

Assignment 2: Database Design

Due on February 14, 2011, at the beginning of class

This assignment is worth 200 points, 20% of your course grade. The assignment can be done in pairs. You can choose any other student as your partner, but remember that you cannot work with the same partner on your final project. If you work as a pair, please submit one assignment with two names. Both partners will get the same grade. Please limit your collaboration to one partner only.

Part A: Designing a Restaurant Database

“Eatr” is a small company that wants to launch a website that would allow people to get information about restaurants in Toronto. Eatr has already done some preliminary user experience research and decided that the app would offer the following features:

1. A visitor to the site, “Alice,” should be able search for restaurants that matching some combination of:
 - The name of the restaurant or a part of a name (e.g., “Maroli”).
 - Tags describing cuisine type / regional origin (e.g., “Uighur”, “Andhra”, “Russian”, “Indian”). Note that the same restaurant could be at the same time “Andhra”, “Indian”, and “Vegetarian”.
 - Specific dishes that available at the restaurant (e.g., “lasagna” or “feijoada”).
 - Location identified as a neighborhood (“the Annex”) or a nearby landmark (“Bloor Theater”).

The search should result in a list of restaurants with a name of each restaurant and its average rating.

2. Alice should be able to click on a restaurant from the list and see the following information about it:
 - The name of the restaurant.
 - Associated tags (see above).
 - The address and phone number.
 - Price range (from “\$” to “\$\$\$\$”).
 - A list of comments from visitors, which would include a rating, from one to five stars.
 - The average of visitor ratings.
 - A list of available dishes.
3. Alice should be able to click on the list of dishes available at the restaurant and see:
 - The rating of this specific dish at this particular restaurant.
 - Comments on with numeric ratings (from one to five stars) for that specific dish.
 - The average rating for the dish
4. When looking at a comment by “Bob”, Alice should be able to click on Bob's name to see Bob's profile and the list of all Bob's comments. The list of Bob's comments show which restaurant / dish each comment refers to.
5. Comments on restaurants and dishes should be “threaded.” That is, Alice should be able to post a comment *in response* to a comment by Bob. The web interface should somehow show which comments are in response to other comments.
6. Alice should be able to mark Bob's comment as “helpful” or “unhelpful.” When applicable, comments would be shown with a note saying something like “24 out of 32 people found this helpful.”

The database does not need to support any features beyond those outlined above.

A.1. A Basic ER Diagram and a Data Dictionary (40 points, week 4 material)

Draw an ER diagram for Eatr's database. At this point, use many-to-many relationships where appropriate and do *not* use associative entities. (We'll consider an entity to be “associative” if it does not have any of its own attributes and just stands for a relationship.) Use the simplified diagram style that we used in class, with either crow’s foot or UML notation for relationships. Label all relationships. Attach a concise explanation for any design choices that may require an explanation. If some choices depend on information not provided in the outline above, then you should explain your assumptions. (Your assumptions should not directly contradict the stated requirements.)

In addition to the diagram, describe each entity in the form of a “data dictionary”: for each entity, provide a short (10-100 words) description of what it represents and fill in a table like this one (one per entity) that lists the attributes, their domains, and their meaning. For example:

World

Describes a planet or moon in the StarWars universe, including planets that are destroyed in the course of the story.

Attribute	Domain	Meaning
world_id	integer	an arbitrary ID to be used as an identifier
world_name	text string, up to 10 characters	the name of the world
percent_water	decimal	the percent of the world's surface that is covered by water.
diameter	decimal	the diameter in kilometers.
destroyed	boolean	true if the world is destroyed in the course of the story.
population	integer	the population of the world (the largest value at any point in the story).

A.2. An ER Diagram with Associative Entities (10 points, week 4 material)

Modify your ER diagram to get rid of all many-to-many relationships, replacing them with associative entities. Explain any decision that may require an explanation.

Please note that you need to provide two different diagrams: one for part 1.1 and another for 1.2.

