3-2008

At Issue

Bill Feireisen

George K. Thiruvathukal
Loyola University Chicago, gkt@cs.luc.edu

Recommended Citation

Bill Feiereisen and George K. Thiruvathukal, "At Issue," Computing in Science and Engineering, vol. 10, no. 2, pp. 60-64, Mar./Apr. 2008, doi:10.1109/MCSE.2008.49

Are We There Yet?

By Bill Feiereisen

COMPUTING IN SCIENCE & ENGINEERING IS A FANTASTIC READ—AT LEAST, THAT'S MY OPINION. I ENJOY IT FROM COVER TO COVER, AND I CONFESS TO SOMETIMES READING IT IN BED. *CiSE* REFLECTS ONE OF MY MOTIVATIONS IN COMPUTATIONAL SCIENCE, WHICH IS TO BE "HANDS ON." IN ADDITION TO MORE THEORETICAL ARTICLES, THERE'S A DO-IT-yourself problem-solving aspect that makes me want to run to the keyboard and try out the latest technique that I've just read about or visit the Web site described and look for software.

However, *CiSE* is ultimately a magazine, and many of these extracurricular activities are left as an exercise for the reader. Despite the fact that the American Institute of Physics and the IEEE Computer Society maintain a superb Web site for *CiSE*, it's basically an electronic reproduction of the magazine itself. Could the magazine's Web presence be more? Could it become a place we come back to every day because it always features something new with which to solve the next problem or innovative material to present in our next class?

The scientific and engineering world is currently experiencing an information technology revolution. It follows on the heels of the general excitement about Web 2.0, but our revolution is perhaps a bit more soberly thought out, as is appropriate to our field. Nevertheless, we can see the experimentation pioneered on the Web taking hold in science. It's in the vast changes to our institutional scientific libraries, in the comprehensive selection of electronic materials beyond journals and books, in the links to similar materials archived all over the world that are accessible instantaneously from the comfort of the keyboard, and in the phenomenal literature search capabilities that go with them. Even Google is making this world of scientific "literature" available to everyone, which can only be good for science and engineering. How exciting!

Let's look more closely at a couple of concrete examples. Like many of you, I subscribe to *Science* magazine, and as with many publications these days, it has an extensive Web site (www.sciencemag.org) with supporting material for each of its articles. You can find downloadable versions of the articles and readable versions of the figures but also an archive of methods and sometimes the data from the studies themselves. If you click on the link at the bottom of the main page, you'll find videos and slideshows as well as a subscription link to podcasts. I bought my first iPod two years ago after discovering podcasts, and these audio articles have become a regular part of my long drive to work.

The Public Library of Science (PLoS; www.plos.org) makes articles in the biosciences freely available without subscription. It recently launched an experiment in online review and feedback for a portion of its Web site called "PLoS One." Visitors are invited to criticize, correct, annotate, and start discussions about articles, with each article now including a clickable block that lets online readers do this. It remains to be seen how well unfiltered feedback will work in the context of a top-quality scientific journal, but it shows the influence that blogs and wikis are beginning to have on science and engineering. There's even an annual conference discussing blogging for science (specifically, the North Carolina Science Blogging Conference).

Communications of the ACM (CACM) just announced a policy change in which articles will first appear on the Web and will no longer be required to conform to "print format." This is a tipping point for the traditional journal that we've read all these years and surely signals a sea change in what we consider to be a scholarly article. One of our own copublishers, the IEEE Computer Society, is also considering a move toward a consolidated modern online presence for its journals.

At the other end of the spectrum are the many software repositories, ranging from commercial to open source. The open source and academic software development communities have been vibrant for as long as I've followed them. I regularly browse Freshmeat (www.freshmeat.net)

 Copublished by the IEEE CS and the AIP 1521-9615/08/$25.00 ©2008 IEEE

CiSE's Place on the Modern Web

In this installment of At Issue, two of *CiSE*'s editorial board members describe the magazine's current and future Web presence. We start with Bill Feiereisen's discussion of science on the Web and what it could mean for *CiSE* and end with George Thiruvathukal's discussion of the challenges therein. *CiSE*'s official homepage is www.computer.org/cise; we gladly welcome comments with suggestions or feedback! You can contact lead editor (Jenny Stout; jstout@computer.org) or write to either Bill or George (their email addresses appear at the end of their respective essays). —*eds.*

and Sourceforge (www.sourceforge.net) for new and interesting projects and search them for existing software before I write my first line of code.

