

# 효과적인 지식확장을 위한 LOD 클라우드에서의 변화수용적 심층검색

김광민

솔트룩스 인공지능연구센터  
(jp9894@naver.com)

손용락

서경대학교 컴퓨터공학과  
(syl@skuniv.ac.kr)

본 연구는 시멘틱 웹의 실질적 구현체인 LOD 클라우드에서 연결정책을 활용함으로써 LOD들간 연결을 효과적으로 제공하고 LOD의 변경된 내용을 검색결과에 빠짐없이 반영할 수 있는 방안을 제시한다. 현재 LOD 클라우드에서는 개체간 연결은 <owl:sameAs>를 이용하여 개체들이 동일함을 명시적으로 기술하는 방식으로 이루어져 있다. 하지만, 이러한 명시적 연결방식은 LOD 클라우드 규모의 방대함에도 불구하고 개체간 동일성을 개체단위에서 파악하여야 하는 어려움이 있으며 주기적으로 LOD에 추가하여야 함에 따라 검색 시 개체들이 누락되는 한계가 있다. 이를 극복하기 위하여 본 연구에서는 명시적 연결을 생성하는 대신 LOD별로 연결하고자 하는 LOD와의 연결정책을 수립하여 LOD와 함께 공개하는 방식을 제안한다. 연결정책을 활용함으로써 연결하여야 할 동일개체를 검색시점에서 파악할 수 있으므로 추가되었던 개체들을 누락됨 없이 검색결과에 포함시킬 수 있고 LOD 클라우드에서의 연결성도 효과적으로 확충할 수 있다. 확충된 연결성은 정보의 지능적 처리의 선행과정인 지식확장의 근간이 된다. 연결정책은 연결하고자 하는 소스와 타겟 LOD의 주어 개체들간의 동일성을 평가하는데 도움이 되는 술어 쌍을 명세하는 방식으로 수립하며 검색 시 이러한 술어쌍에 대응하는 RDF 트리플을 검색하고 이들의 목적어들이 충분히 동일한 것인가를 평가하여 주어개체들의 동일수준을 판단한다. 본 연구에서는 이러한 연결정책을 이용하여 여러 LOD들을 심층적으로 검색하는 시스템을 구현하였다. 검색과정에서는 기존 명시적 연결들도 함께 활용하도록 구현하였다. 검색시스템에 대한 실험은 DBpedia의 주요 LOD들을 대상으로 진행하였다. 실험결과 연결대상 개체들의 목적어들이 0.8 ~ 0.9의 유사수준을 가지는 경우 적절한 확장성을 가지고 충분히 신뢰적인 개체들을 적절하게 포함하는 것으로 확인하였다. 또한, 개체들은 8개 이상의 동일연결을 제공하여야 검색결과가 신뢰적으로 활용될 수 있을 것으로 파악되었다.

**주제어** : 시멘틱 웹, 심층검색, 온톨로지, 오픈 데이터 연결, 연결정책

논문접수일 : 2018년 1월 15일    논문수정일 : 2018년 6월 18일    게재확정일 : 2018년 6월 23일  
원고유형 : 일반논문    교신저자 : 손용락

## 1. 서론

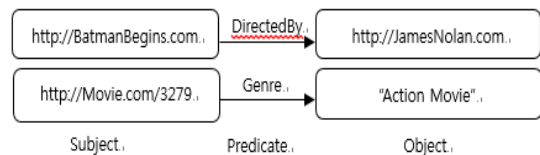
현재의 웹은 ‘문서의 웹(Web of Documents)’ 구조를 근간으로 하고 있다. 작성된 내용을 브라우저를 통하여 제공하는데 초점이 맞추어진 HTML은 데이터 단위로 세부적인 의미를 부여하는 것이 불가능하다. 이와 더불어, 연결도 웹

페이지 단위로만 이루어짐에 따라 기술하고자 하는 개별 개체 단위에서의 정교한 연결을 제공하지 못하고 있다. 문서의 웹이 가지는 이러한 특성은 그 규모의 방대함에도 불구하고 문서에 수록된 세부 내용이 가지고 있는 의미를 컴퓨터가 이해하고 이를 지능적으로 처리하는데 많은 제약이 된다. 이러한 제약을 극복하기 위하여 팀

버너스 리는 ‘데이터의 웹(Web of Data)’ 구조의 시멘틱 웹을 주창하였으며 W3C가 중심이 되어 관련 표준화 작업을 진행하고 있다(Bill and Bary, 2011; Hitzler, 2009; Grigoris, 2012; W3C, 2017). 시멘틱 웹은 개체를 기술하는 세부 데이터 단위로 의미를 부여하고 연결 또한 개체 단위로 이루어져 지식의 확장이 페이지 단위가 아닌 개별 사실단위에서 이루어지는 것을 가능하게 한다. 이러한 구조를 제공함에 따라 컴퓨터는 데이터에 부여된 의미에 기반한 지능적인 정보처리를 가능하게 하며 필요한 경우 하나의 주제에 대한 접근으로 해당 주제에 대한 다양한 관점에서 기술된 내용들을 한 번의 검색으로도 확장 획득하는 것을 가능하게 한다(Heath and Bizer, 2011; Sikos, 2015). 시멘틱 웹의 이러한 확장성은 웹을 지능적 처리가 가능한 정보자원으로 역할 할 수 있도록 한다. 본 연구는 이러한 시멘틱 웹의 실질적인 구현체인 Linked Open Datga(LOD) 클라우드에서의 검색이 보다 풍부한 연결성에 기반하여 진행될 수 있도록 하는 방안을 제시한다. 연결성 확대는 정보의 지능적 처리를 위한 선행과정인 지식확장의 근간이 된다.

LOD 클라우드는 시멘틱 웹의 취지에 공감하는 여러 기관들이 연결성을 가지는 공개된 데이터집합인 LOD와 이에 대한 의미적 명세인 온톨로지를 공개함으로써 구성된다(Abele and McCrae, 2017; Auer et al., 2014; Jeong, 2015). LOD와 온톨로지는 RDF(Resource Description Framework) 모델을 이용하여 표현된다. RDF 모델은 <Figure 1>과 같이 주어(Subject), 술어(Predicate), 목적어(Object)로 구성된 RDF 트리플 형식을 가지며 W3C에서는 이들 각각을 URI(Unique Resource Identifier)로 식별될 수 있도록 할 것을 표준으로 한다. 단, 목적어는 문자

열(Literal)로도 구성하는 것이 허용되며 실제로 많은 목적어들이 문자열로 제공된다(Carol and Shenghui, 2015; Martin et al., 2007). 온톨로지에는 주어와 목적어 개체의 주체가 명세되며 술어에 대해서는 연결할 주어와 목적어의 주체를 각각 정의역(Domain)과 치역(Range)으로 명세한다(Dean and James, 2011; Robert and Barry, 2015). 이러한 명세에는 온톨로지 기술언어들인 RDFS, OWL, OWL2가 제공하는 다양한 문법적 특성들이 반영됨으로써 데이터가 가지는 상세 의미를 풍부하게 기술하는 것이 가능하다. <Figure 1>의 예는 <http://BatmanBegins.com>라는 주어 개체가 <http://JamesNolan.com>라는 목적어 개체를 감독으로 하고 있다(Directed)는 사실과 <http://Movie.com/3279>라는 주어 개체의 장르(Genre)가 “Action Movie”라는 사실을 기술하고 있다(Eric, 2014; Sikos, 2015; W3C, 2017).



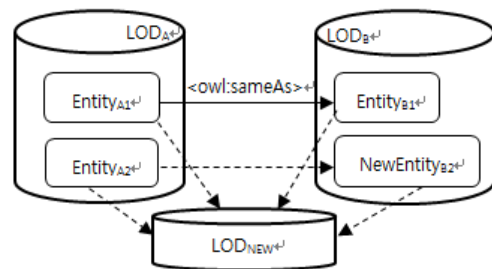
<Figure 1> Structure of RDF triples

LOD 클라우드에서의 연결은 대부분 <owl:sameAs>라는 술어를 사용하여 주어와 목적어 개체가 동일하다는 사실을 LOD에 명시적으로 기술함으로써 이루어진다. 이러한 연결을 통하여 주어 개체는 자신을 설명하고 있는 내용에 더하여 연결대상인 목적어 개체를 설명하고 있는 내용을 확장하는 것이 가능하게 된다(Harth A., et al., 2014). <Figure 1> 예에서 만약 RDF 트리플 {<http://BatmanBegins.com> <owl:sameAs>

<http://Movie.com/3279>}이 존재한다면 <http://BatmanBegins.com>에 대한 검색결과로 감독이 <http://JamesNolan.com>라는 사실과 더불어 장르가 “Action Movie”라는 사실도 함께 획득하는 것이 가능하게 된다. 이와 같이 LOD 클라우드는 서로 다른 RDF 트리플들의 주어 개체들이 동일하다는 주장에 근거하여 연결을 이루고 이를 통하여 확장된 지식을 얻을 수 있는 구조적 특성을 가지고 있다(Abele et al., 2017; Konstantiou, 2015). 하지만, 이러한 연결구조를 갖추고 있음에도 불구하고 현재의 LOD 클라우드에서는 아직도 연결성이 태부족하여 원래의 목표에 제대로 이루지 못하고 있는 상황이다(Machael, 2013; Schmachtenberg et al., 2017). 2017년을 기준으로 약 2,740여개 LOD들이 LOD 클라우드에 참여하고 있지만 이들 가운데 44%는 다른 LOD들과 전혀 연결되지 않은 채 데이터 사일로에 머물러 있다. 2개 이하의 LOD들로만 연결되어 있는 경우도 71%에 이르고 있다. 10개 이상 LOD들과 연결을 가지고 있는 경우는 8%를 넘지 못하고 있다. 연결이 이루어진 경우에도 이들 중 40% 이상은 연결대상을 DBpedia, GeoName, W3C와 같이 유명한 LOD들에 집중되어 있다(Park et al., 2013). LOD 클라우드에서의 이러한 연결의 빈약함은 2015년 이후 LOD 클라우드의 성장을 담보상태에 머물러 있도록 하는 중요한 원인으로 알려져 있다(Bizer, 2017; Jeff, 2017).

