

빅데이터 기반 대용량 시맨틱 웹 검색 기술 동향

윤석찬, 남궁현*, 양성권**, 김홍기***

다음커뮤니케이션, *한국전자통신연구원, **솔트룩스, ***서울대학교

요약

시맨틱 웹 기술은 웹의 초창기부터 다양한 연구와 표준이 개발되었지만 이를 활용한 데이터 서비스 분야는 그 역사에 비해 성공 사례가 부족한 것이 현실이다. 최근 웹 2.0을 시초로 링크드 데이터의 성장, 정부의 개방형 데이터 서비스, 소셜 웹 서비스의 등장으로 인해 웹의 구조적 데이터는 폭발적으로 성장해 왔으며, 대용량 시맨틱 웹 기반 서비스에 대한 요구와 연구가 진행되고 있다. 본 고에서는 킬러 애플리케이션으로서 기존 시맨틱 웹 기반 검색 기술의 문제점들을 알아보고 이를 해결하기 위해 최근 화두로 떠오르는 빅데이터(Big Data) 기술 요소인 하둡(Hadoop) 및 NoSQL을 활용하여 대용량 시맨틱 웹 데이터를 활용한 Daum의 영화/음악/인물 기반 의미 검색 및 의학 LOD를 기반한 검색 서비스 개발 사례를 제시한다. 이를 토대로 이중 모델 데이터간 연결 및 실시간 데이터 리비전 관리 등 한계점들을 살펴보고 향후 대용량 공공 데이터 활용을 위한 방향을 모색해 본다.

바로 간단한 마크업과 이를 복사 및 붙여넣기를 통해 배우고 재생산 할 수 있었다는 점이 크다. 특히, 블로그와 위키 같은 다양한 참여형 애플리케이션이 가능하다. 개발자의 경우도 기존의 서비스 지향 구조(SOA) 기반 기술 즉, WSDL이나 SOAP이나 웹 서비스(Web Services)가 성공하지 못하고 RESTful한 API를 통해서 성공한 오픈 API가 이를 반증한다. 사용하기 어려운 기술은 문제다. 두번째 이유는 킬러 애플리케이션이 없다는 점이다. 기존의 인터넷 산업에서 성공했던 포털, 포럼, 블로그, 사진, 동영상, 검색 등에 비해 성공한 서비스 모델이 나오지 않았다는 점이다. 세번째는 시맨틱 웹은 몇몇 도메인에서만 발전하게 되었다는 점이다. 문헌 정보에서의 서지 정보, 의료 정보 및 영화나 음악 같은 콘텐츠 정보 즉, 기존의 관계형 데이터베이스로 이미 구축되고 있었던 정보들을 시맨틱 웹으로 전환하는 연구에 불과했다.

본 고에서는 이러한 시맨틱 웹의 문제점들을 해결하고 새로운 서비스 모델을 만들어 내기 위해 최근 화두로 떠오르는 빅데이터(Big Data) 기술을 이용한 대용량 분산 처리 기법과 공공 데이터 활용이라는 두 가지 관점에서 현황과 발전 방향을 모색해 본다.

I. 서론

2006년부터 전 세계적으로 웹 2.0(Web 2.0)에 대한 붐이 일어났다. 웹 2.0은 플랫폼으로서 웹이라는 모토에 따라 발전을 해왔으며, 웹을 정보 매개의 수단이 아닌 애플리케이션 플랫폼으로 새로운 조명을 하게 되었다. 이 때문에 많은 사람들이 차세대 웹 기술에 대해 많은 관심을 기울였다. 소위 웹 3.0을 클라우드 소싱(Crowdsourcing) 및 시맨틱 웹의 데이터 구조화라고 이야기하는 경우도 있다. 이를 위해 시맨틱 웹 기술을 이용하여 많은 업체들이 새로운 기술을 내놓고 있다 [1].

