Model-Based Systems Engineering → Semantics + Data Mining

Mark A. Austin

University of Maryland

austin@umd.edu

ENCE 688P, Fall Semester 2020

September 14, 2020

Overview

Part 3

Ontologies and Ontology-Enabled Computing

Definition of an Ontology

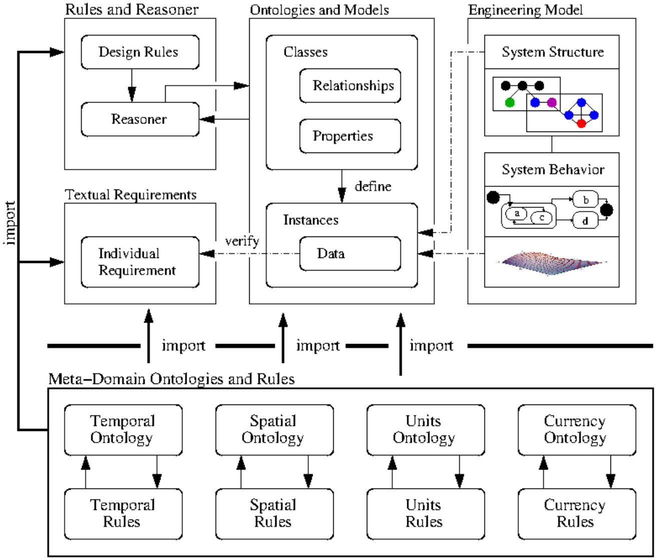
Definition (Ontology)

An ontology is a set of knowledge terms, including the vocabulary, the semantic interconnections, and some simple rules of inference and logic for some particular topic or domain.

Three Goals:

- Provide a semantic representation of each entity and its relationships to other entities;
- Provide constraints and rules that permit reasoning within the ontology;
- Describe behavior associated with stated or inferred facts.

Framework for Model-Based Design



The Data-Ontology-Rule Footing

Building Block for Semantic Modeling and Event-driven Execution of Multi-Domain Systems

MSSE/Ph.D. (Civil Systems) Students

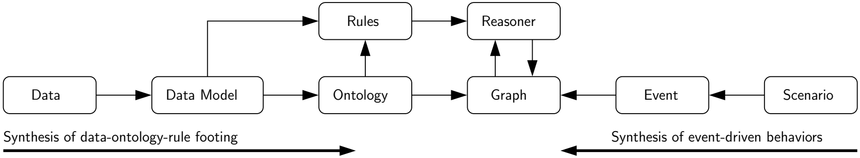
1. Parastoo Delgoshaei (2013-2017);
2. Maria Coelho (2015-present).

Data-Driven Approach

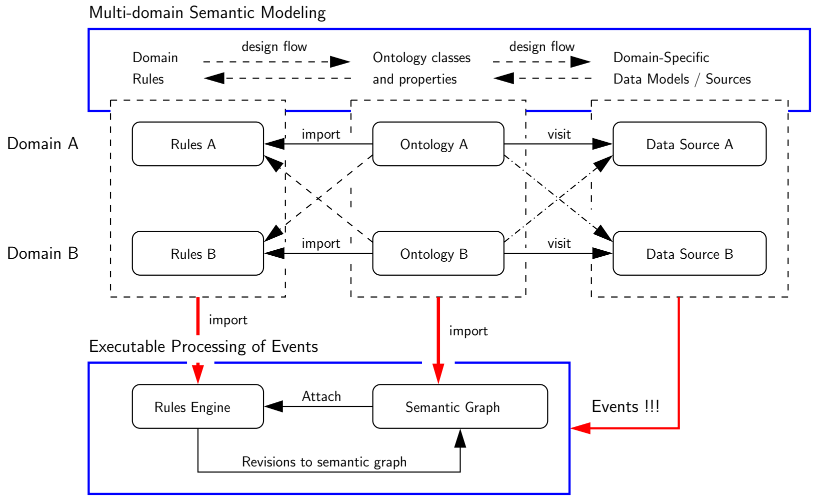
Guiding Principles:

1. One footing for ontologies, rules and data ...
2. Use (but do not extend) foundational level ontologies ...
3. Ontologies visit data models to get individuals ...
4. Semantic graph dynamically responds to incoming events ...
5. Enhance power of rules with backend functions ...

