ΣΥΓΧΡΟΝΕΣ ΔΙΑΔΙΚΤΥΑΚΕΣ ΕΦΑΡΜΟΓΕΣ

Β. Λούμος – Ι. Αναγνωστόπουλος

Αναπαράσταση και ανάλυση ερωτημάτων SPARQL σε μορφή γράφου

Κορμανιώτης Τρύφων – Κουτσανδρέας Αδαμάντιος

RDF

- The Resource Description Framework (RDF) is a framework for representing information in the web

- It is a family of World Wide Web Consortium (W3C) specifications, originally designed as a metadata data model

- It has come to be used as a general method for conceptual description or modeling of information that is implemented in web resources

- Using a variety of syntax notations and data serialization formats

RDF data model description

- The RDF data model is similar to classical conceptual modeling approaches such as entity-relationship or class diagrams
- It is based upon the idea of making statements about resources in the form of subject-predicate-object expressions
- These expressions are known as triples in RDF terminology
- The subject denotes the resource, and the predicate denotes traits or aspects of the resource and expresses a relationship between the subject and the object

RDF dataset

- An RDF dataset comprises one default graph and zero or more named graphs
- RDF graphs are sets of subject-predicate-object triples
- the elements may be IRIs(generalization of URIS), blank nodes, or datatyped literals.
- RDF graphs are used to express descriptions of resources
- There are some controlled vocabularies in common use
- The intent of publishing RDF-based ontologies on the Web is often to establish, or circumscribe, the intended meanings of the resource identifiers used to express data in RDF

RDF Schema- RDFs

- RDFs is a semantic extension of RDF
- RDFs defines classes and relations of resources and organize their hierarchy
- RDFs defines signatures of relations (domain, range), documents them with labels and comments
- RDFs defines associated inference rules

RDF Applications

- DBpedia
- Creative Commons
- FOAF(friend of a friend)
- Haystack Client
- IDEAS Group
- Microsoft RDF-based Profile Management
- MusicBrainz
- NEPOMUK
- Simple knowledge Organization System
- SIOC(Semantically-Interlinked Online Communities)
- Smart-M3

RDF enables you to open your data to applications through the web

SPARQL

- SPARQL (SIMPLE Protocol and RDF Query Language) is an RDF query language (semantic) which is able to retrieve and manipulate data stored in Resource Description Framework (RDF) format.

- It was made a standard by the *RDF Data Access Working Group* (DAWG) of the World Wide Web Consortium

- It is recognized as one of the key technologies of the semantic web

- SPARQL protocol sending queries and their results across the web

SPARQL

- Beside a query language, it is a result format and an access protocol for RDF
- It is declarative and it is based on RDF data model (triples/graph)
- SPARQL queries are executed against an RDF dataset
- SPARQL query language based on the triple model filters to add constraints, optional and alternative parts
- SPARQL allows for a query to consist of triple patterns, conjunctions, disjunctions, and optional patterns.

SPARQL Basic clauses

- select :to identify the values to be returned
- from :to identify the data sources to query
- where :to identify the triple/graph pattern to be matched against the triples/graphs of RDF
- prefix :to declare the schema used in the query
- filter :to add constraints to the graph pattern
- optional :to make the matching of a part of the pattern optional
- union :to give alternative patterns in a query
- order by :to sort
- limit :result number
- offset :rank of first results
- unbound :to test a variable I not bound
- construct :return a spesific RDF graph for each result

Graph patterns for the SPARQL query pattern

- Basic Graph Pattern (BGP)
- Group Graph Pattern
- Optional Graph Pattern-keyword Optional
- Union Graph Pattern-keyword Union
- Graph graph pattern-keyword Graph
- Constraints-keyword Filter

join ordering problem

- The join ordering problem is a fundamental challenge that has to be solved by any query optimizer
- A query is a request for information from a database
- Depending on the order of the join, there is a different computation time
- The query optimizer attempts to determine the most efficient way to execute a given query by considering the possible query plans
- We focus on static query optimization ,a join order optimization of triple patterns performed before query evaluation.

Join ordering problem at RDF systems

- Namely in RDF systems the optimizers are faced with extremely large sizes of queries due to verbosity of the data format
- SPARQL follows a sequential semantics when evaluating join groups
- although in many cases the order of evaluating patterns does not change the outcome of the query

- The optimization goal is:
- to find the execution plan which is expected to return the result set fastest and
- without actually executing the query or subparts

The solution of the problem

- This is typically solved by means of heuristics and summaries for statistics about the data
- Queries results are generated by accessing relevant database data and
- Manipulating it in a way that yields the requested information
- Given a SPARQL query, the query engine starts off with constructing its representation
- The representation is one of the Graph Patterns for the SPARQL query pattern mentioned above

Two different approaches

- RDF-3X:query compilation time (dominated by finding the optimization order) is one order of magnitude higher than the actual execution time
- Virtuoso-7:greedy algorithm for compilation leads to slow run time (sub-optimal order){seems to spend much less time finding the join order (probably employing some kind of greedy heuristics), but pays a high price for the (apparently) suboptimal ordering

✓ Ideally, we would like to have a hybrid of two approaches: a heuristics that spends a reasonable amount of time optimizing the query, and yet gets a decent join order