

Patent Information

- 발명자  
이규철
- Patent number  
10-2021-0009564  
(2021.01.22.)

Keyword

- LOD(Linked Open Data)
- 맵리듀스
- 데이터 변환 시스템

Applications

- LOD 서비스

Patentee & contact point

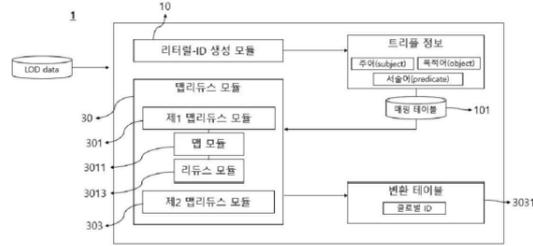


충남대학교  
기술가치센터  
042-821-7174  
cnutlo@cnu.ac.kr

Technology Overview

LOD 저장을 위한 맵리듀스 기반 데이터 변환 시스템 개발

- 대량의 LOD 저장을 위한 맵리듀스 기반 데이터 변환 시스템(1)은 리터럴-ID 생성 모듈(10), 및 맵리듀스 모듈(30)을 포함하며, 맵리듀스 모듈(30)은 제1 맵리듀스 모듈(301) 및 제2 맵리듀스 모듈(303)을 포함할 수 있다. 제1 맵리듀스 모듈(301) 및 제2 맵리듀스 모듈(303)은 모두 맵 모듈(3011) 및 리듀스 모듈(3013)을 포함할 수 있음
- 본 발명은 대량의 RDF 형식으로 저장된 데이터를 효과적으로 압축 저장함으로써, 저장공간비용, 데이터 검색비용, 분산환경에서의 통신비용 등을 줄일 수 있는 효과가 있음



<대량의 LOD 저장을 위한 맵리듀스 기반 데이터 변환 시스템 구성도>

Technology Highlights

Predicate-ID 매핑 테이블

- 리터럴-ID 생성 모듈(10)은 데이터 변환을 위해 Subject/Object(주어 및 목적어)-ID 매핑 테이블, Predicate-ID(서술어) 매핑 테이블, 2개의 ID 매핑 테이블을 생성할 수 있음
- (a)는 Subject/Object-ID 매핑 테이블과 ID-Subject/Object 테이블을 나타내는 도면이고, (b)는 Predicate-ID 매핑 테이블과 ID-Predicate 테이블을 나타내는 도면임. (c)는 Subject/Object-ID 테이블과 ID-Subject/Object 테이블을 나타내는 도면이고, (d)는 Predicate-ID 테이블과 ID-Predicate 테이블을 나타내는 도면임
- Subject/Object-ID 테이블은 RDF 데이터에서 주어(Subject)와 목적어(Object)에 위치하는 리터럴(Literal)과 이에 대응하는 정수 값을 저장하는 ID 매핑 테이블을 나타냄
- 해당하는 데이터를 빠르게 검색하기 위해, 매핑 테이블은 해시 테이블로 구성된다. 주어/서술어 Literal을 테이블의 Key로, 각 Literal에 대응하는 정수 ID를 Value로 저장하며 이때, 데이터의 변환과 매핑 테이블 생성은 동시에 발생함
- 변환하고자 하는 주어/서술어 Literal을 Subject/Object-ID 테이블에 입력하면, 이에 대응하는 ID 값을 검색함

[Subject/Object-ID 테이블]		[Predicate-ID 테이블]	
Subject/Object Data	Hash Value (Subject/Object ID)	Predicate Data	Hash Value (Predicate ID)
http://www.Department0.University0.edu/Lee	1	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	1
http://www.Department0.University0.edu/Professor	2	http://www.lehigh.edu/~zhp2/2004/04/01/univ-bench.owl#EmailAddress	2

[ID-Subject/Object 테이블]		[ID-Predicate 테이블]	
Hash ID (Subject/Object ID)	Hash Value (Subject/Object Data)	Predicate ID	Hash Value (Predicate Data)
1	http://www.Department0.University0.edu/Lee	1	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
2	http://www.Department0.University0.edu/Professor	2	http://www.lehigh.edu/~zhp2/2004/04/01/univ-bench.owl#EmailAddress

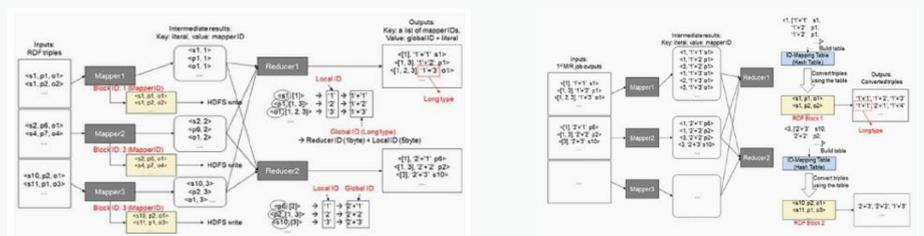
[Subject/Object-ID 테이블]		[Predicate-ID 테이블]	
Hash ID (Subject/Object ID)	Hash Value (Subject/Object Data)	Predicate Data	Hash Value (Predicate ID)
1	1#Professor	3#EmailAddress	2

[ID-Subject/Object 테이블]		[ID-Predicate 테이블]	
Hash ID (Subject/Object ID)	Hash Value (Subject/Object Data)	Predicate ID	Hash Value (Predicate Data)
1	1#Lee	2#type	1
2	1#Professor	3#EmailAddress	2

