

Kevin Stevens  
End-of-semester report  
CSE499 (Medical Natural Language Processing Research Group)

After this semester, I thought it might be useful to spend a few paragraphs helping those who might find themselves in the same situation I did. Namely, joining this research group, or any research group, dealing with knowledge representation and not having the slightest clue what the term "knowledge representation" means.

First, and I don't think that I can emphasize this enough, do not start by going to the SnePS tutorial and hacking your way through it. This is the approach I tried at first and nearly all the words bounced off my head like marbles off concrete. The writers of the tutorial are assuming a good deal of background knowledge that you do not have yet. Don't go near the tutorial until you have a pretty good idea what the following terms mean:

- assertion
- intentsional
- extentional
- base node
- structured node
- unstructred node
- arc labels
- case frame

I recommend that you should be able to look at a knowledge representation network and be able to roughly reconstruct what it says and to be able to take a simple sentence and construct a paper network from it before sitting down with the tutorial in hand to punch assertions into Cassie.

Thankfully, there is plenty of material at hand to help you. I found the links from Dr. Rapaport's site invaluable, especially those that link to the SNeRG page. If you are anything like me, you'll get about 25% of the way into the material before you overload. Stop there, get the next link, and start reading that 25%. Each time through, I was able to get a little more than 25%. But getting that 25% is all you need to be able to start constructing simple networks of your own and entering them into SNePS.

I found this method extremely helpful during the semester. Of significant help to me was going through a piece of text and attempting to come up with simple case frames of my own and then constructing networks with them. The practical experience of that process, and talking through the results with Dr. Rapaport, was of more benefit than all the journal articles I read.

Kevin Stevens  
End-of-semester report  
CSE499 (Medical Natural Language Processing Research Group)  
Summary of results.

My first job with this group was to attempt to derive some simple case frames from the discharge summaries. Secondly, I was to apply those case frames to a discharge summary and parse it from beginning to end. After presenting the work to the SNeRG group, here are what appear to be the major issues:

None of the case frames as I conceived them deal with the issue of multiple visits to the hospital. At some level, each of them need to have an arc pointing to the particular discharge summary they were generated from. Otherwise, there will be no way of reasoning about the relations between separate hospital admissions.

Time. The temporal issues have been left entirely out of these case frames. Mostly because others are dealing specifically on this issue and at least partially because it was beyond my knowledge at this time.

Space. The issues of where people are in the hospital must be dealt with. By this I mean is a patient's location defined by the ward that he is admitted to (ICU, oncology, etc.) or their actual physical location (Cath lab, X-Ray, Operating Room 1, etc.). I would suspect that some combination of both will be required to be represented. A patient can be admitted to ICU, and taken to X-Ray where something else could happen (falling and breaking something, for instance). In that case, both ward and physical location become relevant.

#### Patient History

A great deal more information needs to be gathered on this subject. Some questions:

Do you have a history of a condition if you are still suffering from it? (i.e. can you have a history of diabetes?)

Are all mentioned conditions part of a patient's history?

How do you represent the family medical history?

As to the case frames themselves, they are fairly straightforward. Nearly all of them deal with a property of the patient. For example, the patient has a name, medical record number, age, gender, etc. Each of these properties has a case frame to match. And as these case frames were developed only in regard to the Brendon H discharge summary, it is a certainty that there are numerous other case frames to be found in the material.

There is also a first pass representation of the Brendon H discharge summary inside the "Networks" directory. The strategy was to start at the beginning of the summary and represent everything in sequence, rather than trying to determine a structure for information such as the header and try to put the information into that structure.