Doccurate: A Curation-Based Approach for Clinical Text Visualization

Aniket Dhole - 1215147440
Kshitij Kashettiwar - 1215199947
Piyush Doke - 1215122233
Pranjal Milind Dhawale - 1215153849
Sai Vamsi Ogety - 1215337149

- **Paper Summary and System**

Clinical practice is a complex task that requires a grasp of both the medical issues afflicting a patient as well as contextual factors influencing their health such as family status, economic situation, and mental health and text is format used in all of the medical records for communication because **text** has the ability to preserve dependencies while concisely communicating the "health narrative". Doctors use methods like skimming the notes and asking questions to patients to understand their medical records of a patient. The main challenge comes in when the number of patient records increases and it becomes a complicated task to analyze all those documents in a very short period. The paper takes leverage of Data Visualization and Text summarization to address this challenge.

This paper proposes a more verifiable and customizable approach to leverage semi-automated text processing and unstructured medical knowledge in clinical text visualization. The approach utilizes **user-in-the-loop** iterative refinement supported by visualization and proposes the creation of **Filter collections** that leverage medical taxonomies like **SNOMED-CT** to summarize and aggregate related information in a pre-processed tagged text. The paper presents **Doccurate**, a clinical text visualization prototype to create a high-level view of patient records. **Filtered Collections**: They basically defines a set of related taxonomies pertaining to a shared meaning; by leveraging parent-child relationships in taxonomies, we can extend this meaning to a large collection of children concepts covered under the umbrella and rapidly group related concepts together. They take the leverage of text tagged by automated entity recognition tools like **Apache cTAKES** and their mappings to medical taxonomies such as **SNOMED-CT** or **UMLS**, a medical thesaurus that define these mappings. The Interface of the system presents Fc's associated with a medical record and upon selecting a single Fc, all its children from the taxonomy are shown and **stream graphs** are used to show the occurrences of its children along with their frequencies and timeline as one of the axes. They also provide the flexibility of adding own Fc's. The selection of Fc's and a single Child will give us a visualization of the label in the document along with a timeline.

- **Proposed System Extension**

  - **Tree Taxonomy** - The implemented system takes into account the taxonomy from SNOMED-CT concepts and uses it for Filtered Collections which is used for level-wise deeper inspection of patient data for analysis. Although the visualizations that are implemented give a good in-depth perspective of the diseases and symptoms related to the specific patient there is no visualization of the actual taxonomy itself. We propose an extension to this by creating a tree visualization of the taxonomy of the diseases and their contextual information. We think that the treemap along with the filtered collections will provide a holistic picture of not only the patient's health problems but also will be useful for doctors to analyze the diseases for patients and their relation with many other symptoms, contextual information and potentially other diseases as well.
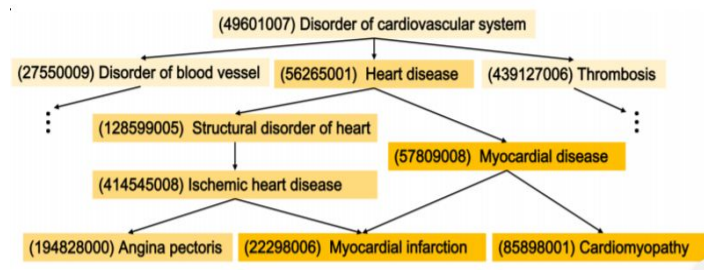
The Figure shows the tree taxonomy of diseases

○ **Frequency of Recorded Allergy vs Time Period:** We found that the dataset of patients containing the medical history contains allergies along with other data. We propose an extension wherein these allergies can be plotted as a bar graph taking into account the frequency of recording them over various time periods by extracting them from the dataset. We think that creating such visualization will be helpful not only for patients to keep track of their allergies but also for the doctors to prescribe medications to patients in the future.

○ **Severeness of diseases** - In addition to the above observations we found that the data also contains information about the hospital transfers, for example, if a patient was transferred from the general ward to ICU. We propose an extension for analyzing this data and extracting pattern for the degree of severeness that was recorded for that patient by possibly using a line chart. We think that by doing this the overall health of the patient can easily be visualized in a single graph and will be potentially helpful for doctors to not only prescribe medications with the specific strengths and doses of medicines but also will be helpful for faster response with respect to transfers and admissions to hospitals.

● **List and Link to Two Datasets**
We have requested the dataset from the following site and are waiting for approval. Two patient records are our two datasets.

-https://physionet.org/content/mimiciii/

● **Work Division between Teammates**

Name	Front-End	Mappings to medical taxonomies like **SNOMED-CT** or **UMLS**	Data-Visualization	Entity recognition Apache cTAKE
Aniket			✔	✔
Kshitij		✔	✔	
Piyush	✔		✔	✔
Pranjal	✔	✔	✔	
Sai	✔		✔	✔

● Expected Timeline
  ○ Mappings to medical taxonomies like **SNOMED-CT** or **UMLS -** 10/11/2019
  ○ Entity recognition Apache cTAKE - 10/11/2019
  ○ Data-Visualization - 11/08/2019
  ○ Front-End - 11/22/2019