<본 발명에 따른 매핑 테이블>

글로벌 ID 생성 첫번째 맵리듀스 잡/변환 테이블을 생성하는 두번째 맵리듀스 잡



<제1 맵리듀스 모듈에서 수행하는 글로벌(Global) ID 생성 맵리듀스 잡의 동작 과정>

<제2 맵리듀스 모듈에서 수행하는 트리플 변환 잡 동작 과정>

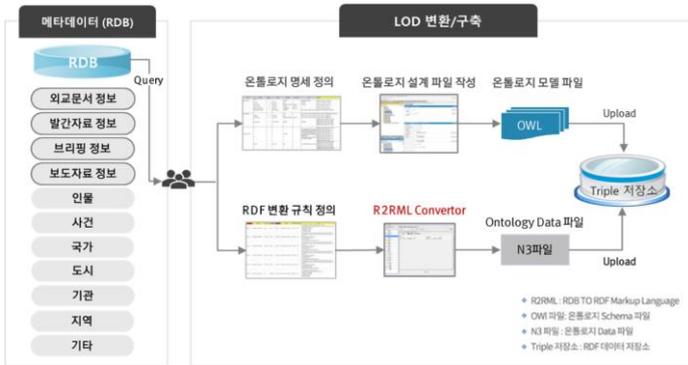
- 글로벌 ID 생성 잡 : 전체 RDF 트리플에 포함된 주어, 서술어, 목적어를 중복 없는 리터럴(Literal)로 추출하여, 각 리터럴(Literal)을 구별할 수 있는 정수 자료형의 글로벌 ID를 부여하는 작업임
- 맵 모듈(3011)은 입력 RDF 트리플 데이터를 일정한 크기의 블록으로 나누어 입력받음. 각 맵 모듈(3011)은 입력받은 트리플 데이터를 주어, 서술어, 목적어 단위로 분할하고, 각 리터럴(Literal)을 중간 결과의 Key로, 맵의 태스크 ID를 중간 결과의 Value로 작성하며, 입력받은 데이터 블록에 맵의 태스크 ID를 부여하여 HDFS에 바로 쓰기를 수행함
- 제2 맵리듀스 모듈 : 제1 맵리듀스 모듈(301)에서 생성한 밸류의 정보 중 특정 RDF 블록에 속하는 리터럴의 글로벌-ID 정보만을 저장하여 변환 테이블(3031)을 구축할 수 있음
- 두 번째 수행되는 맵리듀스 잡은 트리플 변환 잡으로, 첫 번째 맵리듀스 잡에서 생성한 결과, 즉 글로벌 ID 값을 사용하여 RDF 트리플 데이터를 변환하는 작업임

## Technology Readiness Level(TRL)



## Technology Applications

### LOD(Linked Open Data) 서비스



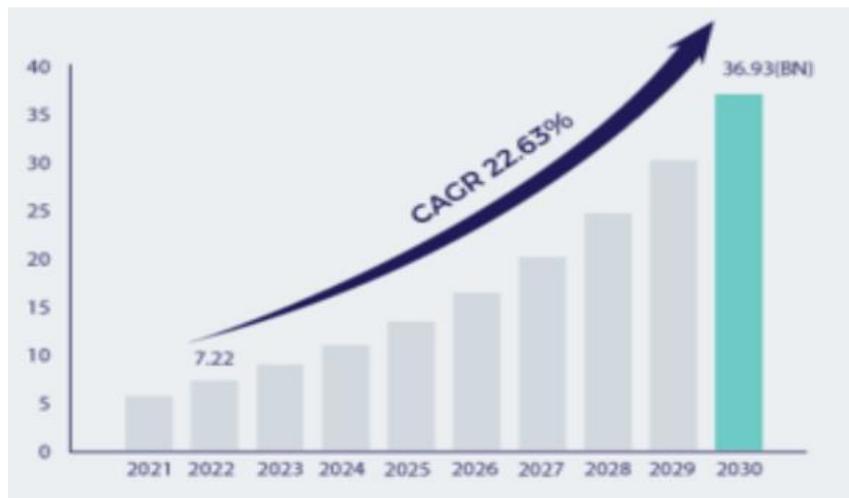
## Market Trends

### 기술의 주요 적용 시장: 오픈 소스 인텔리전스 시장

#### 오픈 소스 인텔리전스 시장규모

- 전 세계 OSINT(오픈 소스 인텔리전스) 시장규모는 2022년 72억 2천만 달러로 측정됐으며, 2023년부터 2030년까지 연평균 성장률(CAGR) 22.63%로 성장해 2030년에는 369억 3천만 달러에 도달할 것으로 예상됨

(단위 : 10억 달러)



[2023-2030 세계 오픈 소스 인텔리전스 시장규모]

(출처: 전 세계 오픈 소스 인텔리전스 시장 전망, 2023~2030년, SNS Insider)

- 오픈 소스 인텔리전스(open-source intelligence)는 공개적으로 사용 가능한 소스에서 얻은 정보를 수집, 분석 및 전파하는 프로세스를 말하며, 비즈니스 계획에 대한 통찰력을 얻기 위해 다양한 부문에서 OSINT에 대한 수요가 증가하고 있음
- R&D의 핵심이 공개적으로 이용 가능한 정보에 대한 전 세계적인 연구로 전환됨에 따라 OSINT 시장이 크게 확대될 것으로 전망됨