

# WebM.D+

## WHAT DISEASE DO YOU HAVE?

Using...  
**SQL** &  
**Machine Learning**

**Final Project**  
— CSCI 403 —  
**Database Management**

### Interesting Queries

```
def create_new_patient(cursor, name, age, sex):  
    sexStr = str()  
    sexStr = "male" if sex == 0 else "female"  
    query = "INSERT INTO patient (name, age, sex) VALUES (%s, %s, %s)"  
    cursor.execute(query, (name, age, sexStr,))
```

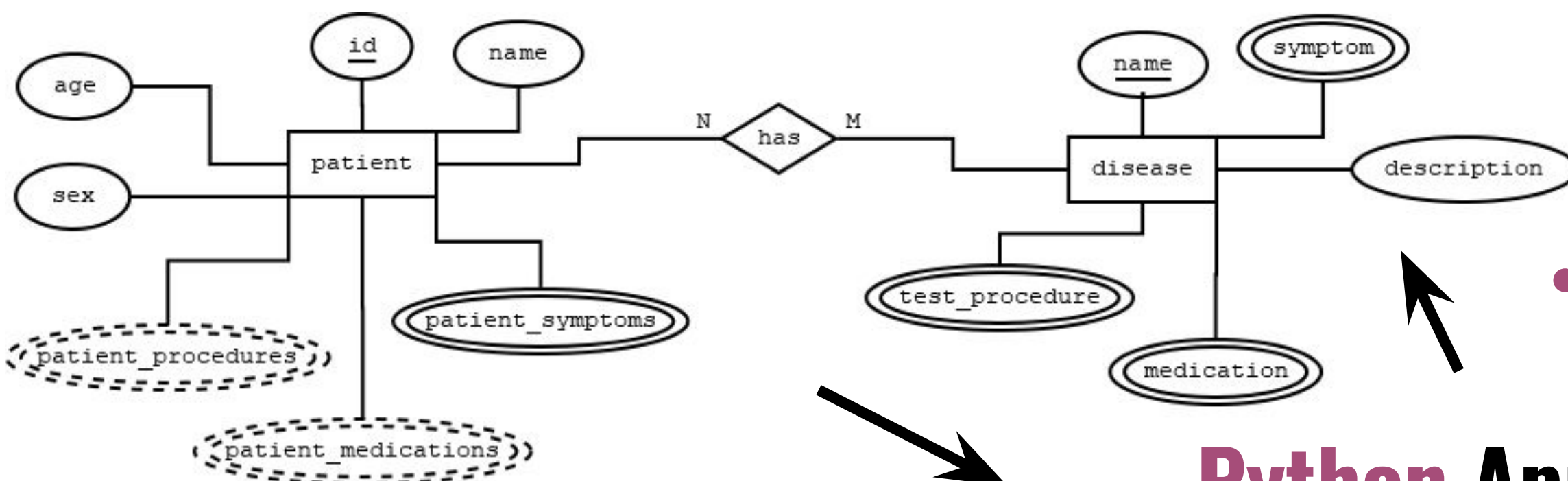
```
SELECT name, description, symptom  
FROM disease d, disease_symptoms ds  
WHERE d.name = ds.disease_name;
```

```
SELECT name, description, symptom  
FROM disease d, disease_symptoms ds  
WHERE d.name = ds.disease_name  
AND d.name ILIKE 'f%';
```

```
SELECT name, COUNT(symptom)  
FROM disease d, disease_symptoms ds  
WHERE d.name = ds.disease_name  
GROUP BY d.name;
```

```
SELECT name, COUNT(symptom)  
FROM disease d, disease_symptoms ds  
WHERE d.name = ds.disease_name  
AND d.name IN  
    (SELECT disease_name FROM medication  
     WHERE medication_name ILIKE '%ine'  
     AND medication_name NOT ILIKE 'obsolete')  
GROUP BY d.name;
```

### Database ERD



• Results of app queried back into database

### Python Application

### Random Forest Classifier



#### Database

- Two main entities, patients and diseases
- Disease table maps 800 diseases to their related symptoms, medications, and procedures
- Patient table stores resulting diagnosis data

Welcome to WebMD+

**Patient Info**

Sex: ☒ male ☐ female

Age:

First Name:

Last Name:

Symptom 1:

Symptom 2:

Symptom 3:

Symptom 4:

Symptom 5:

Predicted Disease/Affliction:

Other Possible Diagnoses:

- Panic disorder
- Eating disorder
- Substance-related mental disorder
- Postpartum depression

#### WebM.D+ App

- Takes up to five of user's symptoms
- Feeds inputs into classifier
- Outputs predicted diagnosis
- Queries for other linked diseases

#### Machine Learning

- Uses sklearn's Random Forest Classifier.
- Predicts disease given list of symptoms
- Labels = 800 unique diseases