Assignment: 4

Course: BMI 506

Date: April 13, 2015

Goal: Testing and Validation. Demonstrate, using working prototype OR mockup, the function of your CDSS. Develop test cases designed to identify various aspects of its operation and possible points of failure. May include visualization methods and feedback from representative users.

Project Title: Patients Like Mine (PLM)

Team Members: Matt Halbert (591), Eric Holden (506), Tara Salehpour (506,591), Michelle Winerep (591)

Contributions:
- Tara: Test Cases: Database Model
- Eric: Prototype: SQL Query

Background

The goal of this project is to implement a tool within the framework of the PLM system that allows a domain expert to author a template for use in the generation of a database query. The templates are composed of an input template, which defines the variables, their ranges, and their relative timeframes in order to determine which patients are considered similar to the patient of interest, and an output template, which defines the variables and their relative timeframe to be extracted from the cohort of similar patients. The authoring tool will allow the domain expert to select the variables to be included in both the input and output templates from an internal knowledge base. The template will then be used in conjunction with a patient of interest's information in order to generate a database query which will then be run on the past patients database(s) to extract information from patients that are determined to be suitably similar.

Test Cases: Database Model

Guided by Dr. Peter Li's expertise, the following variables were selected for use in defining the input and output templates: Blood Pressure, Hematocrit, Hemoglobin, Blood Transfusion, White Blood Cell (WBC) Count, Temperature, Anticoagulants (Medication), Colorectal Surgery (date), Demographics, Age (DOB) and Gender.

A database model was designed to store SQL query-able patient test case information. The model represents a simplified version of the data structure models provided by Dr. Peter. Figure 1 illustrates the relational schema for the database. Each relation represents a table in the database. (See the appendix for database tables with patient data and SQL database model file.) The attributes of interest are stored in the following tables:

Patient

- Date of Birth
- Gender
- Demographics

Procedure:

- Colorectal Surgery
- Blood Transfusion

Observation:

- Blood Pressure
- Hematocrit
- Hemoglobin
- White Blood Cell (WBC) Count
- Temperature

Medication Administration:

- Anticoagulants, Blood Thinners

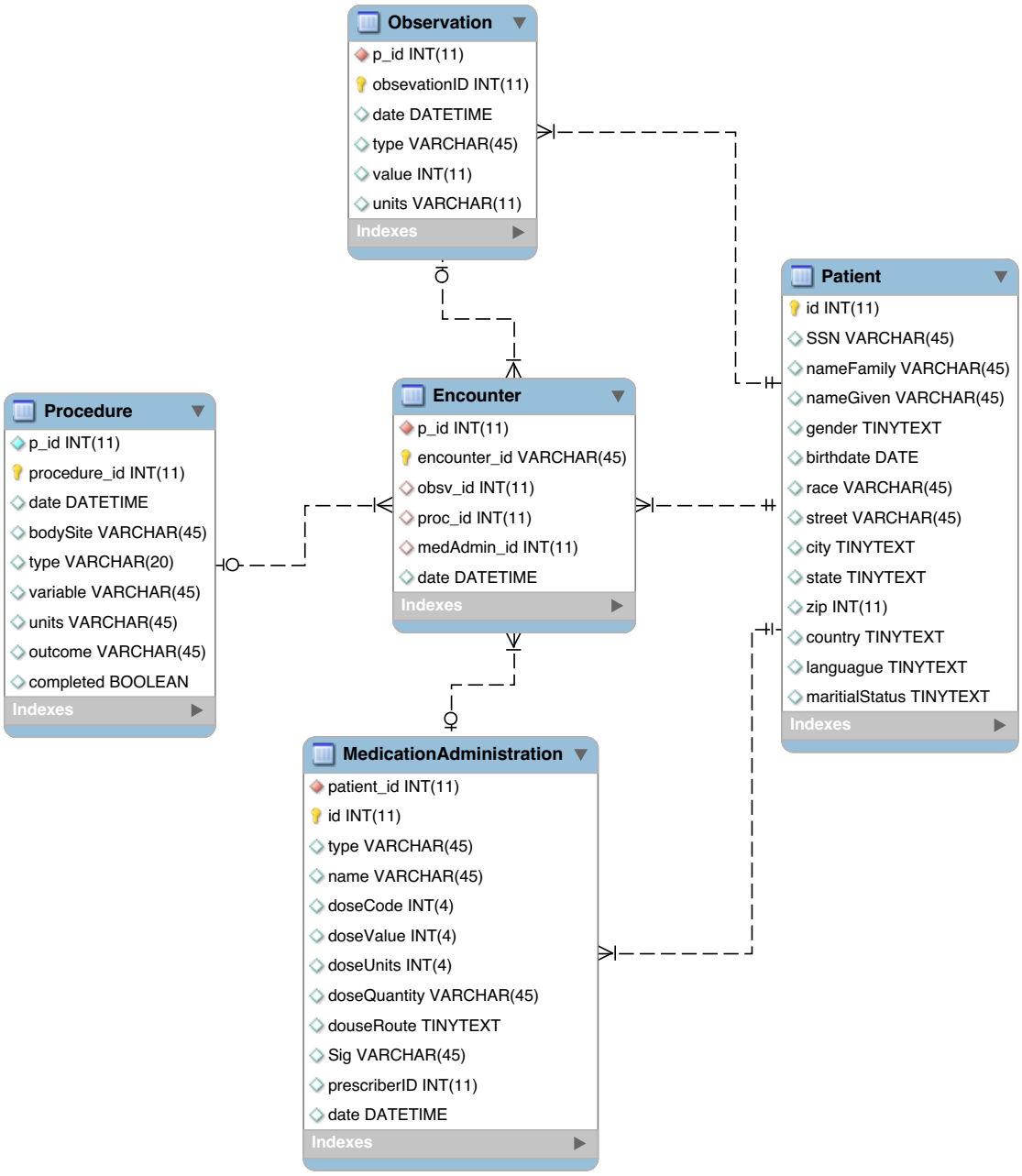

Figure 1: Relational Schema: Patient Database Model

The manufactured mock patient database includes test cases from five patients. Patient data was engineered to represent diverse patient information relevant to the risk of post-op colorectal bleed. Figure 2 summarizes all the relevant variables per patient. Each patient has two sets of values for each variable of interest. The first set of values represents a baseline measurement taken prior to surgery. The second set of values parallels post-operation values within three days of colorectal surgery. The last column displays the percent change relative to the baseline value. Negative percent change values are right justified. Values less than or equal to -10% are in

pID 1 Corona, Joshua **Age: 37 Male**

Observations	Jan-01	Jan-02	% Change
Hematocrit (%)	41	38	-7%
Hemoglobin (g/dL)	17	15	-12%
WBC (mcL)	6500	6399	-2%
Systolic BP (mm Hg)	111	107	-4%
Diastolic BP (mm Hg)	73	73	0%
Temperature (°C)	37	37	0%

Current Medication		
Warfarin	3 mg	1 po daily

pID 2 Bella, Calista **Age: 52 Female**

Observations	Mar-07	Mar-08	% Change
Hematocrit (%)	41	44	7%
Hemoglobin (g/dL)	13	12	-8%
WBC (mcL)	6600	6500	-2%
Systolic BP (mm Hg)	108	110	2%
Diastolic BP (mm Hg)	71	73	3%
Temperature (°C)	37	37.5	1%

pID 3 Kyle, Selina **Age: 74 Female**

Observations	Mar-12	Mar-13	% Change
Hematocrit (%)	57	54	-5%
Hemoglobin (g/dL)	15	13	-13%
WBC (mcL)	6500	6399	-2%
Systolic BP (mm Hg)	108	103	-5%
Diastolic BP (mm Hg)	79	71	-10%
Temperature (°C)		37	