Preliminary Schematic:



Template for Semantic Modeling + Processing of Events



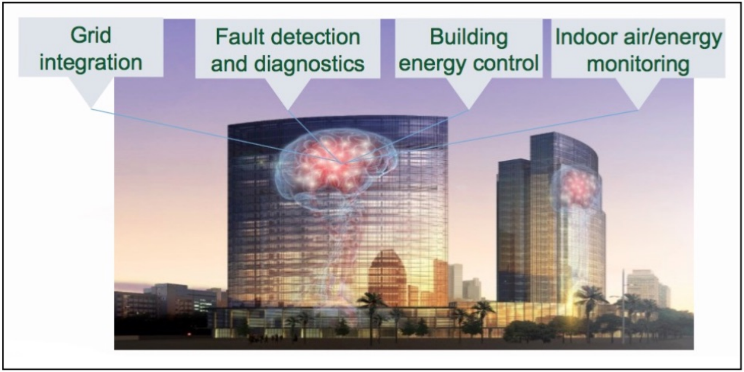
Case Study

Detection and Diagnostic Analysis of Faults in HVAC Equipment

Source: Delgoshaei and Austin, 2017.

Fault Detection in Buildings

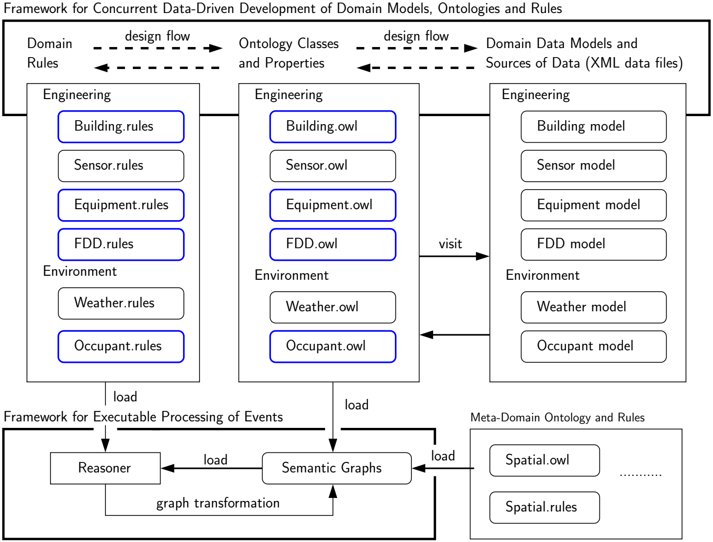
Example 1: Buildings that Think! (Work at NIST / UMD, 2017)



Research Question: How to use AI / Semantics to bring data, context and algorithms together for decision making?

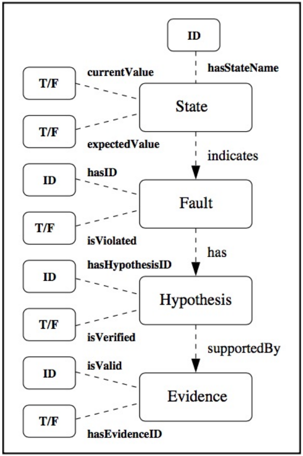
Legend: data = building geometry; context = occupant behavior; algorithms = reasoning.

Multi-Domain Building Semantics

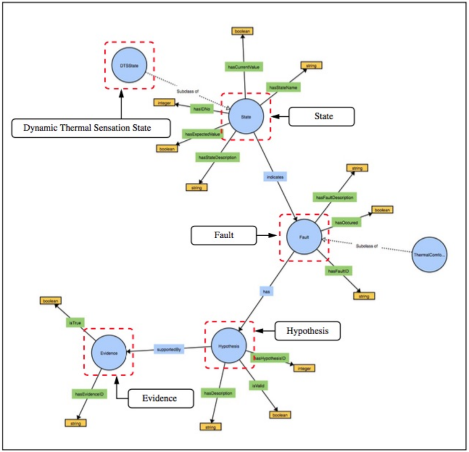


Multi-Domain Rule-based Reasoning

Flowchart for Processing of Faults

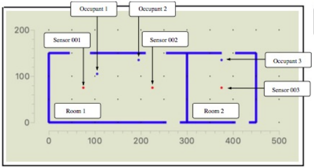


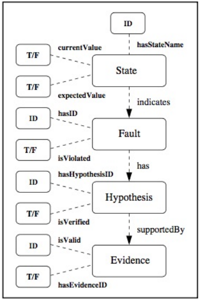
Fault Detection and Diagnostic Analysis Ontology



