

Welcome to the Sentinel Innovation and Methods Seminar Series

The webinar will begin momentarily

Please visit www.sentinelinitiative.org for recordings of past sessions and details on upcoming webinars.

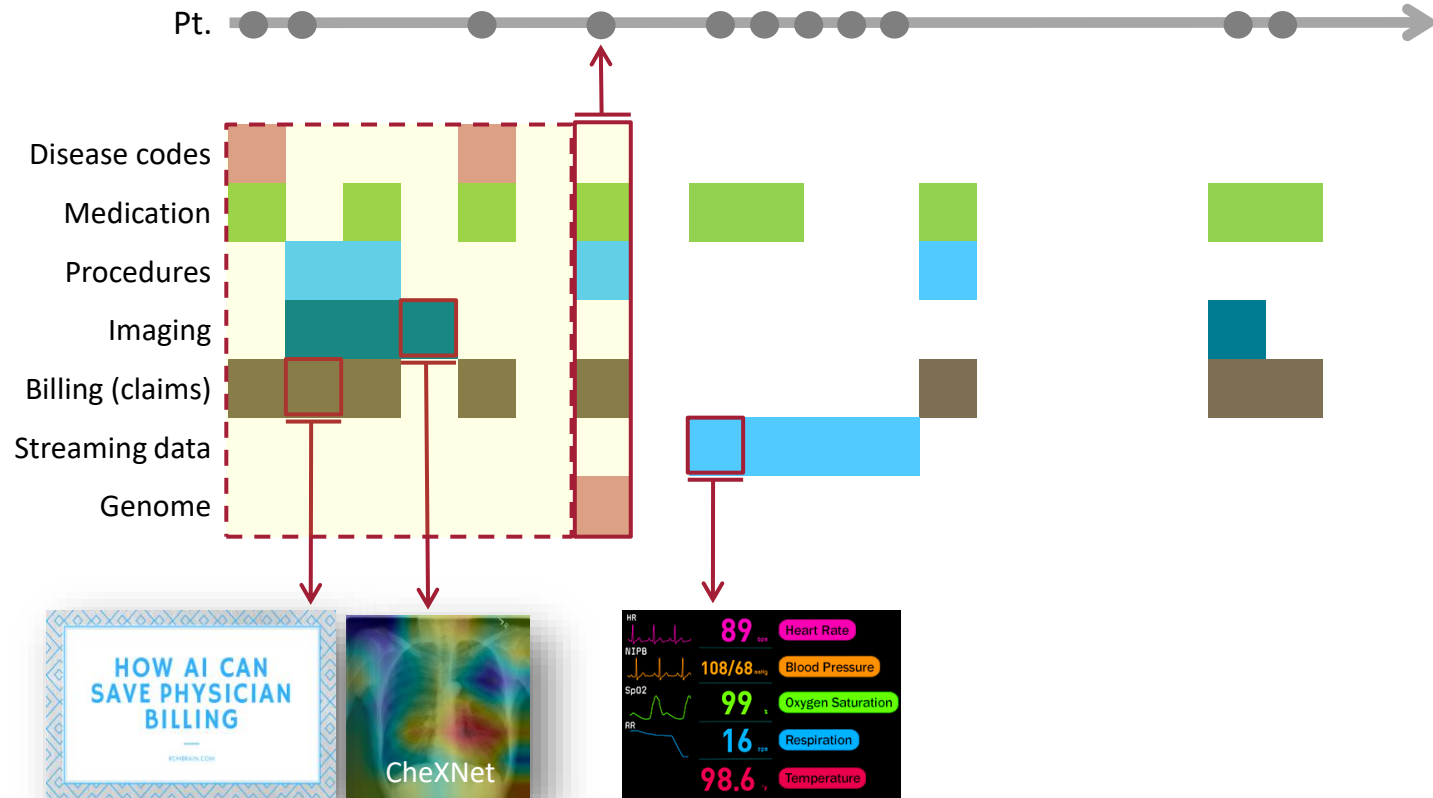
Note: closed-captioning for today's webinar will be available on the recording posted at the link above.




