Foundational Paper on Referential Bases

Paul S Prueitt PhD
October 31, 2006

In this short paper, the key-less hash table is compared with the hash table. The nature of a hash table is described in simple language. It is shown that the key-less hash table is the optimal realization of the well known abstract concept of a hash table. It is also, secondarily, shown that the key-less hash is the most important foundational tool for a developing ".vir" standard. In particular we show that the family of hash tables is essential for the technical constructions called RIBs (referential information bases) and Orbs (Ontological referential bases). [1] It is suggested that these constructions will replace the now very common relational database systems. Additional details are given in the footnotes.

Optimality might be proven in the context of a specific set of definitions that are developed as part of the ".vir" standard. The notion of optimal algorithms is one that is difficult, unless (conjectured) differences between computational processes and "natural processes" have been agreed on. In the second school viewpoint, we adopt the viewpoint of Robert Rosen that abstraction involving human thought produced the notion of mathematics and algorithms and all specific algorithms, and this production of abstraction is less than the category of natural phenomenon.

A proof technique, shown by Dr Harold Szu to myself, was first given at a private meeting at George Washington University in April 1998. The proof involved advanced dynamical theory. Using a similar, but simple, argument optimality for the key-lass hash may be demonstrated. The details have to be discussed in other situations. [2]

The point to the optimality arguments is two fold. The mere fact of provable optimal computational processes is forecasted by this 1998 theorem by Szu. The fact that any category of computational process can be implemented poorly, less poorly and then provable optimal suggests that there is an end to the innovation in the foundations of computer science. This is good; because provable optimality would mean that eventually all algorithms will be public domain. [3] How the optimal technique is used may in fact be patentable in any of a thousand ways. It is thus claimed that the concept developed by Gruenwald and patented in 2002, is one that is fundamental. The larger problem is how to support needed innovation in information science itself.

Optimality weakens an actual patent, as demonstrated in the Apple verses Microsoft lawsuits. For example, the "desktop" concept for building an human computer interface was judged to be not patentable based on the notion that the concept of a desktop was optimal as a human-computer interface and no further deep innovation could occur. So far the Gruenwald patent has not been challenged.

In any hash table, a fixed virtual or physical memory space is allocated based on the range of the hash function. The range of any arithmetic function is the set of numbers that are possible outputs. The hash function is a simple "modulus" arithmetic function that divides by a fixed integer and then places a pointer to that part of the virtual or physical space associated with the remainder after division. The input to the hash function is a string of letters. The string is converted an integer using some standard, reproducible, assignment.

Given any hash function, the mapping between a string and a number is well defined, so a second occurrence of the same string is mapped to precisely the same integer. In the mapping, any natural order to the string may be removed without affecting the abstract construction, since there is simply a one to one correspondence implicit in the rules/formulas. In this sense the hash function is "generative" in precisely the way that fractal image compression and the production of generative digital objects are generative.

In fractal compressions, the generation of the image (decompression) is an iterative process that uses a transform matrix on the colors and positions of patterns. The generator "creates" the output rather than retrieving a value from a look up table. This "generative" process allows the insertion of book keeping code and thus the instrumentation of the expression of the generative objects remains consistent with the ".vir" standard.

The ".vir" standard is simply a specification of the functions of a class of generative digital objects. The standard is needed so that the instrumentation that is used can provably protect all digital property moving within the ".vir" subnet. [4]

The digital rights management properties of the ".vir" standard can be achieved using any reasonable transaction operating system (TOS) such as the ".vir" proposed Acappella Software Inc technology (or even .net or J2EE). However, if the class of hash tables is not used, and used in an optimal fashion; any global transaction system will failure to scale up to billions of transactions per second with no failures.

The failure potential is present any time there is an essential index that is separate from the actual encoded data. In the relational database management systems there are layers and layers of indices. In keyless hash tables we have simpler structure because the core need for indices is remove. The future development of a category based technology builds from the current standards on web-based ontology. Orb and Ribs replace relational database systems as means for providing structure to data.

The simplification of access to records is so profound that the proof that the ".vir" system is ultra stable and provably secure is immediately drawn from an understanding of the nature of the set of positive integers. There is a structural feature related to what Gelhard and Prueitt referred to as "structural holonomy" . [5] This structure feature at first may appear as a spoiler of the possibilities, but on investigation one finds remarkable bypasses to difficult technical issues such as scalability.

The structural holonomy is a "single laying down" of the bins of the keyless hash. In the simplest version, location of the bins is ordered by treating the strings as if a base 64 number. The keyless hash function simply regards "Sam" as equal to $45*(64)^2 + 1*(64)^1 + 13*(64)^0$. In a non-keyless hash function, for example, "Sam" might be assigned the number $26 + 19 + 1 + 13 = 59$ based on the alphabetical order of letters in English and the convention that capital letters are ordered after the small letters. Under this assignment scheme "Rbm" would be given the same number as "Sam". In the base 64 keyless hash, "Rbm" is equal to $44*(64)^2 + 2*(64)^1 + 13*(64)^0$. So "Rbm" is not equal to "Sam". Note also that any six-digit number is greater than any five-digit number.

