



Qing Zeng, PhD



VA
Informatics and
Computing
Infrastructure



Poll Question

- There are over 1 billion clinical notes in the VINCI repository: how would you like to use them? Which types of notes?



Poll Question

- What Text Processing tools have you used (e.g. Metamap, MEDLEE, ARC)?



Overview

- V3NLP is a system that intends to facilitated end-user use of text process functions
- It provides a framework that supports heterogeneous NLP modules
- It is designed to meet the big text processing needs



Components

- User Interface
- Backend Framework
- Repository of processing modules



Sample Text

Diagnoses

Medications

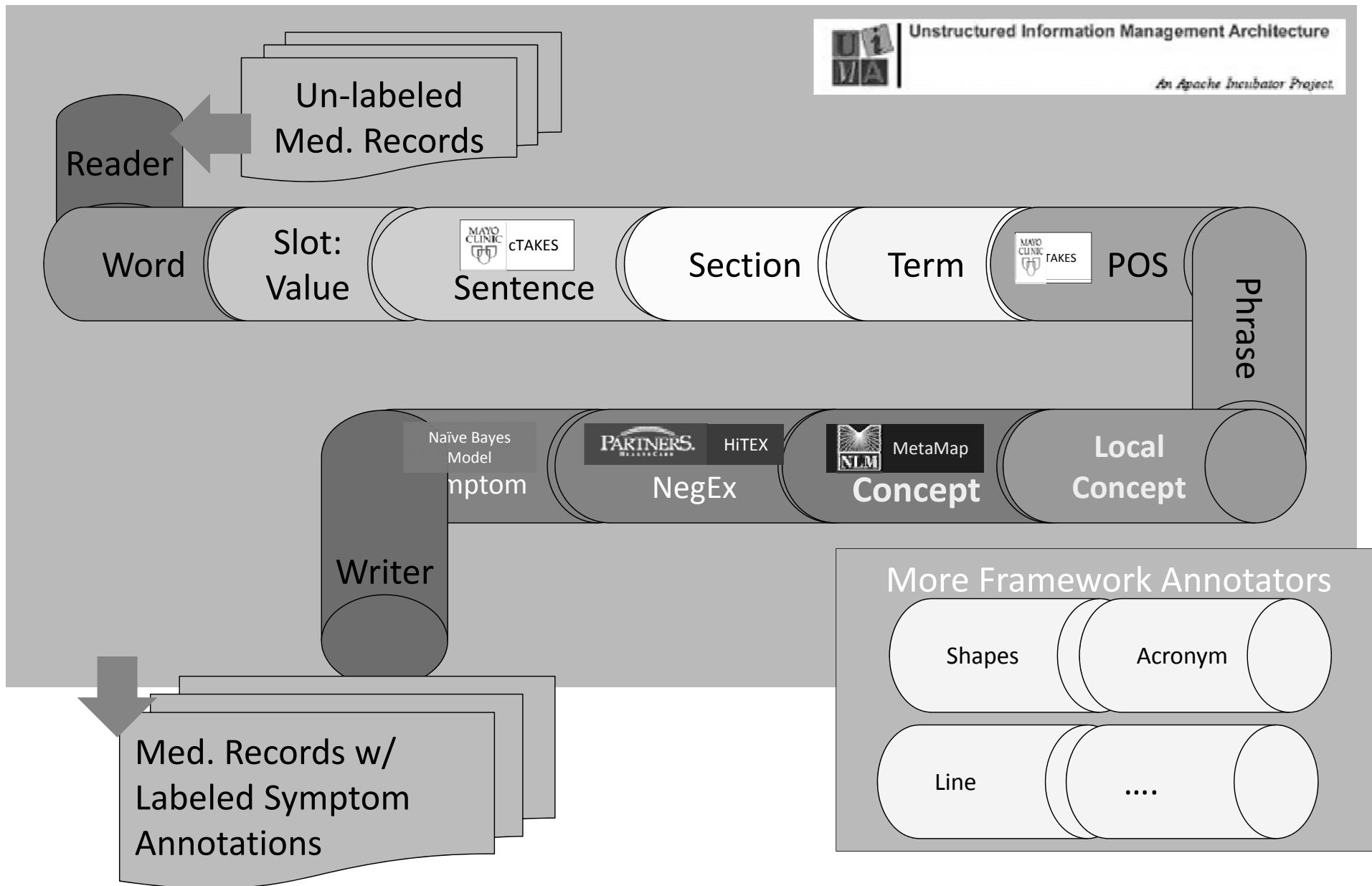
Lab
Orders

Rule out polymyalgia rheumatica versus early inflammatory arthritis versus diffuse osteoarthritis. She will continue on Tylenol as needed. I have urged her to start Aleve, 1 tablet twice a day with usual precautions. I have sent today for sedimentation rate, CRP, rheumatoid factor, ANA, and CBC. She is quite reluctant to have a blood drawn, but I have told her it is very important to help guide her treatment. I will be in touch with her when results are available, and we briefly discussed the possibility of starting steroids. I will arrange office follow up in 1 month.