Ontology-driven weak supervision for clinical entity classification in electronic health records

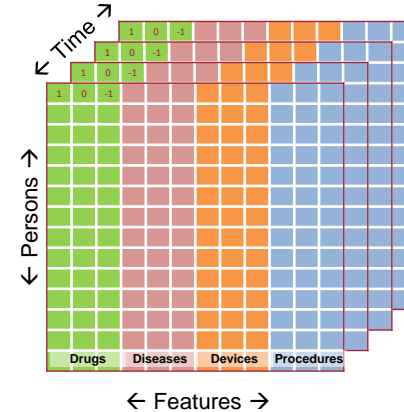
Nigam Shah

A patient timeline view of the data



Converting patient timelines to datasets

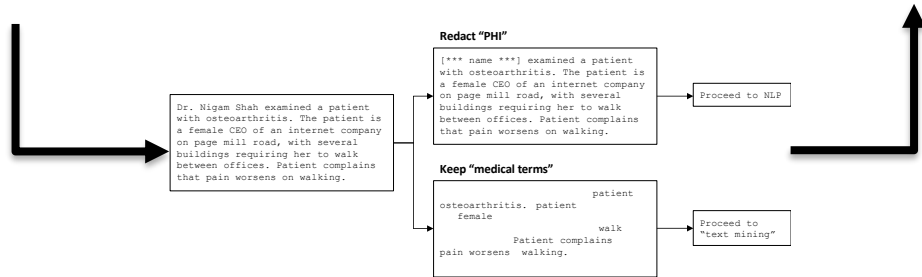
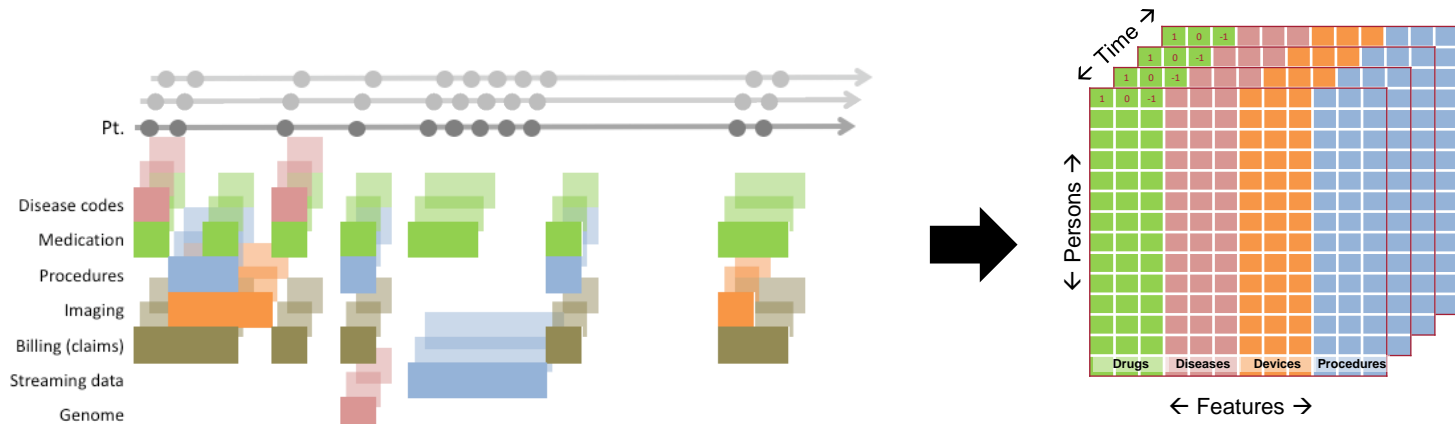
Patient Details		GP Details																																																																												
 <p>GME0000 Smith, Caroline</p> <p>Sex: Female DOB: 1940/01/01 Next of kin: John Smith</p> <p>Phone: 365-565-9090 Address: 19 Provincial Rd. Edmonton AB T6M 1R7</p>		<p>Name: Jones, Evans Phone: 333-465-5545 Address: 11 Terrence Ave., Edmonton, AB T4Y 8U9</p>																																																																												
Patient Record		Other Healthcare Providers																																																																												
<p>Alerts</p> <ul style="list-style-type: none"> Allergies - Sulfa Drugs Pap smear due Td due A1C above target 		<table border="1"> <thead> <tr> <th>Name</th> <th>Disp.</th> <th>Last Encounter</th> <th>Next encounter</th> <th>Right of Access</th> </tr> </thead> <tbody> <tr> <td>Diaz, Ellen</td> <td>Cardiology</td> <td>01/2006</td> <td>07/2006</td> <td>Y</td> </tr> <tr> <td>Fournier, Janice</td> <td>RN</td> <td>08/2005</td> <td></td> <td>N</td> </tr> <tr> <td>Cohen, Richard</td> <td>Dermatology</td> <td>07/2005</td> <td></td> <td>N</td> </tr> </tbody> </table>		Name	Disp.	Last Encounter	Next encounter	Right of Access	Diaz, Ellen	Cardiology	01/2006	07/2006	Y	Fournier, Janice	RN	08/2005		N	Cohen, Richard	Dermatology	07/2005		N																																																							
Name	Disp.	Last Encounter	Next encounter	Right of Access																																																																										
Diaz, Ellen	Cardiology	01/2006	07/2006	Y																																																																										
Fournier, Janice	RN	08/2005		N																																																																										
Cohen, Richard	Dermatology	07/2005		N																																																																										
Diagnosis		Medications																																																																												
<table border="1"> <thead> <tr> <th>Date</th> <th>State</th> <th>Status</th> </tr> </thead> <tbody> <tr> <td>11/1989</td> <td>Hypertension</td> <td>Ongoing</td> </tr> <tr> <td>05/1996</td> <td>Diabetes</td> <td>Ongoing</td> </tr> <tr> <td>02/2002</td> <td>Coronary Artery Disease</td> <td>Ongoing</td> </tr> <tr> <td>12/2005</td> <td>Fasting lipids</td> <td></td> </tr> <tr> <td>1/2005</td> <td>Exercise stress test</td> <td></td> </tr> <tr> <td>02/2005</td> <td>Coronary angiogram / Cellulitis</td> <td>Resolved</td> </tr> <tr> <td>05/1981</td> <td>Cholecystectomy</td> <td>Resolved</td> </tr> <tr> <td>01/1967</td> <td>Cesarian section</td> <td>Resolved</td> </tr> </tbody> </table>		Date	State	Status	11/1989	Hypertension	Ongoing	05/1996	Diabetes	Ongoing	02/2002	Coronary Artery Disease	Ongoing	12/2005	Fasting lipids		1/2005	