GraphSM 2015 Introduction

Iztok Savnik (University of Primorska FAMNIT) and
Kiyoshi Nitta (Yahoo Japan Research)
GraphSM 2015 Introduction

Iztok Savnik (University of Primorska FAMNIT) and
Kiyoshi Nitta (Yahoo Japan Research)
Current state of graph DBs

- There exist a variety of different dictionaries, properties, concepts, ...
- There exist a variety of formats and models for knowledge and data representation

**Wordnet**
- Princeton large lexical database of English
- Cognitive lexical semantic to concepts
- 115,000 synsets
- Synsets are linked by:
  - Conceptual semantic relations (synonymy, hypernymy, ...) and
  - Basic relational links
- Include definitions of synsets

**Freebase**
- Free, knowledge graph
- People, places, and things
- 2.875 billion entities, 40.432 billion entities

**Wikidata**
- Free knowledge base with 14,710,852 items
- Properties of persons, organizations, works, events, etc.

**Linked Open Data**
- Number of triples: 33.8 Giga (2011)
- The number of datasets has grown by 217% (2011 to 2014)

**YAGO**
- 10 Mega (510) concepts
- Meta-Facts included, inferential
- Accuracy of 95%
- Includes:
  - Wikipedia, WordNet, GeoNames
  - Links between Wikipedia taxonomy
  - RDF examples
  - Anchored in time and space

**Cyc**
- Knowledge base
  - Diagrams, trees
  - Conceptual networks (etymology)
  - Higher ontology, basic theorems, specific theorems
  - Primitives semantic net based on
  - Conceptual semantic net based on
  - Based on first-rate calculus
  - Rule-based reasoning
Linked Open Data

- Number of triples: 33 Giga (109) (2011)
- the number of datasets has grown by 271% (2011 -> 2014)
## Survey of RDF Storage Managers

<table>
<thead>
<tr>
<th></th>
<th>$T_s$</th>
<th>$I_s$</th>
<th>$Q_s$</th>
<th>$S_s$</th>
<th>$J_s$</th>
<th>$C_s$</th>
<th>$D_s$</th>
<th>$F_s$</th>
<th>$D_m$</th>
<th>$Q_m$</th>
<th>$S_m$</th>
<th>$A_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3store</td>
<td>v</td>
<td>S</td>
<td>U</td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>4store</td>
<td>v</td>
<td>S</td>
<td>U</td>
<td>o</td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td>h</td>
<td>p</td>
<td></td>
<td>n</td>
</tr>
<tr>
<td>Virtuoso</td>
<td>v</td>
<td>G</td>
<td>S</td>
<td>Ulo</td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td>TA</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>RDF-3X</td>
<td>v</td>
<td>6</td>
<td>S</td>
<td>Ulo</td>
<td>o</td>
<td>R</td>
<td></td>
<td></td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>Hexastore</td>
<td>v</td>
<td>6</td>
<td>o</td>
<td>Ul</td>
<td></td>
<td></td>
<td>n</td>
<td></td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>Apache Jena</td>
<td>p</td>
<td>S</td>
<td></td>
<td>Ulo</td>
<td>R</td>
<td>m</td>
<td>R</td>
<td></td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>SW-Store</td>
<td>h</td>
<td>S</td>
<td>Uo</td>
<td></td>
<td>c</td>
<td>m</td>
<td>c</td>
<td></td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>BitMat</td>
<td>v</td>
<td>m</td>
<td>S</td>
<td>Ulo</td>
<td>p</td>
<td>c</td>
<td></td>
<td></td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>AllegroGraph</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>h</td>
<td>p</td>
<td>m</td>
</tr>
<tr>
<td>Hadoop/HBase</td>
<td>h</td>
<td>c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p</td>
<td>m</td>
</tr>
<tr>
<td>$H_2 RDF+$</td>
<td>v</td>
<td>6</td>
<td>S</td>
<td>Ulo</td>
<td>c</td>
<td>c</td>
<td></td>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TriAD</td>
<td>v</td>
<td>6</td>
<td>S</td>
<td></td>
<td>p</td>
<td>R</td>
<td></td>
<td></td>
<td>h</td>
<td>p</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>TripleBit</td>
<td>p</td>
<td>6</td>
<td>o</td>
<td>Uo</td>
<td></td>
<td>c</td>
<td></td>
<td></td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>Trinity.RDF</td>
<td>g</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>gStore</td>
<td>g</td>
<td>a</td>
<td>S</td>
<td>Ul</td>
<td>g</td>
<td>n</td>
<td>c</td>
<td></td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>Neo4j</td>
<td>g</td>
<td>a</td>
<td>Olo</td>
<td>g</td>
<td>f</td>
<td>c</td>
<td></td>
<td></td>
<td>p</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>big3store</td>
<td>v</td>
<td>3</td>
<td>S</td>
<td></td>
<td>n</td>
<td>R</td>
<td>l</td>
<td>R</td>
<td>n</td>
<td>d</td>
<td>r</td>
<td>p</td>
</tr>
</tbody>
</table>

**Table I. Properties of RDF Storage Managers**

- **$T_s$**: Tuple space
- **$I_s$**: Insertion space
- **$Q_s$**: Query space
- **$S_s$**: Storage space
- **$J_s$**: Join space
- **$C_s$**: Copy space
- **$D_s$**: Delete space
- **$F_s$**: Forward space
- **$D_m$**: Direct mapping
- **$Q_m$**: Query mapping
- **$S_m$**: Storage mapping
- **$A_m$**: Access mapping
Challenges in designing big3store

- Automatic distribution and replication of RDF data
- Intelligent distribution of query processing
- Dynamic updates in RDF storage manager
- Multi-threaded architecture of query executor
- Distributed cache for query executor
GraphSM topics

THEORY
Search in graph databases
Algebra and logic of graphs
Expressive power of graph query languages
Formalizations of graph databases

DATA MODELLING
Graph data modelling
Advanced graph data models
Data modelling for specific graph applications

STORAGE MANAGERS
Indexing methods for graph processing
Storage systems for large-scale graph databases
Automatic distribution and replication of graph databases
Storage managers for specific graph applications

QUERY PROCESSING
Flexible query answering on graph-structured data
SPARQL query processing
Intelligent distribution of SPARQL query processing
Using map-reduce operations for graph processing

ANALYSIS AND MANAGEMENT
Graph pattern matching
Knowledge discovery from graphs
Algorithms for graph database processing
Analysis of graph databases from specific domains
Information retrieval on graph-structured data

APPLICATIONS
Biological and medical graph databases
Graph processing for Social Networks
Visualizing, browsing and navigating graph data
User-interfaces for graph databases
GraphSM 2015 Introduction

Iztok Savnik (University of Primorska FAMNIT) and
Kiyoshi Nitta (Yahoo Japan Research)