I tend to use Perl extensively. I'm still in awe of the tremendous riches of Perl scripts and programs available and point you to the Comprehensive Perl Archive Network (CPAN; www.cpan.org), which is just fun to browse. It has a wealth of software freely available, but you have to know what it's for and how to use it. Almost every project has extensive example programs, but the connection to a particular science area is usually missing.

Which brings me to a common conundrum—a lot of the "hands-on" aspect is still left to the reader to figure out alone. Does this offer an opportunity for *CiSE* and for computational science in general? Is there need or room for an expanded *CiSE* Web site that ties the pieces together for a computational scientist who wants to figure out how to do something?

It's clear that many modern science and engineering opportunities are in multidisciplinary work, but most of us grew up in a particular discipline. The challenge is to broaden your expertise across the boundaries of your traditional discipline, but the option of repeating your schooling doesn't really exist nor should it. You spent years coming up to speed in your chosen discipline, but simply repeating this for a second area leaves you high and dry when the opportunities take you to yet a third area. So how do you become productive in a new area of science and engineering?

Like many others, I've been exploring the boundary between high-performance computing and biology. I'm on a campaign to "get smart" about modern bioinformatics in this new age of the data deluge of sequenced DNA. Up until a few years ago, the total volume of curated DNA data available for study was, at most, a few tens of gigabytes. This wasn't big-data science as we know it, and entire databases could be stored and processed on desktop machines.

Needless to say, this has changed: a revolution is under way in the laboratory with the introduction of the latest generation of sequencing machines. They now produce terabytes in days, but they've also introduced new challenges for the algorithms that assemble sequences into complete genomes and subsequently analyze them. The computational demands are increasing exponentially so there's a true need for high-performance computing in biology.

Take, for example, the field of metagenomics (www.nap.edu/catalog.php?record_id=11902), which studies communities of micro-organisms without the laboratory step of isolating individual species (most microbes can't be cultured individually in a laboratory). Researchers in this field typically gather environmental samples containing thousands of unknown species and attempt to determine their DNA with new laboratory techniques. Because researchers have to start with fragments of DNA from many organisms, it helps to have reliable algorithms that separate these fragments into bins representing each species before combining them into whole genomes. As you can imagine, this involves a great deal of statistics and linear algebra. Perl provides many math modules and interfaces to well-known linear algebra packages such as Lapack, but you're on your own if you want to quickly prime the programming pump and apply these in a scientific area that's new to you.

Naturally, I spend a great deal of time with search engines to avoid reinventing the wheel, and because of this, I've found that many people before me have written modules that tie more general methods together. One example is the set of BioPerl modules that provide many standard bioinformatics services, easy formulation of inquiries to many genetics databases, formatting of results, and delivery to statistical and graphical analysis programs. But how exactly can these analysis programs help me? For someone who didn't grow up in the biological sciences, acquiring this context is difficult and points to a need for a new type of resource. Traditionally, this has been the domain of textbooks, which draw together the collective experience in a scientific area and convey it as a course, albeit in a static manner. To remedy this, book publishers are moving online in a fashion similar to journals (see, for example, the excellent new bioinformatics text *Discovering Genomics, Proteomics and Bioinformatics*, http://wps.aw.com/bc_campbell_genomics_2/43/11232/2875502.cw/). But where is the one-stop-shopping that draws together the experience of many experts in this area?

CiSE is considering an expanded online presence, and

my suggestion to its editorial board is to construct a Web site that incorporates many of the ideas discussed here. I focused on bioinformatics, but an up-to-date compendium of knowledge and links in many areas of computational science would be invaluable to our readers. I can foresee my journey into high-performance bioinformatics finding its place on such a Web site and would look forward to reading about your experiences in new areas as well.

The big question is whether *CiSE* can pioneer a lively gathering place on the Web for those interested in computational science. In addition to the online version of the magazine itself, it would need to support blogs associated with the departments and columns. Perhaps the Web site could be where we gather these blogs, demonstration software, tutorials, visualizations, news, business happenings, and maybe even jobs? It could have a community-contribution aspect in much the same way as Wikipedia, in which the general computational community is invited to share their experience and pool their knowledge.