사실 1998년부터 시맨틱 웹 분야에 다양한 웹 표준이 발전했고 다양한 연구가 진행되었다. 그러나 이에 대한 성공 사례가 나오지 않았다는 점에는 모두 동의하고 있다. 첫번째 이유는 기술이 여전히 어렵다는데 있다. 웹이 발전했던 가장 큰 요인이

II. 시맨틱 웹 검색의 현황

오픈 API를 통한 웹의 플랫폼적인 성장은 반구조적(Semi-structured) 데이터를 생산하는데 큰 기여를 하였다. 이어서 트위터와 페이스북 등의 소셜웹 서비스는 다양한 사람과 데이터를 연결하고 이를 공개적으로 제공하면서 질적 양적 성장을 이루어냈다.

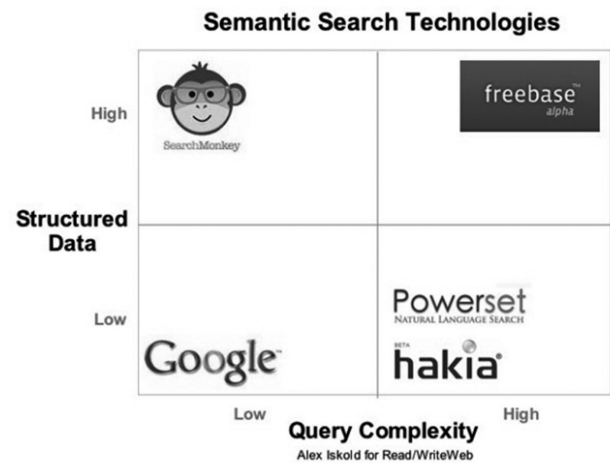
그러다 보니 국내에서도 네이트의 시맨틱 검색을 비롯하여 몇몇 검색 엔진 업체들이 소셜 분석을 가미한 시맨틱 검색 서비스를 내놓고 있다. 이들 모두 기존 데이터에서 구문이나 문장 분석에서 중요 주제를 추출하고, 이에 대한 관련 주제를 연결하여 시각화하고 전통적인 자연어 처리 기법을 사용한다. 블로그,

게시판과 같은 구조화되지 않은 텍스트를 대상으로 주제 분류와 예상 답변을 제시하는 방식으로 그 뼈대는 일반적인 텍스트 기반 정보 검색(IR) 기법을 기반으로 하고 있다.

이는 정보 검색에 대한 가장 일반적인 분야로서 텍스트 분석, 자연어 처리, 기계 학습과 같은 분야는 인공 지능 분야에서 시맨틱 기술을 말하는 것이다. 하지만, 시맨틱 웹은 이와 완전히 다른 접근이다. 시맨틱 웹은 웹을 데이터 단위로 구조화 시켜 상호 관계성을 파악하려는 시도 즉, 웹을 데이터베이스화 혹은 지식 기반으로 만들고 이를 기반으로 추론을 목표로 한다.



(a) 네이트의 시맨틱 검색



(b) 글로벌 시맨틱 웹 검색 분류

그림 1. 시맨틱 웹 기반 검색 서비스 현황

2000년대 중반 들어 위키피디아를 통해 웹에서 구조화된 데이터가 가능하다는 것을 보여준 이후, 이를 RDF(Resource Description Framework) 방식으로 변환 시킨 DBPedia.org 라는 프로젝트로 인해 크게 바뀌었다. 링크드데이터(Linked

Open Data)라고 불리는 이름으로 RDF 기반으로 데이터 웹을 구조화 시키거나 RDFa나 마이크로포맷(Microformat) 같은 방식으로 HTML 기반 의미 마크업을 시도하고 있는 것이다. 구글의 스퀘어드나 MS가 인수해 Bing에 추가한 파워셋(Powerset)의 경우, 위키피디아의 구조화된 DB를 기반으로 한다. 시맨틱 웹 검색의 시초라할 수 있는 하키아(Hakia) 역시 LOD를 기반으로 하고 있고 일부 자연어 처리 기술을 접목하고 있는 상태이다 [2]. 국내에서는 유사한 방식으로 네이버에서 영화 시맨틱 웹 검색엔진을 출시하였다. 영화 콘텐츠 DB의 관계성을 RDF로 추출(Exporting)한 후 URI를 기반으로 그래프를 따라 의미를 쫓아가는 방식이다 [3]. 이와 같은 시맨틱 웹 검색 방식은 구조화된 지식을 담고 있는 그릇이 필수적이다. 포털의 경우, 영화나 음악 같은 DB를 RDF로 관계를 정하고 이를 URI 기반으로 사용자 콘텐츠와 유기적으로 엮음으로서 좀 더 의미적인 시맨틱 웹 검색이 가능할 것이다.

