Project Number: 51-MMX-0501

The Health of Database Research

An Interactive Qualifying Project Report
submitted to the Faculty of
WORCESTER POLYTECHNIC INSTITUTE
in partial fulfillment of the requirements for the
Degree of Bachelor of Science
by

Alin Sirbu

Andrei-Tudor Vasilescu

Date: May 2nd, 2006

dbhealth@wpi.edu

Approved:

Professor Murali Mani, Primary Advisor

Professor Matthew O. Ward, Co-Advisor

1. Abstract

Increasing numbers of database researchers have raised their concerns regarding the “health” of major database conferences. Various problems such as bias in the paper review process and the performance of executive positions are being analyzed. Our team assessed the current state of 3 conferences: SIGMOD, PODS and VLDB through the filter of several “health” parameters by conducting surveys of database researchers. In addition to this, we gathered database information and produced interesting and useful statistics for the community. Finally, we produced this report that includes our conclusions and recommendations to the conference organizers and the new researchers in the database field.

Table of Contents

2. Introduction

2.1 The Subject of This Project

The subject of this project raises the issue of the **Health of Database Research**. For scientists and researchers, conferences are very good venues to present the result of their work. In the database field, researchers should have their papers viewed in the same light as papers in other fields. Given this, the concept of "health" in this case is important to them.

Based on the hypothesis that some communities do not exhibit, in some ways, positive properties such as diversity, quality, quantity, fairness (impartiality) of organization and evaluation of work, this is an attempt to **"measure" the health** of significantly important conferences or journals and **produce clear statistics** from the gathered evidence and a **presentable situation** on the current "health" of the database research community for people just entering this field or for people that are interested and want to get involved in reforming the current system. Our aim is to provide statistics that we believe are interesting, and which have not been done to date. Therefore, we expect our statistical results to be useful to the database research community.

2.2 The Goals and Sub-Goals

Specifically, the goals of this project are to gather electronic and human data from sources within database conferences such as SIGMOD (**S**pecial **I**nterest **G**roup's **M**anagement **O**f **D**ata), PODS (**P**rinciples **O**f **D**atabase **S**ystems) and VLDB (**V**ery **L**arge **D**ata **B**ases) and study the variation of participants, submitted and accepted papers, and of the positions held by people in a certain conference over several years. We will study the overlap between PCs (**P**rogram **C**ommittee members) and their connection to the previous and present research papers, the trends of certain people in holding many positions at once or the same executive position over many years.

In terms of human data collection, surveys targeted at different individuals have been conducted, looking at each of their perception of the problem at hand. The intended survey takers were members of different Executive Committees of the past database conferences, as well as some people who have not had any significant experience leading them. It included questions regarding different parameters of "heath", as well as demographic and opinion questions that are relevant to this study.

2.3 The Target Audience

The target audience for this IQP is, in fact, any person who would like to enter the research field and would like to get some guidance on what is actually happening in the current world of paper publication. It may also be useful for learning what to expect from conferences and journals, and get exposed with the reality of rising paper numbers, stiffer competition and low acceptance rates. As well as newcomers, current and past conference organizers have data, opinions, statistics and recommendations available to them. If mined data will not convince him/her, then results from actual human surveys will shed a more convincing light on the current problems that the database research community is dealing with. This project's outcome will also be useful to people who would like to study the "health of research" and contribute with solutions and experiments of their own. Sister communities might benefit from the data and new organizations can look at what we have done to see what to expect in the future when they will grow and paper reviewing becomes harder, and the actual process more difficult to control.

3. Background (Literature Review)

In recent years, several articles about problems in the "health" of Computer Science research communities have been written by individuals that used to be part of some of the conferences that are mentioned to be having issues.

One such paper, published in 2004 calls for a "**rethinking of the conference reviewing process**" [1]. In the first part of the article, the (panel) authors list some important points leading to the discontent in the presumably "broken" process. They blame the dramatic increase in the Program Committee size along with the number of submitted papers, the impersonal contact between paper authors and their reviewers, the lack of motivation of a Program Committee member and the difficulty of assessing papers in domains that they are not experts in.

Like many articles, this one gives us a list of possible solutions to the abovementioned problems such as reducing the size of the committees, introducing more hierarchy, and phases in reviewing a certain paper.

In another article, Dr. David A. Patterson (the president of the ACM) tackles two big problems, the Health of Computer Science Research, in addition to the "Dearth of Big Idea Papers" [2]. By showing charts of different conference statistics (including one big Database Conference called SIGMOD), he immediately expresses the concern of increasing number of paper submissions, increasing workload of Program Committee members, the rate decrease in the acceptance of new paper, with a focus on Big Idea Papers. Because this type of papers usually gets a poor rating and such a number lands on a PC's review list, a paper with a revolutionary idea is likely to be discarded.

What is different about this article is that it actually proposes an experiment that conference leaders could and should try: to select a distinct Program Committee that meets and reviews the Big Idea Papers exclusively. He also proposes a session of three such papers to be taken into consideration. However, he is aware of the fact that visible

and analyzable results will show after only several years into the experiment and whether this will be a successful thing. Similarly, the OOPSLA conference has developed its own session that takes care of these papers, called Onward! It is the place "to reveal the revolutionary, air the provocative, and expose the subversive" [3]; it is composed of three sub-tracks: papers, presentations, and films.

Back to the article in discussion, the last part presents three perspectives from which the increasing popularity of conferences can cause trouble: **Impact on the Program Committee**, **Impact on the Conference and Field**, **and Impact on the Authors of Rejected Papers**. Regarding the first item, Dr. Patterson is concerned about the size of the committee which is not able to have good conversations on a single paper due to the high volume of submissions. The second argument makes the point that conferences that get a high inflow of submissions, but tend to accept about the same number of papers as some time ago, will miss out on many good papers just because of this practice. The third perspective, the author's perspective, is the one that we will touch upon when conducting our surveys: the frustration of someone whose paper, probably a "good paper", got rejected. The person will not know the actual reason for the rejection (may it be arbitrary, political, or of some other nature). This makes authors write more papers and send them to the same conference, or send the same paper to more of them, to increase their chances of getting something accepted. The strategy thus results in the increase that we have mentioned and could cause real problems in the future, unless a proper research project in the matter is conducted.

As efforts in tackling and/or remedying this situation, ACM's president mentions two interesting initiatives: SIGMOD's collaboration with two big database conferences in reviewing the same paper in one conference after being rejected in another one, with the optional revisions that this author can make; the other one is that a so-called "task-force" has been created with the help of Dr. Patterson and the SIG (Special Interest Group) Governing Board to investigate the often-mentioned issue of rising submissions and trying to document possible outcomes and solutions.

A third article follows the one published in 2004 by a panel of high-position-members, and explains the results of the experiment conducted in the conferences of 2005. The name is "**Database Publication Practices**" [4] and it has been made possible by approximately the same panel as in [1], but with a few changes. This time, we have information about a series of groups interested in reviewing the research publication process, such as the VLDB Board of Trustees, the aforementioned SIGMOD task force, "the SIGMOD Executive Committee, the SIGMOD Advisory Board, journal editors and a plenary panel at SIGMOD 2004".

It has been acknowledged that the quality of reviews has been dropping, and that many good papers are lost "in the noise". The SIGMOD 2004 panel suggested, between others, to "break the pipeline between conferences, by moving deadlines so that rejected papers could not be immediately resubmitted. This would encourage authors to submit more polished work and choose the most appropriate forum" – [8]. Other methods, such as web tools, more reviews on the same paper, and two-tier Program Committees have been suggested as well. Remarkable was also the suggestion of somehow blurring the line between journals and conferences.

A more practical approach has been taken at SIGMOD 2005, where a partitioning of the Program Committee has been tried (9 groups covering different areas) with more communication and a mandatory meeting where everyone showed up. Authors could contact their reviewers and more explanations were given. However, the acceptance percentage was maintained at "the usual 15%".

This article also mentions the VLDB-SIGMOD resubmission process (some papers not accepted at one conference will be able to be transferred to the next one for reviewing), and also some Journal statistics – ToDS (Transactions on Database Systems) and the VLDB Journal.

4. Procedure

The aim of this section is to describe the work done by the IQP team. In order to provide a basis for our results concerning the integrity of the reviewing process, we had to accumulate data from a variety of sources, bring it into a suitable format and then make useful analyses.

The project has involved gathering a large quantity of data. The main purpose is to try to get as close as possible to drawing an unbiased conclusion by correlating multiple sources of information. Two types of data have been gathered:

- Electronic data – the conference organization committees and accepted papers from the three conferences that we focused on: SIGMOD, PODS, VLDB. Such data is manipulated and kept in a database and spans a period of 10 years of conferences. The sources used in building this repository were the Internet home pages of these conferences (See Bibliography section).
- Human-generated data – the results of our survey. The main objective of the survey was to capture points of view of the people that are currently, and will probably continue to be, involved in these conferences. Paralleling the idea that software is made to please the client, these opinions will be useful in understanding what organizers and participants think about the aspects that need to be modified in the current organizational structure of the conferences and the reasons for doing so.

  The methods of amassing this data are described more in Section 5 of this report.

Once a considerable amount of data had been gathered, first attempts at getting some form of analytical results pointed out one other task that required immediate attention. As data came from various sources, from all the three conferences and 10 distinct years, information was not uniform across all our database entries. As an example, people could have their names spelled in different ways, abbreviated, middle names omitted. The institutions that these people were affiliated with had the same

pattern of variance in naming and also changed from one year to another as people made new steps in their careers.

In order to produce even the most basic of statistics, i.e. counting the number of papers a person had published, the data from the organizing committees and the paper submissions had to **uniformized**. This term refers to having the same name for every individual in both sections of the database. Endowed with such information, we are able to follow a person's evolution in the database research environment.

At the moment when the correlation between the committee members and paper submitters was reliable, we were able to draw some conclusions about the situation in these conferences. Numerical, or blind, results are shown in Section 6, while Section 7 includes some comments and charts supported by the presented numbers.

In the final part of our project, we attempted to create some graphical means of displaying the information we consider relevant. It will be for future readers, to understand our conclusions while also being provided with graphs that support those results.

5. Data Collection

5.1 Conference Organization Tracker

In this short section we try to describe the paths that we followed in gathering the data. There is no secret resource that had already collected this data and made it available somewhere else, so, we had to start by reviewing the conference e-proceedings sites and take information from there directly. Following, we describe one possible method for taking the data given on an HTML webpage and transferring it to an SQL database.

Starting point. The formats used to display data differ among conferences and among different years of the same conference as these web pages were created in a chronological order and the format has changed through time. For example: in SIGMOD, the format for displaying conference organization committee members is person's name (first name, middle names, last name), the place where that person works (university, company in business industry). Note also that in the case where external referees are made available their affiliation is not listed.

Using small Java applications in conjunction with a word processor and Excel. The information on the website is copied and pasted into a text file through the use of a word processor. To continue, we have written three small Java applications to help in processing large document files in order to be able to fill in an Excel sheet. From the Excel sheet we will get the whole information into a word processor, replace the end of cell symbols with the splitter symbol the SQL loader looks for. Now, the data can be loaded into the tables.

- *The NameSplitter* – handles person participation files of format <first names, last name, affiliation> where there is a space between each two different fields and the firstName is given as a continuous string. **Benefits**: The user takes advantage of an option value that can be set to 1, 2 or 3, to get the column information for the three fields that will be put into the Excel sheet. **Disadvantages**: The user has to preprocess the file to ensure that the firstNames

do not contain any spaces between them. Example: " John A. Williams" will be transformed into "JohnA. Williams". The user will have to go again in and replace these spaces in the Excel files.

- *The ListSplitter* – handles paper participation files. It follows a single format, paper name and list of authors on the next line, separated by commas. The output will be lines of first name, last name and paper presented. This can be given as input to the NameSplitter (where the affiliation information is replaced with paper title). **Benefits**: Creates a file that can be easily used to get the columns through the NameSplitter. It also is enhanced to make sure that in the firstName all spaces are automatically removed.
- *The PaperUniversity Splitter* – handles files that have the following type of lines: first line contains the paper's name and on the next lines there are authors (full names) and their affiliation, the two being separated by the string " – ". **Benefits**: This is how most papers are represented on these websites. There are two options. The value of 1 corresponds to getting such a file and producing one without the affiliation information (will run the ListSplitter further on). Also, it ensures that there are no spaces in the format of the first name. The value of 2 corresponds to discarding paper information and supplying an output in the form needed by the NameSplitter. Also, the PaperUniversity Splitter, under option 2, will ensure that the first names do not contain spaces, so that the produced file can be used by NameSplitter to get column information.

Result: We have gathered **1160 committee members** from SIGMOD, PODS and VLDB from years 1997-2006 (10 years span) and added all of them to our Oracle database for the uniformization process.

5.2 Research Paper Tracker

Using the bibliography of several major database conferences (http://www.informatik.uni-trier.de/~ley/db/ - The DBLP) that contains an anthology of all accepted research papers, we were able to gather complete Excel data sheets of paper information (including First Name, Last Name of authors, Title of paper, Section of the Conference). These have also been entered into our database for query processing and statistics.

The main tool in retrieving the research paper information has been a Java program that parses entire pages of text and separates authors with their respective pieces of work, accompanied with their paper title and the section where it was presented. The Java program accesses the text from regular files, but we copy-pasted the content from the actual HTML pages to separate text files (SIGMOD##.txt, PODS##.txt, VLDB##.txt).

Several methods in the main paper loader Java class deal with processing strings (converting to upper/lower-case, eliminating spaces and other bad characters from both ends) or with database operations (drop a table, add a new paper to the desired matching table.

Example, the addPaper method:

public static void addPaper(String Table, String lname, String fname,
String univ, String title, String section)

This will take the Table name (e.g. "PAPERSP99") and will make a Load statement consisting of the Last Name (lname), First Name (fname), Institution (univ), Paper Title (title) and Conference Section (section). A string called "query" is generated with the command to Oracle and the statement is executed via the method call:

stmt.execute(query);

The program is also capable of extracting content from many pages by repeating the database injection cycle more often. The user can specify an array of strings called "conf[]" that contains as many table names as wanted, provided that the table names match the files on the disk and agree with the naming standard ("Conference##.txt") that is part of our convention.

A typical conference file contains at first the section/track name, after which groups of 3 lines follow until we reach a blank line. The 3 lines contain the authors separated by commas, the paper title and a residual line with the electronic link to the paper. A format very close to the DBLP has been kept so as to require minimal human modifications. As part of the convention, the line with the author names must end with a colon (":") and the paper title must end with a dot (".). These patterns were established by looking at multiple entries of the DBLP archive.

Thus a sample paper entry in the file looks like:

Index Structures ← Section/Track name

Frank Ramsak, Volker Markl, Robert Fenk, Martin Zirkel, Klaus Elhardt, Rudolf Bayer: Integrating the UB-Tree into a Database System Kernel. 263-272 ← Paper title

Electronic Edition BibTeX

Note: everything after the arrows are comments, not appearing in file.

Parsing of papers and their authors is done via the following fields, which are shown initialized. The section name is persistent and does not change until it reaches a new track, delimited always by a blank line. The title name is maintained for all authors of the paper for which we are making entries in the database. Therefore, a paper with 3 authors will have 3 entries in the table, each with different author names, but with the same paper title and section name. This structure is useful for finding overlaps between the other part of the database that deals with individuals that held conference positions and authors of database papers. The sample fields are on the next page.

```
String firstName = "";
```

String lastName = "";

String university = "";

String title = "";

String section = "";

After parsing from these files is done, the information stored in different fields can be used to either inject to the database or produce output for an Excel file that is also used for ease of processing in this project.

In conclusion, we have made use of 2 classes: DBPaper.java and PaperLoader.java. The former defines what a paper is by specifying the fields (the same as the abovementioned strings), and the latter actually does the actions described throughout this subsection. Every paper gets its own instance which is created right before inserting one entry into the array that is used for sorting all the data.

Result: We have gathered **2174 papers** from SIGMOD, PODS and VLDB from years 1996-2005 (10 years span) and added all of them to our Oracle database for the uniformization process. This precedes our efforts to create interesting statistics that tie DB researchers and the papers together.

5.3 Database Uniformization

Necessity: As mentioned previously, there are occurrences of the same individual in the database records but with different affiliated institutions or with abbreviated first and middle names. In order to be able to count the occurrences of a person in the organizing committees of the conferences we are tracking it is important to have a unique record for one individual throughout the database. Moreover, linking the people in the organizing committees with the ones that submit papers to these conferences will benefit from having unique and as complete as possible records of individuals.

Methodology: The records we had gathered so far had been entered into a database under the Oracle server available to us. Though SQL queries could easily point

out all the records belonging to an individual, updating them with SQL statements would have been cumbersome. Accordingly, we decided to output the information gathered from the 30 conference organization committees and use other applications to filter out the noise in records.

Spooling. The data was collected by using the spool command from the SQL environment. To make it easier to be imported further on, the headers and usual formatting of SQL were abandoned with the command:

" set echo off newpage 0 space 0 pagesize 0 feed off head off trimspool on "

Also, the query delivered the information with column information separated by commas or semicolons that can then be easily replaced with other symbols by using any Find/Replace method. The Select statement in these queries was of the form:

" SELECT col1 || ';' || col2 || ';' || col3 "

MS Excel. After replacing the delimiter we mentioned earlier with an end of cell symbol, the data gathered can be easily copy pasted into a MS Excel worksheet. We added a header for the sheet and the data was ready to be modified.

Modification responsibilities: Ensuring that the name of an individual is correct and as complete as possible. Also, the institution an individual is affiliated with should be constant in all records.

Modification concerns: There exist individuals with like names that participate in the same or different conferences. It is hard for us to check that the records we believe to pertain to one person are actually only for that one person or whether there might be two people with the same name.

Method: A combination of sorts, typically by Last and First name, or Institution, Last and First names, proved good views on the data that enabled us to ensure the same first name is used for a person. Also, there were multiple institution entries, with different names. For one individual we have chosen to add to his/her record the institution where he/she was most recently listed. On the matter of institution names, we have chosen to expand the name of the institutions to our best knowledge and to prevent abbreviations (which can be numerous) from being used.

Reinsertion of data. Once the data is uniform it is time for it to be reentered into the appropiate tables in the Oracle database account. We omitted to mention previously that before the data was spooled, unique table identifiers had been added to each record. This was done by adding a column with an implicit value (distinct among the tables) to each relation. The command for doing so would be:

"alter table <TableName> add <ColumnName> <ColumnType> default <Value>"

" Ex.: alter table pods97 add Conference varchar2(4) default 'PODS';"

"alter table pods97 add Year integer default 1997;"

Now that we know the place where each record came from we can recompose the original entries from each of the relations. Sorting by conference and years and selecting all tuples with the same combination of identifiers gives us the contents of one relation. The data load files were recreated with the "corrected" data and the whole database was updated.

***Note*:** On this occasion, the creation and deletion of relations and views, as well as the reinsertion of tuples into the database was streamlined and made available for the whole database, including people and paper tables. Packages of the required SQL and control data files can be obtained by contacting one of the team members.

5.4 Database Presentation – Views

The data we have gathered and manipulated can be viewed in two ways:

1. Numerical data based on one of the two segments: organizing committee members or paper submitters.
2. Numerical data based on correlations we have tried to make between the two mentioned segments.

