

Award Lays Tracks for ktrain



IDA's Welch Award annually recognizes the best external publication by IDA researchers. This year marks the 10th anniversary of the award, named for retired U.S. Air Force General Larry D. Welch, who served as president of IDA from 1990 to 2003 and again from 2006 to 2009. This summary is the sixth in a series reflecting on the 10 winning publications since the award's inception in 2011. The 2016 Welch Award winner for best external publication was "**Mining Measured Information from Text**," a short paper published in *SIGIR '15: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*.

Much of the work the Department of Defense does requires researchers to search highly technical documents for specific information. However, until the search engine Arun Maiya, Dale Visser, and Andrew Wan created, there was no easy, convenient way to make a fully supported search for measured information.

Maiya, Visser, and Wan recognized that the ability to extract measured information from text was not only a common component of research, but also something without a convenient or readily available solution.

The trio first designed a method to effectively extract measured quantities from text. Then, they developed a search capability called MQSearch, a tool researchers use when looking for information on measured quantities in documents.



“We developed a capability that was very important to the sponsor in characterizing large collections of electronic data quickly,” says Visser. “It was a nice complement to the natural language and machine-learning techniques that were deployed within the same analysis application.”

The Welch Award considers all types of written content from any field. Maiya, Visser, and Wan’s article on measured information extraction was the first and, thus far, only computer science paper to win. Maiya described the pioneering win as a special honor.

The work in “Mining Measured Information from Text” is part of a series of data products being developed at IDA. Most recently, Maiya created the ktrain library, an open-source Python library that makes state-of-the-art machine learning more accessible and easier to apply. Maiya says, “The ktrain package not only supports text analyses but also analyses of images, graphs, and tabular data.” The library, which is available from GitHub, has been downloaded over half a million times. It has a variety of applications in different sectors and is currently being used by Stanford University for a machine-learning-enhanced search engine for publications about coronavirus called CoronaCentral.ai.



Arun S. Maiya (amaiya@ida.org) and **Dale Visser** (dvisser@ida.org) are researchers in the Information Technology and Systems Division of the Systems and Analyses Center. Maiya earned a doctorate degree in computer science from the University of Illinois at Chicago. Visser holds a doctorate in physics from Yale University.