Gillian McClennen, Hamsini Malli Sivakumar, Michelle Wilfred

Dr. Wolfgang Gatterbauer

Database Design - CS3200

7 December 2022
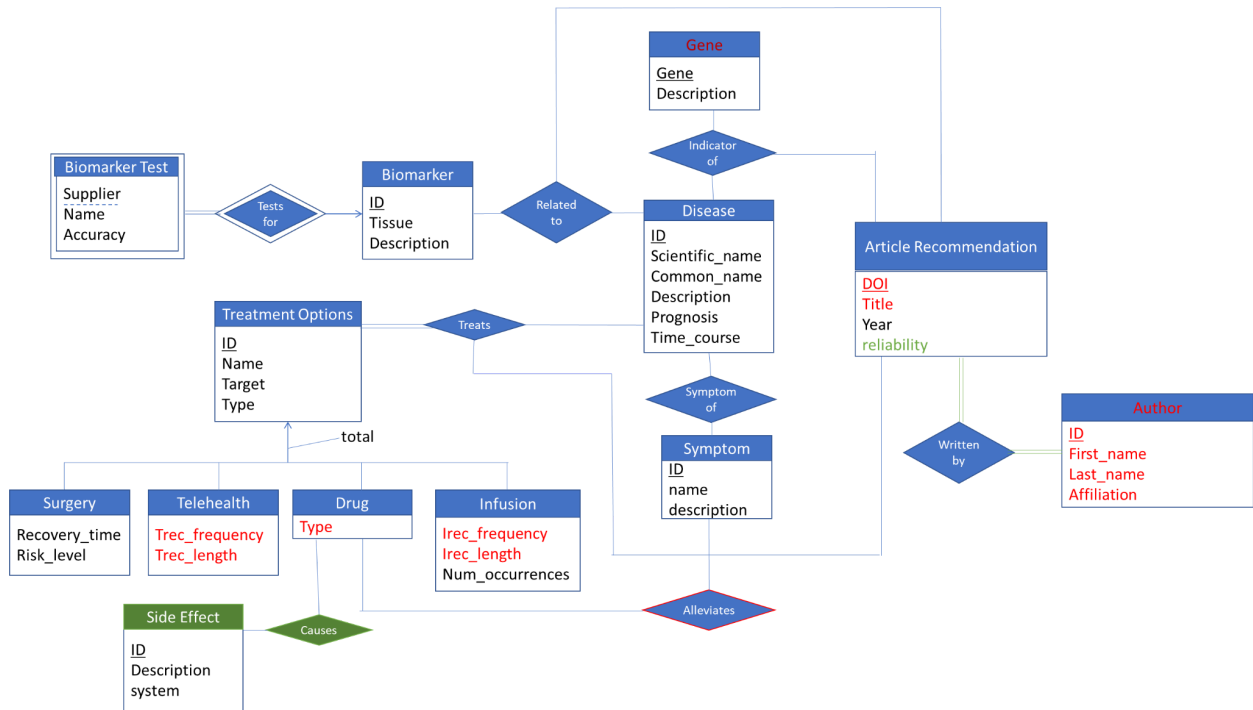
Phase 5: Additions & Corrections On The Doctor's Diagnosis Tool Database

Amendment Notes:

1. The reviewers asked for a list of items that we considered adding to the database that we couldn't add due to time and research limitations. We could have identified some way to connect side effects to diseases, given that some side effects are diseases or symptoms in themselves. For example, some cancer drugs cause headaches, which deserves a treatment of its own, but we chose to separate these two entities for simplicity. Additionally, we chose to have biomarker tests check for only one biomarker, when in reality, the relationship is much more complex and could test for multiple biomarkers.

2. The review team suggested that we enforce a mandatory participation constraint for the author and article recommendation relationship which we adjusted. This makes sense since we would want our database to store only the authors who have written articles relevant to the various treatments, biomarkers, and genes that we are keeping track of.

3. As mentioned by the professor, the initial ER diagram structure prevented different drugs from having the same side effect and also prevented one drug from having various side effects. We corrected this by adding side effect as a separate entity which also allowed us to add more attributes to side effect, making our design more flexible and logically better.

4. Pointed out by both the professor and the review team, Query 1 has no need for article or author relationships as the table 'diseasegenearticle' already ensures that the gene is recommended by an article. Refer to textfile for updated query.

5. The reviewers noted that when running Query 3, the null values such as '2004' were placed above the 'no-recommendation'. To remedy this, we changed the query so that it sorts by X.affiliation instead. Refer to textfile for updated query.

6. The reviewers suggested that in Query 4, adding a nested query every time a symptom needed to be searched for would be inefficient as more symptoms needed to be added. The changes in this query allow the user to get the disease that contains all the symptoms listed, assuming the count of symptoms is changed to be equal to the number of symptoms the user is searching for. Refer to textfile for updated query.

Updated ER Diagram:

** updates are highlighted in green

ER Diagram entities and relationships:

- **Gene**: Gene, Description
- **Biomarker Test**: Supplier, Name, Accuracy — *Tests for* → **Biomarker**: ID, Tissue, Description
- **Biomarker** — *Related to* → **Disease**
- **Gene** — *Indicator of* → **Disease**
- **Disease**: ID, Scientific_name, Common_name, Description, Prognosis, Time_course
- **Article Recommendation**: DOI, Title, Year, reliability
- **Treatment Options**: ID, Name, Target, Type — *Treats* → **Disease**
- **Author**: ID, First_name, Last_name, Affiliation — *Written by* → Article Recommendation
- **Disease** — *Symptom of* → **Symptom**: ID, name, description
- Treatment Options (total): **Surgery** (Recovery_time, Risk_level), **Telehealth** (Trec_frequency, Trec_length), **Drug** (Type), **Infusion** (Irec_frequency, Irec_length, Num_occurrences)
- **Side Effect**: ID, Description, system — *Causes*
- *Alleviates*

Updated Relational Schema:

** updates are highlighted in green

- **biomarkerDiseaseArticle**: Biomarker_id (FK), DOI (FK), Disease_id (FK)
- **article**: DOI, title, Year, reliability
- **writing**: Author_id (FK), DOI (FK)
- **author**: ID, First_name, Last_name, affiliation
- **biomarkerTest**: biomarker_id (FK), supplier, name, accuracy
- **biomarker**: ID, tissue, description
- **disease**: ID, scientific_name, common_name, description, Prognosis, Time_course
- **diseaseGeneArticle**: Gene (FK), Disease_id (FK), DOI (FK)
- **gene**: Gene, Description
- **symptom**: ID, name, description
- **diseaseSymptom**: symptom_id (FK), disease_id (FK)
- **treatmentRecommendation**: Treatment_id (FK), Disease_id (FK), DOI (FK)
- **prevention**: drug_ID (FK), symptom_ID (FK)
- **treatment**: ID, name, target, type
- **sideEffect**: ID, description, system
- **drugSideEffect**: drug_ID (FK), side_effectID (FK)
- **drug**: dtreatment_ID (FK), type
- **infusion**: itreatment_ID (FK), irec_frequency, irec_length, num_occurrences
- **surgery**: streatment_id (FK), recovery_time, risk_level
- **telehealth**: ttreatment_ID (FK), trec_frequency, trec_length