Continuing, we will include descriptions of the views we have created with the purpose of linking the information in the database. Note that results from both categories will be presented in Section 6 and then analyzed in Section 7.

As mentioned, the focus of this subsection is to present the views that have been written in order to underline some important information that will support our project's hypothesis. Below, you will find a series of view names that we made available and descriptions of their actual behavior.

1. **Keepoff (Only Paper Submitters)** – this view shows the people that have submitted papers to the 3 conferences we track and yet have never participated in the organizing committees for any of them. Information is viewed and can be filtered by the year of the last submission and the threshold of submissions that make a person relevant for this view. The schema of the view output is shown here:

<Last Name>	<First Name>	<Total Submissions>	<Year of Last Submission>

2. **Stuck (Only Conference Organizers)** – this view shows the people that participate in conference organization committees and yet haven't submitted a paper in the past 10 years. Information on them can be viewed and filtered based on the number of participations and the year of the last one. The schema of the view output is shown here:

<Last Name>	<First Name>	<Total Positions>	<Year of Last Submission>

3. **Heat (Simultaneous Submitters and Organizers)** – this view shows the people that have submitted a paper and also participated in an organizing committee at the same conference on the same year. Information can be filtered on the number of occasions of this happening. The schema of the view output is shown here:

<Last Name>	<First Name>	<Occurences>

4. **Recognition (Organizers after being Submissions)** – this view shows the people that have submitted a paper to a conference in one given year and the next year have participated in the organizing committee for that conference. Information can be filtered on the number of occasions of this happening. The schema of the view output is shown here:

<Last Name>	<First Name>	<Occurences>

5. **Unjustified Organizers** – this view notes the difference in overlap between the previous two, giving the number of people having a paper and being in the organizing committee in one year and yet not in the previous one. Information can be filtered on the number of occasions of this happening. The schema of the view output is shown here:

<Last Name>	<First Name>	<Occurences>

6. **Newcomers (Organizers after more Submissions)** – this view shows the number of papers that a person has submitted before they were first selected in an organizing committee based on our 10 year tracking of these conferences. Information can be filtered on the number of papers and the year of the debut in a conference organization committee. The schema of the view output is shown here:

<Last Name>	<First Name>	<Conference Organization Debut Year>	<Submissions Before Debut>

5.5 DB-HEALTH Website and Survey

In order to accomplish the goal of gathering human data regarding the health of database research conferences, we set out to build a small-sized website. Its purpose has been to help us conduct a survey by having all of the visitors register themselves by entering their personal login information into our database and having them answer questions. The results have been stored in HTML format for later analysis and publication.

By creating an account for this project, we could link this website to address www.wpi.edu/~dbhealth/index.html. The website has been written in HTML. Since it had to be a dynamic website to allow registration, survey taking and the gathering of results and feedback, we have used Perl/CGI as our scripting language. The survey itself contains questions generated dynamically depending on the users personal information with regard to conference participation (which ones, positions held).

Details about each script that was written are as follows:

Registration.cgi : We used registration.html to retrieve and process information from users. The CGI script validates and enters information into our database. Requested fields include First Name, Last Name, desired Username and Password, Country, E-mail and Institution. An optional part is the address, which people should include in case they win a prize at our IQP raffle. After that, registrants need to specify which conferences they have attended so far (SIGMOD, PODS, VLDB, or OTHER conference), as well as whether they have held any executive positions within these. One mouse click validates the entered information, including e-mail and checking for duplicate registrations and also creates a new account in the Oracle database for subsequent use, each user having a unique, generated ID. Here is the overall table layout for the login information (table name – LOGIN):

Name	Type
FIRSTNAME	Characters

LASTNAME	Characters
USERNAME	Characters
PASSWORD	Characters
INSTITUTION	Characters
COUNTRY	Characters
EMAIL	Characters
ADLINE1	Characters
ADLINE2	Characters
SIGMOD	Number
PODS	Number
VLDB	Number
OTHER	Number
EXECUTIVE	Number
COMPLETED	Number
ID	Number

Members.cgi : This handles the part of the website called the "members section". It tells the survey taker whether he or she took the survey or not and displays the option of taking it or leaving additional feedback. But before all this can happen, the user is prompted with a login form that asks for the person's username and password.

The script is also responsible for generating the survey page dynamically after retrieving the necessary information from the Oracle database (the Login table). There are two ways to call it, with action "Login" (for login) and "Survey" for survey taking.

As mentioned before, the survey is dynamic and questions are based on user information. Below is a Perl code snippet that shows how question 2 is selected based on executive position information:

```
      if ($data->{EXECUTIVE} eq 1) {
            $question2  = "Having served in one or more leadership
roles (PC/PC Chair/Organizing Committee/Executive Committee) ";
            $question2 .= "in the <B>$confs</B> conferences, how would
you rate the experience?"}
```

```
      else {
            $question2  = "How would you rate the performance of the
various leadership roles (PC/PC Chair/Organizing Committee/Executive
Committee) ";
            $question2 .= "in <B>$confs</B> Database conferences?"}
```

Dbs.cgi : Once the survey has been administered, this script gathers the answers in all the textboxes and radio buttons and converts them into readable data, after which it stores the results in a plain text file in HTML format.

Answers from the HTML form are retrieved as follows:

```
$ans1 = $query->param('ans1');
$ans2 = $query->param('ans2');
Etc...
```

There is code inside the script that ensures some safety of the information and of the system itself, and processes all the input data by removing HTML tags (which are obviously not needed nor allowed) and converting the "tainted variables" into untainted ones. The process is needed due to the security weakness of Perl. Malicious information can always be passed (either involuntary or as a hacker attack) and processes can be run on the host machine. Perl tries to counter this problem by introducing the built-in concept of tainted variables, which cannot be passed to any function, and therefore should protect the system until the information gets untainted and ready to use.

Example of removing HTML tags: **`$name =~ s/</</g;`**

Example of getting untainted variable: **`$ans2 =~ /^([^<]*)$/`**)

Here is how we process the survey form in general: For the questions that involve rating, they convert the selection into a numerical value, and for the ones that do not have anything selected the program defaults to "not rated". A similar approach applies to the long-answer questions where empty textboxes are replaced with "no answer". After all

information processing has taken place successfully, the script connects to the Oracle database and retrieves survey taker information, such as name and institution and checks for the validity of the submission (no multiple or clandestine submissions are allowed) and looks at the date and time to place a stamp on the survey answers.

The system is then ready to output everything into the global results file and appends an HTML answer format chunk to the rest of the answers, separated by an HTML separator (<HR>). After all writing is done, the script confirms the completion of the survey by a success screen and by changing the "Completed" variable in the database table from 0 to 1. That is used to know which users took the survey, not just register. Also, the person is given the invitation to provide any feedback to the system.

Below is some sample code that shows an overall format of the survey data that is kept on disk:

```
#Report all answers
$entry = <<"EndOfText";

<STRONG><FONT size=+2>$name</FONT></STRONG> <I>from</I> <STRONG><FONT
size=+2>$inst</FONT></STRONG> <I>contributed at</I>
<STRONG>$now_string</STRONG> : <BR>

<P>
<STRONG>Answer 1 : </STRONG>$ans1<BR><BR>
<STRONG>Answer 2 : </STRONG>$ans2<BR><BR>
```

Admin.cgi : This script deals with the administrative portion of the website. It provides an interface with the Oracle database information and the survey results, including an SQL query executer (executes a query on-site and displays the result in a table format) and access to previous IQP draft documents. We are not going to include details on HTML code since it is very straightforward and pertains to website writing.

Below is an example of how Perl is used to print results in HTML format:

```
open (IN, "$data_file") or die "Can't open $data_file for reading: $!";
```

```
flock(IN, 1)            or die "Can't get LOCK_SH on $data_file: $!";
while (<IN>) {
    print;
}
close IN                or die "Can't close $data_file: $!";
```

Below is another script snippet that displays information on registered users (Perl/DBI details are omitted):

```
      $ct=1;
      print ("<FONT size=+2><table border=1 align=CENTER>\n");
      print ("<TR><TD>Nr.</TD><TD>First Name</TD><TD>Last
Name</TD><TD>Institution</TD>");
      print ("<TD>Country</TD><TD>E-mail Address</TD><TD
align=CENTER>Completed</TD></TR>\n");
      while ($data = $st->fetchrow_hashref()) {
            if ($data->{COMPLETED} eq 1) {
                  $comp = "<FONT color=RED>YES</FONT>";
            }
            else {
                  $comp = "<B>NO</B>";
            }

            print "<TR><TD align=CENTER>$data->{ID}</TD><TD>
$data->{FIRSTNAME} </TD><TD>$data->{LASTNAME} </TD>\n";
            print "<TD>$data->{INSTITUTION}</TD><TD>
$data->{COUNTRY}</TD><TD>$data->{EMAIL}</TD>\n";
            print "<TD align=CENTER>$comp</TD></TR>\n";
            $ct++;}
      print ("</TABLE></FONT>\n");
```

Feedback.cgi : A simple script is used that follows a similar approach to gathering comments on our system and storing them as HTML in a single text file. In the administrative section, a link takes the administrator to a page where the contents of the feedback file are displayed in the web browser for inspection. A similar link is available

to all survey takers who have successfully completed our survey. That is in fact, its purpose.

The information entered into the textboxes is, again, processed by removing HTML tags and get untainted variables in order to eliminate insecurities of the scripting system. Perl helps us with its built-in techniques and functions.

5.6 DB-HEALTH Questionnaire

The questionnaire of the DB-HEALTH survey with all its question variants is included in Appendix 1 at the end of the project report.

6. Results

This section will mainly include numerical pieces of information regarding the data that we have gathered during this project. They will be presented in two separate subsections: data regarding the database entries on the committee organizations and paper submissions, and, results we got through our survey.

6.1 DB-HEALTH Database Results

As mentioned, the electronic data we gathered has been entered into an Oracle database account that we used throughout the project. Here are some of the details of the collected data:

6.1.1 Organizing Committees Data

As a reminder, our organizing committee data spans all three conferences over a period of ten years – 1997 to 2006.

The conference organization data includes external referee positions for three years of SIGMOD – 1998, 1999 and 2001, and two years of PODS – 2001 and 2002. Including these positions, there are 1160 individuals holding organization positions, totaling a number of 3051 positions required by these conferences during 10 years of activity.

However, since data on external referee positions is not available for all conference years, all statistics that follow don't consider these positions.

- 857 people have been involved in these conferences, occupying a number of 2506 positions
- A SIGMOD organization committee includes, on average, 83 positions (minimum: 56 positions in 1997, maximum: 128 positions in 2004).
- A PODS organization committee includes, on average, 24 positions (minimum: 13 positions in 1997, maximum: 33 positions in 2004).

- A VLDB organization committee includes, on average, 145 positions (minimum: 108 positions in 1997, maximum: 204 positions in 2004).

6.1.2 Accepted Papers Data

As a reminder, our paper acceptance data spans all three conferences over a period of ten years – 1996 to 2005.

There were 2174 papers accepted at these conferences, written by 3020 authors, defining 7112 combinations of paper and author. On average, there are 3 authors for one paper, the minimum and maximum number of authors being 1 and 27, respectively.

Some other data concerning each individual conference:

- SIGMOD accepted an average of 303 papers each year; it registered a minimum of 233 papers in 1996 and maximum of 413 papers in 2005.
- A SIGMOD paper has 3 (3.45) authors on average, with a maximum of 18 authors contributing to one paper.
- PODS accepted an average of 77 papers each year; it registered a minimum of 62 papers in 1996 and maximum of 100 papers in 2005.
- A PODS paper has 3 (2.56) authors on average, with a maximum of 7 authors contributing to one paper.
- VLDB accepted an average of 332 papers each year; it registered a minimum of 161 papers in 1996 and maximum of 520 papers in 2004.
- A VLDB paper has 3 (3.31) authors on average, with a maximum of 27 authors contributing to one paper.

6.1.3 View Results

Following are some results extracted from the output of the views we described in Section 5.4.

Keepoff (Only Paper Publishers) – There are 113 people who have published more than 3 papers, 25 people who have published papers more than 5 times and 7 who

have published more than 7 papers, and yet have never participated in a conference organization committee. A person has written 14 papers at most, and yet that person has not participated in a conference organization committee.

Stuck (Only Conference Organizers) – There are 45 people who have held more than 2 organizing positions and 10 who held office more than 4 times, and yet they have published no papers in the past 10 years in these conferences. A person has had at most 11 conference positions, and yet that person never published a paper in the mentioned time span.

Heat (Simultaneous Publishers and Organizers) – There were 58 people who published a paper and held an organizing position in the same year more than 3 times, and 18 people were in the same situation for more than 5 times. The maximum number of times an individual has published a paper and had an organizing position in the same year and conference is 9 occasions.

Recognition (Organizers after being Submissions) – Numbers for people who have been assigned a conference position after publishing papers in the same conference in the previous year are: 43 people in this category were assigned conference positions more than 4 times and 9 individuals were given such positions more than 7 times. The maximum number of such occasions for an individual is 11.

Unjustified Organizers – These numbers are for people who hold an organizing position and submit a paper in one year at a conference but who haven't published a paper in the previous year: 87 people were in this situation at least once and the maximum number of these situations is 3.

Newcomers (Organizers after more Submissions) – These individuals have published a certain number of papers before they were first assigned with a conference position: 96 people published more than 3 papers before filling a conference position, 28 people published more than 5 papers and 13 individuals published more than 7 papers before getting in a conference as organizers. A person has published at most 19 papers before that individual received an organizing position.

Please keep in mind that all these statistics are constrained by the amount of data we have gathered (years 1996 – 2005 for SIGMOD, PODS, and VLDB conferences).

6.2 DB-HEALTH Survey Results

The DB-HEALTH Survey has been conducted over a period of more than a month, the timeline being: **Feb 10th – March 20th, 2006**. The aim was to gather human data regarding various parameters included in short-answer, rating and long-answer questions. The first set of questions has been composed of just personalizing ones, while the next set included ratings (diversity, impartiality, quality, quantity rating and additional factors). The last set, which is the largest, has proven to be the more important of the whole survey since it gave interesting and practical answers to some of the issues and questions raised in this project.

Below is an overall analysis of some of the survey result numbers, including number of responses and statistics to the set of rating questions:

Number of survey invitations: 196
Number of registered people: 32
Number of submissions: 21

Result: We maintained more than 10% responses! The percentage shows the ratio of the number of people that submitted survey responses versus the number of invitations sent. In this case, the number of submissions is 21, and the invitation number raises to about 200.

In general, people rate the executive position experience as good, very good, or OK, but a number of responses mention the hard work required to perform the tough tasks of a PC or PC Chair. The overall response is positive.

In terms of the ratings that include our "health" parameters, database researchers are generally happy with the diversity and with the overall quality of their conferences. Lower marks are given to the ratio of accepted papers versus submitted numbers and to

the impartiality of the paper reviewing inside conferences. In additional criteria, they praise their organization and the amount of new ideas.

Interesting to note is that a response gives an "F" to cronyism (favoritism shown to friends and associates) by a rating of 1. Some of the long-responses are elaborate and should require a careful look. Several good improvement ideas have been raised in this survey, ranging from triage processes, disallowing authors to resubmit rejected or poor papers, double-blind review processes, etc.

6.2.1 Survey taker demographics

Out of the 196 invitations sent, 21 of them responded. Below is a summary of the demographics of invited people, followed by the same type of statistics of the ones who actually responded to our survey invitation and registered on the DB-HEALTH website.

- ***Survey Invitations***

In general, the survey was sent to about 100 different institutions (universities, research centers) and about 196 database researchers. In ranking of position numbers of the invited people, **Program Committee** members were the most numerous – 358 positions held in total in the 3 conferences that we monitored over the 10-year period (SIGMOD, PODS and VLDB). **Specific chairpersons** fell in second place (not including General Chairs, but track/session chairs) with a count of 146 positions from 1997-2006. There were about 60 different types of chair positions in the 3 conferences. **Program Committee Chairs** followed with about 50 occurrences, then General Chairs – 10 positions, and finally some **Co-chairs** (total of 2, 1 in SIGMOD, one in VLDB, none in PODS).

From the big pool of database researchers, more than 30 invitations were sent to the most active people in conferences statistically, and the rest were hand-picked by our team.

Fig. 1 gives a more detailed break-down of the position numbers for each conference (total numbers) in the form of a chart:

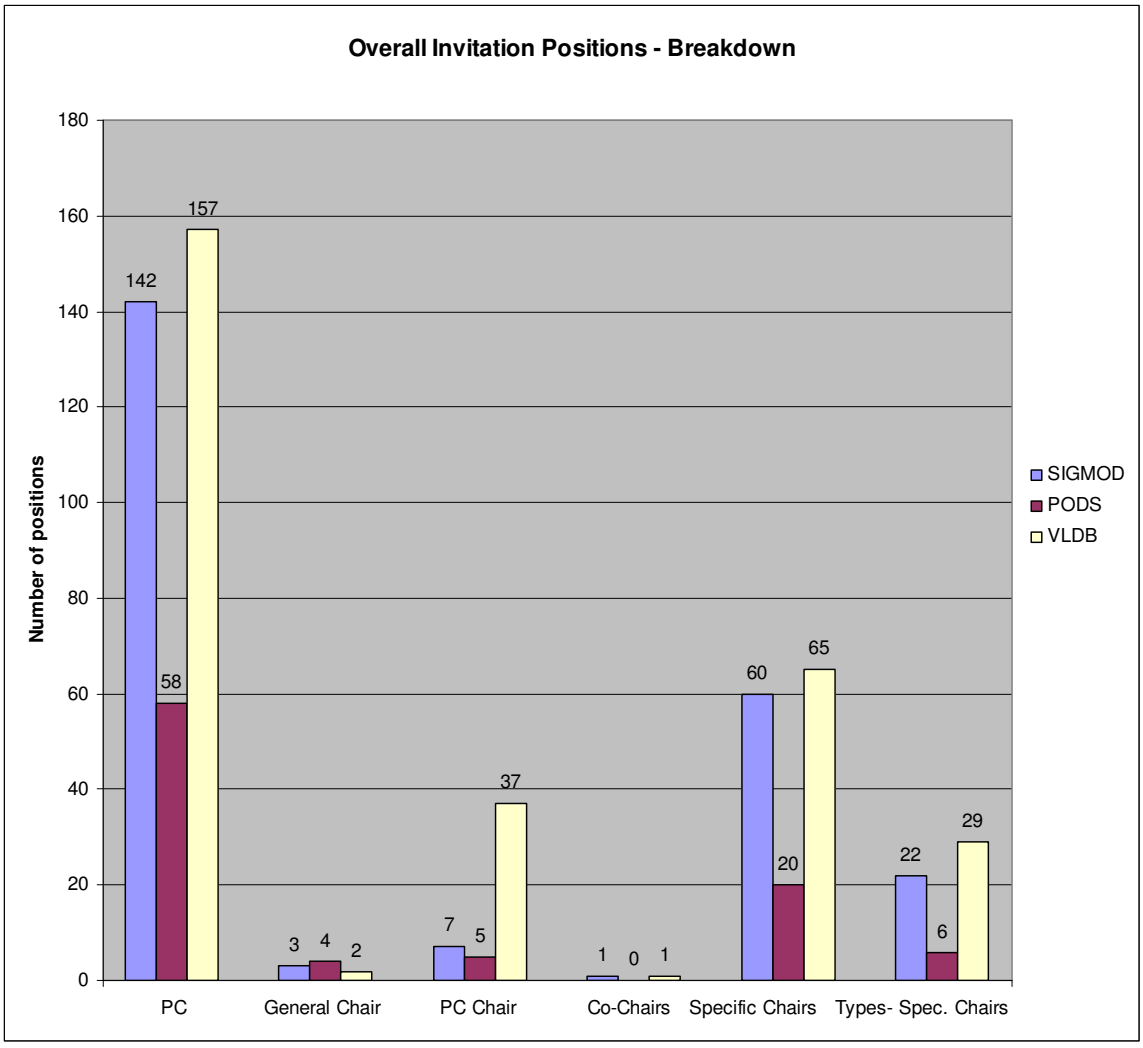


Fig. 1 – Breakdown of conference positions held by people invited to participate in the survey

- ***Survey Invitation Responses***

There were 21 responses to the DB-HEALTH survey. Out of the 21 survey takers, the most common answer for the number of papers published was 10. While the minimum number was 4, the largest amount of papers was 75. The mean was calculated to be 24 and the median 19.