In the next few months, the magazine's editorial board will be considering these questions in earnest, and we'd greatly appreciate your feedback. Please let us know what you think!

Bill Feiereisen is a computational scientist at Los Alamos National Laboratory and one of the founding directors of the New Mexico Computing Applications Center. His technical interests include computational astrophysics and biology and their intersection with high-performance computing. Feiereisen has a PhD in mechanical engineering from Stanford University. Contact him at bill@feiereisen.net.

If We Build It, Will They Come?

By George K. Thiruvathukal

I'VE BEEN INVOLVED WITH *CiSE* MAGAZINE FOR SEVERAL YEARS NOW: IT SEEMS LIKE JUST YESTERDAY, BUT I WROTE MY FIRST ARTICLE FOR IT IN JANUARY 2002. AT THE TIME, I didn't really know much about the magazine. I received a message from my predecessor as its Scientific Programming department editor, Paul F. Dubois, who invited me to write about my work on Java and the Java Grande Forum. For those of you who regularly read my contributions to the department, it should come as no surprise that I thought the Web was cool well before I wrote my first article for *CiSE*. A big part of what made it cool was its potential for interactive applications and community portals, which has become much more compelling with the advent of Web 2.0.

So when our editor in chief, Norman Chonacky, asked me to write a counterpoint to Bill Feiereisen's excellent essay about why and how *CiSE* should be on the Web, suffice it to say I was a bit perplexed about how to present an argument against something I truly believe would be useful for us and that I know will end up being expanded from its present form. Nevertheless, I decided to give it a try. I see this exercise as being helpful to us because I think we'll be able to cull a strategy for enhancing our presence on the Internet (a term I prefer to the Web) that—I hope—will have impact.

Let's begin with an analysis of our current Web efforts, which I think highlight the daunting challenge we face in trying to bring our Web presence up to the standards of the mid 1990s. As current readers hopefully know already, *CiSE* has a Web presence, so you *should* know how to get there. As an experiment, I googled *CiSE* on 7 December 2007, and here's what I got:

- www.cise.nsf.gov (the US National Science Foundation's Directorate for Computer & Information Science);
- www.cise.ufl.edu (the Department of Computer and

Information Science and Engineering, University of Florida);
- www.catholiccincinnati.org (Roman Catholic Archdiocese of Cincinnati);
- www.cise.missouri.edu (Center for Innovations in Education); and
- http://cise.aip.org (the American Institute of Physics' site for *Computing in Science & Engineering* magazine).

Although the workings of Google's search algorithm are mysterious and closely guarded secrets, the results are ordered here by relevance. We're fortunate that a search for our name is relevant, but we'd certainly be a great deal happier if we were at the top of the list of results.

One particularly disturbing aspect, however, is that the site where our content (in particular, the current issue) is most readily available—www.computer.org/cise—doesn't appear in the first couple of *pages* of results. This in itself is revealing: is it possible that most people who do come to our site come by way of the AIP and not the IEEE Computer Society? I wonder whether the Google gods are telling us something.

In any event, this illuminated a great potential for my counterpoint to Bill's essay. As I see it, a significant motivation for a small publication like *CiSE* to have a prominent Web presence in today's online world is to drive traffic to our site based on keywords, which can be readily extracted from our content over the past several years. So I decided to explore the keywords notion in greater depth. I started with the non-noise words in our title: computing, science, and engineering. Lo and behold, http://cise.aip.org comes out on top—but www.computer.org/cise comes out 15th. Thus, when people search for the spelled-out title of our magazine, they find us, but they find us via the AIP site first. There would appear to be at least a glimmer of hope here.

Of course, it's important not to read too much into the fact that *CiSE* contributors find themselves by typing in the non-noise words in our magazine's title. These searches are clearly being performed by folks who already know of and—we hope—love us. But what happens when someone performs a search based on meaningful content-driven keywords or combinations thereof? In other words, how do readers "yet to walk through our doors" find us? I decided to give some other keywords and phrases a try:

- "computational" and "science," not within the first two pages of search results;
- "computational science" as one string, ditto;
- "computational" and "physics," ditto; and
- "computation" and "engineering," ditto.