현재 LOD 클라우드에서의 <owl:sameAs>를 이용한 명시적 연결방식은 연결트리플들을 생성, 관리하여야 한다는 부담과 더불어 LOD 클라우드에서의 변화를 검색결과에 실시간으로 반영하는 것이 불가능하다는 한계를 가진다. <Figure 2>는 명시적 연결이 가지는 이러한 문제점을 도

시한다. 연결을 위한 RDF 트리플 {<EntityA1> <owl:sameAs> <EntityB1>}를 LOD<sub>A</sub>에 명시적으로 기술함으로써 <EntityA1>에 대한 검색결과에 <EntityB1>에 대한 내용을 더하여 제공할 수 있다. 하지만, LOD<sub>B</sub>에 <EntityA2>와 동일한 개체인 <NewEntityB2>가 새로이 등장하였을 경우 연결 트리플 {<EntityA2> <owl:sameAs> <NewEntityB2>}를 LOD<sub>A</sub>에 명시적으로 기술하기 전까지는 <EntityA2>에 대한 접근을 <NewEntityB2>에 대한 접근으로 확장하는 것이 불가능하다. 더불어, LOD<sub>NEW</sub>가 LOD 클라우드에 새로이 참여하였을 경우 LOD<sub>A</sub>나 LOD<sub>B</sub>는 LOD<sub>NEW</sub>가 공개하고 있는 데이터를 모두 파악한 후 <owl:sameAs>를 이용하여 명시적 연결들을 자신의 LOD에 기술한 후에야 LOD<sub>NEW</sub> 개체로부터 확장된 지식을 획득하는 것이 가능하게 된다. 실제 LOD 운영에 있어서 이러한 방대한 명시적 연결들을 주기적으로 갱신, 관리하는 것은 큰 부담으로 작용한다. 결국, 현재의 명시적 연결방으로는 LOD 클라우드에서의 변화된 내용을 검색요청 시점에서 검색결과에 포함시켜 제공하는 것이 불가능하다.



<Figure 2> Explicit Links' Problems

<Figure 2>에서 제시한 문제점을 극복하기 위하여 본 연구에서는 LOD 마다 방대한 규모의 명

시적 연결 트리플들을 생성, 관리하는 대신 서로 다른 LOD들에 존재하는 개체들간의 동일성을 평가하는데 적용할 연결정책을 수립하고 이를 LOD와 함께 공개하는 방안을 제시한다. 연결정책에는 연결하고자 하는 타겟 LOD와 동일성 평가에 적용할 여러 조건들을 명세한다. 본 연구에서는 연결정책을 명세하기 위한 어휘집합을 제시하였으며 수립된 연결정책을 참조하여 LOD 클라우드의 여러 LOD들을 심층적으로 검색하는 변화수용적 심층검색(Change Acceptable In-Depth Searching: CAIDS) 시스템을 구현하였다. CAIDS는 사용자 질의에 지정된 표층LOD로부터 검색한 개체들과 동일한 것으로 평가되는 개체들을 사용자가 인식하지 못한 여러 LOD들로부터 심층적으로 검색하고 이들에 대한 상세 내용을 표층LOD로부터 검색한 개체들에 대한 내용에 확장하여 제공한다. 그 결과 LOD클라우드에 대한 검색이 다양한 관점에서 기술된 지식의 확장으로 이어지게 된다. 심층검색은 LOD의 연결정책에 명세된 타겟 LOD를 따라 진행한다. 또한, 검색된 개체들간의 동일성은 연결정책에 명세된 술어쌍에 연계된 목적어들의 유사도를 평가하여 판단한다. 또한, 검색시점에서 연결정책을 참조하여 타겟LOD로부터 동일개체를 획득함에 따라 타겟LOD에서 이루어졌던 변화들이 검색시점에서 검색결과에 반영되는 것이 가능하여진다. CAIDS는 개체에 대하여 확장적으로 기술된 내용뿐만 아니라 해당 내용의 의미적 구조를 안내하는 온톨로지를 재구성하여 검색결과로 함께 제공한다. 이에 따라 CAIDS의 검색결과는 기존 LOD와 동일한 방식으로 활용하는 것이 가능하다.

본 논문의 2장은 LOD 클라우드에서의 연결성 구축에 관련된 선행연구들과 이들의 한계점을

기술한다. 3장은 LOD 클라우드에서 발생한 변화를 연결정책을 활용하여 검색결과에 실시간으로 반영하는 검색시스템의 개요와 구현된 세부 동작과정을 소개하며 4장에서는 검색결과와 확장 수준과 검증된 사실들의 포함 수준, 검색규모, 검색결과와 신뢰성을 기준으로 성능평가를 진행하고 그 결과를 분석하였다. 5장에서는 향후 기능추가 방향을 제시하였다.

## 2. 관련 연구

LOD 클라우드에서의 LOD간 연결은 기본적으로 표준식별자 기반 연결방식과 목적어 유사성기반 연결방식으로 나누어진다. 표준식별자 기반 연결방식은 ISBN(International Standard Book Number), GTIN(Global Trade Item Numbers), ISIN(International Securities Identification Numbering) 등과 같은 표준식별자에 역함수 술어 <owl:InverseFunctionalProperty>를 적용한다. 즉, 서로 다른 주어 개체들이 역함수가 적용된 동일한 식별자를 가질 경우 이들은 동일한 개체로 판단하는 방식이다. 높은 정확성을 제공하지만 표준식별자가 사용되는 경우로만 그 적용이 한정된다(Heath et al., 2011).

목적어 유사성기반 연결방식은 소스와 타겟 LOD에서 동일한 의미를 가지는 술어들을 선정하고 이들의 목적어들이 유사한 내용일 경우 해당 주어 개체들이 동일한 것으로 판단하는 방식으로 SILK(Volz et al., 2009), LIMES(Ngonga et al., 2011), SALE(Park and Sohn, 2014), TILE(Park and Sohn, 2014), (Choi and Park, 2013) 등이 있다. 이들은 공통적으로 LOD 별로 유사성 평가에 적용할 연결정책을 마련하고

LOD를 공개할 때 연결정책을 참조하여 타겟 LOD에서 동일한 개체를 선별한 후 이들을 향하여 <owl:sameAs>를 이용한 명시적인 연결트리플을 생성하고 이를 LOD에 첨부하여 공개한다. 이들 가운데 SILK는 연결정책을 참조하여 소스와 타겟LOD로부터 <rdfs:labels>와 같은 대표성을 가지는 술어들을 기준으로 유사성이 높은 개체들로 후보군을 구축한 후 세부적인 연결정책을 적용하여 유사성을 재차 평가하는 방식을 운영한다. LIMES는 SILK의 성능을 향상시키는데 초점을 맞추고 있다. SILK와 마찬가지로 연결 후보 개체를 발굴한 후 이를 기반으로 타겟 개체들의 표본집합을 구성하고 각 표본과 가장 인접한 개체들을 사전에 매칭시키는 작업을 한다. 인접성 파악은 개체들에게 공간위치를 부여한 후 삼각부등식을 적용하여 측정한다. 이 과정에서 유사하지 않음이 명확한 개체들은 제외함으로써 유사성 평가의 성능을 향상시킨다. SILK와 LIMES의 유사성 평가방식은 연결정책에 명시된 매칭술어들의 목적어 값들간의 유사성에 국한되어 있다. SALE과 TILE은 이러한 평가방식에 더하여 개체의 술어들이 가지는 문법적 특성을 유사성 평가에 반영하여 <owl:sameAs>로 기술되는 개체 동일성에 대한 신뢰도를 향상시키는데 주안점을 두었다. 이들은 LIMES의 방식을 기본적으로 진행한 후 그 결과 트리플들의 술어들에 대하여 계층성, 연관성, 동일성, 역함수, 정의역 등의 특성을 파악하고 이를 유사성 평가에 반영하였다. SALE은 RDFS, OWL 문법을 수용하였고 TILE은 이들과 더불어 OWL2까지 수용함에 따라 추론을 위한 문법적 특성도 유사성 평가에 반영하였다. SILK, LIMES, SALE, TILE 모두 공통적으로 명시적 연결을 생성하고 이를 LOD와 함께 공개하는 방식을 채택하고 있다. 이러한 방

식은 전술한 바와 같이 LOD 클라우드에서의 변화된 내용을 검색시점에서 결과에 반영할 수 없다는 근본적인 한계를 가진다.

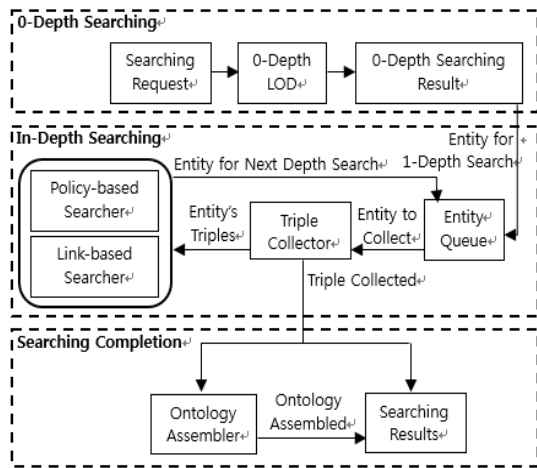
### 3. 변화수용적 심층검색

<owl:sameAs>를 이용한 기존의 명시적 연결 방식이 가지고 있는 LOD 클라우드에서의 변화에 대한 비수용적 문제를 해결하기 위하여 본 장에서는 LOD 마다 연결정책을 명시하고 검색이 이루어지는 시점에서 연결정책들을 참조하여 동일개체들을 도출하고 이들을 기술하고 있는 내용을 획득함으로써 검색결과가 LOD 클라우드의 변화된 내용을 반영할 수 있도록 하는 방안을 제안한다.

#### 3.1 시스템 개요

CAIDS 시스템의 동작개요는 <Figure 3>과 같다. LOD 클라우드에 대한 질의요청은 사용자가 특정 LOD를 0-Depth LOD로 지정하는 것으로 시작한다. 사용자는 CAIDS에 {Query, SPARQL Endpoint, Depth Level, Similarity Level}을 제시한다. Query은 RDF 데이터집합에 대한 W3C 표준 질의문법인 SPARQL로 작성된 질의문이다. SPARQL Endpoint는 Query를 받아들이고 검색 결과를 제공하는 프로세스의 URI로 0-Depth LOD에서 제공하는 기능이다. Depth Level은 0 이상 정수이며 심층검색을 진행하는 깊이에 대한 제한을 지정한다. Depth Level이 0인 경우 CAIDS는 사용자 질의를 0-Depth LOD에서만 실행하는 것으로 국한하며 1 이상일 경우에 심층 검색을 진행한다. Similarity Level은 0.0 ~ 1.0 값

을 가지며 소스와 타겟 RDF 트리플들의 술어쌍이 소스LOD의 연결정책에 부합할 경우 이들의 목적어들에 대한 유사 정도를 지정한다. 1.0인 경우 목적어들이 완전히 동일하다는 조건을 만족하여야만 소스와 타겟 주어개체들을 동일후보개체집합에 포함시킨다.



