MANCHESTER 1824

The University of Manchester

QuickTime™ and a decompressor are needed to see this picture.

Problems arising in applications: Background Knowledge Representation, Data modelling & Ontologies

Alan Rector

BioHealth Informatics Group
School of Computer Science &
Northwest Institute of BioHealth Informatics
University of Manchester, Manchester M13 9PL

rector@cs.manchester.ac.uk

QuickTime™ and a decompressor are needed to see t

NIBHI

Our applications

- Ontology driven (clinical) software
  - Ultra-adaptable & very late bindings
    - Ontologies for indexing context carrying "payload"
      - ***Reduce the effort to modify, localise and adapt software to niches***
        - *Medical institutions are a mass of niches - conglomerates of cottage industries*
- Debugging adapting large clinical terminologies - SNOMED-CT
  - How to debug and make useful a DL model with 450,000 classes, ≥ 1M terms, ≥1.5M axioms
    - Built by many people with many idiosyncrasies
  - Can it be used as part of development of new WHO ICD Revision 11?
- Developing standards
  - Ontology for Clinical Trials
    - Must fit into a UML / MDA based framework

Ontology & Knowledge Representation

- Is ontology all of knowledge representation?
  - If not, where are the boundaries? The Interfaces?
- What is / should be the relation of different technologies/Languages?
  - RDF(S) / SPARQL / Linked open data?
  - SKOS? / Thesauri
  - OWL? OWL / DL Reasoning? / “Meaning” / “Ontologies”
- How should ontologies related to data modelling, UML & MDA?
  - Is there an added value for “Ontology Drive Architectures”?
  - Can UML/MDA tools be adapted / linked to ontologies? to OWL?
  - How to fit into the broader standards community
- How should languages link to ontologies? KR more generally
  - Especially in a multi-lingual environment?

OWL/DLs and Our Representations

- We require DL/OWL for composition and indexing
  - All our domains are too big to enumerate explicitly
  - Almost all involve context
    - Cancer of X organ associated with Y Gene
    - Initial management protocol for male over 14 in (our) emergency department with acute head injury and alcohol/drug intoxication
    - Trial of X intervention compared to Y intervention in Z disease using W methodology in patients over 50 controlled for sex, & severity, excluding prior treatment with any form of \V or U.
    - Diagnostic probabilities for acute abdominal pain in a woman of child bearing age? An infant? An elderly male?
- We need ontological distinctions to avoid confusions
  - Things and their roles
    - Person, doctor and patient
  - Things in the world and information about them
    - The event, the data, the copy of the data on the database…

But OWL is really a logic language rather than an ontology language

- Not everything in OWL is an ontology
  - e.g. a formal representation of a UML diagram
- Not all ontologies can be expressed in OWL
  - Many require second order constructs, cyclical constructs, ...
- Slowly defining boundaries and interfaces

And most knowledge is not ontological

- "Appendicitis may cause acute abdominal pain"
- We can represent this if we transform each property to a class - "reify" each association
  - but our tools to view the results are the awful
  - and should we then call it an "ontology"?
- Example
  - Association of acute abdominal pain in a woman of child bearing age? An infant? An elderly male?
    - Association that has_topic some Woman_of_child_bearing_age & has_target some Appendicitis.
    - Association that …

Treating associations as classes links naturally to other formalisms

- Really just existential statements analogous to Conceptual Graphs
- Links naturally into UML
  - Every UML association identified implicitly with a class
  - UML validators using DLs work this way
    - *Berardi, D., Calvanese, D., and De, G., Giuseppe 2005. Reasoning on UML Class Diagrams. Artificial Intelligence. 168, 70-118.*
  - … but no tooling
- But issue of closed world instance instance validation remains
  - Missing values have no meaning in OWL
    - it just infers them to be present
  - See work on “constraints” by Motik et al.
    - *Motik, B., Horrocks, I., and Sattler, U. 2007. Adding integrity constraints to OWL. Third OWL Experiences and Directions Workshop (OWLEd-2007).*

...and still we have problems with other knowledge that arises naturally

- Defaults & exceptions
- Strengths of association / uncertainty
  - Really second-order knowledge - about the class of associations not the individual association
- Higher order statements - “same kind as”
- Cyclical statements - “same”, “different”,
- “All…All” statements
  - All licensed drivers are authorised to drive all cars
    - the “Cat lovers problem”
      - *A horrible work-around exists, but it only really works for instances*
- How best to get additional layers of reasoning

and with Language - especially for medical terminologies

- Most applications need to present expressions in pseudo-natural language
  - Experts won't read even outlines -
    - although they will read spreadsheets
- More than just rdf:label or even the family of skos:xLabels
  - To get useful language need gramatical information e.g.
    - Plural and singular forms
      - *Ontology and natural language conventions differ in use of singular and plural*
    - Information on how to handle modifiers
      - *Ball that has_colour some Red --> "Red ball"*
      - *Ball that belongs_to value Alan --> "Alan's Ball"*
      - *Ball that has_status some In_play --> "Ball that is in play"*
    - In other languages, much other information on inflexions, prepositions, alternatives, etc.

Note on Defaults & Exceptions

- Early frame systems had simple inheritance with over-riding
  - Worst case intractable, but…
    - Constructing the set of most specific values in an ontology is cheap
    - In a well normalised ontology it is either
      - *A singleton*
      - *A set for which a "conflict resolution rule" - e.g. priority - can easily be used*
      - *A set which is a natural union - e.g. all eligible candidates by any criterion*
  - Particularly useful for managing user interfaces, strength's of association, etc. that simply don't follow logic
    - But must keep such payloads that don't conform to DL reasoning away from any DL reasoner.

Summary: Some of what I need integrated in my toolkit

- Composition, Definitions and classification
  - All the fields are too big to enumerate explicitly
- Lexicons and thesauri and language generation
- Easy ways to construct & view existential / contingent statements
- Links to other reasoning and rules
  - Simple defaults and exceptions
- Cleaner between OWL, RDF, SKOS, Linked Open Data
  - With definitions of boundaries and interfaces
- Bidirectional links between OWL& UML
  - Strong links to closed world rules and instance validation

… and is it helpful to call it all "ontology"?