WeCare ...dedicated towards better public health...

Priyekant Aghi | Sagar Patni | Chandni Shrivastava | Kshitij Khandelwal

IRA A. FULTON SCHOOLS OF engineering
ARIZONA STATE UNIVERSITY

Overview

- Our intelligent and interactive system based on WebMD dataset aims to cater primarily to the users in the remote and rural areas where there is a scarcity of medical resources.
- Our tool can also be useful for researchers providing them with easily accessible information about the topics and other relevant sources.
- Our tool provides information about similar topics, the most relevant workplaces and departments making use of simple interactive visualizations that can be easily interpreted by any user.

Research Questions

- What are the most trending topics over a particular period of time?
- What are the topics similar to the one's selected by the user?
- Which workplaces are answering the most number of questions related to a topic?
- Which departments are answering the most number of questions related to a topic?
- Which already answered questions might be useful for the user ?

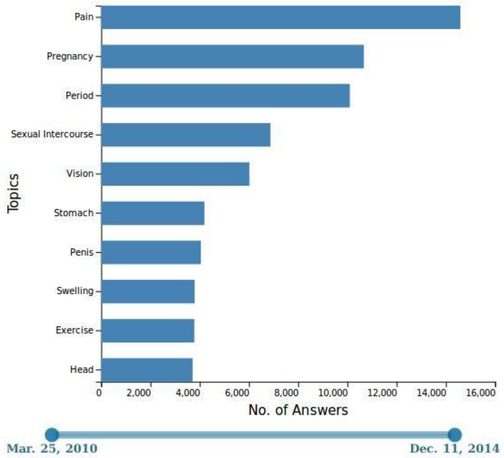
Data

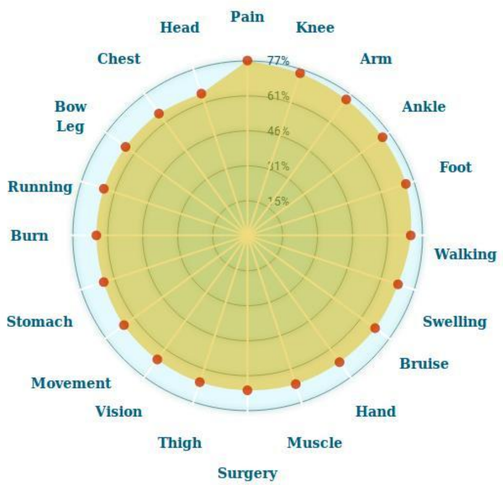
We used the WebMD dataset which consists of following data:

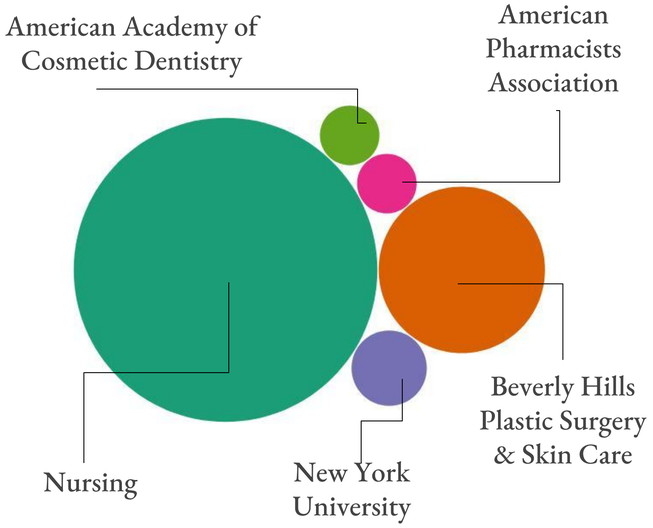
- ➢ Time period (Mar. 2010 - Dec. 2014)
- ➢ 1704 Topics
- ➢ 5402 Users
- ➢ 25319 Questions
- ➢ 74101 Answers
- ➢ 69 Workplaces
- ➢ 58 Departments

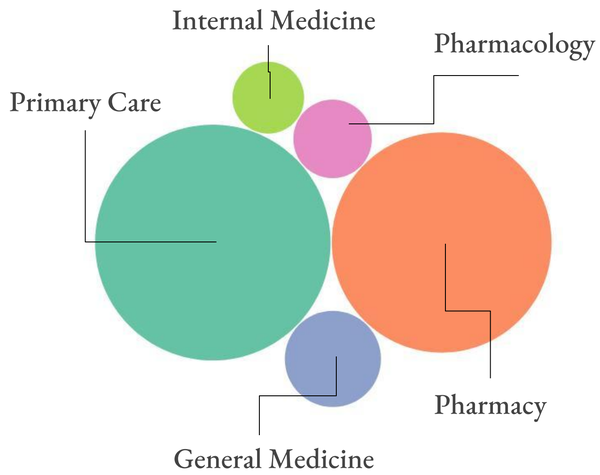
Search :

Topics Name	Count
Pain	2568
Pregnancy	2214
Period	2044
Vision	1070
Skin	866
Stomach	709
Swelling	684
Drinking	661









Methodology

- To process the data we combined all the data files into a single dataframe using 'questionId' as a key column.
- The WebMD dataset as a whole has a lot of missing values. We ignored missing values at different stages of data processing as most of the missing values were categorical, for which default values cannot be added.
- The data obtained after the pre processing is used to display the visualizations.
- The topics and their counts are listed that enables the user to search through any/all the topics and select the one that he/she is interested in.
- The bar chart component displays the top trending topics over the selected time period.
- The radar chart displays the topics that are similar to the selected topic. It shows the similarity measure between similar topics using cosine similarity. The similarity increases radially outwards and vice-versa.
- The bubble charts display the most relevant workplaces and departments related to the selected topic.

Future Work

- Our project can be made more useful to the user by listing the departments and workplaces in and around the geographical location of the user.
- For this, we aim to collect more data from specific regions and aggregate with our existing one.