Special Insert

Computing Research Association

Best Practices Memo

Evaluating Computer Scientists and Engineers For Promotion and Tenure

The evaluation of computer science and engineering faculty for promotion and tenure has generally followed the dictate "publish or perish," where "publish" has had its standard academic meaning of "publish in archival journals" [*Academic Careers*, 94]. Relying on journal publications as the sole demonstration of scholarly achievement, especially counting such publications to determine whether they exceed a prescribed threshold, ignores significant evidence of accomplishment in computer science and engineering. For example, conference publication is preferred in the field, and computational artifacts — software, chips, etc. — are a tangible means of conveying ideas and insight. Obligating faculty to be evaluated by this traditional standard handicaps their careers, and indirectly harms the field.

This document describes appropriate evidence of academic achievement in computer science and engineering.

Computer Science and Engineering — Structure of The Field

Computation is synthetic in the sense that many of the phenomena computer scientists and engineers study are created by humans rather than occurring naturally in the physical world. As Professor Fred Brooks of the University of North Carolina, Chapel Hill observed [*Academic Careers*, 94, p. 35],

When one discovers a fact about nature, it is a contribution *per se,* no matter how small. Since anyone can create something new [in a synthetic field], that alone does not establish a contribution. Rather, one must show that the creation is better.

Accordingly, research in computer science and engineering is largely devoted to establishing the "better" property.

The computer science and engineering field in academe is composed of faculty who apply one of two basic research paradigms: theory or experimentation. Generalizing, theoreticians tend to conduct research that resembles mathematics. The phenomena are abstract, and the intellectual contribution is usually expressed in the form of theorems with proofs. Though conference publication is highly regarded in the theoretical community, there is a long tradition of completing, revising, and extending conference papers for submission and publication in archival journals. Accordingly, faculty who pursue theoretical work are often more easily evaluated by traditional academic mechanisms. Nevertheless, the discussion below regarding "impact" will apply to theoretical work, too.

As a second generalization, experimentalists tend to conduct research that involves creating computational artifacts and assessing them. The ideas are embodied in the artifact, which could be a chip, circuit, computer, network, software, robot, etc. Artifacts can be compared to lab apparatus in other physical sciences or engineering in that they are a medium of experimentation. Unlike lab apparatus, however, computational artifacts embody the idea or concept as well as being a means to measure or observe it. Researchers test and measure the performance of the artifacts, evaluating their effectiveness at solving the target problem. A key research tradition is to share artifacts with other researchers to the greatest extent possible. Allowing one's colleagues to examine and use one's creation is a more intimate way of conveying one's ideas than journal publishing, and is seen to be more effective. For experimentalists conference publication is preferred to journal publication, and the premier conferences are generally more selective than the premier journals [*Academic Careers*, 94]. In these and other ways experimental research is at variance with conventional academic publication traditions.

The reason conference publication is preferred to journal publication, at least for experimentalists, is the shorter time to print (7 months *vs* 1-2 years), the opportunity to describe the work before one's peers at a public presentation, and the more complete level of review (4-5 evaluations per paper compared to 2-3 for an archival journal) [*Academic Careers*, 94]. Publication in the prestige conferences is inferior to the prestige journals only in having significant page limitations and little time to polish the paper. In those dimensions that count most, conferences are superior.

Impact — The Criterion for Success

Brooks noted that researchers in a synthetic field must establish that their creation is better. "Better" can mean many things including "solves a problem in less time," "solves a larger class of problems," "is more efficient of resources," "is more expressive by some criterion," "is more visually appealing in the case of graphics," "presents a totally new capability," etc. A key point about this type of research is that the "better" property is not simply an observation. Rather, the research will postulate that a new idea — a mechanism, process, algorithm, representation, protocol, data structure, methodology, language, optimization or simplification, model, etc. — will lead to a "better" result. For researchers in the field, making the connection between the idea and the improvement is as important as quantifying how much the improvement is. The contribution is the idea, and is generally a component of a larger computational system.

The fundamental basis for academic achievement is the impact of one's ideas and scholarship on the field. What group is affected and the form of the impact can vary considerably. Often the beneficiaries of research are other researchers. The contribution may be used directly or be the foundation for some other artifact, it may change how others conduct their research, it may affect the questions they ask or the topics they choose to study, etc. It may even indicate the impossibility of certain goals and kill off lines of research. Clearly, it is not so much the number of researchers that are affected as it is how fundamentally it influences their work. Users are another group that might feel the impact of research.