〈Figure 3〉 Architecture of CAIDS

표층검색(0-Depth Searching)은 사용자가 질의를 요청한 LOD(0-Depth LOD)에서 이루어지는 검색이다. 표층검색 결과 개체들은(0-Depth Searching Result) 향후 심층검색(In-Depth Searching)이 확장시켜야 할 기준이 되기 위하여 개체대기열(Entity Queue)에 등록된다. 심층검색 단계는 개체대기열로부터 선입선출방식으로 개체(Entity to Collect)를 제공받으면서 시작한다. 해당 개체에 대하여 심층검색을 진행하기 전에 트리플수집기(Triple Collector)는 우선 그 개체가 주어진 URI에 적시된 LOD에서 RDF 트리플의 주어개체로 실제로 존재하는 것인가를 HTTP 수행으로 확인한다. 실존이 확인된 개체의 RDF 트

리플은 검색완료(Searching Completion) 단계로 전달(Triple Collected)되어 검색결과(Searching Results)의 일부분이 된다. 이와 동시에, 실존이 확인된 개체는 새로운 소스개체(Entity's Triples)가 되어 정책기반검색기(Policy-based Searcher: PS)와 연결기반검색기(Link-based Searcher: LS)에게 전달된다. 개체가 저장되어 있는 LOD가 연결정책을 운영하고 있는 경우에는 해당 개체를 PS에 전달하고 그렇지 않은 경우나 해당 개체가 <owl:sameAs>로 연결되었던 것이었을 경우에는 LS로 해당개체를 전달한다.

PS는 전달받은 개체가 저장되어 있는 LOD(소스LOD)가 공개하고 있는 연결정책을 참조하여 타겟LOD에 접근한 후 명세된 매칭술어의 목적어들간 유사성으로 접근 중인 개체와의 동일성을 평가하여 타겟개체로 선별한다. 검색한 타겟개체(Entity for Next Depth Search)들은 다음 깊이의 심층검색을 위하여 개체대기열에 등록한다. LS는 제공된 개체에 <owl:sameAs>로 기술된 명시적 연결이 있을 경우에 한하여 진행되며 <owl:sameAs>의 목적으로 기술되어 있던 타겟개체(Entity for Next Depth Search)를 개체대기열에 등록한다. 이후 개체대기열은 다시 선입선출로 개체를 제공하고, 이의 실존을 검증하고, 유효하다면 해당 RDF 트리플을 검색완료로 전달한다. 심층검색 단계에서는 이러한 과정을 최초 질의요청에서 명세한 검색깊이 만큼 반복한다.

검색완료(Searching Completion)는 심층검색으로 수집한 RDF 트리플들을 최초 검색을 요청한 사용자에게 제공하는 단계이다. 검색결과인 RDF 트리플 집합을 사용자가 활용하기 위해서는 이에 대한 온톨로지도 함께 제공되어야 한다. 따라서, 온톨로지 조립기(Ontology Assembler)는 심층검색 과정에서 RDF 트리플들을 제공한

LOD들의 온톨로지들을 참조하여 검색결과 의 의미적 구조를 제공할 수 있는 온톨로지를 재구성한다. 따라서, 최종 사용자에게 제공되는 검색결과(Searching Results)는 재구성된 온톨로지(Ontology Assembled)와 이를 준수하는 RDF 트리플들의 집합으로 구성된다. 본 연구에서는 이들을 N-triple 형식으로(Hart and Catherine, 2013; Sikos, 2015) 직렬화하였다.

### 3.2 연결정책

LOD 클라우드에서 정책기반검색(PS)이 진행되기 위해서는 LOD마다 연결정책이 명세되어야 한다. 이를 위하여 LOD(소스LOD) 관리자는 연결하고자 하는 타겟LOD가 공개하고 있는 주제들에 대한 내용과 그에 대한 의미적 구조인 온톨로지를 충분히 검토하여야 한다. LOD의 연결정책은 주제한정 정책과 술어매칭 정책으로 구성된다. 주제한정 정책은 연결하고자 하는 개체의 주제를 명세하는 'type'(<<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>>)에 한정을 둠으로써 타겟LOD에서의 검색이 소스 개체와 동일한 주제에 해당하는 개체들만을 대상으로 이루어지도록 한다. 예를 들어 소스LOD인 한국 DBpedia LOD의 <<http://ko.dbpedia.org/class/영화>> 주제와 타겟LOD인 프랑스 DBpedia LOD의 <<http://fr.dbpedia.org/class/cinema>> 주제를 주제한정 대상으로 소스LOD의 연결정책에 명세함으로써 소스LOD의 '영화' 개체와 동일한 것으로 연결할 개체를 타겟LOD에서 검색할 때 오직 'cinema' 형식의 개체들만을 대상으로 하게 된다.

술어매칭 정책은 소스와 타겟LOD의 온톨로지로부터 동일한 의미를 가지는 술어쌍을 파악하여 연결정책에 명세함으로써 PS로 하여금 이

들의 목적어들간의 유사성을 평가하도록 하는 것이다. 예를 들어 상기 예의 주제한정 정책과 더불어 소스LOD의 술어로 <<http://ko.dbpedia.org/property/제목>>과 타겟LOD의 술어로 <<http://fr.dbpedia.org/property/title>>를 동일한 의미를 가지는 매칭술어 쌍으로 연결정책에 명세할 경우 소스LOD의 '영화' 주제 개체와 연결할 개체를 타겟LOD에서 검색할 때 주제가 'cinema'이고 매칭술어인 '제목'과 'title'의 목적어들이 충분히 유사한 경우로 한정할 것이다. 일반적으로 동일한 것으로 한정된 주제에 대하여 복수의 매칭술어 쌍들을 연결정책으로 명세하는 것이 보다 신뢰적인 연결을 제공할 수 있다.

LOD와 함께 공개하는 연결정책도 LOD들간 연결을 위한 온톨로지에 해당한다. 따라서, 이를 명세하기 위한 어휘들로 <Table 1>을 제안한다. 네임스페이스 'lp'는 <http://linkpolicy.org/ontoligy>로 정의한다. 연결정책은 LOD와 마찬가지로 RDF 모델을 근간으로 구성된다. LOD(SourceLOD)는 lp:linkpolicy를 사용하여 다수의 상세 연결정책들을 명세한다. lp:linkpolicy는 목적어로 공백노드(BlankNode\_A)를 가지며 이 공백노드는 SourceLOD 연결정책 명세의 출발점이 된다. BlankNode\_A의 술어 lp:targetLOD는 SourceLOD와 연결시킬 타겟LOD의 SPARQL EndPoint 주소를 등록한다. 주제한정 등록은 lp:regRestrictTopic으로 명세되며 목적어로 공백노드(BlankNode\_B)를 가진다. BlankNode\_B는 lp:sourceRestrictPredicate와 lp:sourceRestrictTopic을 이용하여 소스LOD에서 주제한정에 사용될 술어(대부분은 <<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>>을 사용함)와 동일한 것으로 한정하고자 하는 주제(예: <<http://ko.dbpedia.org/class/영화>>)를 명세한다. 타겟LOD에 대한 주제한정 등록은 lp:reg

〈Table 1〉 Vocabularies for Link Policy

Vocabulary	Role	Subject	Object
lp:linkpolicy	Link policy registration	SourceLOD	BlankNode_A
lp:targetLOD	Target LOD registration	BlankNode_A	Target LOD
lp:regRestrictTopic	Topic restriction registration	BlankNode_A	BlankNode_B
lp:sourceRestrictPredicate	Predicate for source LOD's topic restriction registration	BlankNode_B	Predicate for source LOD's topic restriction
lp:sourceRestrictTopic	Type for source LOD's topic restriction registration	BlankNode_B	Type for source LOD's topic restriction
lp:regTargetRestrict	Target LOD's topic restriction registration	BlankNode_B	BlankNode_C
lp:targetRestrictPredicate	Predicate for target LOD's topic restriction registration	BlankNode_C	Predicate for target LOD's topic restriction
lp:targetRestrictTopic	Type for target LOD's topic restriction registration	BlankNode_C	Type for target LOD's topic restriction
lp:predicateMatching	Predicate matching registration	BlankNode_B	BlankNode_D
lp:sourcePredicate	Source LOD's predicate registration	BlankNode_D	Source LOD's predicate
lp:targetPredicate	Target LOD's predicate registration	BlankNode_D	Target LOD's predicate

TargetRestrict로 명세하며 목적어로 공백노드(BlankNode\_C)를 가지며 여기에 lp:targetRestrictPredicate와 lp:targetRestrictTopic을 사용하여 타겟 LOD에 적용할 주제한정 술어와 정보를 명세한다. 술어매칭은 lp:predicateMatching으로 명세한다. 술어매칭의 경우 주제한정에 귀속되어야 하므로 주어는 lp:regRestrictTopic의 목적어인 BlankNode\_B가 된다. lp:predicateMatching은 목적어로 공백노드(BlankNode\_D)를 가지며 이를 공동주어로 하는 lp:sourcePredicate과 lp:targetPredicate를 이용하여 소스LOD와 타겟LOD에서 동일한 의미를 가지는 매칭술어 쌍을 명세한다.