There was some regional diversity, specifically people from Europe, USA, Canada, India and Australia registered in the system. The predominant number was

composed of 17 registrations from the United States. Brazil, Italy and Greece were among other countries listed. Three survey takers came from Canada, and two came from India and two researchers from Greece.

Surprisingly, every invited survey taker chose the "held executive positions" option at registration time, although a few responses denied the fact that they held them.

In terms of institution diversity, most of the places that people were from were North American universities, along with some top research facilities. Any one of them had at most two affiliated researchers responding to the survey invitation. We are not disclosing any names in this section. The rest of the institutions that had actual survey takers have a count of one each. In total, there were representatives from 19 different institutions, among the ones who responded.

Top counts were also noticed in conference participations like SIGMOD and VLDB (19 each), followed by PODS with 13 and OTHER conferences – 16 people.

6.2.2 Summary of numerical answers

Below is a summary of the only numerical part of this survey, the rating question number 3. For each rating, we provided the average, median, minimum and maximum rating, along with a frequency histogram plotting the frequency of each grade by using **2D bars. The first four parameters were designed by us, while the last rating was optional** and there was no need to represent individual answers visually.

Question 3. Ratings - Statistics and Frequency Histograms:

Q 3.1 : **Diversity** (geographical, background-wise)
(1-no diversity, 5-acceptable, 10-very diverse)
Average rating : 6.7
Median : 7
Minimum : 3
Maximum : 10

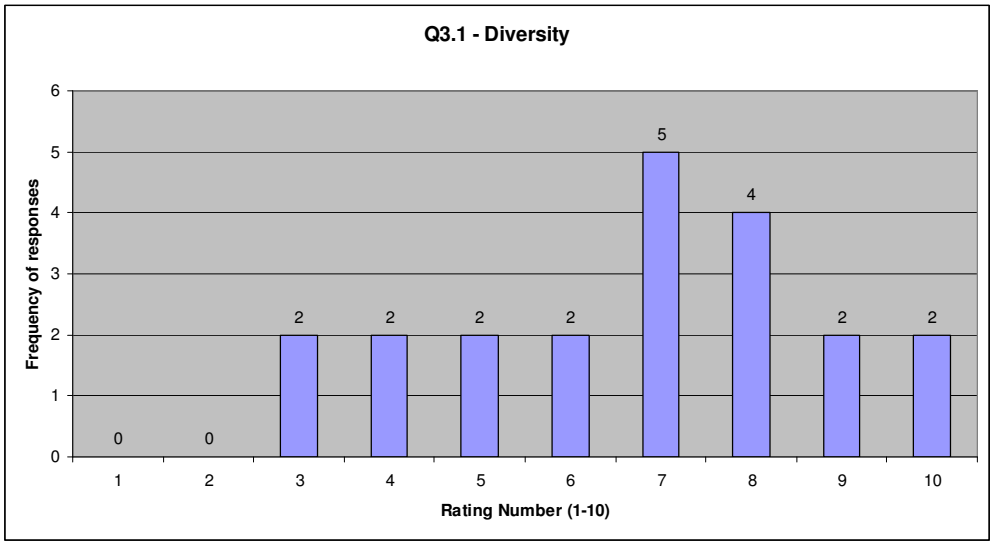


Fig. 2 – The frequency of responses for survey question 3.1 - Diversity

Q 3.2 : **Overall Quality (conference, papers)**

(1-poor, 5-fair, 10-excellent)

Average rating : 7

Median : 8

Minimum : 3

Maximum : 9

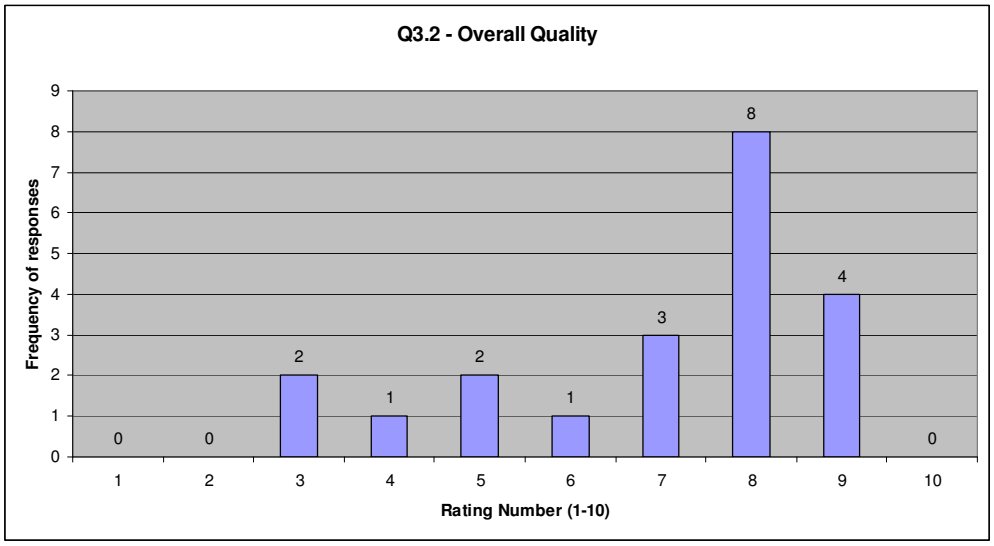


Fig. 3 – The frequency of responses for survey question 3.2 – Overall Quality

Q 3.3 : **Quantity of accepted vs. submitted papers**

(1-very low 0-10%, 5-fair 40-50%, 10-really high 90-100%)

Average rating : 4.2

Median : 3

Minimum : 1

Maximum : 9

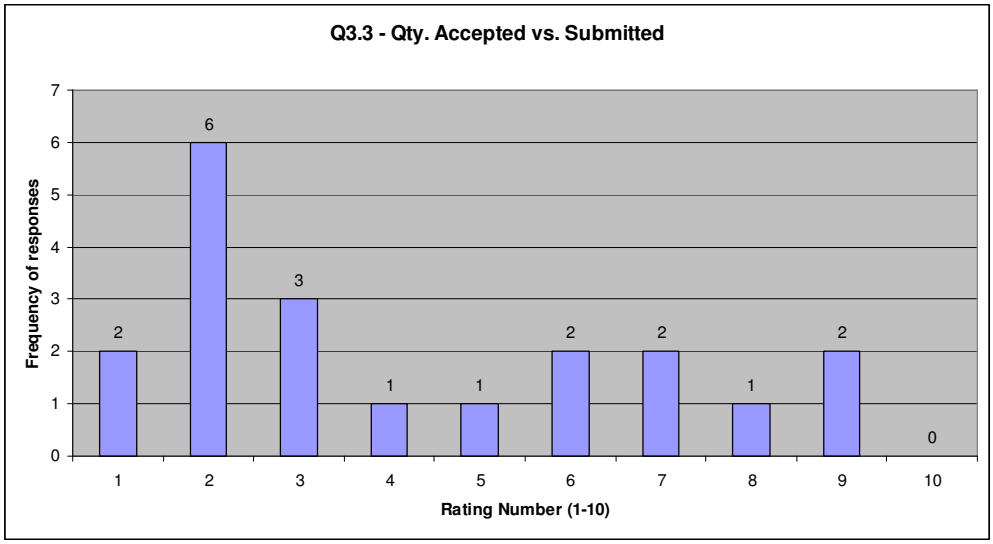


Fig. 4 – The frequency of responses for survey question 3.3 – Qty of Accepted vs. Submitted

Q 3.4 : **Impartiality (fairness)**

(1-very unfair, 5-sufficiently fair, 10-very fair)

Average rating : 6

Median : 6

Minimum : 1

Maximum : 10

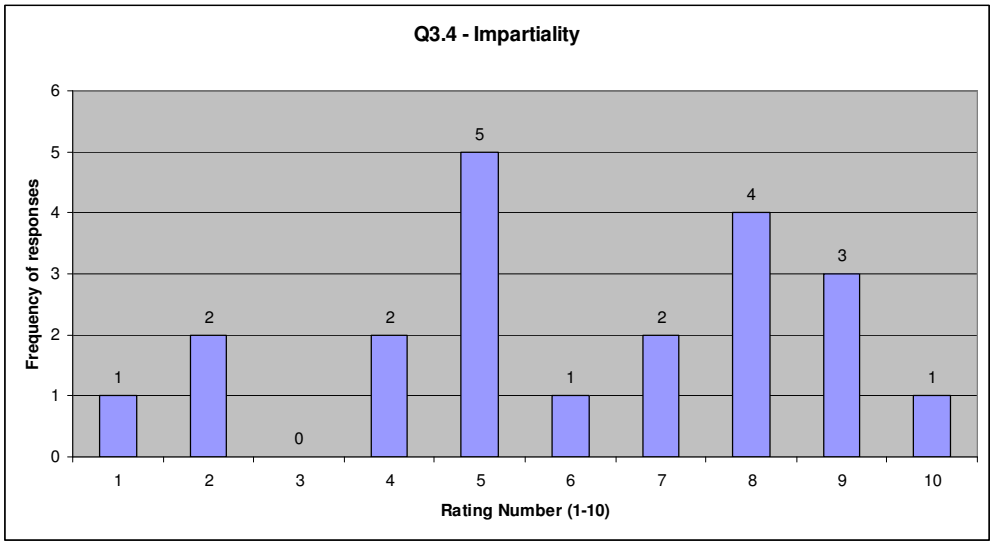


Fig. 5 – The frequency of responses for survey question 3.4 - Impartiality

Q 3.5 : Additional Ratings. Individual answers are included below (some of them are low-rated!). Please note that these are terms that the survey takers decide to use and we refrain from commenting them.

(1-poor, 5-fair, 10-excellent)

Answer 3.5 - Other factors - *Amount of new ideas*: **7**

Answer 3.5 - Other factors - *Quality of reviews*: **6**

Answer 3.5 - Other factors - *Sanity*: **2**

Answer 3.5 - Other factors - *Excessive cronyism*: **1**

Answer 3.5 - Other factors - *Organization*: **8**

7. Data Analysis

7.1 Analysis of database results

The purpose of this section is to provide some graphical representation for the data that we have summarized in Section 6. Along with the presented charts, we will include some comments that try to impartially characterize the data being viewed.

The statistics presented here have been made possible by combining SQL queries over individual or pairs of paper and position tables, and selecting the important information. Using MS Excel, the data was either shown as lines (Sections 7.1.1 and 7.1.2), displayed as individual plots (Section 7.1.3), or frequency bars (Section 7.1.4).

7.1.1 Summary of Conference Positions

We will provide on the next page a chart with the number of conference positions in these past ten years. The purpose of the chart in Fig. 6 is to indicate an overall trend and possibly show some difference in the sizes of the organizational committees for these conferences.

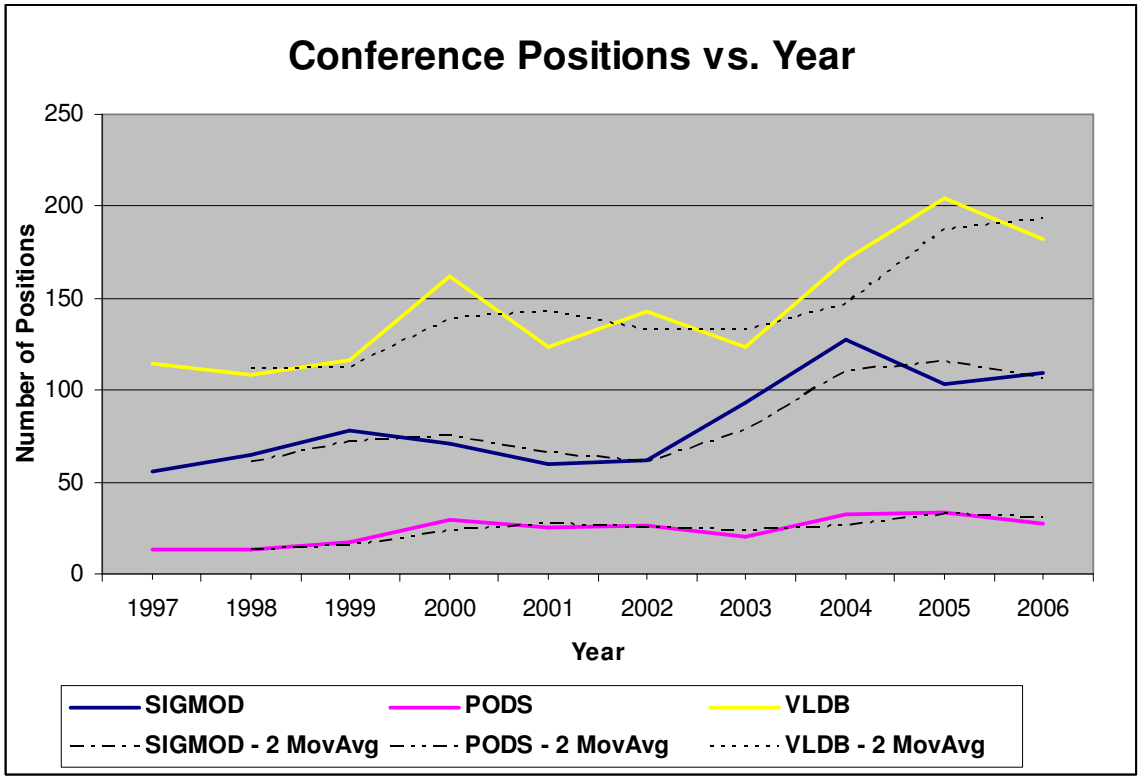


Fig. 6 – The chart shows the number of organizing positions in each conference for the past ten years. The trend lines (2 period moving averages) are relevant for the pattern of the distribution of these numbers.

This chart shows that all three conferences have experienced an upward trend in the last decade. The increase seems to be lowest in the case of PODS, where the number of positions has modified insignificantly in the period 2000-2006. SIGMOD registered a significant increase from 2002 until 2004. Comparing the past two years, the number of conference positions has been increasing once again, however the 2 times moving average is still decreasing at the 2006 value. VLDB has been in the last decade the largest conference of the three. Also, it has experienced the largest degree of variability, experiencing periods of increasing and decreasing size. However, the trend since 2003 is compelling of the increasing size of this conference. Overall, SIGMOD and VLDB have increased in size at an approximately similar pace, while the size of PODS has doubled, even though the conference keeps its smaller scale compared to the other two.

Our guess is that regardless of attempts to decrease the number of conference positions, the non-decreasing trend will be confirmed in the case of all three conferences. Future information remains to prove or disprove our prediction.

7.1.2 Summary of Conference Papers

While we do not possess information of submitted versus accepted number of papers ratio, which seems to be the hottest debate topic of the previous conference reviews, we show in Fig. 7 a representation of the accepted papers for every year in our data history.

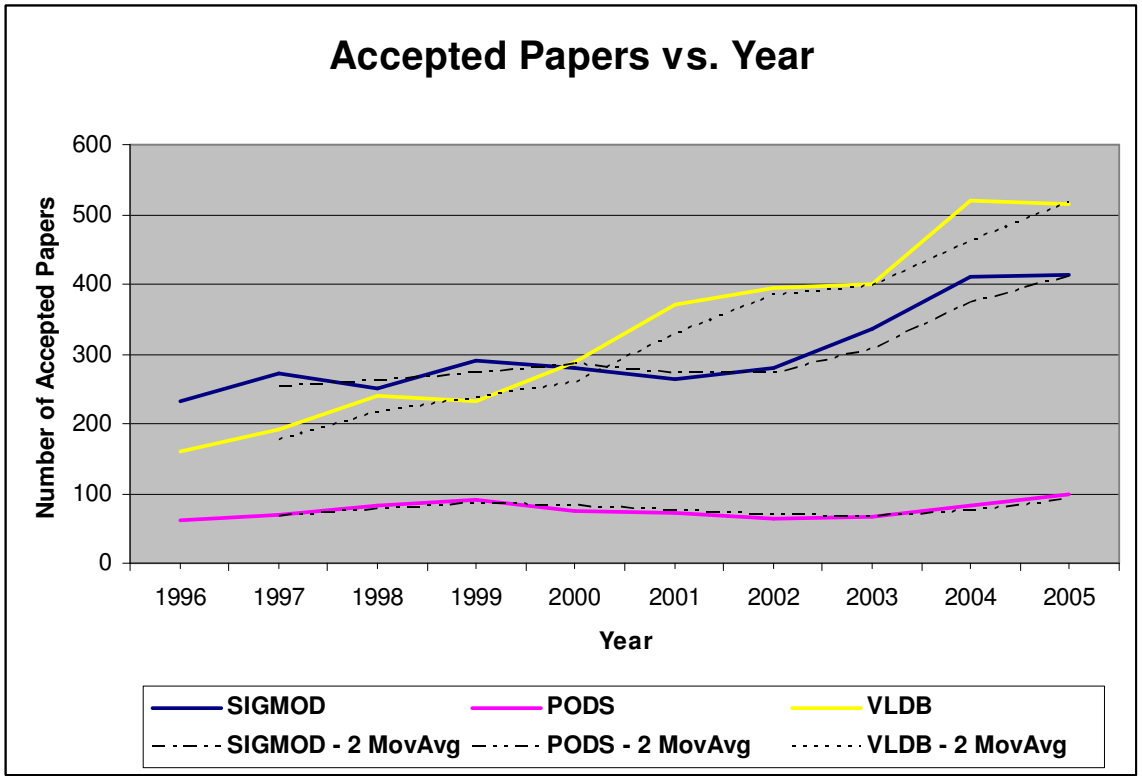


Fig. 7 – The chart shows the number of papers accepted by each conference during the past ten years. The trend lines (2 period moving averages) are relevant for the pattern of the distribution of these numbers.

This chart shows that the number of accepted papers experienced an upward trend in the last decade. Just like in the case of conference organizing positions, the increase

seems to be lowest for PODS, where the number of accepted papers has modified insignificantly in the last decade. SIGMOD registered a significant increase from 2002 until 2005, while the pattern kept changing in the period 1996 – 2000. Even if it accepted fewer papers in 1996 than SIGMOD, VLDB has increased almost constantly and at the highest pace. Overall, SIGMOD and VLDB have increased in size, while the size of PODS has increased but to a lower extent.

