

# CIS 560 Database System Concepts

## Fall 2006

### Homework 7 of 10: Problem Set (PS7) Entity-Relational Data Modeling

Assigned: Sun 12 Nov 2006  
Due: Fri 17 Nov 2006 (before midnight)

The purpose of this assignment is to exercise your basic understanding of entity-relational data modeling and normalization, and get you started on the term project.

This homework assignment is worth a total of 20 points.

Each problem is worth 5 points; your score for Problems 2 – 5 will be averaged with your hour exam scores.

Type (as PDF) or hand-write and scan your solution and upload it to your KSOL drop box.

#### References

- Chapters 6 & 7, Database System Concepts, 5e
  - <http://dev.mysql.com/tech-resources/articles/introto-normalization.html>
  - [http://en.wikipedia.org/wiki/Database\\_normalization](http://en.wikipedia.org/wiki/Database_normalization)
  - [http://en.wikipedia.org/wiki/First\\_normal\\_form](http://en.wikipedia.org/wiki/First_normal_form)
  - [http://en.wikipedia.org/wiki/Second\\_normal\\_form](http://en.wikipedia.org/wiki/Second_normal_form)
  - [http://en.wikipedia.org/wiki/Third\\_normal\\_form](http://en.wikipedia.org/wiki/Third_normal_form)
1. **Term Project.** Turn in a relational schema for your term project, and make an appointment to see the instructor about the design this week (the week of Mon 13 Nov 2006). The revised schema and design interview will count for the first part of the project (50 points or 5% of your course grade) in addition to the 5 points (0.5%) for this problem.

#### **Notes:**

- Your schema should be in Third Normal Form (3NF) as specified on MP6. **Show your work** in converting from the First Normal Form (1NF) version that you specified for the flat file format.
- Type your solution. Use MS Word (with Visio 2003 or Access 2003 figures as needed) or OpenOffice and turn in a PDF file titled PS7-1.pdf.

**Exam redo.** Consider the following schema for Problems 2 – 5. The contents are just given as examples.

**Drug:** (drug-id, drug-name, syndrome)

**Doctor:** (doctor-id, first-name, last-name, area-of-practice)

**Patient:** (patient-id, first-name, last-name, syndrome)

#### Drug

drug-id	drug-name	syndrome
a101	Ovidria	infectious eczema
a101	Ovidria	delusional American pox
a101	Ovidria	hematologic interstitial elephantiasis
c3po	Zaphat	Bell-Guillain syndrome
d555	Betoxi	acquired insomnia
d555	Betoxi	peripheral epilepsy
h323	Soloxand	delusional American pox
r2d2	Heurridrine	peripheral epilepsy
r2d2	Heurridrine	Bell-Guillain syndrome
x123	Tozenol	hematologic interstitial elephantiasis

#### Doctor

doctor-id	first-name	last-name	area-of-practice
a50	Roberta	Hanz	General practitioner
b39	Rufus	Garside	Ear, nose, and throat
c76	Sakamae	Akamatsu	Radiologist
d57	Fabio	Apassionado	Sports medicine

#### Patient

patient-id	first-name	last-name	syndrome
a239	John	Locke	delusional American pox
b398	James	Ford	peripheral epilepsy
c185	Desmond	Hume	acquired insomnia
d940	Hugo	Reyes	acquired insomnia
e093	Kate	Austen	Bell-Guillain syndrome
f452	Tunde	Eko	peripheral epilepsy

#### References:

Drug name generator: [http://www.wordlab.com/tools/t\\_index.cfm](http://www.wordlab.com/tools/t_index.cfm)

Disease name and symptom generator: <http://www.microfiction.org/disease/>

Name generators: <http://rumandmonkey.com/widgets/toys/namegen/>





4. **Database Design (2 parts).** Consider the original schema for Problems 2-3 and in particular the ternary relationship set Prescription referred to in problem 1(d). This relation should include at least the date on which the prescription is written, the duration of the prescription in days, and the dosage (quantity and frequency). In this problem, you will take the redesigned Drug from 3(d) and use it to sketch the E-R diagram.

a) **E-R data modeling.** Using your revised table, write down an entity-relational (E-R) data model for Patient, Drug (the new version containing the ID as primary key), and Doctor. Your E-R diagram should show the cardinality of each mapping, written on the link, with a directed arrow denoting "one-to-one relationship".

b) **Normalization and E-R data model.** Explain (i.e., sketch) how the E-R diagram changed when you changed Drug so that it has drug-id as primary key.

**Note:** In the last part of the course, you will put these concepts together with others we have covered (queries, semistructured data), are covering (query processing), or will cover (transactions) and look at *aggregation* (mappings between relationship sets).

## 5. Queries and XML (2 parts).

- a) **Queries on relationship sets.** A medical insurance company, seeking to reduce the frequency of malpractice cases, asks you to write queries for checking that a doctor is writing prescriptions in areas where he or she is qualified to do so. Suppose you have a populated table Prescription that does not contain non-primary key attributes syndrome and area-of-practice. Write a query in relational algebra or SQL that will return the set of prescriptions along with all syndromes of the patient and areas of practice of the prescribing physician. This should involve one or more natural joins or Cartesian products, to put together the fields you need by retrieving information from Doctor and Patient, and to match them against Prescription.

- b) Consider the XML record:

```
<drug drug-type = "prescription" >
  <drug-id> c3po </drug-id>
  <drug-name> Zaphat </drug-name>
  <syndrome> Bell-Guillain syndrome </syndrome>
</drug>
```

Give an example of how fields such as `drug-id` can be represented either using *attributes* or *subelements*, and explain how to decide for a schema such as Drug.

**Class participation.** Post your answer to extra credit student feedback question from Exam 2 to the instructor ([CIS560TA-L@listserv.ksu.edu](mailto:CIS560TA-L@listserv.ksu.edu)) before class time Friday, 17 Nov 2006.