴스 간의 관계를 표현하는 ObjectProperty의 계층 구조이며, 가장 오른쪽은 인스턴스가 가질 수 있는 값과의 관계를 표현하는 DatatypeProperty의 계층 구조이다.



[그림 5] 속성 계층 구조

그림에서 보는 바와 같이 "조연배우" 클래스는 충분조건으로 "콘텐츠" 클래스의 한 인스턴스와 "조연"이라는 관계를 갖으며, "조연"이라는 관계는 중앙의 ObjectProperty에서 "출연"의 서브 프로퍼티로 나타난다. "조연배우" 클래스가 가질 수 있는 프로퍼티는 "조연", "기획", "원작", "각본", "소속", "배역", "출연" 등이 있으며, "이름", "별명" 등의 어트리뷰트를 가질 수 있다.

온톨로지 지식베이스의 구축은 앞 절에서 설명한 멀티미디어 콘텐츠를 표현하기 위해 설계된 온톨로지 스키마에 인스턴스를 채우는 과정이라고 할 수 있다. 본 논문에서는 포털 사이트로부터 영화, 드라마, 예능에 대한 콘텐츠 메타데이터를 수집하여 온톨로지를 구축하였으며 프로토타입의 구현 및 성능 평가를 위해 57,000개의 인스턴스로 지식베이스를 구축하였다. <표 2>는 이러한 지식베이스의 개체 수를 나타낸다.

인스턴스의 적재는 웹상의 콘텐츠 메타데이터를 수집기를 통해 데이터베이스로 구축하고, 데이터베이스를 다시 온톨로지로 마이그레이션하는 과정을 거쳤다.

〈표 2〉 구축된 지식베이스의 개체 수

	Classes	Object-Properties	Datatype-Properties	Instances
개수	91	24	17	74,233 (영화: 2,235, 드라마: 22,463, 예능: 10,504)

3.3 SWRL 정의

SWRL은 서술 논리 추론 규칙 언어로서 최소한의 관계 정의로 스키마를 설계하고, OWL-DL 공리를 추론할 수 있다. 이러한 추론을 통해 일관성, 포함관계, 동등관계, 멤버관계 등을 추론할 수 있다.

한 예로, [그림 6]의 첫 번째 SWRL을 살펴보면, "y"는 "x"의 시리즈 관계이며, "y"는 "z"라는 제목을 가질 수 있다. 따라서 "x"는 "z"라는 제목을 가질 수 있음을 정의한 것이다. 이렇게 정의된 SWRL을 온톨로지 스키마에 추가하면 지식베이스 구축 시에 인스턴스들의 포함관계, 동등관계, 멤버관계를 추론 엔진을 통해 추론할 수 있다. 논문에서 구축된 온톨로지 추론은 명세 언어로 SWRL을 사용하고 있으며, 실제 추론은 Pellet 추론 엔진을 활용하고 있다.

Rules +

시리즈(?x, ?y), 제목(?y, ?z) -> 제목(?x, ?z)
조연(?z, ?y), 시리즈(?x, ?y) -> 조연(?z, ?x)
시리즈(?x, ?y), 등장(?z, ?y) -> 등장(?z, ?x)
시리즈(?x, ?y), 기획(?z, ?y) -> 기획(?z, ?x)
장르(?y, ?z), 시리즈(?x, ?y) -> 장르(?x, ?z)
시리즈(?x, ?y), 소재(?y, ?z) -> 소재(?x, ?z)
시리즈(?x, ?y), 배경(?y, ?z) -> 배경(?x, ?z)
원작(?z, ?y), 시리즈(?x, ?y) -> 원작(?z, ?x)
주연(?z, ?y), 시리즈(?x, ?y) -> 주연(?z, ?x)
시리즈(?x, ?y), 방송시작일(?y, ?z) -> 방송시작일(?x, ?z)
시리즈(?x, ?y), 출연(?z, ?y) -> 출연(?z, ?x)
각본(?z, ?y), 시리즈(?x, ?y) -> 각본(?z, ?x)
방영(?z, ?y), 시리즈(?x, ?y) -> 방영(?z, ?x)
시리즈(?x, ?y), 주제(?y, ?z) -> 주제(?x, ?z)
시리즈(?x, ?y), 방송종료일(?y, ?z) -> 방송종료일(?x, ?z)
시리즈(?x, ?y), 방영요일(?y, ?z) -> 방영요일(?x, ?z)
시리즈(?x, ?y), 단역(?z, ?y) -> 단역(?z, ?x)
연출(?z, ?y), 시리즈(?x, ?y) -> 연출(?z, ?x)