Exercise stress test		02/2005	Coronary angiogram / Cellulitis	Resolved	05/1981	Cholecystectomy	Resolved	01/1967	Cesarian section	Resolved	<table border="1"> <thead> <tr> <th>Date</th> <th>Medications</th> <th>Prescriptions</th> <th>Last Filled</th> </tr> </thead> <tbody> <tr> <td>11/1989</td> <td>Hydrochlorothiazide 25 mg</td> <td>One tab at breakfast</td> <td>12/2005</td> </tr> <tr> <td>03/1999</td> <td>Glyburide 5 mg</td> <td>One tab twice daily</td> <td>12/2005</td> </tr> <tr> <td>01/2001</td> <td>Metformin 500 mg</td> <td>Two tabs twice daily</td> <td>12/2005</td> </tr> <tr> <td>03/2001</td> <td>Atorvastatin 20 mg</td> <td>One tab at supper</td> <td>12/2005</td> </tr> <tr> <td>02/2002</td> <td>Atenolol 50 mg</td> <td>One tab at breakfast</td> <td>12/2005</td> </tr> <tr> <td>02/2002</td> <td>ECASA 325 mg</td> <td>One tab at breakfast</td> <td>12/2005</td> </tr> <tr> <td>02/2006</td> <td>Ramipril 10mg</td> <td>One tab at supper</td> <td>02/2006</td> </tr> <tr> <td>06/2005</td> <td>Cloxacillin 500 mg</td> <td>Discontinued</td> <td></td> </tr> <tr> <td>05/2004</td> <td>Beclomethasone Cream</td> <td>Discontinued</td> <td></td> </tr> </tbody> </table>		Date	Medications	Prescriptions	Last Filled	11/1989	Hydrochlorothiazide 25 mg	One tab at breakfast	12/2005	03/1999	Glyburide 5 mg	One tab twice daily	12/2005	01/2001	Metformin 500 mg	Two tabs twice daily	12/2005	03/2001	Atorvastatin 20 mg	One tab at supper	12/2005	02/2002	Atenolol 50 mg	One tab at breakfast	12/2005	02/2002	ECASA 325 mg	One tab at breakfast	12/2005	02/2006	Ramipril 10mg	One tab at supper	02/2006	06/2005	Cloxacillin 500 mg	Discontinued		05/2004	Beclomethasone Cream	Discontinued									
Date	State	Status																																																																												
11/1989	Hypertension	Ongoing																																																																												
05/1996	Diabetes	Ongoing																																																																												
02/2002	Coronary Artery Disease	Ongoing																																																																												
12/2005	Fasting lipids																																																																													
1/2005	Exercise stress test																																																																													
02/2005	Coronary angiogram / Cellulitis	Resolved																																																																												
05/1981	Cholecystectomy	Resolved																																																																												
01/1967	Cesarian section	Resolved																																																																												
Date	Medications	Prescriptions	Last Filled																																																																											
11/1989	Hydrochlorothiazide 25 mg	One tab at breakfast	12/2005																																																																											
03/1999	Glyburide 5 mg	One tab twice daily	12/2005																																																																											
