Transformative Power of Natural Language Processing of Health Data: Tales from VA

Merry Ward, PhD
Scientific Program Manager
VA Health Services Research & Development

Matthew Samore, MD
VA Salt Lake City
IDEAS Center
Consortium for Healthcare Informatics Research

Outline

- Prologue
- Chapter one
  - The set-up: what makes health data challenging
- Chapter two
  - Turning the corner: grappling with the dimensions of the problem
- Chapter three
  - Examples: approaches and solutions

Intramural Research Program in Nation's Largest Healthcare System

CHIR

Our Healthcare System

- Veterans Health Administration’s health care system has 8.34M enrollees served by 807 clinics and 152 hospitals.

- VHA has a fully integrated HIT system, VistA/ CPRS

VISTA Data through March 2012

	Total	Average New Entries Primetime
Orders	3,192,843,576	1,169,632
Images	2,738,564,838	2,376,431
TIU Documents	2,006,236,305	961,440
Medication Administration	1,720,520,235	608,026
Vital Sign Measurements	2,240,762,101	898,219

HSRD Health Care Informatics History

- 2007
  - Launched HSRD's Health Care Informatics research with publication of RFA for Natural Language Processing research.
- 2008
  - Funded Consortium for Healthcare Informatics Research (CHIR), a multi-site collaborative research program.
  - Funded VINCI, a high performance analytic environment with secure access to corporate data warehouse and other VA data sources.
  - Recruited informatics research expertise in investigator-initiated scientific merit review.
  - Established Healthcare Informatics Research portfolio.

Chapter 1:
What Makes Health Data Challenging

Conventional Methods Fail to Scale

A Hypothetical Comparative Effectiveness Study

- In management of hospitalized patients with community-acquired pneumonia:
  - Should empirical therapy against methicillin-resistant *Staphylococcus aureus* (MRSA) and *Pseudomonas aeruginosa* be included?
- The minimum that is necessary to do this research:
  - Assess whether community-acquired pneumonia is "healthcare-associated"
  - Extract treatments & outcomes
  - Adjust for confounding by indication

An opportunity presented by “big data”

- Pneumonia is a common diagnosis in hospitalized patients
  - ~ 15-20,000 per year in VA
  - Manual chart review is not up to the task!
- If national VA data could be fully utilized to support observational comparative effectiveness:
  - Evidence base for treatment guidelines could be significantly strengthened!