Because there is no index; there can be no add or delete of a bin. The whole cannot be modified without a re-laying down of the entire information space. This turns out not to be a problem, in fact this need to re-write an entire information space is the open door to category theory, and formative ontology. [6] A powerful conceptual relationship between the Orb technology and how natural events are formed is also established, since any birth or emergence of a "something" has to be treated as a "whole".

Let us now look at the limitations on the non-keyless hash table, and this notion of instant search that does not need an index.

In any of the class of hash tables, a memory space is developed by allocating a fixed size, in bytes, for a "hash bin". In the non-keyless hash tables a standard bin is associated with each integer from 0 to n-1 where n is the largest number in the output. This memory space is either physical, having specific hardware allocation of space, or has an allocation that is created using the nature ordering of integers.

The non-keyless hash table suffers in two immediate ways, and does not allow immediate index free search. The first limitation on non-keyless hash tables is that there are collisions where more than one string is mapped to the same place, as in "Rbm" and "Sam" in the example above. The second is that to avoid collisions the hash table memory allocation has to have many more bins than what are needed.

An index free-search is due to an ability to go physically or virtually to the middle of the linearly ordered bins where information has been placed. The question can be asked and answered in a single machine cycle, "is this search string smaller or larger than the string that is in the middle bin". If the answer is "smaller" than one can immediately go to the middle of all the linearly ordered bins to the right of the middle bin. If the answer is "larger" then one can immediately go to the middle of all the linearly ordered bins to the left of the middle bin.

There is a similar search process in non-keyless hash tables called "binary search". However, with the foundational Orb and Rib theory the identification of where to move is directly targeted on the informational elements that are the text elements when regarded as strings. The use of binary search on the containers of a keyless hash table is part of a reduction in complicatedness resulting in scalability features that are not present in indexed systems [7] .

This process stops after n-1 steps where n is the smallest integer where 2 to the power n is larger than the number of bins in the linearly ordered set of bins, or the strings are exactly the same. This means that search is completed in twenty machine clock cycles when the information space has 1,048,576 primary elements.

In thirty machine clock cycles a Orb or Rib having 536,870,912 informational elements can be searched. This speed is independent of the size of the individual elements, as long as standard bins sits in a linearly ordered memory space (physical or virtual).

For the record, in theory the virtual space can be the information encoded in two electromagnetic pulses. This theoretical prediction places the keyless hash as a model for memory recall in human brains, based on the work by myself as an extension of Karl Pribram's [8] neurowave equation and on related new works on harmonics of oscillators. [9]

---

[1] URLs for Orbs and Ribs are scattered in the OntologyStream Inc and BCNGroup Inc web sites, in particular:
http://www.bcngroup.org/area1/2005beads/GIF/RoadMap.htm and
http://www.ontologystream.com/OS/clientTechnology.htm

[2] Prueitt, Paul S. (1996a). Is Computation Something New?, published in the Proceedings of NIST Conference on Intelligent Systems: A Semiotic Perspective. Session: Memory, Complexity and Control in Biological and Artificial Systems. October 20-23.

[3] Just to make the implications clear, this means that innovation will occur in the application of optimal process to real world problems. This gives rise to the "second school" of information science, building on the first school where optimality was often not available.

[4] In addition to the standards related to keyless hash design and function, the ".vir" standard
    (1)   is a realization of the stratified theory as a specification
    (2)   has well defined specification for the development of ontological modeling tools based on the topic map standard,
    (3)   has well defined specification for a service architecture that separates the push advertising and pull advertising as well as creates service repositories related to commodity transactions.
(see www.bcngroup.org/beadgames/TaosDiscussion/index.htm)

[5] Gelhard and Prueitt (2001) URL: http://www.ontologystream.com/OS/G-PArch.htm

[6] Prueitt (2002) URL: http://www.bcngroup.org/area2/KSF/Notation/notation.htm

[7] The company Hilbert Technology describes a technology based on keyless hash tables at
URL: http://www.hilbertcompany.com/

[8] Pribram, K. H. (1991). Brain and Perception: Holonomy and Structure in Figural Processing. Hillsdale, NJ: Lawrence Erlbaum Associates.

[9] Kowalski, J. ; Ansari, A. ; Prueitt, P. ; Dawes, R. and Gross, G. (1988.) On Synchronization and Phase Locking in Strongly Coupled Systems of Planar Rotators. Complex Systems 2, 441-462.
Prueitt, P.S. (1995) A Theory of Process Compartments in Biological and Ecological Systems. In the Proceedings of IEEE Workshop on Architectures for Semiotic Modeling and Situation Analysis in Large Complex Systems; August 27-29, Monterey, Ca, USA; Organizers: J. Albus, A. Meystel, D. Pospelov, T. Reader.