01/2001	Metformin 500 mg	Two tabs twice daily	12/2005																																																																											
03/2001	Atorvastatin 20 mg	One tab at supper	12/2005																																																																											
02/2002	Atenolol 50 mg	One tab at breakfast	12/2005																																																																											
02/2002	ECASA 325 mg	One tab at breakfast	12/2005																																																																											
02/2006	Ramipril 10mg	One tab at supper	02/2006																																																																											
06/2005	Cloxacillin 500 mg	Discontinued																																																																												
05/2004	Beclomethasone Cream	Discontinued																																																																												
Immunizations		Encounter History																																																																												
<table border="1"> <thead> <tr> <th>Type</th> <th>Most Recent</th> <th>Number Received</th> </tr> </thead> <tbody> <tr> <td>Influenza</td> <td>1/2/2005</td> <td>7</td> </tr> <tr> <td>Pneumovax</td> <td>03/2005</td> <td>1</td> </tr> <tr> <td>Twinrix</td> <td>08/2002</td> <td>3</td> </tr> <tr> <td>Td</td> <td>04/1996</td> <td>1</td> </tr> </tbody> </table>		Type	Most Recent	Number Received	Influenza	1/2/2005	7	Pneumovax	03/2005	1	Twinrix	08/2002	3	Td	04/1996	1	<table border="1"> <thead> <tr> <th>Date</th> <th>Facility</th> <th>Speciality</th> <th>Clinician</th> <th>Reason</th> <th>Type</th> </tr> </thead> <tbody> <tr> <td>02/2006</td> <td>GP</td> <td></td> <td></td> <td>Hypertension</td> <td>Outpatient</td> </tr> <tr> <td>01/2006</td> <td>Cardio Assoc</td> <td>Cardiology</td> <td>Diaz, E.</td> <td>CAD</td> <td>Outpatient</td> </tr> <tr> <td>12/2005</td> <td>GP</td> <td></td> <td></td> <td>Diabetes</td> <td>-</td> </tr> <tr> <td>10/2005</td> <td>General Hosp</td> <td>Dietician</td> <td>Johnson, H.</td> <td>Diabetes teaching</td> <td>Outpatient</td> </tr> <tr> <td>08/2005</td> <td>GP</td> <td></td> <td></td> <td>Diabetes</td> <td>-</td> </tr> <tr> <td>08/2005</td> <td>GP</td> <td></td> <td></td> <td>Cellulitis</td> <td>-</td> </tr> <tr> <td>08/2005</td> <td>Home Visit</td> <td>RN</td> <td>Fournier, J.</td> <td>Cellulitis</td> <td>-</td> </tr> <tr> <td>08/2005</td> <td>GP</td> <td></td> <td></td> <td>Cellulitis</td> <td>-</td> </tr> <tr> <td>07/2005</td> <td>Polyclinic</td> <td>Dermatology</td> <td>Cohen, R.</td> <td>Stasis dermatitis</td> <td>Outpatient</td> </tr> </tbody> </table>		Date	Facility	Speciality	Clinician	Reason	Type	02/2006	GP			Hypertension	Outpatient	01/2006	Cardio Assoc	Cardiology	Diaz, E.	CAD	Outpatient	12/2005	GP			Diabetes	-	10/2005	General Hosp	Dietician	Johnson, H.	Diabetes teaching	Outpatient	08/2005	GP			Diabetes	-	08/2005	GP			Cellulitis	-	08/2005	Home Visit	RN	Fournier, J.	Cellulitis	-	08/2005	GP			Cellulitis	-	07/2005	Polyclinic	Dermatology	Cohen, R.	Stasis dermatitis	Outpatient
Type	Most Recent	Number Received																																																																												
