Linked Data Indexing Methods: A Survey

Martin Svoboda, Irena Mlýnková

Charles University in Prague

The Czech Republic

21st October 2011

SWWS@OTM, Crete, Greece

Outline

- Introduction
- Dimensions
- Approaches
- Observations
- Challenges
- Conclusion

Introduction

- Motivation
  - Web of Documents
  - Web of Data
- Linked Data
  - Principles
    - Unique identifiers (URIs)
    - Useful description (HTTP, RDF)
    - Links

Introduction

- RDF (Resource Description Framework)
  - Triples
    - Subject Predicate Object.
  - Graph
    - Directed labeled multigraph
    - Vertices for subjects and objects
    - Edges for particular triples

Intent

- Querying framework
  - Architecture
    - Compromise between local and distributed approaches
  - Issues
    - Physical storage
    - Index structures
    - Query processor
  - Problems
    - Data scalability, distribution and dynamicity

Intent

- Architecture
  - Local
    - Efficient processing
    - Independent data
    - Storage requirements
  - Distributed
    - Runtime requests
    - Up-to-date data
    - Network throughput

Dimensions

- Aspects
  - Data
  - Index
  - Querying
- Dimensions
  - Not all combinations make sense

Dimensions

- Data distribution
  - *Local*, *distributed* or *global* data
- Data units
  - *Triples*, *quads*, *documents* or other *sources*
- Data dynamicity
  - *Durable*, *changeable* or *volatile* data
- Index organization
  - *Local* or *distributed* model

Dimensions

- Index items
  - *Keywords*, *triples*, *quads*, *trees*, *paths* or *areas*
- Index content
  - Pure *data*, *statistics* or *summaries* about data
- Index dynamicity
  - *Dynamic* or *static* structures
- Access patterns
  - *Universal* or *limited* approaches

Dimensions

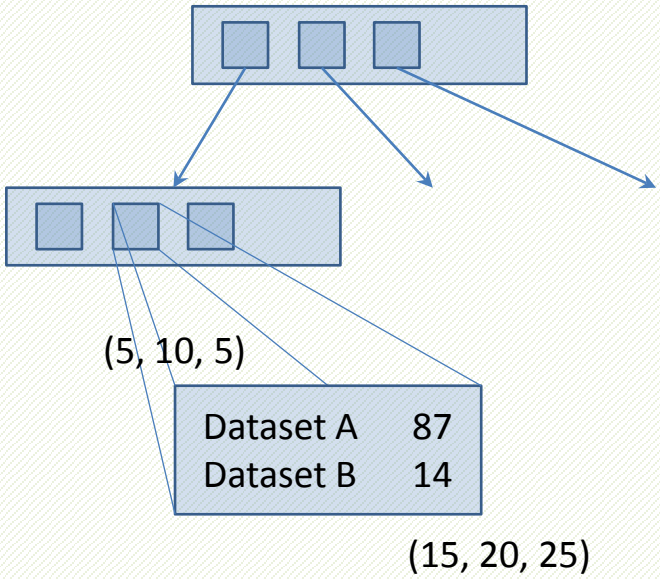
- Querying layer
  - *Syntactic*, *structural* or *semantic* querying
- Query models
  - *Full text* querying or graph *patterns*
- Query evaluation
  - *Local* or *distributed* processing
- Query results
  - *Complete* or *incomplete* results

Categories

- Main approach types
  - Querying systems
    - Local or distributed data
    - Structural queries
    - Complete results
  - Searching engines
    - Global data cloud
    - Full text queries
    - Imprecise results

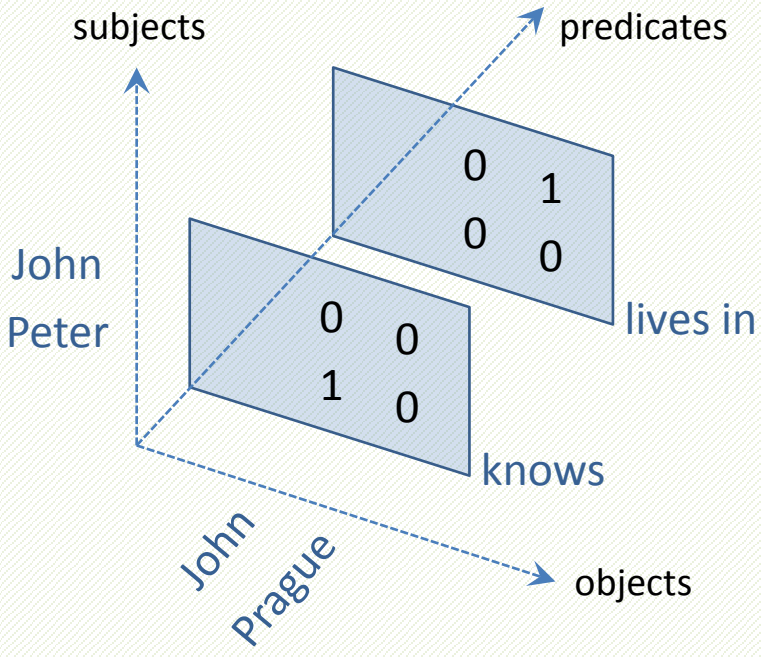
Approaches

- Source selection
  - Andreas Harth et al.: **Data Summaries for On-Demand Queries over Linked Data**
    - Data transformation
      - 3-dimenisonal space
      - Hash functions
    - Q-trees based on R-trees
      - Overlapping bounding boxes
      - Buckets with summaries



Approaches

- BitMat index
  - Medha Atre et al.: **Matrix "Bit"loaded: A Scalable Lightweight Join Query Processor for RDF Data**
    - 3-dimensional matrix
      - Bit values 0 or 1
    - 2-dimensional slices
      - S-O, O-S, P-O, P-S slices
    - Implementation
      - Compressed bit runs



Observations

- String compression
  - Repeating string values
    - URIs and literals
  - Unique integer identifiers
    - Efficient processing
    - Space requirements
  - Translation maps
    - Both directions
    - Based on B-trees

Observations

- Data pruning
  - Idea
    - Query optimization
    - Relevant data
  - Methods
    - Filtering selections
    - Join ordering
  - Problem
    - Partial knowledge

Challenges

- Data distribution
  - Motivation
    - Datasets are distributed
    - Appropriate compromise
  - Problems
    - Network drawbacks
    - Space requirements
    - Independent datasets

Challenges

- Data scalability
  - Motivation
    - Web of Data size explosion
      - September 2011:
      - 295 datasets, 31 billion triples, 504 million links
  - Problems
    - Scalable storages and indices
    - Efficient query evaluation
    - Quality, provenance and trust

Challenges

- Data dynamicity
  - Motivation
    - Data tend to ageing
  - Problems
    - Continuous updates
    - Dynamic structures

Conclusion

- Problem
  - Linked Data indexing methods
- Contributions
  - Approaches comparison
    - Dimensions
    - Observations
    - Challenges

Thank you for your attention…

Faculty of Mathematics and Physics
Charles University in Prague