

## My research areas

Michael Zakharyashev <http://www.dcs.bbk.ac.uk/~michael/>

(with all my publications and funded research projects)

- Knowledge representation and reasoning in AI
  - ontologies, description logics
  - query languages
  - ontology-mediated queries
  - temporal and spatial representation, reasoning and querying
- Semantic technologies
  - ontology-based data access
  - Statoil use case
  - Siemens use case
- Mathematical logic
  - modal logic, temporal logic, spatial logic, dynamic logic
  - complexity theory

# Querying resources on the Web

## Resource Description Framework (RDF) <http://www.w3.org/RDF/>

a general method for conceptual description or modelling of information for web resources

- represents data and metadata as triples (subject, predicate, object)

- Siemens phone data

dbr:Siemens\_C25 dbo:manufacturer dbr:Siemens .  
dbr:Siemens\_C25 rdf:type yago:SiemensMobilePhones .

- IBM phone data

dbr:IBM\_Simon dbo:manufacturer dbr:IBM .  
dbr:IBM\_Simon rdf:type yago:IBMMobilePhones .

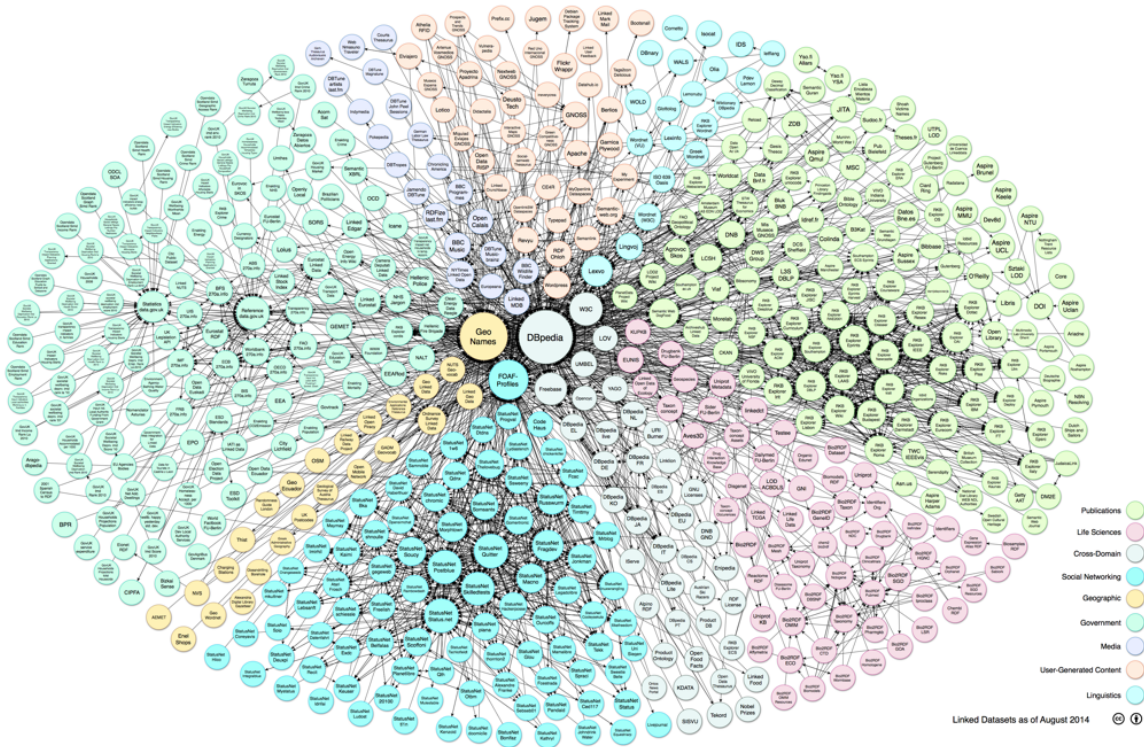
## SPARQL query:

```
SELECT ?C
WHERE
{
  ?C rdf:type dbo:Organisation .
  ?C foaf:made ?P .
  ?P rdf:type yago:Telephone .
}
```

## Ontology

```
 $\forall x, y (manufacturer(x, y) \rightarrow maker(x, y))$ 
 $\forall x, y (maker(x, y) \rightarrow made(y, x))$ 
 $\forall x, y (manufacturer(x, y) \rightarrow organisation(y))$ 
 $\forall x (SiemensMobilePhone(x) \rightarrow Telephone(x))$ 
 $\forall x (IBMMobilePhone(x) \rightarrow Telephone(x))$ 
```

# Linked Data





## Querying relational data: Statoil case

(from Norwegian Petroleum Directorate's FactPages)

show me the wellbores completed before 2008 where Statoil as a drilling operator sampled less than 10 meters of cores



