Ontology-Based Data Access in the Medical Informatics Platform

Human Brain Project
Unifying our understanding of the human brain.

European Commission

Ontologies play a key role in representing the **knowledge/semantics** of a domain of interest. Their use in the medical domain has been extensively studied in the past since they provide a standard terminology with well-defined meaning enabling **semantic access** and promoting **interoperability**. Bridging ontologies and data is of paramount importance for the MIP. Given a query defined using ontology terminology, provide answers that reflect both the data and the knowledge captured by the ontology. In this poster we present how Ontology Based Data Access can aid MIP regarding data integration.

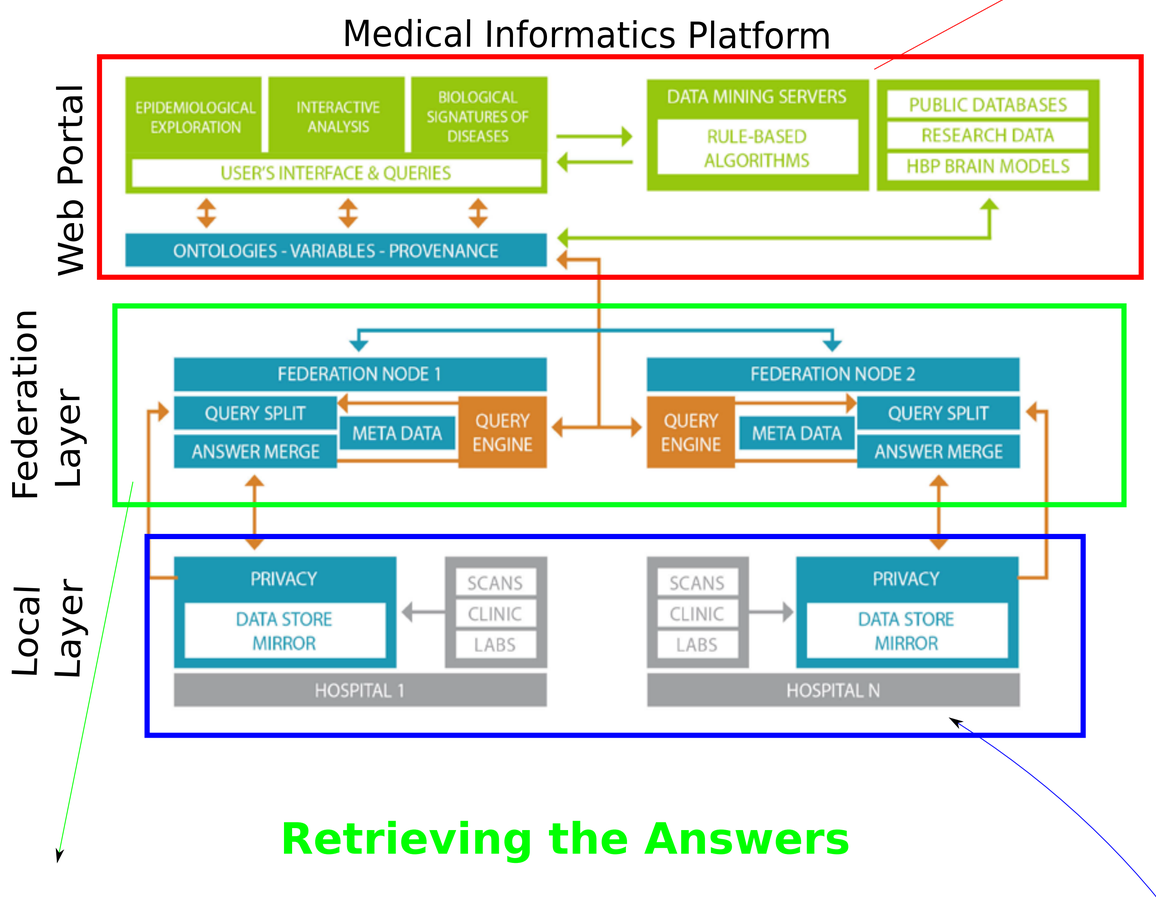
Vasilis Vassalos
Athens University of Economics and Business

Giorgos Stoilos
Athens University of Economics and Business

Tassos Venetis
Athens University of Economics and Business

I want to know the number of cases suffering from some form of Dementia.

So many possibilities:
Dementia can be caused by diseases with ICD-10 codes G30.9 (for Alzheimers), or G20 (for Parkinsons), or G31 (for Pick's dementia), or G10 (for Huntington's), or ...



Retrieving the Answers

– Low level mappings dictate how data are mapped to ontology terms

$$\text{diagnostic}(\mathbf{pid}, G30.9, \mathbf{date}, \text{“CHUV”}) \rightsquigarrow_{m_1} \text{Patient}(\mathbf{pid}) \wedge \text{hasEffect}(\mathbf{pid}, ad) \wedge \text{AlzheimersDementia}(ad)$$

$$\text{diagnostic}(\mathbf{pid}, G20, \mathbf{date}, \text{“CHUV”}) \rightsquigarrow_{m_2} \text{Patient}(\mathbf{pid}) \wedge \text{hasEffect}(\mathbf{pid}, pd) \wedge \text{ParkinsonsDementia}(pd)$$

$$\cdots$$

– Translate to low-level SQL using the mappings and evaluate it over the data

```
SELECT d.pid
FROM diagnostic AS d
WHERE d.code="G30.9"
   UNION
SELECT d.pid
FROM diagnostic AS d
WHERE d.code="G20"
   UNION
...
```

Ontologies

– Represent domain knowledge in an abstract way e.g., SNOMED CT

$$\text{AlzheimersDementia} \sqsubseteq \text{Dementia}$$

$$\text{ParkinsonsDementia} \sqsubseteq \text{Dementia}$$

$$\text{PicksDementia} \sqsubseteq \text{Dementia}$$

$$\text{ClostridiumNovyi} \sqsubseteq \exists \text{hasActiveIngredient}.\text{Toxoid}$$

$$\cdots$$

Ontology-Based Query Answering

– Issue queries using ontology vocabulary and not low data-level details (ICD codes, etc)

$$\mathcal{Q} = Q(x) \leftarrow \text{Pat}(x) \wedge \text{hasEff}(x, y) \wedge \text{Dem}(y)$$

– Translate/Rewrite/Expand User Query Using a **Reasoning System**

Compute so-called rewriting $\mathcal{R}$

$$\mathcal{Q}_1 = Q(x) \leftarrow \ldots \wedge \text{AlzheimersDementia}(y)$$

$$\mathcal{Q}_2 = Q(x) \leftarrow \ldots \wedge \text{ParkinsonsDementia}(y)$$

$$\mathcal{Q}_3 = Q(x) \leftarrow \ldots \wedge \text{PicksDementia}(y)$$

$$\cdots$$

$$\mathcal{R} = \{\mathcal{Q}, \mathcal{Q}_1, \mathcal{Q}_2, \mathcal{Q}_3, \ldots\}$$

Benefits

✔ Semantically rich access to data

✔ High-level/intuitive way to form queries

✔Interoperability
easier to exchange information/knowledge with other platforms/institutions

Challenges

1. Investigate and adopt medical ontologies
   SNOMED, DiseaseOntology, GenesOntology, ...
2. Define mappings from data to ontologies.
   Use mapping languages such as R2RML
3. Extend AUEB tools to support:
   R2RML, query language of RAW system, ...
4. User query construction, answering, efficiency, ...

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS