WATERLOO
CHERITON SCHOOL OF
COMPUTER SCIENCE

An Ontological Approach to Data-Mining for Emergency Medicine

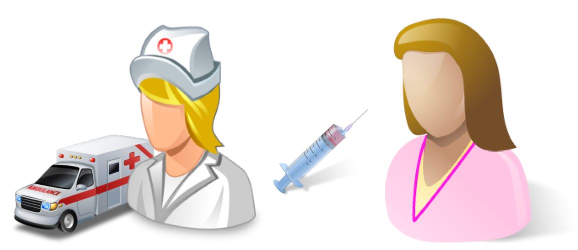
Atif Khan, John Doucette, Changjiu Jin, Lijie Fu, Robin Cohen

Abstract

Evidence-based **personalized decision making** AI system targeted for time and knowledge **constrained environments**.

Utilizes **semantic data representation** & **reasoning** to answer user queries. **Verifiable logical proofs** for system made decisions.

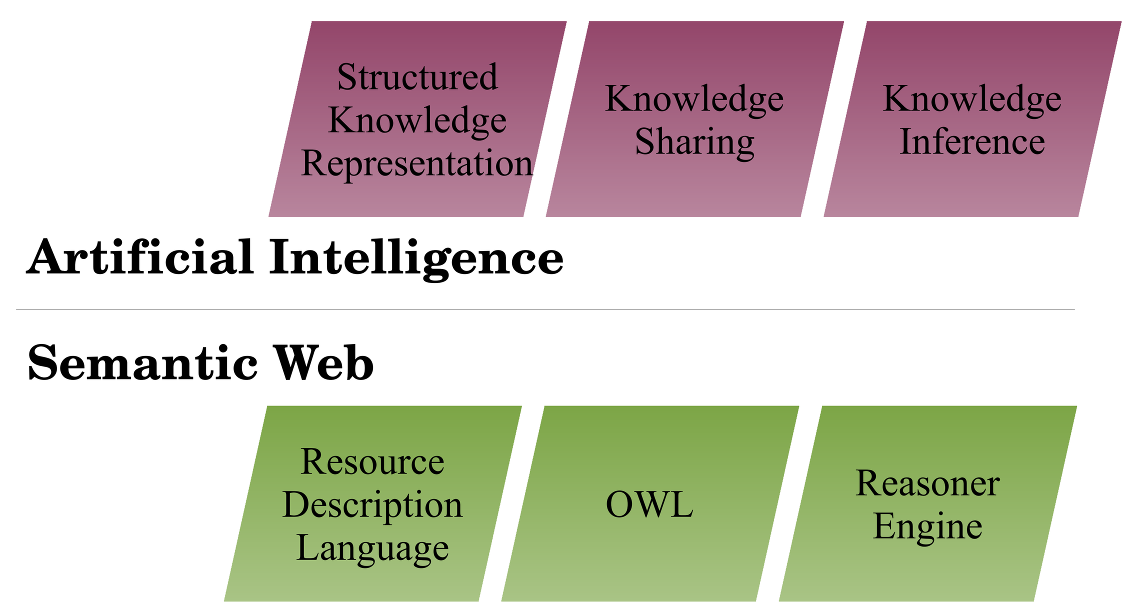
Motivational Example



query — Can I administer?

constraints — time, patient information

Building Blocks



Knowledge Sharing

Ontology

- "*an ontology is an explicit and formal specification of a conceptualization*"[1]
- Adds meaning to information

