

A 'holy grail' of computing hidden in human speech

December 5 2017, by Adrienne Berard

Imagine all 30,557 words of Shakespeare's "Hamlet" being written simultaneously by tens of thousands of people.

To maintain the elegance of the prose, there is a necessary interplay between words. Each line has to coordinate with the next to keep the overall meaning of the play intact.

For some of the most advanced computer programmers, that imaginary scenario is close to reality. Only instead of writing "Hamlet," they're writing software for self-driving cars, updating smartphones, and securing banking systems.

"I think about it like trying to build a complete building on sand, without having any strong foundation. Things are constantly shifting," said Denys Poshyvanyk. "It's [language](#) on top of language."

The way humans use language informs how we process information and the same goes for computers, Poshyvanyk says. Just like human speech, computer language—or [source code](#)—has its own syntax and semantics.

Poshyvanyk, an associate professor in William & Mary's Department of Computer Science, has been working to bridge that human-to-computer language gap for the better part of the past decade. He and a team of nine W&M students are researching the ways code can mirror human communication.

"What we're doing is taking some of the techniques which have been very successful in the area of [natural language](#) processing, information retrieval and machine learning and adopting them in the field of software development," Poshyvanyk said.

From Poshyvanyk's lab in McGlothlin-Street Hall, Ph.D. student Kevin Moran has been exploring what he calls the "holy grail of software engineering"—direct translation between source code and human language. His aim is to translate between people reporting cellphone glitches and the source code responsible for the glitch. It turns out getting humans and computers talking is no small task.

"If I had to pick one word to sum up the whole field of computer science, it would be abstraction," Moran said. "As a computer scientist, you are trying to get a computer to do what you want. The way you interact with source code is an abstraction of the way you would normally communicate in natural language."

In September, Moran and Poshyvanyk presented their research at the International Conference on Software Maintenance and Evolution in Shanghai. The paper was one of 12 co-authored by the duo in the past two years, a level of productivity that has been a hallmark of Poshyvanyk's career.

Over the past decade as a researcher, Poshyvanyk has published 107 refereed conference papers, including 89 co-authored with William & Mary students. He has also published 25 journal papers, 20 of them co-authored with W&M students.

"I believe that having a truly synergistic and collaborative environment is very prolific," Poshyvanyk said. "When students feel they are part of a successful team, that encourages a lot of creativity."

Students say Poshyvanyk's focus on teamwork has taught them accountability and fostered a collaborative lab environment, where they can have pride in what they produce.

"You have this sense of ownership over the work," said Marty White, a 2017 Ph.D. graduate of the department and senior lead scientist at Booz Allen Hamilton. White has co-authored five papers with Poshyvanyk. "Yes, it's a frenetic pace, but you don't even really think about it. It's like one of those things, if you want something done, give it to a busy person."

Even with a high volume of research projects, students say the majority of their ideas are left on the cutting room floor.

"At any given time, we have a backlog of 20 projects we want to be doing," Moran said. "As a result, only the best ideas make it to the research stage."

Some of that research has recently garnered international acclaim. For the past two years, Poshyvanyk was recognized for the Most Influential Paper at the International Conference on Program Comprehension. The awards are given 10 years after a paper is published. Poshyvanyk's 2007 paper introduced the idea of creating a search engine for source code.

"The idea we had was to create a Google for developers," Poshyvanyk said. "We wanted to develop an engine that would index your source code. You would be able to write a natural language query without any restrictions and simply search your code base."

The paper led software developers to explore entirely new ways of searching source code, Poshyvanyk said. Hundreds of papers have followed on the topic and several companies created their own Google-like code search engines.

Poshyvanyk's more recent work has also gained recognition. In November, he won the Automated Software Engineering Distinguished Paper Award for a paper he co-authored with his current Ph.D. student Christopher Vendome, former Ph.D. [student](#) Mario Linares Vasquez and three Italian collaborators.

Their [paper](#) evaluated metrics to test how well programmers understand the code they are tasked with reading. They found that readable pieces of code do not have a direct correlation with understanding. While programmers are all writing in the same language, the meaning of text hinges on the author.

"The abstractions you create while writing your program make it hard for me to easily jump in and read what you wrote," said Poshyvanyk. "Sometimes for me to understand how one piece of source [code](#) is used, I have to look through an entire library written by other people."

Provided by The College of William & Mary

Citation: A 'holy grail' of computing hidden in human speech (2017, December 5) retrieved 7 May 2024 from <https://phys.org/news/2017-12-holy-grail-hidden-human-speech.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.