Our guess is that regardless of attempts to decrease the number of accepted papers (a process thought to be artificial by a group of people), that number will nevertheless continue to increase in the domain of database research, a domain of increasing importance and involvement. Future information/statistic will either prove or disprove our prediction.

7.1.3 Conference Positions and Conference Papers Comparisons

In the team's effort to produce interesting and useful statistics using the database information that we gathered, some additional charts relating the data on the published papers in each of the 3 conferences (SIGMOD, PODS, VLDB) and the holders of executive positions have been made.

Fig. 8 provides a chart that shows each person who both held executive positions and published papers in SIGMOD in the year range of 1997-2005:

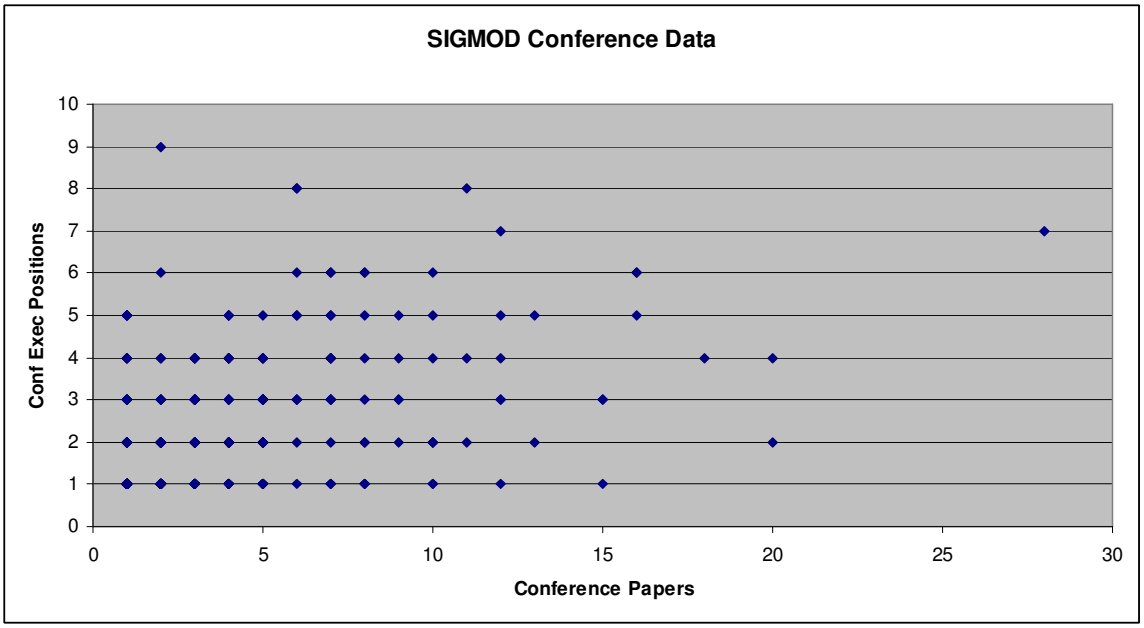


Fig. 8 – Correlation of the people involved in organizing SIGMOD and the number of papers they publish.

Result: Most of the cases appear to be balanced in the number of papers and the number of executive positions, however, there are a few people who have held a high number of leadership roles and have barely published a paper in the SIGMOD conference in the last 10 years (Example, from the chart: 9 executive positions, 1-2 papers). The reverse also exists with people that get a considerable amount of accepted papers and don't really serve much on the executive board (Example, from the chart: 20 papers, 2 executive positions). Most of the data is concentrated on the lower-left side of the chart, with few executive positions and few papers published during 1997-2005.

Fig. 9 provides a chart that shows each person who both held executive positions and published papers in PODS in the year range of 1997-2005:

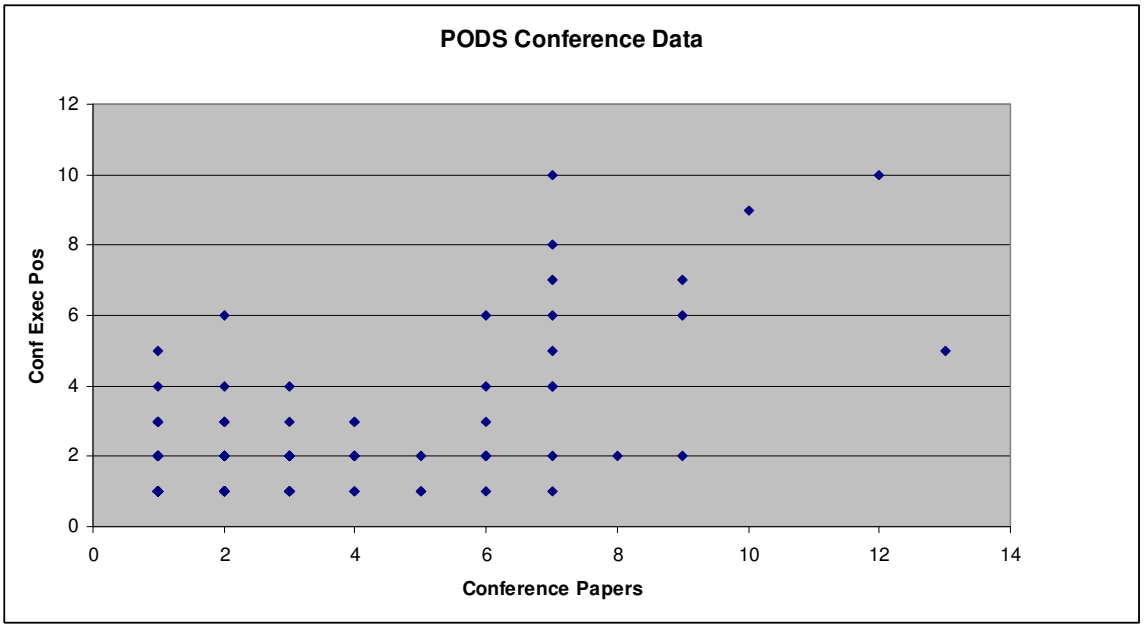


Fig. 9 – Correlation of the people involved in organizing PODS and the number of papers they publish.

Result: Since there is less data on the PODS conference (it is also smaller), it shows clearly a better distribution of the two parameters. Therefore, there are more people in the upper-right region of the graph, closer to the diagonal of equal values. The absence of points in the upper-left part indicates that people have published equal or more papers than they held executive positions, which is a good sign. Probably this is an indicator of the "healthiness" of the database conference. A rotation of positions might be a candidate for the cause of this distribution. Again, the majority of data points are centered in the lower-left region, just as in SIGMOD, with few papers published and with few leadership positions served.

Fig. 10 is a chart that shows each person who both held executive positions and published papers in VLDB in the year range of 1997-2005:

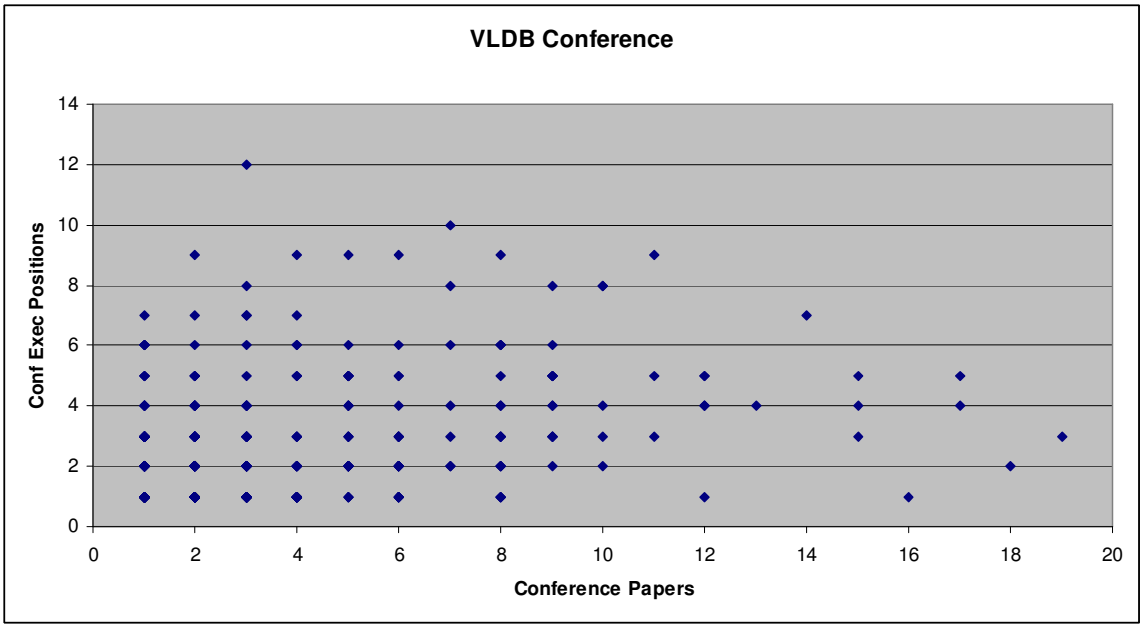


Fig. 10 – Correlation of the people involved in organizing VLDB and the number of papers they publish.

Result: The situation in VLDB is similar to SIGMOD on the left side of the chart, with more positions held than papers published. However, more cases exist where people have many papers in their vitae, but barely held any position. Unfortunately or not, there are not really any data located in the upper-right part along the main diagonal. Here as well the majority of people are situated on the lower-left region, with a low number of papers and a few positions held. Altogether, the representation seems to suggest a more balanced situation than the one in SIGMOD.

7.1.4 View Results Details

Keepoff (Only Paper Publishers) – The following charts (fig. 11 and 12) show values for the people who only submit papers but don't participate in organization committees.

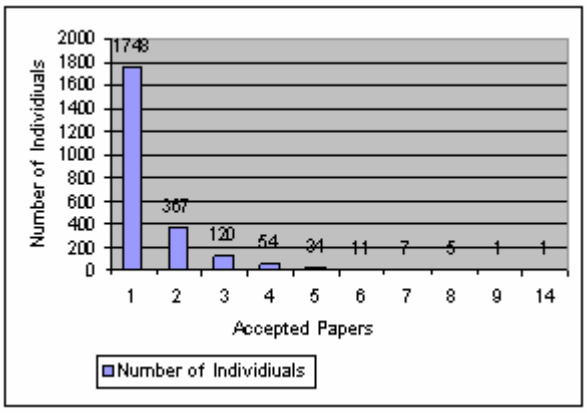


Fig. 11 – No. of People VS No. of Publications

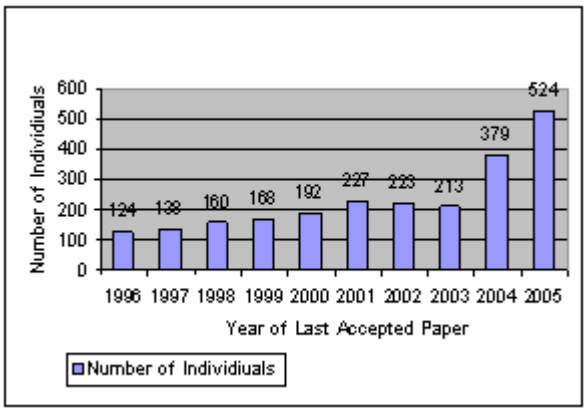


Fig. 12 – No. of People VS Year of Latest Paper

There are a few interesting pieces of information that can be read from these two charts. Firstly, it comes natural to infer the majority of people in this category are not frequent contributors to the database field, as they submit only a few papers in these conferences.

Secondly however, there seems to be an alarmingly increasing number of people who only publish papers in recent years (2004 and 2005). This can be attributed to two likely causes: the field has drawn many adepts in recent years or some people continue writing papers even if they don't participate in the conference organization committees.

Stuck (Only Conference Organizers) – The following charts (fig. 13 and 14) show values for the people who only participate in organization committees but don't have accepted papers.

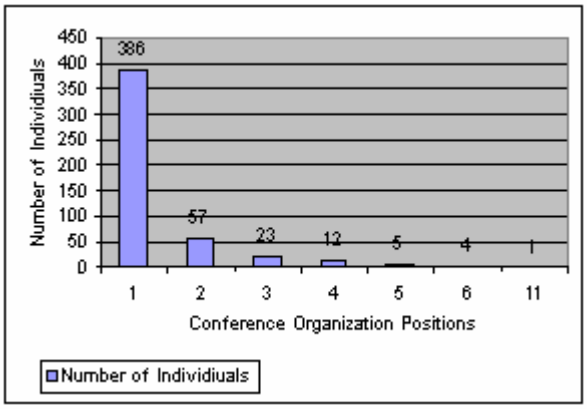


Fig. 13 – No. of People VS No. of Position

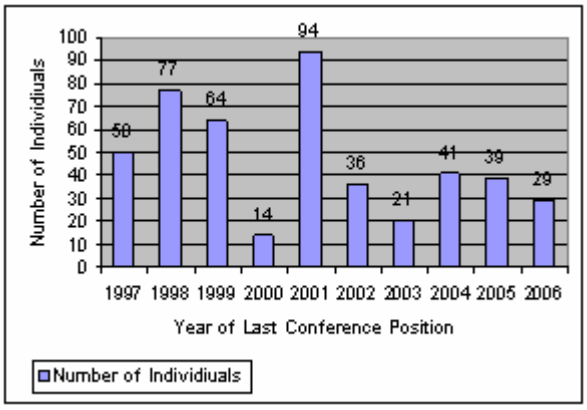


Fig. 14 – No. of People VS Year of Latest Position

The fact that the first chart is skewed left, with the majority of individuals grouped in the first column, shows that the problem of people only participating as organizers is somewhat under control. However, even if large numbers of individuals appear in rare cases, the existence of people holding organizing office positions for more than 4 times and not submitting any papers should raise a serious warning signal.

The second chart informs that this possible problem was highest in 2001. The situation dramatically decreased in amount by 2003, however it is rather alarming that since then the process tends to increase again.

Heat (Simultaneous Publishers and Organizers) – The following chart (fig. 15) show values for the people who helped organize and published papers to a conference in the same year.

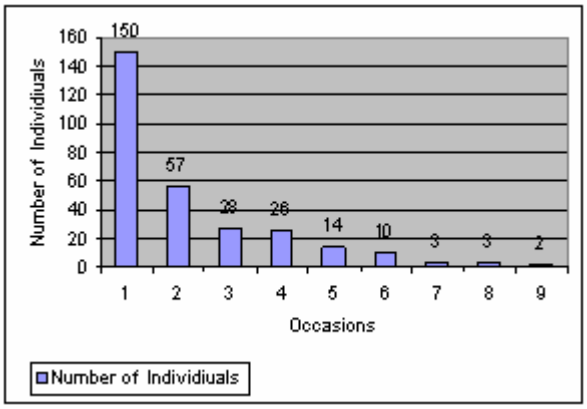
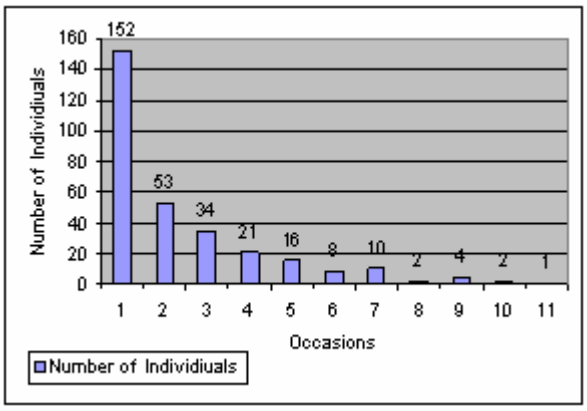


Fig. 15 – No. of People VS No. of Occasions

This chart shows that there is a relatively significant number of people who submit and participate in organizational committees for the same conference, in the same year. While the largest value, 150, describes only the situation of one occurrence, the number related to 4 or more occasions should not be easily overlooked.

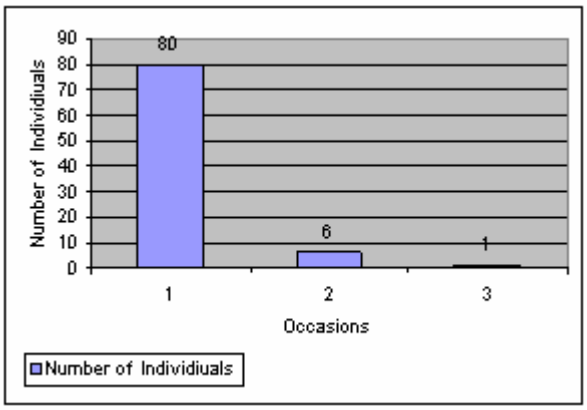
Recognition (Organizing after Publishing) – The following chart (fig. 16) show values for the people who helped organize a conference after publishing papers to it in the previous year.



This chart shows that generally most people are assigned conference positions after only a few accepted papers. There are also a few people who are frequent participants in these conferences and have received 7 or even more such organizing positions.

Fig. 16 – No. of People VS No. of Occasions

Unjustified Organizers – The following chart (fig. 17) show values for the people who helped organize and published papers to a conference in the same year and had not submitted a paper to that conference in the previous year.



This chart shows that even if there are quite a few such people, the number of occasions this situation occurs is usually limited to one per person. This indicates that the situation might not be relevant and/or the person probably had submitted in previous years but in other conferences or at least 2 years before.

Fig. 17 – No. of People VS No. of Occasions

Newcomers (Organizers after more Publications) – The following charts (fig. 18 and 19) show values for the people who were asked to help organize a conference for the first time after submitting a given number of papers.

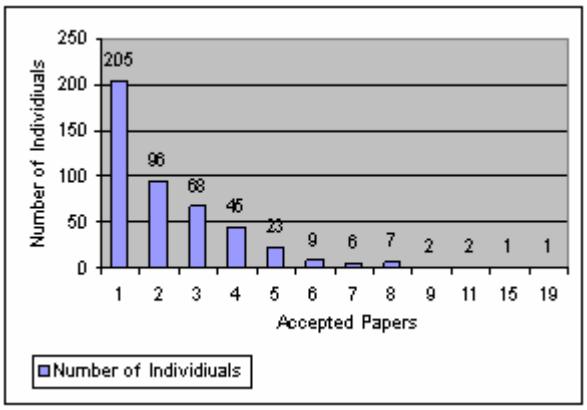


Fig. 18 – No. of People VS No. of Publications

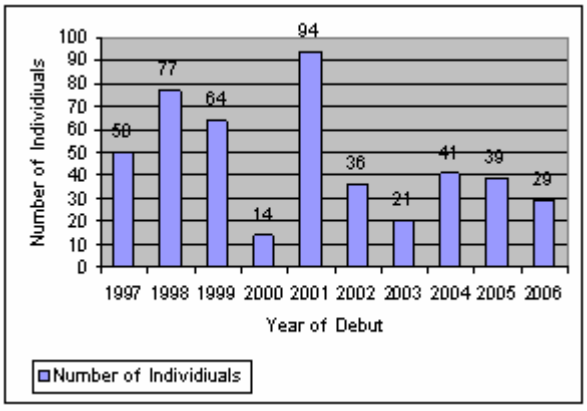


Fig. 19 – No. of People VS Year of Debut

In the case of this view the impact of the fact that our history includes only 10 years and doesn't reflect the career of database researchers is most influential on the results.