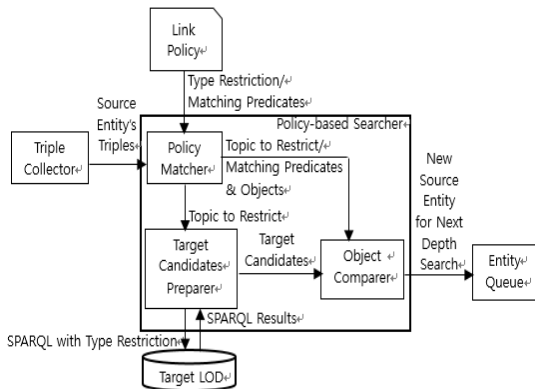
### 3.3 정책기반 심층검색

연결정책에 기반하여 심층검색을 진행하는 정

책기반검색기(PS)의 세부적인 동작구조는 <Figure 4>와 같다. 정책매치(Policy Matcher)는 트리플 수집기(Triple Collector)로부터 전달받은 소스개체의 트리플들(Source Entity's Triples)로부터 확보한 술어들이 연결정책(Link Policy)에 부합하는가를 확인한다. 부합이 확인된 경우 정책매치는 소스개체의 술어와 목적어(Matching Predicates & Objects), 그리고 그 술어가 귀속되어 있는 주제(Topic to Restrict)를 목적어비교기(Object Comparer)에 전달한다. 한편, 정책매치는 주제한정 명세(Topic to Restrict)를 타겟후보준비기(Target Candidates Preparer)에게 전달하여 타겟LOD로부터 주제한정에 부합하는 개체들과 이들의 트리플들을 가져와 준비할 수 있도록 한다. 타겟후보준비기는 주제한정에 근거한 SPARQL(SPARQL with Type Restriction)로 타겟LOD로부



터 주제한정 정책에 부합하는 내용(SPARQL Results)을 획득하고 이를(Target Candidates) 목적어비교기에 전달한다. 목적어비교기는 소스개체의 술어와 매칭되는 술어를 가지는 타겟개체를 검색하고 이들로부터 매칭술어 쌍에 연결된 목적어들간의 유사성을 비교한다. 유사성 비교는 N-Gram 거리측정 방식을 적용한다(Brown et al., 1992). N-Gram 방식은 기존의 문자열 비교 방식과는 달리 내부 문자열들의 순서가 다르더라도 적정한 보정을 통하여 유사성을 보다 정밀하게 측정할 수 있다는 장점이 있다. 유사성 비교 결과가 사용자 질의요청에서 요구되었던 유사수준을 초과하는 경우 타겟개체는 소스개체와 충분히 동일한 것으로 평가한 후 이를 새로운 소스개체(New Source Entity for Next Depth Search)로 하여 다음 깊이의 검색을 위하여 개체대기열(Entity Queue)에 등록한다.

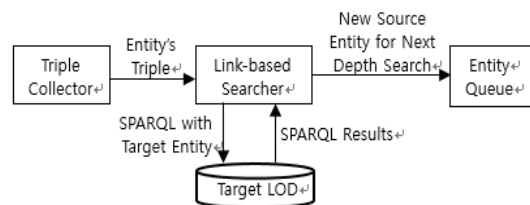


<Figure 4> Execution of Policy-based Searching

### 3.4 연결기반 심층검색

트리플 수집기로부터 제공된 RDF 트리플의 술어가 <owl:sameAs>인 경우 연결기반검색기

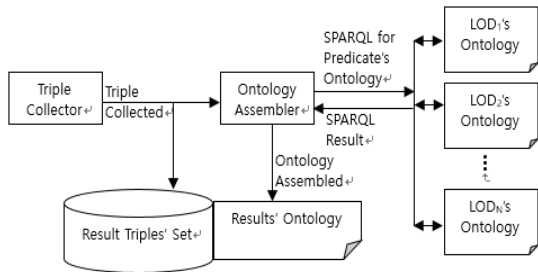
(LS)로 전달된다. LS의 세부적인 동작구조는 <Figure 5>와 같다. LS는 트리플 수집기로부터 전달받은 RDF 트리플(Entity's Triple)의 목적어 URI를 이용하여 SPARQL(SPARQL with Target Entity)을 구성한 후 이를 타겟LOD에 요청한다. <owl:sameAs>로 연결된 타겟개체의 경우 SPARQL Endpoint에 대한 정보가 없기 때문에 LS가 타겟LOD를 접근하기 위해서는 SPARQL Endpoint를 유추하여야 한다. 대부분의 경우 SPARQL Endpoint는 <http://ko.dbpedia.org/sparql> 형식으로 유추하는 것이 가능하다. SPARQL로부터 타겟개체의 존재가 확인되었을 경우 타겟개체는 다음 깊이 검색의 소스개체(New Source Entity for Next Depth Search)가 되기 위하여 개체대기열(Entity Queue)에 등록된다.



<Figure 5> Execution of Link-based Searching

### 3.5 검색결과 구성

LOD 클라우드에서의 심층검색 결과가 효율적으로 활용되기 위해서는 일반적인 LOD와 동일한 형식으로 검색을 요청한 사용자에게 제공되어야 한다. 따라서, 검색결과 데이터집합과 함께 이에 대한 의미적 접근을 가능하도록 하는 온톨로지가 함께 제공되어야 한다. <Figure 6>은 CAIDS가 검색결과를 구성하는 과정을 도시하고 있다.



<Figure 6> Constitution of Searching Results

PS와 LS가 진행한 심층검색을 통하여 여러 LOD들로부터 수집된 RDF 트리플들의 주어개체들은 최초 표층검색에서 검색된 주어개체와 동일한 것으로 평가된 것들이다. 따라서, 트리플 수집기로부터 전달된 RDF 트리플(Triple Collected)들의 집합인 결과트리플 집합(Result Triples' Set)은 최초 검색된 주어개체와 심층검색과정에서 접근한 LOD들로부터 동일한 것으로 파악되어 획득한 주어개체들에 연결된 술어와 목적어들로 구성된다.

결과트리플 집합이 활용되기 위해서는 이를 명세하는 온톨로지가 함께 제공되어야 한다. 이를 위하여, 온톨로지 조립기(Ontology Assembler)는 트리플 수집기로부터 전달받은 RDF 트리플에서 술어를 발췌한 후 해당 트리플이 존재하는 LOD의 온톨로지에서 해당 술어에 명세된 의미와 역할을 수집하고(SPARQL for Predicate's Ontology/SPARQL Result) 이를 조립하여(Ontology Assembled) 결과온톨로지(Results' Ontology)로 구성한다. 이 과정을 보다 구체적으로 설명하자면 발췌한 술어들에 대하여 먼저 심층검색과정에서 접근하였던 LOD들이 공개하고 있는 온톨로지로부터 해당 술어에 대한 온톨로지 명세를 검색한다. <Figure 7>은 술어 <http://ko.dbpedia.org/property/유형>에 대한 온톨

로지를 ko.dbpedia.org의 온톨로지로부터 SPARQL을 통하여 검색한 결과이다. 이러한 방식으로 심층검색 결과에 등장하는 모든 술어들에 대한 온톨로지 명세 내용을 수합하여 결과온톨로지 구성한다.

<http://ko.dbpedia.org/property/유형>	Subject
<http://www.w3.org/2000/01/rdf-schema#label>	Predicate
"유형@ko"	Object
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	Predicate
<http://www.w3.org/1999/02/22-rdf-syntax-ns#Property>	Object

<Figure 7> Fragment of ontology searched with <http://ko.dbpedia.org/property/유형>

본 연구의 구현에서는 이러한 결과트리플 집합과 결과온톨로지를 N-Triple 형식으로 직렬화시켰다. 검색결과는(Result Triples' Set/Results' Ontology) 일반적인 LOD를 접근하는 것과 마찬가지로 차후 SPARQL로 접근하여 다양하게 재사용할 수 있다. 이와 더불어, 저작권이 허용하는 범위 안에서 결과 내용을 수정, 보완하여 새로운 LOD로 구축하고 이를 LOD 클라우드에 공개하는 것도 가능할 것이다.

#### 4. 실험 및 분석

CAIDS에 대한 실험은 DBpedia 프로젝트에 참여하고 있는 LOD들 가운데 한국, 프랑스, 이탈리아, 스페인, 포르투갈의 DBpedia LOD들을(koDBpedia, frDBpedia, itDBpedia, esDBpedia, ptDBpedia) 대상으로 진행하였다. 이들 LOD들의 규모는 <Table 2>와 같다. SPARQL 사용을 위하

여 Apache Jena 3.1.0 API를 사용하였다. LOD마다 연결정책을 부여하였으며 매칭술어간의 유사도는 1.0, 0.9, 0.8, 0.75, 0.73을 적용하였다. 5개 LOD들 가운데 koDBpedia를 표층검색 대상으로 하였으며 ‘배트맨’, ‘아이언맨’, ‘스파이더맨’, ‘분노의 질주’를 검색조건으로 16개 개체들이 koDBpedia로부터 검색되었다. 이들 검색된 개체들로부터 4개 LOD들로 향하는 명시적 연결들은 68개로 파악되었다. 이러한 연결 상황 하에서 실험 및 분석은 정책기반검색(PS) 수행결과가 얼마나 확장된 내용을 제공하는가와 신뢰적인 내용들을 얼마나 포함하고 있는가를 파악하는 것으로 진행하였다. 또한, 연결정책만을 적용하는 경우에서의 검색결과를 명시적 연결들도 함께 적용하는 경우에서의 검색결과와 트리플 수를 기준으로 비교함으로써 개체들간의 연결을 연결정책만으로 파악하는 것이 충분한 규모의 검색결과를 제공하는지를 파악하였다.

<Table 2> Scale of LODs(2017. 12. 21)

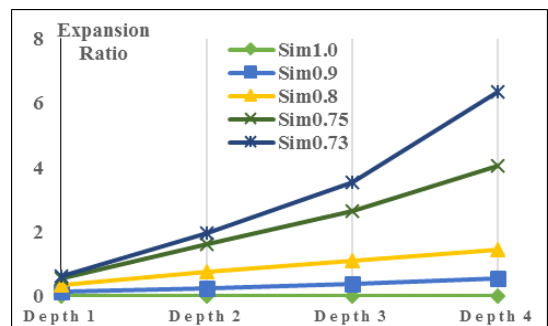
LOD	개체 수
http://ko.dbpedia.org	310,811
http://fr.dbpedia.org	1,591,318
http://it.dbpedia.org	968,794
http://es.dbpedia.org	1,120,144
http://pt.dbpedia.org	865,889

#### 4.1 CAIDS의 확장 규모 평가

확장률은 CAIDS가 연결정책을 이용하여 심층검색을 진행할 경우 <owl:sameAs>를 이용한 명시적 연결만으로 검색되는 개체규모와 비교하여 어느 정도의 비율로 개체들을 검색하는가는 나

타낸다. 이러한 확장률에 대한 각 깊이에서의 계산은 (식1)과 같이 이루어진다.

