https://github.com/Vocamp/DetectorFinalState

The Detector Final State pattern: Using the Web Ontology Language to describe a Physics Analysis

David Carral
Wright State University

Pascal Hitzler
Wright State University

Charles Vardeman
University of Notre Dame

Gordon Watts
University of Washington/Seattle

Abstract

The Data and Software Preservation for Open Science (DASPOS) collaboration has developed an ontology for describing particle physics analyses. The ontology, a series of data triples, is designed to describe dataset, selection cuts, and measured quantities for an analysis. The ontology specification, written in the Web Ontology Language (OWL), is designed to be interpreted by many pre-existing tools, including search engines, and to apply to both theory and experiment published papers. This paper gives an introduction to OWL and this branch of library science from a particle physicist's point of view, specifics of the Detector Final State Pattern, and how it is designed to be used in the field of particle physics primarily to archive and recall analyses. A general introduction to DASPOS and how its other work fits in with this topic will also be described.

What Would It Take to Build a Search Engine With Direct Knowledge of the Paper Content

Representing an analysis is pretty easy on a pen and paper. Even translating it into a few data structures is pretty easy, and creating a search engine to run over the data isn't that hard.

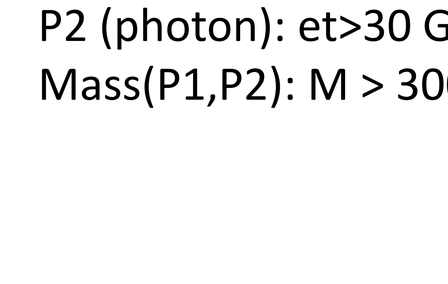
But this problem has been solved already!

This is called ***knowledge representation***, and is an active field of research in the library and artificial intelligence communities.

Of course, they are tackling a much bigger problem: how to codify all the knowledge in the world.

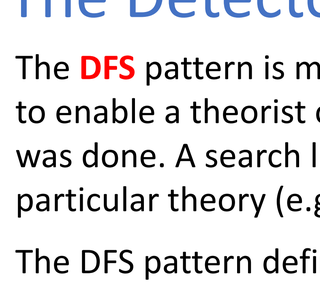
This is done in two steps. First you define an ontology that can describe the knowledge you wish to encode. The ontology is similar to a class hierarchy. Second, you populate it with the knowledge – the facts and relationship between facts.

Because this is all done with standards it is possible to use a common set of tools to *store, serve, and "reason"* about the knowledge. References like "Semantic Web" are common in this space



Where is that paper that searched for a bump in the 2 photon mass spectra?

1 Search the Archive (arxiv.org)?



Search for the word "diphoton" in the archive?
Search for the words "two photons" in the archive?
Search for the word "photon" and window?
What if the title doesn't contain the word photon?

2 A Search Engine with semantic knowledge

What would it take to make this a valid search engine query:

P1 (photon): et>30 GeV;
P2 (photon): et>30 GeV;
Mass(P1,P2): M > 300 GeV

Paper Knowledge
Paper Knowledge
Paper Knowledge
Paper Knowledge

Search Engine

The Detector Final State OWL Pattern

The **DFS** pattern is meant to describe only the base sample selection criteria. Its goal is to enable a theorist or another experimentalist to look for samples on which a search was done. A search like this will quickly window down the analyses, regardless of a particular theory (e.g. SUSY vs a two higgs doublet model).

The DFS pattern defines **physics objects** and **simple relations** or cuts based on their properties

Physics Objects

A jet or an electron, an event level quantity like missing energy

- Objects can have properties (E_T, p_T, p_x, etc.)
- Object types are declared by a BaseDefinition
  - *E.g. anti-kt-04:* which points to a jet built out of the anti k_T algorithm with $R = 0.4$
- An object can be a function of other items in the event
  - As an example: NTrack(J1, 0.2, pt>1.0 GeV) would count the number of tracks of at least 1 GeV within $\Delta R < 0.2$ around J1.

Relations

A relation is a series of selection criteria which are "and" or "or"'d together.

- Each binary relation ($>, <, \geq, \leq, ==, !=$, etc.) takes two arguments
- At least one has to be a *PhysicalQuantity* (but both can be if desired).
- One can be a number with units. We use a well established ontology to represent a number with units, called *quantities, units, dimensions and data types.*

How Do You Organize What You Know?

- Facts
- Relationships
- Domains
- Inter-Domain relationships

How can we encode this for a computer to use?

Ontology!

Web Ontology Languages

"OWL" - Encoding the knowledge of the web

- Information or facts or assertions
- How the various facts in one domain relate to each other
- Rules, or axioms, that limit the relationships appropriately
- Can embed one ontology in another

This is a Global Effort

Knowledge Representation

Lots of researches are working on this

http://ontologydesignpatterns.org/wiki/Main_Page

The main repository page for patterns. Authors submit and they are available and stored in a machine readable format.

http://www.openphacts.org/index.php

Pharmacy site with lots of data on drugs. The back end's knowledge is built on ontologies described by OWL.

https://pubchem.ncbi.nlm.nih.gov/rdf/

Semantic Web technologies are emerging as an increasingly important approach to distribute and integrate scientific data.

