TODS Perceptions and Misconceptions

Richard Snodgrass

rts@cs.arizona.edu

Ling Liu asked that I write a quarterly column on the ACM *Transactions on Database Systems* (*TODS*). I welcomed this opportunity to engage in a dialog with the database community. *TODS*' raison d'être is to disseminate scientific results and as such is a partnership between authors and readers, mediated by ACM publications staff and volunteers, primarily the *TODS* Editorial Board, the *TODS* Information Director (Curtis Dyreson) and reviewers worldwide, and propagated by libraries and companies via institutional subscriptions, by individuals via their subscriptions, and by SIGMOD via the SIGMOD Anthology and DiSC.

I came on board as Editor-in-Chief a year ago, inheriting a strong and responsive journal expertly led by Won Kim. This is a time of pivotal change within the STM (Scientific/Technical/Medical) publishing enterprise.

- As with other areas, the web is also transforming publishing, enabling and encouraging a structural move from print subscriptions sent by postal mail to electronic subscriptions accessible via the internet.
- Mergers between STM publishers are replacing many small, independent publishers serving niche markets with just a few behemoths yielding great power.
- A revolt is growing among libraries and readers against what is viewed by many as excessive subscription rates and excessive increases in rates.
- A conflict is escalating between those (primarily readers) who believe in the free availability of scientific literature and those (primarily commercial publishers) who want greater protection of their digital rights, and thus less accessibility.
- New technology available now or on the horizon may initiate even more jarring changes: precipitously declining storage costs, ebooks, medium bandwidth wireless internet connectivity, tablet computers, electronic paper, shops offering on-demand printing from the web.

These changes have and will continue to impact *TODS*. I will track these changes in subsequent columns.

In this introductory column, I briefly remark on some of the prevalent perceptions and misconceptions about *TODS*. As will many interesting subjects, the misconceptions outnumber the valid perceptions.

Perceptions

I've encountered some perceptions concerning *TODS* which I feel continue to be valid.

TODS is of high quality.

> Summarizing a citation analysis of database literature, considering over 100,000 citations, the web page `http://www.acm.org/sigmod/dblp/db/about/top.html` lists the top-cited papers and books. Thirty *TODS* papers appear on this list; 31 papers were from all other journals combined. *TODS* also dominated all conferences. *TODS* fares similarly well in an summary of estimated impact from the Research Index database (`http://citeseer.nj.rec.com/impact.html`), which ranked journals according to their average citation rate. *TODS* was judged the database journal with the most impact, appearing in the top four percent of the 800-odd journals and conferences analyzed.

TODS is accessible.

TODS has thousands of print and electronic subscribers. Of course, the journal (including all past issues) appears in the ACM Digital Library and is thus available to the many individual and institutional DL subscribers. *TODS* is also included in the SIGMOD *Anthology* and the annual *Digital Symposium Collection* CDROM publications, 5000 copies of which have been sent all over the world. These disparate media (print, web, CDROM, DVDROM), widely distributed, ensure that *TODS* articles are easily available to database researchers.

TODS papers are too long.

Let's look at the past twenty five years of *TODS*. The average article length has more than doubled, from 19.2 pages in 1976 to 41.9 pages this year. The average article last year was longer than the longest article in 1976. The shortest article last year, at 31 pages, was longer than the average article for the entire first decade of *TODS*' existence. In five separate years an article of at least 60 pages appeared (one weighed in at a whopping 79 pages). Details are available in an editorial in the December 2002 issue of *TODS*.

The result is that readers are confronted with less diverse and more ponderous papers in each issue, of concern to the Editorial Board. So we recently changed *TODS* policy to encourage "shorter submissions, including even very short (say, five page) submissions. The primary criterion for acceptance is improving on the state-of-the-art in some significant way." Indeed, it is probable that a soon-to-appear issue will contain a paper shorter than ten pages. Our hope is that we can return to those halcyon days when the average was closer to 20 pages.

Misconceptions

I've also repeatedly overheard or been told things about *TODS* that are not true, or no longer hold.

It takes forever to get a paper reviewed by TODS.

