

Building Mental Health Knowledge Graph

Group Members: Xiaoying Zhang, Yunrui Shao

Project Domain and Goal

With the rapid development of the contemporary world, young people's mental health problems have gradually become one of the focuses of our society. As graduate students, except for taking care of all the coursework, how to properly relieve stress and prevent the potential occurrence of mental illness has become a significant topic for us.

We want to build a knowledge graph to treat mental health disorders and associated symptoms. The knowledge graph will include information about detailed descriptions of different illnesses, their categories, their treatments, healing drugs (including generic-name, brand-name, drug class, description, etc.), and recommended therapists/ psychiatrists sorted by locations (including name, email, tel number, introduction, etc.). This knowledge graph could not only help patients to seek proper treatments and professionals in their locations but also offer ways for those potential patients to detect whether they need further help or not. Utilizing the string-match method, we also plan to write a service that could detect and extract keywords from the user-input symptoms, and return the feedback on the possible illness that the user might be faced with.

Datasets and Representation

The knowledge graph will contain data from 3 sources. We will use [Wikidata](#) as our structured source, and we will crawl our other data from [Psychology today](#) and [Drugs.com](#). From Wikidata, we plan to get basic information about different mental health illnesses, and the related drugs; from Psychology today, we plan to get therapists/ psychiatrists, their contact information, and their area of expertise from different locations in the US; from Drugs.com, we plan to find detailed information about specific drugs. There is no existing ontology for our project, so we will design an ontology that includes all the functions we need (There will be 3 main elements: mental health illness, therapists/ psychiatrists, and drugs).

Technical Challenge

In this project, we need to learn new technologies and new theories quickly and solve challenges that may arise in each part.

We will use Scrapy to crawl data and SPARQL queries to collect data from websites, then use SpaCy to extract meaningful information. It is a challenge to ensure data accuracy, especially for unstructured data. We need to remove noisy data and dedupe. Different datasets may use different words for the same disease and drugs. For this issue, we will use the string-match method and check part of the data manually in the process of entity linking.

After constructing the knowledge graph database by Neo4j, We will provide users with search and recommendation services. In order to adapt to the questions of all kinds of people, we will provide keyword search and user-input symptoms search, and we will process the users' text descriptions. To ensure the quality of our knowledge graph, we will build a suitable set of test cases and continuously improve its believability.

The above are the foreseeable technical challenges, we will also deal with issues that arise in various stages at any time.