(식1) 깊이  $i$ 에서의 확장률(Expansion Ratio):  
 (PS와 LS를 함께 수행하여 깊이  $i$ 까지 검색한 개체규모 - LS만을 수행하여 깊이  $i$ 까지 검색한 개체규모) / 깊이  $0$ 에서 명시적연결로 검색된 개체규모 ■



<Figure 8> Expansion Ratio of CAIDS

<Figure 8>에서는 CAIDS의 심층검색 단계에서 PS만을 진행하였을 경우 표층검색 LOD인 koDBpedia에서 검색한 16개 개체들을 기준으로 추가적으로 도출한 개체들의 규모를 깊이 단위에서 유사도별 확장률로 파악하여 제시하였다. 검색한 16개 개체로는 ‘배트맨’, ‘아이언맨’, ‘스파이더맨’, ‘분노의 질주’를 koDBpedia에 질의하여 ‘배트맨\_(1966년\_영화)’, ‘배트맨\_(1989년\_영화)’, ‘배트맨\_2’, ‘배트맨 3:forever’, ‘배트맨\_4: 배트맨과 로빈’, ‘아이언맨(영화)’, ‘아이언맨\_2’, ‘아이언맨\_3’, ‘스파이더맨2’, ‘어메이징\_스파이더맨(영화)’, ‘어메이징\_스파이더맨\_2’, ‘분노의 질주’, ‘분노의 질주 4’, ‘분노의 질주 5’, ‘분노의 질주 6’, ‘분노의 질주 7’를 얻었으며 이들로부터 64개의 명시적 연결들을 확보하였다. 이들을 시작으로 하여 CAIDS에 대한 실험을 진행한 결

과 전체적으로 유사도가 낮을수록, 깊이를 더하여 검색할수록 검색결과와 확장률이 증가하는 것으로 확인되었다. 구체적으로는 유사도 1.0(Sim1.0)에서는 모든 깊이의 LOD들로부터 전혀 확장을 이루지 못하였다. 이는 다음 깊이 LOD로의 연결을 위해서는 유사도가 1.0이므로 매칭술어의 목적어 값들이 완전히 동일하여야 하는데 이러한 조건에서는 확장이 거의 이루어질 수 없기 때문이다. 유사도 0.9(Sim0.9)에서는 매칭술어의 목적어 값들이 유사하여야 하는 것에 대한 요구가 완화됨에 따라 깊이가 더하여지면서 확장률이 조금씩 증가하였다. 유사도 0.8(Sim0.8)부터는 확장률이 더욱 증가하였지만 한편으로는 koDBpedia에서 최초 검색된 개체들과는 무관한 개체들이 일부 등장하기 시작하였다. 유사도 0.75(Sim0.75)부터는 확장률이 크게 증가하여 koDBpedia의 명시적 연결로 획득한 개체들과 상당히 무관한 개체들이 다량 검색결과에 포함되었다. 유사도 0.73(Sim0.73)을 적용할 경우 깊이<sub>4</sub>에서 6.35의 확장률을 보이면서 결과를 더욱 왜곡시키는 것을 확인하였다. 이러한 과도한 확장은 이전단계 깊이에서 낮은 유사도로 검색된 개체에 대하여 낮은 유사도를 재차 적용함에 따른 것으로 판단된다. 따라서, 적절한 확장률을 유지하기 위해서는 0.8 ~ 0.9 정도의 유사도를 적용하여야 할 것으로 판단된다.

#### 4.2 CAIDS의 포함 규모 평가

본 실험에서 표층검색 LOD로 사용된 koDBpedia의 경우 사용자가 이를 직접 선정하여 질의를 요청하였으므로 사용자가 koDBpedia를 충분히 신뢰하고 있는 것으로 전제하였다. 따라서, koDBpedia로부터 나머지 4개 LOD들로 향하

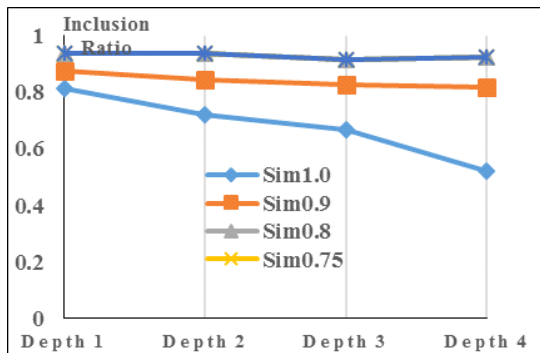
는 명시적 연결들 역시 koDBpedia에 포함된 RDF 트리플들이므로 충분히 신뢰적인 것으로 전제하였다. 따라서, 본 절에서 고찰하는 포함률은 CAIDS가 심층검색을 진행하면서 PS를 실행한 결과가 koDBpedia와 신뢰적으로 연결된 것들로 전제된 개체들을 얼마나 포함하고 있는 것인가를 나타낸다. 각 깊이에서의 포함률은 (식1)과 같이 계산된다.

(식2) 깊이<sub>i</sub>에서의 포함률(Inclusion Ratio):  

$$\frac{LS\text{만을 수행하여 깊이}_i\text{까지 검색한 개체집합} \cap PS\text{만을 수행하여 깊이}_i\text{까지 검색한 개체집합}}{LS\text{만을 수행하여 깊이}_i\text{까지 검색한 개체집합}}$$

<Figure 9>는 포함률을 각 깊이에서 검색된 개체들이 표층검색 LOD인 koDBpedia에서 검색된 개체들과 명시적으로 연결된 개체들을 어느 정도 포함하고 있는가를 유사도별로 계산하여 나타낸다. 유사도 1.0의 경우 심층검색을 깊게 진행할수록 포함률이 빠르게 감소하였다. 이는 매칭술어의 목적어 값들이 완전히 동일한 경우에만 다음 깊이에서 개체를 검색하므로 koDBpedia에서 해당 깊이 LOD로 명시적으로 연결되어 있던 개체들이 PS에서는 검색되지 않는 경우가 빈번하게 발생하기 때문이다. 더욱이, 특정 깊이에서 검색되지 않을 경우 다음 깊이의 연결이 단절되기 때문에 포함률의 감소를 촉진하게 된다. 유사도 0.9에서는 동일조건이 조금 완화되었음에도 불구하고 유사도 1.0과는 확연히 달리 깊이<sub>4</sub>에서도 82% 정도의 포함률을 보이고 있다. 이러한 높은 포함률은 완화된 유사도로 인하여 검색되는 개체들이 증가한 것과 더불어 깊이<sub>i+1</sub> LOD로 연결을 이루는 깊이<sub>i</sub> LOD에서의 개체가 검색되지 않았을 경우에도 그와 유사한 개체가 깊이<sub>i</sub> LOD에서 검색되고

그로부터 깊이  $i+1$  LOD에서의 개체가 검색되는 경우들이 발생하는 것에도 원인이 있다. 유사도 0.8은 유사도 0.9보다 높은 90% 이상의 포함률을 보이고 있다. 유사도 0.75, 073인 경우 유사도 0.8와 비교하여 포함률의 증가가 이루어지지 않고 거의 유사한 결과를 보이고 있다. 하지만, 이들의 경우 전술한 바와 같이 확장이 과도하게 이루어져 최초 검색된 내용과 상이한 결과들이 다수 등장한다. 따라서, 적절한 확장성과 신뢰적 결과들에 대한 적절한 포함성을 함께 얻기 위해서는 유사도를 0.8 ~ 0.9 정도로 설정하는 것이 바람직한 것으로 판단된다.

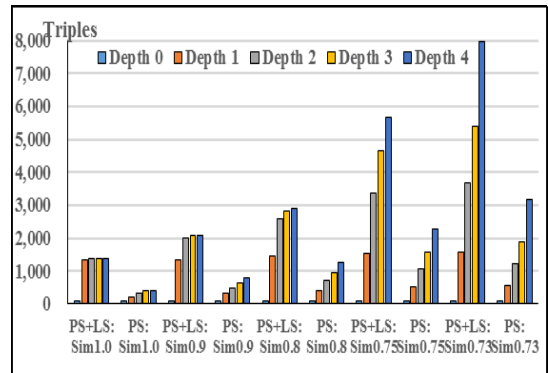


(Figure 9) Inclusion Ratio of CAIDS

### 4.3 CAIDS의 검색 트리플 규모 평가

CAIDS의 검색규모는 명시적 연결을 이용한 연결기반검색(LS)을 운영하지 않고 정책기반검색만(PS) 운영하여 검색한 개체들의 트리플 수와 PS, LS를 함께(PS+LS) 운영하여 검색한 개체들의 트리플 수를 비교하여 파악하였다. 즉, 연결정책만으로 CAIDS를 운영할 경우 명시적 연결들이 제공되는 상황에서 CAIDS를 운영하는 경우 대비 어느 정도 규모의 검색결과를 제공할 수

있는가를 <Figure 10>에서 제시하였다. 유사도 1.0에서 PS+LS는 깊이\_1로 검색을 진행하면서 트리플 수가 급격하게 증가한 후 다음 깊이의 LOD들에서는 거의 증가하지 않는 양상을 보인다. 이는 DBpedia 프로젝트의 특징이 반영된 것으로서 DBpedia LOD들은 명시적 연결들의 대부분을 LOD 공개자가 다른 DBpedia LOD들로 직접 생성하였기 때문에 각 LOD에서의 명시적 연결대상 규모가 거의 동일하다. 또한, 유사도 1.0에서는 매칭술어의 목적어들이 완전히 일치하는 경우에만 연결이 이루어지고 정책적 연결로 검색된 트리플들도 대부분이 명시적 연결을 통한 검색결과에 이미 포함되어 있다는 것도 트리플 증가가 거의 없는 것에 주요 원인이 된다. 유사도 1.0에서 PS만을 실행하였을 경우 깊이를 더할수록 검색 트리플 수가 증가하지만 유사도를 낮춘 경우와 비교해서는 낮은 증가 추이를 보이고 있다.