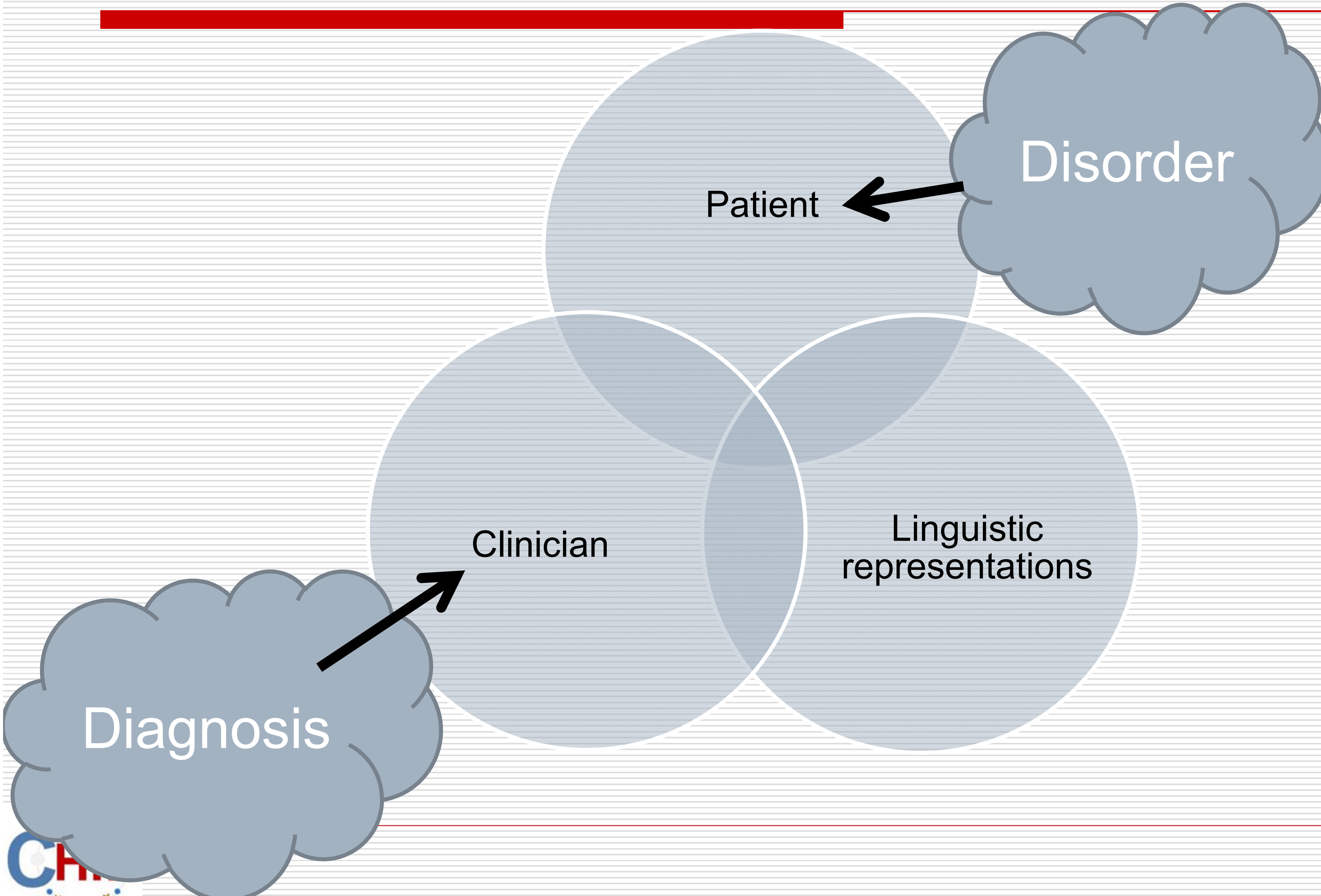
However, for even the simplest task

- Identification of who has pneumonia
  - The conventional approach is flawed

ICD-9 483.8, 485-6, 487.0 ≠ **Pneumonia**

What is meant by “pneumonia”?

Going beyond ICD-9 codes



Limitations of ICD-9 Codes

- The coded diagnosis is not necessarily the diagnosis asserted in the record
- The asserted diagnosis is not necessarily the disorder (disease)
- May not be reliably applied across time and space
- Do not reflect information available at the time management decisions are made

By themselves ICD-9 codes do not:

- Allow examination of differential diagnosis, stage, severity, phenotype, social context, disease course, treatment response
- What also needs to be accessible for analysis:
  - Diagnostic test results
    - Including text reports
  - Clinician notes
    - Including consultations, discharge summaries, progress notes, history & physical, nursing notes

Chapter 2: Turning the Corner

- The question is not whether natural language processing (NLP) alone is perfectly accurate
  - The question is whether NLP plus structured data (e.g., ICD-9 CM codes) is better than structured data codes alone

Taking Forward Steps: Define Target

- Pneumonia as a disorder
  - Pneumonia as a disorder:
    - E.g., inflammation and infection in the lung

  Or
  - Pneumonia as a diagnosis made in the context of clinical care:
    - E.g., an assertion made by an expert clinician who performs a comprehensive evaluation of the patient

  Or
  - Pneumonia as a diagnosis to decide on cases to be included in an epidemiological analysis
    - E.g., an explicit criterion that relies on clinical findings and diagnostic tests

Decide Role for Text Data

"Tag" text elements as codes

- Concept indexing or concept extraction
- Often, incorporated into explicit rules
- **Example:**
  - **Detection of post-operative complications***

