I am Pan Zhenyang(A0078075X).

The summary:

E.F Codd wrote a paper about the relational model for the data banks in 1985, which is the foundation of the modern database model. In the paper, Codd first list the disadvantage of data dependencies in the data system at his time. From the aspects of ordering, indexing and access path dependencies, he points out that the problems of current database with the applications and systems and suggests that the new database model should apply the information hiding principle. Then he begins to define the relational model of data. The relation is a subset of the Cartesian of the product of each data set. A relation which has degree n is n-ary. The row in the subset represents an n-tuple of relations, which should be distinct and not affected by the order. Each column which represents a domain (unordered) is sensitive with the order. If the relation is non-simple, then normalizing it by taking the primary key of its parent to expand to its child’s primary key and taking away the non-simple column of the parent. Then he demonstrates the operations on the relations. Permutation is to consider the domain to have order so that it will be easy for the internal operations. Projection π(i,j), means take out the column i, j of the relation, strike out the others and delete the duplication. Join operation use the projection operation. For example, R join with S in union U has the relationπ(i,k)U = R,π(k,j)U= S. It also can be applied to more than two relations. The next is the composition. Two relations can be composed if they can be joined. At last, he mentions about the redundancy of the named set should be distinguished from the stored set and the consistency among the relations should be kept.

The comments:

Knowing briefly what the paper is about, in my opinion, Codd has defined a very well organized relational database, which is still widely used nowadays. With applying the hiding information principle, the database does not need to change the implementation or data organizing when the outside application and system have changes. To view the data as different sets is very creative and efficient. Data is an abstract concept, Codd used the sets model to represents it is really a big leap. Once it is can be modeled as the sets, we can apply the all the operations on sets such as join, project, union etc. With the normalization, the data set can be simple enough to store and manipulate. These are the obvious advantages. I just imagine that if the data do not have relations, do we still need to make it to fit the relational model or just separately store them. For example, all the data are discrete points; we should have a lot of tables for each point. Is it still suitable for a great number of information? Maybe, in the future there will be different types of database to handle different kinds of data (not the relational one dominating). Anyway, Codd’s relational model is really efficient.