I write this in mid-July 2002. I personally have a few papers currently in submission at journals (not to be identified, but not *TODS*), including one submitted last January, seven months ago, and one submitted in June 2001, 13 (!!) months ago. In contrast, the average turnaround time for a *TODS* submission is around five months, with a maximum hovering around six months. Except for a few papers that were submitted in December 2001 that are nearing a decision, all of the papers submitted to *TODS* before February 2002 have now been processed.

In comparison, papers submitted to the SIGMOD *conference* require about three months to review, only about two months shorter than the average *TODS* submission. So while anecdotal evidence indicates that other database journals are very slow, not so for *TODS*, which is approaching conference speed in reviewing.

TODS publishes only theory papers.

The June 2002 issue of *TODS* includes a superb paper by Nicholas Bruno, Surajit Chaudhuri and Luis Gravano on top-k selection queries, which have the potential to be more efficiently evaluated than queries that return all results. This paper shows how to exploit statistics already maintained by a relational DBMS to map such queries to efficient range queries. This paper has a detailed empirical evaluation using a real DBMS. This paper contains not a single theorem; I doubt anyone would consider it a theory paper. However, it *does* contain several lemmas and a careful mathematical analysis of various distance functions. As such it exemplifies the best kind of *TODS* paper, a systems paper that exhibits a "fusion of theory and systems", to "use existing theoretical results and [to point] to possible theoretical research issues." (`http://www.acm.org/tods/Authors.html`).

TODS doesn't publish theory papers. (Misconceptions don't have to be consistent!)

The September 2001 issue of *TODS* includes the very interesting paper by Wilfred Ng proposing an extension of the relational model to incorporate ordered domains. This paper proves five deep theorems using some thirteen lemmas, five propositions, and 27 definitions. Along the way, the paper proposes a complete and sound axiom system for ordered functions dependencies (OFDs) and establishes a set of sound and complete chase rules for OFDs. But the paper also proposes an extension to SQL (DDL and DML) for ordered domains and provides a grammar and many example queries. As such it too exemplifies the best kind of *TODS* paper, a theory paper that considers "applications and implementation considerations/consequences."

TODS publishes only research papers.

This used to be the case. However, a recent innovation is to solicit a new type of paper: focused surveys on topics relevant to *TODS*. These should be deep and will sometimes be quite narrow, but should make a contribution to our understanding of an important area or subarea of databases. *TODS* surveys should be educational to the database audience by presenting a relatively well-established body of database research.

Surveys can summarize prior literature on a theoretical or systems research topic, or can explain approaches implemented in commercial systems. A survey of the former type summarizes a literature on a particular subject, presenting a new way of understanding how the papers in this literature fit together. A survey of the latter type summarizes the best industrial art, and can be acceptable even if it represents no new contribution over what has been used in industry for years, if the paper's content is not to be found in the published literature. (Sometimes such practice is ahead of the published research; especially in such cases it is important for researchers to know about these engineering advances.) See the editorial that appeared in the September 2001 issue of *TODS* for more details.

By the time they come out, TODS papers are irrelevant.

The March 2002 issue of *TODS* contains three papers, which were originally submitted in August 1997, March 2001, and August 2001. So two of these papers are a year or less old. Submit to SIGMOD in October and your paper appears the following June, an interval not much shorter than *TODS*. There is no longer such a big difference between *TODS* and conferences in terms of relevancy.

One important component of the interval from first submission to appearance is the time needed for production of an accepted paper. *TODS* is in the first group of ACM journals to adopt a new production process. In the past, accepted articles were converted from their submission format, LaTeX in most cases, to SGML for printing, which took considerable time and introduced errors. Now accepted articles are typeset directly from their LaTeX source into PDF, for both the printed version and the version in the ACM Digital Library, provided that the source is prepared using the *TODS* style file (see the author information on the *TODS* web page for more details). This approach results in fewer errors, faster production and thus a more timely publication. Moreover, since authors will know what the final version will look like, fonts and all, they can better fine-tune the appearance of their papers. And a faster production is one part of the effort to ensure relevancy of *TODS* papers.

Request for Feedback

I mentioned at the start that I want this column to be a dialog with the database community. I've summarized some of the changes *TODS* is undergoing. I would be very interested in hearing from you as to what *TODS* can do better in disseminating scientific results. As two examples, the innovations of industrial surveys and short contributions originated in suggestions from readers. If you have a suggestion for how to improve *TODS*, please contact me or anyone on the *TODS* Editorial Board.