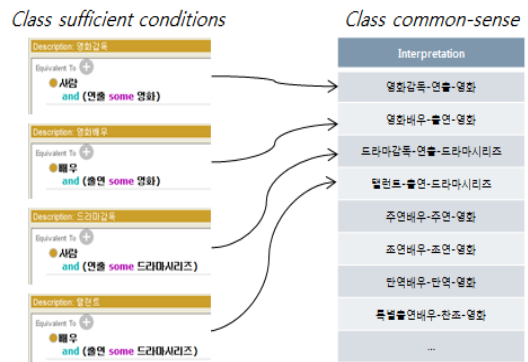
[그림 6] SWRL 정의

4. 성능 평가 및 비교 분석

4.1 실험환경 구성

사용자 검색 의도에 가장 부합하는 의미해석 결과를 제공하기 위해 본 논문에서 제안한 온톨로지 지식베이스 기반의 사용자 질의 의미 해석을 검증하는 실험으로 스키마로부터 추출한 common sense 모델의 유사도 실험을 예제로 가졌다. "류승완 영화"에서의 "류승완"을 개체 식별하면, "영화감독", "영화배우", "극장가"의 속성을 동시에 가질 수 있다. 하지만 "류승완"은 "영화감독"으로써 "연출"의 속성으로 연결된 "영화"의 인스턴스가 다른 속성보다 많이 나타나게 된다. 즉, "류승완"은 영화감독으로써 더 유명한 common sense와 같은 정보를 지식베이스로부터 획득할 수 있다.

[그림 7]은 "영화감독", "영화배우" 등에 대한 충분조건을 스키마로부터 추출하여 common sense 모델을 구성한 것이다.



[그림 7] class common sense 모델

시스템에서 제공하는 의미해석 결과를 한 단계 추상화하면 "류승완 연출 영화"는 "감독 배우 작가 연출 영화"라는 형태의 의미해석 결과를 생성할 수 있다. 이와 같이 의미해석 결과를 모두 추상화 하여 common sense 모델과 유사도를 계산하면 특정 인스턴스에 적합한 속성 순으로 의미해석 결과를 평가하고 순서화할 수 있게 된다. 유사도 측정 알고리즘은 협력적 추천 방법의 Confidence 알고리즘을 활용한다. 아래 식과 같이 의미해석 결과에 대한 유사도 측정을 위해 Confidence를 계산할 수 있다.

$$Confidence(I, Cs) = \frac{\sigma(I \cap Cs)}{\sigma(I)} \quad (1)$$

Confidence 알고리즘은 지식베이스 기반의 common sense 모델과 질의 의미해석 결과 간의 유사도 측정 시에 동일한 패턴의 의미해석 결과는 정확도 높게 구분할 수 있다.

4.2 실험 결과

<표 3>은 의미해석 결과에 대한 Confidence 값을 계산한 결과이다.

<표 3> common sense와 의미해석 유사도 결과

의미해석 결과	common sense	유사도	순위
영화배우 출연 영화	영화배우 출연 영화	3/3 = 1.00	1
영화배우 연출 영화		2/3 = 0.67	2
영화배우 각본 영화		2/3 = 0.67	2
영화감독 출연 영화	영화감독 연출 영화	2/3 = 0.67	2
영화감독 연출 영화		3/3 = 1.00	1
영화감독 각본 영화		2/3 = 0.67	2
영화배우 영화감독 출연 영화	영화배우 출연 영화	3/4 = 0.75	1
영화배우 영화감독 연출 영화		2/4 = 0.50	2
영화배우 영화감독 각본 영화		2/4 = 0.50	2
영화배우 영화감독 출연 영화	영화감독 연출 영화	2/4 = 0.50	2
영화배우 영화감독 연출 영화		3/4 = 0.75	1
영화배우 영화감독 각본 영화		2/4 = 0.50	2

표에서 보는 바와 같이, "영화배우"의 특성을 나타내는 인스턴스는 common sense 모델과의 Confidence 유사도 계산 시, "출연" 속성이 나타나는 의미해석 결과가 가장 유사도가 높다는 것을 알 수 있다. 마찬가지로 "영화감독"의 특성을 나타내는 인스턴스는 "연출" 속성이 나타나는 의미해석 결과가 가장 높은 유사도를 보인다. 다른 예로, "영화배우"와 "영화감독"의 특성을 모두 가지고 있는 인스턴스의 경우 "출연"과 "연출" 속성이 동일한 유사도로 나타난다. 이러한 경우, 동일한 유사도를 나타내는 속성에 대해 순서화할 수 있는 모델을 필요로 하게 된다.

