From the Editor in Chief...

Collective Wisdom

A Modest Proposal to Improve Peer Review, Part 2

Fred Douglis • IBM T.J. Watson Center • f.douglis@computer.org

In my last column (September/October 2007), I reflected on the trouble IC had when encountering a submission that was substantially similar to another submission received elsewhere — a definite no-no. I described some tools for detecting self-plagiarism and considered a possible way to get authors to make submitted works available alongside those already published, so that we could more easily detect overlapping submissions and other violations of submission guidelines. One key requirement would be that either the submissions themselves be stored and managed securely, with the same guarantees as submissions in the review process or that only some derived data be stored in a centralized fashion. In the latter case, we could readily tell that two manuscripts were similar without actually having access to either one: the arbitrator detecting the overlap wouldn’t need to be trusted to protect the content, either.

This arbitrator is an example of a broader class of neutral agents that can serve many functions on the Internet. In this column, I’ll focus on a second aspect of the reviewing process — reviewer integrity — and then return briefly to a few other real-world examples.

Reviewing the Reviewers

Author misconduct (even if intentional) isn’t the only concern a periodical editor or conference program chair has: the flip side of the process is inappropriate referee behavior. For the most part, I believe such behavior is rare — instances of reviewers showing favoritism or disclosing confidential submissions, for instance, don’t often arise. (Of course, subtler impacts on a confidential submission are certainly possible but difficult to quantify — say, a reviewer learns an idea from a manuscript and doesn’t consciously build on it but is influenced by it.) However, one particular form of misconduct seems positively rampant: reviewers who agree to perform a particular task by a particular time and then don’t deliver, either by failing to provide their reviews or by providing such substandard reviews that the evaluations are worthless.

For transactions, delays are part of the game because without a specific date by which to publish a specific paper, deadlines are typically lax. For magazines such as IC, however, we try to keep a closer eye on reviews for special issues because, like conferences, we have specific deadlines. Any one reviewer for a periodical is typically responsible for just one or two reviews, and if those slip, it isn’t too hard to recover, but conferences are another story: I’ve been a program chair or a program committee (PC) member many times, and it seems that almost every conference has at least one PC member who does none of the reviews by the appointed deadline. Sometimes this results in a scramble to get each of the remaining PC members to review an extra few papers in order to have enough evaluations for every submission, which in turn can lead to slippage in the acceptance notification deadline. In short, everyone loses.

Sometimes, people have legitimate excuses for not fulfilling their obligations to a PC, but it’s truly inexcusable to fail to review a large number of papers without any advance warning. Although last-minute emergencies can happen, PC reviews usually take place over a span of several weeks, if not a few months, so if a reviewer hasn’t started reviewing the papers as the deadline approaches and an emergency arises, the reviewer is still at fault. If the deadline comes and goes without even an explanation, as is so often the case, such behavior is reprehensible. Although conferences overcome this by requiring reviews to be submitted in small batches — for instance, a third at a time — to impose earlier deadlines and identify problem reviewers early in the process, this approach does not completely avoid such problems.

I keep a list of the handful of people I know to have shirked their reviewing obligations, either when I was a PC chair or simply a member of the
Committee, and I will never ask them to be on a PC I organize. But others don’t have my list, and I don’t have anyone else’s. Only occasionally have I been asked my opinion about a potential PC member, and most recently the request came after the PC member had already been invited. So what’s the right way to ensure a reputation system for academic reviewers that’s both useful for avoiding these problems and fair to the reviewers? (Not to mention that you must fully consider any legal and ethical ramifications: I’m not exactly willing to post an online list of people I wouldn’t put on a PC.) After some recent plagiarism incidents, the IEEE has developed a policy of initially warning authors and then formally banning them after repeated incidents. It knows about repeated incidents because the editor or program chair notified the IEEE about the occurrence. Professional organizations such as the IEEE, the ACM, and Usenix could conceivably collect statistics about reviewers — in fact, for referees of periodicals, some societies maintain information about the number of manuscripts reviewed and how many reviews were on time or late, but to my knowledge, they don’t maintain the more interesting statistic of reviews that were so late they were rescinded. Given that PCs are distributed and have a lot to lose from problematic PC members, it’s probably more important to avoid bad PC members than to avoid bad magazines or transaction referees. Could each professional organization start tracking its PC members, turning what has been until now a poor, ad hoc, word-of-mouth reputation system into something organized into institutional memory? Tracking how people fulfill their reviewing obligations would not only help avoid relying on people who prove to be untrustworthy, it would also help find the best reviewers to use on future PCs and ultimately as program chairs. But why does the specific professional organization matter? If someone drops the ball for an IEEE conference, he or she is probably just as likely to leave an ACM conference in the lurch. So, we’re back to the question I raised in the last issue: how might different publishers share enough information to avoid reliance on people who prove time and again to be unreliable, without falling back on word-of-mouth anecdotes or impugning the integrity of someone publicly in what is normally a very private process?

Collective Wisdom: A Possible Solution

In essence, I think something like a credit-reporting agency might work. A strawman proposal is as follows:

• Create an independent agency, jointly sponsored by the professional organizations that run conferences.
• Provide a mechanism for reporting to this agency when reviewers have unjustifiably failed to satisfy their obligations. This would cover missing reviews, as well as other problems such as quality. Positive reports would also be encouraged.
• Send a report to the reviewer that an incident has been logged, giving that reviewer the opportunity to dispute the report. Some oversight would be necessary to purge inappropriate reports, possibly with additional records from the person who logged the initial complaint (to encourage whistle blowers to not “cry wolf”).
• Provide a way for program chairs, editors, and others who select reviewers to interrogate this database.
• Set policies for aging out information, so that minor infractions are quickly ignored and even major ones don’t affect a referee indefinitely.