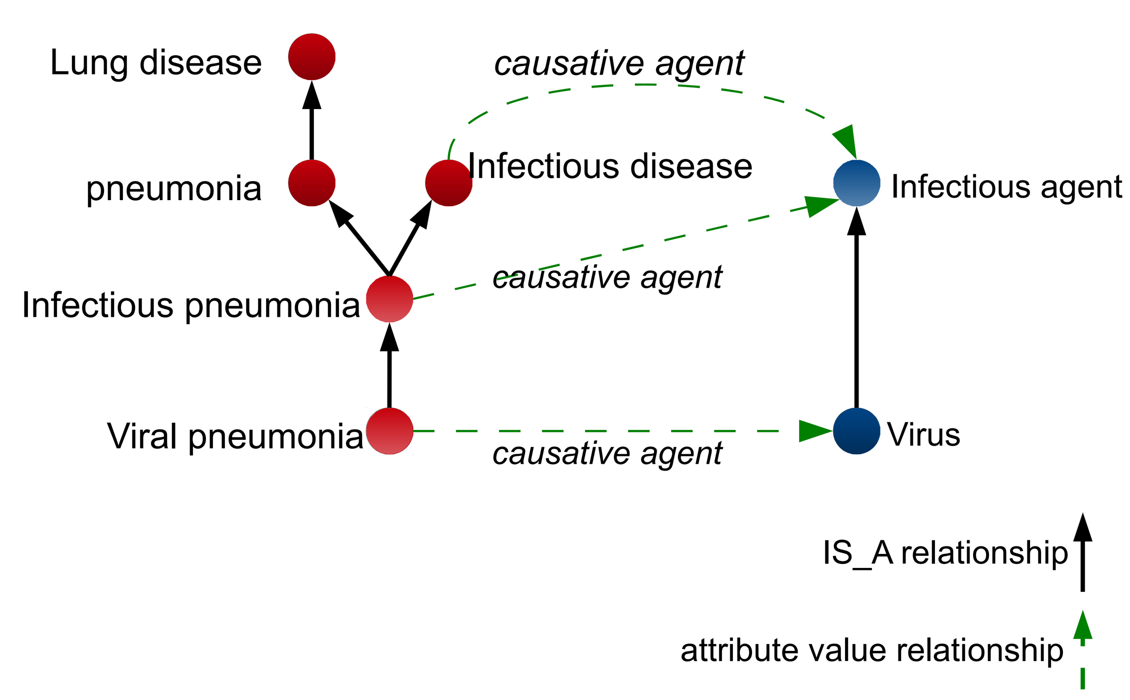
Storage

- Knowledge is stored in triple stores
- Triple store is
  - collection of facts/statements
  - represented in triples

[1] S. S. Sure, Y., R. Studer, Methodology for Development and Employment of Ontology based Knowledge Management Applications, in: SIGMOD Record 31 (4), 2002, pp. 18–23.

