Infrastructure and Methodologies to Facilitate Semantic Graph-search on Technological Information in a Social-media Style

CONTACT DETAILS:
NAME: TALAPADY N BHAT, JOHN ELLIOTT, ANNE PLANT, CARELYN CAMPBELL, URSULA KATTNER
ORGANISATION: NIST
EMAIL: BHAT[AT]NIST.GOV

Objectives

Internet and social media play a significant role in modern information exchange. Every day one uses social-media both to distribute and receive information. Some features of social media like Face-book are

- Community has both data-providers and data-users
- Information is stored in a pre-defined 'data-shelf' of a data-graph
- Core infrastructure for managing information is reasonably language-free

Science has truly evolved to become a community activity involving every country and almost every household. We routinely 'tune-in' to internet-resources to share and seek scientific information.
Creating a social media of scientific information needs an infrastructure where scientists world-wide can deposit results. Some of the issues that one has to resolve prior to establishing scientific social media are:

- Minimize challenges related to local language and its grammar when possible
- Determining the best 'data-graph' to place information in an intuitive way without knowing too much about the data management?
- How to find relevant scientific data without spending too much time on the internet?

We have been developing best-practice, discriminating 'root'-based terms for semantic, scalable, use-case-specific, high-value, on-demand data-graphs using the method used by many languages and in particular by the root-based Sanskrit and Latin.

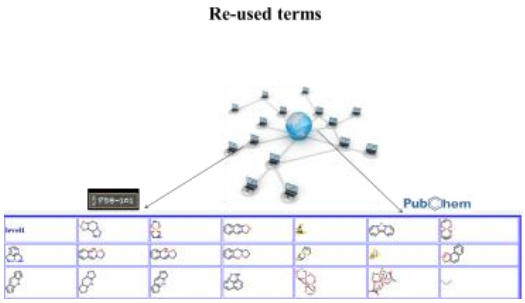
On-going activities:

We consider the above challenge as related to that of super-store that manages a continuously evolving set of items and customer preferences in disparate environments.
How store-managers determine the right shelf to place a- item, say ibuprofen?
o Store manager determines that ibuprofen is a pharmacy product of the type pain-killer to decide in which shelf to place it.
o A customer determines that the article he is looking for is a pharmacy product and he goes to the pharmacy aisle, finds the medicines section, then finds the pain-killers section etc. If ibuprofen is not available in the rack that he is looking at, then he picks up a substitute.
This way a shop-manager easily shelves and a customer intuitively searches without having to know much about the way the items are organized. He can even manage in situations where he does not know exactly what brand names or medicine that he wants or available in a particular store. He can find a substitute item without spending the whole day in a store. The 'shelf-names' are called 'roots' in linguistic sense from which many data-graphs can emerge on-demand. We have been developing standard 'roots' for scientific data exchange.

Results:

Our efforts to develop automated and rule and root-based methods to identify and use best-practice, discriminating terms in generating semantic data-graphs for science started almost a decade back with a chemical structure database. This database has millions of structures obtained from the Protein Data Bank and the PubChem used world-wide. A sample of the roots selected for chemical

structures and a mechanism for global network of computer to exchange data using roots is shown figuratively below.



All databases use these 'structural-roots' as 'critical-globally-shared-nodes'. Any database may extend, trim, mutate or combine roots to form data-graphs on-demand. A user can search through these data-graphs using visual images of the nodes. For this reason, all Web-interfaces are language independent too.

Subsequently we extended our efforts to build root-based terms to text-based data of cell-images. In this work we use few simple rules to define and extend terms based on best-practice as decided by weaning through millions of best practices use-cases chosen from over hundred biological ontologies. Currently we are working on extending this method to publications of interest to Material Genome, Open-Gov and NIST-wide publication archive - NIKE.

URL:

- http://xpdb.nist.gov/chemblast/pdb.pl
- http://www.eurekalert.org/pub_releases/2013-07/aiop-ffm071813.php