Index Number: UE20027715 **Programme:** BSc COMP. ENG

UNIVERSITY OF ENERGY AND NATURAL RESOURCES, SUNYANI
SCHOOL OF ENGINEERING
DEPARTMENT OF COMPUTER AND ELECTRICAL ENGINEERING

LEVEL 300, END OF SECOND SEMESTER EXAMINATION, 2017/2018
Bachelor of Science (Electrical and Electronics Engineering)
CENG 312 Database and Information Retrieval

May, 2018 **Time: 2 Hrs : 45 Mins**

Instructions: *The paper consist of TWO SECTIONS, A and B. Answer ALL questions in SECTION A and TWO (2) questions completely in SECTION B.*

SECTION A – *Shade well the letter corresponding to the correct option from A to D on the Shading Answer Sheet. 0.5 mark per any correct answer.*

1. The relational database environment has all of the following components except _____.
   A. separate files
   B. database
   C. query languages
   D. database tables
2. Database management systems are intended _.
   A. eliminate data redundancy
   B. establish relationship among records in different files
   C. maintain data integrity
   D. all of the above
3. What language does application programs use to request data from the DBMS?
   A. DML
   B. DDL
   C. query language
   D. any of the above
4. The highest level in the hierarchy of data organization is called _____.
   A. data bank
   B. data base
   C. data file
   D. data record
5. Which are Data Manipulation Language (DML) statements?
   I. SELECT II. ALTER
   III. DROP IV. DELETE
   A. I, IV
   B. II, III
   C. I, II, III
   D. II, III, IV
6. Report generators are used to _____
   A. store data input by a user
   B. retrieve information from files
   C. answer queries
   D. both B and C
7. A form is defined _____
   A. where data is placed on the screen
   B. the width of each field
   C. both a and b
   D. none of the above
8. A top-to-bottom relationship among the items in a database is established by a _____.
   A. hierarchical schema
   B. network schema
   C. relational schema
   D. all of the above
9. The management information system (MIS) structure with one main computer system is called a _____.
   A. hierarchical MIS structure
   B. distributed MIS structure
   C. centralized MIS structure
   D. decentralized MIS structure
10. Which of the following hardware component is the most important to the operation of database management system?
    A. High resolution video display
    B. High speed, large capacity disk
    C. Plotter
    D. Mouse

Index Number: _____ **Programme:** _____

11. Batch processing is appropriate if _____.
    A. large computer system is available
    B. only a few transactions are involved
    C. all of the above
    D. none of the above
12. Large collection of files is called _____.
    A. fields
    B. records
    C. database
    D. sectors
13. In order to use a record management system, _____.
    A. you need to understand the low-level details of how information is stored
    B. you need to understand the model the record management system uses
    C. Both A and B
    D. None of the above
14. If a piece of data is stored in two places in the database, then _____.
    A. storage space is wasted
    B. changing the data in one spot will cause data inconsistency
    C. in can be more easily accessed
    D. both A and B
15. Which of the following fields in a student file can be used as a primary key?
    A. Class
    B. Social Security Number
    C. GPA
    D. Major
16. Which of the following is not an advantage of the database approach?
    A. Elimination of data redundancy
    B. Ability of associate deleted data
    C. Program/data independence and increased security
    D. All of the above
17. Which of the following contains a complete record of all activity that affected the contents of a database during a certain period of time?
    A. Report writer
    B. Query language
    C. Data manipulation language
    D. Transaction log
18. In the DBMS approach, application programs perform the _____.
    A. storage function
    B. processing functions
    C. access control
    D. all of the above
19. A set of programs that handle a firm's database responsibilities is called _____.
    A. database management system (DBMS)
    B. database processing system (DBPS)
    C. data management system (DMS)
    D. all of above
20. A record management system _____.
    A. can handle many files of information at a time
    B. can be used to extract information stored in a computer file
    C. always uses a list as its model
    D. Both A and B

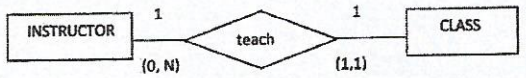
21. A command that lets you change one or more fields in a record is _____.
    A. insert
    B. modify
    C. lookup
    D. None of above
22. An attribute that can be calculated from related attribute values is called a(n) _____ attribute.
    A. simple
    B. composite
    C. multivalued
    D. derived

