

# Facebook for molecules

EurekaAlert!

## ***Scientists design easily searchable Facebook-like networks that could speed the development of new drugs and designer materials***

Social media has expanded to reach an unlikely new target: molecules. Scientists at the National Institute of Standards and Technology (NIST) have created networks of molecular data similar to Facebook's recently debuted graph search feature. While graph search would allow Facebook users to find all their New York-living, beer-drinking buddies in one quick search, the NIST-designed networks could help scientists rapidly sift through enormous chemical and biological data sets to find substances with specific properties, for example all 5-ring chemicals with an affinity for enzyme A. The search approach could help speed up the development of new drugs and designer materials.

The NIST team will present their research at the upcoming American Crystallographic Association Meeting, held July 20-24 in Honolulu.

### **Choosing the Right Words**

Molecules don't maintain their own online profiles, so a key challenge for the NIST research team was to develop a standard language for scientists to describe their research subjects. For example, one research group may describe a material's properties as glassy while another team might use the word vitreous, even though the two words have the same meaning, explained Ursula Kattner, a researcher in the Materials Science and Engineering Division at NIST.

One approach to the problem could be to define a standard set of words, but NIST scientists opted for a more flexible approach that could evolve with time. The search language they developed is similar to Indo-European languages like Sanskrit and Latin, which use short roots to build words based on a set of rules, said Talapady Bhat, a research chemist at NIST who has been leading the effort to develop a shared vocabulary for NIST's scientific databases. He gives the example of the Sanskrit word "yoga," which is based on the roots "Y(uj)", which means to join, "O", which means creator, God, or brain, and "Ga", which means motion or initiation. Similarly, scientists could take the three simple root words "red", "laser," and "light," and combine them into a single compound word "red-laser-light" that conveys a new concept. Using the root and rule-based approach will mean that scientists who know the roots can figure out the meaning of unfamiliar terms, and it also gives scientists flexibility to develop easily understandable new terms in the future.

The NIST team has already applied their root-based vocabulary rules to the chemical structures in PubChem, a "monstrous database" of millions of compounds and chemical substances, to the world wide protein data bank (PDB), and to specific

## Facebook for molecules

Published on Electronic Component News (<http://www.ecnmag.com>)

---

NIST-based databases, said John Elliot, a biophysicist and another member of the team. While the scientific databases haven't reached a Facebook-like level of more than a billion users, they are actively used by many scientists in the NIST community and beyond.

### Graphing Relationships

Once the preliminary vocabulary was established, the NIST team also worked to categorize the descriptions of molecules and scientific experiments in a hierarchical fashion that would allow a search to return comprehensive, yet precise results. A common problem with many search approaches is that they get too many results, said Elliot. Elliot described his team's approach as similar to the problem of locating the Doritos in a large Walmart store. "First you find the grocery market section, then the next level of hierarchy is snacks, after which you go to the chips section, and then you'll quickly know if they have Doritos or not," said Elliot. "So even if a store has a million products, you can find out if they have your product really quickly." The team said the hierarchical approach could also guide scientists who need to pick out key words to index in their research papers.

Organizing the huge amounts of data generated by science is a big challenge, the team said, but it has potentially huge payoffs. Effective graph searches could allow scientists to rapidly identify chemical structures and properties that are needed for the development of new drug agents or advanced materials (such as high efficiency jet turbines or flexible solar cells), results more than worthy of a Facebook like.

Original release:

[http://www.eurekalert.org/pub\\_releases/2013-07/aiop-ffm071813.php](http://www.eurekalert.org/pub_releases/2013-07/aiop-ffm071813.php) [1]

**Source URL (retrieved on 02/01/2015 - 9:54pm):**

<http://www.ecnmag.com/news/2013/07/facebook-molecules>

### Links:

[1] [http://www.eurekalert.org/pub\\_releases/2013-07/aiop-ffm071813.php](http://www.eurekalert.org/pub_releases/2013-07/aiop-ffm071813.php)