What it's like to be a POPL Referee
or
How to write an extended abstract so that it is more likely to be accepted.
Mark N. Wegman

IBM Research
Yorktown Hts, N.Y. 10598

I've just served on my second POPL program committee, and thought that some people may be interested in my observations. A POPL extended abstract should be designed to be read by the program committee and should not be the final paper. As with all writing, it should be written with its audience in mind. An author who has not served on the committee may not have a clear conception of the refereeing process and what effect that has on the acceptance of papers.

The papers start arriving the beginning of August and the committee meets the second week of September (those were the dates this year). There are nine referees each of whom tries to read them all or asks someone they know to read the paper if they can't.

This year there were 165 submissions. If I had nothing else to do, I would have had six weeks to read them. Unfortunately, August was a busy month for me and I did not start working on the POPL papers until August 26. This left me with two and a half weeks to read the papers. (For my previous POPL committee there were only 125 and I was not busy, so I had six weeks.) This worked out to be a schedule of 10 to 15 papers a day, or about half an hour apiece, with breaks in between. While reading the papers I got quite bleary minded and caught myself a number of times confusing two papers.

After the program committee members have read the papers we get together and discuss them. This process normally takes one or two days. Before we meet most of the ratings have been compiled (There's always someone who doesn't send them in quite in time.) At the meeting the most knowledgeable people in a given area are often able to convince the rest, even if they started in the minority.

The two committees' meetings I've been on have worked in somewhat different ways. Both times we accepted between one fifth and one sixth the number of papers. Both times we spent an hour or two rejecting all the papers that were fairly clear rejects (those that could be rejected because most of the referees agreed, and the one or two remaining ones didn't know about the area and didn't feel strongly).

In the first committee we divided the papers by area and designated one expert on the committee to discuss all the papers in the area. This was fairly enjoyable, and took two days. Several papers were re-read over night. On all but about four papers we reached a consensus about whether papers were worthwhile. On those four we had to vote. We decided that twenty-five papers were worthwhile, and then noticed that that was a good number. At the end we asked people if there were any regrets and the four papers were brought up again. The decisions were not changed. I had always voted with the winning side and on several was the

deciding vote. On the way home I changed my mind on several, but by then it was too late.

There was one mistake we made that I know of. One of my colleagues from Yorktown wrote a paper that one of the committee members claimed was the same as a previous piece of work. I didn't believe it, but was not familiar with that previous piece of work and so could not argue the case.

In the second committee we went through the papers more or less in the order they were received. We deviated from that order when two or three were nearly identical or had conflicting claims. This process took only one exhausting day. Because of the larger load, this was not an educational process for the referees. We did not spend enough time on each paper for each referee to be certain that what we were doing was right. In the first committee we had the leisure time to educate the committee members who were not specialists in an area as to why a particular paper was good or bad. In the second committee each referee went along with the majority. We also did not review our votes at the end. I have less confidence in this procedure, although I'm not sure that there are any papers whose fate would have been decided differently. I don't have enough of a feeling about the papers to think of errors that we made, except in my area of expertise, where I think we made a couple. But I think the only serious one falls in the category of the type of paper that is inherently hard to accept at POPL (see below).

The major worry of a referee is rejecting a good paper, or accepting an embarrassing paper. There are two types of good papers that will be rejected: those that are poorly written (when the audience of spaced out referees is considered) and those that are necessarily more than they appear at first reading.

What constitutes a poorly written paper? Let me give some examples:

1. The authors of the paper decide that the referees are all famous scientists who can see implications, motivations, and know the literature, and thus since there is a 10 page limit only include the details of the algorithm and/or proofs and theorems.

2. The author who says this work is similar to another paper (often by the same author appearing in a recent paper), but does not explain the differences.

3. The paper ignores other recent work in the area. Then someone on the committee who notices a resemblance can believe that you are not aware of all the relevent work.

4. The paper by someone with a practical background, who dresses up a fine practical paper with theoretical trappings, and gets a bit of it wrong. Then the referees who would appreciate the practical results get lost and the theoreticians on the committee don't know whether the errors can be patched.

5. The paper that only says what will be accomplished in the full paper, without convincing the committee that they will really be able to deliver.

6. Papers that present only glowing results are likely to be viewed as highly suspect by the referees. I haven't seen this in POPL, but other conferences have had problems with papers that do not present a balanced view of the work.

With any new method there are advantages and disadvantages that must be weighed.

Its hard to avoid the above pitfalls in 10 pages. One thing that a number of authors did was to send in a full paper and circle the parts the referees should read. I liked this approach, although one other member of the committee gave them automatic bottom scores because they were too long. (He didn't stick to that). I may be in the minority on this as another referee thought we should automatically reject papers that were not true extended abstracts. One thing I found very helpful were examples. An author who at the beginning of the paper says "here's something I can do, and there are no published papers that can do this" make the refereeing process much easier. Numbers are also useful. "The techniques in this paper saved x% of the instruction on these benchmarks". Any paper that contains a statement like the above contains new information (possibly not valuable new information).