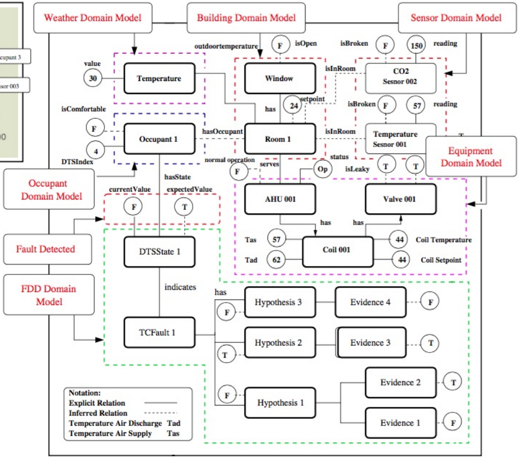
Multi-Domain Rule-based Reasoning

Case Study Problem



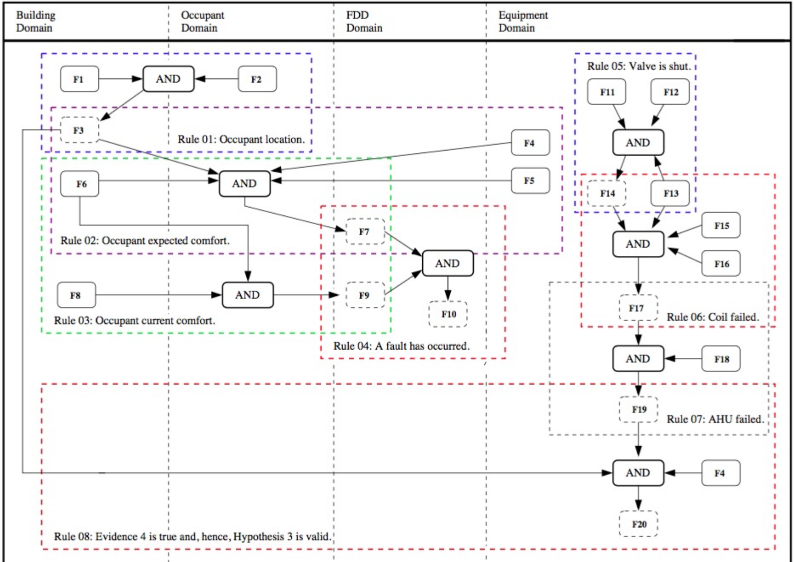


Snapshot of Fully Assembled Semantic Graph



Multi-Domain Rule-based Reasoning

Snapshot of Multi-Domain Evaluation and Forward Chaining of Rules

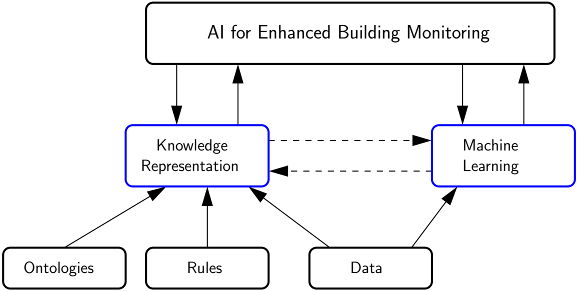


Multi-Domain Semantic Modeling + Data Mining

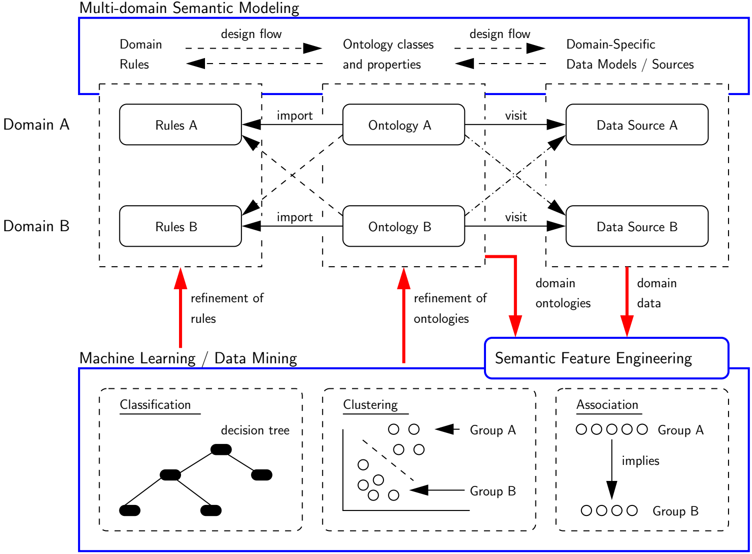
Multi-Domain Semantic Modeling + Data Mining

Initial Idea: Ditch semantic modeling → focus on machine learning instead.

Much Better Idea: Understand how can semantic modeling and data mining work together as a team?