Treatment Options

Follow Up Plans

Sample Processing Pipeline

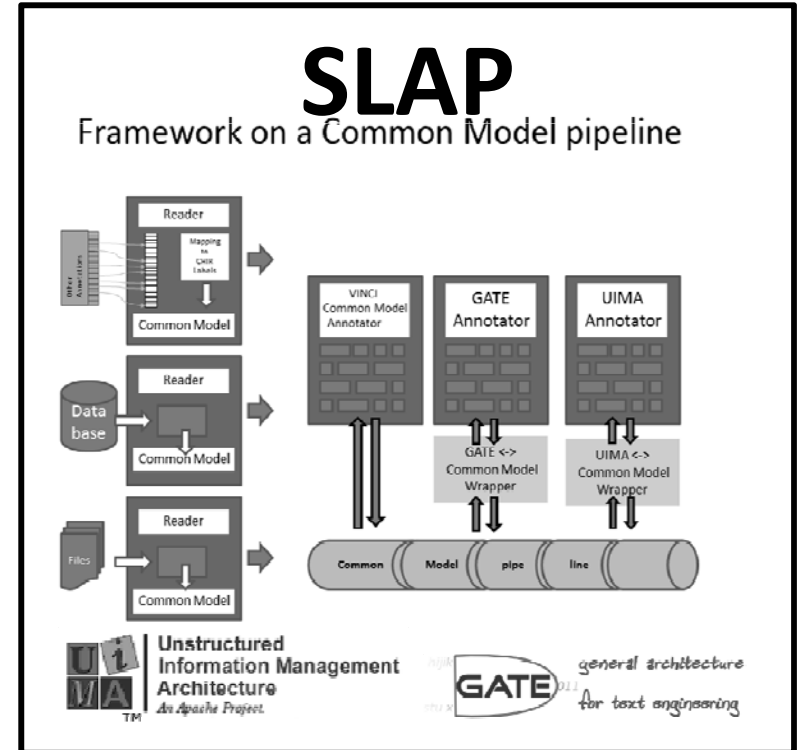
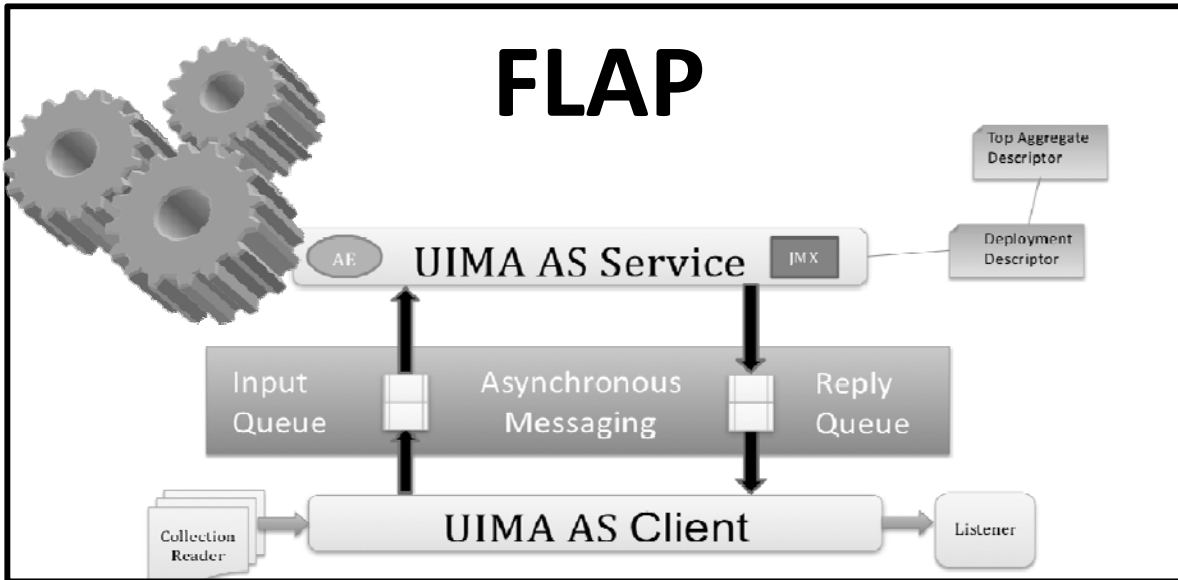




Interface

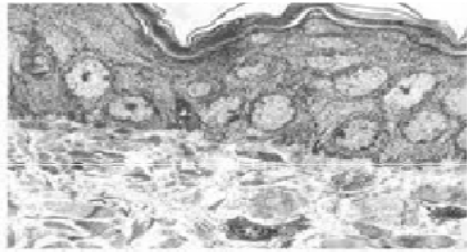
- Using existing processing pipelines
- Customize existing processing pipelines
- Create new existing processing pipelines
- Review NLP results