Knowledge Representation

SNOMED-CT Ontology Example

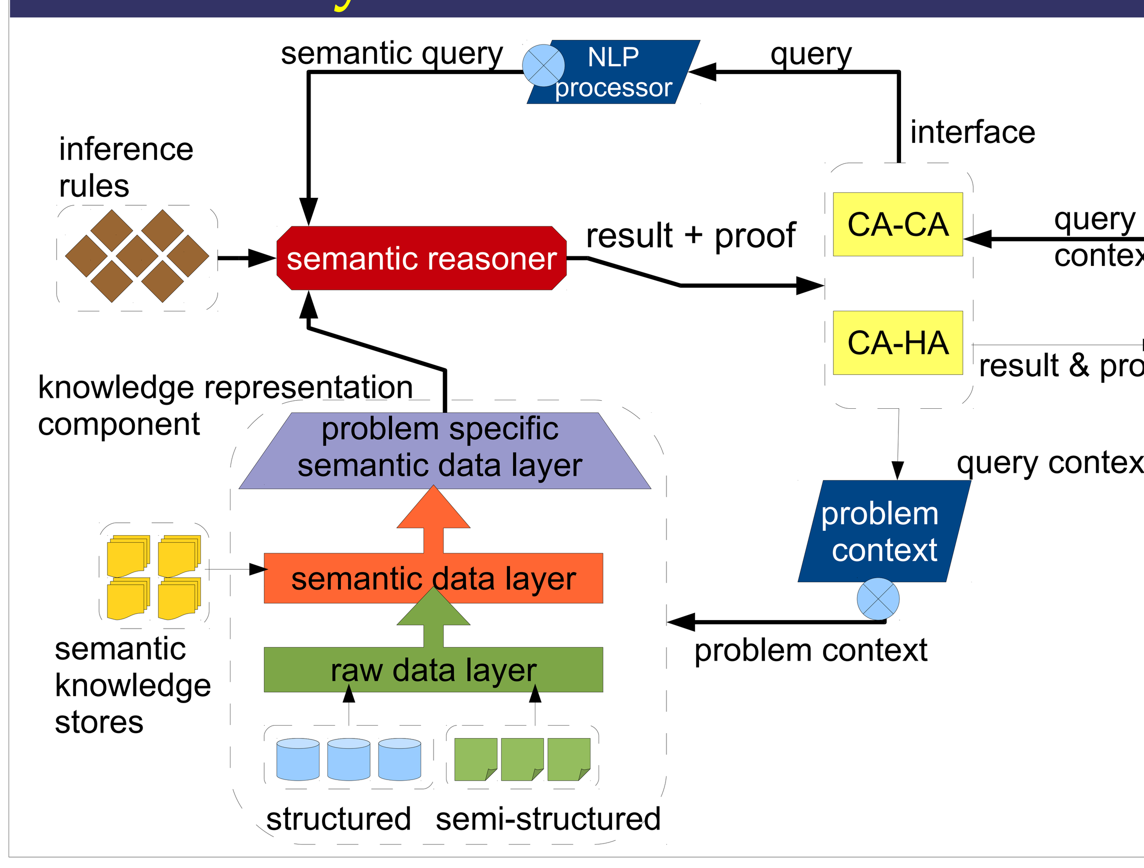


Proposed Framework

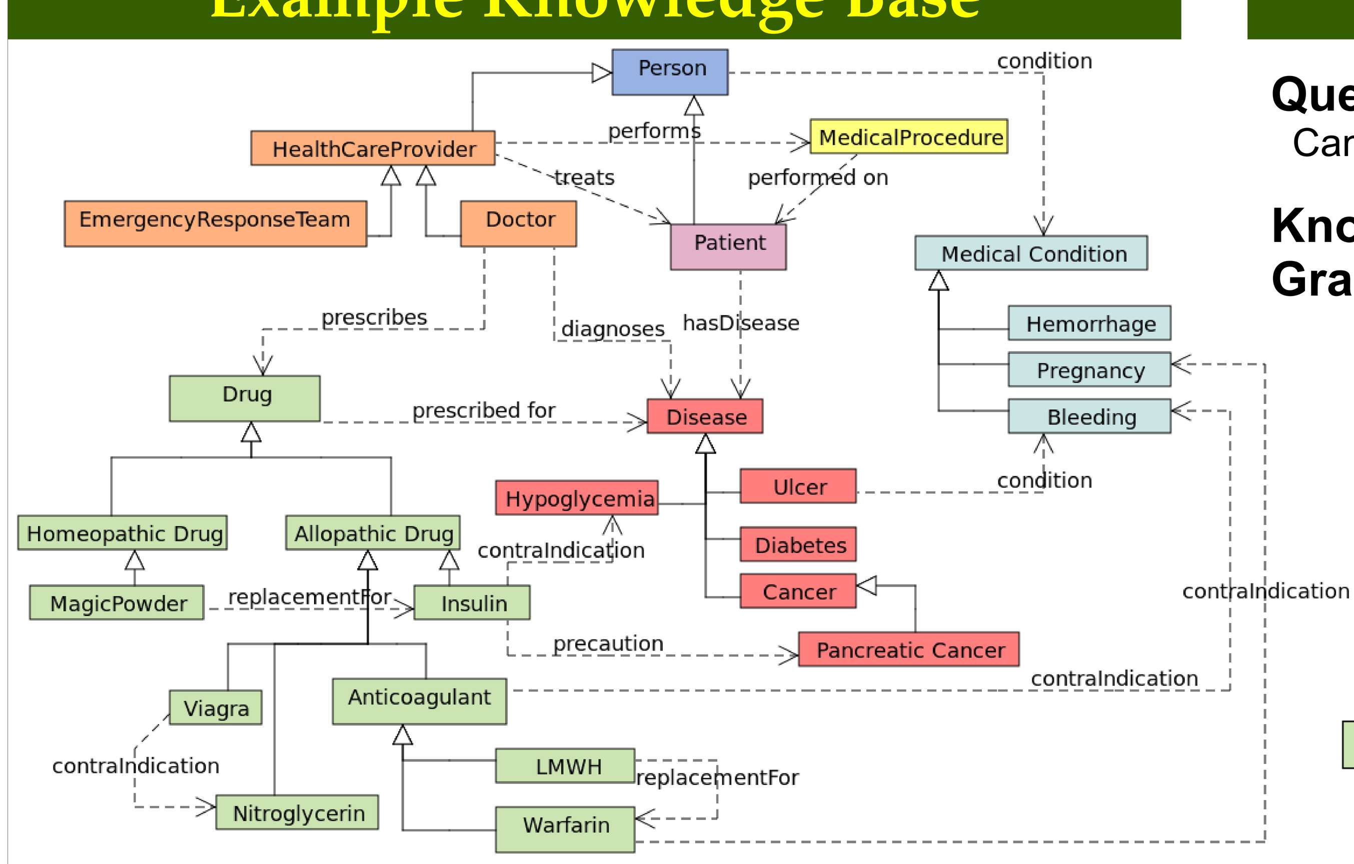
Semantic Data Mining

- Description
  - analysis of **large data-sets** and finding patterns and correlations through the **use of semantic technologies** for knowledge **representation** and **reasoning**.
- Factual rather than statistical
  - decisions are made based on evidence found in the data
- Accuracy depends on richness of the knowledge model
  - semantic concepts & their **relationships**