아직까지 기존 텍스트 기반 웹 검색에서 출발한 '시맨틱 검색'과 웹을 DB로 보고 이를 의미 기반으로 연결할 '시맨틱 웹 검색' 어느쪽이 성공할지는 알 수 없다. 소셜 웹 시대가 되면서 대용량 데이터가 더욱 커지고 있고, 트위터와 같은 소셜 네트워크 데이터를 이용하면 검색에서 더 많이 구조화된 데이터로 사용하는 사례가 더욱 많아지고 있다.

III. 링크드데이터 기반 검색 기술의 발전

링크드데이터(LOD)는 웹 상에서 이전에 연계되지 못했던 데이터들을 상호 연계하기 위해서 데이터 출판 및 접근 원칙을 제 공함으로써 동질적인 도메인 데이터와 이질적인 도메인 데이터 연계에 중요한 역할을 수행할 수 있게 된다. 2007년 12개 데이터 셋으로 시작하여 지속적인 양적, 질적 성장을 통해서 2011년 295개 데이터 셋, 316억 트리플 규모로 커졌으며, 매년 새로운 데이터 셋이 추가되면서 지속적인 성장을 계속하고 있다 [4].

표 1. LOD 분야별 데이터 통계

분야	데이터셋	RDF 트리플(억개)	비중
미디어	25	18	5.8%
지리정보	31	61	19.4%
정부	49	133	42.0%
출판	87	29	9.3%
교차분야	41	41	13.2%
의과학	41	30	9.6%
사용자생성 콘텐츠	20	1	0.5%
계	295	316	100%

현재, BBC, Thomson Reuters, EliLilly 등을 중심으로 산업계 정보 공동 활용을 이끌어내었고, 선진국을 중심으로 활발하게 진행중인 정부 데이터까지 포함함으로써 양적, 질적인 성장이 가능해졌다.

〈표 1〉은 LOD 를 구성하는 분야별 데이터 셋 개수, RDF 트립플 개수를 보여주며 LOD 는 시맨틱 웹 데이터의 주된 배포 경로로서 온톨로지 연계 및 활용에 중요한 역할을 수행하고 있으며, LOD 에 참여하는 분야들이 확대되고, 분야별 온톨로지가 복잡해지고 규모가 커짐에 따라 온톨로지 내 데이터 간 관계 규모가 기하급수적으로 늘어나고 있다. 이러한 환경 변화에 따라 데이터에 대한 일관성, 품질, 관리, 활용 및 도구 개발에 대한 필요성이 아울러 높아지고 있다.

링크드데이터 애플리케이션은 RDF로 만들어진 데이터와 그 사이의 링크를 이용하여 만들어지는 것으로, 먼저 Tabulator(MIT), OpenLink RDF Browser, Fenfire(DERI) 같은 RDF 브라우저가 있다. 특정 검색어에 대해 RDF 데이터를 표시하거나 시각적으로 보여주는 기능을 제공한다. 두번째는 이들 RDF 사이에 데이터를 혼합한 매쉬업 서비스로서, 대표적으로 DBpedia Mobile과 같이 위키피디아를 RDF로 변환한 후 이를 지도에 시각적인 모습으로 보여준다. 세번째는 LOD 검색엔진으로 기존 검색엔진과 같이 키워드에 대해 랭킹 모델 및 링크를 모아서 보여주는 것이다. Sig.ma(DERI)가 여기에 포함된다.

