CS4604 Midterm Exam

March 09, 2000

Please enter the following information:

- **Name:**

- **ID:**

GOOD LUCK!
Have a good break!
Do not write below this line

Problem	Max Score	Score
1	25	
2	10	
3	15	
4	20	
5	24	
6	6	

1. (25 points) Design an E/R diagram to model the following information about the web: The world wide web is a global network consisting of numerous web sites. Sites are uniquely distinguished by their primary URL (like `http://www.cs.vt.edu`). Each site contains thousands of individual web pages and the identifier (URL) for a web page is obtained by concatenating the URL of the site (it belongs to) and the filename encoding the individual page (if necessary, path information from the site's top directory is included). For example, the identifier for the `faculty.html` page (file) under the `http://www.cs.vt.edu` site is given by `http://www.cs.vt.edu/faculty.html`. Web pages point to one another, via hyperlinks. Search engines are devices used to index the web to aid in content retrieval. For each web page indexed by a search engine, information about keywords (present in the page) should be stored. Two or more search engines may index the same page with different choices for keywords. One web page may be indexable via hundreds of keywords and a given keyword might appear in thousands of pages. Different search engines do not return the same results for a given keyword. For each web page returned by a search engine (for a given keyword), rankings (of relevance) are also typically provided. In addition, search engines treat keywords differently. Some search engines return web pages indexed by 'Galleries', even if the input query was for 'Museums.' Some search engines are more strict and will only return web pages if the exact keyword is present. A search engine creates its store of web pages by starting with a set of 'hub pages' and following hyperlinks from them via a crawler, indexing everything on its path, till it either reaches a dead-end or arrives at a page already indexed. Try to model as much as you can and write *Notes*, if necessary. Failing to write *Notes* indicates that your E/R diagram is *exact*.

2. • (3 points) It is possible for a relationship to be declared many-one and be symmetric. Under what conditions will this be true? *Note:* A symmetric relationship is one which is its own inverse.

• (4 points) Consider a n-way relationship R between entity sets E_1, E_2, ..., E_n. What is the minimum and maximum number of attributes that could form the key for a relation derived from R?

• (3 points) Give example data for a relation $R(A, B, C)$ for which the following hold: $\{A \rightarrow\rightarrow B, CB \rightarrow\rightarrow A\}$.

3. (10+5=15 points) An ODL schema when converted to relations (using the method that we now know so well), gives rise to the following schema (without normalization): `R(a), S(b), T(c), U(d), V(a,b,c,d)`. You may assume that the same symbols refer to the same attribute and that different symbols refer to different attributes (e.g. the `a` in `R` and `V` are the same). Your task is to (i) give a possible ODL schema that could have produced these relations, and (ii) describe what kind of situation would have caused this ODL schema, in the first place. Full points will be awarded to the most insightful answer.

4. (20 points) Consider the relation $R(A, B, C, D, E, G)$ with FDs given by $\{AB \rightarrow C, AC \rightarrow B, AD \rightarrow E, B \rightarrow D, BC \rightarrow A, E \rightarrow G\}$. Consider the decomposition of R into $\{AB, BC, ABDE, EG\}$. Is this decomposition dependency-preserving? Is it a lossless decomposition? Just saying 'Yes' or 'No' to these two questions, without explanations, is worth zero points.

5. (24 points) Assume that we have the following four tuples in a relation $S(A, B, C)$: $(1, 2, 3), (4, 2, 3), (5, 3, 3), (5, 3, 4)$. Which of the following can you infer *does not* hold over S? Give short one-line reasons, for each.

   - $A \rightarrow B$

   - $A \rightarrow\!\!\!\rightarrow B$

   - $BC \rightarrow A$

   - $BC \rightarrow\!\!\!\rightarrow A$

   - $B \rightarrow C$

   - $B \rightarrow\!\!\!\rightarrow C$

6. (6 points) When we go from 3NF to BCNF, notice that we might be splitting a key across two decomposed relations. Why is this such a good idea? If we split the key attributes across two relations, isn't that a cause for worry? Why then, is BCNF such a good normal form?