From the chart, the fact that the largest bars corresponds to a small amount of accepted papers before participating in an organizing committee is proof that conferences generally welcome newcomers. However, the occurrence of people with more than 6 publications who were yet to participate in organizing committees shows that unfortunate exceptions do occur.

The fact that the number of newcomers occurring each year is rather variable might be a sign that consistency and length of relationship with the conferences are valued in some years. While in 2000 and 2003 there were only 14 and 21 newcomers respectively, there are years that stand out in the number of received newcomers, such as 2001 when 94 newcomers were recorded.

Once again, please keep in mind that all these statistics are constrained by the amount of data we have gathered.

7.2 Summary/Analysis of non-numerical survey results

Below are the summary and commentary of the long answers to Questions 2, 4, 5, 6, 7, 8, 8-1, 9 and 10 of the DB-HEALTH survey:

Question 2. Of all the positions (PC/PC Chair/Organizing Committee/Executive Committee) have you served in any of these roles? How would you rate the experience?

While most of the survey takers rated the experience as being good or very good, a good number mentioned that serving in these positions is exhausting, but rewarding to them and to the community. The overall response has been positive, as we expected.

One interesting quote in this question has been: "A PC Chair has incredible freedom which can be used wisely or unwisely. A chair is also given little guidance, our community does not do a good job communicating <<best practices>>".

Question 4. Are you aware of any flaws in the paper review process or the conference organization itself in Database conferences? If so, could you please describe them?

A lot of the responses deal with the PC workload and the impact on the quality of a review, these two being considered inversely proportional. In major conferences with a very high number of submissions, that constitutes a big problem, including bias and superficiality as well.

Concerns have been expressed about the randomness of the selection process and the consistency of the reviews from year to year. The fault is mainly put on the "human" side of the process and the randomness of the reviewers, which acknowledge the process to be therefore imperfect.

Another common "flaw" as seen by some survey takers is the excessive emphasis on presentation, relevance and syntactic correctness, and too little on the actual ideas. A few times the experience of some reviewers is brought into discussion, including the new people (such as graduate students) who might not be able to recognize some new ideas, revolutionary papers or to provide a quality review.

Finally, a good number of answers approve the double-blind review process, along with author feedback, which they feel it should increase. A few survey takers report that many experiments undertaken in big conferences such as SIGMOD/PODS/VLDB fail and should rather be conducted in smaller ones, and incorporate the successful ones in larger organizations later.

Question 5. Recent studies have shown that while paper submissions in recent Database conferences have significantly risen, the percentage of accepted papers tends to maintain the same approximate value (such as 17%). Have you noticed this? What is your opinion? (Here is your chance to address question 3.3 in detail)

Note: We are also providing a link to a website with interesting statistics on acceptance rates at major database conferences: [...]

Most of the answers to this question wish that the current acceptance ratio to stay the same, since efforts have been made to compensate for the growth of the submission number, but not to the full extent (20-25% acceptance in much earlier years). Some fear that if the percentage rose, an indication of paper quality will be lost and that one or two bad reviews can automatically reject a paper. In the view of the survey taker, the number of submissions is not correlated to the increase in quality of papers. They think that it is important and normal for big conferences to maintain a high selectivity, since the number of poor submission is big. All in all, they think that a ratio such as 17% is fine.

On the other hand, some responses are against the low number of acceptance with respect to number of submissions, arguing mainly that a lot of good/acceptable papers are getting rejected in the conference. They would like to see the acceptance go up by about 10 more percent (roughly one quarter of submissions).

Question 6. How well does your research community welcome new researchers and their papers every year? If you can also describe this issue quantitatively, please do so. (e.g. percent papers by first-time authors)

Almost all answers provide a positive attitude to the idea of newcomers and say that there is no discrimination or preferential treatment given to the new authors. They do not see any problems, even with a high number of new researchers. The general opinion is that the process is

pretty fair and that conferences are open from this point of view; the community welcomes newcomers very well.

A minority reports problems with papers of new authors, but report the usefulness of the blind reviewing process. A suggestion is also the mentorship of authors or getting an opinion before the paper submittal.

Question 7. Do you think that it is good practice for some people to hold executive roles in conferences for several years in a row? Why? Why not? (Executive Positions: General Chair, Vice-Chair, PC Chair, Executive Committee, etc.)

The general view towards holding executive roles "for several years in a row" is that it should not happen, and that change is always good, provides diversity and helps remove bias. In any case, all answers that mention PC's and PC chairs agree that they are the first positions to be diverse every year. Executive positions are given more positive judgment, affirming that (organizational) continuity and strategy should be kept longer.

Some replies talk about a maximum of 2 or 3 years, one year for learning how to do the job, and the next to do it right and pass on the lessons on how to do things properly. One answers complains about the lack of guidance given to a PC chair regarding the "best practices" – what works well and what does not.

Question 8. In your opinion, should people who have had several papers presented at the conference be considered or given priority for leading positions at the conference in subsequent years?

A few main ideas have been expressed on this question. First of all, their opinion is that being a "leading researcher" is not correlated to being a good organizer. However, having both qualities is the best thing.

Another idea is that priority given for leading positions is a fair thing to do since "leading researchers" should be recognized for their status in the community, along with their experience on submission and review process.

In addition, a concern is expressed regarding the specialization of some people with research papers. They think it is a good idea for the researcher to know "many areas reasonably well" in order to better qualify for leading positions. An opinion that may tie in is: "Leading

positions should go to people who have been successful with the conference and therefore know it well."

Question 8-1. How are executive positions elected in conferences in your area? (we are interested in PC, PC Chair, General Chair, and other leadership positions)

Since this is, indeed, public information, the question actually tested how many of the survey takers were familiar with the process and how well. Most of the answers were the same regarding SIGMOD and VLDB elections, and some criticisms fell on SIGMOD in favor of VLDB. Formally, chairs and other high-ranking positions are elected by some sort of steering committee who, in turn, name lower positions within the conference.

It is good that some informal remarks have been made, including the fact that leaders should have done some important work for the community, or criteria such as: experience, "visibility on the field", and the ability to run/perform that position. Diversity is also an important factor, especially in selecting program committee members.

Although all of the survey takers chose the option of having served in the executive committee of any conference, one response lacked the knowledge of the election of any position. This contrast also appears in question 2 of this survey, showed by the results.

Question 9. If you were given the chance to do so, what would you change in the conference paper reviewing process?

This is a very important question and some good suggestions have been made. Some more frequently mentioned ideas have been the double-blind reviewing process, along with author feedback on papers. Consequently, authors should be able to respond to the reviewer's comments (one comment even talks about responses during the reviewing process) in order to make things clearer.

Another well shared opinion is the disallowance of rejected papers in one conference to be resubmitted to other ones immediately. Arguments in support of this view have been the time to allow the author to improve the paper using the reviewer feedback before resubmission and the dropping of paper submission numbers to help reduce workload and to make a more quality review of papers. Moreover, "memory" should be built into the system by having previously

reviewed papers carry the comments from the reviewers and the responses from the authors as well.

One more interesting suggestion is that the PC chair needs to check all reviews for meaningfulness and any flaws (a so-called "quality control"). In a response to the number of reviews that a single chair needs to check, it is mentioned that this is, nowadays, a many persons' process with conferences of such scale. Talking about PC, a cap should be placed on the number or papers a PC member should review, according to another opinion, as well as the restriction of PC members to serve in other conferences in parallel.

A triage process and the anonymity of abstracts, titles of papers have also been suggested. PC members and PC chairs should be anonymous, the same as the papers (no author names). A response would like the reviews to be critiqued publicly, allowing the option of early reject, but also permitting the paper to be reconsidered, and introduce some kind of "pass mark", an acceptance threshold on papers.

All in all, many want to see the double-blind review process and other good processes implemented in all major database conferences and become standards that ensure the consistency of the whole review process. The reviewer load should be reduced so as to enable them to make a better job reviewing.

Question 10. In an honest assessment, how would you rate yourself as a reviewer, and how would you rate your peers in terms of toughness, fairness and consistency?

In a self-assessment pretty much all of the responses confirmed the fact that they are fair and try to remain so, and are usually tough in reviewing, but try to be more lenient with papers that contain novel ideas. Consistency is a problem for a lot of reviewers and this is mentioned throughout many of the responses.

A complain that was expressed is that some reviewers tend to look for the flaws in a paper, rather than "seeing the potential impact", recognizing good ideas and things that authors did to the best of their ability. In addition, many cannot make a clear decision on papers or do the job superficially. Some of them try to be positive and "find the best in every paper".

It seems, from the responses, that doing a good PC job is very time-consuming and that the workload should be reduced to keep up the quality and consistency of reviews.

Finally it is worth mentioning the issue that some of the responses are tackling: some papers are SIGMOD papers, and so on, geared towards presentation and relevance, and that people have strict notion of what a SIGMOD paper should “look like”. Consequently “some novel ideas/concepts often do not fit the accepted molds”.

8. Conclusions

In this section, we are making an attempt to produce a valid conclusion with respect to the goal of our project. As mentioned in Section 2.1, our goal was to create an overview on the "health" of the research section of the database community at the curent moment.

In the section mentioned above, we have roughly described the concept of health as being a balanced combination of a few general factors: diversity, quality, quantity and fairness (impartiality). We again assumed that conferences are interested in improving this factor, and undertook the task of finding out the quality of the efforts being made in this direction.

Striving to achieve useful results, our position had to be as neutral as possible. Combining blind, data-generated results with the subjective answers we received through our survey, we consider our analysis in Section 7 to be objective.

On the other hand, different readers are going to argue that some variable thresholds are acceptable and relevant, while others might think the same limits are not appropriate and could be overlooked. One basic example is given in the case of the analysis of people who do not participate in organizing committees even if they have made several publications. For us and based on our current knowledge, a person with 8 publications and no conference positions should've been given at least a PC position by now. From experience or personal knowledge and observations, some readers could argue that such a situation is perfectly normal, and thus, the source of subjectivity related to these variables. Once again, our position is that of an observer, and, accordingly, we mentioned for each variable the values associated with different thresholds.

One more important aspect of our analysis is that it is based on a ten year span of conference activity. However, regardless of the size of the considered history, this is only a static glimpse on the community. The health of research conferences in the database community is an evolving descriptor. Consequently, we will stress in the following

section the importance of continuing this observatory work. If this is achieved, people will be able to more accurately understand the direction in which an individual conference might be going.

Our Point of View

After stating our goals, assumptions and limitations, we are now going to expose the point of view that can be deduced from the data we have gathered. Let us see what a general survey would probably look like:

[Category : ***Personalization***]

Question 1. Number of papers in the past 5 years.

Answer: 1 – 2 papers. **Explanation:** Our data indicates there were 1264 papers and 2080 authors in the past 5 years. The value we have given is an expected one based on these two parameters.

Question 2. Leadership roles.

Answer: No, I have not been involved in any conference as organizer. **Explanation:** Given that 3020 people have had papers published papers, out of which 238 have not been involved in organizing committees, we incline to say that an average individual has not had leadership roles.

[Category : ***General health of DB - opinions***]

Question 3. Answers are based on the means or medians for each category based on the distribution of answers observed in each situation.

- **Diversity** (geographical / background-wise)
  7 – acceptably high diversity.
- **Overall quality** of the papers and the conference itself
  8 – good quality.

- **Quantity of accepted vs. submitted papers**

  3 – approximately 20-30%.
- **Impartiality**

  6 – sufficiently fair
- **Other factors** (feel free to include any)

  No answer.

Question 4. Conference flaws.

Answer: I am rather discontent with the situation in database conferences. I feel my paper has been reviewed by a person not totally interested or aware of the approached topic, who may have not judged the paper based on its contents. **Explanation:** Our answer is based on the results from our survey that have been summarized in Section 7.2.

[Category : ***Specific, target questions***]

Question 5. 17% acceptance ratio.

Answer: I am knowledgeable of the acceptance ratio. While a higher rate would guarantee more people get their works published, too high ratios might mean poor papers are accepted. **Explanation:** Again, our answer is based on the summary presented in Section 7.2.

Question 6. Newcomers.

Answer: In recent years, the larger number of accepted papers brought new researchers to the community. This is a good thing, and there is no special treatment applied to newcomers. **Explanation:** This answer is based on the summary in Section 7.2. Also, the summary for question 10 in the same section mentions that reviewers are interested in novel ideas. However, if we look at fig. 13, we can see that there is a very large number of people who only had one accepted paper. This could indicate that newcomers generally do not maintain a long profile in the database research community.

[Sub-category : ***Conference leadership***]

Question 7. Do you think that it is good practice for some people to hold executive roles in conferences for several years in a row?

Answer: No, this should not happen in the case of PC and PC Chair positions. Executive positions could be held for more years, but a limit should be set at 2-3 years. **Explanation:** Answer based on Section 7.2.

Question 8. In your opinion, should people who have had several papers presented at the conference be considered or given priority for leading positions at the conference in subsequent years?

Answer: Yes, such people can be considered for leading positions more often, but it should be clear that they accept a broad selection of topics and are impartial in their judgments. The quality of their papers does not always reflect in their ability to run executive positions. **Explanation:** Answer is based on Summary 7.2. According to figure 11, it is visible that certain individuals are preferred for conference positions even if they don't have more recent accepted papers.

Question 8-1. How are executive positions elected in conferences in your area? (we are interested in PC, PC Chair, General Chair, and other leadership positions)

Answer: Chairs and other high-ranking positions are elected by some sort of steering committee who, in turn, name lower positions within the conference. **Explanation:** Answer is based in the summary in Section 7.2.

[Category : ***Possible solutions, improvement***]

Question 9. If you were given the chance to do so, what would you change in the conference paper reviewing process?

Answer: Conferences should first implement a triage stage, in which papers are given to members of the reviewing committee who have knowledge of the topic. Once replies are created by reviewers, committee members should validate that these replies are meaningful. In case these responses are of good quality, they should be recorded for

future reference and sent to the paper submitter. In case the paper is rejected, it should not be allowed to be submitted again before relevant modifications are made. **Explanation:** Answer summarizes Section 7.2.

[Category : ***Self-assessment***]

Question 10. In an honest assessment, how would you rate yourself as a reviewer, and how would you rate your peers in terms of toughness, fairness and consistency?

Answer: I consider myself a tough reviewer, but I am trying to find the good parts in all papers before passing a judgment. As for my peers, doing a good PC job is very time-consuming and the workload should be reduced to keep up the quality and consistency of their reviews. **Explanation:** Answer summarizes Section 7.2.

In conclusion, the reviewing process in particular, and conference organizations in general, could benefit from the implementation of techniques that balance the workload on each reviewer in order to produce good feedback. While such techniques are currently experimented, it is understandable that there are people either pleased with the overall mechanism or frustrated with it. Overall, the health of the database conferences is almost good, with some exceptions that will, hopefully, be addressed in future years.

9. Recommendations

This section of the report deals with suggestions that come out of the survey responses, along with some of our opinions on the improvement of the database conference review process and on helping new researchers that are entering the field. Since there are many newcomers every year, it is useful to have a system in place for them. The sub-sections below discuss these problems, the third one proposing an extension of this project for more years and more conferences.

9.1 Final Recommendations for Conference Organizers

Most of the suggestions come directly from the survey responses, but we are putting all good ideas together in this section which we think might be useful for conference organizers. They include things about the review process, workload of the Program Committees and a few criteria to consider when electing new positions.

If it does not yet exist, the conference should find a way to provide help for the paper authors that need it, such as a mentorship program. Mentors are very useful for newcomers in the research publication field and can provide information on how to support, present and develop their ideas in a paper.

To maintain the perception of fairness (even though it might not have a large impact on the acceptance rates), double-blind review could be useful. One more idea would be to try and remove all forms of cronyism and increase diversity of Program Committee members and their chairs. Reduce bias and elect new members more often.

An essential method to help fellow researchers is to implement an author feedback system and the possibility to respond to the comments and give chance to improvement. Moreover, it is in the interest of both sides to have some kind of “memory” built into the reviewing system, something that will carry with the current review previous feedback, responses and improvement. As mentioned in the survey results, a triage process could also be included where papers can be filtered out based on title pages and abstracts (all in anonymous form, no names whatsoever) and include an acceptance threshold for the given marks in this process.

We acknowledge that Program Committee members have a significant amount of workload when it comes to paper reviewing, so a cap on the number of papers for each PC member would make sense. It has been mentioned that the PC Chairs have a lot of work doing quality check on paper reviews, so spreading out this task is something to take into account. We are not including any specific suggestions for this issue.

In the DB-HEALTH survey a fact has been brought up, that people know what a "SIGMOD" paper should look like. If there are such conceptions in the conference, it would be preferred for those to be removed. In addition, try to accept papers whose ideas and their support significantly compensate for superficial presentation and/or relevance, as long as they have any (bad presentation or not relevant papers would continue, of course, to be unacceptable). However, review feedback should still include suggestions on the need and/or methods for presentation and relevance improvement.

Finally, to make sure that members of the executive committee do their job successfully and impartially, elected people should know their attributions very well and should be given priority if they had success previously with the respective conference. a good idea for the researcher to know "many areas reasonably well" in order to better qualify for leading positions. "Leading positions should go to people who have been successful with the conference and therefore know it well".

9.2 Final Recommendations for New Researchers

These suggestions may not be entirely original or expected to bring significant changes, however we think they are worthwhile to consider. For now, until conference policies change (if ever), new authors without much experience should try in addition to work on presentation and relevance on their paper. Big ideas should be very well supported, but there is not much to do in order to convince a PC member to accept it. Presentation will help in their efforts.

A good way to gain experience might be the usefulness of conference mentors. Some of the conferences have already tried it (such as **SIGIR** – The **S**pecial **I**nterest **G**roup on **I**nformation **R**etrieval conference **[5]**) and it is a good way to get a glimpse on how PC members think and judge papers. Apprenticeships are also a valid idea. As well

as this, (informal) opinion of peer researchers or previous/current PC members could be valuable and should be done before the author submits his or her paper. If the conference allows it, new authors should ask for feedback and for help on strategies for improving the paper quality and relevance.

The bottom line is that database conferences accept new researchers well. However they should get some initial help if they feel they need it or do not have a good idea about the way that paper for the conference should be written or presented until they get more experienced.

9.3 Project Extension

This project and its accompanied report studied the "health" of database conferences only based on hypothetical assumptions. Based on the relevant results of our study, we can conclude that it is worthwhile to expand this project initiative to other research fields as well. Moreover, the target conferences were only major ones, SIGMOD, PODS and VLDB. Considering that it is very likely to come across different levels of "health", it is recommended that the project is extended to other database conferences as well (such as ICDE - International Conference on Data Engineering, CIKM - Conference on Information and Knowledge Management).

In our efforts to produce results in a short period of time, the team decided to collect data from only the last 10 documented years of the three conferences, both in terms of conference organization and research papers. Consequently, some results in this project report are constrained to this time window and may not capture the entire activity of some researchers that have been active before 1996 or 1997. Therefore, it is advisable, in case this effort is to be continued, to update the conference database with new conference data each year and keep the results of all queries up-to-date. Statistics are then easy to extract in order to display more recent trends. As well as this, old conference records should be used to augment the people and papers database to make up for the lack of data on people that have been active before our time window start – 1996.

