

Domain Vocabulary Development: Improved Semantics BoF (2)

Breakout 5 – Thursday 5 April 11:30 - 1:00.

research data sharing without barriers

Prepared for RDA 9th Plenary Barcelona, 2017 Gary Berg-Cross (Ontolog)

Domain Vocabulary BoF Session Agenda

- Facilitating Domain Vocabulary Development & Harmonization
 - Recap from P8 BoF Gary Berg-Cross
 - Follow up since P8 ontology engineering document, virtual meetings, vocamp, experimental work, RDA fellow...
- Ontology engineering & ISO efforts to create standards for City information: Mark Fox
- Using domain ontologies in CEDAR to standardize the development of metadata in order to describe experimental datasets: Mark Mussen
- Other....?
- Community discussion regarding the state of:
 - Vocabulary development, methods & formalization, harmonization, infrastructure
 - Common interests such as vocabulary development & services,
 - issues and best practice solutions in the domain vocabulary space
 - Follow up ideas, IG?, session at P10?

Recap from P8 BoF - Gary Berg-Cross

Info at: https://www.rd-alliance.org/domain-vocabularies-rda-8th-plenary-bof-meeting The Problem(s) are we trying to help with?

Help systematize the large body of domain definition work on terms & their meaning Reality is very heterogeneous & there is lots of it We lack of foundational grounding

<u>Speakers from Several Specific Domains Discuss their Status, Goals and Issues</u> •Chem & Materials Research,

• We could use a better understanding of the relations of vocabularies, taxonomies & ontologies. Ontologies inform a vocabulary.

•Earth Science domain,

•Mark Fox on quality of urban life interest group & Smart Cities – role of foundational ontologies, patterns & competency questions – see Ontological Engineering & Development 101 briefing.

 See https://www.rd-alliance.org/sites/default/files/attachment/Ontology %20Engineering.pdf

Strategy Leverage existing discussion and work such as better foundations and tools to improve the quality of definitions and access

Follow up since P8: Ontolog sessions on Domain Vocabularies

Topic and Session Overview Gary Berg-Cross (Ontolog Board Members)

http://ontologforum.org/index.php/DomainVocabularies

Presenters:

- Mark Fox (University of Toronto) An Upper Level Ontology for Global City Indicators
- Torsten Hahmann (University of Maine): Domain Reference Ontologies vs. Domain Ontologies: What's the Difference? Lessons from the Water Domain
- Boyan Brodaric (Research Scientist at Natural Resources Canada): What's a river? A foundational approach to a domain reference ontology for water

2nd Session Nov. 17, 2016

Speakers

Simon Scheider (Human Geography and Spatial Planning, Universiteit Utrecht) Challenges in ontological prerequisites for meaningful spatio-temporal analysis (maps, statistics).

Olivier Bodenrieder (NIH) Vocabulary experience with SNOMED CT.

Mike Bennett (FIBO) Semantics of FIBO Business Vocabulary

Cory Casanave (Model Driven) Cross domain sharing and federation of threat and risk information, an application of Semantic Modeling for Information Federation (SMIF a.k.a. SIMF)

From Ontolog Forum 2016, Domain Vocabulary Semantics

Introduction to DC-Vocamp 2016 UMD's Center for Geospatial Information Science Gary Berg-Cross

http://vocamp.org/wiki/GeoVoCampDC2016

RDF vocabulary for Chemical Safety & Chemical Terminology (Leah McEwen) A pattern to support Materials Research vocabularies (Kimberly Tryka & Alden Dima) Topography - basic terrain primitives, slope, length, shape, curvature (Usery & Sinha)



Data from Reported Incident Strings

Arsine, phosphine, and tetraborane are all oxidized explosively by fuming nitric acid.

- Phosphine, hydrogen sulfide, and selenide all ignite when fuming nitric acid is dripped into the gas.
- ^{3.} Hydrogen telluride ignites with cold concentrated nitric acid, sometimes exploding.