Input text elements into machine learning algorithms

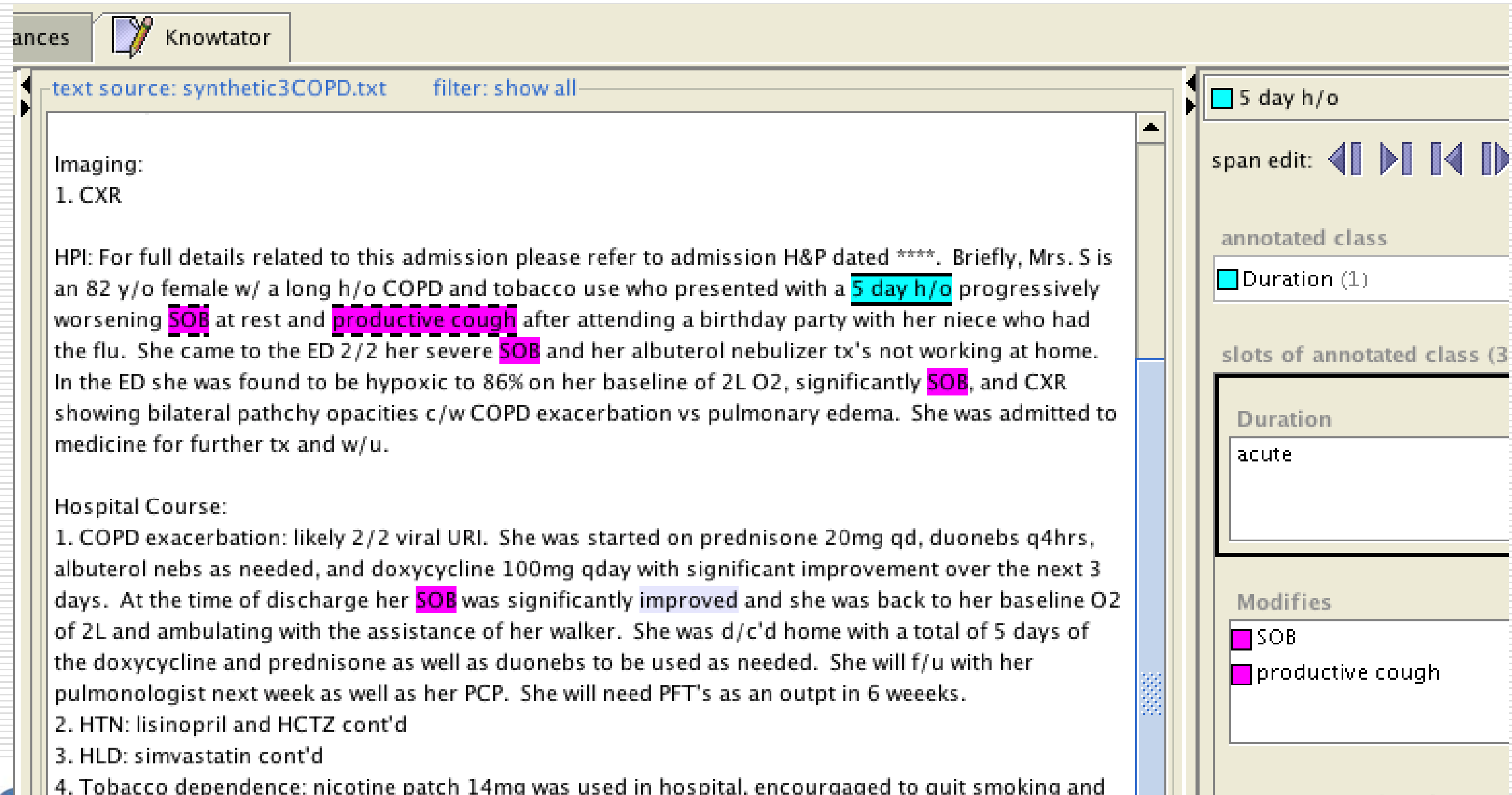
- Document- or patient-level classification
- Often, a black box approach
- **Example:**
  - **Classification of CXRs as non-normal**

*JAMA, 2011. **306**(8): p. 848-55.

Create Human Annotations

- Annotations are labels which assign meaning to data.
- Start with exploration
  - Use key word searches to explore text notes
  - Identify relevant document types
- Follow iterative approach
  - Rigorously test inter-rater reliability
  - To the extent possible, simplify cognitive task for human annotator

Search & Annotation Tools in VINCI



Train and Apply NLP System
"See one, do one, teach one"

DISCHARGE SUMMARY REPORT

Principal Diagnoses

PRINCIPAL DIAGNOSIS: (1) **DIABETES** (2) **ASTHMA.**

Additional diagnosis

HISTORY: The patient is a 47-year-old woman with a history of **diabetes, hypertension and asthma**. She was having increasing asthma on and off during this time…..The patient didn't experience any **shortness of breath** or **chest tightness**…. Her mother has a history of **asthma and diabetes**. He[r father] died of a **myocardial infarction**.