23. A transparent DBMS _____.
    A. cannot hide sensitive information from users
    B. keeps its logical structure hidden from users
    C. keeps its physical structure hidden from users
    D. Both B and C

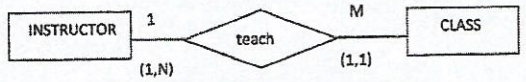
24. Choose the correct Entity Relationship Model diagram based on the scenario below:
    - "Each INSTRUCTOR may teach one or more CLASS"

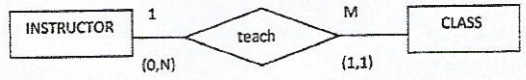
Index Number: _____ **Programme:** _____

- "Each CLASS is taught by one INSTRUCTOR"

A. 

B. 

C. 

D. 

25. The ascending order of a data hierarchy is_.
    A. bit-byte-record-field-file-database
    B. byte-bit-field-record-file-database
    C. bit-byte-field-record-file-database
    D. bit-byte-file-record-field-database
26. The common types of entities are: _____.
    A. strong entities
    B. weak entities
    C. associative entities
    D. All of the above
27. Which of the following is true of a network structure?
    A. It is a physical representation of the data
    B. It allows a many-to-many relationship
    C. It is conceptually simple
    D. It will be dominant data base of the future
28. Which of the following is a problem of file management system?
    A. Difficult to update and program dependence
    B. Lack of data independence
    C. Data redundancy
    D. All of above
29. Primitive operations common to all record management system include _____.
    A. print
    B. sort
    C. look - up
    D. All of above
30. The logical representation of an organization's data is called a(n): _____.
    A. database model
    B. entity-relationship model
    C. relationship systems design
    D. database entity diagram
31. An entity types whose existence depends on another entity type is called a(n) __ entity.
    A. strong
    B. weak
    C. co-dependent
    D. variant
32. A _____ is the set of allowable values for one or more attributes.
    A. cardinality
    B. tuple
    C. degree
    D. row
33. A person's name, birthday, and social security number are all examples of:
    A. attributes
    B. entities
    C. relationships
    D. descriptors
34. A relationship between the instances of a single entity type is called a(n) _____ relationship.
    A. ternary
    B. primary
    C. binary
    D. unary
35. Which of the following conditions should exist if an associative entity is to be created?
    A. All the relationships for the participating entities are many-to-many
    B. The new associative entity has independent meaning
    C. The new associative entity participates in independent relationships
    D. All of the above
36. Internal Schema is one of the three-schema architecture. Which one is true about Internal Schema?
    A. It hides the details of physical storage structures and concentrates on describing entities, data types, relationships, user operations, and constraints.
    B. Uses a physical data model and describes the complete details of data storage and access paths for the database.
    C. Implementation of data model

D. The part of the database that a particular user is interested in and hides the rest of the database from user

37. ____ works on a single relation R and defines a relation that constrains only those tuples (rows) of R that satisfy the specified condition (predicate).
    A. Cartesian product
    B. Difference
    C. Selection
    D. Intersection

38. _____ works on a single relation R and defines a relation that contains a vertical subset of R, extracting the values of specified attributes and eliminating duplicates.
    A. Projection
    B. Cartesian Product
    C. Difference
    D. Intersection

39. You need to produce a report for mailing labels for all customers. The mailing label must have only the customer name and address. The CUSTOMERS table has these columns:

```
CUST_ID NUMBER(4) NOT NULL
CUST_NAME VARCHAR2(100) NOT NULL
CUST_ADDRESS VARCHAR2(150)
CUST_PHONE VARCHAR2(20)
```

Which SELECT statement accomplishes this task?