pID 4 Kerr, Joe **Age: 82 Male**

Observations	Apr-01	Apr-03	% Change
Hematocrit (%)	50	55	10%
Hemoglobin (g/dL)	14	13.5	-4%
WBC (mcL)	6399	6200	-3%
Systolic BP (mm Hg)	119	118	-1%
Diastolic BP (mm Hg)	79	77	-3%
Temperature (°C)		37.11	

pID 5 Wayne, Bruce **Age: 30 Male**

Observations	Apr-01	Apr-03	% Change
Hematocrit (%)	50	39	-22%
Hemoglobin (g/dL)	17.5	16	-9%
WBC (mcL)	7000	6000	-14%
Systolic BP (mm Hg)	119	117	-2%
Diastolic BP (mm Hg)	77	77	0%
Temperature (°C)			

Current Medication		
Warfarin	5 mg	1 po daily

Figure 2: Test Cases Summary: Designed to show relevant attribute ranges

Prototype: SQL Query

In order to demonstrate how a template can be used in conjunction with patient information in order to generate an SQL query, we developed a rough tool to serve as a proof-of-concept. This tool was developed in Java and uses input template, output template, and patient text files to generate the appropriate query. Of course, being a proof-of-concept tool, certain details of the implementation such as the exact format of the templates and patient information, as well as the structure of the database to be queried are all placeholder values that have little bearing on practical use. An example set of inputs and outputs will be shown below to illustrate how the query is generated. The input template illustrated in Figure 3 below follows the same basic schema described by the program mock-up in assignment 3. The structure is as follows: variable name, lower bound, upper bound, and relative timeframe. The lower and upper bound fields can be either percentages or predefined values. Percentages will be taken in relation to the patient of interest's particular values. Predefined numbers, however, will be treated differently depending on the presence of a numerical sign. Numbers without a sign are considered absolute and will be included to the query as-is, with no regard to the patient's value. Numbers with a sign are considered relational and will be added or subtracted to the patient of interest's value as needed. For demographic variables such as gender where there is no clear range of similarity or timeframe, the value N/A is acceptable. As described in assignment 3, the output template (Figure 4) only defines the variable to be extracted and its relative timeframe. There is no defined similarity in this case, since the variables of interest will be extracted as-is.

```
Hemoglobin, -10%, +10%, Post-OP Day 2;
WBC, 5000, 7000, Post-OP Day 2;
Hematocrit, 0%, +5, Post-OP Day 2;
Gender, N/A, N/A, N/A;
Temperature, -10%, +10%, Post-OP Day 2;
Blood Pressure, -10%, +10%, Post-OP Day 2;
```

Figure 3: Input Template (Text Representation)

```
Transfusion, Post-OP Day 3;
Temperature, Post-OP Day 3;
```

Figure 4: Output Template (Text Representation)

For the purposes of this prototype, the patient information (Figure 5) has been encoded using the text identifiers EVENT, TEST, and DEMOGRAPHIC. Events are used to calculate the relative timeframe by denoting important milestones in the patient's care. In this case, the event of interest is colorectal surgery. Tests denote numerical values that can have a range of similarity applied via the input template. Demographics denote text information that has no real range of similarity. The exception to this is the date of birth variable that can be used to calculate the patients age if needed. Again, dates do not apply to demographic information, but do apply to test and event information.

```
EVENT, Colorectal Surgery, SURGERY, 01/01/2015;
DEMOGRAPHIC, Race, Caucasian, N/A;
DEMOGRAPHIC, Gender, Male, N/A;
DEMOGRAPHIC, DOB, 01/01/1985, N/A;
TEST, WBC, 6000, 01/03/2015;
TEST, Hemoglobin, 16, 01/03/2015;
TEST, Hematocrit, 50, 01/03/2015;
TEST, Temperature, 98.6, 01/02/2015;
```

Figure 5: Patient Information (Text Representation)

