

Natural Language Processing for Clinical Progress Notes

The LibreHealth Project seeks to strengthen our projects ability to serve the clinical and education user base by added interfaces to cTakes, a well known, open source, Natural Language Processing service for clinical data. This project will be further improved by using ActiveMQ asynchronous communication (also open source) as one of the protocols for sending and receiving secure progress notes and the cTakes resulting coded data.

NLP support will allow clinical users to improve the ability to report and track quality, outcomes and issues, while (ideally) letting the clinician enter progress notes in a more textual and comfortable format. The educational users can use de-identified progress notes data from real providers to do research and classroom simulations in conjunction with the NHANES based clinical data we are currently using.

cTakes interface NLM area of focus, effort estimates

- Collect and De-Identify a large number of Progress notes from several provider's existing EHR. These will be used both for testing and as a basis for the educational use cases. 30 man hours
- Develop export interfaces from LibreHealth EHR and LH Toolkit to send freeform/ notes to a cTakes instance, which then uses NLMS to map terms and and temporal relationships. 40 man hrs
- Map cTakes results to discrete data elements in the EHR and Toolkit data structures. 40 man hrs
- Create services for managing the import/export data using best practices for scalability with ActiveMQ and cTakes projects. 60 man hrs
- QA, Testing and Documentation. 40 man hrs

Success will be demonstrated when the project can show that freeform/text based progress notes from the EHR and toolkit in real time and receive, map and import the discrete data into their databases such that they information is available in "normal" views from the user's perspective.

Side benefits to this project include core infrastructure to use ActiveMQ for other information communication needs, such as peer-to-peer and doctor-patient data sharing.

All code will be open source, made available under the appropriate licences for the component projects, ie: Mozilla Public License, Gnu Public License v2+ and Apache License v2.