A.
```
SELECT *FROM CUSTOMERS;
```

B.
```
SELECT CUST_NAME,
CUST_ADDRESS, CUST_PHONE
FROM CUSTOMERS;
```

C.
```
SELECT CUST_ID, CUST_NAME,
CUST_ADDRESS, CUST_PHONE
FROM CUSTOMERS;
```

D.
```
SELECT CUST_NAME,
CUST_ADDRESS FROM CUSTOMERS;
```

40. Within a table, the primary key must be unique so that it will identify each row. When this is the case, the table is said to exhibit _____ integrity.
    A. referential
    B. entity
    C. enforced
    D. key

41. A _____ is the set of allowable values for one or more attributes.
    A. Cardinality
    B. Tuple
    C. Degree
    D. Domain

Index Number: _____ **Programme:** _____

SECTION B

ANSWER ANY TWO (2) QUESTIONS. EACH QUESTION CARRIES 20 MARKS

QUESTION ONE

a. What is database?

b. Consider the following relations:

```
Emp(eno, ename, title, city)
Proj(pno, pname, budget, city)
Works(eno, pno, resp, dur)
Pay(title, salary)
```

where the primary keys are underlined, and `Emp.title` is a foreign key to `Pay.title`, `Works.eno` is a foreign key to `Emp.eno`, and `Works.pno` is a foreign key to `Proj.pno`. For each part of this question (considered independently of the other parts), write a single SQL statement that accomplishes the following requirements.

i. For each city, how many projects are located in that city and what is the total budget over all projects in the city?

ii. For each project, what fraction of the budget is spent (in total) on salaries for the people working on that project? Sort your answer by the value of the budget.

c. i. Formulate the following SQL queries in relational algebra. The SQL queries are expressed over the same schema as given in (b).

```
SELECT pname, budget
FROM Proj, Works, Emp
WHERE title = 'Programmer'
AND Works.eno = Emp.eno
AND Works.pno = Proj.pno
```

ii. Explain referential-integrity constraint.

iii. Describe the differences in meaning between the terms *relation* and *relation schema*.

d. Use **an** Entity-Relationship Diagram to depict the following enterprise. Explain any additional assumptions you make and list any aspects you were not able to depict.

- ✓ A house is identified by a three-part address consisting of a number, street, and city. Each house also has a style (e.g., bungalow) and a set of colours.
- ✓ A person is identified by a social insurance number. For each person, record his/her name, age, and sex.
- ✓ Persons who are at least 18 years old may own zero or more houses, and every house is owned by at least one person.
- ✓ Any person (regardless of age) lives in at most one house as his/her principal residence, and a house can have zero or more persons living there.

Index Number: UE2602718 **Programme:** _____

QUESTION TWO

a. Consider two relations namely ***Sailors*** and ***Reserves***. Two instances, **FirstSailors** and **SecondSailors** are created from ***Sailors*** relation and **ReserveSailors** from the ***Reserves*** relation. The resulting tables are listed below:

FirstSailors			
sid	sname	rating	age
22	Dustin	7	45.0
31	Luber	8	55.5
58	Rusty	10	35.0
61	Rubby	8	37.0

SecondSailors		
sid	sname	rating
28	Yuppy	9
31	Luber	8
44	Guppy	5
58	Rusty	10

ReserveSailors			
sid	sname	rating	age
22	Dustin	7	45.0
31	Luber	8	55.5
58	Rusty	10	35.0
57	Hunter	7	43.0

For each of the algebraic expression below, provide in a tabular form the result of each expression when applied to the appropriate relations (tables) above.

i. $\sigma_{rating>8}(\text{FirstSailors})$

ii. $\pi_{sname,rating}\left(\sigma_{rating>8}(\text{SecondSailors})\right)$

b. Express the relational algebra $\left(\rho_{(enr,ename,dept)}(E)\right) \triangleright\triangleleft_{dept=dnr} \left(\rho_{(dnr,dname)}(D)\right)$ into SQL query statement.

c. The schema below are for two relations of UENR Information Management System:

`Instructor-schema = (InstrID, InstrName, DeptName, InstrSalary);`	`InstrTeaches-schema = (ID, courseID, sectionID, semester, year );`
i. Primary key (ID), ii. Foreign key (DeptName) references department	i. Primary key (ID, course id, sec id, semester, year), ii. Foreign key (course id, sec id, semester, year) references section, iii. Foreign key (ID) references instructor);