Multi-Domain Semantic Modeling + Data Mining

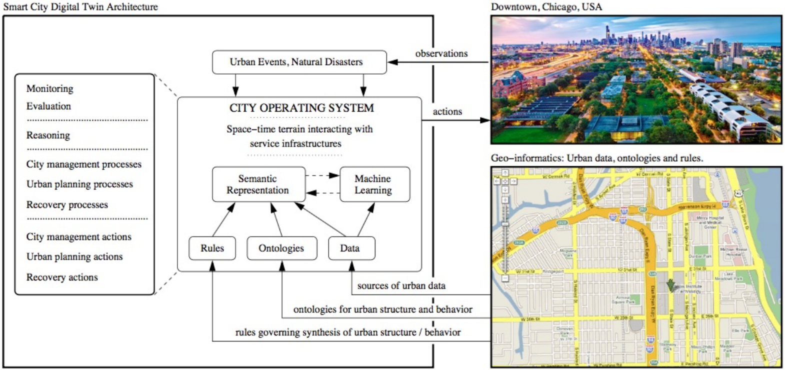


Case Study

Energy Consumption of 2,500 Buildings in Chicago

Energy Consumption of Buildings in Chicago

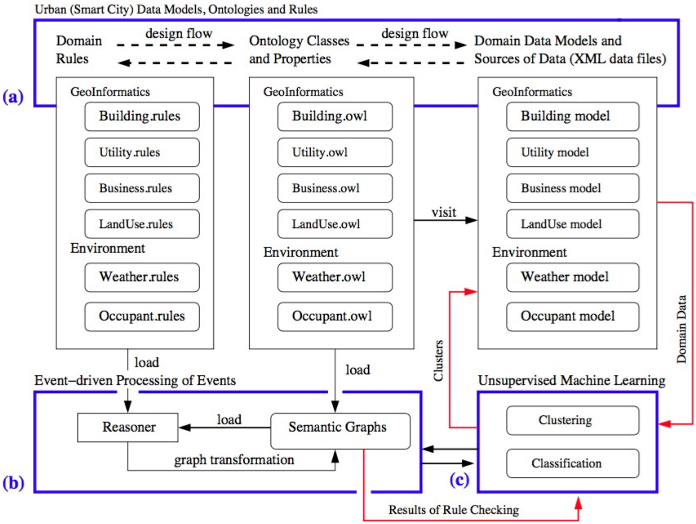
Example 2: Energy Consumption of 2,500 Buildings in Chicago (NIST / UMD / IIT) (2018)



Research Question: What factors – e.g., age, location, floor area, functionality – are strong indicators of energy consumption in buildings?

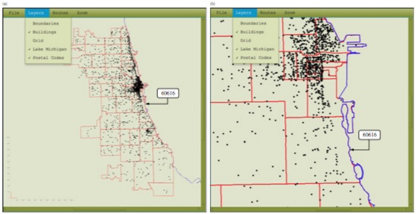
Energy Consumption of Buildings in Chicago

