ABSTRACT

Imperfection is found in variety of forms in different real-life situations. The imperfect information is further classified as incomplete, imprecise or uncertain information. Such information may also arise in the context of applications of artificial intelligence and various information systems. The current commercial relational database management systems are not capable of handling and manipulating imperfect information. The present thesis surveys typical sources of imperfect information in databases and highlights the usefulness of such imperfect information as a basis for efficient and effective decision-making.

The thesis extends the structure of a typical relational database management system for storing and manipulating uncertain information using the theory of probability. A comprehensive probabilistic relational database management system is proposed and various theoretical issues related to developing an exact prototype of such a system are addressed including the probabilistic integrity constraints, probabilistic relational algebra, probabilistic normal forms and its probabilistic query language called PSQL (Probabilistic Structured Query Language).

Various issues related to transaction processing for the comprehensive probabilistic relational database management system are addressed. The PLACID properties that a transaction should possess are enumerated and an algorithm for concurrency control using probabilistic locking techniques is presented.

To avail the benefits associated with distributed systems, the concept of probabilistic databases is extended to a distributed environment and various issues related to query execution are studied. An algorithm is proposed that will minimize the amount of data transfer over a network. Other issues related to distributed concurrency control of the proposed system are studied.