Using the tool we developed with the information provided above produces the output found in Figure 6. Again, it should be noted that the database structure (the FROM and JOIN statements) is merely a placeholder, and would change in actual implementation. The main idea is that the JOIN statements will encompass the entirety of a patient's information so as to eliminate complexity during query generation. The downside is that performing the search would take slightly longer for large databases. The date value for each particular variable may also be less clearly defined in actual practice, but it was determined that simplicity was best for the purposes of this prototype. This particular example covers a number of potential complications. For example, the temperature and blood pressure values in the input template have not been included in the query. The blood pressure value was not included because there was no recorded blood pressure measurement in the patient of interest's record. The temperature value was not included because, even though a temperature value was in the patient's record, the date did not match the relative timeframe defined in the input template. The code below also takes into account similarity ranges created using a combination of percentage values as well as absolute and relative predefined values.

```
SELECT Transfusion, Temperature FROM Colorectal
JOIN Patient ON Colorectal.PatientID = Patient.PatientID
JOIN Observation ON Patient.PatientID = Observation.PatientID
JOIN Procedure ON Patient.PatientID = Procedure.PatientID
JOIN Encounter ON Patient.PatientID = Encounter.PatientID
JOIN MedicationAdministration ON Patient.PatientID = MedicationAdministration.PatientID
WHERE TransfusionDate = "01/04/2015"
AND TemperatureDate = "01/04/2015"
AND Hemoglobin BETWEEN 14.4 AND 17.6
AND HemoglobinDate = "01/03/2015"
AND WBC BETWEEN 5000.0 AND 7000.0
AND WBCDate = "01/03/2015"
AND Hematocrit BETWEEN 50.0 AND 55.0
AND HematocritDate = "01/03/2015"
AND Gender = "Male";
```

Figure 6: Sample SQL Output

Appendix: Database Patient Tables

These tables represent how the patient data is stored in the relational database.

Patient

id	SSN	nameFamily	nameGiven	gender	birthdate	race	street	city	state	zip	country	languague	maritialStatus
1	111-11-1111	Corona	Joshua	Male	08/08/1977	Caucasian	123 E 1st St.	Corvallis	OR	97333	USA	english	single
2	222-22-2222	Bella	Calista	Female	06/11/1962	Other	123 E 2nd St.	Corvallis	OR	97333	USA	english	married
3	333-33-3333	Kyle	Selina	Female	07/12/1940	Hispanic	123 E 3rd St.	Corvallis	OR	97333	USA	english	single
4	444-44-4444	Kerr	Joe	Male	11/06/1932	Caucasian	123 E 4th St.	Corvallis	OR	97333	USA	english	married
5	555-55-5555	Wayne	Bruce	Male	01/01/1985	Caucasian	124 E 5th St.	Corvallis	OR	97333	USA	english	married

Procedure

p_id	procedure_id	date	bodySite	type	variable	units	outcome	completed
1	1	1/1/15 3:01	Colorectal	SURGERY	NULL	NULL	complication	TRUE
2	2	3/7/15 7:01	Colorectal	SURGERY	NULL	NULL	success	TRUE
3	3	3/12/15 7:01	Colorectal	SURGERY	NULL	NULL	complication	TRUE
4	4	4/1/15 7:01	Colorectal	SURGERY	NULL	NULL	success	TRUE
5	5	1/1/15 0:00	Colorectal	SURGERY	NULL	NULL	success	TRUE
5	6	1/2/15 0:00	subq	Transfusion	1	unit	NULL	TRUE

Medication Administration

patient_id	id	type	name	doseCode	doseValue	doseUnits	doseQuantity	douseRoute	Sig	prescriberID	date
1	1	anticoagulant	warfarin	completed	3	mg	30	by mouth	1 po daily	2222	01/01/2011
1	2	anticoagulant	warfarin	completed	3	mg	30	by mouth	1 po daily	2222	01/01/2012
1	3	anticoagulant	warfarin	completed	3	mg	30	by mouth	1 po daily	2222	01/01/2013
1	4	anticoagulant	warfarin	completed	3	mg	30	by mouth	1 po daily	2222	01/01/2014
2	5	analgesic	hydrocodone	completed	5	mg	120	by mouth	1-2 po q 4-6 h prn	3333	02/15/2014
1	6	antibiotic	amoxicillin	completed	120	mg	60	by mouth	1 po bid	1111	11/06/2014
1	7	anticoagulant	warfarin	in progress	3	mg	30	by mouth	1 po daily	2222	01/01/2015
5	8	anticoagulant	warfarin	in progress	5	mg	30	by mouth	1 po daily	4444	01/01/2015

