



eCOMMONS

Loyola University Chicago  
Loyola eCommons

---

Computer Science: Faculty Publications and  
Other Works

Faculty Publications and Other Works by  
Department

---

3-2012

## Digging into Data

George K. Thiruvathukal

Follow this and additional works at: [https://ecommons.luc.edu/cs\\_facpubs](https://ecommons.luc.edu/cs_facpubs)



Part of the [Computer Sciences Commons](#)

---

This Article is brought to you for free and open access by the Faculty Publications and Other Works by Department at Loyola eCommons. It has been accepted for inclusion in Computer Science: Faculty Publications and Other Works by an authorized administrator of Loyola eCommons. For more information, please contact [ecommons@luc.edu](mailto:ecommons@luc.edu).



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License](#).

## DIGGING INTO DATA

By George K. Thiruvathukal, Associate Editor in Chief



**T**HE STORIED LIFE OF A PROFESSOR OR ACADEMIC IS OFTEN PORTRAYED AS ONE WHERE THIS INDIVIDUAL HAS COPIOUS FREE TIME ON HIS OR HER HANDS, NOT ALTOGETHER REMOVED FROM THE VISION OF A CAREFREE AND GLAMOROUS HOLLYWOOD LIFESTYLE. IT MIGHT JUST BE ME, BUT TODAY'S

academics seem to be working more than ever to maintain what's still a great career and work-life balance. It's not a question of whether we work the hours, because we do; it's just a matter of when.

The past several months have been a bit of a whirlwind for me. For the first time ever, I organized a conference from conception to realization—the Chicago Colloquium on Digital Humanities and Computer Science. Suffice it to say, I have a new appreciation for the incredible work others put into conferences, and I'm still recovering. Being the organizer in chief, I truly can't do justice to describe in detail what actually happened at this conference, given the array of topics that were covered. The technical presentations reminded me a great deal of some of my earlier work in high-performance computing and supercomputing, where seemingly every person present was thinking about how to apply computer science and computational methods to understand just about every problem in the humanities. I know that the digital humanities field might not be familiar to many readers, but this area has really taken off in the past few years and seems to be experiencing something similar to what happened in the 1990s with computational science, where just about every corpus (body of work) is being analyzed in one way or another using algorithmic and data-driven methods—the same methods we're applying to computational science and engineering.

Many people might wonder what the field of digital humanities actually is. In a nutshell, it's the application of computational methods to the humanities. To understand why anyone would want to do this, consider the following question: What do you do with a million books? This, of course, refers to the major human undertaking by Google Books to digitize seemingly all of the books on the planet. Your first reaction, if you enjoy reading—as I once did when I had free time—might be to say, “Read them.” If only life were so simple. (And even if you read 500 or more

words per minute, you couldn't read all of them anyway.) We no longer “just read” things; in our technology-driven world there are now so many ways to present, read, perceive, and analyze text. In particular, the use of text analytics and visualization can greatly guide how a person reads a text, especially if the text isn't well understood or actively studied.

To the end of using emerging methods from digital humanities, I decided to do a wacky experiment. This experiment, in the end, had nothing to do with digital humanities per se, yet would make use of one of the tools from this community, so to speak. To describe my experiment concisely, I wanted to determine whether a given funding opportunity was relevant to my research by examining the text of various US National Science Foundation solicitations. So I paid a visit to Wordle.net, which is a toy for generating word clouds from text that you provide. The site is extremely easy to use. You simply click on the create button and the site gives you a form to enter the text that you'd like to visualize, which you can then copy and paste or use the URL for your records. Because I only wanted to analyze the text of the solicitation related to the actual research being targeted, I opted to copy and paste. I tried the text for a number of research solicitations that were currently “open,” and ultimately found one that generated the word cloud shown in Figure 1.

Suffice it to say, we had a match for my research interests, which will be evident to readers who have read *CiSE*'s Scientific Programming department.

Wordle is a rather neat tool that basically uses the results of one of the first computer science programs—word count—in a rather novel way. It's a trivial algorithm per se, but the visualization is less trivial. Careful thought has been given to the layout and presentation, so that your eyes see (at a glance) which words are truly emphasized in the solicitation. Given that my work tends to be focused on the

