Peruvian Computing Society (SPC)
School of Computer Science
Sillabus 2021-I

1. COURSE

CS272. Data Management II (Mandatory)

2. GENERAL INFORMATION

2.1 Credits	:	3
2.2 Theory Hours	:	1 (Weekly)
2.3 Practice Hours	:	4 (Weekly)
2.4 Duration of the period	:	16 weeks
2.5 Type of course	:	Mandatory
2.6 Modality	:	Face to face
2.7 Prerrequisites	:	CS271. Data Management. (4^{th} Sem)

3. PROFESSORS

Meetings after coordination with the professor

4. INTRODUCTION TO THE COURSE

Information Management (IM) plays a leading role in almost every area where computers are used. This area includes the capture, digitization, representation, organization, transformation and presentation of information; Algorithms to improve the efficiency and effectiveness of access and update of stored information, data modeling and abstraction, and physical file storage techniques.

It also covers information security, privacy, integrity and protection in a shared environment. Students need to be able to develop conceptual and physical data models, determine which IM methods and techniques are appropriate for a given problem, and be able to select and implement an appropriate IM solution that reflects all applicable constraints, including scalability and Usability.

5. GOALS

- To make the student understand the different applications that the databases have, in the different areas of knowledge.
- Show appropriate ways of storing information based on their various approaches and their subsequent retrieval of information.

6. COMPETENCES

a) An ability to apply knowledge of mathematics, science. (**Usage**)

b) An ability to design and conduct experiments, as well as to analyze and interpret data. (**Assessment**)

d) An ability to function on multidisciplinary teams. (**Assessment**)

i) An ability to use the techniques, skills, and modern computing tools necessary for computing practice. (**Usage**)

7. SPECIFIC COMPETENCES

a14) Properly use files for storage and retrieval of information.

b4) Identify and efficiently apply various algorithmic strategies and data structures for the solution of a problem given certain space and time constraints.

b5) Identify and efficiently apply diverse algorithmic strategies and data structures for the solution of a problem in parallel and distributed environments.

d2) Developing group presentations and reports on specific topics.

d3) Develop group work on each course topic.

i3) Properly use the query optimization, performance, indexing and table fragmentation modules for distributed DBs using an open source database engine such as PostgreSQL, Cassandra or MongoDB

8. TOPICS

Unit 1: Physical Database Design (10)	
Competences Expected: b,j	
Topics	**Learning Outcomes**
• Storage and file structure • Indexed files • Hashed files • Signature files • B-trees • Files with dense index • Files with variable length records • Database efficiency and tuning	• Explain the concepts of records, record types, and files, as well as the different techniques for placing file records on disk [Usage] • Give examples of the application of primary, secondary, and clustering indexes [Usage] • Distinguish between a non-dense index and a dense index [Usage] • Implement dynamic multilevel indexes using B-trees [Usage] • Explain the theory and application of internal and external hashing techniques [Usage] • Use hashing to facilitate dynamic file expansion [Usage] • Describe the relationships among hashing, compression, and efficient database searches [Usage] • Evaluate costs and benefits of various hashing schemes [Usage] • Explain how physical database design affects database transaction efficiency [Usage]
Readings : [Bur04], [Cel05]	

Unit 2: Transaction Processing (12)	
Competences Expected: b,j	
Topics	**Learning Outcomes**
• Transactions • Failure and recovery • Concurrency control • Interaction of transaction management with storage, especially buffering	• Create a transaction by embedding SQL into an application program [Usage] • Explain the concept of implicit commits [Usage] • Describe the issues specific to efficient transaction execution [Usage] • Explain when and why rollback is needed and how logging assures proper rollback [Usage] • Explain the effect of different isolation levels on the concurrency control mechanisms [Usage] • Choose the proper isolation level for implementing a specified transaction protocol [Usage] • Identify appropriate transaction boundaries in application programs [Usage]
Readings : [Phi97], [Ram04]	

