Title of Paper: Observation on Database Research Trends via Publication Statistics

Authors: Kuo-Wei Hsu, Amanuel Godefa

Reviewer Team (Name): Stuart Ness

Date Review Completed: 11/12/2006

SUMMARY:

FOCUS:

Does the paper clearly identify the problem it is addressing?
The stated problem is to classify the research from VLDB and SIGMOD to produce a trend of database research in the past 5 years through a statistical measure.

Does the paper clearly explain related work and their limitations?
The paper presents a number of survey based papers that present current research issues as of 2003 to 2004. The only note that is presented on their limitations is that these papers have not been supported by statistics. Another issue that should possibly be mentioned is that there have not been any new directions in the past two to three years. Another limitation of the previous work may be the difficulty in determining popular areas because of the need for expert knowledge to determine interest. The proposed method may provide an easier method to determining interesting areas.

Does the paper identify its key contributions?
The paper's key contribution is suggested to be a projection of the most popular research areas. The paper also attempts to claim that this method is able to show interesting areas in research.

Does the paper present any evidence to support the contribution claim?
The paper uses statistical evidence to show popular topics, however, it lacks any comparison between the more traditional survey approach to popular areas to the statistical method. One method of evidence which could be beneficial would be to provide a comparison between the statistical results and the database research surveys.

TECHNICAL EVALUATION:

Is the literature survey complete?
Because of the extremely wide scope of the article (covering all database research topics) there are some other possible areas which could be explored. There is a fair amount of work in surveys of particular areas, but there is not a broad over-reaching article used to pull all the database areas together. In addition, the latest source is 2004.

Is the work novel relative to the literature? Explain.
The work is novel in the sense that it provides a trend analysis of database research and a broad classification of current research interests. However, because it simply uses a counted paper number, it does not offer any new statistical method for determining the information, and it does not provide any comparison to survey methods to show which areas are popular areas of work.

As a reviewer do you agree with the contribution claims? Explain.
I agree that they have created a trend analysis, but they have not taken the work far enough to show how it truly shows popular areas of focus. In addition, it lacks complexity which would make the solution interesting. It still requires someone to examine the data to “evaluate” the popular areas. Adding a section or method providing some statistical quantifier would benefit the ability to read the results and show why outliers exist within the data sets.
READABILITY AND ORGANIZATION:

Is the paper easy to read and understand to students in this course (Csci 8701)?

The subject of the paper is easy understandable to students in this course. There are some grammatical errors which detract from the readability of the paper, primarily in the early paragraphs, however, the majority of the paper is relatively easy to read, and conveys the information fairly well.

Is the paper self-contained?

The paper is self-contained. However, some explanation of the different areas of research, what each area consists of in a broad manner would be beneficial for the naive (possibly intended) audience.

Is the paper length reasonable?

The paper is of reasonable length, however, extending the paper by another 2 pages of text would be beneficial if adding some discussion of how the results compare to existing surveys, extending the trend analysis to consist of ten years as opposed to five years, and provides some explanation of the different areas.

Does it include sufficient number of figures and tables?

The paper provides two tables, which show the counts of papers from the two areas. Instead of presenting these two tables, it might be beneficial to present one graph that shows the topics of each conference separately, and together, and also a figure that shows the areas of interest from the survey based literature review. If the tables are kept, they should be put into an appendix table.

STRENGTHS:

What are the strengths of this paper?

The strengths of this paper are that it provides an easy to use method to discover what have been popular focuses of topics that have been published in a particular journal or conference in the past few years. It presents it in a quick manner that can be used to show what are open areas, what are hot topics, and what areas are slowly not becoming as interesting within research work. In addition, it provides a method which can be used for a more dynamic outlook over needing a domain expert to classify work and give an expert opinion, which may be biased.

AREAS FOR IMPROVEMENT:

How can this paper be improved? If you were to rewrite this paper, what revisions would you consider?

A few things that you may wish to consider for improving this paper:

1. Compare results of the statistical method to the results from the literature survey. (Shows validation that your method performs as good as the traditional method.) Explain why your results are superior to the literature.

2. Present a method to explain the statistical survey. Use graphs rather than tables to present the information. Consider using some method of time variant statistics to explain the information.

3. Expand the time period to be from the past 10 years rather than the past 5 years. 5 years doesn't provide much trend in the scope of this work.

4. Consider narrowing focus on areas of interest, such as query processing. There are many topics within each area. It may be beneficial to say which areas within those are interesting in current literature.

5. Explain how category labels are applied to each work. Is it arbitrary, or could it be easily replicated by someone without domain knowledge.