표 2. 기존 검색 기술 과 시맨틱 웹 검색 비교

기존 웹 검색	방법	시맨틱 웹 검색
외부 웹 문서 및 사내 콘텐츠 DB	대상	링크드데이터(LinkedData) 및 사내 콘텐츠 DB
웹 크롤러를 통해 수집	수집	RDF 수집
랭킹에 따라 문서 인덱스	저장	관계에 따라 RDF Triple 변환
IR 알고리즘	결과	SPARQL 쿼리 응답
키워드 기반 랭킹 기반 검색	서비스	그래프 기반 의미 검색
Google(1조)	데이터 용량	LinkedData (320억)
Google	대표 기업	Bing, Hakia

기본의 웹 검색 엔진과 시맨틱 웹 검색엔진은 〈표 2〉와 같은 차이를 가진다. 시맨틱 웹 검색이란 HTML내 RDFa, 마이크로포맷 혹은 HTML5 Microdata와 같은 반구조적 데이터와 XML, JSON과 같은 구조적 데이터 그리고 RDF나 RDFs 같은 구조적 데이터를 크롤링하여 Virtuoso, Sesame, Jena 같은 RDF 레포지터리에 데이터를 저장 및 색인 하고 이를 시맨틱웹 질의 언어인 SPARQL을 이용하여 사용자의 요청에 대한 결과를 제공하거나 랭킹 모델을 취합하여 검색 서비스로 제공하는 것이다.

하지만, 검색 엔진 및 서비스는 정보 수집, 저장, 서비스 모든 면에서 주기적으로 대용량 처리 능력이 필요하며, 웹 기반 데이터가 기하 급수적으로 늘어나면서 클라우드나 분산 처리 플랫폼이 필수 불가결하다. 2009년을 기점으로 LOD 기반의 대용량 시맨틱 웹 데이터 처리가 문제가 되면서 기존 검색 엔진과 마찬가지로의 플랫폼이 필요해졌다.

IV. Hadoop 기반 대용량 LOD 검색 개발 사례

기존의 시맨틱 웹 검색 엔진은 특정 도메인의 소규모 데이터에 국한하여 다음과 같은 데이터 처리 과정을 거친다. 즉, 1) 모델 정의(개념과 관계 속성에 대한 정의, 최대한 현실에 부합하는 모델을 만들며 확장 유연성), 2) RDF 생성 및 색인(대개 기존 DB에서 변환 및 RDF, Triple, N-Triple 형태 저장), 3) SPARQL 질의를 통한 결과 추론 등으로 이루어진다. 하지만, 이들 데이터 처리 시스템들은 대개 싱글 머신에서 돌아가도록 개발되는 경우가 많아 조금만 복잡한 질의라도 처리 시간이 길어지는 단점이 있었다. 이를 해결하기 위해 RDF 저장소의 색인 기법을 다양화하거나 메모리에 색인 데이터 전량을 올린다거나 그것도 안되면 아예 다시 관계형 DB에 저장하는 방법으로 성능 문제를 해결해 왔다.

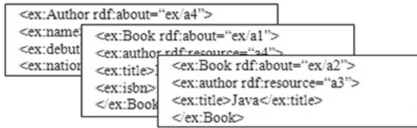
최근에 클라우드 및 분산 처리 기술에 큰 변화를 가져온 하둡(Hadoop)을 이용하여 이러한 문제를 획기적으로 바꾸는 계기가 되고 있다. Hadoop은 구글의 검색 데이터 처리에서 주로 사용한 맵리듀스(Map/Reduce) 방식을 오픈 소스 소프트웨어로 구현한 것으로 2006년 야후!의 더그커팅이 처음 시작하여, 지금은 대용량 데이터 처리 비용을 감소 시키고 빅데이터(BigData)'라는 새로운 IT 기술 조류를 만들어 내는 주요 기술로 자리잡았다.

하둡을 통해 분산 환경의 데이터 처리와 저장을 쉽게 처리할 수 있다. Map/Reduce 분산 환경에서 데이터 작업을 아주 작은 단위로 쪼갠 후, Map 함수에서 데이터를 처리를 하고 Reduce 함수에서 원하는 결과값을 계산시켜 결과를 얻어낸다. 이를 자체 파일 시스템(HDFS, Hadoop Distributed File System)을 이용해서 결과 파일을 적당한 블록 사이즈 (64MB)로 나눠서 각 노드 클러스터에 저장한다.