5 days later:

```
SELECT DISTINCT cores.wlbName, cores.lenghtM, wellbore.wlbDrillingOperator, wellbore.wlbCompletionYear
FROM
```

```
( (SELECT wlbName, wlbNpdidWellbore, (wlbTotalCoreLength * 0.3048) AS lenghtM
  FROM wellbore_core
  WHERE wlbCoreIntervalUom = '(ft)' )
```

```
UNION
```

```
(SELECT wlbName, wlbNpdidWellbore, wlbTotalCoreLength AS lenghtM
  FROM wellbore_core
  WHERE wlbCoreIntervalUom = '(m)' )
```

```
) as cores,
```

```
( (SELECT wlbNpdidWellbore, wlbDrillingOperator, wlbCompletionYear
  FROM wellbore_development_all
```

```
UNION
```

```
(SELECT wlbNpdidWellbore, wlbDrillingOperator, wlbCompletionYear
  FROM wellbore_exploration_all )
```

```
UNION
```

```
(SELECT wlbNpdidWellbore, wlbDrillingOperator, wlbCompletionYear
  FROM wellbore_shallow_all )
```

```
) as wellbore
```

```
WHERE wellbore.wlbNpdidWellbore = cores.wlbNpdidWellbore
```

```
...
```

**In STATOIL:**

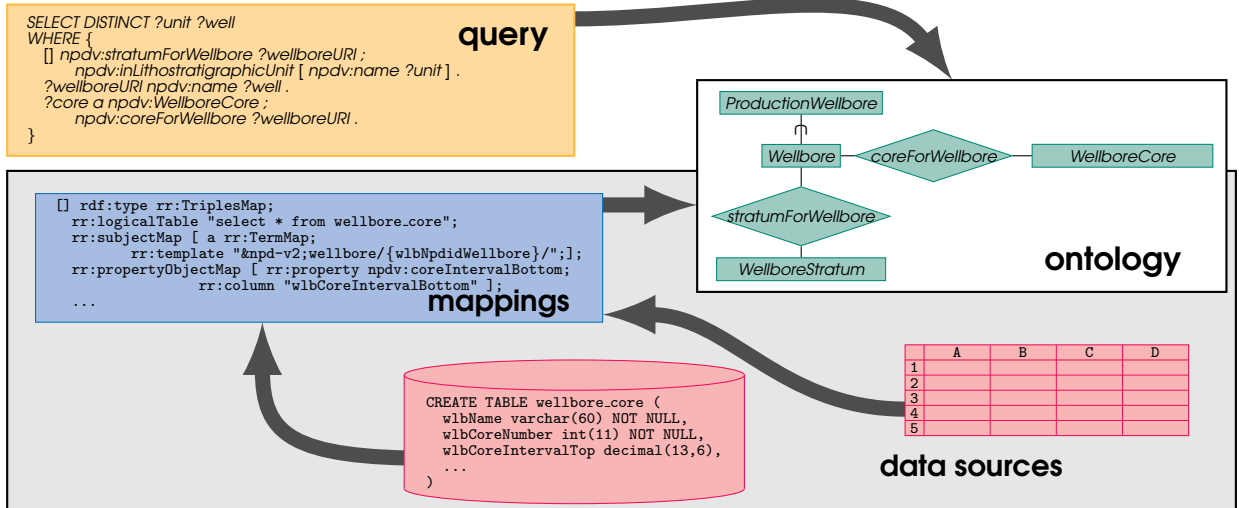
**1,000 TB of relational data**

**2,000 tables**

**different schemas**

**30-70% of time on data gathering**

# Ontology-based data access (OBDA)



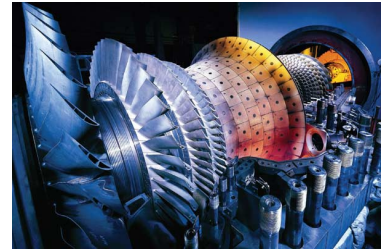
## Ontology

- gives a high-level conceptual view of the data
- provides a convenient & natural vocabulary for user queries
- enriches incomplete data with background knowledge

## Temporal OBDA: Siemens case

Siemens remote-diagnostic centres analyse data by querying aggregated sensor measurements

turbine.Id	dateTime	activePower	rotorSpeed	mainFlame	...
t03	2015-04-04 12:20:48	2	1550	0	
t03	2015-04-04 12:20:49	1.8	1400	null	
t01	2015-04-04 12:20:52	1.7	1350	1	
		...			



**Typical query:** when an **active power trip** occurred, i.e., the active power was above 1.5MW for a period of at least 10 seconds, 3 seconds after which there was a period of at least one minute where active power was below 0.15MW

**Metric temporal logic ontology, II**, with axioms such as

$$\text{ActivePowerTrip}(v) \leftarrow \text{Turbine}(v) \wedge \exists_{[0,1m]} \text{ActivePowerBelow0.15}(v) \wedge \diamond_{[60s,63s]} \exists_{[0,10s]} \text{ActivePowerAbove1.5}(v)$$

**Ontology-mediated query** (II,  $\text{ActivePowerTrip}(t)@x$ )

(find turbines  $t$  and moments  $x$  at which  $t$  had an active power trip)

# Ontology Applications: Healthcare

**SNOMED CT** is a formal ontology of medical terms, synonyms and definitions used in clinical documentation and reporting most comprehensive, multilingual clinical healthcare terminology in the world