(Figure 10) Triples from CAIDS

유사도를 0.9, 0.8로 완화시키면 PS+LS인 경우와 PS만을 수행한 경우 모두 깊이를 더할수록 검색 트리플 수가 증가하는데 유사도 완화에 더 많

은 비율로 영향을 받는 PS가 PS+LS보다 상대적으로 높은 증가 추이를 보이고 있다. 특히, PS+LS에서는 깊이\_1에서 깊이\_2로 진행하는 경우 보다 깊이\_2에서 깊이\_3, 4로 진행하면서 검색된 트리플 수의 증가 추이가 더 완만해지는 것은 깊이를 진행할수록 중복된 트리플들이 다수 검색되어 이들을 결과에서 제거하였기 때문이다. 유사도 0.75, 0.73에서는 깊이\_3, 4부터 트리플들이 과도하게 검색되고 있다.

한편, PS만을 진행하는 경우 연결정책으로 연결된 5개 LOD들에 국한하여 검색이 이루어지지만 PS+LS를 진행하는 경우 연결정책에 명시된 LOD들 외에 명시적연결로 접근하게 되는 LOD들도 존재함에 따라 깊이\_1부터 개체별로 15개나 16개 LOD들이 검색에 참여하는 것으로 파악되었다. 따라서, PS만 진행하는 경우 참여하는 LOD 수는 PS+LS 경우와 비교할 경우 깊이\_1, 2, 3, 4에서 각각 13%, 19%, 25%, 31% 정도이다. 각 깊이에서 PS+LS가 각 깊이별로 해당 비율만큼의 LOD들에서 수행하였다면 유사도 1.0, 0.9, 0.8 경우에서의 평균을 기준으로 분석한다면 PS 경우가 PS+LS 경우보다 깊이\_1, 2, 3, 4에서 70%, 32%, 24%, 20% 정도 더 많은 트리플들을 검색할 것으로 예상된다. 따라서, LOD 클라우드에서 대규모의 명시적 연결들을 생성, 관리하는데 많은 어려움이 있는 현재와 같은 상황을 극복하기 위하여 본 논문에서 제시하듯이 LOD 마다 연결정책만을 수립하고 이를 검색시점에서 참조하도록 하는 환경을 구축할 경우 상대적으로 적은 부담으로도 충분한 규모의 동일연결들을 생성하여 보다 풍부한 검색결과를 획득하는 것이 가능할 것으로 판단된다.

#### 4.4 CAIDS의 확장 신뢰수준 평가

CAIDS 실행으로 획득한 확장된 결과는 깊이\_0의 LOD에서 검색된 개체들과 동일한 것으로 파악된 개체들에 대하여 기술된 내용으로 구성된다. 더불어, LOD를 구성하는 내용은 주장된 사실(Assertion)이라는 근본적인 특성을 가지고 있다. 따라서, CAIDS의 PS 혹은 LS가 진행되는 과정에서 연결을 통한 검색확장을 위하여 사용하는 “개체\_A와 개체\_B가 동일하다”는 내용 또한 주장이라는 근본적인 특성을 가지고 있다. 주장은 완전히 신뢰적인 사실인 공리(Axiom)와는 달리 해당 내용의 신뢰수준에 대한 추가적인 평가가 필요하다. 이에 본 절에서는 CAIDS가 동일한 것으로 파악한 개체들간의 동일성에 대한 신뢰수준을 0부터 1까지 값으로 평가하였다.

우선 사용자의 SPARQL 검색요구가 최초 접근하는 깊이\_0의 표층검색 LOD에서 획득되는 개체들의 신뢰수준은 1로 평가하였다. 즉, 사용자가 특정 LOD를 충분히 신뢰하여 검색대상으로 직접 선정한 것이었으므로 해당 LOD는 충분히 신뢰적인 내용들을 제공하는 것으로 전제하였다. 심층검색이 이루어지는 깊이\_1부터의 LOD들은 사용자가 전혀 인지하지 못하였던 것들이므로 이들로부터 검색된 개체들이 깊이\_0에서 검색한 개체들과 얼마만큼이나 동일한 것인가에 대한 신뢰수준에 대해서는 보다 구체적으로 고찰할 필요가 있다. 이를 동일수준이라고 하고 아래와 같이 정의하였다.

**(정의1) 동일수준(Identity Ratio):** 심층검색 과정에서 깊이\_i의 LOD<sub>i</sub>에서 검색된 개체\_i가 표층검색으로 깊이\_0의 LOD<sub>0</sub>에서 검색된 개체\_A와 동일하다는 것에 대한 신뢰 정도 ■

예를 들어 연결트리플 {개체\_A owl:sameAs

개체\_B}, {개체\_B owl:sameAs 개체\_C}가 존재하고 개체\_A는 깊이\_0의 LOD<sub>A</sub>에 개체\_B와 개체\_C는 깊이\_1의 LOD<sub>B</sub>, 깊이\_2의 LOD<sub>C</sub>에 저장되어 있을 때 개체\_B와 개체\_C에 대하여 이들이 각각 어느 정도로 표층검색으로 검색된 개체\_A와 동일한 것인가를 동일수준으로 표현한다. 또한, 개체간의 동일성 주장은 단방향으로 이루어진다. 즉, {개체\_A owl:sameAs 개체\_B}라는 사실은 이 내용을 주장하는 측(LOD<sub>A</sub> 공개자)의 상황이며 이것이 {개체\_B owl:sameAs 개체\_A}를 추론이 가능하다고 의미하는 것은 아니다. 단방향성이 적용됨에 따라 {개체\_A owl:sameAs 개체\_B}에서 개체\_A는 소스동일개체가 되고 개체\_B는 타겟동일개체가 된다. 두 개체가 동일하다는 주장은 LS가 처리하는 명시적 연결트리플뿐만 아니라 PS의 진행에 따른 정책적 연결을 통하여도 이루어지며 이들에 대해서도 본 절의 실험을 통하여 동일수준을 평가하였다.

이와 더불어 “LOD<sub>B</sub>에 저장된 개체\_B가 LOD<sub>C</sub>에 저장된 개체\_C와 동일하다”는 주장은 LOD<sub>B</sub> 공개자에 의하여 이루어진 것이다. 따라서, LOD<sub>B</sub>에 저장된 모든 주장된 사실들에는 LOD<sub>B</sub>의 신뢰도가 반영되어야 하며 “개체\_B가 LOD<sub>C</sub>에 저장된 개체\_C와 동일하다”는 주장도 이에 포함된다. 따라서, 개체\_C가 개체\_A와 어느 정도로 동일한 것인가를 의미하는 개체\_C의 동일수준은 이러한 개체\_B가 개체\_A와 어느 정도로 동일한 것인가를 의미하는 개체\_B의 동일수준에 “개체\_B가 LOD<sub>C</sub>에 저장된 개체\_C와 동일하다”는 주장을 한 LOD<sub>B</sub>의 신뢰도가 반영되어야 한다. “LOD<sub>B</sub>에 저장된 개체\_B가 LOD<sub>C</sub>에 저장된 개체\_C와 동일하다” 주장에 따른 타겟동일개체인 개체\_C의 동일수준은 (식3)으로 평가된다.

(식3) 타겟동일개체 개체\_C의 동일수준: 소스

동일개체 개체\_B의 동일수준 × 소스동일개체를 포함하고 있는 LOD<sub>B</sub> 신뢰도 ■

단, 전술된 논거에 의거하여 표층검색 대상인 LOD<sub>A</sub>의 신뢰도는 1로 전제하며 LOD<sub>A</sub>에서 등장하는 동일성 주장들에서의 소스동일개체들의 동일수준도 1로 전제한다. (식3)에 의거하여 심층검색의 첫 LOD인 깊이\_1에서의 모든 소스동일개체들의 동일수준들도 1이 된다.

심층검색이 이루어지는 LOD에 대한 신뢰도는 구글의 페이지랭킹 전략과 유사하게 해당 LOD를 타겟으로 하는 동일성 주장에 따른 진입연결이 많을수록 높은 것으로 평가하였다. 하지만, 현재 LOD 클라우드에서는 각 LOD에 대한 진입연결의 규모를 파악하는 것이 불가능한 상황이다. 따라서, 본 실험에서는 사용자 질의를 심층검색으로 진행하는 과정에서 진입연결이 많이 이루어지는 개체를 저장하고 있는 LOD가 해당 검색요청에 대하여 높은 신뢰도를 가지는 것으로 하였다. 즉, 많은 LOD로부터 동일연결을 통한 참조가 이루어지는 경우 해당 LOD의 신뢰도를 높게 평가하는 전략을 채택하였다. LOD에 대한 신뢰도는 (식4)에 제시하였다.

(식4) 계산대상 LOD의 신뢰도(LOD Confidence):  $(\text{InComming}_{\text{cur}} - \text{InComming}_{\text{min}}) / (\text{InComming}_{\text{max}} - \text{InComming}_{\text{min}})$

InComming<sub>max</sub>: 현재 검색된 개체들 가운데 진입연결이 가장 많은 개체의 진입연결 수

InComming<sub>min</sub>: 현재 검색된 개체들 가운데 진입연결이 가장 적은 개체의 진입연결 수

InComming<sub>cur</sub>: 현재 신뢰도 계산 대상 LOD에 저장된 타겟개체로 향하는 진입연결 수 ■

LOD의 신뢰도는 0.7 ~ 1.0 으로 정규화하였다. 이는 일반적인 웹과는 달리 LOD 클라우드에는 개별 개체에 대한 상세 내용을 지식명세인 온

톨로지를 기반으로 구성된 LOD를 기관들이 구축하여 공개하였으므로 이들의 신뢰수준은 0.7 이상이 되는 것으로 전제하였다.

