

INSTRUCTIONS FOR REVIEWERS -- VLDB 2001

VLDB has built an outstanding reputation as highly selective, high-quality conference. VLDB 2001 aims at further improving the VLDB quality standards concerning the review process.

Each reviewer is expected to immediately download the papers, to read and understand them in their entirety, and to provide high quality reviews. The submitted reviews should be of adequate length and include sufficient details about the paper; in no cases should reports give to submitting authors the impression of lack of care and depth of analysis. High quality reviews will satisfy, for example, the following requirements:

1. When a referee gives a high score on significance or relevance, the referee report must explain why and how exactly the paper contributes to improving the state of the art (as opposed to merely judging on the basis of a hot topic's overall relevance, or not giving any textual explanation at all). Likewise, low scores on significance or relevance need to be explained as well.
2. When a referee criticizes the originality of a paper, the referee report must give concrete pointers to prior work that has solved the issue that is considered in the paper or closely related issues.
3. Scores on technical depth must be substantiated by pointing out the paper's specific methods, techniques, and results on which the score is based.
4. The overall score must be compatible with (at least one of) the scores on relevance, significance, originality, technical depth, and presentation (not necessarily an average of these sub-scores, however).

Each reviewer is expected to be aware of the **BROADENING FACTOR** and of the **SPECIFICITY FACTOR**, defined below. Broadening and Specificity Factors need to be entered in the review form and will be taken into account during the paper selection process, although not in a formal and predefined way (e.g., by quotas).

All referee reports must be entered in the PC support database 3 (three) weeks before the PC meeting. This schedule will allow for a two-week consolidation round of electronic discussion (via e-mail or some other electronic forum, and possible revision of referee reports) of controversially rated papers. If during the electronic discussion it appears that there is not enough information on a given paper, then (based on his own initiative or upon explicit request of any reviewer) the PC chair can request another independent review.

In very specific cases, where controversies cannot be reconciled and there are critical issues to be clarified, PC chairs may contact the authors of a paper for clarification during the consolidation round. This may involve exposing specific parts of the reviewers' comments to the authors, but without exposing scores and the overall assessment. Such provision should apply to a marginal fraction of the papers and should be considered as experimental.

All referee reports should be consolidated (to the extent that controversies can be reconciled by electronic discussion) 1 (one) week before the PC meeting. The PC support database should then be frozen and may be made accessible to all PC members (with the exception of papers co-authored by PC members), so that all PC members can familiarize themselves with all referee reports before the PC meeting.

Duplicate submissions

Duplicate submissions are not allowed for VLDB conferences. A VLDB conference submission is considered to be a duplicate submission if there is another paper with all of the following properties:

1. the paper and VLDB submission have at least one author in common
2. the paper is more than 4 pages long, when formatted for the VLDB Proceedings
3. the main technical content of the paper substantially overlaps that of the VLDB submission
4. the paper is published or under consideration to be published in a refereed journal or proceedings (electronic or printed) that is generally available (e.g., not limited to conference attendees).

Authors submitting papers to VLDB conferences are expected to agree to the following terms:

"I understand that the paper being submitted must not contain substantial overlap with any other paper currently submitted elsewhere. Furthermore, previously published papers with any overlap must be cited prominently in this submission. "

Reviewers are expected to check that the papers cited by authors as overlapping with their submission cannot be considered as duplicates, based on the above properties.

Broadening Factor

The VLDB encourages broadening by soliciting papers that are (1) on a broader range of topics than those considered by previous database conferences; (2) on riskier and more novel challenges, as opposed to incremental improvements on existing results; (3) from a broader range of contributors (e.g., from across the spectrum of developing and deploying database technology and from those outside the field who pose new requirements and challenges); and (4) in novel formats such as reports on case studies, systems development and testing, and product evaluations relative to new application requirements.

A VLDB submission is considered to contribute to broadening the database field if it addresses issues beyond conventional database topics and technology. It must contribute to expanding database technology or methods beyond conventional databases and applications to the full scope of data management—its applicability, its challenges, and its future directions. The program committee evaluates the degree of broadening on a three-point scale.

- **Strong:** Papers that produce new, substantial results related to issues that clearly contribute to broadening data management beyond core topics (e.g., address new, advanced application challenges or broadening issues such as the impacts of data management on new applications or business processes, or vice versa).
- **Modest:** Papers that explicitly address and contribute to a broadening topic or issue or apply exciting or new solutions to a topic or issue that is beyond core data management.
- **Little or none:** Applies to most VLDB papers in the past, which have generally dealt with core data management technologies or topics. Papers may mention or reference applications requirements or contexts or broadening issues but not contribute to the broadening objectives, as defined for VLDB2001.