Framework for Integrated Semantics + Data Mining



Energy Consumption of Buildings in Chicago

Buildings in Chicago Metropolitan Area



Mining Data For Association Relationships ....

```
Rules and Associations for Zoning of Residential Buildings

Association 1: Site EUI (kBtu/sq ft)='(-inf-56.55]'
  ==> Building Type=Multifamily Housing   <conf:(0.84)>

Association 2: Community Area=NEAR NORTH SIDE
  ==> Building Type=Multifamily Housing <conf:(0.8)>

Association 3: Gross Floor Area - Buildings (sq ft)='[150354-513214]'
  ==> Building Type=Multifamily Housing <conf:(0.83)>

Association 4: Year Built='(1999.5-inf)'
  ==> Primary Property Type=Multifamily Housing <conf:(0.85)>

Rule 1:Building(?x), hasFloorAreaRatio(?x,?a), greaterThan(?a,6.6)
       isType(?x,?t), equal(?t,"multi-family") -> hasSubCat(?t,"RM6.5")

Rule 2:Building(?x) hasAge(?x,?a)
       greaterThan(?a,20) -> isType(?x,"multi-family")
```

Mining Data for Classification Hierarchies + Rules

```
Experiment A
eui <= 171.5    <-- first breakpoint
|   zip = 60616
|   |   age <= 102: MULTIFAMILY HOUSING (59.0)
|   |   age > 102
|   |   |   eui <= 118.2: MULTIFAMILY HOUSING (5.0)
|   |   |   eui > 118.2: OFFICE (3.0/1.0)
eui > 171.5
|   zip = 60616
|   |   eui <= 269.2: MULTIFAMILY HOUSING (11.0/1.0)
|   |   eui > 269.2: COLLEGE/UNIVERSITY (3.0/2.0)

Number of Leaves: 138
Size of the tree: 153
Correctly Classified Instances   1443 --> 80.7047%
Incorrectly Classified Instances 345 --> 19.2953%
```

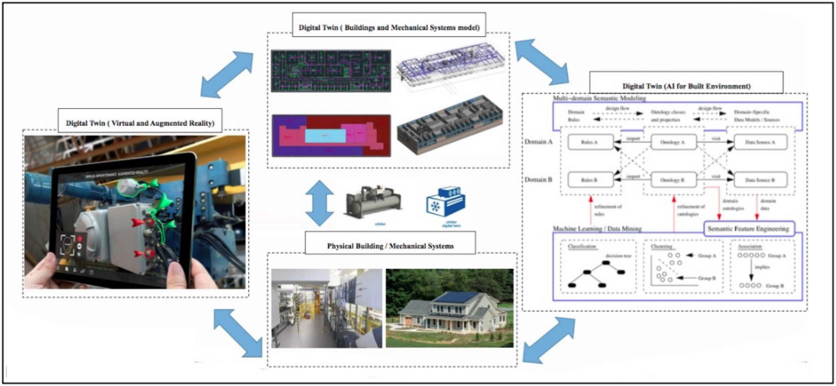
```
Experiment B
zip = 60616     <-- first breakpoint
|   age <= 86
|   |   age <= 53
|   |   |   area <= 115066: NEAR SOUTH SIDE (13.0/6.0)
|   |   |   area > 115066
|   |   |   |   age <= 12: NEAR SOUTH SIDE (5.0)
|   |   |   |   age > 12
|   |   |   |   |   eui <= 130.2: DOUGLAS (3.0)
|   |   |   |   |   eui > 130.2: NEAR SOUTH SIDE (8.0/2.0)
|   |   age > 53: DOUGLAS (18.0/2.0)
|   age > 86: NEAR SOUTH SIDE (7.0/2.0)

Number of Leaves:  82
Size of the tree: 102
Correctly Classified Instances   1399  --> 78.2438%
Incorrectly Classified Instances 389   --> 21.7562%
```

Software. WEKA (Waikato Environment for Knowledge Analysis).