10. References

The references, in order of their citation, are:

[1] Rethinking the Conference Review Process, Franklin, Widom et al. (2004)
http://www.sciences.univ-nantes.fr/lina/SIGMODPODS04/panelcamera.pdf
Source: Proceedings of the 2004 ACM SIGMOD
PANEL SESSION: Panel
Pages: 957 - 957
Year of Publication: 2004

[2] The Health of Research Conferences, David A. Patterson (2004)
http://portal.acm.org/citation.cfm?id=1035134.1035153
Source: Communications of the ACM
Volume 47 , Issue 12 (December 2004)
The Blogosphere
Pages: 23 - 24
Year of Publication: 2004

[3] OOPSLA Website
http://www.oopsla.org/2006/

[4] Database Publication Practices, Bernstein et al. (2004)
http://www.vldb2005.org/program/paper/wed/p1241-bernstein.pdf
Source: Proceedings of the 2005 VLDB
Pages: 1241-1246
Year of Publication: 2005

[5] SIGIR 2006 – Mentorship Program
http://www.sigir2006.org/mentoring.asp

11. Bibliography

11.1 Literature

1. Garcia-Molina et. al. *Database Systems, The Complete Book.* Prentice Hall ISBN: 0-13-031995-3

11.2 The Internet

2. SIGMOD Conferences – Online Collection
   http://www.SIGMOD.org/SIGMOD/conferences/index.html

3. PODS Conferences – Online Collection
   http://www.SIGMOD.org/SIGMOD/conferences/index.html

4. Very Large Databases (VDLB) Conferences – Online Collection
   http://www.vldb.org/

5. The DBLP (Database Conference Archive)
   http://www.informatik.uni-trier.de/~ley/db/

6. Rethinking the Conference Review Process, Franklin, Widom et al. (2004)
   http://www.sciences.univ-nantes.fr/lina/SIGMODPODS04/panelcamera.pdf

7. The Health of Research Conferences, David A. Patterson (2004)
   http://portal.acm.org/citation.cfm?id=1035134.1035153

8. Database Publication Practices, Bernstein et al. (2004)
   http://www.vldb2005.org/program/paper/wed/p1241-bernstein.pdf

9. SIGMOD PC Chair's website on Conflicts of Interest:
   http://infolab.stanford.edu/~widom/SIGMOD-coi.html

10. Perl / CGI Tutorials and Sample Scripts
    http://www.hotscripts.com/CGI_and_Perl/index.html

11. Perl.com : The Source for Perl
    http://www.perl.com/

12. SIGMOD, PODS, VLDB – Organizing Committees (Websites) *
    * Please see **Appendix 4**.

12. Appendices

12.1 Appendix 1 – DB-HEALTH Survey

The actual **survey** is comprised of about 10 main questions, divided in several categories, as follows: 2 questions for **personalization** (survey ease-in), 2 rough questions on the **overall "health" of database research conferences** (including the question with the multiple ratings of 1-10), 2 questions on targeted, specific matters (**paper acceptance vs. submission numbers**, **conference tolerance towards new-comers**), 3 questions on the delicate topic of **conference leadership**, and 1 question each in the categories of **possible solutions and improvement**, and **self-assessment**.

The final questions of the surveys can be found below, along with their dynamic alternatives:

[Category : ***Personalization***]

Question 1. For demographic purposes only, how many database papers have you published in the past 5 years?

Question 2. Having served in one or more leadership roles (PC/PC Chair/Organizing Committee/Executive Committee) in the [...] conferences, how would you rate the experience?

OR (depending on user information)

How would you rate the performance of the various leadership roles (PC/PC Chair/Organizing Committee/Executive Committee) in [...] Database conferences?

[Category : ***General health of DB - opinions***]

Question 3. On a scale of 1 to 10, how would you rate the "health" of database research in your community, based on the following factors that we determined to define

the term "health"? (We define health here as: diversity, quality, quantity, impartiality of organization and evaluation of papers)

- **Diversity** (geographical / background-wise)
  (1-no diversity, 5-acceptable, 10-very diverse)
- **Overall quality** of the papers and the conference itself
  (1-poor, 5-fair, 10-excellent)
- **Quantity of accepted vs. submitted papers**
  (1-very low 0-10%, 5-fair 40-50%, 10-really high 90-100%)
- **Impartiality**
  (1-very unfair, 5-sufficiently fair, 10-very fair)
- **Other factors** (feel free to include any)
  (1-poor, 5-fair, 10-excellent)

Question 4. Are you aware of any flaws in the paper review process or the conference organization itself in Database conferences? If so, could you please describe them?

[Category : ***Specific, target questions***]

Question 5. Recent studies have shown that while paper submissions in recent Database conferences have significantly risen, the percentage of accepted papers tends to maintain the same approximate value (such as 17%). Have you noticed this? What is your opinion? (Here is your chance to address question 3.3 in detail)

Note: We are also providing a link to a website with interesting statistics on acceptance rates at major database conferences: [...]

Question 6. How well does your research community welcome new researchers and their papers every year? If you can also describe this issue quantitatively, please do so. (e.g. percent papers by first-time authors)

[Sub-category : ***Conference leadership***]

Question 7. Do you think that it is good practice for some people to hold executive roles in conferences for several years in a row? Why? Why not? (Executive Positions: General Chair, Vice-Chair, PC Chair, Executive Committee, etc.)

Question 8. In your opinion, should people who have had several papers presented at the conference be considered or given priority for leading positions at the conference in subsequent years?

Question 8-1. How are executive positions elected in conferences in your area? (we are interested in PC, PC Chair, General Chair, and other leadership positions)

[Category : ***Possible solutions, improvement***]

Question 9. If you were given the chance to do so, what would you change in the conference paper reviewing process?

[Category : ***Self-assessment***]

Question 10. In an honest assessment, how would you rate yourself as a reviewer, and how would you rate your peers in terms of toughness, fairness and consistency?

12.2 Appendix 2 – DB-HEALTH Survey Results

The following survey responses are all in anonymous form and some references to real names have been removed. We have attached all 21 responses, regardless the completeness of any answer (including no answers to some of the questions).

Survey Response 1 *at* **Fri Feb 10 11:04:34 2006**:

Answer 1 : 10

Answer 2 : OK

[Question 3 - Ratings]

Answer 3.1 - Diversity (geographical, background-wise) :	**7**
Answer 3.2 - Overall quality of the papers and the conference itself :	**8**
Answer 3.3 - Quantity of accepted vs. submitted papers :	**6**
Answer 3.4 - Impartiality (fairness) :	**10**
Answer 3.5 - Other factors - :	**no answer**

Answer 4 : Too many submissions -- make good reviews hard to do in a limited time.

Answer 5 : It would be a real problem if percentage acceptance fell below current levels

Answer 6 : Don't have data. Certainly no intentional discrimination. Even some mild "affirmative action" to favor newcomers.

Answer 7 : No. Change is good.

Answer 8 : Yes -- this should be a primary criterion.

Answer 8.1 : Appointed by governing bodies.

Answer 9 : Not sure.

Answer 10 : As good as is humanly possible under the circumstance -- both for me and for most of my peers.

Survey Response 2 *at* **Mon Feb 13 10:17:29 2006**:

Answer 1 : 25

Answer 2 : exhausting

[Question 3 - Ratings]

Answer 3.1 - Diversity (geographical, background-wise) : **10**

Answer 3.2 - Overall quality of the papers and the conference itself : **9**

Answer 3.3 - Quantity of accepted vs. submitted papers : **9**

Answer 3.4 - Impartiality (fairness) : **9**

Answer 3.5 - Other factors - : **no answer**

Answer 4 : Reviewer overload leads to lesser quality reviewing

Answer 5 : I have noticed this. It is a conscious effort on the part of the committee by increasing the number of papers accepted.

Answer 6 : very well

Answer 7 : To some extent, yes. Experience is important. However, it is also important to give younger researchers a chance. Younger leaders could have backup from more experienced folks.

Answer 8 : Yes, if the leaders should be the "leading researchers". On the other hand, these people might not be the best organizers.

Answer 8.1 : VLDB does it with the consent of the Board of Directors (Endowment). SIGMOD does it with the consent of the SIGMOD executive committee.

Answer 9 : I would institute a triage process. Authors should prepare an abstract of no more than 3 pages that will convince the committee that their work is interesting. The committee would invite the best <reasonable number> of those.

Answer 10 : toughness = 9 fairness = 10 consistency = 7

Survey Response 3 *at* **Fri Feb 24 11:19:28 2006**:

Answer 1 : 19

Answer 2 : good

[Question 3 - Ratings]

Answer 3.1 - Diversity (geographical, background-wise) :	**10**
Answer 3.2 - Overall quality of the papers and the conference itself :	**9**
Answer 3.3 - Quantity of accepted vs. submitted papers :	**1**
Answer 3.4 - Impartiality (fairness) :	**7**
Answer 3.5 - Other factors - :	**no answer**

Answer 4 : Superficial or highly biased reviews are very frequent. The high workload of PC members is a major problem. At some conferences, especially those with huge PCs where very few papers have common referees, selection is often perceived as a rather arbitrary process, hihly dependent on the specific referees assigned and on their effectiveness in defending their papers.

Answer 5 : High selectivity is a good thing. Many conferences receive huge numbers of poor submissions.

Answer 6 : I believe new authors are generally treated the same as other authors.

Answer 7 : It depends on the position. Exec committee yes (for organizational continuity), PC chair no.

Answer 8 : What do you mean by "leading positions" ?

Answer 8.1 : This depends on the conference. At Pods, the PC chair and General Chair are elected by the Executive committee, which consists of the 3 most recent PC chairs and 3 most recent General chairs. General chairs serve 2 year terms.

Answer 9 : Hard problem, it depends on the conference. I do not alime the anonymization practiced by SIGMOD, I like the author feedback tried by Jennifer Widom at SIGMOD. There should be a way to penalize large numbers of repeated submissions of poor quality by the same person. However, this is hard to do.

Answer 10 : I consider myself a thorough, fair and rather tough reviewer. Rating my peers is too broad to answer. There is the whole spectrum of reviewers.

Survey Response 4 *at* **Fri Feb 24 14:11:08 2006**:

Answer 1 : 50-75

Answer 2 : rewarding, but a lot of work

[Question 3 - Ratings]

Answer 3.1 - Diversity (geographical, background-wise) : **9**

Answer 3.2 - Overall quality of the papers and the conference itself : **9**

Answer 3.3 - Quantity of accepted vs. submitted papers : **9**

Answer 3.4 - Impartiality (fairness) : **9**

Answer 3.5 - Other factors - *amount of new ideas* : **7**

Answer 4 : Senior, established, well-known, well-connected researchers have it easier. Researchers with good affiliations have it easier. Double-blind reviewing helps counter these.

Answer 5 : I have noticed this. I think it is a fact. More conferences get more selective. Self-refereeing is also an important aspect. 15% at CIKM is not as tough as 15% at SIGMOD.

Answer 6 : Most first-time authors author papers with authors who are not first-time authors, as part of their Ph.D. studies. I am not sure that there is a need for special, or preferential, treatment. VLDB used to handle different regions of the world differently. So papers from some regions got preferential treatment. The community has no special arrangements for authors from other communities.

Answer 7 : No. I consider it healthy for these to change. This broadens ownership and improves on diversity adn helps ensure variation/change.

Answer 8 : Not necessarily. People should be given positions based on their reputation

and track record...

Answer 8.1 : In many cases, those who bid for a conference propose a team to a steering committee or the like. Then a negotiation can be initiated. Many considerations go into the preparation of the initial proposal and any subsequent negotiations. It is often the case that there are special mechanisms for ensuring different diversities and young people when composing PCs.

Answer 9 : I would introduce double blind reviewing for the major conferences.

Answer 10 : I try to be positive when in doubt. I like to reward works where I can tell that the authors (all of them) have done their best. I like to reward work with specific, novel contributions. I like to reward work that demonstrates technical excellence. I consider it essential to be fair and consistent.

Survey Response 5 *at* **Fri Feb 24 22:56:42 2006**:

Answer 1 : 12

Answer 2 : very good but also exhausting

[Question 3 - Ratings]

Answer 3.1 - Diversity (geographical, background-wise) :	**6**
Answer 3.2 - Overall quality of the papers and the conference itself :	**8**
Answer 3.3 - Quantity of accepted vs. submitted papers :	**not rated**
Answer 3.4 - Impartiality (fairness) :	**9**
Answer 3.5 - Other factors - *quality of reviews* :	**6**

Answer 4 : The quantity of submitted papers has increased the degree of randomness in a

paper's acceptance or rejection. PC chairs have tried various ways to improve this but have not been completely successful so far.

Answer 5 : First of all, I didn't understand the meaning of the score in questions 3 and didn't give an. We are judging the acceptace ratio, but next to the score you also give a percentage. Is this the acceptance ratio itself (which doesn't make sense)? Or is it something else, which I can't figure out? As for question 5, I believe it's the right thing to do to maintain roughly the same ratio. The community is larger, more good people write more good papers, we have to accept more in absolute numbers and keep the acceptance ratio the same.

Answer 6 : No hard numbers, but I see new names all the time.

Answer 7 : No, the positions should move around. This is pretty much what is happening in all the major conferences. Steering committees must have people for a few years to maintain some memory, but that's something else.

Answer 8 : Leading positions should go to people who have been successful with the conference and therefore know it well.

Answer 8.1 : General and Program chairs are picked by the organization running the conference, either directly or among competing proposals by the organizers themselves. PC is picked by the pc chair. Other leadership positions by the general and program chair mostly.

Answer 9 : There are many things that can be changed, but there is nothing I believe will fundamentally improve the quality of the process given its scale.

Answer 10 : Not very tough, fair, and consistent. Same for most of my peers, although some don't read papers as carefully as they should and are quick in accepting or rejecting them.

Survey Response 6 Sat Feb 25 04:21:53 2006:

Answer 1 : 15

Answer 2 : Very Good

[Question 3 - Ratings]

Answer 3.1 - Diversity (geographical, background-wise) : **9**

Answer 3.2 - Overall quality of the papers and the conference itself : **8**

Answer 3.3 - Quantity of accepted vs. submitted papers : **1**

Answer 3.4 - Impartiality (fairness) : **5**

Answer 3.5 - Other factors - : **no answer**

Answer 4 : Virtual PC meetings do not help PC members to get a good idea on all papers being discussed.

Answer 5 : Yes, one of the reasons is that the DB field is becoming more and more open to other areas with so diverse applications! The conferences may move to a new scheme: accept more papers but offer presentations to a fraction of them.

Answer 6 : Quite well, I don't see a problem there.

Answer 7 : It is good to not hold executive roles for several years. New people bring new ideas and fresh, new methods to make the conference better.

Answer 8 : No

Answer 8.1 : Based on visibility on the field, and on how well they can run things

Answer 9 : Introduce a better "quality control" method for the reviews. All reviews

should be checked for flaws

Answer 10 : I believe that overall a good job is done. With so many papers to review, in so diverse areas, the community is in general doing a good job.

Survey Response 7 *at* Wed Mar 1 07:24:27 2006:

Answer 1 : 10

Answer 2 : good

[Question 3 - Ratings]

Answer 3.1 - Diversity (geographical, background-wise) : **5**

Answer 3.2 - Overall quality of the papers and the conference itself : **3**

Answer 3.3 - Quantity of accepted vs. submitted papers : **2**

Answer 3.4 - Impartiality (fairness) : **1**

Answer 3.5 - Other factors - : **no answer**

Answer 4 : no answer

Answer 5 : I think that the conferences should accept more papers. I think that an acceptance of 25%-30% would be reasonable

Answer 6 : The research community does not

Answer 7 : No; people who have executive roles should also no be allowed to submit papers.

Answer 8 : No; the community should diversify. It is not healthy to have a restricted set

of authors presenting work on a few topics

Answer 8.1 : no answer

Answer 9 : no answer

Answer 10 : Very fair and when I do not feel confident about my review, I usually say so, rather than simply rejecting the paper.

Survey Response 8 *at* **Fri Mar 3 11:46:08 2006**:

Answer 1 : 10

Answer 2 : good

[Question 3 - Ratings]

Answer 3.1 - Diversity (geographical, background-wise) :	**8**
Answer 3.2 - Overall quality of the papers and the conference itself :	**8**
Answer 3.3 - Quantity of accepted vs. submitted papers :	**2**
Answer 3.4 – Impartiality (fairness) :	**8**
Answer 3.5 - Other factors - :	**no answer**

Answer 4 : The basic issue is that databases are used everywhere, so the boundaries of the subject have expanded. I think this is good, but it does make for extremely specialized papers.

Answer 5 : 17% is about right.

Answer 6 : I think there are many newcomers.

Answer 7 : No, three or four at most.

Answer 8 : Yes, I think that's fair.

Answer 8.1 : I think appointed from on high.

Answer 9 : no answer

Answer 10 : I don't think I'm tough, but I'm fair and consistent. I try to find the best in every paper.

Survey Response 9 *at* **Fri Mar 3 13:03:28 2006**:

Answer 1 : Database data ming papers 19 (2000-2005)

Answer 2 : no.

[Question 3 - Ratings]

Answer 3.1 - Diversity (geographical, background-wise) : **4**

Answer 3.2 - Overall quality of the papers and the conference itself : **4**

Answer 3.3 - Quantity of accepted vs. submitted papers : **2**

Answer 3.4 – Impartiality (fairness) : **2**

Answer 3.5 - Other factors - *Sanity* : **2**

Answer 4 : The core flaw is too much emphasis on relevance and presentation. A decent paper might get three weak accepts and will not go through top conference. That is a flaw. There is a small difference between weak accept and accept. Second, absolute lack of tolerance to 'out of fashion' research. The same work when 'presented' in new vocabulory by epsilon modification of idea by 'biggies' goes through. That is, unjust (insanity) shown towards quality of presentation. The term 'RELEVANCE" should be

thrown out of the review form and so should presentation. Who is a reviewer who is to jud ge relevance of some work? Who knows it will be mighty useful few years down the line when the reviewer is no-more. If you want to criticise -- show errors in the idea/technical aspects, coverage of solution, etc. What is debilitating to new researchers and old researchers is this UNCANNY need to learn new way of presenting the paper to get reasonable reviews. So a researcher instead of perfecting teh idea (or takes weak idea) and mechanically makes it much more presentable by sheer presentation skills. In SIGMOD/VLDB?ICDE the lifespan of a buzzword is 3 years. A good number of papers would be accepted in new areas -- wherein most PC members or regulars have papers. A reasonable good but old problem will be killed for lack of relevance or presentation skills. ---------- I sent the following query to ***** and *** respons eis attached when 'conflict of interest' was speficied. I did not submit the paper. ********** Dear Prof. I have a paper deadling with 'skyline queries optimization & execution' to which primary and secondary areas does it fall. There is no clear area as optimization and processing of queries. Closest areas are either too specific or not having the main focus as this. Second, suppose I know a PC member does not like this area, can I place him/her under conflict of interest? ********** Sometimes the broad areas are so badly worded that your papers are automatically irrelevant. ********* *** response You need to choose one primary area-- pick what's closest. Pick as many secondary areas as you like. You must follow the conflict of interest guidelines (so the answer to your question is no). ---------- So, I lose on relevance & I cannot also get the known reviewer who does not like this area NOT to review my paper. Lose-lose situation. So, I did not submit. ---------- See point 9 for more on this below.

