DOCTORAL SCHOOL OF INFORMATICS
COMPLEX EXAM SUBJECT

Databases, knowledge bases (main subject)

Notes:
Subjects A) recommended for student completed Advanced Database courses I-II.
Subjects B) recommended for students interested in practical database applications and implementations

A) Theoretical fundamentals of databases

1. The relational model and query languages

1. The relational model
2. Conjunctive queries (rule based, tableau queries, conjunctive calculus, SPC/SPJR Algebra)
3. Adding Negation: Non recursive Datalog, Algebra, relational (domain) calculus, equivalence theorems
4. Query optimization,
5. Query equivalence and containment of tableau queries, undecidability of query equivalence,
6. Acyclic joins

2. Constraints and design theory

7. Functional and Key dependencies; implication, algorithm for deciding implication, Armstrong axioms.
8. Join and multivalued dependencies, decompositions. Axiomatizations: yes for FD and MVD; no for FD and JD.
9. Deciding implication for FD-s and JD-s in common, The Chase algorithm.
10. Inclusion dependencies, implication, axiomatization, complexity of deciding implication. Inclusion dependencies and FD-s: undicadability theorem.
11. Design theory of relational databases, Normal Forms

3. Datalog

12. Recursive, non negated Datalog programs
13.Model-theoretic ,Fixpoint semantics, Proof-theoretic spproach
14.Seminaive evaluation of Datalog, Top-down evaluation techniques

4. Recursion and negation

15.Inflationary and destructive semantics of recursive queries with negation
16.The While and While$^+$ languages for relational algebra
17. Partial fixpoint and fixpoint extension for relational calculus, Negated and double negated Datalog,
18. Equivalence theorem for inflational and noninflational languages

5. Expressiveness and complexity

19.Typed, computable and generic queries
20. Complexity of Queries and Languages
21.Complexity of First-Order Queries
22.Expressiveness of First-Order Queries
23. Fixpoint and *While* Queries
24.The Impact of Order - PTIME and PSPACE queries

6. Complex Values

25.Complex Value Databases, types, schemas and instances 511

26.The Algebra and the Calculus for complex velues
27 Equivalence Theorems , Expressive Power and Complexity

7. Object Databases

28. Informal presentation and formal definition of an OODB Model
29. Languages for OODB Queries
30.The Object Relational model in SQL
31.XML as data and XML query languages

Literature:

Abiteboul S. –Hull R.– Vianu V.: Foundation of Databases, Addison-Wesley Publ. Comp. 1995.,
Ullman J.D., Widom J. : A First Course in Database Systems, Pearson Prentice Hall, 2008.

B) Database Technologies

1. The relational database model, data modelling, Entity-Relationship model, UML, mapping to relational model
2. Design theory of relational databases, functional dependencies, decomposition, implication problem of dependencies, the Chase algorithm, Boyce-Codd Normal Forms, 3NF, multivalued dependencies, 4NF and algorithms for normalization.
3. Operations for relations. Basic relational algebra, extended algebra, sets and multisets . Independent set of operations. Logics on relations, Datalog rules and programs, recursion and minimal fixpoint.
4. The SQL database language. Queries in SQL, mapping between algebra and SQL. The SELECT statement, set operations, subqueries, aggregation and grouping.
5. Data definition in SQL, CREATE, ALTER, DROP statements, modifications of tables: Insert-Delete-Update, transactions in SQL
6. Constraints (keys, referential integrity, CHECK constraints, global constraints) and triggers in SQL
7. VIEWS and indexes in SQL
8. Embedded SQL, dynamic SQL, PSM, interfaces
9. Recursion is SQL (WITH option), nested relations, object-relational model
10. XML and XML query languages
11. On-Line Analytic Processing, star schemas, data cubs
12. Relational data organization on secondary storage
13. Index structures, B-trees, hash-tables, multidimensional indexes, bitmap indexes
14. Query execution plans, cost model, one-pass, two-pass algorithms for relational operations, multipass algorithms, index-, hashing- and sort-based algorithms
15. Query compiler, optimization strategies, physical execution plans
16. Coping with system failures, transactions and logging based recoveries (undo, redo), archives
17. Concurrency control, ACID transactions, schedules, serializable schedules, lock-based and time-stamp based protocols for enforcing serializability
18. Serializability and recovery, dirty data and cascading roll-back, deadlock, long transactions
19. Transaction management in distributed databases; distributed locks, distributed serializability, commit protocols, recovery and synchronization
20. Distributed query processing, executing full relational operations, distributed joins (semijoins and acyclic joins),
21. Parallel processing (Map-Reduce framework) and pear-to pear serach (distributed hash-tables, Chord-Circles and Finger Tables

Literature:

Garcia-Molina H., Ullman J.D., Widom J.: Database Systems – The Complete Book, 2nd edition. Pearson Prentice Hall, 2009.