Backend Framework

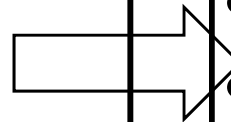


Repositories

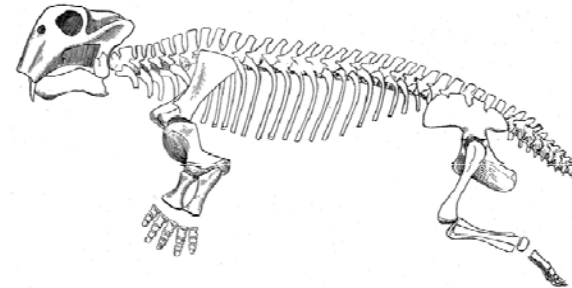
Derma



- Git Repositories
- Maven Projects
- Collaboration Quality
- Working but not tested



Bones



- Git Repositories
- Maven Projects
- Production Quality
- Tested
- Documentation
- Solar, Huston hooked in

VINCI/CHIR NLP Team



- Diana Carter
- Ryan Cornia
- Guy Divita
- Scott DuVall
- Tom Ginter
- Shirl Nielson
- Doug Redd
- Ben Sunder



V3NLP Poll Question

- How likely are you to use a tool like V3NLP for processing clinical notes
 - Extremely likely
 - Probably
 - Maybe
 - Not very likely
- What additional features would you like to see?

4/30/2012	12:00pm	New Natural Language Processing tools available on VINCI	VA Informatics and Computing Infrastructure	Cornia, Ryan and Qing Treitler Zeng
-----------	---------	--	---	-------------------------------------

Please find below the written responses to the unanswered questions from the VINCI Cyberseminar on 04/30/12; follow up questions should be directed to the VINCI staff.

Question: Is there a way to view all of the available modules that perform the same task (e.g., all of the pos taggers that are available) so that a more sophisticated user can select a particular module?

-Yes, you can view and select the available modules (e.g. we have two POS modules).

Question: Great work, I look forward to learning more. For cohort identification does the MRN need to be within the document for one to retrace to the patient, or will there be a field in the production version?

-The notes do have MRN associated with them, though V3NLP does not support retrieval directly. We have developed a query tool for structured and free text called Voogo, which we hope to release soon.

Question: How viable is NLP for large-scale data extraction for national research. For example, is it possible to extract every EF listed from every echo for research use?

-It is feasible to use NLP to process all VINCI data, especially for less demanding tasks like extracting EF values. We have tested the speed of the several pipelines and estimated the most computationally intensive tasks would take several weeks for the 1.8 billion notes. However, we are currently experiencing a network connection problem.

Question: When will the pathology report module become available?

-We hope to bring those modules to VINCI in the next 6 to 12 months.

Comment from VINCI staff: This is from VINCI. V3NLP will be working on radiology text. The report text in radiology is similar to that in TIU notes.

Question: Is the output of VNLN stored in flat files or database?

-Currently the result is stored in flat files through the interface. In batch processing, both options have been implemented.

Question: Are there processes to word crunch files to determine what terms lie in unfamiliar files? Ways to figure out what is in the file on which to base your search? For instance, if you want to search "heart disease" you need to see how to code the concept. Can we see the terms existing in the TIU to figure out what search terms to use?

4/30/2012	12:00pm	New Natural Language Processing tools available on VINCI	VA Informatics and Computing Infrastructure	Cornia, Ryan and Qing Treitler Zeng
-----------	---------	--	---	-------------------------------------

- We have developed a query tool called Voogo that would support such exploration, which we hope to release soon.

Comment: We are extremely likely to use something like V3NLP, but would probably need more orientation 1-to-1.

Question: Does this system work with dental records?

- The tool is expected to work on all note types, but is likely to require some customization and refinement to achieve high extraction accuracy.

Question: Can V3NLP be used for a completely different type of data, obviously with different terms and concepts? I am looking for a tool to use to analyze free-text responses to the all-employee survey. The high number of responses speaks to having software to assist in identifying concepts in the comments.

- The tool is expected to work on all note types, but is likely to require some customization and refinement to achieve high extraction accuracy.

Question: Neophyte though interested in this area; I need some NLP background to be able to see how i would use this. Plus, there are restrictions to the VINCI data; would like to use it on other VA data, either local or CDW, would that be possible somehow?

- We are making the V3NLP open source and we are happy to distribute it within the VA.

Question: Can one use v3NLP's extraction to obtain ejection fractions > 50%, that is to say, use formulas, rather than specific values.

- Yes, this can be easily done using regular expressions.

Question: Are notes classified by note type: e.g. can you differentiate between Pathology notes vs. admission notes?

- Each note has a type. While there are pipelines specializing on different note types, the processing modules we have are not type-specific.

Question: How do I bring notes the problem related notes??

-If the question is about uploading the notes, VINCI staff could help. If the quest is about searching for notes related to a particular problem, we hope to release a query tool soon.