Week 1 Reflection

•    Why do we need clinical terminologies? Why not use just English for documentation?

Natural language is, for the most part, an extremely unstructured data type from which it is relatively difficult (compared to something like a spreadsheet) to uniformly extract machine interpretable information. The reason it would be useful to have a machine interpretable version of clinical text is that it could be used for tasks like decision support, both at the individual level and institutional level, or for conducting observational data studies.

•    What is concept-orientedness?

Within the context of clinical terminologies, concept-orientedness is the idea of having terminologies that are symbolic representations of ideas, a way of representing conceptual notions that people have in their head. In the ideal case each concept is a unique and atomic concept; these concepts do not have to be human understandable as long as there is a way to map natural language to the terminology (though even this is not totally necessary); this would allow different words to map to different terms. Concepts can be composed together to identify relationships between them, or even to make new concepts

•    What are the desiderata for terminologies (in your own words)? Do you agree with these criteria? Is there anything missing?

The desiderata outlines the ideal clinical terminology. This terminology would be concept oriented, with each individual term associated with a single concept. The terms should be organized into a poly-hierarchy; the poly-hierarchy allows for terms to be part of multiple conceptual relationships, and will facilitate modelling of new relationships between terms. The organization of the terms in the poly-hierarchy should allow for different levels of association, at both high-level views capturing broader relationships and lower-level granular relationships. Once a term has been incorporated into the terminology, it should be there forever, even if the term is no longer used. The terminology should be complete, essentially to prevent creating missing and unusable data, while at the same time make it simple to add new terms into the terminology. Finally, the terminology should be able to capture the context in which the terms are used. This last idea is the one I think is the least important, simply because I think it is implicitly a part of the other parts of the desiderata. I think if a terminology has atomic concepts perfectly mapped to each other as part of a poly-hierarchy, then the context of group of terms should be able to be inferred (though this would require some set minimum for the amount terms included in a patient record). I think the graph structure between terms would be particularly amenable for this. You could have concepts be nodes, and have different “verbs” (‘symptom\_of’, ‘causes\_disease’) be edges, and then walk-through different paths on the subgraph of a subset of terms to find out different relationships and contexts. (My hot take of the day)

•    Using what you have learned so far, in your own words explain: what makes clinical terminology so hard? What are the challenges?

I think the central challenge is that medicine is conceptually an incredibly complicated field that rapidly changes over time. To me, it seems that clinical terminologies are essentially trying to distill the knowledge of clinicians into an abstract, machine interpretable form. I often hear that the human mind is magnitudes better at learning complicated relationships than computers; despite this, it still takes at minimum 10 years to become a doctor, and that’s not counting the education before college. The terminology that could accurately capture this vast amount of information would be massive. At the same time, it would need to be flexible enough to keep pace with how fast medicine changes. These two ideas, having a complete terminology that captures all concepts and a terminology that is flexible to the inclusion of new concepts, seems diametrically opposed.