Specificity Factor

Many have noted that the papers appearing in recent database conferences such as VLDB and SIGMOD are getting more and more specific. Ten years ago, there were papers introducing new data models, query languages, and conceptual notations; one encounters few such papers today. This is in some way inevitable, as the field matures and as a literature and set of accepted concepts and paradigms becomes prevalent. Therein also lies the danger that the field is becoming ossified,

that papers counter to the prevailing wisdom are rejected in favor of thorough studies of narrow topics of interest to only a few people.

The [Asilomar report](#) labels these latter papers "delta-X" papers, and recommends a radical approach of going to poster sessions and invited papers for conferences. We do not agree with such a disruptive strategy, preferring instead a more evolutionary approach that encourages broader papers. The Spf is a rating that is a single digit, with a larger Spf indicating the paper is more specific, and thus appeals to a smaller portion of the community. While more specific papers should not be rejected out of hand, they need to be particularly compelling to be selected over less specific alternative papers.

The Spf is designed to be determined quickly, from a paper's title and abstract. Studies will be needed to determine whether the Spf is a well-defined metric; for this reason, we will not include it in the overall ranking computed for each paper, but will instead just have it available as one of the many considerations kept in mind when the paper is evaluated, both by the individual program committee members and during the final discussion at the program committee meeting. Note that "specificity" can be considered the inverse of "broadening". Thus, the specificity factor is another attempt to encourage broadening, by discouraging overly narrow papers. We view this as an experiment and will attempt to determine post facto whether the three-point broadening factor or the nine-point specificity factor, or some combination, is best.

Very generally, 1 (one) is added to the Spf for each significant reduction in the paper's domain of applicability. To provide an absolute scale, we list some rough guidelines. The guidelines are incomplete, and are intended only to be illustrative. Each successive level assumes everything in the previous levels fixed and known.

0. The paper introduces a new benefit to humanity (after all, that is ultimately why we are in this business).
1. The paper introduces a new means to effect a known benefit to humanity.
2. The paper introduces a new class of software to implement a known means to help humanity. This software manages data in some way, but the paper itself is not particular to a data model or query language.
3. The paper introduces a new or altered data model to support an existing class of software.
4. The paper introduces a new or improved query language or design notation or conceptual modeling technique for an existing data model.
5. The paper introduces a new or improved query optimization or evaluation technique to support an existing query language, or a new construct for an existing conceptual modeling technique.
6. The paper introduces a new or improved input to an existing query optimization or evaluation technique, or a new or improved way to determine the configuration of an existing conceptual modeling construct.
7. The paper introduces a new or improved way to calculate or estimate or tune an existing input to an existing query optimization technique.

One determines the Spf by first identifying where the paper falls vis-a-vis this range from 0 to 7, then adding one unit for each significant restriction on applicability (such as a paper applying to only one or a few operators of a query language) or not including a major part of the space (such as working only on select-project-join queries).

As examples, here are nonsensical (we hope!) titles at some of the Spf levels.

3 *The "Hysterical" Data Model to Support Time-varying but Space-constant Data*

4 *The Hyperbolic Query Language to Support the Hysterical Data Model*

5 New Space-constant Optimizations for the Hyperbolic Query Language

6 Circular Histograms for Use in Space-constant Optimizers

7 Fast Reconstruction of Circular Histograms

5+1=6 New Space-constant Optimizations for Aggregates in the Hyperbolic Query Language

5+1+1=7 New Space-constant Optimizations for Correlated Subqueries in the Hyperbolic Query Language on Shared-Nothing Multiprocessors

7+1+1=9 Tuning Circular Histograms for Use with Non-Materialized Views in a Low-Memory Environment

This last paper has the following significant restrictions in its domain of applicability:

- Applies only to applications managing time-varying but space-constant data
- Assumes the hysterical data model used for such applications; unclear whether it would apply to other data models
- Assumes the hyperbolic query language for that data model
- Assumes space-constant optimizers for this language; unclear whether it would apply to other optimizer approaches for this language
- Assumes circular histograms for such optimizers
- Considers only the tuning of such histograms
- Considers reconstruction only for use with non-materialized views
- Applies only in a low-memory environment; unclear whether it would apply when memory was prevalent

Given that this paper has a quite high Spf, it had better be pretty exciting to be preferred over, say a paper introducing new optimizations for the hyperbolic query language. The reason that Spf is only a single digit is that it is difficult to imagine a paper with a multi-digit Spf that anyone would want to read, though we're sure we'll be proven wrong some day.