Table 2c attached is the data produce after running a certain query done on the relations.

i. Write the SQL query that will generate the data in **Table 2c**

ii. What SQL query will produce the data in the **Table 2c-ii** below

Table 2c-ii

semester	salary
Fall	107978.47
Fall	103146.87
Spring	103146.87

iii. Provide in a tabular form, the data produced when the query below are executed on the relations in (c) with data in **Table 2c**

```
SELECT      *
FROM        Instructor, InstrTeaches
WHERE       (
                Instructor.InstrID = InstrTeaches.ID AND
                Instructor.salary < 100230) AND
                (Instructor.DeptName = 'Psychology' AND
                Instructor. DeptName = 'Mech. Eng.');
```

d. i. List four significant differences between a file-processing system and a DBMS.
   ii. Name the main steps in database design. What is the goal of each step? In which step is the ER model mainly used?

QUESTION THREE

a. i. What is an information retrieval system? Give two examples.
   ii. Explain inverted index in information retrieval.

   ii. Name three components in a typical retrieval system, where *MapReduce* are potentially helpful. (A MapReduce program is composed of a *Map()* procedure that performs filtering and sorting and a *Reduce()* procedure that performs a summary operation in parallel).

b. i. Twitter is building its next generation of tweet search engine. Engineers are arguing that classical IR techniques are totally sufficient to build it so that there is no need for them to work overtime. As Twitter's chief research scientist in IR, do you agree with them? If not, what are the major technical challenges? What has to be innovated and what has to be adapted in such a system? (**Hint**: *think about the system architecture of a classical IR system and the nature of Twitter*).
   ii. In retrieval evaluation, all the documents are independently labelled against the query and the labels are assumed to be static when being used to evaluate different retrieval systems. List three limitations of such an assumption?

c. i. What are the main approaches to evaluating selections? Discuss the use of indexes, in particular.
   ii. What is the goal of query optimization? Is it to find the best plan?
   iii. Describe the iterator interface for operators and access methods. 'What is its purpose?

SECOND SEMESTER EXAMINATION, 2017/2018

CENG 312 - DATABASE AND INFORMATION RETRIEVAL

Table 2c: *Table for QUESTION TWO (c)*

InstrID	InstrName	DeptName	salary	InstrSalary	courseID	sectionID	semester	year
14365	Lembr	Accounting	32241.56	14365	200	1	Spring	2007
14365	Lembr	Accounting	32241.56	14365	843	1	Fall	2010
15347	Bawa	Athletics	72140.88	15347	457	1	Spring	2001
19368	Wieland	Pol. Sci.	124651.41	19368	545	1	Fall	2001
19368	Wieland	Pol. Sci.	124651.41	19368	581	1	Spring	2005
19368	Wieland	Pol. Sci.	124651.41	19368	591	1	Spring	2005
22591	DAgostino	Psychology	59706.49	22591	338	1	Spring	2007
22591	DAgostino	Psychology	59706.49	22591	338	2	Spring	2006
22591	DAgostino	Psychology	59706.49	22591	352	1	Spring	2006
22591	DAgostino	Psychology	59706.49	22591	400	1	Spring	2007
22591	DAgostino	Psychology	59706.49	22591	400	2	Fall	2003
22591	DAgostino	Psychology	59706.49	22591	482	1	Fall	2005
22591	DAgostino	Psychology	59706.49	22591	599	1	Spring	2003
22591	DAgostino	Psychology	59706.49	22591	642	1	Fall	2004
22591	DAgostino	Psychology	59706.49	22591	663	1	Spring	2005
22591	DAgostino	Psychology	59706.49	22591	867	1	Fall	2006
22591	DAgostino	Psychology	59706.49	22591	962	1	Spring	2008
22591	DAgostino	Psychology	59706.49	22591	972	1	Spring	2009
22591	DAgostino	Psychology	59706.49	22591	991	1	Spring	2008
25946	Liley	Languages	90891.69	25946	192	1	Fall	2002
28097	Kean	English	35023.18	28097	366	1	Fall	2005
28097	Kean	English	35023.18	28097	808	1	Fall	2003
28400	Atanassov	Statistics	84982.92	28400	603	1	Fall	2003
28400	Atanassov	Statistics	84982.92	28400	604	1	Spring	2009
3199	Gustafsson	Elec. Eng.	82534.37	3199	169	1	Spring	2007
3199	Gustafsson	Elec. Eng.	82534.37	3199	169	2	Fall	2002
3199	Gustafsson	Elec. Eng.	82534.37	3199	561	1	Fall	2006
3199	Gustafsson	Elec. Eng.	82534.37	3199	631	1	Spring	2007
3335	Bourrier	Comp. Sci.	80797.83	3335	949	1	Fall	2007
3335	Bourrier	Comp. Sci.	80797.83	3335	960	1	Fall	2009
34175	Bondi	Comp. Sci.	115469.11	34175	274	1	Fall	2002
34175	Bondi	Comp. Sci.	115469.11	34175	571	1	Spring	2004
34175	Bondi	Comp. Sci.	115469.11	34175	747	1	Spring	2004
36897	Morris	Marketing	43770.36	36897	242	1	Fall	2009
36897	Morris	Marketing	43770.36	36897	313	1	Fall	2010
36897	Morris	Marketing	43770.36	36897	696	1	Spring	2002
36897	Morris	Marketing	43770.36	36897	791	1	Spring	2006
36897	Morris	Marketing	43770.36	36897	795	1	Spring	2004
41930	Tung	Athletics	50482.03	41930	401	1	Fall	2003
41930	Tung	Athletics	50482.03	41930	421	1	Fall	2004
41930	Tung	Athletics	50482.03	41930	692	1	Spring	2010