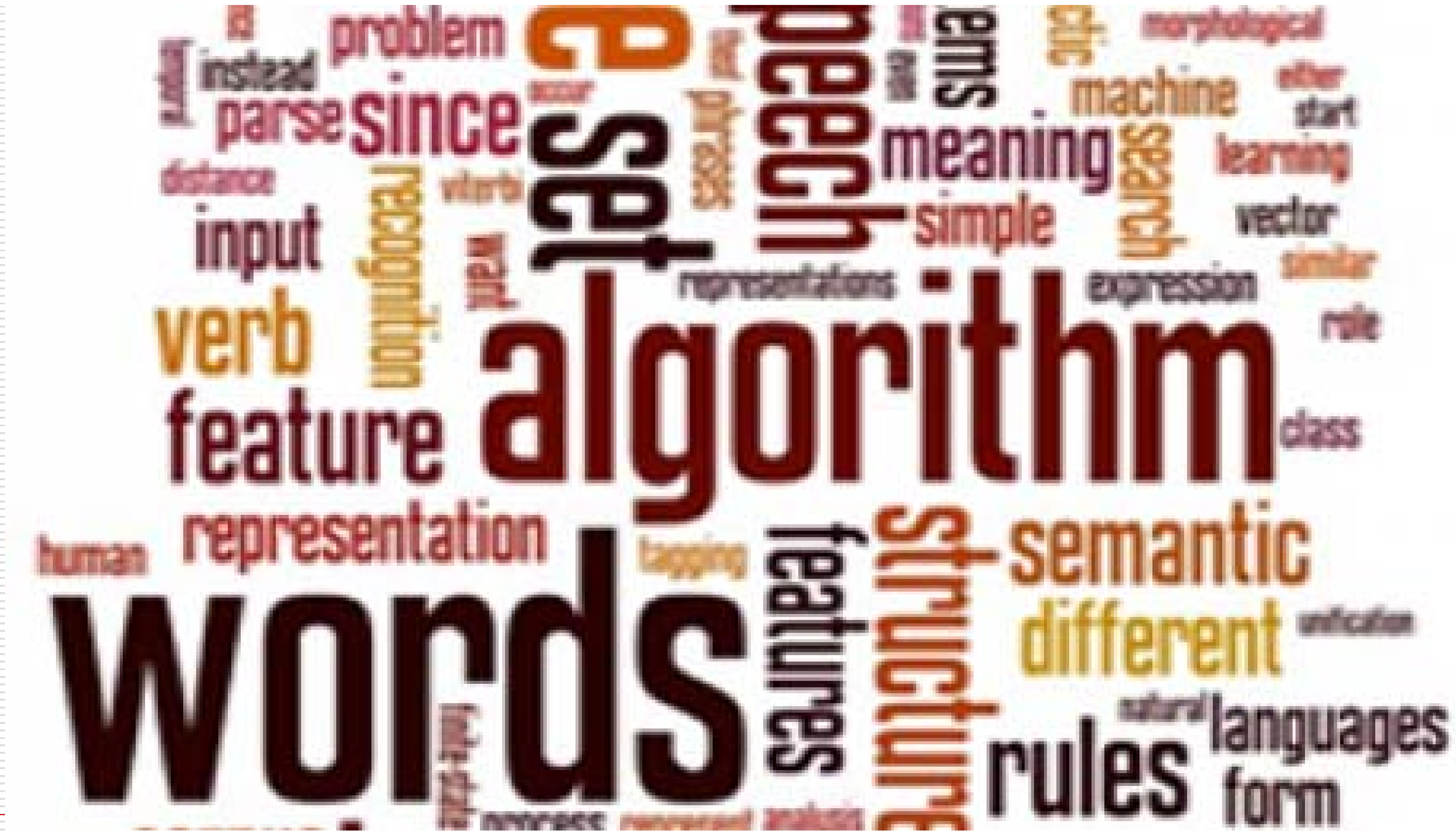
Asthma Medication

MEDICATIONS: **Albuterol** p.r.n.; Vanceril 2 puffs b.i.d.; Zestril 10 mg p.o. q.d.; insulin 70/30 33 U q.a.m., 15 U q.p.m.