Note also that the number of possible papers goes up exponentially with Spf. There are probably only a few dozen papers legitimately at Spf 2, and perhaps a few hundred papers at Spf 3. But at Spf 7, the number of possible papers, most of which are uninteresting, is mind-boggling.

Our informal experience with papers submitted to previous VLDB's is that most papers have an Spf between 4 and 7, with the particular ones at 7 quite narrow. Also, the title often doesn't reveal major restrictions to the applicability, but the abstract generally does (and should). Sometimes, the Spf increased by a unit when reading the paper, as a major restriction became apparent that wasn't mentioned in the title or abstract, indicating that the abstract and perhaps the title should be changed to make the restriction more explicit.

Our experience also is that many papers with a high Spf are excellent papers according to the accepted criteria. Their proposed approach is often fully described, the empirical studies quite thorough, with a large range of parameters that are varied. This makes sense, as it is easier to be thorough in a narrow domain than in one that is broader and more varied. This is one of the reasons that prototypical papers in VLDB and other high quality conferences have over time evolved into detailed studies of highly specific and well-defined questions, of interest to only a few people.

Paper Classification

Each paper submission must classify the paper along two orthogonal dimensions, the topic area ("Core DB Technology" or "IS Infrastructure") and its category ("Research", "Vision" or

"Application/Experience"). While categories have been used already in past VLDB Conferences, area assignments are introduced mainly for the purpose of gathering summary information about papers – possibly to be used during PC meetings - and of easing the transition to VLDB2002 - where PCs are expected to be thematic (by area) rather than regional.

Reviewers should be aware of the area and category of each paper. In general, the area selection may help PC Chairs in selecting appropriate reviewers for each paper, while the category may influence the evaluation criteria used by each reviewer. Areas and categories are further described below.

Core DB Technology Area

Papers in the *Core Data Technology* area report on technology that is meant to be incorporated in the DB system itself. This includes DB engine functions, such as query languages, data models, query processing, views, integrity constraints, triggers, access methods, and transactions -- in centralized, distributed, replicated, parallel, mobile, and wireless environments. It also includes extended data types, such as multimedia, spatial, and temporal data, and system engineering issues, such as performance, high availability, security, manageability, and ease-of-use.

Information Systems Infrastructures Area

Papers in the *Information Systems Infrastructures* area report on methods, issues, and problems faced during the design, development and deployment of innovative solutions for information management. Examples include digital libraries, E-commerce, scientific and engineering systems, computer-supported cooperative work, federated DBs, data warehousing and other types of enterprise knowledge management. IS Infrastructure also covers middleware and tools that exploit DB technology but are typically not part of a DB system itself. Examples include workflow, TP monitors, application servers, services in support of E-commerce, mediators and other web-oriented data facilities, meta-data repositories, data and process modeling, user interfaces and data visualization, data translation and migration, data cleaning, multi-agent systems, mediators, and system management. The above examples are meant to be suggestive. Papers on other topics that fit the conference's goals are welcome.

Research Category

Papers in the *Research* category contain significant and original research results. Papers reporting and evaluating new architectures, structures, methods, systems and models are encouraged, particularly where these are carefully validated against realistic criteria, such as typical test loads and data volumes. Papers should develop and demonstrate the value of new ideas that are relevant to the future development of databases and their applications. They may make their contribution by developing new implementation methods, new design techniques, new theoretical results, new algorithms, or in any other way that makes databases more powerful, more effective, more widely applicable or easier to use. Developments in specialized databases or applications are also of interest to the VLDB community.

Vision Category

Papers in the *Vision* category provide a medium for discussion of expected technological, economic and social developments and their impact on databases. These papers are inevitably speculative; however they are expected to present clearly a scientifically and technically convincing argument of relevance to the database community.

They are normally written by authors very experienced in the database industry or in database research, or else with some other deep experience which they bring to bear on database issues. Papers in this category help to formulate directions in which the database industry or database research should develop. Insight and perspective are particularly highly valued.

Application/Experience Category

Papers in the *Application/Experience* category provide a forum for discussion of experiences in implementing DBMSs or in applying DBMS technology in challenging situations. Reports of industrial experience of using databases are particularly welcome. Papers in this category are written by authors who have been involved in building major Database Management Systems or in managing the use of databases in demanding applications. It is hoped that such papers will have one or both of the following effects:

- Demonstrating the effectiveness of new database methods or technology so that others may appreciate its value and possibly re-use the approach in their own context.
- Demonstrating the deficiencies in our current understanding and technology as a challenge to researchers and implementers.