Vision for Future Capability

Future Vision: Digital Twins + Virtual and Augmented Reality

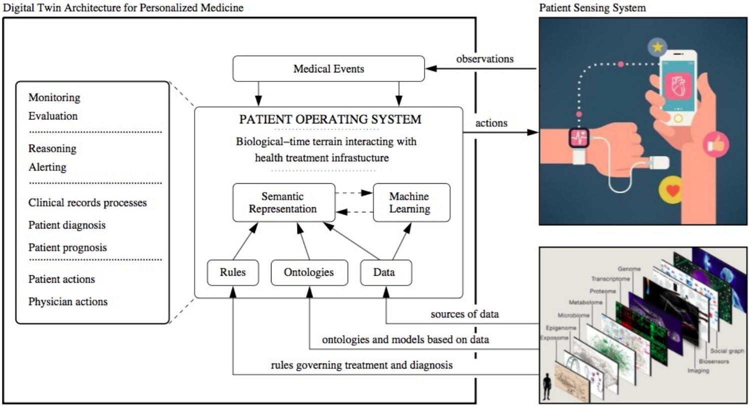


Case Study

Semantics + Data Mining for Precision Medicine

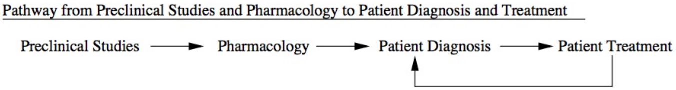
Semantics + Data Mining for Precision Medicine

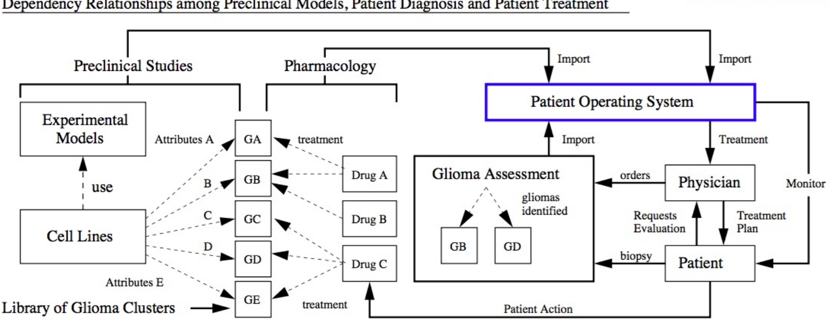
Example 3: Semantic Foundations for Precision Medicine (NCI / UMD) (2017-2019)

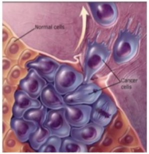


Long-Term Objective: Digital Twin Architecture for Improved Management of Symptoms and Treatment of Brain Cancer Patients.

Semantics + Data Mining for Precision Medicine







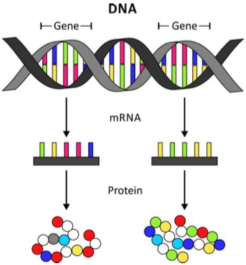
Semantics + Data Mining for Precision Medicine

Problem Complexity

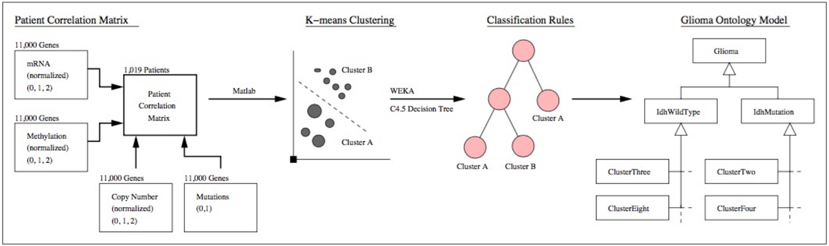
Human Genome: 19,000-20,000 individual genes

Patient data extracted from Cancer Genome Atlas

- 1,019 Patients
- Each patient described by 44,000 units of data assembled from 11,000 gene attributes from 4 sequencing method.



Data-to-Rules Flowchart



References

- Delgoshaei P. and Austin M.A., Framework for Knowledge-Based Fault Detection and Diagnostics in Multi-Domain Systems: Application to Heating Ventilation and Air Conditioning Systems, International Journal on Advances in Intelligent Systems, Vol. 10, No 3 and 4, December 2017, pp. 393-409.
- Austin M.A., Delgoshaei P., Coelho M. and Heidarinejad M. , Architecting Smart City Digital Twins: Combined Semantic Model and Machine Learning Approach, Journal of Management in Engineering, ASCE, Volume 36, Issue 4, July, 2020.
- Abraham J., Semantic Foundations for Formalizing Brain Cancer Profiles, MS Thesis in Systems Engineering, University of Maryland, April 2019.
- Wagner et al., An Ontology for State Analysis: Formalizing the Mapping to SysML, Proceedings of 2012 IEEE Aerospace Conference, Big Sky, Montana, March, 2012.