Influenza	1/2/2005	7																																																																												
Pneumovax	03/2005	1																																																																												
Twinrix	08/2002	3																																																																												
Td	04/1996	1																																																																												
Date	Facility	Speciality	Clinician	Reason	Type																																																																									
02/2006	GP			Hypertension	Outpatient																																																																									
01/2006	Cardio Assoc	Cardiology	Diaz, E.	CAD	Outpatient																																																																									
12/2005	GP			Diabetes	-																																																																									
10/2005	General Hosp	Dietician	Johnson, H.	Diabetes teaching	Outpatient																																																																									
08/2005	GP			Diabetes	-																																																																									
08/2005	GP			Cellulitis	-																																																																									
08/2005	Home Visit	RN	Fournier, J.	Cellulitis	-																																																																									
08/2005	GP			Cellulitis	-																																																																									
07/2005	Polyclinic	Dermatology	Cohen, R.	Stasis dermatitis	Outpatient																																																																									
Diabetic Indices																																																																														
<table border="1"> <thead> <tr> <th>Type</th> <th>Value</th> <th>Most Recent</th> </tr> </thead> <tbody> <tr> <td>A1C</td> <td>0.071</td> <td>12/2005</td> </tr> <tr> <td>LDL</td> <td>2.41</td> <td>12/2005</td> </tr> <tr> <td>BP</td> <td>135/75</td> <td>02/2006</td> </tr> <tr> <td>Urine Microalb</td> <td>0.02</td> <td>08/2005</td> </tr> <tr> <td>Eye Exam</td> <td></td> <td>05/2005</td> </tr> <tr> <td>Home Gluc (average)</td> <td>7.4</td> <td>01/2006</td> </tr> </tbody> </table>		Type	Value	Most Recent	A1C	0.071	12/2005	LDL	2.41	12/2005	BP	135/75	02/2006	Urine Microalb	0.02	08/2005	Eye Exam		05/2005	Home Gluc (average)	7.4	01/2006																																																								
Type	Value	Most Recent																																																																												
A1C	0.071	12/2005																																																																												
LDL	2.41	12/2005																																																																												
BP	135/75	02/2006																																																																												
Urine Microalb	0.02	08/2005																																																																												
Eye Exam		05/2005																																																																												
Home Gluc (average)	7.4	01/2006																																																																												



