

New curation tool a boon for genetic biologists

June 21 2011

With the BeeSpace Navigator, University of Illinois researchers have created both a curation tool for genetic biologists and a new approach to searching for information.

The project was a collaboration between researchers at the Institute for Genomic [Biology](#) and the department of computer science. Led by Bruce Schatz, professor and head of medical information science at the U. of I., the team described the software and its applications in the web server issue of the journal [Nucleic Acids Research](#).

When biologists need information about a gene or its function, they turn to curators, who keep and organize vast quantities of information from academic papers and scientific studies. A curator will extract as much information as possible from the papers in his or her collection and provide the biologist with a detailed summary of what's known about the gene – its location, function, sequence, regulation and more – by placing this information into an online database such as FlyBase.

"The question was, could you make an automatic version of that, which is accurate enough to be helpful?" Schatz said.

Schatz and his team developed BeeSpace Navigator, a free online software that draws upon databases of scholarly publications. The semantic indexing to support the automatic curation used the Cloud Computing Testbed, a national computing datacenter hosted at U. of I.

While BeeSpace originally was built around literature about the bee genome, it has since been expanded to the entire Medline database and has been used to study a number of insects as well as mice, pigs and fish.

The efficiency of BeeSpace Navigator is in its specific searches. A broad, general search of all known data would yield a chaotic myriad of results – the millions of hits generated by a Google search, for example. But with BeeSpace, users create "spaces," or special collections of literature to search. It also can take a large collection of articles on a topic and automatically partition it into subsets based on which words occur together, a function called clustering.

"The first thing you have to do if you have something that's simulating a curator is to decide what papers it's going to look at," Schatz said. "Then you have to decide what to extract from the text, and then what you're going to do with what you've extracted, what service you're going to provide. The system is designed to have easy ways of doing that."

The user-friendly interface allows biologists to build a unique space in a few simple steps, utilizing sub-searches and filters. For example, an entomologist interested in the genetic basis for foraging as a social behavior in bees would start with insect literature, then zero in on [genes](#) that are associated in literature with both foraging and social behavior – a specific intersection of topics that typical search engines could not handle.

This type of directed data navigation has several advantages. It is much more directed than a simple search, but able to process much more data than a human curator. It can also be used in fields where there are no human curators, since only the most-studied animals like mice and flies have their own professional curators.

Schatz and his team equipped the navigator to perform several tasks that

biologists often perform when trying to interpret gene function. Not only does the program summarize a gene, as a curator would, but it also can perform analysis to extrapolate functions from literature.

For example, a study will show that a gene controls a particular chemical, and another study will show that chemical plays a role in a certain behavior, so the software makes the link that the gene could, in part, control that behavior.

BeeSpace can also perform vocabulary switching, an automatic translation across species or behaviors. For example, if it is known that a specific gene in a honeybee is analogous to another gene in a fruit fly, but the function of that gene has been documented in much more detail in a fruit fly, the navigator can make the connection and show a bee scientist information on the fly gene that may be helpful.

"The main point of the project is automatically finding out what genes do that don't have known function," Schatz said. "If a biologist is trying to figure out what these genes do, they're happy with anything. They want to get as much information as possible."

More information: The paper, "BeeSpace Navigator: Exploratory Analysis of Gene Function Using Semantic Indexing of Biological Literature," is available online at [nar.oxfordjournals.org/content ... /nar.gkr285.abstract](http://nar.oxfordjournals.org/content.../nar.gkr285.abstract)

Provided by University of Illinois at Urbana-Champaign

Citation: New curation tool a boon for genetic biologists (2011, June 21) retrieved 25 April 2024 from <https://phys.org/news/2011-06-curation-tool-boon-genetic-biologists.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.