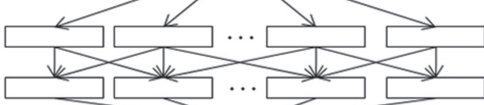
Hadoop을 검색 데이터 처리 기술에 사용하면 사회적 이슈가 발생했을 때, 분산 환경의 동적 제어 API를 이용하여 크롤링 및 인덱싱 작업을 비주기적으로 시행 가능하고, 사용자 생산 콘텐츠

즈 DB에 대해서 분산 데이터 처리를 위한 Hadoop 뿐만 아니라 대용량 데이터 저장을 위한 Hbase와 같은 NoSQL 플랫폼 등 각종 분산 컴퓨팅 자원을 필요 시 이용할 수 있어 실시간 웹 (Realtime Web) 검색을 대응하기 위한 검색 엔진 시스템으로서 적합하기 때문이다.

RDF Statements



#1 Parsing & Splitting Statement



#2 Triple sorting

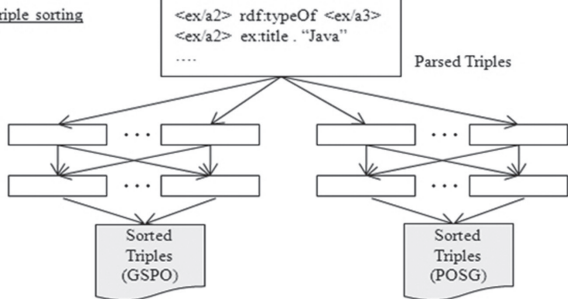


그림 2. Map/Reduce를 이용한 RDF 데이터 전처리 과정

새롭게 제안한 대용량 시맨틱 웹 데이터 포털 서비스 방법은 1) 구조적 데이터 수집 후 2) 데이터 저장 (Hadoop을 이용한 대용량 RDF 변환 및 처리, NoSQL을 이용한 검색 데이터 저장소) 및 3) 쿼리 및 데이터 분석을 시행한다. 이 때, 사용자 쿼리에 해당하는 질의어를 SPARQL Query로 생성하여, 쿼리에 대한 서브 쿼리를 자동 생성하여 작업을 통해 해당 색인(Answer-set)을 주기적으로 추출하고 이를 랭킹 및 결과로서 제공 하며, 관계 기반 질의어 확장 및 추천을 통해 사용자에게 좀 더 친밀한 사용자 경험을 줄 수 있다 [5].

〈그림 2〉는 맵/리듀스 방식을 이용한 RDF 데이터 전처리 과정을 간략히 보여주고 있으며, 〈그림 3〉은 전체 시스템 구성도이다.

본 연구 결과는 Daum의 영화/음악/인물 데이터를 기반으로 5억건의 RDF 트리플을 처리하는데 몇 십분 안에 원하는 해당 색인을 얻어낼 수 있었다. 특히, 사용자에게 친숙한 검색어의 질의 확장을 속성 중심으로 자동 완성 기능으로 제공하여 검색 체류 시간을 증가하는 부수적인 결과도 얻었다 [6].