Answer 5 : They say we will accept as many good papers as we can. And the result is the same 12%. See my answe to Q4 above. I no longer really envision submitting a paper SIGMOD or VLDB for just the want of relevance and presentation work (the beautification) that needs to be done. I can use that time to teach 100 illeterate children or adults how to read or write. Unless SIGMOD/ VLDB remove this emphasis it will not work. Look at old VLDB/SIGMOD papers the presentation was not really great, some of the papers were tough to read, but they are ever green papers. Now who cites the papers

on Deductive database, Association rule mining, or object oriented stuff. Majority of SIGMOD papers will have extended last two SIGMOD/VLDB works, that is what teh reviewers have done, understand, and can acertain relevance of work. A very SAD state of affairs indeed.

Answer 6 : I have NSF proposal which is funded. I need 3 VLDB papers, so my students will get those. Now I can get my next grant. Rich get richer and poor students of rich people unless they are extremely bright are disgusted with the whole affair, and go and join some company. It is a cycle funding requires papers, promotion requires papers, paers gives funding, PC chairship gives promotion. How many PC chairs of VLDB/SIGMOD got promoted to Associate professorships or Professorships right after they were pC chairs. You can find the data and check it your selves. Look at the papers of biggies - students just before tehy are up for tenure -- some thier weakest papers will appear in major conferences. Again check it out. You need to join the gang. Check for any new reseacher who has 3 or more papers in any SIGMOD/VLDB/ICDE he would have worked with a biggie just before this phase, and some biggies are involved with his work just earlier to this. I accept your paper you accept my paper. <Do not go by my words, check this connection between biggies and their students, grants, and tenures/promotions, do your home work>

Answer 7 : No. It forms a reason for "I help you, you help me" scenario. The same thing with ACM. VLDB is slightly better. How many PC chairs of SIGMOD are from outsise USA and are from so called wise researchers. If you see ^ above, if you give PC chairship to people who are in it for promotion then you have lost it.

Answer 8 : The problem is the people with mor epapers ar emore specialized. They know that area only and they see other papers through those 'Jaundiced' eyes. The reviwer should be firs of all a person who knows many reas reasonably well. He/She should be a teacher who teaches seminar courses in DBMS and knows old and recent work. And no paper should be given to Graduate student for reviewing. They see - comma and fullstop, and miss the big picture. They need also to be people who look at bigger issues and who

are not narrowly focused on next months applicability of the paper. For that, patents are there which can be submitted.

Answer 8.1 : Based on wide exerience, absolute clarity on quality of work and not thoroughly focused on presentation or relevance.

Answer 9 : I gave following suggestions to both ***** & ******** in 2004, when they asked for suggestions in dbworld. Nothing was done, as expected :-(. ---------- Date: Wed, 09 Jun 2004 15:41:52 +0530 From: removed To: ****** ****** <*****@**.*******.***>, *******@**.********.*** Subject: Conference Reviewing --- Suggestions My suggestions are the following:

1. Anonymous PC-Chair and PC. Only the SIGMOD/ACM Chair or few others (who are not submitting papers and act as governing council) know who exactly is the PC-Chair/PC-Chairs, and PC members.
2. Anonymous PC. The PC-chair does not know who the other members of PC are. They are numbered and no names. That way more rationale criticism of papers takes place during discussion. Also, a more unbiased distribution of papers to PC takes place.
3. Text summarization of abstracts and titles. This way even if the abstract is available through other means, the buddy system of selecting papers would be made difficult. Of course, the papers are blind, no author names displayed anywhere. In fact, only abstracts should be submitted, no titles, titles for accepted papers can be given later.
4. Any PC member or PC Chair reveals the information that he/she is in PC or was PC chair before the conference, should lead to a professional disgrace.
5. Text summarization or some tool used to rewrite the reviews so that identification of reviewers is difficult.
6. With all of the above early reject is good idea. Otherwise, people will do early rejects based on non-pertinent factors.
7. There should be a rebuttal system in place for early rejects. So that some of them can be reconsidered.
8. Declare an acceptance threshold for a paper (like a pass mark), and any paper that gets that level of review is accepted, so inherent quality limit, but no quantity limit. We accept

as many good papers as that get submitted. Since reviews are public, it is difficult for a reviewer to be leniant in grading, see below.
9. Once the list of accepted papers is decided -- the reviews must be put in public domain for accepted papers rebuttal. Sometimes some early related work gets missed out. After this rebuttal phase final list of papers is selected. No rejected paper can be accepted in this phase, but an accepted paper can get rejected.
10. All papers are treated to be pot-pourri session, no quotas for any sub-area. That way no compulsion to fill up sessions in an area with sub-standard papers. Finally, create a mailing list where these suggestions can be further discussed and fine tuned.
Thanks, ********** No mailing list was created and these issues were not discussed. One way to check quality of reviewing is to put reviews of all accepted papers for public perusal. If you relaly think the paper is good why are you so worried about it. BTW, this email was not acknowledged. I thought of going to conference and speaking out. If the people who sought the comments do it so casually --- who will listen to me. Finally, this survey is a good idea -- but I do not expect anything tangible to come out, neither should you. You are all just wasting your time, unfortunately.

Answer 10 : I look at the technical value of the paper, and if the idea is good I am okay with it. I have not used any other considerations, except quality of paper in accepting a paper. Depending on the quality of conference I have been consistent in my reviewing. You attempt is good --- send all this to SIGMOD chair and ACM Chair. As far as I know -- nothing will happen. I hope more than statistical results these comments are made public and announced so that everyone gets to see the real feeling about state of affairs. If you do not do that then your intentions will be questioned. Finally, good luck!!! The real sad part is that from doing research -- many of people have become solid 'paper writers' and 'relevance finders'. They do not work problems they like -- they work on problems that can get them SIGMOD papers. One prominent reseracher told me -- why are you still working in area X, I left it long time ago -- there is no VLDB/SIGMOD paper in that area now.

Survey Response 10 *at* **Fri Mar 3 13:17:44 2006**:

Answer 1 : Over 30

Answer 2 : Fascinating! A PC Chair has incredible freedom which can be used wisely or unwisely. A chair is also given little guidance, our community does not do a good job communicating "best practices".

[Question 3 - Ratings]

Answer 3.1 - Diversity (geographical, background-wise) : **3**

Answer 3.2 - Overall quality of the papers and the conference itself : **8**

Answer 3.3 - Quantity of accepted vs. submitted papers : **2**

Answer 3.4 - Impartiality (fairness) : **2**

Answer 3.5 - Other factors - *Excessive cronyism* : **1**

Answer 4 : Yes. 1) Lack of consistency in the review process from year to year. Each PC Chair is free to "experiment" with new processes, most of which fail. I don't understand why we experiment with our top conferences (SIGMOD/VLDB) rather than trying things out on smaller venues, figuring out what works and then scaling it to our largest venues. 2) Excessive cronyism and no rules tot control this. 3) lack of clear, shared values for how to run an ethical and fair conference.

Answer 5 : I think the major venues have tried to keep the percentage approximately the same in order to maintain a consistent quality. I think this is a good trend. We don't want the top venues to be devalued and we also want top papers to not get in, even if there are a large number of submissions.

Answer 6 : I have not seen any studies on this. Anecdotally, I don't think we do a good job. No venues (except SIGMOD for the last few years) practice double-blind reviewing. As a result, the identity of authors plays a major role in reviewing (positively and negatively). In addition, in my experience on SIGMOD, most PC members do not take

the double blind issue seriously (perhaps because it is not a community standard). They view it as a "hindrance" and the identify of authors is often openly disucssed during the review process.

Answer 7 : Yes, provided the election procedure is a broad and fair one. I think our conferences do not have enough "memory". Having served as PC Chair, I was given little guidance or advice on what works well and what doesn't. We end up experimenting too much with a major venues (especially SIGMOD and VLDB -- PODS has a more stable review process). And often the "experiments", like this year's numerous SIGMOD experiments with a new form of author feedback (different from last year) and having just two reviews for most papers, were not communicated to the community before submission. Way too many decisions get made behind closed doors by one or two people. Importantly, each PC Chair decides for him or herself how conflicts of interest (with PC members and importantly with the PC Chair) are handled, and I don't feel that conflicts are always handled ethically or fairly -- and they certainly aren't handled consistently.

Answer 8 : I believe this is already done.

Answer 8.1 : Each venue does this differently. VLDB has the 21 person VLDB Endowment. The Endowment itself is elected (albeit only by current members). SIGMOD has a small (5-6) person executive where the chair and vice-chair are elected through ACM. Looking at the membership, SIGMOD could clearly benefit from more diversity on its executive. I think it is a mistake for there to be a single, all-powerful, PC Chair as there is in SIGMOD. VLDB has a team of 3-4 chairs which automatically makes people more responsible. I would like to see SIGMOD institute a system where the PC was lead by a team from diverse backgrounds (not a team tht one person gets to choose!)

Answer 9 : Greater emphasis and consistency in the ethics of reviewing. I've seen dbgroups at major universities discuss papers that are under review in their seminars. This is grossly unethical. There is much too much open discussion of confidential reviews (who reviewed what, who killed which paper, who accepted which bad paper

written by their friend...) I would like to see double-blind reviewing become standard for all major venues (as it is in other fields of CS). I would like to see consistent policies implemented for handling conflict-of-interest papers. I also think that one person can no longer run a PC alone. There are too many papers. An important job of the PC chair is to ensure all reviews are fair and thorough. No one person can do this for over 200 papers (which require at least 600 reviews). So this job is no longer being done, hence authors get frustrated when they get inappropriate, non-expert or bias reviews. We study scalability, but as a community have not studied how to make our review process scale. Chairing a PC is a many person job for any venue with over 100-200 submissions!

Answer 10 : One thing you learn being a PC Chair is how to write a great review. When you have to read hundreds of reviews, you get inspired by the deep quality and thoughfulness of some reviewers, and annoyed by the thoughtlessness of others. I think every graduate program should teach seminars in "how to write a great review". To be honest, I wasn't a very good reviewer at the start of my career, and I wish some PC Chairs had taken the time to correct my mistakes. We need PC Chairs who take fairness and consistency very seriously if we want to improve the quality of our reviewing process.

Survey Response 11 *at* **Fri Mar 3 13:21:40 2006**:

Answer 1 : 30

Answer 2 : Good

[Question 3 - Ratings]

Answer 3.1 - Diversity (geographical, background-wise) : **4**

Answer 3.2 - Overall quality of the papers and the conference itself : **5**

Answer 3.3 - Quantity of accepted vs. submitted papers : **8**

Answer 3.4 - Impartiality (fairness) : **5**

Answer 3.5 - Other factors - : **no answer**

Answer 4 : I like the double-blind reviewing and the authors feedback system. It would be good to have it with every conference. I think PC chairs should follow more closely reviewers' work and reject 2-line reviews.

Answer 5 : I do not think we should accept more papers. It would become untractable and may lower the quality of accepted papers.

Answer 6 : This is an issue. I do not have exact numbers but it does not seem to be easy for newcomers to publish papers. Blind reviewing helps. In addition, an approach similar to the SIGIR author mentoring program would be very useful. Newcomers could seek advice from a mentor and get an opinion before submitting a paper to a DB conference.

Answer 7 : I think it is good to change and give a chance to new people also because it is a good way to make sure that a good number of PC members is different from year to year.

Answer 8 : No.

Answer 8.1 : I do not know.

Answer 9 : - Ask the PC chair to check each review and make sure it is meaningful - Make sure authors of accepted papers implement reviewers' comments.

Answer 10 : I always try to give a constructive review and I do not like reviewers who seem to look for flaws in a paper and do not see the potential impact beyond the flaws.

Survey Response 12 *at* **Fri Mar 3 13:25:28 2006**:

Answer 1 : 17- though some would not call them all DB papers

Answer 2 : ok- hard work but interesting

[Question 3 - Ratings]

Answer 3.1 - Diversity (geographical, background-wise) : **7**

Answer 3.2 - Overall quality of the papers and the conference itself : **7**

Answer 3.3 - Quantity of accepted vs. submitted papers : **2**

Answer 3.4 - Impartiality (fairness) : **6**

Answer 3.5 - Other factors - : **no answer**

Answer 4 : Too much emphasis by reviewers on "syntactic correctness"-- e.g. graphs, theorems, etc. And not enough emphasis on importance of ideas. There is a fair bit of randomness in the review process, largely because there is a fair bit of randomness in who reviews which paper. This is why results for a paper can change so much from one conference to the next. Reviewing is a "human" process. We need to be prepared for the fact that it is imperfect. The good news is-- a good paper will be accepted at a conference (eventually). A very bad paper will usually not get accepted at any conference.

Answer 5 : Acceptance rates have gone down to, e.g. 17% or less from 20-25% (I go back a long time). The DB community has grown and we are competing more intensively for conference "slots". Slots have increased-- we are accepting more papers, but not enough to compensate fully for the growth. However, database research is more mature now, and it is really unclear to me that the papers we are publishing are as important as they used to be.

Answer 6 : I do not have figures on this. Instead of doing surveys, you might want to seriously consider doing a study of the rate at which new authors enter our community (i.e. how many are entering per year who have not published in prior years). I think we

are pretty open, but having numbers would quantify this. How about it!!!

Answer 7 : In my viewing of the field, this does not seem to happen to any significant extent. That is "several years in a row". It is true that a fair number of people have run conferences in multiple different capacities-- but almost universally, this has not been "in a row". Frankly, it takes too much work, and most of these folks have "day jobs". These folks are the leaders of the field. It is quite important to the success of a conference that known people have leadership roles. It does take a while to become established-- and to know a large number of people, but this is important in organizing program committees, etc.

Answer 8 : Not automatic leadership positions. People who do publish multiple papers almost universally participate as PC members. But leadership is more than just publishing papers. If someone is interested in a leadership position, then they can talk to the conference organizing committees about making a proposal for a conference. In my experience, organizing committees sometimes have to go out looking for/soliciting proposals. You'd be surprised at how receptive these committees would be to a conference proposal (within the geographic constraints that they work in, such as changing the venue from Europe to N. America to Asia, on a regular basis).

Answer 8.1 : This is public information-- so you do not need me to describe the process. SIGMOD, VLDB, and ICDE all have committees that review proposals from people who are interested in organizing the conference. In some cases, the chair of the proposing group names the PC chairs, in other cases it is the organizing committee that does this. The PC chair then chooses the committee. SIGMOD has general elections. ICDE and VLDB organizing committees elect their own new members. In all cases, there are term limits. In each case, it is a "political" process. That seems unavoidable. This survey has a certain "disgruntled" aspect. As if somehow, some folks should be given a fast track to leadership positions by virtue of having published several papers. But the only fast track is to do important work (which is different from getting papers into conferences). And how does one do that-- and who decides what work is important? We all knew that

Rakesh Agrawal's data mining work was important. There is no process-- usually important work speaks for itself. Rakesh is one of our "leaders", being on both the VLDB and the ICDE organizing committees.

Answer 9 : I'd break the chain that permits people whose papers are rejected at one conference to immediately turn around and submit it at the next. This would reduce the number of submissions substantially (without changing the number of papers). This would reduce the reviewing load per conference, and let us do a more thorough job on the papers. It would also give authors additional time to digest the comments of referees and do more significant revisions before resubmission.

Answer 10 : I think I'm a terrible reviewer :-)!!! The hardest thing is to get reviewers to MAKE A DECISION. Many waffle too much. Many are too impressed by math, graphs, equations, proofs, and experiments. They need to give more thought to what things mean, and what things matter. WRT work habits-- A small number are unreliable-- i.e. don't do the job they commit to or do it in a slipshod manner. But in my experience, most are pretty conscientious in this work, which is very time consuming. Reviewers are probably "self consistent" but there is not great consistency across reviewers. Most try to be fair. Most are not decisive enough, which is different from toughness-- it is a matter of making clear decisions.

Survey Response 13 *at* **Fri Mar 3 20:00:31 2006**:

Answer 1 : 40

Answer 2 : Excellent

[Question 3 - Ratings]

Answer 3.1 - Diversity (geographical, background-wise) : **8**

Answer 3.2 - Overall quality of the papers and the conference itself : **8**

Answer 3.3 - Quantity of accepted vs. submitted papers : **3**

Answer 3.4 - Impartiality (fairness) : **8**

Answer 3.5 - Other factors - : **no answer**

Answer 4 : 1) The double-blind reviewing should be scrapped. It makes for very awkward writing of papers, and also seems to suggest that PC members cannot be trusted to be objective. 2) The conference periods of ICDE, SIGMOD and VLDB should be overlapped so that papers that are rejected at one forum cannot be immediately forwarded to the next forum as is the current practice. If we can submit to only two out of the three conferences, then researchers will be far more careful of sending in quality papers and less inclined to "just take a chance".

Answer 5 : I think it is fine to have a low acceptance rate - acceptance should be based only on quality considerations and it adds to the prestige of the conference and the value of acceptance. I don't think the current acceptance practice should be tampered with in any way.

Answer 6 : I think the process is overall fair, barring the stray case here and there. I don't think there is any need for special welcome of new researchers and their papers - they also need to learn how to fight it out in the competitive research world.

Answer 7 : I don't think it matters with regard to organizational committee members. However, with regard to PC Chair and PC committees, it is good to have a different mix each year to ensure visibility and fairness.

Answer 8 : Yes, nothing wrong in this - it is just a recognition of their leadership status in the research community.

Answer 8.1 : The conferences usually have a steering board that makes the decisions.

Answer 9 : 1) Stop double-blind reviewing process 2) Overlap submission and review process for ICDE,SIGMOD,VLDB series of conferences 3) Keep a cap of at most 6 papers per reviewer 4) No person can be on the PC of more than two of the three conferences

Answer 10 : I think I am a conservative but careful reviewer. I think that many reviewers nowadays, due to the reviewing load, tend to do a superficial job - this can only be addressed by minimizing the load so that a thorough job can be accomplished.

Survey Response 14 **Sat Mar 4 01:58:08 2006**:

Answer 1 : 10

Answer 2 : a lot of work, but helpful to the community

[Question 3 - Ratings]

Answer 3.1 - Diversity (geographical, background-wise) :	**7**
Answer 3.2 - Overall quality of the papers and the conference itself :	**7**
Answer 3.3 - Quantity of accepted vs. submitted papers :	**7**
Answer 3.4 - Impartiality (fairness) :	**8**
Answer 3.5 - Other factors - :	**no answer**

Answer 4 : I don't like anonymous submissions, as is done in SIGMOD. Too much useful information is lost to the program committee.

Answer 5 : no answer

Answer 6 : no answer

Answer 7 : yes, for continuity and to deal with the learning curve.

Answer 8 : they tend to be.

Answer 8.1 : no answer

Answer 9 : no answer

Answer 10 : I feel that I am fair, and that I do a good analysis of the papers to reach a good conclusion about scceptance or rejection.

Survey Response 15 *at* Sun Mar 5 10:37:28 2006:

Answer 1 : 30 plus

Answer 2 : interesting

[Question 3 - Ratings]

Answer 3.1 - Diversity (geographical, background-wise) : **5**

Answer 3.2 - Overall quality of the papers and the conference itself : **3**

Answer 3.3 - Quantity of accepted vs. submitted papers : **3**

Answer 3.4 - Impartiality (fairness) : **4**

Answer 3.5 - Other factors - : **no answer**

Answer 4 : There is no way of garanteeing one has actually gave enough thought to a submitted paper.