4233	Luo	English	88791.45	4233	679	1	Spring	2010
42782	Vicentino	Elec. Eng.	34272.67	42782	793	1	Spring	2002
43779	Romero	Astronomy	79070.08	43779	105	1	Fall	2009
43779	Romero	Astronomy	79070.08	43779	105	2	Fall	2002
43779	Romero	Astronomy	79070.08	43779	476	1	Fall	2010
43779	Romero	Astronomy	79070.08	43779	489	1	Fall	2007
48507	Lent	Mech. Eng.	107978.47	48507	626	1	Fall	2006
48570	Sarkar	Pol. Sci.	87549.8	48570	867	2	Fall	2010
50330	Shuming	Physics	108011.81	50330	468	2	Fall	2007
63287	Jaekel	Athletics	103146.87	63287	334	1	Fall	2009
63287	Jaekel	Athletics	103146.87	63287	852	1	Spring	2008
65931	Pimenta	Cybernetics	79866.95	65931	875	1	Spring	2005
73623	Sullivan	Elec. Eng.	90038.09	73623	694	1	Fall	2002
74420	Voronina	Physics	121141.99	74420	239	1	Fall	2006
74420	Voronina	Physics	121141.99	74420	376	1	Fall	2006
74420	Voronina	Physics	121141.99	74420	443	1	Spring	2010
74420	Voronina	Physics	121141.99	74420	443	2	Spring	2002
74420	Voronina	Physics	121141.99	74420	612	1	Fall	2007
74420	Voronina	Physics	121141.99	74420	959	1	Fall	2006
77346	Mahmoud	Geology	99382.59	77346	486	1	Fall	2009
77346	Mahmoud	Geology	99382.59	77346	493	1	Spring	2010
77346	Mahmoud	Geology	99382.59	77346	704	1	Spring	2008
77346	Mahmoud	Geology	99382.59	77346	735	1	Spring	2003
77346	Mahmoud	Geology	99382.59	77346	735	2	Spring	2010
77346	Mahmoud	Geology	99382.59	77346	864	1	Spring	2006
79081	Ullman	Accounting	47307.1	79081	200	2	Fall	2002
79081	Ullman	Accounting	47307.1	79081	345	1	Spring	2008
79081	Ullman	Accounting	47307.1	79081	408	1	Spring	2007
79081	Ullman	Accounting	47307.1	79081	408	2	Spring	2003
79081	Ullman	Accounting	47307.1	79081	760	1	Spring	2004
79081	Ullman	Accounting	47307.1	79081	974	1	Fall	2003
80759	Queiroz	Biology	45538.32	80759	559	1	Fall	2002
81991	Valtchev	Biology	77036.18	81991	415	1	Fall	2010
81991	Valtchev	Biology	77036.18	81991	702	1	Spring	2001
90376	Bietzk	Cybernetics	117836.5	90376	158	1	Fall	2008
90643	Choll	Statistics	57807.09	90643	461	1	Fall	2002
95709	Sakurai	English	118143.98	95709	258	1	Fall	2007
95709	Sakurai	English	118143.98	95709	270	1	Spring	2010
95709	Sakurai	English	118143.98	95709	468	1	Fall	2005
95709	Sakurai	English	118143.98	95709	960	2	Fall	2006
99052	Dale	Cybernetics	93348.83	99052	158	2	Spring	2008
99052	Dale	Cybernetics	93348.83	99052	237	1	Spring	2008
99052	Dale	Cybernetics	93348.83	99052	237	2	Fall	2009
99052	Dale	Cybernetics	93348.83	99052	496	1	Fall	2001
99052	Dale	Cybernetics	93348.83	99052	629	1	Spring	2003
99052	Dale	Cybernetics	93348.83	99052	748	1	Fall	2003
99052	Dale	Cybernetics	93348.83	99052	802	1	Spring	2003
99052	Dale	Cybernetics	93348.83	99052	893	1	Fall	2007
99052	Dale	Cybernetics	93348.83	99052	927	1	Fall	2002