SOCIAL HISTORY: She has two children. **She has smoked one pack every three days for the last 35 years, but quit 2 months ago.** She does not drink alcohol.

Past Smoker

Chapter 3: Examples of use of NLP

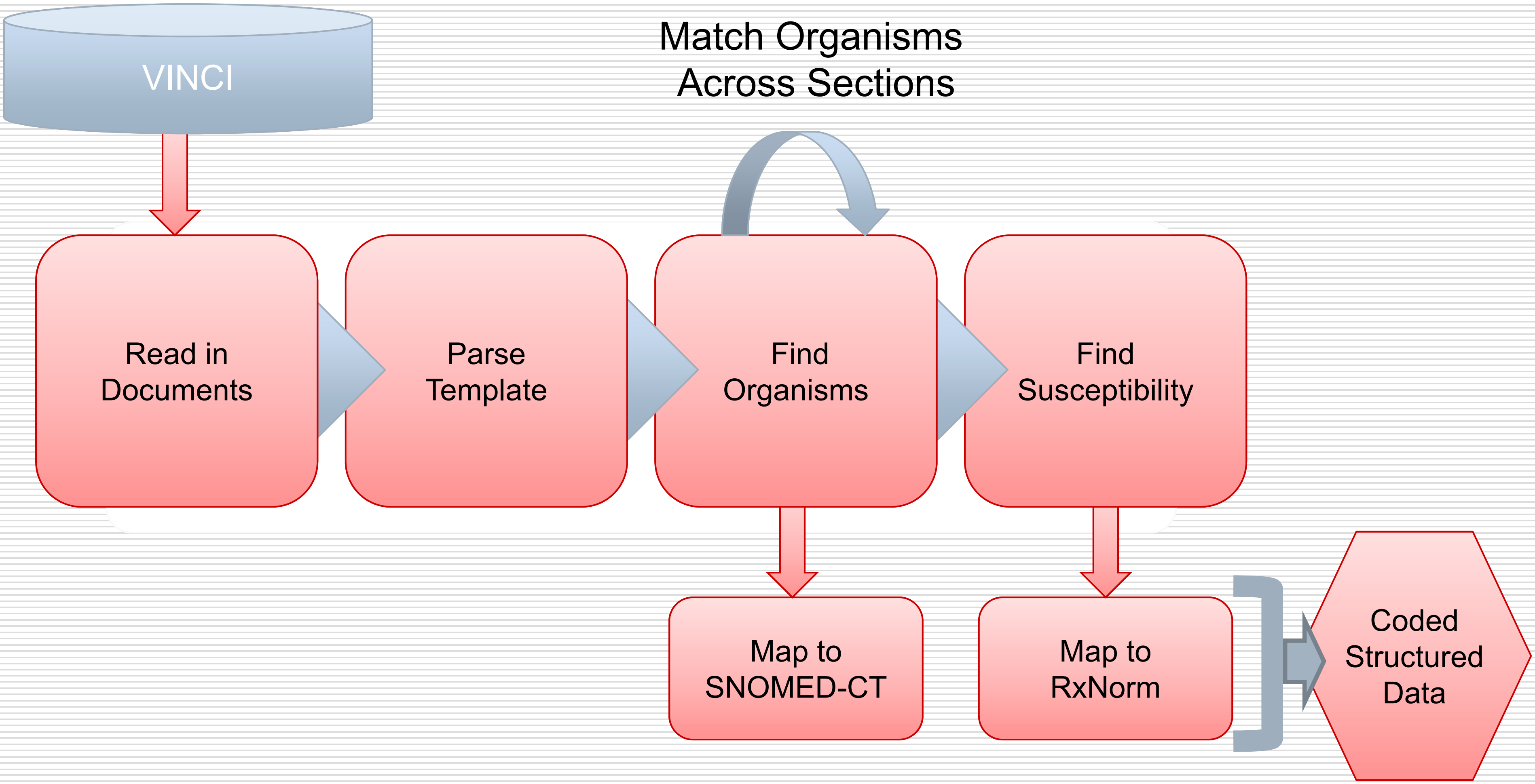


Application of NLP to Microbiology Data

- Statement of the problem:
  - Raw VA microbiology laboratory text reports are not usable
- Enormous station-to-station variability
  - Lack of national standardization in practices, tests, fields, terms, methods

