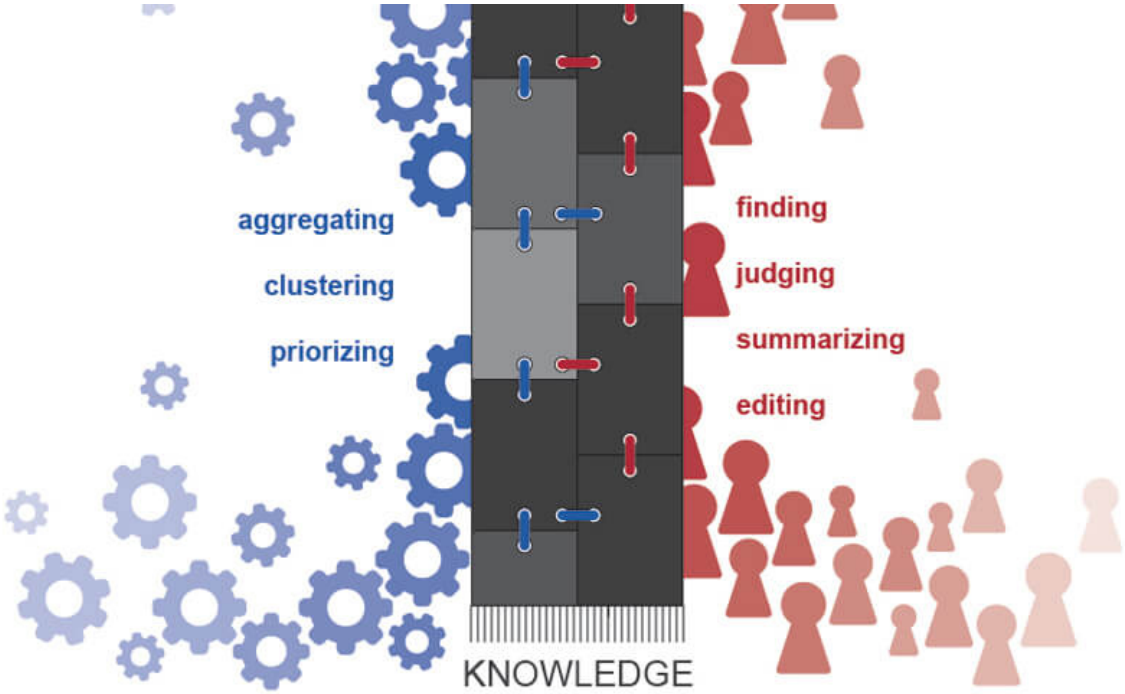# Big thinking in small pieces: Computer guides humans in crowdsourced research

May 10 2016, by Byron Spice

The Knowledge Accelerator, uses a machine-learning program to sort and organize information. Credit: Carnegie Mellon University

Getting a bunch of people to collectively research and write a coherent report without any one person seeing the big picture may seem akin to a group of toddlers producing Hamlet by randomly pecking at typewriters. But Carnegie Mellon University researchers have shown it actually works pretty well—if a computer guides the process.

Their system, called the Knowledge Accelerator, uses a machine-learning program to sort and organize information uncovered by individuals focused on just a small segment of the larger project. It makes new assignments according to those findings and creates a structure for the final report based on its emerging understanding of the topic.

Bosch, which supported and participated in the study, already is adapting the Knowledge Accelerator approach to gather diagnostic and repair information for complex products.

Relying on an individual to maintain the big picture on such projects creates a bottleneck that has confined crowdsourcing largely to simple tasks, said Niki Kittur, associate professor in the Human-Computer Interaction Institute (HCII). It poses a major challenge for Wikipedia, which has been losing core editors faster than it can replace them, he added.

"In many cases, it's too much to expect any one person to maintain the big picture in his head," Kittur said. "And computers have trouble evaluating or making sense of unstructured data on the Internet that people readily understand. But the crowd and the machine can work together and learn something."

Ji Eun Kim of the Bosch Research and Technology Center in Pittsburgh, a co-author of the study, said this learning capability is why the company is extending Knowledge Accelerator for use in diagnostic and repair services.

"We think the Knowledge Accelerator is a powerful new approach to synthesizing knowledge, and are applying it to a variety of domains within Bosch to unlock the potential of highly valuable but messy and unstructured information," she said.

The researchers will present their findings May 10 at CHI 2016, the Association for Computing Machinery's Conference on Human Factors in Computing in San Jose, Calif. In addition to Kittur and Kim, the team includes Nathan Hahn, a Ph.D. student in HCII, and Joseph Chang, a Ph.D. student in the Language Technologies Institute.

In their study, the Knowledge Accelerator tackled questions such as:

- How do I unclog my bathroom drain?;
- My Playstation 3 displays a solid yellow light; how do I fix it?; and
- What are the best day trips possible from Barcelona?

Though compiled from non-experts hired through the [Amazon Mechanical Turk](#) labor marketplace, the resulting articles were rated significantly higher by paid reviewers than comparison Web pages, including some written by well-known experts.

The appeal of crowdsourcing is work can be done quickly and often cheaply by dividing tasks into microtasks and distributing them to workers across the Internet. Lots of people working simultaneously can accomplish many microtasks. But someone has to divvy up the tasks and develop an outline or structure for how all the pieces will finally come together.

Until now, a human supervisor would do those duties, often making choices about research topics and how the findings would be presented without much insight about the subject.

"That's not how the real world works," Kittur said. "You don't just pick an arbitrary structure and fill-in the blanks. You gather information, decide what's important and what's missing and finally arrange the results depending on what you learn. Otherwise, you're limited to very

simple tasks."

Human workers remain critical to the process because of their innate ability to understand unstructured data, such as advice on automobile repairs that can be gleaned from online car enthusiast forums.

Kittur, Chang and Hahn developed a system called Alloy that uses machine learning to provide structure and coherence to the information collected by workers. It automatically recognizes patterns and categorizes the information.

In testing, the researchers used their system to research and write reports on 11 questions. Evaluators found the Knowledge Accelerator output to be significantly better than comparison Web pages, including the top five Google results and sources cited more than three times. One exception was travel sites, Kittur said, which probably is because travel is a strong Internet commodity.

The articles cost about $100 to produce, but Kittur noted the cost might be reduced substantially if the research results could be presented as a list or in some other simple form, rather than as a coherent article.

 More information: [chi2016.acm.org/wp/](chi2016.acm.org/wp/)

Provided by Carnegie Mellon University