For the purposes of evaluating a faculty member for promotion or tenure, there are two critical objectives of an evaluation:

Special Insert

(a) Establish a connection between a faculty member's intellectual contribution and the benefits claimed for it, and

(b) Determine the magnitude and significance of the impact.

Both aspects can be documented, but it is more complicated than simply counting archival publications.

Assessing Impact

Standard publication seeks to validate the two objectives indirectly, arguing that the editor and reviewers of the publication must be satisfied that the claims of novelty and ownership are true, and that the significance is high enough to meet the journal's standards. There is obvious justification for this view, and so standard publication is an acceptable, albeit indirect, means of assessing impact. But it can be challenged on two counts. First, the same rationale can be applied to conference proceedings provided they are as carefully reviewed as the prestige conferences are in the computer science and engineering field. Second the measure of the impact is embodied in the quality of the publication, i.e. if the publication's standards are high then the significance is presumed to be high. Not all papers in high quality publications are of great significance, and high quality papers can appear in lower quality venues. Publication's indirect approach to assessing impact implies that it is useful, but not definitive.

The primary direct means of assessing impact — to document items (a) and (b) above — is by letters of evaluation from peers. Peers understand the contribution as well as its significance. Though some institutions demand that peer letter writers be selected to maximize the peer's stature in the field, e.g. membership in the National Academy, a more rational basis should be used.

From the point of view of documenting item (a), the connection between the faculty member's contribution and its effects, evaluators may be selected from the faculty member's collaborators, competitors, industrial colleagues, users, etc. so that they will have the sharpest knowledge about the contribution and its impact. If an artifact is involved, it is expected that the letter writers are familiar with it, as well as with the candidate's publication record. These writers may be biased, of course, but this is a cost of collecting primary data. The promotion and tenure committee will have to take bias into consideration, perhaps seeking additional advice.

The letter writers need to be familiar with the artifact as well as the publications. The artifact is a self-describing embodiment of the ideas. Though publications are necessary for the obvious reasons — highlighting the contribution, relating the ideas to previous work, presenting measurements and experimental results, etc. — the artifact encapsulates information that cannot be captured on paper. Most artifacts "run," allowing evaluators to acquire dynamic information. Further, most artifacts are so complex that it is impossible to explain all of their characteristics; it is better to observe them. Artifacts, being essential to the research enterprise, are essential to its evaluation, too.

Some schools prohibit letters of evaluation from writers not having an academic affiliation. This can be a serious handicap to experimental computer scientists and engineers because some of the field's best researchers work at industrial research labs and occasionally advanced development centers. Academic-industry collaborations occur regularly based on common interests and the advantage that a company's resources can bring to the implementation of a complex artifact. Letters from these researchers are no less informed, thoughtful, or insightful because the writer's return address is a company.

In terms of assessing item (b) the significance of impact, the letter writers will generally address its significance, but quantitative data will often be offered as well. Examples include the number of downloads of a (software) artifact, number of users, number of hits on a Web page, etc. Such measures can be sound indicators of significance and influence, especially if they indicate that peers use the research, but popularity is not equivalent to impact.

Specifically, it is possible to write a valuable, widely used piece of software inducing a large number of downloads and not make any academically significant contribution. Developers at IBM, Microsoft, Sun, etc. do this every day. In such cases the software is literally new, as might be expected in a synthetic field, but it has been created within the known state-of-the-art. It is not "better" by embodying new ideas or techniques, as Brooks requires. It may be improved, but anyone "schooled in the art" would achieve similar results.

Quantitative data may not imply all that is claimed for it, and it can be manipulated. Downloads do not imply that the software is actually being used, nor do Web hits imply interest. There are techniques, such as the Googol page-rank approach [http://www.google.com], that may produce objective information about Web usage, for example, but caution in using numbers is always advised.

Summary

Computer science and engineering is a synthetic field in which creating something new is only part of the problem; the creation must also be shown to be "better." Though standard publication is one indicator of academic achievement, other forms of publication, specifically conference publication, and the dissemination of artifacts also transmit ideas. Conference publication is both rigorous and prestigious. Assessing artifacts requires evaluation from knowledgeable peers. Quantitative measures of impact are possible, but they may not tell the implied story.

References

Academic Careers for Experimental Computer Scientists and Engineers, 1994, National Academy Press

Googol Page Rank System

Approved by the
Computing Research Association
Board of Directors
August 1999

Prepared by:
David Patterson (University of California,Berkeley)
Lawrence Snyder (University of Washington)
Jeffrey Ullman (Stanford University)