There is a kind of paper that will be rejected by the committee, some of the time, no matter how well written and no matter how good. These are papers on apparently simple ideas that say approximately that if we combine some simple techniques, then we can construct a programming language that allows programming of a new kind. The problem is that some of these papers cannot be judged in half an hour. An example of this is a hypothetical original PROLOG paper submitted to POPL. I've no idea whether there was an original PROLOG paper or if a logic programming paper was sent to POPL early on or not. My initial reaction, and that of most computer scientist that I've talked to, when originally confronted with a description of PROLOG is that it is a pile of _____. PROLOG has only a few simple ideas, unification and backtracking, and some other things which look like warts -- cut, add axiom and delete axiom. It is only when you try to solve some problems using the paradigm that PROLOG suggests that you appreciate that there is something there.

It took me days or weeks of thinking before I realized that these ideas hang together in an interesting way. (Some people still think that PROLOG is not a good idea, but most agree that POPL should contain interesting research which after exploration by many researchers may still turn out to be a bad idea. And most agree that we are not done exploring PROLOG). The POPL referees don't have the leisure time to sit back and think about examples to prove or disprove the utility of an idea. Even if a minority are convinced that it is an idea worth exploring, they will not have the background to convince the others. Someone will say, "How do you represent things which change over time in PROLOG?" The defender (if he/she is a very quick study) might reply, "Well, you can change axioms." The attacker will respond, "Doesn't that strike you as ugly?" The defender will agree and that will be the end of it. I see no reasonable solution that would allow a paper presenting a radically new way of working to be accepted, unless that way of working were proven better, at least in a small domain.

One of the points of this discussion is to encourage authors whose papers are rejected. The referees are human. They make mistakes. There are other conferences and there are journals. If your paper is rejected and you think it is a fundamentally

good idea, see if it can be re-written. If it is an idea you can do further work on, do so. If you are implementing it, finish and describe what you found. That will make a stronger paper. On the other hand don't take this discussion to mean that you should just submit next year. The exact same paper will most likely not make it on the second go round. Show your paper to someone who has gotten papers into conferences and ask for honest criticism.

I passed the above discussion around to a number of people for comments. Many people suggest that I had not adequately stressed that submittals should not be just the same as full papers. They should be summaries. Less may be more, and more may be less. A simple, powerful idea may get lost in a paper that discusses all its ramifications. If the fundamental ideas can be stated concisely you will be doing yourself and the committee a favor.

Some people took this discussion as an attack on the quality of our deliberations. In some sense it is. We are not perfect, and this discussion acknowledges that. However, I think the members of the program committee do as good a job as can reasonably be expected. As the number of papers submitted to a conference goes up, the quality of the judgements must go down. For example, there are conferences with very large numbers of submittals that establish completely separate subcommittees with the effect that there is no single standard of judgement. The way POPL papers are judged, the more submittals, the less time the referees spend on each. Authors should be aware of the system, and write submissions which are relevant to this type of audience.

A few words may be in order on matters that do not enter the committees deliberations. Papers are not accepted, at least directly, because they come from first rank institutions. It may be harder to get a paper in from a lesser institution because there are fewer colleagues around to give good advice and thus the papers are not as polished. Nor are papers accepted because they are from top people. One might read a little further and more carefully, before deciding to reject a paper by one of the top ten living computer scientists but if it is not up to standards it will be rejected. I'm impressed by some of the names we have rejected.

Papers are not often rejected on the grounds that they are on "compiler construction", or "object oriented programming", etc. and the committee is uninterested in those topics. There is some variation in the committee's interests over the years. But, if one committee is harsh on an area there are fewer submittals the next year. Hence fewer papers on the topic are accepted the second year and a whole discipline can abandon POPL. This is unfortunate. It has happened from time to time with almost every topic POPL is interested in. To find out what disciplines POPL covers, just look at its title, or the call for papers.

I have heard several complaints that the POPL committee has been unfair. There are many ways the POPL committee could be unfair. Since POPLs have rejected 5/6ths of the papers submitted, it becomes easy to believe the committee is being unfair. This has not been my experience. But, since we only accept one sixth of

the papers and if you think POPL is being unfair to you, you will only get contrary evidence one sixth of the time (on average).

I have tried to give an explanation of how the committee functions. In the above I have not given an outline for a paper. One of my colleagues suggested one and added a few comments of his own. He was on the Compiler Construction program committee. I'm not sure his outline applies quite as well to problems which solve outstanding open problems or on the other side to retrospective looks at systems or languages, where you talk about problems that a language has when used to write some programs. You can make variations on this outline, but you ought to have good reasons to do so.

"A common pitfall in writing abstracts is to assume that they are just shortened papers. Abstracts produced this way tend to be inferior to those that are written from scratch to be abstracts. A good formula for writing an abstract is to partition it into three parts.

1. "The introduction. This section should be limited to at most three pages. It should state the claims of the paper and give a few examples of what is new and different about this paper.

2. "The body. This section should be the bulk of the abstract. It should contain enough text to make the committee believe the results claimed in the previous section. Any new algorithms should be stated in detail.

3. "The comparison. The last section should contain the pointers to other relevant work. This section should be terse and is there primarily to inform the committee that you know the basic work in the area. In the final paper this section should be greatly expanded but in the abstract this can normally be less than a page.

"In some papers a fourth section may be required. Some papers require hairy proofs of small results to be believable. The reason for including such proofs is ONLY to protect yourself from having the paper rejected because some referee does not believe some trivial detail. If these proofs are not central to the main result it is perfectly acceptable to factor these off to an appendix that extends past the normal limit. A note should be included saying it should not be necessary for most of the committee to labor through these sections."