Semantic Query Answering with Time-Series Graphs

Leo Ferres ¹, Michel Dumontier ²,³, Natalia Villanueva-Rosales ³
¹Human-Oriented Technology Laboratory, ²Department of Biology, ³School of Computer Science, Carleton University

3rd International Workshop on Vocabularies, Ontologies and Rules for The Enterprise (VORTE 2007)
October 15 2007, Annapolis, Maryland, USA
Outline

- Motivation
- Methodology
- Statistical Graph Ontology (SGO)
- Augmenting the SGO
- Ontology population
- Semantic Query Answering over Time-Series Graphs
- Conclusions and Future Work
Motivation

Statistical Graphs

- Used for data visualization in most enterprises.
- Created with a communicative purpose.
- Stored as unstructured images or binary objects (lack of expressive semantics).
- Limited exchange, integration, discovery and query answering of graph information.
- Not easily accessible in certain contexts (small devices, visually impaired people)
In XML/XMLS?
Datatypes, restricted values, arbitrary name elements, no meaning.
This solution

**OWL ontology** for representation, exchange, and semantic query answering of statistical graph data using automated reasoning.
A Graph has a Primary title and a Plot.
A Plot has X and Y axis.
X axis has a Primary category and a Category data axis.
X axis might have a Secondary Category.
A Plot has Series.
Series have at least 2 Data points.
Data points have category data and value data.

Operating profits up slightly in second quarter
$ billions  Seasonally adjusted

Methodology

Requirements

- **iGraph-Lite** (NLP interface to an enriched XML representation) [Ferres et. al., ICCHP 06]

- **Competency questions** [Uschold et. al., KE 96]
  - What is the graph title?
  - Do the x-axis categories of a graph correspond to years?
  - ...
Methodology (2)

Corpus of graphs

- Statistics Canada “The Daily”
- 10 years (1996 – 2006)
- 5060 graphs
  - 2883 line Graphs (63.3%)
Methodology (3)

- Deconstruction of graphs, collection of concepts.
- Added concepts in the OOXML, ODF and GraphRep ontology (coverage).
- Creation of simple taxonomy (is-a relation).
- Definition of relations between objects
- Mapped to upper level ontologies (BFO, BRO).
- Three layer modelling approach.
Modelling tools

**OWL-DL**, Web Ontology Language based in Description Logics (DL).

- Balance between **decidability** and **expressivity**
- **Sound** and **complete** algorithms implemented for reasoning problems.
- Complex concepts are defined out of simpler ones.
- **OWL 1.1** specification, *SHOIQ* (*D*)
- Manchester OWL syntax [Horridge et al., OWLED 06].
Layer 1: Statistical Graph Ontology (SGO)

- Describes a graph (its components and how they are related).
- 62 classes (concepts)
  - E.g. Graph, Title, CategoryData, etc.
- 22 properties
  - 16 object properties. E.g. hasTitle.
  - 6 datatype properties. E.g. hasValue.
- Definitions of the form:
  A Graph hasPart one or more Plot and may hasTitle one or more Title and hasSource Source of origin.

http://ontology.dumontierlab.com/statistical-graph-primitive
Layer 2: Augmenting the SGO

- More complex definitions using more restrictive operations like disjunction, union, intersection, class equivalence, existential and universal restrictions and qualified cardinality restrictions (inferences using classification and realization).

  GraphTitle is equivalent to a Title that isTitleOf some Graph

- Contextual Knowledge
  - Ontology mapping to upper level ontologies (BFO and BRO) and time interval ontology. Ontology reuse.
    E.g. FirstQuarter, Quarter, Year.

- Restrictions hold in any application.

http://ontology.dumontierlab.com/statistical-graph-complex
http://ontology.dumontierlab.com/time-interval
Layer 3: Augmenting the SGO (2)

iGraph Requirements

- **Application dependent** restrictions, not expected to hold outside the application.
  - E.g. LineGraph

- Useful for data exchange.

http://ontology.dumontierlab.com/statistical-graph-igraph
Ontology Population

- From **iGraph-Lite**
  - Enriched XML documents after being annotated by iGraph-Lite with x-axis categories, titles, etc.

- To concept **instances** in the ontology using OWL RDF/XML syntax. E.g. “1997” is an instance of Year.

- **XLST** transformations.
Q1: Retrieve all the datapoints in the graph.

DataPoint that isPartOf some Graph

E.g.

datapoint1 isPartOf series1, series1 isPartOf graph

Transitivity.

Using Protégé 4 alpha (build 53), FACT++ DL reasoner and Manchester Syntax.
Q2 : Retrieve all the value data for the second quarter of any year.

ValueData that isPartOf some
(DataPoint that hasPart some (SecondQuarter))

E.g.

y7 isPartOf datapoint7, datapoint7 hasPart x7,
x7 type SecondQuarter

Ontology mapping.

Using Protégé 4 alpha (build 53), FACT++ DL reasoner and Manchester Syntax.
Q3: Retrieve all series that contain time intervals (time series).

\[ \text{TimeSeries} \equiv \text{Series that hasPart some TimeInterval} \]

E.g.

\[ \text{series1 hasPart datapoint7, datapoint7 hasPart x7, x7 type SecondQuarter, SecondQuarter subClassOf Quarter, Quarter subClassOf TimeInterval} \]

Class def.

Using Protégé 4 alpha (build 53), FACT++ DL reasoner and Manchester Syntax.
Q4: Retrieve all time series graphs.

\[ \text{TimeSeriesGraph} \equiv \text{Graph that hasPart some TimeSeries} \]

E.g.

\[
\begin{align*}
\text{series1 hasPart datapoint7, datapoint7 hasPart x7,} \\
\text{x7 type SecondQuarter,} \\
\text{SecondQuarter subClassOf Quarter,} \\
\text{Quarter subClassOf TimeInterval,} \\
\text{graph hasPart series1}
\end{align*}
\]

Across graphs. Using Protégé 4 alpha (build 53), FACT++ DL reasoner and Manchester Syntax.
Other implications

Data exchange

- **Standard** (syntactic) representation of the data for machine consumption.
- Meaning (semantics) unambiguous.
- Three layer **flexible model**.
- Exchange between statistical agencies.
- XSLT to return **new** information to applications.
Conclusions and Future work

- Provide semantic query answering over time-series graphs.
- Data model for graph information exchange.
- Increased explicit knowledge in statistical graphs for certain demographics (accessibility).
- **Enhanced** iGraph-Lite using **Semantic Web**.
- Still, some challenges:
  - Spatial and temporal reasoning.
  - “Type” of graphs according to their content.
Acknowledgements
Semantic Query Answering with Time-Series Graphs

Leo Ferres ¹, Michel Dumontier ²,³, Natalia Villanueva-Rosales ³
¹Human-Oriented Technology Laboratory, ²Department of Biology,
³School of Computer Science, Carleton University

3rd International Workshop on Vocabularies, Ontologies and Rules for The Enterprise (VORTE 2007)
15 October 2007, Annapolis, Maryland, USA