A Visualization Tool for Navigation of Online Disease Literature

Jimeng Sun, PhD[1], David Gotz, PhD[1], Nan Cao, MCS[2]
[1]IBM Research; [2]HKUST

Abstract. Online health information portals provide valuable content to casual consumers. Used properly, this medical information can help patients better manage their health. However, the page-oriented nature of these resources makes it difficult for users to understand the overall information space and navigate the complex relationships between various diseases. We have developed a visualization tool that helps users navigate a large set of disease-related documents and understand multidimensional relationships for key semantic concepts such as symptoms and treatments.

Introduction. Patients and caregivers have access to many internet-based health information portals. Examples include WebMD, MayoClinic.com, and Google Health. These resources provide concise, easy-to-understand descriptions of diseases with information on symptoms, diagnosis, prevention, and treatment. Unfortunately, while studies have shown that non-experts turn to online information to explain undiagnosed symptoms [2], the organization of these document-based resources is best suited for situations where patients are searching for a specific disease after receiving a diagnosis.

In our work, we aim to assist patients during self-diagnosis. In these cases, patients must navigate a web of inter-related disease properties (e.g., symptoms, treatment, or causes) to identify which diseases may be relevant. Therefore, our visualization is designed to help provide decision support for patients [1] prior to diagnosis.

Methods. As a pre-process, we apply entity extraction and topic modeling techniques to a set of disease-centric documents taken from a web health information portal such as Google Health. The extracted entities (e.g., symptoms, treatments, etc.) and topic models are then indexed for use within the visualization tool. The visualization itself presents an overview by rendering diseases as nodes that are organized spatially into topic clusters. Overlaid on top of the topic map are a series of graphs connecting diseases. Each graph displays relationships within a single data dimension (e.g., symptoms) and is constructed based on the results of entity extraction. Users can control the visualization to navigate the information and apply filters to consider only certain relationships. Upon the identification of a disease of interest, users can easily navigate to the original web page for more detailed information.

Results and Discussion. Our visualization is shown in Figure 1. Each large yellow node represents a single disease. Disease nodes are positioned closer in space if they are related according to the topic model. The size of a disease node is proportional to its prominence within the topic. Secondary facets for each disease (such as cause, diagnosis, prevention, prognosis, symptom and treatment) are displayed as small colored nodes surrounding each disease node. The secondary nodes for various diseases are connected with edges if the underlying documents' text relating to a specific facet is similar. This allows users to understand how different diseases are connected. For example, cancer and tumor are connected with multiple edges because they have similar symptoms, treatments and prognosis. The color buttons on the top left allow users to control which facets are displayed. Beneath the disease nodes, a contour map gives an overview of the disease clusters, allowing users to understand global patterns. For example, Figure 1 illustrates two main disease clusters for cancer and pneumonia.
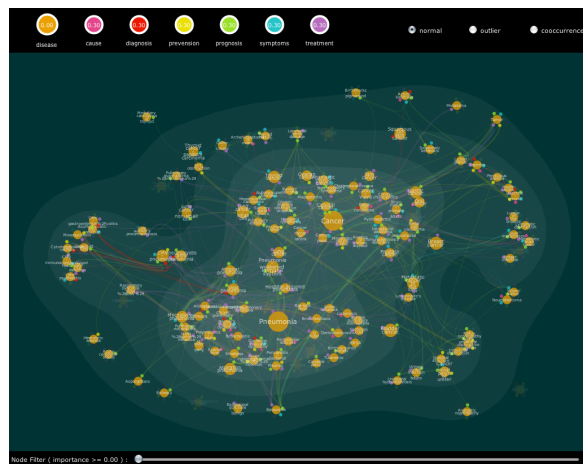

Figure 1: A screenshot of our visualization tool.

References

[1] H. B. Jimison, P. P. Sher, and J. J. B. Jimison. *Clinical Decision Support Systems: Theory and Practice*, chapter Decision Support for Patients, pages 249–261. Springer, 2nd edition, 2007.

[2] R. W. White and E. Horvitz. Experiences with web search on medical concerns and self diagnosis. In *AMIA Annual Symp Proc*, 2009.