CS 410: Text Information Systems  
Final Report

Search System For Hospital Transparency Data

Tony Law, Morgan Whitman, Aareana Reza

## [tonylaw2@illinois.edu](mailto:tonylaw2@illinois.edu) | [whitman5@illinois.edu](mailto:whitman5@illinois.edu) | [areza5@illinois.edu](mailto:areza5@illinois.edu)

**Table of Contents**

Overview Page 1

Implementation and Usage Page 2-4

Contributions Page 5

Overview

**An overview of the function of the code (i.e., what it does and what it can be used for).**

Healthcare costs make up one of individuals and employers largest expenses and prices continue to rise. This puts a significant burden on individuals in health plans. New legislation in recent years has required hospitals to provide standard prices on publicly available websites. This data is accessible on hospital or healthy system websites but not easily accessible to individuals searching by service.

This project allows users to query and search medical services (“shoppable” services are of particular interest) and returns a list of relevant treatments and hospitals with these reported services.

Datasets:

· “Cost of care” and “price transparency” found on some hospital websites. Hospitals post prices on websites. Many times, in an Excel file format. (e.g., https://www.rush.edu/patients-visitors/billing/cost-care, https://www.uchicagomedicine.org/patients-visitors/patient-information/billing/price-transparency)

· Kaggle Dataset hospital-price-transparency link

· CPT codes / Procedure code long and short descriptions for inverted search mapping/interchanging

Implementation and Usage

**Documentation of how the software is implemented with sufficient detail so that others can have a basic understanding of your code for future extension or any further improvement. AND. Documentation of the usage of the software including either documentation of usages of APIs or detailed instructions on how to install and run a software, whichever is applicable.**

Front-end starts in our main by prompting the user and obtaining the query. We pass this query to our search which implements meta’s inverted index to create a document list of matches. From there our driver establishes our data structures and provides a dictionary of treatments with pertinent hospitals providing that service. We utilize OO with Treatment, Patient, Hospital, and Price classes. To build indexes for the records and those relationships, we intersected multiple datasets. We established a threshold for our output to limit the results to 20 matched treatments as well as limiting the number of hospitals to no more than 5. We then output the list of ranked treatments and hospitals that best match the user’s query.

Our code has the additional functionality to input a city and limit the results of the search to hospitals within a certain city. We opted to comment this out due the limited dataset we were able to obtain.

* Requirements:

o Python 3.5

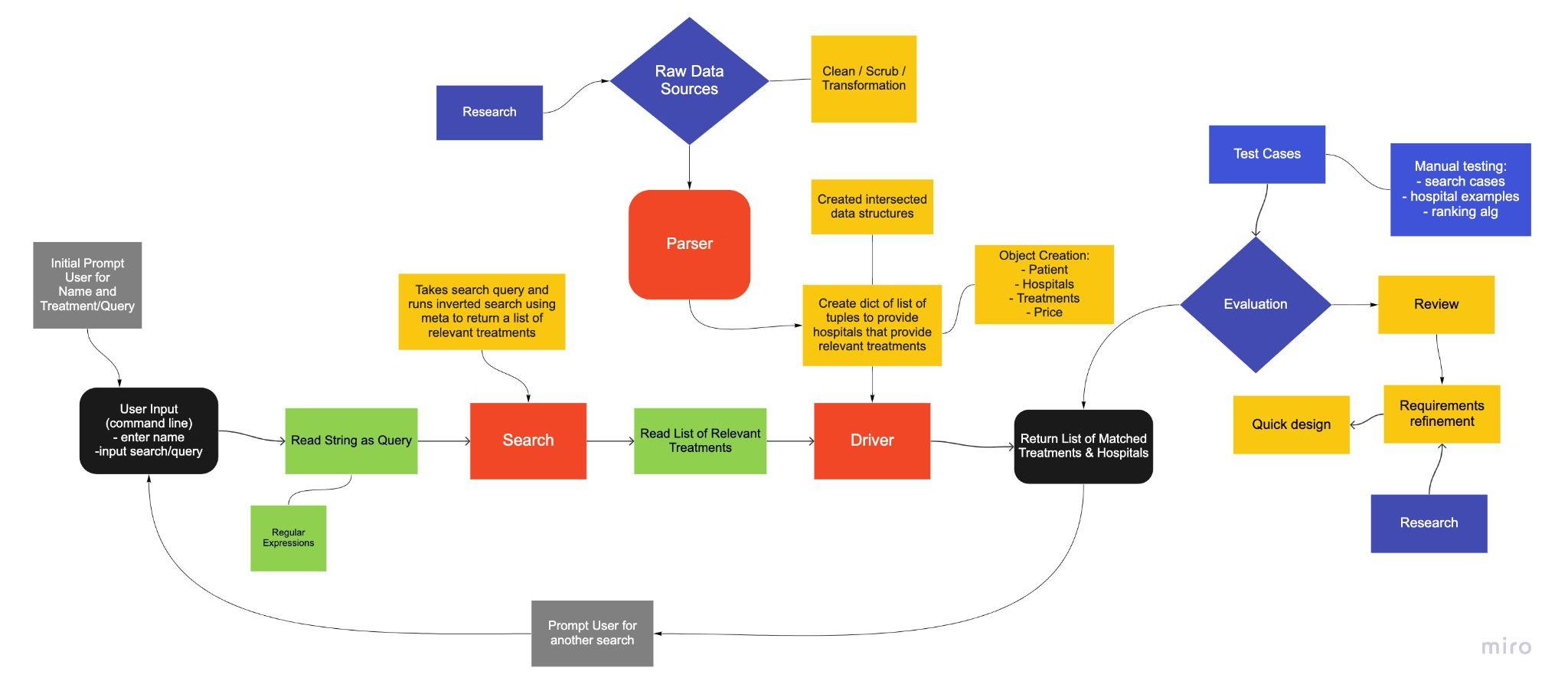
o MetaPy

* To run in command line:

|  |
| --- |
| Pip install -r requirements.txt  python main.py |

* Suggested treatments to search:
  + o Mri
  + o Ct scan
  + o Air
  + o Transport
  + o Ambulance
* Sample output

|  |
| --- |
| Hello! Welcome to the Search System For Hospital Transparency Data. Please enter your name: User What kind of treatment are you looking for today? mri  We found treatments. Now we're now checking the hospitals offering this service.  Best Matched Treatments:   Treatment Name: Magnetic resonance imaging with contrast breast; bilateral  Hospital Name: CarolinaEast Medical Center  NPI: 1801852835  Cash Price: 2288.15  Gross Price: 3813.59  Min Price: 343.67  Max Price: 3546.65   Treatment Name: Computer-aided detection including computer algorithm analysis of breast mri image data for lesion detection/characterization pharmacokinetic analysis with further physician review for interpretation (list separately in addition to code for primary...  Hospital Name: Atrium Health Cabarrus  NPI: 1487866315  Cash Price: 21.37  Gross Price: 0.0  Min Price: 0.0  Max Price: 0.0   Treatment Name: Magnetic resonance imaging without contrast followed by with contrast breast; unilateral  Hospital Name: Atrium Health Wake Forest Baptist Davie Medical Center  NPI: 1154326379  Cash Price: 1665.5  Gross Price: 3331.0  Min Price: 365.87  Max Price: 3164.44 ...  ... ... Do you want to find another treatment? (Yes/Y or No/N) |

See below for a Miro documenting the flow of our code.

Additional Improvements/Features:

· Scrapper functionality to extract useful data from hospital websites.

· Provide a recommendation functionality to match results based on similar content

Contributions

**Brief description of contribution of each team member in case of a multi-person team.**

## 

## Team Name: TAM

Our team approach was to build our general structure and then iteratively build functionality. We partitioned the initial portion of coding and then met weekly to test, discuss design, data structures, bugs, and divided the weekly work accordingly. Tasks not listed with a specific member were collaboratively worked on.

## Division of Tasks:

* Object Oriented design for entities and class structures (Aareana)
* Find datasets/API that provide a list of hospital price transparency (Tony, Morgan)
* Data cleaning, mapping, and transformation (Tony, Aareana, Morgan)
* Building query (Tony, Morgan, Aareana)
* Output, printing the search results (Morgan)
* Interactive command line interface for user input (Aareana, Morgan)
* Implement Data Parser (Tony)
* Testing and Debugging (Tony, Morgan, Aareana)
* Configure MetaPy class, index and corpus file (Tony, Morgan, Aareana)
* Implement search functionality (Tony, Morgan, Aareana)
* Search functionality to allow users to sort through and filter data based on a query (Tony, Morgan, Aareana)
* Intersection of datasets and data structure (Tony)