Decisions to make:

- About source and choice of features, including text
- About handling of time
- About defining an electronic phenotype
- About building a cohort

Text is one source to use [be careful how deep you go!]



Core concepts

- Processing of Text vs. Natural Language Processing
- Handling of knowledge graphs
- Ascertaining whether the patient had an “event” or not (and when)
 - We can choose to do this task just using a textual document
 - We can do this task as a classification problem, where text-extracted content is just one element on a patient timeline.

Handling of text – two world-views

Dr. Nigam Shah examined a patient with osteoarthritis. The patient is a female CEO of an internet company on page mill road, with several buildings requiring her to walk between offices. Patient complains that pain worsens on walking.

[*** name ***] examined a patient with osteoarthritis. The patient is a female CEO of an internet company on page mill road, with several buildings requiring her to walk between offices. Patient complains that pain worsens on walking.

Redacting “PHI”

osteoarthritis. patient
female patient
walk
Patient complains
pain worsens walking.

Keeping “medical terms”

Simple text processing

1

Input Text

PAST MEDICAL/SURGICAL HISTORY: Positive for atrial fibrillation. The patient had AVR 6 years ago. Peripheral arterial disease with hypertension, peripheral neuropathy, atherosclerosis, hemorrhoids, proctitis, CABG, and cholecystectomy.

FAMILY HISTORY: Positive for atherosclerosis, hypertension, autoimmune diseases in the family.

REVIEW OF SYSTEMS: Weight loss of 25 pounds within the last 6 months, shortness of breath, constipation, bleeding from hemorrhoids, increased frequency of urination, muscle aches, dizziness and faintness, focal weakness and numbness in both legs, knees and feet.

LABORATORY DATA AND RADIOLOGICAL RESULTS: The patient had a chest x-ray, which showed cardiomegaly with atherosclerotic heart disease, pleural thickening and small pleural effusions, a left costophrenic angle which has not changed when compared to prior examination, COPD pattern. The patient also had a head CT, which showed atrophy with old ischemic changes. No acute intracranial findings.

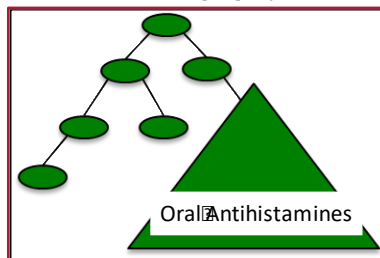
DISCHARGE DIAGNOSIS: Syncope.

DISCHARGE MEDICATIONS: The patient was discharged on the following medications: Cardizem 90 mg p.o. thrice daily, digoxin 0.125 mg p.o. once daily, allopurinol 100 mg two times daily, Coumadin 4 mg p.o. q.h.s., and Remeron 15 mg p.o. q.h.s.

NegEx and ConText Patterns

Text processing

Knowledge graph



53 8
44 5
4539 8
3 16 (colon)
496 3
1363 13 (atrial fibrillation)
1 19 (period)
8 7
5087 12
129 6
158 6
14091 3 (peripheral arterial disease)
254 33
2 12
4624 2
4198 2 (coma)
2 15 (coma)
2835 2
2 11
10647 2 (proctitis)
2 9
2026 2 (cabg)
2 4
11 2
1907 4 (cholecystectomy)
1 15 (period)
...

past medical surgical history: positive for atrial fibrillation. patient avr years ago. peripheral arterial disease hypertension, peripheral neuropathy, atherosclerosis, hemorrhoids, proctitis, cabg, and cholecystectomy.

family history: positive for atherosclerosis, hypertension, autoimmune diseases

review of systems: weight loss pounds months, shortness of breath, constipation, bleeding hemorrhoids, increased frequency of urination, muscle aches, dizziness and faintness, focal weakness and numbness both legs, knees and feet.

laboratory data and results: patient chest x-ray, which cardiomegaly atherosclerotic heart disease, pleural thickening and small pleural effusion, left costophrenic angle which not changed compared prior examination, copd pattern. patient head ct, which atrophy old ischemic . no acute intracranial findings.

discharge diagnosis: syncope

discharge medications: patient following medications: Cardizem 90 mg p.o. thrice daily, digoxin 0.125 mg p.o. once daily, allopurinol 100 mg two times daily, Coumadin 4 mg p.o. q.h.s., and Remeron 15 mg p.o. q.h.s.