System Architecture



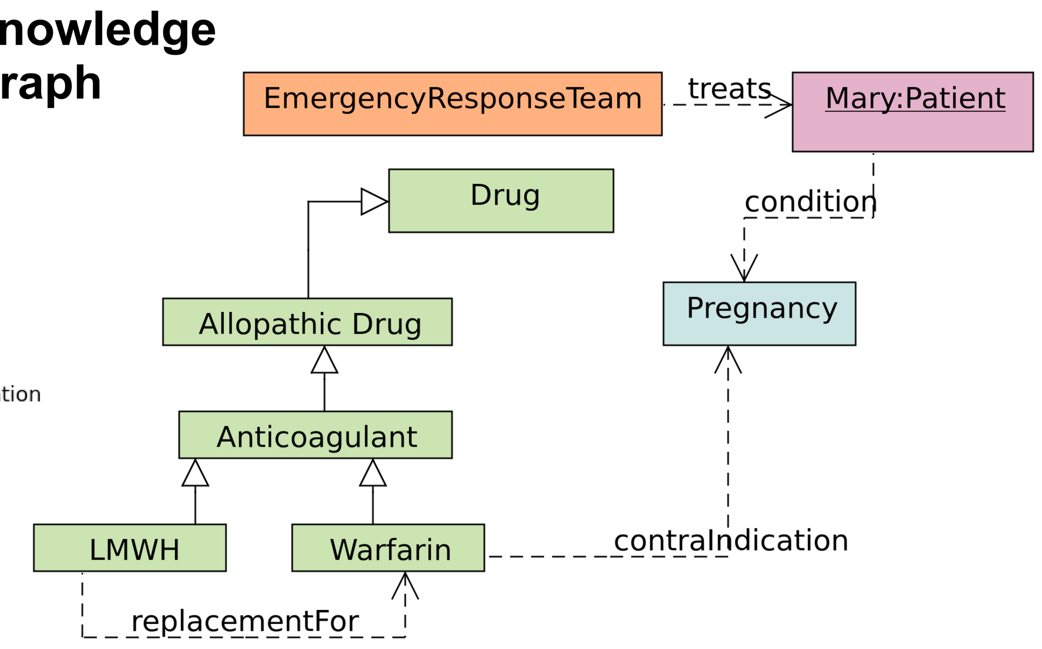
Example Knowledge Base



Example Scenario

Query:
Can we administer Warfarin drug to Mary?

Knowledge Graph



Inference Rule

If a drug is contraindicated in a condition, and any entity has that condition, then that entity cannot be given the drug.	{?ANY :condition ?CONDITION. ?DRUG :contraIndication ?CONDITION. } => {?ANY :canNotBeGiven ?DRUG}.

Semantic Proof

{{:Mary :condition :Pregnancy} e:evidence <facts.n3#_76>.
{:Warfarin :contraIndication :Pregnancy} e:evidence <facts.n3#_28>}
=>
{{:Mary :canNotBeGiven :Warfarin} e:evidence <rules.n3#_18>}.

Proof found in 18 steps (17821 steps/sec) using 1 engine (56 triples) }.

No Free Lunch Theorem

Description

- *"in the absence of prior knowledge about the properties of the function, all possible strategies for optimization must perform precisely the same on average"* [2]

[2] Wolpert, D.H., and Macready, W.G. No Free Lunch Theorems for Optimization, IEEE Transactions on Evolutionary Computation, 1997

No Free Lunch Theorem

Impact on Data Mining Approaches

- Non-semantic
  - most are subject to NFL theorem
    - as they require some sort of optimization
- Semantic
  - domain/problem specific knowledge is already included
  - use of ontology adds the domain knowledge
  - not subject to NFL theorem

Advantage of Semantic Approach

Optimization Independence

- Do not require new models/optimizations to answer new queries
  - knowledge is represented as factual statements
  - queries are answered based on explicit or implicit (inference) facts from the knowledge store

Transparency

- Explicit meaning
- Separation of concerns
  - inference rules can be defined independent of the data
- Logical proof of each decision made
  - each decision can be logically validated

Peer Reviewed Publication

Atif Khan, John Doucette, Changjiu Jin, Lijie Fu, Robin Cohen. *"An Ontological Approach to Data Mining for Emergency Medicine."* Northeast Decision Sciences Institute 2011 Annual Conference Montreal, Canada - April 14-16, 2011