OBDA: Theory and Practice

Ian Horrocks
Information Systems Group
What is an Ontology?
What is an Ontology?

A fundamental branch of **metaphysics**

- Studies “being” or “existence” and their **basic categories**
- Aims to find out what **entities** and **types of entities** exist
What is an Ontology?

A conceptual model of (some aspect of) the world
What is an Ontology?

A conceptual model of (some aspect of) the world

- Introduces **vocabulary** relevant to domain, e.g.:
  - Anatomy
What is an Ontology?

A conceptual model of (some aspect of) the world

- Introduces **vocabulary**
  - relevant to domain, e.g.:
    - Anatomy
    - Cellular biology
What is an Ontology?

A conceptual model of (some aspect of) the world

- Introduces **vocabulary**
  relevant to domain, e.g.:
  - Anatomy
  - Cellular biology
  - Aerospace
What is an Ontology?

A conceptual model of (some aspect of) the world

- Introduces **vocabulary** relevant to domain, e.g.:
  - Anatomy
  - Cellular biology
  - Aerospace
  - Cell Phones
What is an Ontology?

A conceptual model of (some aspect of) the world

- Introduces **vocabulary** relevant to domain, e.g.:
  - Anatomy
  - Cellular biology
  - Aerospace
  - Cell Phones
  - Oil and gas
What is an Ontology?

A conceptual model of (some aspect of) the world

- Introduces **vocabulary** relevant to domain
- Specifies **meaning** (semantics) of terms

Oil pipeline is a pipeline from a facility that is an oil facility
What is an Ontology?

A conceptual model of (some aspect of) the world

- Introduces **vocabulary** relevant to domain
- Specifies **meaning** (semantics) of terms

Oil pipeline is a pipeline from a facility that is an oil facility

- **Formalised** using suitable logic

\[
\forall x. [\text{OilPipeline}(x) \rightarrow \text{Pipeline}(x) \land \\
\exists y. [\text{fromFacility}(x, y) \land \\
\text{OilFacility}(y)]]
\]
What is an Ontology?

A conceptual model of (some aspect of) the world

- Introduces **vocabulary** relevant to domain
- Specifies **meaning** (semantics) of terms

  Oil pipeline is a pipeline from a facility that is an oil facility

- **Formalised** using suitable logic

  \[
  \text{OilPipeline} \sqsubseteq \text{Pipeline} \sqsupseteq \exists \text{fromFacility.OilFacility}
  \]
Applications: Semantic Web

Olympics modern pentathlon: Samantha Murray wins silver

By Ollie Williams

Samantha Murray won modern pentathlon silver for Great Britain in the final event of the London 2012 Olympic Games.

The 22-year-old world bronze medallist started the run-shoot finale eight seconds off the pace, but caught up to finish a superb second.

Lithuania’s Laura Asadauskaitė won as GB’s Mhairi Spence, the world champion, finished 21st.
Applications: Semantic Web
Semantic Technologies

- SemWeb motivated development of robust infrastructure:
Semantic Technologies

- SemWeb motivated development of robust infrastructure:
  - Languages

![W3C OWL](image1.png)
![RDFa](image2.png)
![POWDER](image3.png)
![SKOS](image4.png)
![RDF](image5.png)
![SPARQL](image6.png)
Semantic Technologies

- SemWeb motivated development of robust infrastructure:
  - Languages
  - Storage and querying

- Hermit
- FaCT++
- Racer
- pellet
- uOnto Querying ONTOlogies
- OWLIM
- CEL
- jena
- TrOWL

Optique
Semantic Technologies

- SemWeb motivated development of robust infrastructure:
  - Languages
  - Storage and querying
  - Development tools
Semantic Technologies

- SemWeb motivated development of robust infrastructure:
  - Languages
  - Storage and querying
  - Development tools

- Increasingly used in “Intelligent Information Systems”, and in particular **Ontology Based Data Access (OBDA)**
Applications: HCLS

- **OBO foundry** includes more than 100 biological and biomedical ontologies
- **Siemens** “actively building OWL based clinical solutions”
- **SNOMED-CT** (Clinical Terms) ontology
  - used in healthcare systems of more than 15 countries, including Australia, Canada, Denmark, Spain, Sweden and the UK
  - also used by major US providers, e.g., Kaiser Permanente
  - ontology provides common vocabulary for recording clinical data
Applications: Energy Supply Industry

- **EDF Energy** offer personalised energy saving advice to every customer

- **OWL ontology** used to model relevant environmental factors

- **Oxford’s HermiT reasoner** used to match customer circumstances with relevant pieces of advice
Applications: Intelligent Mobile Platform

- **Samsung** developing Intelligent Mobile Platform to support context-aware applications

- IMP monitors environment via **sensor data** (GPS, compass, accelerometer, ...)

- **OWL ontology** used to model environment and **infer context** (e.g., coffee with friends)

- Applications exploit context to enable more **intelligent behaviour**
Applications: OBDA
Applications: OBDA
Applications: OBDA

Pipelines from oil facilities?
Applications: OBDA

Pipelines from oil facilities?
Applications: OBDA

Q(x) ← Pipeline(x) ∧
fromFacility(x, y) ∧
OilFacility(y)

Pipelines from oil facilities?
Applications: OBDA

Pipelines from oil facilities?

\[ Q(x) \leftarrow \text{Pipeline}(x) \land \text{fromFacility}(x, y) \land \text{OilFacility}(y) \]
Applications: OBDA

Pipelines from oil facilities?

Pipeline(p1)
fromFacility(p1, f1)
OilFacility(f1)
OilPipeline(p2)
fromFacility(p2, f2)
OilFacility(f2)
OilPipeline(p3)

Q(x) ← Pipeline(x) ∧
fromFacility(x, y) ∧
OilFacility(y)
Applications: OBDA

Pipelines from oil facilities?

Pipeline(p1)
fromFacility(p1, f1)
OilFacility(f1)
OilPipeline(p2)
fromFacility(p2, f2)
OilFacility(f2)
OilPipeline(p3)

OilPipeline ⊆ Pipeline ⊊ ∃ fromFacility.OilFacility
Applications: **OBDA**

Pipeline\((p_1)\)

fromFacility\((p_1, f_1)\)

OilFacility\((f_1)\)

OilPipeline\((p_2)\)

fromFacility\((p_2, f_2)\)

OilFacility\((f_2)\)

OilPipeline\((p_3)\)

OilPipeline \(\sqsubseteq\) Pipeline \(\sqcap\)

\(\exists\) fromFacility. OilFacility

\(Q(x) \leftarrow \text{Pipeline}(x) \land\)

fromFacility\((x, y) \land\)

OilFacility\((y)\)

Pipelines from oil facilities?
Applications: **OBDA**

- Pipeline(p₁)
  - fromFacility(p₁, f₁)
- OilFacility(f₁)
- OilPipeline(p₂)
  - fromFacility(p₂, f₂)
- OilFacility(f₂)
- OilPipeline(p₃)

OilPipeline ⋐ Pipeline ⋐ ∃fromFacility.OilFacility

Pipelines from oil facilities?

p₁, p₂, p₃

Q(x) ← Pipeline(x) ∧ fromFacility(x, y) ∧ OilFacility(y)
OBDA: Practical Issues

- Scalability is always an issue, particularly when we have large and heterogeneous data sources.

- Can be a large (startup) cost for building and maintaining ontologies (and mappings).

- Users may still have difficulty formulating queries.
Optique Objectives

- Provide **semantic end-to-end connection** between users and data sources
Optique Objectives

- Provide **semantic end-to-end connection** between users and data sources
- Enable users to rapidly formulate **intuitive queries** using familiar vocabularies and conceptualisations
Optique Objectives

- Provide **semantic end-to-end connection** between users and data sources
- Enable users to rapidly formulate **intuitive queries** using familiar vocabularies and conceptualisations
- Return **timely answers** from large scale and heterogeneous data sources
Optique Solution

Diagram showing the process flow of Optique Solution with stages such as Application, Query Formulation, Ontology & Mapping Management, Query Transformation, Query Planning, Stream Adapter, Query Execution, and Q.E.
Query rewriting:
- uses ontology & mappings
- computationally hard
- ontology & mappings small
Query rewriting:
- uses ontology & mappings
- computationally hard
- ontology & mappings small

Query evaluation:
- ind. of ontology & mappings
- computationally tractable
- data sets very large
Query Rewriting

Given ontology $\mathcal{O}$ query $\mathcal{Q}$ and mappings $\mathcal{M}$:
Query Rewriting

Given ontology $O$ query $Q$ and mappings $M$:

- **Rewrite** $Q \rightarrow Q'$ s.t. answering $Q'$ without $O$ equivalent to answering $Q$ w.r.t. $O$ *for any dataset*
Query Rewriting

Given ontology $\mathcal{O}$ query $Q$ and mappings $\mathcal{M}$:

- **Rewrite** $Q \rightarrow Q'$ s.t. answering $Q'$ without $\mathcal{O}$ equivalent to answering $Q$ w.r.t. $\mathcal{O}$ for any dataset

$$Q(x) \leftarrow \text{Pipeline}(x) \land \text{fromFacility}(x, y) \land \text{OilFacility}(y)$$
Query Rewriting

Given ontology $\mathcal{O}$ query $Q$ and mappings $\mathcal{M}$:

- **Rewrite** $Q \rightarrow Q'$ s.t. answering $Q'$ without $\mathcal{O}$ equivalent to answering $Q$ w.r.t. $\mathcal{O}$ for any dataset

\[ Q(x) \leftarrow \text{Pipeline}(x) \land \text{fromFacility}(x, y) \land \text{OilFacility}(y) \lor \text{OilPipeline}(x) \]
Query Rewriting

Given ontology $O$ query $Q$ and mappings $M$:

- **Rewrite** $Q \rightarrow Q'$ s.t. answering $Q'$ without $O$ equivalent to answering $Q$ w.r.t. $O$ for any dataset

- **Map** ontology queries $\rightarrow$ DB queries (typically SQL) using mappings $M$ to rewrite $Q'$ into a DB query

$Q(x) \leftarrow \text{Pipeline}(x) \land \text{fromFacility}(x, y) \land \text{OilFacility}(y) \lor \text{OilPipeline}(x)$
Query Rewriting

Given ontology $O$ query $Q$ and mappings $M$:

- **Rewrite** $Q \rightarrow Q'$ s.t. answering $Q'$ without $O$ equivalent to answering $Q$ w.r.t. $O$ for any dataset
- **Map** ontology queries $\rightarrow$ DB queries (typically SQL) using mappings $M$ to rewrite $Q'$ into a DB query

```
SELECT name FROM pipeline,fromfac,facility . . .
UNION SELECT name FROM oilpipeline
```
Query Rewriting

Given ontology $\mathcal{O}$ query $Q$ and mappings $\mathcal{M}$:

- **Rewrite** $Q \rightarrow Q'$ s.t. answering $Q'$ without $\mathcal{O}$ equivalent to answering $Q$ w.r.t. $\mathcal{O}$ for any dataset
- **Map** ontology queries $\rightarrow$ DB queries (typically SQL) using mappings $\mathcal{M}$ to rewrite $Q'$ into a DB query
- **Evaluate** (SQL) query against DB

```
SELECT name FROM pipeline,fromfac,facility . . .
UNION SELECT name FROM oilpipeline
```

---

Optique
**Optique Solution**

**Query rewriting:**
- uses ontology & mappings
- computationally hard
- ontology & mappings small

**Query evaluation:**
- ind. of ontology & mappings
- computationally tractable
- data sets very large

---

![Diagram of Optique solution](image-url)
**Query rewriting:**
- uses ontology & mappings
- computationally hard
- ontology & mappings small

**Query evaluation:**
- ind. of ontology & mappings
- computationally tractable
- data sets very large

**Other features:**
- support for query formulation
Query Formulation

Optique
**Optique Solution**

### Query Rewriting:
- Uses ontology & mappings
- Computationally hard
- Ontology & mappings small

### Query Evaluation:
- Ind. of ontology & mappings
- Computationally tractable
- Data sets very large

### Other Features:
- Support for query formulation
- Ontology & mappings extended “on the fly”
Query rewriting:
- uses ontology & mappings
- computationally hard
- ontology & mappings small

Query evaluation:
- ind. of ontology & mappings
- computationally tractable
- data sets very large

Other features:
- support for query formulation
- Ontology & mappings extended “on the fly”
- IT-expert oversees O&M management

Optique Solution
Optique Solution

Query rewriting:
- uses ontology & mappings
- computationally hard
- ontology & mappings small

Query evaluation:
- ind. of ontology & mappings
- computationally tractable
- data sets very large

Other features:
- support for query formulation
- Ontology & mappings extended “on the fly”
- IT-expert oversees O&M management
- Adapter to support streaming data

Diagram:
- Application
- Query Formulation
- Ontology & Mapping Management
- Query Transformation
- Query Planning
- Stream Adapter
- Query Execution
- Q.E.

Optique
Query rewriting:
- uses ontology & mappings
- computationally hard
- ontology & mappings small

Query evaluation:
- ind. of ontology & mappings
- computationally tractable
- data sets very large

Other features:
- support for query formulation
- Ontology & mappings extended “on the fly”
- IT-expert oversees O&M management
- Adapter to support streaming data
- Distributed query execution
Thank you for listening

Any questions?

Optique