CS848 Paper Submission Site

Go To CS848 Site

Submit or Update A Review For Paper #18

Go To CS848 Submissions

It is currently Wednesday 24th of November 2004 04:42:50 PM EST

Your review has been finalized!

Paper # 18 (Download paper of type application/pdf, 1800432 bytes)	
Title:	Improved Histograms for Selectivity Estimation of Range Predicates
Abstract:	

You have already finalized your review for this paper.
You can no longer modify it, but you may view it.

If you made a mistake in your review and you want it "unfinalized",
you may send mail to the program chair asking them to unfinalize it

Send yourself this review by email

Attribute	Value
Are you finished with this review?	Finalize, I am done editing
Provide a short summary of the paper	This paper presents an overview of histogram concepts and discusses several different histogram types. An overview of novel ways of generating these types of histograms is also presented. The paper begins by discussing the various concepts that form the basis of a histogram: domain, splitting methods, etc. The authors provide a good mathematical background for histograms. An overview of the topics that will be discussed in later sections of the paper are presented. Equi-width histograms partition the domain into buckets of the same size. Thus, each bucket contains an equal portion of the possible domain values. Frequencies within each bucket can change to represent the values that fall within that bucket. Equi-depth histograms partition the domain into buckets of varying size. Each bucket must contain the same frequency value. Thus, bucket widths can change to represent a variable portion of the domain. Spline-based histograms attempt to represent the frequency of values within the range as a piecewise linear curve. Although these types of histograms have not been applied to database systems yet, they have solid foundations in numerical analysis. Unfortunately, they are computationally expensive to construct and maintain. V-optimal histograms attempt to group contiguous frequencies into groups in order to minimize the overall variance in frequency values experienced between buckets. These histograms require a large amount of overhead to build as all possible simple histograms must be enumerated, taking exponential time.

	End-biased histograms place some number of highest and lowest frequency values in their own buckets, and represent all other frequencies in a single bucket. This requires an exhaustive enumeration of possible frequencies as well, but can be computed in slightly over linear time. The authors next present some sort orderings and discuss some new trends in histogram generation. Some of the new proposals are compressions in frequency and area, max difference of frequency and area, etc. The authors next turn to a discussion of the parameters that must be calculated to build histograms. Some techniques for generating quantiles, frequencies, number of distinct values, and the spread of each column are discussed. A treatment of the required sample size for generating histograms and parameters is presented, as well as the time taken to compute various presented histogram types. Before concluding, the authors present a comprehensive performance analysis that shows, among other things, the relative error experienced using different histogram types, and the frequency skew.
What is the strength of the paper? (1-3 sentences)	This paper presents an overview of the different types of histograms that have been proposed and are in use in many database management systems. A thorough performance analysis of each type, as well as comparisons between types, is presented.
What is the weakness of the paper? (1-3 sentences)	This paper does not present sufficient detail or the strengths of different techniques. The mathematical background is not organized and difficult to follow.
Your qualifications to review this paper	I know the material, but am not an expert
Writing Quality	Good
Relevance to query processing?	The paper is relevant to query processing
Experimental Methodology	**Good**
Novelty of paper	Incremental improvement
Overall paper merit	**The paper is a novel or new contribution with average/weak methodology, or an incremental contribution that has good methodology. Someone in the area should read it**
In your opinion, will this paper be important over time?	Average
Provide additional detailed comments to the author	You have presented a comprehensive overview of histogram concepts and types. Your paper is generally well written and organized. I have a few comments for improvement. -Mathematical basics were difficult to follow and confusing. Better formatting and explanations would have been useful. -Why are none of the algorithms described in section 4 appropriate to use for range query result size estimation? -What are some of the strengths of the old algorithms? Why are your new algorithms better than the old ones? -Better descriptions of the new techniques is required. Not enough information about them to determine their usefulness has been presented. Overall, well done. Your performance analysis was quite comprehensive and your writing style was good.

Additional comments to PC (not seen by author)	**None.**

Goto Main Index Close Window