Powerful new ways to electronically mine research may lead to scientific breakthroughs

February 10 2011

A "Science" Perspective by Assistant Professor James Evans and Post-doctural scholar Jacob Foster, both of the University of Chicago, argues that electronically-mined research may lead to future scientific breakthroughs. The computational production and consumption of metaknowledge--the knowledge about knowledge itself--can uncover potential biases inherent in research, as well as information on the backgrounds and previous work of researchers in order to learn more about content, implications and context. Credit: © 2011 Jupiter Images Corporation

The Internet has become not only a tool for disseminating knowledge through scientific publications, but it also has the potential to shape scientific research through expanding the field of metaknowledge—the study of knowledge itself.

The new possibilities for metaknowledge include developing a better understanding of science's social context and the biases that can affect

research findings and choices of research topics, according to an article published by University of Chicago researchers in the journal *Science*. Pooling research-related information online can shed light on how scientists' personal backgrounds or funding sources shape their research approaches, and could open up new fields of study, wrote James Evans, assistant professor in sociology at the University of Chicago, and Jacob Foster, a post-doctoral scholar at the University, in an analysis supported with a National Science Foundation grant.

"The computational production and consumption of metaknowledge will allow researchers and policymakers to leverage more scientific knowledge—explicit, implicit, contextual—in their efforts to advance science," the two wrote in the Perspectives article "Metaknowledge," published in the Feb. 11 issue of Science. Metaknowledge is essential in a digital era in which so many investigations are linked electronically, they point out.

An important new tool for metaknowledge researchers seeking previously hidden connections is natural language processing, one of the rapidly emerging fields of artificial intelligence. NLP permits machine reading, information extraction and automatic summarization.

Researchers at Google used computational content analysis to identify the emergence of influenza epidemics by identifying and tracking related Google searches. The process was faster than other techniques used by public health officials. These content analysis techniques complement the statistical techniques of meta-analysis, which typically incorporate data from many different studies in an effort to draw a larger conclusion about a research question, such as the influence of class size on student achievement.

For scientific research, meta-analysis can trace the connections between data and conclusions in ways that might not otherwise be noticed. For

example, the availability of samples from the Southern Hemisphere related to continental drift has influenced the way in which geologists have made conclusions about plate tectonics.

Metaknowledge also has unveiled the possibility of "ghost theories" implicit assumptions that may undergird scientific conclusions, even when researchers do not acknowledge them. For example, psychologists frequently use college students as research subjects and accordingly publish papers based on the behavior of a group that may or may not be typical of the entire population. Scholars using traditional metaknowledge techniques found that 67 percent of the papers published in the Journal of Personality and Social Behavior were based on studies of undergraduates. The use of computation could accelerate and widen the discovery of such ghost theories.

Entrenched scientific ideas can develop when studies repeatedly find conclusions that support previous claims by well-known scholars and also when students of distinguished researchers go on to do their own work, which also reinforces previous claims. Both of those trends can be uncovered by scholars working in metaknowledge, Evans and Foster said.

Metaknowledge also helps scholars understand the role funding plays in research. "There is evidence from metaknowledge that embedding research in the private or public sector modulates its path," they write. "Company projects tend to eschew dogma in an impatient hunt for commercial breakthroughs, leading to rapid but unsystematic accumulation of knowledge, whereas public research focuses on the careful accumulation of consistent results."

The promise of metaknowedge is its capacity to steer researchers to new fields, they said.

"Metaknowledge could inform individual strategies about research investment, pointing out overgrazed fields where herding leads to diminishing returns as well as lush range where premature certainty has halted promising investigation," Evans and Foster said.

Provided by University of Chicago