The Web Ontology Language (OWL)

A way to organize knowledge

The Web Ontology language is designed to represent *structure of information on the internet.*

OWL → Formally Defines → Ontology → Knowledge

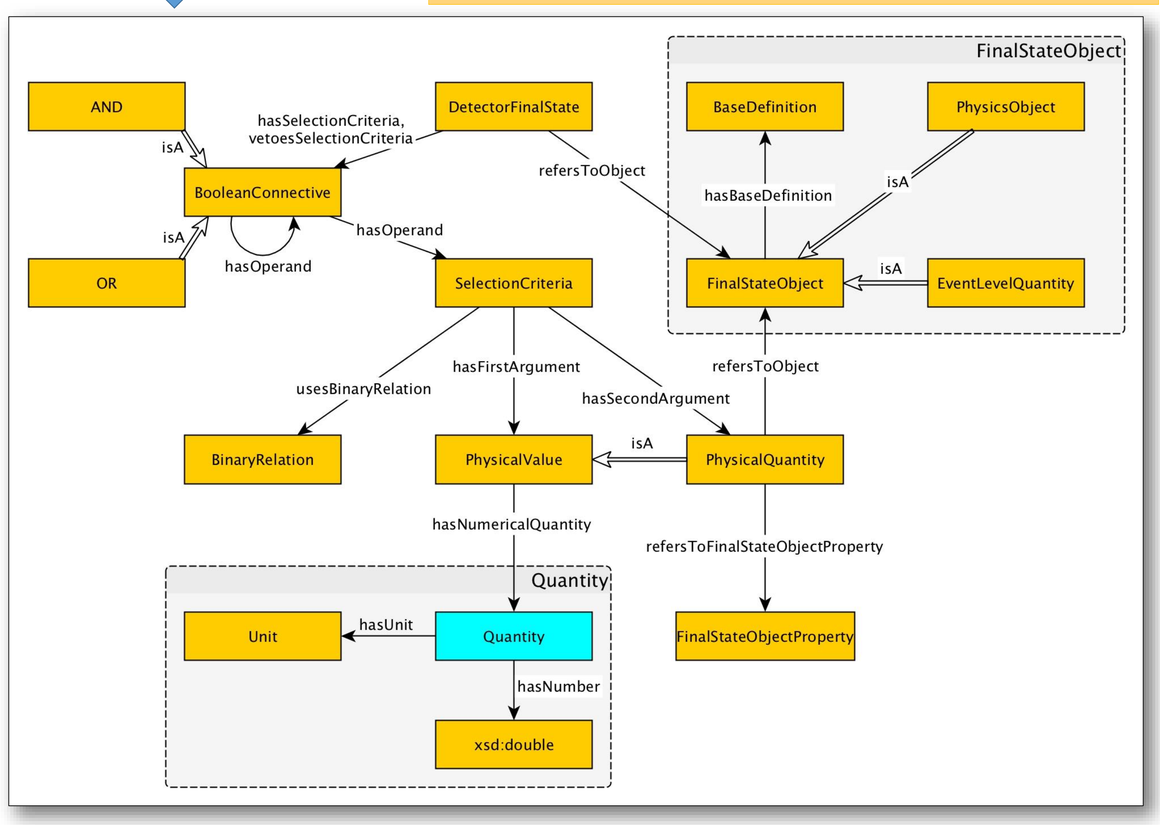
The knowledge is expressed in various forms text forms (including XML). This is syntax is called RDF/Turtle – which is an XML variant. It is **verbose**.

```
atlas:J1 rdf:type dfs:PhysicsObject ;
  dfs:hasBaseDefinition "anti-kt-04" .

atlas:physicalQuantity0 rdf:type dfs:PhysicalQuantity ;
  dfs:refersToObject atlas:J1 ;
  dfs:refersToFinalStateObjectProperty dfs:et .

atlas:number1 rdf:type qudt:QuantityValue ;
  qudt:numericValue "60"^^xsd:decimal ;
  qudt:unit unit:GigaElectronVolt .

atlas:selectionCriteria2 rdf:type dfs:SelectionCriteria ;
  dfs:usesBinaryRelation dfs:greaterThan ;
  dfs:hasFirstArgument atlas:physicalQuantity0 ;
  dfs:hasSecondArgument atlas:number1 .
```

$J_1 > 60$ GeV

J_1 is built using the Anti-k_T clustering algorithm with an 0.4 radius

The definition of the ontology



Current Status and What is Next?

- The ontology has been shown to be flexible enough to handle simple analysis specifications required to catalog an analysis by the final state.
- A simple physicist domain specific language has been created to easily author a DFS.
- Next, we will accumulate twenty such analysis specifications.
- Configure them to work with SPARQL, and test searching against this small sample database.
- This ontology is meant to be embedded in a much larger one that connect simulation and data samples, commands to reproduce the analysis. Work will continue along this front as well.
- Work with CERN to help with their internal representation of their analysis stores and the OpenData project.