이러한 연구 결과는 여전히 데이터 사이즈가 너무 작아서 분산 처리 기술에 대한 도움을 얻기에는 부족함이 있다. 그러나

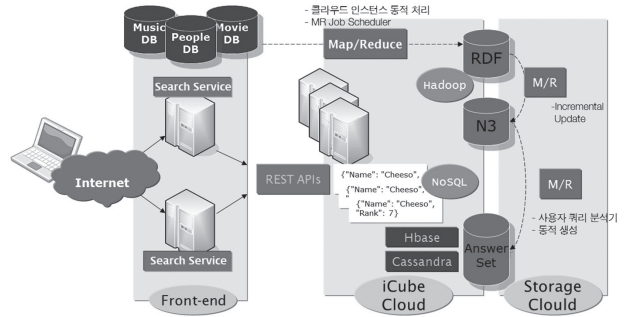


그림 3. Hadoop 기반 시맨틱 검색엔진 시스템 구성도

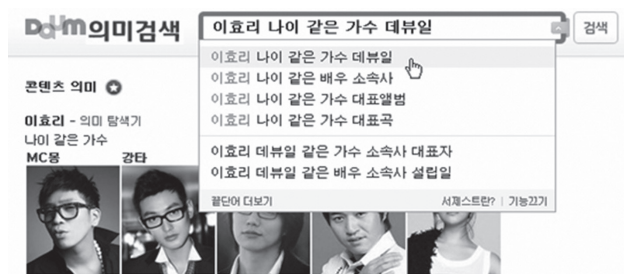


그림 4. 다음 영화-음악-인물 의미검색

의학 데이터 영역은 LOD 영역에 수십억 RDF 트리플의 데이터를 제공하는 분야이다.

기존의 시스템을 확장하여 PubMed, DrugBank, Linked Life Data 등의 레포지터리로 부터 크롤링한 90억건의 RDF 트리플을 처리하는 시스템을 개발하였다. 데이터 사이즈가 커지고 온톨로지 모델이 다양해짐에 따라 기존의 SPARQL 해당 색인(AnswerSet) 모형에서 데이터 처리 횟수가 급격히 늘어나는 단점이 생기게 되었고 이에 따라 RDF 데이터 처리 시 Hadoop을 통해 Apache Solr라는 분산 검색 엔진에 색인함으로써 작업 효율성을 배가하였다.

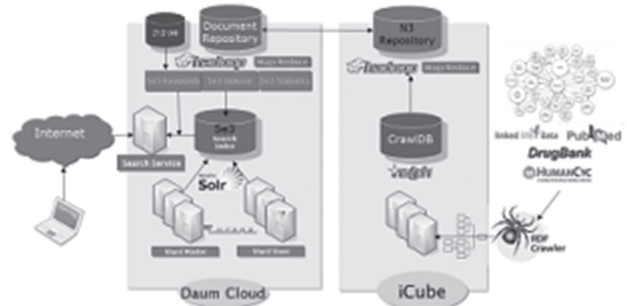


그림 5. 의학 LOD 검색 서비스 시스템 구성도

Solr 기반 RDF 검색 색인은 검색어(Search), 문서(Document), 키워드 추천(Suggestion)으로 나누어 만들어지며, 연관된 트리플을 빠르게 검색하여 실시간으로 데이터 검색 서비스가 가능하다는 것을 보여주었다.

V. 공공 데이터 서비스로의 확장 및 한계

2009년 들어 정부가 만든 공공 데이터의 개방 활동이 본격화되고 있으며 웹 창시자인 팀 버너스리(Tim Berners-Lee)의 주도하에 영국의 data.gov.uk를 통해 데이터 공개가 시작되었다.

미국의 경우, 오바마 정부의 Open Government Initiative에서 data.gov를 통해 LOD 데이터 개방이 추진되고 있으며, 정부가 공공 정보에 대한 접근을 용이하게 하여 국민의 창의적 혁신을 지원하고, 개인과 커뮤니티가 삶의 질을 개선하고 더 나은 정보에 의한 판단을 통해 솔루션을 만들어 낼 수 있도록 돕고 있다 [7]. 국내에서도 국가 공유자원 포털(data.go.kr)을 통해 각 부처의 다양한 데이터를 통합 구축하여 2011년부터 제공을 시작하였다.

하지만, 공공데이터는 그 특성상 데이터의 출처가 다양할 뿐만 아니라 크기도 다양하다. 정책 법안 같은 문서로부터 스프레드 시트 및 기상청의 날씨 자료처럼 실시간으로 끊임없이 생산되는 대용량 데이터도 있다. 따라서, 이들을 통합적이고 유기적으로 다루기 위해 시맨틱 웹 온톨로지 모델과 대용량 분산 처리 시스템은 필수 불가결하다. 또한, 이를 기존 비정형 혹은 반정형 웹 데이터와 결합해 데이터를 분석하는 것은 더 큰 요구 사항을 만들어낸다.