Substances Outcomes Consequences Conditions **Operations** Apparatus/ equipment

https://pubchem.ncbi.nlm.nih.gov/compound/nitric_acid





Text mining PubChem Software Tools

- NLTK (Python)
- Chemicaltagger (Java)
- Jupyter Notebooks

– <NounPhrase> <.I.J>Incompatible</JJ> <NN-CHEMENTITY>materials</NN-CHEMENTITY> </NounPhrase> <COLON>:</COLON> – <NounPhrase> <NNPS>Bases</NNPS> </NounPhrase> <COMMA>,</COMMA> – <VerbPhrase> <VBG>Oxidizing</VBG> </VerbPhrase> – <NounPhrase> <NNS>agents</NNS> </NounPhrase> <COMMA>,</COMMA> – <VerbPhrase> <VBG>Reducing</VBG> </VerbPhrase> – <NounPhrase> <NNS>agents</NNS>

Ahmed Eleish - eleisa@rpi.edu

Knowledge from data

- Extracting useful structured information from flowing text
 - Natural Language Processing understanding role/category of word in sentence
 - Semantic Modeling structuring and storing knowledge for question answering
 - Software tools environment for

manipulating and analyzing text data



PUBCHEM > COMPOUND > ACETONE > LCSS

LCSS Laboratory Chemical Safety Summary for CID 18

Acetone

e.g.

PubChem CID:	180
Chemical Names:	Acetone; 2-propanone; Propanone; Dimethyl ketone;
Molecular Formula:	C ₃ H ₆ O
Molecular Weight:	58.08 g/mol

9.1 Reactivities and Incompatibilities

```
X 0
```

Incompatible materials: Bases, Oxidizing agents, Reducing agents, Acetone reacts violently with phosphorous oxychloride.

Sigma-Aldrich; Material Safety Data Sheet for Acetone. Product Number: 270725, Version 4.10 (Revision Date 11/24/2014). Available from, as of January 5, 2015: http://www.sigmaaldrich.com/safety-center.html

from HSDB

A mixture of acetone and chloroform in a residue bottle exploded. Since addition of chloroform to acetone in presence of a base will result in a highly exothermic reaction, it is thought that a base may have been in the bottle.

National Fire Protection Association; Fire Protection Guide to Hazardous Materials. 14TH Edition, Quincy, MA 2010, p. 491-7

Safety Report to Structured Data



reaction.

Start with free text, subjected to NLP analysis and harmonized with Safety Template Model

Situation: Chemical Incident Substance : acetone Substance Role: Substance Parameter Form : Phase : solid Concentration : Conditions : Process/operation: SYNTHESIZE Objects: Chloroform Objects: potassium hydroxide Objects: calcium hydroxide Objects: 1,1,1-trichloro-2-hydroxy-2methylpropane Obejct Role: (base) Outcome: (exothermal reaction) Operator: Reaction: interact Chain of Events: reaction with Ontology Summit 2017: AI, Learning, Reasoning, and Ontologies

<u>Track A:</u> "Using Automation and Machine Learning to Extract Knowledge and Improve Ontologies"



Estevam Hruschka (Associate Prof at Federal University of Sao Carlos DC-UFSCar & adjunct Professor at CMU): "Never-Ending Language Learning (NELL)"

Valentina Presutti, (Semantic Technology Laboratory of the Institute of Italian National Research Council (CNR)): "Semantic Web machine reading with FRED"

Alessandro Oltramari (Research Scientist at Bosch) :

"From machines that learn to machines that know: the role of ontologies in machine intelligence"

Champion: Gary Berg-Cross Ontolog Board Member Ontology Learning/Automation Tasks: A Layer Cake of Ontological Primitives (Buitelaar, Cimiano)



Learn concepts & relations from text etc. Select text fragments and assign them to 8/27/16 an ontological concept.. 14