Unit 3: Information Storage and Retrieval (10)

Competences Expected: b,j

Topics	Learning Outcomes
• Documents, electronic publishing, markup, and markup languages • Tries, inverted files, PAT trees, signature files, indexing • Morphological analysis, stemming, phrases, stop lists • Term frequency distributions, uncertainty, fuzziness, weighting • Vector space, probabilistic, logical, and advanced models • Information needs, relevance, evaluation, effectiveness • Thesauri, ontologies, classification and categorization, metadata • Bibliographic information, bibliometrics, citations • Routing and (community) filtering • Multimedia search, information seeking behavior, user modeling, feedback • Information summarization and visualization • Faceted search (e.g., using citations, keywords, classification schemes) • Digital libraries • Digitization, storage, interchange, digital objects, composites, and packages • Metadata and cataloging • Naming, repositories, archives • Archiving and preservation, integrity • Spaces (conceptual, geographical, 2/3D, VR) • Architectures (agents, buses, wrappers/mediators), interoperability • Services (searching, linking, browsing, and so forth) • Intellectual property rights management, privacy, and protection (watermarking)	• Explain basic information storage and retrieval concepts [Usage] • Describe what issues are specific to efficient information retrieval [Usage] • Give applications of alternative search strategies and explain why the particular search strategy is appropriate for the application [Usage] • Design and implement a small to medium size information storage and retrieval system, or digital library [Usage] • Describe some of the technical solutions to the problems related to archiving and preserving information in a digital library [Usage]

Readings : [Pet98], [Ram04]

Unit 4: Distributed Databases (36)	
Competences Expected: b,j	
Topics	**Learning Outcomes**
• Distributed DBMS – Distributed data storage – Distributed query processing – Distributed transaction model – Homogeneous and heterogeneous solutions – Client-server distributed databases • Parallel DBMS – Parallel DBMS architectures: shared memory, shared disk, shared nothing; – Speedup and scale-up, e.g., use of the MapReduce processing model – Data replication and weak consistency models	• Explain the techniques used for data fragmentation, replication, and allocation during the distributed database design process [Usage] • Evaluate simple strategies for executing a distributed query to select the strategy that minimizes the amount of data transfer [Usage] • Explain how the two-phase commit protocol is used to deal with committing a transaction that accesses databases stored on multiple nodes [Usage] • Describe distributed concurrency control based on the distinguished copy techniques and the voting method [Usage] • Describe the three levels of software in the client-server model [Usage]
Readings : [M T99]	

9. WORKPLAN

9.1 Methodology

Individual and team participation is encouraged to present their ideas, motivating them with additional points in the different stages of the course evaluation.

9.2 Theory Sessions

The theory sessions are held in master classes with activities including active learning and roleplay to allow students to internalize the concepts.

9.3 Practical Sessions

The practical sessions are held in class where a series of exercises and/or practical concepts are developed through problem solving, problem solving, specific exercises and/or in application contexts.

10. EVALUATION SYSTEM

********* EVALUATION MISSING ********

11. BASIC BIBLIOGRAPHY

[Bur04] Donald K. Burleson. *Physical Database Design Using Oracle*. CRC Press, 2004.

[Cel05] Joe Celko. *Joe Celko's SQL Programming Style*. Elsevier, 2005.

[M T99] Patrick Valduriez M. Tamer Ozsu. *Principles of Distributed Database Systems, Second Edition*. Prentice Hall, 1999.

[Pet98] Julita Vassileva Peter Brusilovsky Alfred Kobsa. *Adaptive Hypertext and Hypermedia, First Edition*. Springer, 1998.

[Phi97] Eric Newcomer Philip A. Bernstein. *Principles of Transaction Processing, First Edition*. Morgan Kaufmann, 1997.

[Ram04] Shamkant B. Navathe Ramez Elmasri. *Fundamentals of Database Systems, Fourth Edition*. Addison Wesley, 2004.