즉, 이중 모델을 가진 공공 데이터와 웹 데이터를 취합하여 일반 사용자가 쉽게 질의할 수 있는 원스톱 대용량 지식 기반 시스템이 필요하다고 할 수 있다. 이에 따라 <그림 6>과 같이 Hadoop을 이용한 기존 LOD 검색 엔진 플랫폼을 이용하여 다양하게 들어오는 공공 데이터를 통합적으로 빠르게 색인하여 이를 SPARQL 쿼리 엔진, XML/JSON 기반의 오픈 API로 제공하여 다양한 매쉬업 사례가 나오도록 하는 연구가 추진 중이다.

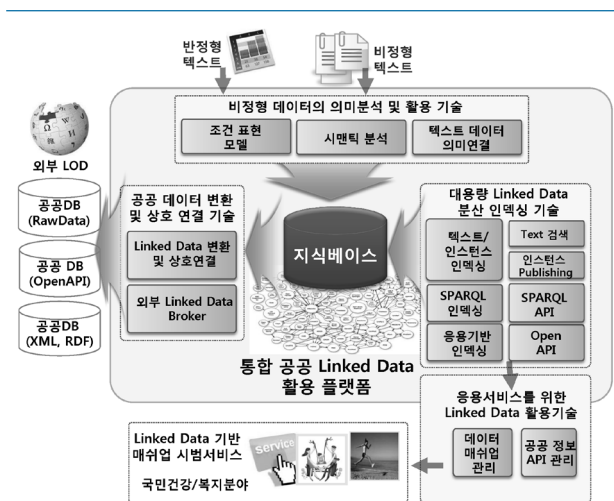


그림 6. 통합 공공 LOD 활용 플랫폼 모식도

이렇듯 LOD를 이용한 대용량 데이터 처리는 많은 활용 가능성이 있으나 아직도 풀어야 할 숙제 또한 안고 있다. 첫번째로 데이터의 규모가 여전히 작다는 점이다. 상용 LOD 대용량 데이터 처리에 대한 관심이 높아지자 국제 시맨틱 웹 콘퍼런스(ISWC)에서는 매년 Billion Triples Challenge라는 콘테스트를 통해 수십억 RDF 트리플의 데이터를 제공하여 다양한 서비스 실험을 하도록 장려하고 있다. 두번째로는 LOD간 연결 링크가 적다. 링크는 이중 모델간의 데이터 연결을 통해 더 확장된 지식을 생산할 수 있는 가장 기초적인 작업이지만, 전체 LOD의 5%만이 링크를 가지고 있다. 전문가에 의한 수작업 및 데이터 마이닝 같은 고전적인 방법을 벗어나야 한다 [8].

즉, 여러 영역에서 LOD 생산은 많이 되고 있으나 그 사이의 연관 관계 및 링크와 가중치를 높여 주는 작업이 여전히 어렵고, 시간에 따라 바뀌는 데이터 변경 사항에 대한 처리 부분이다 [9]. 전자를 해결하기 위해 신뢰 기반의 데이터 쿼리 확장인 T-SPARQL과 스트리밍 RDF 데이터 처리를 위한 C-SPARQL과 같은 확장이 나와 있다. 이를 Hadoop 같은 대용량 데이터 분산 처리 플랫폼과 연계하는 연구가 필요하다.

마지막으로 독특한 서비스로의 확장이 어렵다는 점이다. 기존의 데이터셋은 너무 전문적이어서 킬러 앱이 나오기가 힘들다. 국민 생활에 밀접한 공공 데이터를 이용하여 검색과 연계한 다양한 서비스가 나와야 하는 이유이다.

VI. 결론

최근 빅데이터 기술의 성장은 우리에게 많은 가능성을 안겨주고 있다. 기존에 활용하지 못했던 버려진 데이터를 값싸고 효율적으로 처리할 수 있는 도구를 제공해 주기 때문이다. 본 고에서는 기존 시맨틱 웹 연구의 한계점을 살펴보고 대용량 데이터 처리를 위해 Hadoop 및 NoSQL과 같은 빅데이터 기술을 활용한 몇 가지 연구 사례를 소개하였다. 최근에 공공 데이터를 활용한 빅데이터 활용에 대한 다양한 요구가 대두되고 있으며, 기존의 연구 결과를 토대로 이중 모델을 가진 공공 데이터와 웹 데이터를 취합하여 일반 사용자가 쉽게 질의할 수 있는 원스톱 대용량 지식 기반 시스템이 필요하다. 국민 생활에 밀접한 공공 데이터를 이용한 서비스 산업 발전을 위해서는 다양한 공공 데이터간 링크 확장, 실시간 데이터 처리 및 리비전 관리 등의 연구가 진행되어야 할 것이다.