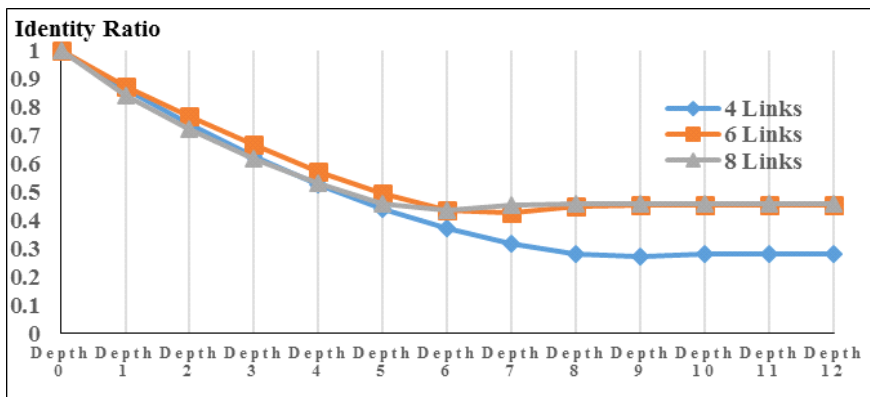
CAIDS가 심층검색을 진행하면서 도출한 개체들에 대해서는 (식3)과 (식4)를 적용함에 따라 동일수준이 전파되어 LOD가 깊이\_0에서 멀어질수록 동일수준이 낮게 평가되었다. 이를 확인하기 위하여 본 절의 실험에서는 100개의 LOD로 구성된 가상의 LOD 클라우드를 구축하였다. 각 LOD는 1,000개의 개체를 저장하여 총 100,000개의 가상 데이터를 구성하였으며 각 개체마다 타겟동일개체로의 연결이 4개, 6개, 8개인 경우로 실험을 진행하여 <Figure 11>과 같은 동일수준 변화를 확인하였다. 동일수준은 깊이를 진행할수록 감소하다가 일정 깊이 이후부터는 일정하게 유지되는 것으로 나타났다. 이는 타겟동일개체로 전파되는 동일수준이 해당 개체가 기존에 가지고 있던 동일수준보다 낮을 경우 전파된 동일수준을 수용하지 않고 기존의 동일수준을 유지하기 때문이었다.

<Figure 11>의 실험상황과는 달리 CAIDS가 심층검색을 진행하는 과정에서 이미 검색하였던

개체에 대한 새로운 동일연결들이 등장하는 경우가 빈번하게 있다. 본 절에서는 이러한 경우를 동일동조연결이라고 정의하고 이를 실험에 반영하였다.

**(정의2) 동일동조연결(Identity Agreement Link):** 심층검색으로 새로이 검색된 개체를 소스 동일개체로 하여 이미 검색되었던 개체를 타겟 동일개체로 연결 주장 ■

예를 들어 {개체\_A <owl:sameAs> 개체\_B}, {개체\_B<owl:sameAs> 개체\_C}, {개체\_C <owl:sameAs> 개체\_D}의 주장들은 개체\_A에서 개체\_B, 개체\_C, 개체\_D로의 단방향 연결을 이룬다. 이러한 연결은 이행성에 따른 {개체\_A=개체\_B, 개체\_A=개체\_C, 개체\_A=개체\_D}라는 추론결과를 제공한다. 이 상황에서 {개체\_D <owl:sameAs> 개체\_B}라는 동일동조연결이 등장할 경우 개체\_B, 개체\_C, 개체\_D의 연결은 사이클 구조를 가지게 된다. 이러한 사이클 구조는 개체\_B, 개체\_C, 개체\_D가 개체\_A와 동일하다는 것에 대하여 추가적인 동조가 이루어진 것이다. 이러한 동일동조연결로 인하여 개체\_A에서 개체\_D까지의 동일성이 반복적으로 추론되므로



<Figure 11> Identity Ratio of CAIDS' results



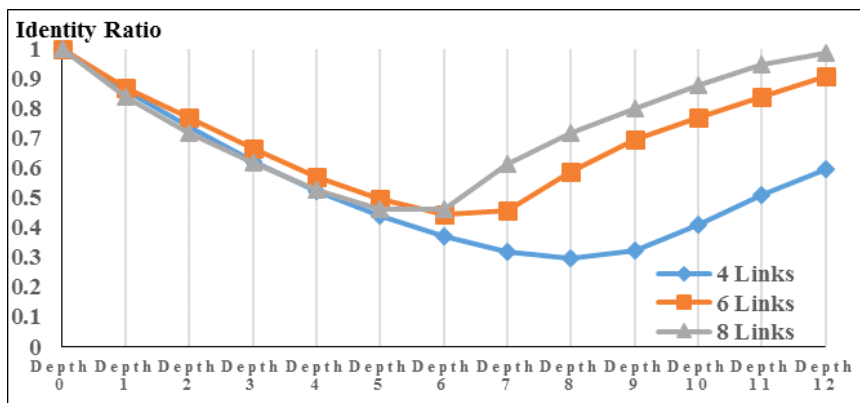
동조에 참여하는 개체\_B, 개체\_C, 개체\_D에 대한 동일수준도 보완되어야 한다. 이를 위하여 본 절의 실험에서는 개체\_D의 동일수준을 동일동조로 연결되는 개체\_B에 전파시킴으로써 개체\_B의 동일수준을 보완한다. 개체\_B의 동일수준이 보완됨에 따라 개체\_C와 개체\_D의 동일수준도 갱신된다. 단, 개체\_C와 개체\_D의 경우 새로이 전파되는 동일수준이 기존에 가지고 있던 동일수준보다 높을 경우에만 이를 새로운 동일수준으로 수용한다.

동일동조연결을 포함하는 경우에서의 동일수준의 변화를 <Figure 12>에서 제시하였다. 동일연결의 수와는 무관하게 LOD 신뢰도 감소에 따라 지속적으로 감소하던 개체의 동일수준이 심층검색을 어느 정도 깊이 이상 진행하면 다시 증가하는 모습을 보였다. 동일수준의 복귀는 동일연결의 수가 많을수록 보다 이른 깊이에서 시작된다. 특히, 동일연결이 8개인 경우 깊이\_12까지 심층검색을 진행할 경우 표층검색으로 확보한 개체와 거의 동일한 수준의 신뢰도를 제시하였다. 깊이\_9에서 동일연결이 4일 경우 0.32에 머무른 동일수준이 동일연결이 6이나 8인 경우

각각 0.69, 0.80에 이르고 있음에 따라 확장된 내용에 대하여 충분히 신뢰를 가지고 사용하는 것이 가능하게 된다. 따라서, LOD에 참여하는 개체들에게 다른 LOD에 저장된 개체들에 대한 동일연결을 8개 이상을 부여할 경우 사용자들로 하여금 LOD 클라우드로부터 확보한 확장된 지식들에 대하여 충분히 신뢰하면서 활용하는 것이 가능하도록 할 수 있을 것이다. 본 논문에서 제시한 연결정책을 통한 동일성 제공은 이러한 동일연결들을 수월하게 부여하는 방안이 될 것이다.

## 5. 결론 및 향후 연구

본 연구에서는 LOD 클라우드에서의 동일개체에 대한 연결을 <owl:sameAs>를 이용하여 명시적으로 기술하는 기존방식 대신 LOD 마다 연결정책을 수립하여 LOD와 함께 공개할 것을 제안하였다. 기존의 명시적 연결을 이용한 LOD 클라우드로부터의 검색은 변화된 내용을 검색결과에 충실하게 반영하지 못한다는 한계가 있다. 연



<Figure 12> Identity Ratio of CAIDS' results with Identity Agreement Links

결정책을 운영할 경우 검색시점에 연결정책을 참조하여 동일개체들을 검색함에 따라 명시적 연결이 가지는 한계를 근본적으로 해결할 수 있다. 이러한 연결정책이 RDF 모델에 입각하여 명세될 수 있도록 하게 위하여 본 연구에서는 연결정책 명세어휘 집합을 제시하였다. 연결정책에는 소스와 타겟 LOD의 개체들간 동일성을 평가하는데 사용할 소스, 타겟 술어쌍들이 명세된다. 검색이 진행되는 시점에서 연결정책에 명세된 소스, 타겟 술어쌍에 결부되는 소스, 타겟 목적어들의 유사도를 측정함으로써 소스, 타겟 주어 개체들의 동일여부를 판단하고 이를 통하여 최초 검색된 주어 개체에 대한 다양한 관점의 상세내용들을 검색하는 시스템 CAIDS를 구현하였다. 본 연구는 LOD 클라우드에서 연결정책을 검색시점에서 참조하여 검색에 적용하는 최초의 시도이며 CAIDS 구현을 통하여 이러한 방식이 실제 운영 가능한 것임을 확인하였다.

CAIDS에 대한 실험은 가장 성공적인 LOD 프로젝트로 평가되는 DBpedia에 참여하는 LOD들 가운데 5개 국가(한국, 프랑스, 이탈리아, 스페인, 포르투갈) LOD를 대상으로 하였다. 실험결과 소스와 타겟 목적어들간의 유사수준을 0.8 ~ 0.9로 요구할 경우 검색결과가 적정하게 확장되고 신뢰가 확인된 개체들이 충분히 포함되는 것으로 확인하였다. 이와 더불어, 연결정책만으로 CAIDS를 운영하는 경우에도 충분한 규모의 동일개체들을 획득할 수 있는 것으로 확인되었다. LOD 클라우드가 풍부한 연결성을 제공할 경우 검색으로부터 획득하는 내용에 대한 신뢰를 확인하는데 도움이 된다. CAIDS에 대한 실험을 통하여 개체별로 8개 이상의 동일연결들을 부여하는 것이 검색결과의 신뢰확인에 도움이 되는 것으로 파악되었다. 실험에서 확인된 목적어들간

의 적정한 유사수준과 개체별 동일연결 규모는 LOD 클라우드에 참가할 LOD를 구축하는데 주요한 기준점이 될 것으로 판단된다. 본 연구에서 제안한 LOD별 연결정책 수립방안과 연결정책을 이용한 심층검색 시스템 구현은 LOD 마다 <owl:sameAs>로 방대한 규모의 명시적 연결들을 생성하고 이를 갱신관리하여야 하는 부담 없이도 효율적으로 LOD 클라우드에서의 연결성을 풍부하게 제공하는 환경을 구축하는데 기여할 수 있을 것으로 사료된다.

향후 연구에서는 연결정책에 명세된 술어들의 문법적 특성을 개체동일성 평가에 반영할 수 있는 방안을 고찰하고자 한다. 또한, 연결정책으로부터 동일한 것으로 판단된 개체들에 대한 상세기술내용을 검토하여 이들 개체들이 결코 동일할 수 없는 경우와 동일한 것이 명확한 경우를 구분함으로써 검색결과를 정제하는 방안을 마련하고자 한다. 이를 위해서는 상세내용을 구성하는 주어, 술어, 목적어에 적용된 어휘들의 문법 구조에 대한 분석이 중요한 역할을 할 것으로 예상된다. 이와 더불어, CAIDS 실험에 적용한 동일동조연결 개념을 CAIDS의 기능으로 추가하여 사용자로 하여금 확장된 검색결과들 가운데 기재된 신뢰기준에 부합하는 결과만을 최종적으로 획득할 수 있도록 할 계획이다.

