

“Drug - Adverse Reactions” knowledge graph generator

Introduction:

In our project, we built a knowledge graph for drugs and their adverse reactions that have been reported. The knowledge graph will explain the possible reasons why a drug causes a side effect.

The knowledge graph contains four types of nodes: drug, adverse reactions, medical subject headings retrieved from the MeSH database, GO terms database. We extract the reported drug-adverse reaction pair from FDA database, and mine knowledge for each edge from PubMed literature database. Finally, we use the networks to give a possible explanation of the relation between drugs and adverse reactions.

Members:

Huamin Zhang
Sixuan Shen
Shijie Guo
Yichang Liang

Data source

Medical Subject Headings (MeSH) database: MeSH is a comprehensive controlled vocabulary for the purpose of indexing journal articles and books in the life sciences; it serves as a thesaurus that facilitates searching.

The Gene Ontology (GO) project provides the most comprehensive resource currently available for computable knowledge regarding the functions of genes and gene products

FDA Adverse Event Reporting System (FAERS) database: Raw patient case report from one quarter in xml format, roughly 350,000 cases for each quarter.

PubMed database: PubMed comprises more than 28 million citations for biomedical literature from MEDLINE, life science journals, and online books.

Data Retrieval & Processing

Mesh, GO term, FDA are all in XML format.

Use python script to extract all the necessary information in these XML files, and only keep the string titles.

Combine Mesh and GO term as term set; Use FDA to generate drug set and adverse reaction set.

Relation Matrix

Build a dictionary for all the medical terms appeared in literatures in PubMed database. The key is literature ID and the value is a list of all medical terms appeared in the literature.

Based on the dictionary, calculate the relationship between every two terms and form a Relation Matrix, which will be used for searching knowledge graphs.

Search Algorithm

Use IDDFS (Iterative Deepening Depth First search) to search the matrix and output paths.

Convert paths to a json-style file for knowledge graph visualization.

visualization

Built a website.

User can search based on drug and ADR, the result is a network of the resulted knowledge graph.