True Internal Representation
(with some keys shown for illustration)

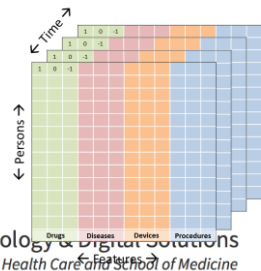
Reconstructed Representation

Defer binding to concepts

2

- Event / Outcome concepts**
- Juvenile Idiopathic Arthritis
 - :
 - Uveitits
 - Iridocyclitis

3

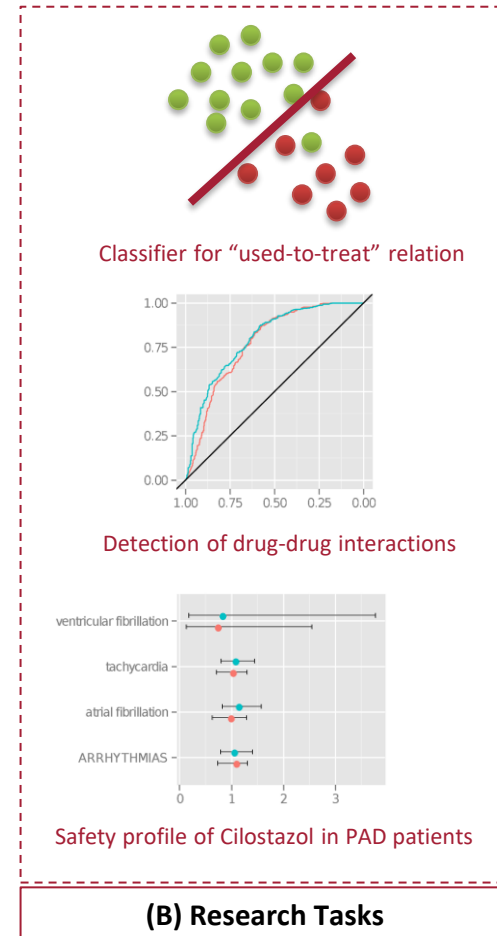
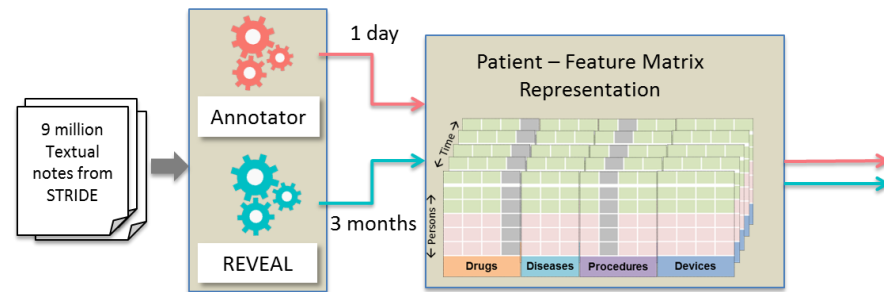
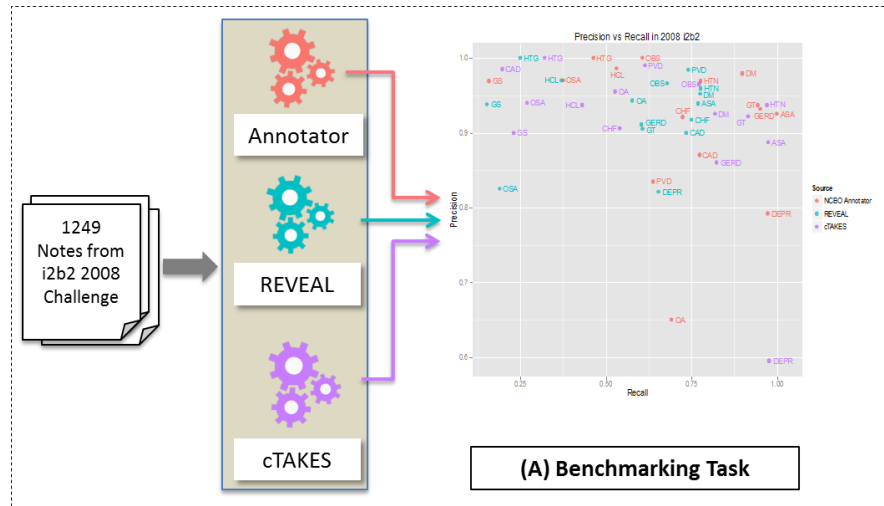


Count present, positive mentions, about the patient

Juvenile Idiopathic Arthritis
ICD 9 codes
696.0, 714.0, 714.2, 714.3, 714.9, 720.2, 720.9