---- MICROBIOLOGY ----|Accession: MI1234 Received: 1/1/2012 10:04|Collection sample: FLUID Collection date: 1/1/2012 09:15|Site/Specimen: SKIN|Provider: Dr. Jekyll|Test(s) ordered: CULTURE, FLUID/TISSUE completed: |* BACTERIOLOGY FINAL REPORT => TECH CODE: Mr. Hyde|CULTURE RESULTS: 1. HEAVY METHICILLIN RESISTANT STAPHYLOCOCCUS AUREUS|2. E-COLI|ANTIBIOTIC SUSCEPTIBILITY TEST RESULTS:|METHICILLIN RESISTANT STAPHYLOCOCCUS AUREUS|:|SUSC INTP|AMPICILLIN >8 R IV 1.0-2.0 gm Q4h PO 250-500mg Q6h|AMPICILLIN/SUL<8/4 R IV 1/0.5-2/1 gmQ6h|CEFEPIME >16 R IV 0.5-5.0 gm Q12Hr|CEFOTAXIME <8 R IV 1.0-2.0 gm Q8-12Hr|CEFTRIAXONE 32 R IM/IV 1.0-2.0 gm Q24H|CEPHALOTHIN <8 R IV .5-2 gm Q4-6H/PO 250-500mg Q6H|CIPROFLOXACIN <1 S IV 200-400 mg Q12H/PO 250-750 mgQ12H|CLINDAMYCIN <0.5 S IV 600-900 mg Q8H/PO 150-300 mg Q6H|ERYTHROMYCIN >4 R PO 250-500 mg Q6H|GENTAMICIN <4 S IM/IV:1.7 mg/Kg Q8H|IMIPENEM <4 R IV:250-500mg Q6H-8H|LEVOFLOXACIN <2 S PO/IV:250-500 mg Q24Hr|LINEZOLID <2 S IV 600 mg Q12H/PO 400 or 600 mg Q12H|OXACILLIN >2 R IV 0.5-2.0 gm Q4H|PENICILLIN >8 R IV:2-24 MU/Day (divided Q4-6H)|RIFAMPIN <1 S DO NOT use alone for Chemotherapy!|TETRACYCLINE <4 S PO 250-500 mg Q6Hr|TRIMETH./SULFA<2/38 S IV 3.3-6.6 mg/Kg Q8Hr\PO 1-2 Tabs Q12H|VANCOMYCIN <2 S IV:1.0 gm Q12H-24H|Bacteriology Remark(s):|Preliminary Report:|HEAVY STAPH SPECIES|ID & sens. to follow.|Final Report:|1. HEAVY METHICILLIN RESISTANT STAPHYLOCOCCUS AUREUS|2. ESCHERICHIA COLI

Approach: Extraction & Coding*



*BMC Med Inform Decis Mak. 2012 Jul 11;12:34.

Microbiology Has Become a Valuable Data Resource in VA

- Operational partnerships
  - National Infectious Disease Service
  - Antimicrobial Stewardship Task Force
  - Office of Public Health
- Epidemiology and health services research
  - Analyze of MRSA and other types of resistant organisms
  - Analyze of variation in antibiotic prescribing