This is a problem! Although our content is directly linked to our page on a regular basis, our site is seldom found when searching for obvious keyword combinations. Whatever we do on the Web must truly begin with an understanding of how to remedy this situation (assuming we're really serious about attracting nonsubscribers).

A possible explanation for this is that all of our content is "locked up" in the digital library, which is only available to paid digital library subscribers (many of whom are institutional). Although this might work for some journals published by the IEEE, it doesn't work for *CiSE*, which is fundamentally a hybrid magazine/journal. Most of our articles and columns appear without abstracts and lack keywords, although that information is provided in the background metadata. This precludes proper indexing by search engines and, therefore, keeps prospective readers away, but interestingly enough, this phenomenon would seemingly make it difficult for paid subscribers to find our site as well.

Although our content is directly linked to our page on a regular basis, our site is seldom found when searching for obvious keyword combinations.

It's abundantly clear that as a first step toward generating interest in our site, we must start by looking at a content model before doing anything else. *The New York Times* created quite a stir recently when it decided to make all of its "TimesSelect" content available for free (www.nytimes.com/marketing/ts/). Many established magazines do a slight variant: they offer a print subscription, subscribers get the latest issue before online freebie hunters, and, after a month (or two, in our case), the current issue becomes freely available online. This allows anyone to read the content for free at some point, but with an appropriate lag-time inserted.

As I argue in my day-to-day life (as a professor, technologist, and hacker), most problems—and solutions to those problems—aren't technical but human in nature. And

humans are much more difficult to move than computers, even when taking a purely physical perspective (average human weight versus average computer weight). The question is, given that *CiSE* is published jointly by the AIP and the IEEE Computer Society, are the powers that be open to models that work in the Web 2.0 landscape? Just about everyone is offering one free service or another, so what about us?

In addition to my concern about not being able to drive traffic to our existing Web sites anytime soon, I'm concerned about our audience. Today's Web is much different than it was in the early days, which is evident just by taking a look at social networking sites. It strikes me that today's "younger" readers grow up within their own communities, which take on an increasingly ad hoc nature (bottom up) versus the top-down model many magazines and newspapers use. We'll truly need to experiment, which isn't a problem for a community that does computational science and engineering.

It's worth a momentary digression to consider what the early days were all about. Tim Berners-Lee—a physicist with strong knowledge of computer science—invented a little thing called the World Wide Web. The central idea of his work, lest we forget, was to create a system wherein scientists could share documents and research results. Ironic, isn't it? Here we are in 2007, and our current approach to the Web fails to make articles, results, code, and so forth readily available, except for a few rare occasions (the Scientific Programming department being one of them, and I still feel like I should be doing so much more as its editor). Yet, we're in the same community as the man who started it all. For this and many other reasons, I'm worried that the computational science community is failing to attract the next generation of students who are already hooked on the Web. We need to have a plan to engage them, possibly through volunteer initiatives and competitions. There's a generation gap of sorts, and I'm not sure the Web in its own right will generate interest in *CiSE* unless it is—pardon my '90s urban radio reference—crazy, sexy, cool.

So as not to rain on anyone's parade, it's clear that we need to become savvier if we want to participate in the Internet space. Moreover, virtual worlds and games could well supersede the Web in a disruptive way in the near future. Even if we expand our Web presence, what's our plan to engage today's youth, who spend their time in this virtual world, even when sitting in my lectures?

I've recently started teaching a "History of Computing" class at Loyola, in which I have many young students—all of whom are among the first generation to spend most of their formative years on the Web—and it strikes me that we have a very discerning audience in the making. These students are more than happy to tell you that a site is "so Web 1.0" or "so yesterday," so how do we strike a balance between attracting young, aspiring computational scientists but not alienate our present readership or community? There's no silver bullet, but I believe that whatever strategy we evolve will need to be multipronged, will take several years, and require significant investments (at least, of time). It won't be trivial, and it'll require us to engage the community—which means, among other things, that we need to have a diverse group of older and younger editorial board members and undertake several experimental initiatives, some of which will challenge virtually every assumption underlying our existence.

If we simply build it, they won't come.

George K. Thiruvathukal is a professor of computer science at Loyola University Chicago. His research interests include programming languages, operating systems, and distributed systems. Thiruvathukal has a PhD from the Illinois Institute of Technology. Contact him at gkt@cs.luc.edu.