- electronic record systems
- catalogues of services
- clinical decision support
- laboratory reporting
- genetic databases
- ...

used in most developed countries in the world

The screenshot displays the Protege OWL editor interface. On the left, the 'SUBCLASS EXPLORER' shows the 'Asserted Hierarchy' for the project 'Thesaurus-small'. The hierarchy is as follows:

- owl:Thing
  - Abnormal\_Cell\_Kind
  - Activity\_Kind
  - Anatomy\_Kind
    - Anatomic Structure, System, or Substance
      - Body Cavity
      - Body Fluid or Substance
      - Body Part
        - Anatomic Surface
        - Brain part
          - Brain tissue structure
          - Brain ventricle part
          - Cavum vergae
          - Extrapyramidal system structure
          - Infratentorial brain structure
          - Interstitial tissue of brain
          - Lateral proper fasciculus
          - Lateral spinorubral tract
          - Meninges of cerebral convex surface
          - Optic nerve chiasma
          - Optical layer and corpus striatum
          - Periventricular gray matter
          - Prosencephalon structure
          - Structure of basal meninges
          - Structure of dorsal spinocerebellar tract
          - Structure of falx cerebelli
          - Structure of falx cerebri
          - Structure of gyrus of brain
          - Structure of half of brain lateral to midsagittal plane
          - Structure of lateral root of optic tract
          - Structure of lateral spinothalamic tract
          - Structure of lobe of brain

On the right, the 'CLASS EDITOR for 'Brain part'' (Instance of owl:Class) is shown. It displays the following properties and values:

Property	Value
rdfs:comment	
bp:ontologyid	44777
Preferred_Name	Brain part
prov:importCreator	ttania
prov:importDate	2011-07-15T17:32:51
prov:ontologyName	SNOMED Clinical Terms
prov:ontologyVersion	2010_07_31
prov:termUrl	<a href="http://bioportal.bioontology.org/visualize/44777?conc=05">http://bioportal.bioontology.org/visualize/44777?conc=05</a>
Synonym	Part of brain
Synonym	Brain part (body structure)

Below the table, there are sections for 'Body Part' and other related classes, with icons for adding and removing instances.

# Semantic Web Applications: Media

**BBC** website for the Football World Cup 2010: 32 teams, 8 groups, 776 players  
too many pages to create, too few journalists

Semantic Web technologies used:

- ontology describes the interrelation between facts of the World Cup
- such metadata stored as RDF triples

- e-governments
- libraries
- news papers
- dbpedia
- . . .

The screenshot shows the BBC website's 'SPORT' section for the 'WORLD CUP 2010'. The main navigation bar includes 'SPORT', 'FOOTBALL', 'WORLD CUP 2010', 'GROUPS & TEAMS', 'FIXTURES & RESULTS', 'VIDEO', and 'BBC COVERAGE'. The main content area is titled 'England' and features a 'Match report' section with a table of recent matches:

Match	Date	Match report
England 1-1 United States	Saturday, 12 June	Match report
England 0-0 Algeria	Friday, 18 June	Match report
Slovenia 0-1 England	Wednesday, 23 June	Match report
Germany 4-1 England	Sunday, 27 June	Match report

Below the match reports is a 'Latest stories' section with two main items:

- Cerrard consults future to England** (with a 'NEW' tag):
  - England sponsorship likely to end
  - Capello to remain England manager
  - Mueller blames England imbalance
  - Capello receives Gartside baking
- Pressure got to Rooney - Ferguson**:
  - FA unfit for purpose says Caborn
  - England's fear of crossing borders
  - England duo bypass London event
  - Barwick baffled by dismal England

On the right side, there is a 'Group C' table:

A	U	C	D	E	F	G	H
Group C Teams	W	D	L	GD	PTS		
USA	1	2	0	1	5		
England	1	2	0	1	5		
Slovenia	1	1	1	0	4		
Algeria	0	1	2	-2	1		

Below the table is a 'Features' section with a sub-heading 'German lessons: Jürgen Klinsmann on how to revolutionise England'. It includes a small image of Jürgen Klinsmann and a list of related content:

- A German view on English football
- Redknapp backs England to shine
- BBC pundits on England
- Roy Hodgson Q&A
- World Cup goals analysis

At the bottom right, there is an 'Around the web' section with links to 'BBC Search+ country page' and 'England Fifa Profile'.



## Research projects

### 'Practical'

- Extracting RDF data from unstructured texts
- Visualising SPARQL queries
- Extracting ontologies and mappings from databases
- Application for querying DBLP (computer science bibliography)
- Data integration via RDF
- Algorithms for answering temporal ontology-mediated queries
- Extracting 'modules' from ontologies
- Ontology-based information systems in your favourite area  
(movies, football, food, travels, etc.)

### 'Theoretical'

- Characterise ontology-mediated queries that can be reduced to SQL queries or SPARQL queries (with or without transitive closure)
- Given an ontology in an expressive language, determine whether it can be 'equivalently' rewritten in a simpler language