Acknowledgement

본 연구는 방송통신위원회의 방송통신 인프라 원천 기술 개발 사업의 연구결과로 수행 되었음. (KCA-2012-(12-911-05-004))

참고 문헌

- [1] 박윤정, "웹 3.0 정보서비스의 사례와 시사점", 방송통신 정책, 21(20), pp. 25-43. 2009
- [2] Alex Iskold, "Semantic Search: The Myth and Reality", http://www.readwriteweb.com/archives/semantic_search_the_myth_and_reality.php, 2008
- [3] 네이버랩 시맨틱 영화검색, <http://semantic.lab.naver.com/>, 2008
- [4] 김평, 김태홍, 정한민, "개체 식별 관점에서 바라본 링크드 데이터 동향", 주간 기술 동향, 정보통신산업진흥원, 2011.12.2
- [5] Seokchan Yun, Mina Song, Hyun Namgung, Sungkwon Yang, Harshit Kumar, Hong-Gee Kim, "Improving the Workflow of Semantic Web Portals using Map/Reduce in Cloud Platform", International Conference on Knowledge Engineering and Ontology Development, pp.485-488, 2009
- [6] 양성권, 정진욱, 윤석찬, 김흥기, "영화 콘텐츠 검색에서의 관계 기반 질의 확장", HCIKorea 2011 Conference
- [7] 팀 오라일리, "열린 정부 만들기(모든 시민이 참여하는 투명한 정부 2.0 프로젝트)", 에이콘 출판사, 2012
- [8] Nansu Zong, Dong-Hyuk Im, Sungkwon Yang, Hyun Namgoong, Hong-Gee Kim, "Dynamic Generation of Concepts Hierarchies for Knowledge Discovering in Bio-medical Linked Data Sets", ACM International Conference on Ubiquitous Information Management and Communication, 2012.
- [9] Dong-Hyuk Im, Nansu Zong, Eung-Hee Kim, Seokchan Yun, Hong-Gee Kim, "A Hypergraph-based Storage Policy for RDF Version Management System." ACM International Conference on Ubiquitous Information Management and Communication, 2012

약 력



윤 석 찬

1996년 부산대학교 지질학과 이학사
1999년 부산대학교 지질학과 이학석사
2010년 서울대학교 치의학과 박사 수료
1996년~2004년 나인포유(주) 부설 연구소 실장
2004년~현재 ㈜다음커뮤니케이션 DNALab 팀장
2007년~현재 제주대학교 컴퓨터공학과 겸임교수
관심분야: 웹2.0, 오픈소스, 시맨틱웹, 의료정보학



남궁 현

2004년 전북대학교 컴퓨터공학 학사
2006년 과학기술연합대학원 컴퓨터소프트웨어 및 공학석사
2008년 서울대학교 치의학과 박사 수료
2011년~현재 한국전자통신연구원 스마트서비스 연구팀 연구원
관심분야: 소셜웹, 시맨틱웹, 의료정보학



양 성 권

2007년 선문대학교 컴퓨터정보학부 이학학사
2010년 서울대학교 치의학과 치의학석사
2012년 서울대학교 의료정보학 박사수료
2012년~현재 솔트룩스 전략사업연구센터 연구원
관심분야: 빅데이터, 시맨틱웹, 의료정보학



김 흥 기

1985년 고려대학교 학사
1993년 미국 조지아대학 이학석사
1996년 미국 조지아대학 이학박사
1998년~2005년 단국대학교 경영정보학과 부교수
2005년~현재 서울대학교 치과대학 교수
관심분야: 인공지능, 시맨틱웹, HCI, 의료정보학