5. 결 론

본 논문에서는 사용자 질의 의미 해석을 위한 온톨로지 지식베이스 스키마 구축을 제안하였다. 또한, 제안된 스키마로부터 질의 의미의 유사도를 측정함으로써 생성된 의미해석 후보들을 평가 및 순서화하는 실험을 통하여 사용자 검색 의도에 만족하는 검색 결과를 제공하는 것을 확인할 수 있었다. 제안된 시스템은 common sense 모델에 존재하지 않는 인스턴스의 구분은 어렵지만, 좀 더 다양한 도메인의 지식을 활용하여 온톨로지를 구축한다면 사용자가 높은 만족감을 가지는 질의 의미 해석 시스템을 제공할 수 있을 것이다.

향후 연구 과제로 사용자 검색 패턴 분석을 통해 사용자 선호 기반 질의 그래프 평가 기법에 대한 연구를 진행할 것이다.

This research was financially supported by the Ministry of Education, Science Technology (MEST) and National Research Foundation of Korea(NRF) through the Human Resource Training Project for Regional Innovation

참 고 문 헌

- [1] 이경일, 최광선, 정용일, "다시 보는 시맨틱 웹 그리고 시맨틱 기술Ⅱ", White Paper, 솔트룩스, 2010.
- [2] E. Mäkelä, "Survey of Semantic Search Research" Proc. Seminar on Knowledge Management on the

Semantic Web, 2005.

[3] G. Zenz, X. Zhou, E. Minack, W. Siberski, and W. Nejdl, "From keywords to semantic queries - Incremental query construction on the Semantic Web," J. Web Sem., vol.7, no.3, pp.166-176, 2009.

[4] H. Fu and K. Anyanwu, "Effectively Interpreting Keyword Queries on RDF Databases with a Rear View" LNCS, vol.7031, 2011.

[5] OWL 2 Web Ontology Language Profile (<http://www.w3.org/TR/owl2-profiles/>), W3C, 2009.

[6] S. Kang, D. Kim, Y. Lee, S. Hyun, D. Lee, and B. Lee, "A Semantic Service Discovery Network for Large-Scale Ubiquitous Computing Environment," ETRI Journal, vol.29, no.5, 2007.

[7] T. Tran, P. Haase and R. Studer, "Semantic Search - Using Graph-Structured Semantic Models for Supporting the Search Process" LNCS, vol.5662, 2009.

[8] T. Tran, P. Cimiano, S. Rudolph and R. Studer, "Ontology-Based Interpretation of Keywords for Semantic Search," LNCS, vol.4825, pp.238 - 345, 2007.

도 하 나



- 2013년 2월 : 한남대학교 컴퓨터공학과(공학사)
- 2013년 3월 ~ 현재 : 한남대학교 컴퓨터공학과(공학석사)
- 관심분야 : 시맨틱 검색, 클라우드 컴퓨팅, 상황인식 컴퓨팅, 데이터베이스 보안

· E-Mail : ooooduck@me.com

이 무 훈



- 2002년 8월 : 한남대학교 컴퓨터공학과 컴퓨터공학과(공학사)
- 2004년 8월 : 한남대학교 컴퓨터공학과(공학석사)
- 2008년 10월 ~ 현재 : 한국전자통신연구원 차세대스마트TV연구단 선임연구원

· 관심분야 : 시맨틱 검색, 온톨로지, 정보검색, 상황인식 컴퓨팅

· E-Mail : leemh@etri.re.kr

정 훈



- 2012년 2월 : 한남대학교 컴퓨터공학과(공학사)
- 2012년 3월 ~ 현재 : 한남대학교 컴퓨터공학과(공학석사)
- 관심분야 : 클라우드 컴퓨팅, 상황인식 컴퓨팅, 멀티모달

· E-Mail : jeonghoon0322@hanmail.net

최 의 인



- 1982년 2월 : 한남대학교 계산통계학과(학사)
- 1984년 8월 : 홍익대학교 전자계산학과(석사)
- 1995년 8월 : 홍익대학교 전자계산학과(이학박사)
- 1996년 3월 ~ 현재 : 한남대학교 컴퓨터공학과 교수

· 2003년 2월 : UCLA 방문 교수

· 관심분야 : 시맨틱 웹, 유비쿼터스 컴퓨팅, 모바일, 클라우드 컴퓨팅

· E-Mail : eichoi@hnu.kr