Part B: Madame Z’s Fortune-Telling Center

Madame Z runs a fortune-telling business, employing a number of fortune tellers. The business has many customers, some of whom come back on a regular basis. Before a visit customers can brows the list of fortune-tellers by name or method used (tarot, astrology, etc.) and can schedule a session with a specific fortune-teller. Repeated customers are usually assigned to one person as their “primary” fortune-teller. When customers come in for a session, a number of predictions can be made, which are saved for future reference. (This ensures that this information is available when making predictions in later sessions.) The customer is billed for the services provided during the session, which can include multiple billable items. They can either receive an invoice after the completion of the session to be paid immediately, or have an invoice sent to them by mail at the end of the month, in which case the invoice would include all services for that month. Customers who receive an invoice by mail can pay it by mailing a check. The payments are recorded when they are received.

Madame Z wants to start using a database for tracking the relevant information, so she hired a consultant who produced the ER shown on the next page.

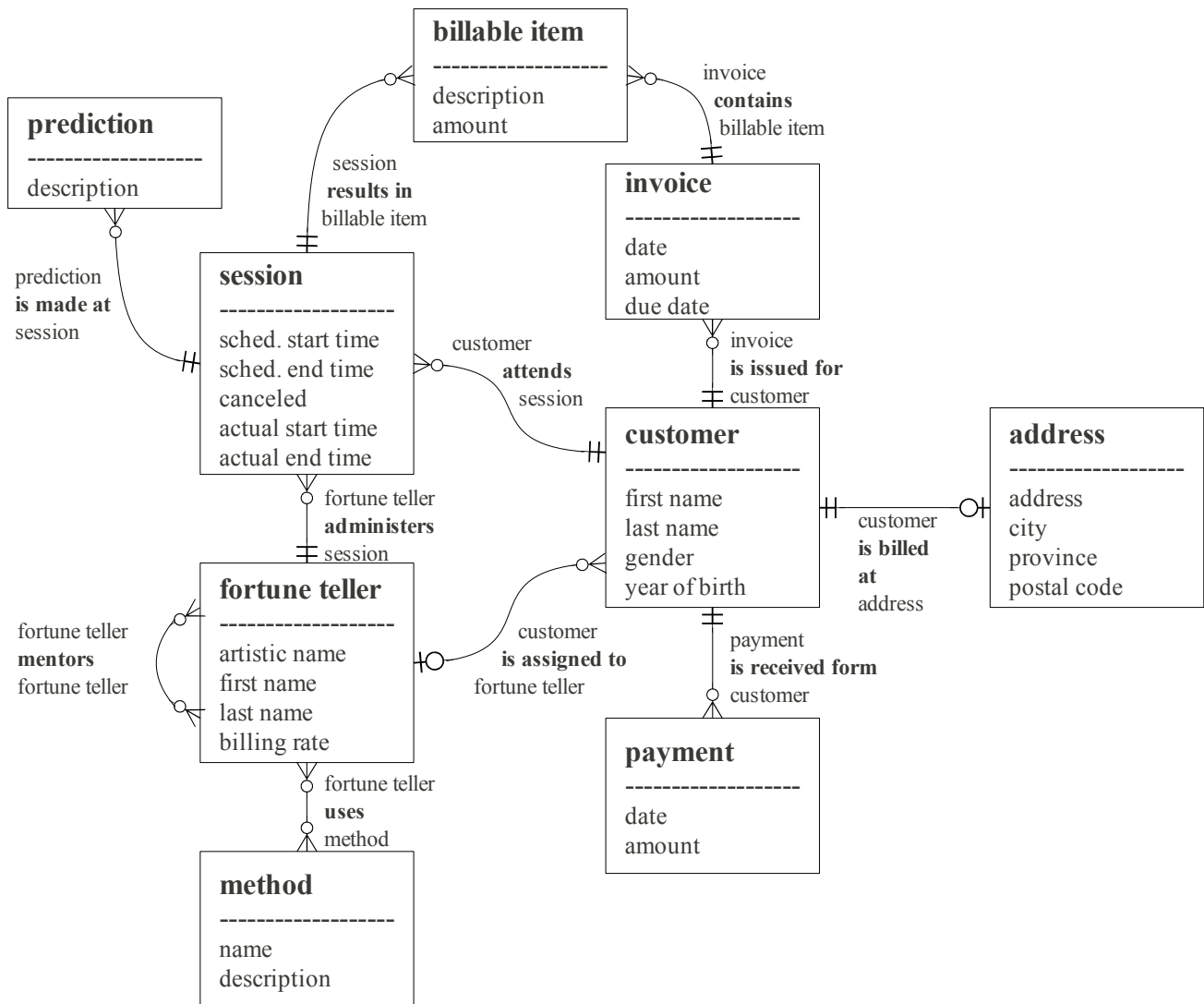


Figure 1. An ER diagram for Madame Z's Fortune-Telling Center.

B.1. Analyzing the Diagram (20 points, mostly week 4 material, but we'll revisit it in week 5)

Using the diagram, please answer the following questions (5 points per question):

- Looking at a particular prediction, is it possible to know who made it, what client it was for, and what method of fortune-telling was used for it? Explain.
- Is it possible to identify to which session a particular payment applies? Explain.
- Invoices are issued at the end of each month. When preparing an invoice for a particular customer, how would the system determine the amount to bill? Would it be possible to send customer an invoice specifying to what session each item corresponds?
- Does the diagram contain any relations that would need to be broken-up with associative entities? If so, which ones? Explain.

Do not use more than 100 words per question.

B.2. Converting the ER Diagram to a Relational Representation (20 points, week 5 material)

Convert the diagram into relational form. Use the notation from Harrington's RDD, chapter 6. E.g.:

Customer (customer_number, customer_first_name, customer_last_name, customer_gender, customer_year_of_birth, customer_primary_teller)

Identify primary keys. Briefly summarize how you arrived at your representation and explain any non-obvious choices.

B.3. Building the Database (20 points, week 6 material)

Write SQL to create the tables for Madame Z's database. Test your code using the database server and then include it in your assignment submission. **Specify the name of the database on the database server in which you are creating the tables.**

B.4. Populating the Database (20 points, week 6 material)

Write SQL to populate the database with the following information:

Jason Smith, male, born in 1972, is signed up for a session with "Dr Rain", from 2-3 pm on Oct. 29, 2010.

Pete Brown, male, has Dr Rain as the primary teller but signed up for a session with Madame Z herself. The session was scheduled to take place from noon to 2 pm on Oct. 6, but Jason showed up late. The session began at 12:30 and went until 2 pm. Pete was charged \$500 for the session. He received his invoice immediately upon the completion of the session and paid it in full right away.

Dr Rain's real name is Robert Green. He uses two methods: Tarot cards and astrology. No description is available for Tarot cards, but astrology is described as "Determining the future by the position of celestial bodies." Dr Rain is mentored by Madame Z.

Populate the database you created on the server and test it. Include your SQL in the assignment.

Part C: Normalizing World Development

Go to the Millennium Development Goals dataset in the World Bank Data Catalog website:

<http://data.worldbank.org/data-catalog/millennium-development-indicators>

Download the Excel version of the datafile:

http://databank.worldbank.org/databank/download/MDG_Excel.zip

The files contain 5 sheets, not counting the "Description" sheet.

C.1. Analysis (8 points per sheet, 40 points total, week 5 material)

For each of the five sheets as a database table, identify all functional dependencies and candidate keys and determine whether the requirements for the 1NF, 2NF, and 3NF are satisfied. Use the format from Harrington's RDD (p. 112.) for expressing functional dependencies. Explain why you believe the normalization requirements are satisfied or violated.

C.2. Normalization (30 points, week 5 material)

Show how the same data can be represented as a set of tables that satisfy 3NF.