CENG 312 DATABASE AND INFORMATION RETRIEVAL
MIDSEM EXAMS, 2017/2018
Answer **any** question. **[40 mins]**

Question One **[25 marks]**

a) You are asked to design a database for the following scenario. A research laboratory is running several drug trials on healthy volunteers to check whether drugs have side effects. Each drug has a unique name. Each trial involves exactly one drug and several volunteers (who take the drug and report if they had any side effects). For each volunteer in each trial it needs to be recorded whether the volunteer had any side effects, and if yes, what those side effects were (there could be several side effects experienced by the same person, for example headache, dry mouth, and fever). It is important that side-effects are described using some standard terminology, so that the laboratory can report what proportion of volunteers had the same side effect. For example, the researchers may tell you that headache should always be recorded as 'headache' and not sometimes as 'pain in the head' and sometimes as 'sore head'. There is however no fixed pre-defined set of possible side effects, because new effects can always be discovered (for example the drug may turn people a bright green colour). For simplicity, assume that each volunteer takes part in at most one drug trial. Data stored about volunteers is their National Insurance number, name, age, gender, address and telephone number.

1. Identity all the entity sets
2. Draw an entity-relationship diagram for the drug trial scenario.
3. List resulting database tables. State what are the candidate keys in each table and what are foreign keys (if any).

b) What are three main functions of a database administrator?

Question Two

a. State two advantages views have over tables.
b. Explain referential-integrity constraint?
c. List four significant differences between a file-processing system and a DBMS.
d. Name the main steps in database design. What is the goal of each step? In which step is the ER model mainly used?
e. Consider the E-R diagram in **Figure 1** below:

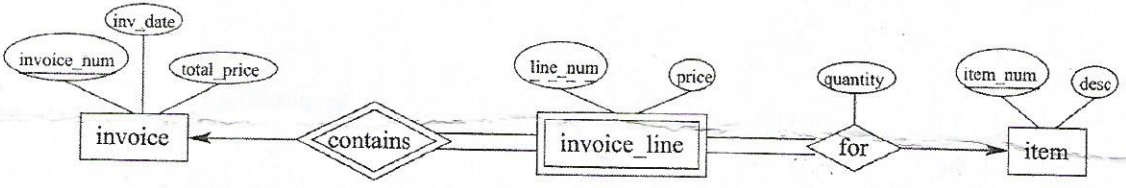


Figure 2

i. Reduce the diagram to a relational database schema.
ii. Express your relational schema using a schema diagram (that indicates primary and foreign keys).

referential tuple is deleted or updated

When relations are updated referential integrity can be violated