Answer 5 : Unless this is done, the number of papers accepted will be no indication of quality. When you get to 1/10 papers accepted out of 500+ submissions all it takes to

reject a paper is one badly done review.

Answer 6 : I think that's not a problem.

Answer 7 : Absolutely not. Different people will help avoiding bias in the process.

Answer 8 : Yes.

Answer 8.1 : A larger body (Executive Committee) usually invites people to be PC chairs, General Chairs are usually chosen based upon some kind of application (or show of interest) to host the conference.

Answer 9 : Have always a double-blind review and with sub-committees. A single PC model, IMHO, doesn't work anymore.

Answer 10 : I try to do a good job, and that costs me time, hence I'm accepting less and less PC invitations. As for my peers it seems that sometimes they're just after what's "wrong" with the paper, once it's found reject. There seems to not have any concern as to whether that "problem" is major or not, i.e., whether it could still be a useful research result.

Survey Response 16 *at* Sun Mar 5 15:41:14 2006:

Answer 1 : 7

Answer 2 : no answer

[Question 3 - Ratings]

Answer 3.1 - Diversity (geographical, background-wise) : **8**

Answer 3.2 - Overall quality of the papers and the conference itself : **6**

Answer 3.3 - Quantity of accepted vs. submitted papers : **7**

Answer 3.4 - Impartiality (fairness) : **4**

Answer 3.5 - Other factors - : **no answer**

Answer 4 : The workload is high on each pc member; hence many members ask for help from unexperienced graduate students, and do not have time to shephard or mentor in the review process. One problem with unexperienced reviewers is that they lack the depth and the vision to recognize revolutionary papers. As we look at the acceptence rate for delta papers, this is pretty evident.

Answer 5 : Last year's SIGMOD (2005) was an interesting exercise, which did not limit the no of papers to be accepted but made decisions on quality only. Nevertheless at the end about the same number of papers were accepted. I think the acceptance rate is about right to maintain the quality of the conferences.

Answer 6 : Do not have these statistics.

Answer 7 : These roles should rotate and they do. Executive committees change less frequently, and it may be good to have rotation but it is also good to have same people in consequitive years to establish and maintain strategy and provide continuity.

Answer 8 : One year is not a good indicator.

Answer 8.1 : no answer

Answer 9 : I would only have pc members who will review the papers themselves, or be involved in the process if they resort to external reviewers.

Answer 10 : no answer

Survey Response 17 *at* Mon Mar 6 05:32:57 2006:

Answer 1 : 40

Answer 2 : Very good

[Question 3 - Ratings]

Answer 3.1 - Diversity (geographical, background-wise) : **7**

Answer 3.2 - Overall quality of the papers and the conference itself : **9**

Answer 3.3 - Quantity of accepted vs. submitted papers : **2**

Answer 3.4 - Impartiality (fairness) : **5**

Answer 3.5 - Other factors - : **no answer**

Answer 4 : Conferences are being used as the main publishing medium, which stresses them and the PC members. They do not need to be as selective as they are, which filters out some good conceptual papers in favor of papers that are better developed, but more marginal. The review process can be opened up to more author feedback. Along with reduced review loads, this would improve the quality.

Answer 5 : Yes, the acceptance rates have stayed the same. With increasing submissions, this means that more papers have been accepted. Given fixed time and resources, it is hard to increase acceptance rates, unfortunately.

Answer 6 : I don't know the statistics of first-time author acceptances. Typically, the new researchers are not distinguished as a group in the evaluation process.

Answer 7 : No, it is not good and typically, this does not happen at SIGMOD and VLDB conferences.

Answer 8 : Yes, but over a longer time period than a few years. As researchers have more publications accepted at the major conferences, they slowly assume leadership positions and are given these roles.

Answer 8.1 : For SIGMOD, the SIGMOD Executive Committee, with input from SIGMOD Advisory Committee, makes the decision on the PC Chair. Typically a short list is developed and discussed in detail before a name is selected. PC Chair then selects PC members, which is scrutinized by SIGMOD Chair for diversity (geographic, gender, etc). General Chairs are usually people who volunteer to hold the conference at their sites. For VLDB, the General Chairs are selected the same way. The PC Chair(s) are proposed as part of the proposal that is considered by the VLDB Endowment Board. The final names are dependent on the Board's approval. ICDE works similarly to VLDB, as far as I know.

Answer 9 : I like the experiments that SIGMOD is involved in: double-blind reviews, author feedback, etc.

Answer 10 : I am tough, but fair and try to be consistent. Overall, I am satisfied that the review process works as well as it possibly can.

Survey Response 18 *at* **Tue Mar 7 07:49:02 2006**:

Answer 1 : 4

Answer 2 : Have not served

[Question 3 - Ratings]

Answer 3.1 - Diversity (geographical, background-wise) : **3**

Answer 3.2 - Overall quality of the papers and the conference itself : **5**

Answer 3.3 - Quantity of accepted vs. submitted papers : **3**

Answer 3.4 - Impartiality (fairness) : **5**

Answer 3.5 - Other factors - : **no answer**

Answer 4 : no answer

Answer 5 : no answer

Answer 6 : no answer

Answer 7 : no answer

Answer 8 : no answer

Answer 8.1 : no answer

Answer 9 : no answer

Answer 10 : no answer

Survey Response 19 *at* Wed Mar 15 10:07:31 2006:

Answer 1 : 35

Answer 2 : 10 (very positive)

[Question 3 - Ratings]

Answer 3.1 - Diversity (geographical, background-wise) : **8**

Answer 3.2 - Overall quality of the papers and the conference itself : **7**

Answer 3.3 - Quantity of accepted vs. submitted papers : **6**

Answer 3.4 - Impartiality (fairness) : **8**

Answer 3.5 - Other factors - *organization* : **8**

Answer 4 : The number of conferences has substantially increased over the past years. This has caused some difficulty to assemble good and homogeneous program committees. Experienced reviewers are always committed to some conference. Thus, sometimes pc chairs have to rely on non- experienced pc members. This has had a bad impact on the reviews received by authors.

Answer 5 : This is true. Important DB conferences such as SIGMOD, VLDB, and ICDE have an very low acceptance rates which usually prevents good papers from being accepted for presentation at these conferences.

Answer 6 : This is a very difficult question to answer considering conferences in general. However, I can tell about my own experience with JCDL - the IEEE/ACM Joint Conference on Digital Libraries. Recently I moved my research interest to DLs and I had no problems so far to get involved with the DL community and to have my papers accepted at JCDL.

Answer 7 : From the experience view point it is. Experienced people are usually a guarantee for a good conference organization. On the other hand, bringing fresh people to hold some executive roles is very important to encouraging new ideas. Thus, I believe we should have a balacing number of experienced people and newcomers holding executive roles in our conferences.

Answer 8 : I think so. They might bring new ideas and also their experience with submission and reviewing processes of that conference.

Answer 8.1 : In general, conferences have a steering committee that selects those who will play the major roles in the conference organization (PC Chair, General Chair, Workshop Chair, etc.). Then, the corresponding program committes are assembled by

their respective chairs.

Answer 9 : I believe authors should be given the opportunity to answer the reviewers´ comments on their papers before the end of the reviewing process, particularly when one specific reviewer has a strong but not very conclusive opinion against them. However, this might increase the reviewing time considerably. SIGMOD has tried this policy recently, but I don´t think it worked properly. Improvements should be made in this process in order to rise the quality of the accepted papers and give a better chance to young researchers not used to large conferences. Another suggestion would be to assist young researchers in the preparation of their papers for submission. SIGIR has such a policy.

Answer 10 : I rate myself as a good reviewer. When I was a pc member for VLDB´98 I was among the 5 best rated reviewers. Wrt my peers, in general I have received good reviews, but from time-to-time I have got very inconsistent reviews which usualy express a strong position of the reviewer against the submitted paper. Reviews like these are very disappointing and do not make any contribution to the authors.

Survey Response 20 *at* **Sun Mar 19 10:01:49 2006**:

Answer 1 : 23

Answer 2 : no answer

[Question 3 - Ratings]

Answer 3.1 - Diversity (geographical, background-wise) : **6**

Answer 3.2 - Overall quality of the papers and the conference itself : **8**

Answer 3.3 - Quantity of accepted vs. submitted papers : **4**

Answer 3.4 - Impartiality (fairness) : **5**

Answer 3.5 - Other factors - : **no answer**

Answer 4 : SIGMOD 2006 provided only 2 reviews for many papers

Answer 5 : this has led to many good (acceptable) papers being rejected. this is not healthy for the area.

Answer 6 : no answer

Answer 7 : it is important to for these to change regularly.

Answer 8 : yes.

Answer 8.1 : the executive committee chooses these.

Answer 9 : - give authors a chance to respond to reviews - build memory into the system---if a paper is re-submitted, the previous reviews should be included, together with the authors answers to those

Answer 10 : i am often less tough when i get papers that have very new (revolutionary) ideas. i think we are getting fewer of those as time goes by. people have strict notions of what a "SIGMOD" paper should look like, and some novel ideas/concepts often do not fit the accepted molds. again, this is not healthy for our area.

Survey Response 21 *at* Sun Mar 19 18:27:01 2006:

Answer 1 : about 50

Answer 2 : good

[Question 3 - Ratings]

Answer 3.1 - Diversity (geographical, background-wise) : **7**

Answer 3.2 - Overall quality of the papers and the conference itself : **8**

Answer 3.3 - Quantity of accepted vs. submitted papers : **5**

Answer 3.4 - Impartiality (fairness) : **7**

Answer 3.5 - Other factors - : **no answer**

Answer 4 : no answer

Answer 5 : an increase in the number of paper submissions tends not to be correlated with an increase in their quality. if we assume that program committee members are being intellectually honest in their opinions about the overall quality of papers (and there is no reason to doubt that), the papers that are being rejected are those that at least one reviewer found somewhat lacking in overall quality.

Answer 6 : no answer

Answer 7 : it's good to do it for a couple of years - the first year when you are learning how to do it properly and the second when you are teaching someone else how to do it properly.

Answer 8 : being a good researcher (i.e., having several papers presented at the conference) is not necessarily correlated with being a good conference organizer. it's good to have both qualities before being considered for a leading position at a conference.

Answer 8.1 : no answer

Answer 9 : no answer

Answer 10 : no answer

12.3 Appendix 3 – DB-HEALTH Survey Feedback

Below you can find the feedback drawn from the past survey takers. Some of them mention interesting ideas:

I'm not sure in what way your survey will provide information on the health of database research. There are many dimensions to this question, not simply the state of database conferences.

Good you are doing the survey. Let us hope the people who ought to pay attention and take action -- know the comments made in this survey and actions are taken appropriately.

Unfortunately, the publications in major database conferences determines which way your career is headed. So, this survey tackles the core of the problem. We should get away from the notion of relevance of database research. We do research to solve problems we think are important. As long as it is formulated well and solved to completion with stated limitations, it should get into the major conference. It SHOULD NOT be curtailed by relevance score. Majority of papers are shot down by relevance scores in major conferences. Research is exchange of ideas/techniques and results. Relevance is the last thing in a researcher's mind. Passion for the problem being solved should be paramount. Unfortunately modern conference publishing is putting relevance at fore-front and passion possibly at the last or not even considered.

I think you should seriously consider doing a study analysing the entry of new people into our field, based on the number of authors who have not published previously getting

papers into major conferences. It is surely a doable study-- and more effective than asking about it in a survey. The "health" aspect suggests that there may be something seriously wrong with the field. While we are not perfect, we are, in my assessment pretty much doing ok. I hope you will publish the results of the survey.

I am a researcher with PhD in early 1990s under the guidance of one of the very acknowledged database system researchers. 1. What is worrying is the mismatch between number of papers submitted and accepted. Even assuming that about 60% of papers are absolutely unacceptable, there are still 40% papers with some solid work, and at least 25% (overall) of papers that can be accepted outright. 2. In many top conferences the instructions to reviewers are we will accept about 15% of papers, so review the papers accordingly. It is not try to review the papers and let the PC chairs know which of have taken a problem, addressed it, and provided evaluation of the solution. 3. Another worrying factor is the number of good database researchers who consistently cannot publish in SIGMOD/VLDB because of the problems they are addressing or the work they are doing does not fit into top 15% but comes in top 25%. Pretty soon they stop attending the conferences and DBMS community loses yet another good researcher. 4. ICDCS is one conference that I found has long review time, still has sessions in many of the original areas. My experience with this conference even when papers got accepted or rejected has been great. I have not been able to get the same kind of confidence in any of the top database system conferences when papers were accpeted or rejected. The core belief of something 'fishy' going on lingers. And believe me this moves you out of the DBMS community for good. 5. There are areas of DBMS most notably conceptual modeling which does not get any respect in major conferences. It is extremely difficult to get a ER++ paper into SIGMOD. Similarly, every three years SIGMOD/VLDB/ICDE spawn of one sub area of to another conference. And the whole top layer of people move onto different problem set and this loop repeats. Think deductive dbms, think object-relational, think schema integration work that has remanifestated itself with logical framework of materialized views, think histograms for selectivity estimations, think XML. 6. Corporatization of conferences. Early and middle 1990s it was tough for

SIGMOD/VLDB to breakeven. Corporate sponsors were needed, since then there is a need to 'justify' the relevance of conference. Otherwise, the corporates won't fund (the managers who sponsor need to be justified why this conference and work presented there is relvant to what the company is doing) -- this meant the traditional researchers who do reserach for their passion (on other research problems) should either join the band wagon of corporate research or publish elsewhere. This is more of the case with DBMS and less with other area conferences. It was interesting one of the papers was shot down because it was not relevant - why because ORACLE/DB2/SQLSERVER does not do its query processing in this manner and your solution cannot fitwell. The question is should we decide on the fact these are only DBMSs that will exist for ever and only solve problems that can fit into these products. If so, make those products and source-codes open so that all researchers can contribute to it. Now there are huge research groups in major corporations doing reserach they have to show they are doing good work -- the corporatarization of the research manifests itself in not only sponsoring conferences but also getting large number of papers in. Coporates bring in best researchers who decide on relevance of others work based on whether it fits into their game plan of what needs to be done. Unless, there is drastic change in the way ACM is structured and the way in which conferences are conducted, and the completely undue importance given to publishing in top conferences, things will not change. There has to be respect given to the problem being solved, and whether it has been solved reasonably, it has been evaluated, and its limitations stated. If all these are done - accept the paper. period. do not look into other aspects. If this community wants to do a thorough introspection let us have a full scale discussion on it, instead of saying all is healthy and fine.

12.4 Appendix 4 – SIGMOD, PODS, VLDB Committee Websites

12.4.1 SIGMOD

1997 – SIGMOD 1997 Proceedings Book (no online reference found)

1998 – Organizing Committee, Program Committee:
http://www.sigmod.org/sigmod/record/issues/9806/index.html

1999 – Organizing Committee, Program Committee:
http://www.sigmod.org/sigmod/sigmod99/eproceedings/index.html

2000 – Organizing Committee, Program Committee:
http://www.sigmod.org/sigmod/sigmod00/eproceedings/index.html

2001 – Organizing Committee, Program Committee:
http://www.sigmod.org/sigmod/sigmod01/eproceedings/index.html

2002 – Organizing Committee, Program Committee:
http://www.sigmod.org/sigmod/sigmod02/eproceedings/index.html

2003 – Organizing Committee, Program Committee:
http://www.sigmod.org/sigmod/sigmod03/eproceedings/index.html

2004 – Organizing Committee
http://www.sigmod.org/sigmod/sigmod04/eproceedings/content/organisation.html
Program Committee:
http://www.sigmod.org/sigmod/sigmod04/eproceedings/content/committees.html

2005 – Organizing Committee, Program Committee:
http://cimic3.rutgers.edu/~sigmod05/SIGMODcfp.htm

2006 – Organizing Committee, Program Committee:
http://tangra.si.umich.edu/clair/sigmod-pods06/

12.4.2 PODS

1997 – Organizing Committee, Program Committee:
http://www.sigmod.org/sigmod/pods/podsorg.html

1998 – Organizing Committee, Program Committee:
http://www.sigmod.org/sigmod/pods/podsorg.html

1999 – Organizing Committee, Program Committee:
http://www.sigmod.org/sigmod/pods/podsorg.html

2000 – Organizing Committee, Program Committee:
http://www.sigmod.org/pods/proc00/

2001 – Organizing Committee, Program Committee:
http://www.sigmod.org/pods/proc01/

2002 – Organizing Committee, Program Committee:
http://www-db.cs.wisc.edu/sigmodpods2002/organization.html

2003 – Organizing Committee, Program Committee:
http://www.db.ucsd.edu/SIGMODPODS03/PODSorg.html

2004 – Organizing Committee, Program Committee:
http://www.sigmod.org/sigmod/pods/proc04/content/committees.html

2005 – Organizing Committee, Program Committee:
http://cimic3.rutgers.edu/~sigmod05/PODScfp.htm

2006 – Organizing Committee, Program Committee:
http://tangra.si.umich.edu/clair/sigmod-pods06/PODS-org.htm

12.4.3 VLDB

1997 – Organizing Committee

http://sunsite.informatik.rwth-aachen.de/dblp/db/conf/vldb/organizers97.html

Program Committee:

http://sunsite.informatik.rwth-aachen.de/dblp/db/conf/vldb/pcmembers97.html

1998 – Organizing Committee

http://www.research.att.com/conf/vldb98/organization.html

Program Committee:

http://www.research.att.com/conf/vldb98/pcmembers.html

1999 – Organizing Committee

http://www.dcs.napier.ac.uk/~vldb99/vldb99.htm

Program Committee:

http://www.dcs.napier.ac.uk/~vldb99/PCMembers-99.html

2000 – Organizing Committee

http://vldb.org/archive/vldb2000/keyofficials.htm

Program Committee:

http://vldb.org/archive/vldb2000/pcmember.htm

2001 – Organizing Committee

http://www.dia.uniroma3.it/vldb2001/officials.html

Program Committee:

http://www.dia.uniroma3.it/vldb2001/cfp.html#PC

2002 – Organizing Committee

http://www.cs.ust.hk/vldb2002/organization/officers.html

Program Committee:

http://www.cs.ust.hk/vldb2002/organization/pc.html

2003 – Organizing Committee

http://www.vldb.informatik.hu-berlin.de/org_offic_conorg.html

Program Committee:

http://www.vldb.informatik.hu-berlin.de/org_progcom.html

2004 – Organizing Committee

http://www.cs.toronto.edu/vldb04/officers.html#Conference%20Organization

Program Committee:

http://www.cs.toronto.edu/vldb04/db-pc.html

http://www.cs.toronto.edu/vldb04/is-pc.html

http://www.cs.toronto.edu/vldb04/indust-pc.html

http://www.cs.toronto.edu/vldb04/workshop-pc.html

2005 – Organizing Committee

http://www.vldb2005.org/organization/orgcomm.php

Program Committee:

http://www.vldb2005.org/organization/progcomm.php

2006 – Organizing Committee

http://aitrc.kaist.ac.kr/~vldb06/organization.html

Program Committee:

http://aitrc.kaist.ac.kr/~vldb06/programCommittees.html