Making Sense of Medical Search Result

Research Aims
Our research is motivated by the problems of searching health related information on the web. Search engines produce diverse sets of results when queried for health related topics, like disease, symptom, or a drug. Aggregating information collected from different websites and inferring summary of findings about a treatment for example, is a difficult task without automation. We are developing techniques of information extraction, opinion polarity detection and aggregating summaries of factual information from different health related websites.

Research Methodology
A probabilistic graphical model will be used to recognize entities, features and opinion polarity from textual data. In general, a disease has several symptoms and symptoms are treated by one or more drugs. Drugs might have improved or worsen the symptom or may result in side effect. Taking these relations into account we have developed the model shown in the Figure 1. The constructed model has nodes as D (Disease), T (Drug/Treatment), S (Symptom), and SE (Side Effect). We call nodes as entities in our experiment. These entities have several attributes associated with it. For example, drug has attributes like doses, duration and timing. Attributes are omitted in the figure for simplicity. Each symptom and side effect entity has opinion polarity associated with it.

Research Approach
To build a proof of concept system we are investigating social media text for summarizing opinions regarding Parkinson’s, its symptoms and treatments. A text corpus by scraping text from a Parkinson’s forum website is created. Using medical ontology and different lexicons we have annotated disease, symptom, drug and context in the text. Text patterns to be used in a semi-supervised setting are being created for automatically recognizing entities and context around the entities from text. Text preprocessing like stemming and spelling checks are yet to be done. Figure 2 shows our annotation result. Here, “metamap” is the mapping of text to medical ontology, “lookup” is the polarity and other lexicon mapping. The properties “orch;phsu” are semantic concepts in medical ontology and “negative” is the polarity of the word.

Figure 1. Data Structure Model

Figure 2. Sample Annotation of Text