## 참고문헌(References)

- Abele A. and McCrae J., *The Linked Open Data cloud diagram*, 2017. Available at <http://lod-cloud.net/> (Downloaded 21 December, 2017)
- Auer, S., et al., T., *Linked open data - creating*

- knowledge out of interlinked data*, Springer, U.S.A., 2014
- Bill, C. and Mary, K., *Towards a semantic web*, Chandos Publishing, U.K., 2011
- Bizer C., *Is the Semantic Web what we expected*, 2017. Available at <https://www.slideshare.net/bizer/is-the-semantic-web-what-we-expected-a-doption-patterns-and-contentdriven-challenges-iswc-2016-keynote> (Downloaded 21 December, 2017)
- Bob, D., *Learning SPARQL*, O'REILLY, U.S.A., 2013
- Brown, Peter F., et al, "Class-based n-gram models of natural language" *Computational linguistics*, Vol.18, No.4(1992), 467~479.
- Carol, G. and Shenghui, W., *Library linked data in the cloud*, Morgan & Claypool Publishers, U.S.A., 2015
- Choi, Y., Park, J., "The Need for Paradigm Shift in Semantic Similarity and Semantic Relatedness : From Cognitive Semantics Perspective", *Journal of Intelligence and Information Systems*, Vol.19, No.1(2013), 111~123.
- David, W., et al., *Linked data: A Structured Data on the Web*, Manning Publication, U.S.A., 2014
- Dean, A. and James, H., *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*, Elsevier, 2011
- Erik, M., *Library Linked Data: Research and Adoption*, ALA TechSource, U.S.A., 2014
- Frank, C. and Lars, S., *Linked Data and User Interaction*, ILFA Publications, 2016
- Grigoris, A., et al., *A Semantic Web Primer*, MIT Press, 2012
- Hart, G. and Catherine, D., *Linked data: A Geographic Perspective*, CRC Press, U.S.A., 2013
- Harth, A., et al., *Linked Data Management*, 1<sup>st</sup> Ed., 20-25. CRC Press, U.S.A., 2014
- Heath, T. and Bizer, C. *Linked Data: Evolving the Web into a Global Data Space*, Morgan & Claypool, U.S.A., 2011.
- Hitzler P., et al., *Foundation of Semantic Web Technologies*, CRC Press, U.S.A., 2009
- Jeff, Z. and Guido, V., *Exploiting Linked Data and Knowledge Graphs in Large Organizations*, Springer, Switzerland, 2017
- Jeong, H., "A Study on Ontology and Topic Modeling-based Multi-dimensional Knowledge Map Services", *Journal of Intelligence and Information Systems*, Vol.21, No.4(2015), 79~92.
- Konstantinou, N., *Materializing the Web of Linked Data*, 1<sup>st</sup> Ed., 51. Springer, U.S.A., 2015
- Michael, D., *The Great Cloud Migration: Your Roadmap to Cloud Computing, Big Data and Linked Data*, Outskirts Press, U.S.A., 2013
- Martin H., et al., *Ontology Management: Semantic Web, Semantic Web Services, and Business Applications (Semantic Web and Beyond)*, Springer, U.S.A., 2007
- Ngonga, A. and Auer, S., "LIMES - A Time-Efficient Approach for Large-Scale Link Discovery on the Web of Data", *Proc. of the 22<sup>nd</sup> IJCAI*, (2011), 2312~2317.
- Park, H. J., Lee, H. J., Kim, J. W., "Participation Level in Online Knowledge Sharing: Behavioral Approach on Wikipedia", *Journal of Intelligence and Information Systems*, Vol.19, No.4(2013), 97~121.

- Park J. and Sohn, Y., “A Syntax Added Link Evaluation Technique for Improving Trustworthiness of LOD’s Linkages”, *Journal of KIISE: Databases*, Vol. 41, No. 1 (2014), 45~61.
- Park J. and Sohn, Y., “Trustworthiness Improving Link Evaluation Technique for LOD Linkages giving Considerations to the Syntactic Properties of RDFS, OWL, and OWL2”, *Journal of KIISE: Databases*, Vol. 41, No. 4 (2014), 226-241.
- Robert, A. and Barry S., *Ontologies with Basic Formal Ontology*, MIT Press, 2015
- Schmachtenberg, M. and Bizerand, H., *State of the LOD Cloud*, 2017, Available at <http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/> (Downloaded 21 December, 2017)
- Sikos, F., *Mastering Structured Data on the Semantic Web*, Apress, U.S.A., 2015
- Volz J., et al., “Silk - A Link Discovery Framework for the Web of Data”, *Proc. of the 2<sup>nd</sup> Workshop on Linked Data on the Web 2009 (LDOW2009)*, (2009), 238~247.
- W3C, *What is Linked Data*, 2017, Available at <https://www.w3.org/standards/semanticweb/data> (Downloaded 21 December, 2017).

Abstract

## Change Acceptable In-Depth Searching in LOD Cloud for Efficient Knowledge Expansion

Kwangmin Kim\* · Yonglak Sohn\*\*

LOD(Linked Open Data) cloud is a practical implementation of semantic web. We suggested a new method that provides identity links conveniently in LOD cloud. It also allows changes in LOD to be reflected to searching results without any omissions. LOD provides detail descriptions of entities to public in RDF triple form. RDF triple is composed of subject, predicates, and objects and presents detail description for an entity. Links in LOD cloud, named identity links, are realized by asserting entities of different RDF triples to be identical. Currently, the identity link is provided with creating a link triple explicitly in which <owl:sameAs> associates its subject and object with source and target entities. Link triples are appended to LOD. With identity links, a knowledge achieves from an LOD can be expanded with different knowledge from different LODs. The goal of LOD cloud is providing opportunity of knowledge expansion to users.

Appending link triples to LOD, however, has serious difficulties in discovering identity links between entities one by one notwithstanding the enormous scale of LOD. Newly added entities cannot be reflected to searching results until identity links heading for them are serialized and published to LOD cloud. Instead of creating enormous identity links, we propose LOD to prepare its own link policy. The link policy specifies a set of target LODs to link and constraints necessary to discover identity links to entities on target LODs. On searching, it becomes possible to access newly added entities and reflect them to searching results without any omissions by referencing the link policies. Link policy specifies a set of predicate pairs for discovering identity between associated entities in source and target LODs. For the link policy specification, we have suggested a set of vocabularies that conform to RDFS and OWL. Identity between entities is evaluated in accordance with a similarity of the source and the target entities' objects which have been associated with the predicates' pair in the link policy.

---

\* AI Labs, Saltlux

\*\* Corresponding Author: Yonglak Sohn

Department of Computer Engineering, Seokyeong University,

124 Seogyong-ro Seongbuk-gu, Seoul, 02173, Korea

Tel: +82-2-940-7741, Fax: +82-2-940-7009, E-mail: syl@skuniv.ac.kr

We implemented a system “Change Acceptable In-Depth Searching System(CAIDS)”. With CAIDS, user’s searching request starts from depth\_0 LOD, i.e. surface searching. Referencing the link policies of LODs, CAIDS proceeds in-depth searching, next LODs of next depths. To supplement identity links derived from the link policies, CAIDS uses explicit link triples as well. Following the identity links, CAIDS’s in-depth searching progresses. Content of an entity obtained from depth\_0 LOD expands with the contents of entities of other LODs which have been discovered to be identical to depth\_0 LOD entity. Expanding content of depth\_0 LOD entity without user’s cognition of such other LODs is the implementation of knowledge expansion. It is the goal of LOD cloud. The more identity links in LOD cloud, the wider content expansions in LOD cloud. We have suggested a new way to create identity links abundantly and supply them to LOD cloud.

Experiments on CAIDS performed against DBpedia LODs of Korea, France, Italy, Spain, and Portugal. They present that CAIDS provides appropriate expansion ratio and inclusion ratio as long as degree of similarity between source and target objects is 0.8 ~ 0.9. Expansion ratio, for each depth, depicts the ratio of the entities discovered at the depth to the entities of depth\_0 LOD. For each depth, inclusion ratio illustrates the ratio of the entities discovered only with explicit links to the entities discovered only with link policies. In cases of similarity degrees with under 0.8, expansion becomes excessive and thus contents become distorted. Similarity degree of 0.8 ~ 0.9 provides appropriate amount of RDF triples searched as well.

Experiments have evaluated confidence degree of contents which have been expanded in accordance with in-depth searching. Confidence degree of content is directly coupled with identity ratio of an entity, which means the degree of identity to the entity of depth\_0 LOD. Identity ratio of an entity is obtained by multiplying source LOD’s confidence and source entity’s identity ratio. By tracing the identity links in advance, LOD’s confidence is evaluated in accordance with the amount of identity links incoming to the entities in the LOD. While evaluating the identity ratio, concept of identity agreement, which means that multiple identity links head to a common entity, has been considered. With the identity agreement concept, experimental results present that identity ratio decreases as depth deepens, but rebounds as the depth deepens more. For each entity, as the number of identity links increases, identity ratio rebounds early and reaches at 1 finally. We found out that more than 8 identity links for each entity would lead users to give their confidence to the contents expanded. Link policy based in-depth searching method, we proposed, is expected to contribute to abundant identity links provisions to LOD cloud.

**Key Words** : Semantic Web, In-depth searching, Ontology, Linked Open Data, Link policy

Received : January 15, 2018 Revised : June 18, 2018 Accepted : June 23, 2018

Publication Type : Regular Paper Corresponding Author : Yonglak Sohn

## 저 자 소개



**김 광 민**

현재 (주)솔트룩스 인공지능연구센터 연구원으로 재직 중이다. 서경대학교 컴퓨터공학과에서 학사와 전자컴퓨터공학 대학원에서 석사를 취득하였다. 주요 관심분야로 Linked Open Data, 온톨로지, 머신러닝, 빅데이터 등이 있다.



**손 용 락**

현재 서경대학교 컴퓨터공학과 교수로 재직중이다. 경북대학교 전자공학과 전산전공으로 학사, 고려대학교 전자및전산학과 정보처리전공으로 석사, KAIST 정보및통신공학과에서 박사를 취득하였다. (주)데이콤에서 친리안II를 개발하였고 미국 Purdue 대학교 전산학과와 보안연구센터 CERIAS에 방문학자로 활동하였다. 주요 연구분야는 Linked Open Data, 온톨로지, 데이터모델링, 정보이론, 정보보안, 트랜잭션관리, 회복관리 등이다.