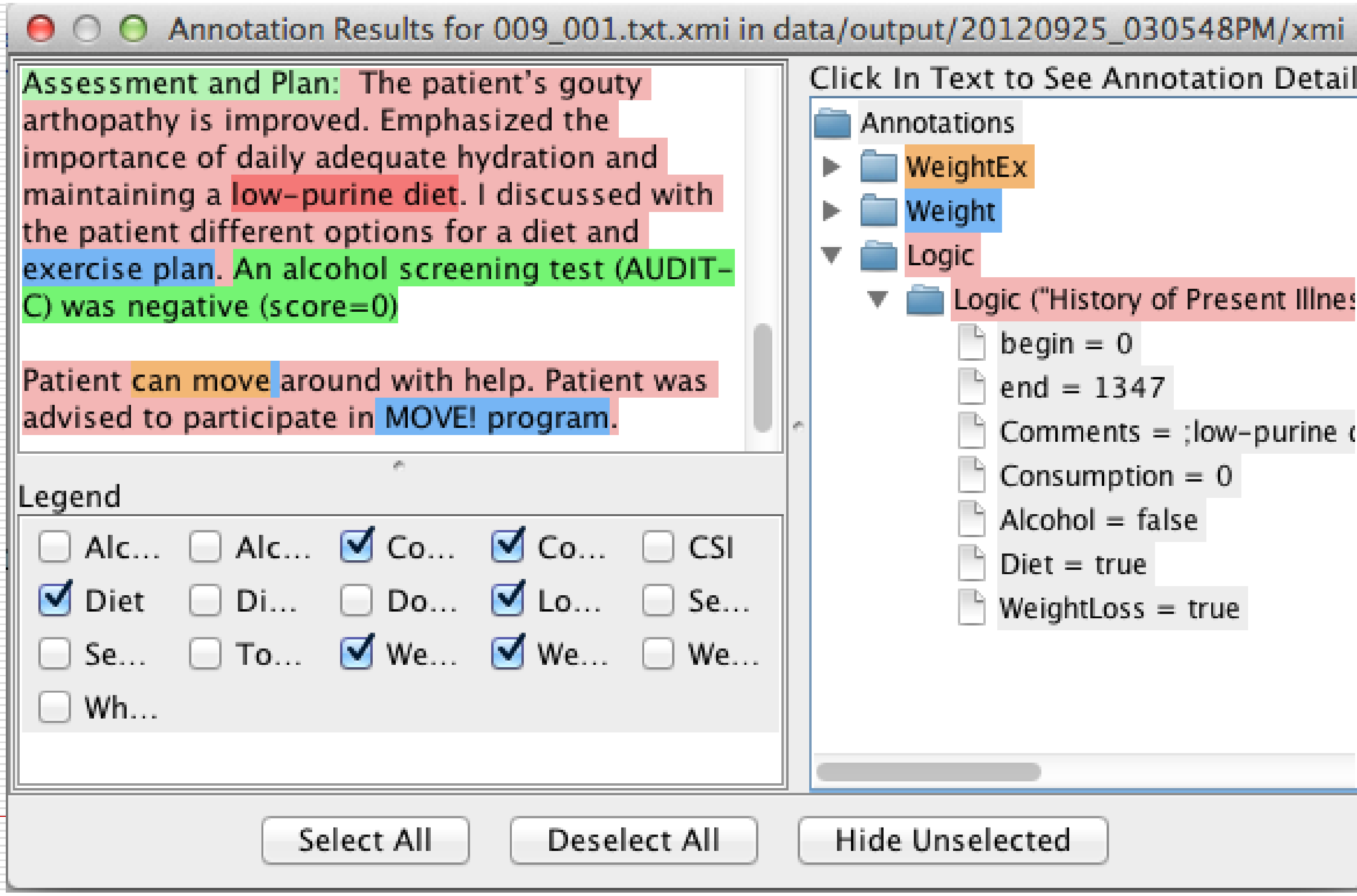
Applying NLP to other diagnostic reports

- Mentions of devices in CXR reports
  - Performance:
    - Sensitivity (Recall) 95%
    - Positive Predictive Value (Precision) 98%
    - Citation: AMIA Annu Symp Proc, 2010. **2010**: p. 692-6
- Ejection fraction in echocardiogram reports:
  - Performance:
    - Sensitivity (Recall) 96%
    - Positive Predictive Value (Precision) 95%
    - Citation: J Am Med Inform Assoc, 2012. **19**(5): p. 859-66
- Positive lymph nodes in pathology reports
  - Performance:
    - Sensitivity (Recall) 88.9%
    - Positive Predictive Value (Precision) 94%
    - Citation: J Am Med Inform Assoc. 2010 Jul-Aug;17(4):375-82

CHIR

Guideline adherence

- Dietary and weight loss counseling
  - NLP correctly classified 98.5% of records



Creating a symptom repository

- Initial rationale:
  - Characterize medically unexplained symptoms in deployed veterans
- Broader goal
  - Establish resource to support other types of research studies

Chapters Summarized

- Scale beyond manual chart review
- Probe deeper than ICD-9 codes
- Be explicit and clear about the semantics
- Iteratively link human annotation to computer annotation
- Create re-usable, shareable resources