3월 4주차 연구노트

RDF datasets 에서의 SPARQL 쿼리 최적화 관련 논문 조사.

관련논문 : [An analytical approach for query optimization based on hypergraph](http://ieeexplore.ieee.org/document/7207087/)

([Sangeeta Sen](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Sangeeta%20Sen.QT.&newsearch=true); [Anisha Agrawal](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Anisha%20Agrawal.QT.&newsearch=true); [Ankit Rathi](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Ankit%20Rathi.QT.&newsearch=true); [Animesh Dutta](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Animesh%20Dutta.QT.&newsearch=true); [Biswanath Dutta](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Biswanath%20Dutta.QT.&newsearch=true) [2015 12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)](http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=7180538) Year: 2015)

RDF(Resource Description Framework)는 시멘틱 웹상의 정보를 저장하기위한 코어데이터 모델이다. 이 것은 triples(subject or resource, predicate or property, and object or property value)형태로 정보를 저장하는데, RDF datasets 에서의 SPARQL 쿼리의 최적화에 대한 이슈가 떠오르고 있다. 여기서 SPARQL이란 쿼리 언어를 기초로한 그래프로, 주요 역할은 패턴을 매치시키는 것이다.

이 논문의 목적은 hypergraph로, rdf-3x, AllegroGraph, Jena라는 기술보다 계산 처리 시간을 단축하기 위한 쿼리 최적화와 데이터 관리 기술이다. Hypergraph는 두 개이상의 vertices연결하는 edges들로 구성된 그래프로 이러한 edges들을 hyper-edges라고 한다.

이 논문의 주요 연구내용은 아래와 같다.

1. Hypergraph based database of RDF data

특정 predicates(술어)와 연결된 모든 subjects(주체)와 objects(개체)는 hyper-edges아래에 있기 때문에 RDF 트리플 검색이 쉬워진다.

2. Query path based query Processing

쿼리를 처리하기 위해, 첫번째 모델은 재배열된 쿼리 predicates(술어)에 따른 경로에서 발생하고 이 경로에 따라서 쿼리는 처리된다.

이 논문에서 제안된 시스템 아키텍처는 크게 세 개의 계층으로 구성되어 있다.

1. User Interface Layer

이 계층에서 유저는 쿼리를 발생시키고, 시스템에서 처리된 결과를 유저에게 제공한다.

2. Query Processing Layer

이 계층은 세 개의 요소들로 구성되어 있다.

- Query parser : 쿼리로부터 쿼리를 매칭

- Query optimizer : hyperedges의 크기에 따라 패턴 매칭을 재배열

- Query processor : 저장된 정보에 따라 쿼리를 실행

3. Storage Layer

이 계층은 두개의 요소를 구성되어 있다.

- RDF data : 트리플 그래프를 표현.

- Indexer and Storage : 싱글 유닛안에 특정 predicate(술어)의 subjects(주체)와 objects(개체)을 가져오고, 키-값쌍으로 된 데이터를 저장함으로써 인덱스를 구성