Encounter

p_id	enc_id	obsv_id	proc_id	nedAdm_id	date
1	1	51	NULL	1	01-01-2011
1	2	52	NULL	2	01-01-2012
1	3	53	NULL	3	01-01-2013
1	4	54	NULL	4	01-01-2014
2	5	55	NULL	5	02-15-2014
1	6	56	NULL	6	11-06-2014
1	7	NULL	NULL	7	01-01-2015
5	8	NULL	NULL	8	01-01-2015
5	9	NULL	5	NULL	01-01-2015
1	10	1	NULL	NULL	01-01-2015
1	11	2	NULL	NULL	01-01-2015
1	12	3	NULL	NULL	01-01-2015
1	13	4	NULL	NULL	01-01-2015
1	14	5	NULL	NULL	01-01-2015
1	15	6	NULL	NULL	01-01-2015
1	16	7	NULL	NULL	01-01-2015
1	17	NULL	1	NULL	01-01-2015
5	18	NULL	6	NULL	01-02-2015
1	19	8	NULL	NULL	01-02-2015
1	20	9	NULL	NULL	01-02-2015
1	21	10	NULL	NULL	01-02-2015
1	22	11	NULL	NULL	01-02-2015
1	23	12	NULL	NULL	01-02-2015
5	24	13	NULL	NULL	01-03-2015
5	25	14	NULL	NULL	01-03-2015
5	26	15	NULL	NULL	01-03-2015
5	27	16	NULL	NULL	01-03-2015
5	28	17	NULL	NULL	01-03-2015
1	29	18	NULL	NULL	01-03-2015
2	30	19	NULL	NULL	03-07-2015
2	31	20	NULL	NULL	03-07-2015
2	32	21	NULL	NULL	03-07-2015
2	33	22	NULL	NULL	03-07-2015
2	34	23	NULL	NULL	03-07-2015
2	35	NULL	2	NULL	03-07-2015
2	36	24	NULL	NULL	03-08-2015
2	37	25	NULL	NULL	03-08-2015
2	38	26	NULL	NULL	03-08-2015
2	39	27	NULL	NULL	03-08-2015
2	40	28	NULL	NULL	03-08-2015
2	41	29	NULL	NULL	03-09-2015
3	42	30	NULL	NULL	03-12-2015
3	43	31	NULL	NULL	03-12-2015
3	44	32	NULL	NULL	03-12-2015
3	45	33	NULL	NULL	03-12-2015
3	46	34	NULL	NULL	03-12-2015
3	47	NULL	3	NULL	03-12-2015
3	48	35	NULL	NULL	03-13-2015
3	49	36	NULL	NULL	03-13-2015
3	50	37	NULL	NULL	03-13-2015
3	51	38	NULL	NULL	03-13-2015
3	52	39	NULL	NULL	03-13-2015
3	53	40	NULL	NULL	03-13-2015
4	54	41	NULL	NULL	04-01-2015
4	55	42	NULL	NULL	04-01-2015
4	56	43	NULL	NULL	04-01-2015
4	57	44	NULL	NULL	04-01-2015
4	58	45	NULL	NULL	04-01-2015
5	59	46	NULL	NULL	04-01-2015
5	60	47	NULL	NULL	04-01-2015
5	61	48	NULL	NULL	04-01-2015
5	62	49	NULL	NULL	04-01-2015
5	63	50	NULL	NULL	04-01-2015
4	64	NULL	4	NULL	04-01-2015

