Research on Data Semantics

Ken Baclawski
College of Computer and Information Science
Northeastern University
What is data semantics?
What is data semantics?

To answer this we will look at software semantics in general.
What is data semantics?
What is data semantics?

Program

Data Structure

separate concerns

Software
What is data semantics?

Program

Data Structure
What is data semantics?
What is data semantics?
What is data semantics?

Program

abstraction

Algorithm
What is data semantics?

Program

abstraction

Algorithm

analyze
What is data semantics?

Program

abstraction

Algorithm

analyze

Complexity
Termination
Correctness
What is data semantics?

Program

abstraction

Algorithm

analyze

Complexity
Termination
Correctness

Data Structure
What is data semantics?

Program
- abstraction
- Algorithm
  - analyze
  - Complexity
  - Termination
  - Correctness

Data Structure
- abstraction
What is data semantics?

- Program
  - abstraction
  - Algorithm
    - analyze
      - Complexity
      - Termination
      - Correctness
  - Data Structure
    - abstraction
    - Ontology
What is data semantics?

Program

abstraction

Algorithm

analyze

Complexity
Termination
Correctness

Data Structure

abstraction

Ontology

analyze
What is data semantics?

Program

abstraction

Algorithm

analyze

Complexity
Termination
Correctness

Data Structure

abstraction

Ontology

analyze

Complexity
Consistency
What is data semantics?

Program
  abstraction
    Algorithm
      analyze
        Complexity
        Termination
        Correctness

Data Structure
  abstraction
    Ontology
      analyze
        Complexity
        Consistency

Inference
What is data semantics?

- Program
  - abstraction
  - Algorithm
    - analyze
    - Complexity
    - Termination
    - Correctness

- Data Structure
  - abstraction
  - Ontology
    - analyze
    - Complexity
    - Consistency

- Inference
- Integration
What is data semantics?

- Program
  - abstraction
  - Algorithm
    - analyze
    - Complexity
      - Termination
      - Correctness
  - Data Structure
    - abstraction
    - Ontology
      - analyze
      - Complexity
        - Consistency
  - Inference
  - Integration
  - Interoperability
What is data semantics?

- Program
  - abstraction
  - Algorithm
    - analyze
    - Complexity
    - Termination
    - Correctness
- Data Structure
  - abstraction
  - Ontology
    - analyze
    - Complexity
    - Consistency
- Inference
- Integration
- Interoperability
- Question Answering
What is data semantics?

- Program
  - abstraction
  - Algorithm
    - analyze
    - Complexity
    - Termination
    - Correctness

- Data Structure
  - abstraction
  - Ontology
    - analyze
    - Complexity
    - Consistency

- Inference
  - Integration
  - Interoperability
  - Question Answering
  - Data Mining
Application Domains

- Health Sciences
- Collaboration tools
- Social networking
- Web services
- Standards: OASIS, W3C, OMG, ISO, ...
- eGovernment
- Knowledge capture

- Software engineering
  - Versioning
  - Modularization
  - Formalization
  - Repositories
  - Higher-order logic

- Wireless communication
- Situation awareness
Open Data Standard
Background

- Vast amount of available data
  - Databases (usually, but not always, relational)
  - Scientific data streams (such as Big Data)
  - Internet (such as websites)
- Large number of formats
  - Often unique to a database, stream or site
- Complex formats
- Underspecified semantics
  - Implicitly specified by programs, or
  - Databases specified with SQL
- Difficult to access and to integrate
  - Often accessible only through applications such as browsers
Open Data Protocol (OData)

- Web protocol for data access
  - Support for CRUD operations
- Built on existing standards
  - HTTP protocol
  - RESTful web services and principles
  - ATOM/XML and JSON formats
- Open standard
  - OASIS Technical Committee
  - Open license
  - Open source libraries
History of OData

• Microsoft Role
  • Original developer of OData
  • Development process was open (Microsoft Open Specification Promise).
  • Donated to OASIS
• Founding organizations of OASIS TC include:
  • SAP AG, IBM, Citrix
  • Progress Software, WSO2
  • Northeastern University
• OData is not a Microsoft product!
Momentum

- OData is already in use
  - Major websites: Netflix, eBay
  - Open government initiative
  - Commonly used for accessing internal data sources within large companies
  - See odata.org for more examples.
- Not just a proposed standard
  - Already up to the third version
  - Based on extensive commercial experience
  - Both proprietary and open source software