

Sample Midterm and Final questions for INFO-340

Database Material (Weeks 1-5)

1. (3 points) Using a figure, explain the **lost update problem**. What feature of databases is used to prevent lost updates?
2. (2 points) a) Who invented relational databases and in what year were they invented? b) What central, beautiful idea underlies them?
3. (6 points) Explain the differences between **conceptual**, **logical** and **physical** database design.
4. (2 points) Using a diagram and an explanation, explain what an **outer join** is.
5. (3 points) Write a relational algebra expression for this query: **Get supplier names for suppliers who do not supply part id, p001.**
6. (6 points) An investment bank brokers a deal between a buyer and a seller of a publicly traded company. Behind the scenes of each of these players are lawyers, accountants, and negotiators. Draw an **EER diagram** that represents this structure. Please include a brief caption that outlines your assumptions.
7. (4 points) Give a formal definition of a **relational database schema**.

Information Retrieval Material (Week 6-10)

8. (6 points) IR is often represented as a **communication process**. Using a diagram, describe this process including the **major phases** and **entities** that make it up.
9. (2 points) Explain the difference between **open** and **closed** vocabularies.
10. (2 points) What is a **document proxy**? Why are document proxies used?
11. (2 points) Define **recall-enhancing** and describe one example technique.
12. (2 points) Why do IR systems make use of an **inverted file**?
13. (4 points) Draw a diagram showing the structure and elements of an **inverted file**
14. (6 points) Describe a basic algorithm for **indexing a corpus of documents**.
15. (6 points) Suppose a high school teacher wants to detect plagiarism in her students work. Describe an IR system that might be designed to solve this problem and list 3 IR technical issues that would have to be addressed.