Observation

p_id	rationID	date	type	value	units
1	5	1/1/15 1:01	diastolic blood pressure	73	mm Hg
1	7	1/1/15 1:01	Hematocrit	41	%
1	2	1/1/15 1:01	Hemoglobin	17	g/dL
1	4	1/1/15 1:01	systolic blood pressure	111	mm Hg
1	6	1/1/15 1:01	temperature	37	°C
1	3	1/1/15 1:01	WBC	6500	WBC/mcL
1	11	1/2/15 1:01	diastolic blood pressure	73	mm Hg
1	12	1/2/15 1:01	Hematocrit	38	%
1	8	1/2/15 1:01	Hemoglobin	15	g/dL
1	10	1/2/15 1:01	systolic blood pressure	107	mm Hg
1	9	1/2/15 1:01	WBC	6399	WBC/mcL
1	18	1/3/15 1:01	temperature	37	°C
2	22	3/7/15 3:01	diastolic blood pressure	71	mm Hg
2	23	3/7/15 3:01	Hematocrit	37	%
2	19	3/7/15 3:01	Hemoglobin	13	g/dL
2	21	3/7/15 3:01	systolic blood pressure	108	mm Hg
2	20	3/7/15 3:01	WBC	6700	WBC/mcL
2	27	3/8/15 1:01	diastolic blood pressure	73	mm Hg
2	28	3/8/15 1:01	Hematocrit	44	%
2	24	3/8/15 1:01	Hemoglobin	12	g/dL
2	26	3/8/15 1:01	systolic blood pressure	110	mm Hg
2	25	3/8/15 1:01	WBC	0	WBC/mcL
2	29	3/9/15 1:01	temperature	37.5	°C
3	33	3/12/15 3:01	diastolic blood pressure	79	mm Hg
3	34	3/12/15 3:01	Hematocrit	57	%
3	30	3/12/15 3:01	Hemoglobin	15	g/dL
3	32	3/12/15 3:01	systolic blood pressure	108	mm Hg
3	31	3/12/15 3:01	WBC	6500	WBC/mcL
3	38	3/13/15 3:01	diastolic blood pressure	71	mm Hg
3	39	3/13/15 3:01	Hematocrit	54	%
3	35	3/13/15 3:01	Hemoglobin	13	g/dL
3	37	3/13/15 3:01	systolic blood pressure	103	mm Hg
3	40	3/13/15 3:01	temperature	37	°C
3	36	3/13/15 3:01	WBC	6500	WBC/mcL
4	44	4/1/15 3:01	diastolic blood pressure	79	mm Hg
4	45	4/1/15 3:01	Hematocrit	50	%
4	41	4/1/15 3:01	Hemoglobin	14	g/dL
4	43	4/1/15 3:01	systolic blood pressure	119	mm Hg
4	42	4/1/15 3:01	WBC	6399	WBC/mcL
4	54	4/3/15 3:01	diastolic blood pressure	77	mm Hg
4	55	4/3/15 3:01	Hematocrit	55	%
4	51	4/3/15 3:01	Hemoglobin	13.5	g/dL
4	53	4/3/15 3:01	systolic blood pressure	118	mm Hg
4	56	4/3/15 3:01	temperature	37.11	°C
4	52	4/3/15 3:01	WBC	6200	WBC/mcL
5	49	4/1/15 3:01	diastolic blood pressure	77	mm Hg
5	50	4/1/15 3:01	Hematocrit	50	%
5	46	4/1/15 3:01	Hemoglobin	17.5	g/dL
5	48	4/1/15 3:01	systolic blood pressure	119	mm Hg
5	47	4/1/15 3:01	WBC	7000	WBC/mcL
5	16	4/3/15 0:00	diastolic blood pressure	77	mm Hg
5	13	4/3/15 0:00	Hemoglobin	16	g/dL
5	15	4/3/15 0:00	systolic blood pressure	117	mm Hg
5	17	4/3/15 0:00	temperature	39	°C
5	14	4/3/15 0:00	WBC	6000	WBC/mcL