Certainly, various issues could arise with such a system, such as ensuring that it’s used only for PCs and similar functions and not (for example) to check on a professor’s integrity before granting tenure. If the system is too onerous or McCarthy-like (http://en.wikipedia.org/wiki/Joseph_mccarthy), people will be reluctant to serve as referees. The referee has an important task, so we need more good ones and fewer bad ones.

In discussing this topic with my colleagues, a general reaction has been that the academic peer-reviewing system doesn’t properly reward people for good behavior as referees. Some magazines recognize especially detailed reviews with an annual list of outstanding reviewers, but not very many periodicals do this. Providing a reputation system in the form of some independent, neutral agent offers the opportunity to recognize and reward favorable behavior. For example, the organizations sponsoring such a system could tithe some small amount from each conference or periodical’s budget (say, US$50) to provide a few small monetary prizes to highly rated reviewers at random. If the odds and amounts are low enough, this won’t be an incentive for collusion, but just having a reward could encourage people to participate. How many people respond to online surveys that offer them some unspecified chance to win a cash prize, relative to those who are simply asked to take the survey?

I also learned, unsurprisingly, that I’m not the first in the academic CS community to propose rating reviewers. In 2006, S. Keshav, a professor at the University of Waterloo, proposed letting authors rate reviewers (see http://keshav-essays.blogspot.com/2006/09/reviewing-reviewers.html). This approach aimed to address poor or biased reviews rather than the complete no-shows that have troubled me so much, but the gist is similar.

Finally, some discussion has occurred of open reviewing and even open publication (submissions would be public, something that would ad-
dress the self-plagiarism issue I raised in the previous column). A panel at a Computing Research Association workshop discussed publication models in 2006, and recently Global Internet ‘07 experimented with an “open reviewing” policy. Simply put, authors could see who wrote each review. Open review would add an onus of responsibility to the process of writing reviews, although extra work would be necessary to determine that a given individual entirely failed to submit promised reviews, and a bystander would have no way to know whether that failure was excusable or not.

**Agents, Agents Everywhere**

Agents to mediate the detection of overlapping submissions and the behavior of conference or periodical referees are just two examples of a general class. Much of the work on Internet agents focuses on computer programs acting on a person or organization’s behalf, but these agents are intended to act autonomously. In a sense, they serve society as a whole rather than any one organization, providing a trustworthy service and central point of contact.

What other sorts of intermediation might these agents serve, and what types already exist? Some sites such as bizrate.com and epinions.com let people rate businesses and products, and companies such as Equifax provide credit histories. Product recommendation engines can suggest music and movies you might like based on similarity ratings made by other people (in fact, such social search is the theme of this issue of *IC*).

Generally speaking, the thing these systems have in common is matchmaking and reputation management. Some of the information is more overt than others — for example, I know exactly who reports what about me to Equifax — but some is more sensitive. How much information is appropriate in any given context? I might report that Spongebob Squarepants was a no-show at the Fourth Conference on Aquatic Comic Characters, but when the Sixth CACC comes along, can the new program chair find out that Spongebob was on a past PC? My PC? Will my program chair find out that Spongebob sent a lower score than it deserves?

Reputation is important, and I wouldn’t have been inspired to write about referees dropping the ball if it weren’t so commonplace. But until we have a reporting agency, people should bear in mind that there’s an informal mechanism already in place: word of mouth. Let the program chairs and editors who ask for your help have only good things to say about you and continue to volunteer. The system needs you.

**Acknowledgments**

Please refer to the previous column for the list of acknowledgments of people who contributed to both installments. My predecessor as editor in chief of *IC*, Bob Filman, deserves particular thanks for his detailed comments and suggestions on the topic of neutral agents. The opinions expressed in this column are my personal opinions. I speak neither for my employer nor for *IEEE Internet Computing* in this regard, and any errors or omissions are my own.

**References**


---

**IC Welcomes New Board Members**

Azer Bestavros is a professor in the computer science department at Boston University. His research interests include networking and real-time systems. He chairs the IEEE-CS technical committee on the Internet. He served on program committees and editorial boards for major conferences and journals in networking and real-time systems and received distinguished service awards from the ACM and the IEEE.

Enning Schulzrinne is a professor in the Department of Electrical Engineering and chair of the Department of Computer Science at Columbia University. His research interests include Internet multimedia and telephony services, signaling, network quality of service, scheduling, multicast, and performance evaluation. Schulzrinne has a PhD from the University of Massachusetts. He is the coauthor of the Internet standards-track protocols RTP, RTSP, SIP, and GIMPS.

Shengru Tu is a professor and the director of the Visual Computing Research Lab in the Computer Science Department at the University of New Orleans. His research interests include service-oriented architectures, Internet applications and Web services for geographic information systems (GIS), GIS system integration, software testing techniques such as regression test selection for Web services, concurrency analysis using Petri net, and Web-based training and instruction systems. Tu has a PhD in Electrical Engineering and Computer Science from the University of Illinois at Chicago.