National Institute of Standards and Technology

RE-USABLE AND SEMANTIC FRIENDLY METADATA

Dr. Talapady N. Bhat, Dr. John T. Elliott, Dr. Ursula R. Kattner, Dr. Carelyn E. Campbell and Dr. Eswaran Subrahmanian

National Institute of Standards and Technology, Gaithersburg, MD, USA

Current challenges involving forensics data management include organization and integration of vast amounts of information from disparate resources. Ad hoc assigned metadata by data resources are vague and in-consistent and therefore they may lead to mixed search results on automated or manual reasoning. We present a 'root' and rule-based method to build re-usable metadata for both machine and manual processing. A web resource of such terms for scientific publications of NIST has been established . http://xpdb.nist.gov/nike/term.pl

Background

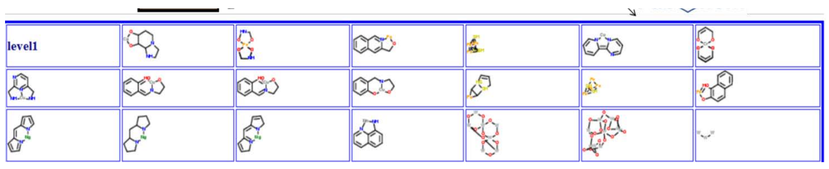
Many Indo-European languages build words by combining a limited set of highly re-used, non-synonymous, short words called 'roots'. Though, such combined words are commonly found in all languages including English, they are more common in some, e.g., German, Hebrew, Latin and Sanskrit, than in others. All the words in Sanskrit are built from 'roots'.

- **Sanskrit, -'Samskrta', is composed from two 'roots' – 'Sam' and 'skara', means 'made ready by combining roots'.**
- **'yoga' is a Sanskrit word – 'Ya-O-Ga' – means join (Ya) brain (O) and motion (Ga)**

Implementation:

- **Chemical Structures- Chem-BLAST.**
- **Cell Images.**
- **Material Genome Initiative.**
- **Natural Language Processing.**

http://xpdb.nist.gov/chemblast/pdb.pl



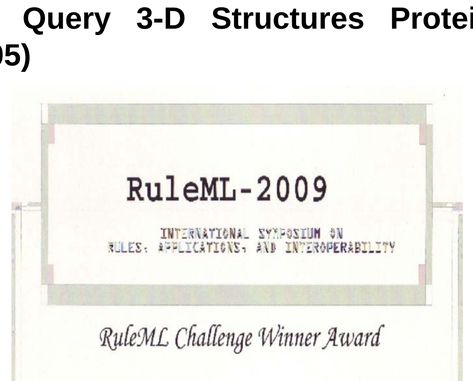
References:

Bhat, T. N. (2010). Building Chemical Ontology for Semantic Web Using Substructures Created by Chem-BLAST. *International Journal on Semantic Web and Information Systems (IJSWIS),* *6*(3), 22-37. doi:10.4018/jswis.2010070102

Plant AL, Elliot JT, Bhat TN: New concepts for building vocabulary for cell image ontologies. *BMC Bioinformatics* 2011, 12:487.

Prasanna, M.D., Vondrasek, J., Wlodawer A., Rodriguez H., Bhat, T. N. Chemical compound navigator: A Web-based Chem-BLAST, chemical taxonomy-based search engine for browsing compounds. Proteins 63(4), 907-917(2006).

Prasanna, M.D., Vondrasek, J., Wlodawer A., Bhat, T. N. Application of InChI to Curate, Index, and Query 3-D Structures Proteins (60):1-4 (2005)



Need Analysis - NIST SRD Terms Assessment

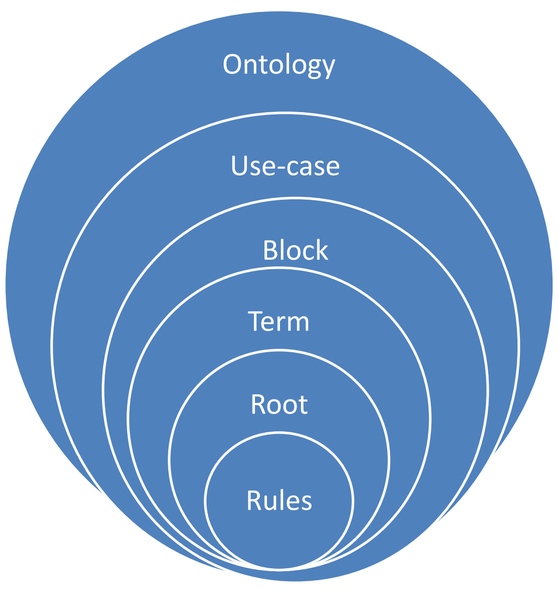
http://xpdb.nist.gov/nike/srd.pl

nist.gov/nike/srd.pl — Untitled Document

Tools Help

SRD#	SRD Name	SRD keyword	Google-auto-complete	Google link
SRD 20	SRD 20 X-ray Photoelectron Spectroscopy Database (Version 4.1)	XPS	xps	XPS Viewer Microsoft Windows
SRD 20	SRD 20 X-ray Photoelectron Spectroscopy Database (Version 4.1)	XPS	xps to pdf	Convert XPS to PDF online free Online PDF Converter
SRD 20	SRD 20 X-ray Photoelectron Spectroscopy Database (Version 4.1)	XPS	xps viewer	What is the XPS Viewer Windows Help
SRD 20	SRD 20 X-ray Photoelectron Spectroscopy Database (Version 4.1)	XPS	xps file	What is the XPS Viewer Windows Help
SRD 23	SRD 23 NIST Reference Fluid Thermodynamic and Transport Properties	air	airtran	AirTran Airways Best Deals on Flights Air Tickets and

Approach



- **Establish rules to create terms from 'roots'**
- **Select a representative set of words from existing terminology for phase based materials from the NIST file repositories as initial 'roots'.**
- **Generate re-usable terms from the 'roots' using the rules.**
- **Establish a web resource of the latest set of 'roots' and terms for review, adoptions, and adaption.**
- **Enable community to suggest additional terms to be incorporated into the web resource.**

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

Root-Based Rules

1. **Choose highly re-used short words as roots.**
2. **keep 'roots' in singular form.**
3. **Avoid including special characters (such as "'", ':','_','-','=') in a root.**
4. **Avoid the use of superfluous words, including stop words such as 'of', 'with' etc, in a term.**
5. **Concatenate 'roots' by a hyphen (-) to form a term.**
6. **Create reasonably discriminating terms. If needed add additional 'roots' to a term to increase its discriminating power.**
7. **Avoid creating too discriminating terms by limiting the number of 'roots' in a term, say to <5 or by splitting a term into multiple terms.**
8. **Maintain a public resource of roots, terms, usage frequency, and examples of usage to help to re-use existing or to construct new terms on demand.**

SUMMARY

Rule and root-based method has the potential to help Forensics science to build re-usable, best-practice, use-case-friendly, scalable metadata by a crowd-sourced approach in a federated setup.