A short tutorial to map biomedical free-text into UMLS concepts using MetaMap

UMLS is a comprehensive thesaurus and ontology of biomedical concepts, and has been widely used. While in text mining, we often want to map the free-text into a list of UMLS concepts, but the mapping process is not trivial (at least to those folks like me who is not an expert in text mining). Thanks to the MetaMap (http://metamap.nlm.nih.gov/), it is not that hard at all as long as you know how to run the tool. I spent sometime playing with this tool to extract diseases from the free text in different sources (e.g., OMIM, DrugBank) It worked pretty well! Here I would like to document the process in case somebody else would like to use this tool as well.

MetaMap includes two major parts (its server including the knowledge-base and various internal algorithms, and its APIs used to call the server to finish your job). You can install the server locally and use Java APIs (or the commands) to call the server. The Jave APIs look very neat but unfortunately, they did not work in my MAC. I suspected it had something to do with the permission, but I am not a MAC guy. Anyway, I figured out that using the commands could do the tricks as well, so I did not try to address the issue of using JAVA api, if someone found out the way, please update here!

Install and start MetaMap

1. Download MetaMap full version and extract into the directory called (public\_mm)
2. Install it locally. It will ask the directory you want install and the JAVA-HOME (/System/Library/Frameworks/JavaVM.framework/Versions/1.6.0)
   1. cd public\_mm
   2. ./bin/install.sh
3. Start the server
   1. ./bin/skrmedpostctl start
   2. ./bin/wsdserverctl start
   3. ./bin/metamap12

Call MetaMap and process the output.

1. python call\_metamap.py input\_file, output\_file, tempt\_file, meta\_map path

e.g., python call\_metamap.py ttd\_dz.txt ttd\_dz\_out.txt test /Users/User/Downloads/public\_mm/bin/metamap12