Terms:
Juvenile idiopathic arthritis, JIA
Juvenile rheumatoid arthritis, JRA
Psoriatic arthritis
Juvenile spondyloarthropathy, spondyloarthritis, enthesitis related arthritis, sacroiliitis, reactive arthritis

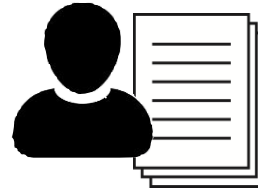
Trade-off: simple or advanced [text-processing]



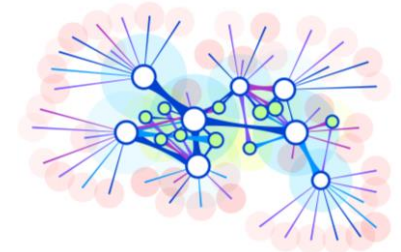
Weak supervision

The State of Clinical Concept Extraction in 2020

- Largely rule-based + manual feature engineering
- Many clinical entity types do not have expert labeled training sets available
- For existing labeled datasets we're locked into their annotation choices
- **PHI / Privacy considerations when sharing labeled data**



not many hand-labeled datasets



but many ontologies!

Fu et al. 2020 *Journal of Biomedical Informatics*.
“Clinical concept extraction: A methodology review”

Ontology-based Labeling

Semantic Labeling

Term mapped to entity type probability

```
aortic stenosis -> [0.0, 1.0]
  present -> [1.0, 0.0]
  aortic -> [1.0, 0.0]
```

Synonym Labeling

Use entire document context to tag **synonymous** entities

```
{aortic stenosis, AS} -> [0.0, 1.0]
{mitral regurgitation, MR} -> [0.0, 1.0]
```

AORTIC VALVE: Mildly thickened aortic valve leaflets (3). No AS.
MITRAL VALVE: Normal mitral valve leaflets with trivial MR.

CONCLUSIONS
The aortic valve leaflets (3) are mildly thickened but aortic stenosis is not present. Trace aortic regurgitation may be present. The mitral valve appears structurally normal with trivial mitral regurgitation.

Ontologies —> **Type** and **Synonymy** Information

Weakly Supervised: Ontologies + Task Rules

Task	+Task-specific rules				Hand-labeled	
	LFs	MV	LM	WS	FS	SOTA
Chemical	+9	81.1	89.2 ± 0.2 [†]	91.1 ± 0.1*	92.4 ± 0.2	93.5 ²⁴
Disease	+6	76.4	79.8 ± 0.3 [†]	79.9 ± 0.2	84.5 ± 0.2	87.2 ²⁴
Disorder	+11	71.2	75.0 ± 0.2 [†]	76.3 ± 0.1*	79.6 ± 0.3	80.1 ⁶⁵
Drug	+11	82.2	85.8 ± 0.4 [†]	88.3 ± 0.3*	93.2 ± 0.3	91.4 ⁶⁶
Negation	17	92.5	93.0 ± 0.0[†]	92.7 ± 0.6*	96.1 ± 0.2	~
DocTimeRel	27	67.8	69.2 ± 0.0 [†]	72.9 ± 0.5*	86.2 ± 0.1	83.4 ⁶⁷

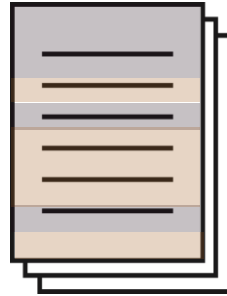
NER performance within 4.1% F1 of models trained with hand-labeled data

Continuous Symptom Monitoring for COVID-19

Could we use presenting symptoms to prioritize who should be tested for COVID-19?



Daily feed of all notes + timestamped edit deltas



Use first 30 minutes of note edits

Virus	AUROC
Adenovirus	0.68 (0.60–0.76)
Influenza virus A	0.73 (0.68–0.77)
Metapneumovirus	0.64 (0.57–0.71)
Parainfluenza virus	0.60 (0.53–0.68)
RSV	0.77 (0.73–0.80)
Rhinovirus	0.62 (0.58–0.66)
SARS-CoV-2	0.64 (0.49–0.79)

Limited ability to discriminate respiratory viruses from symptoms alone (Callahan et al. 2020)

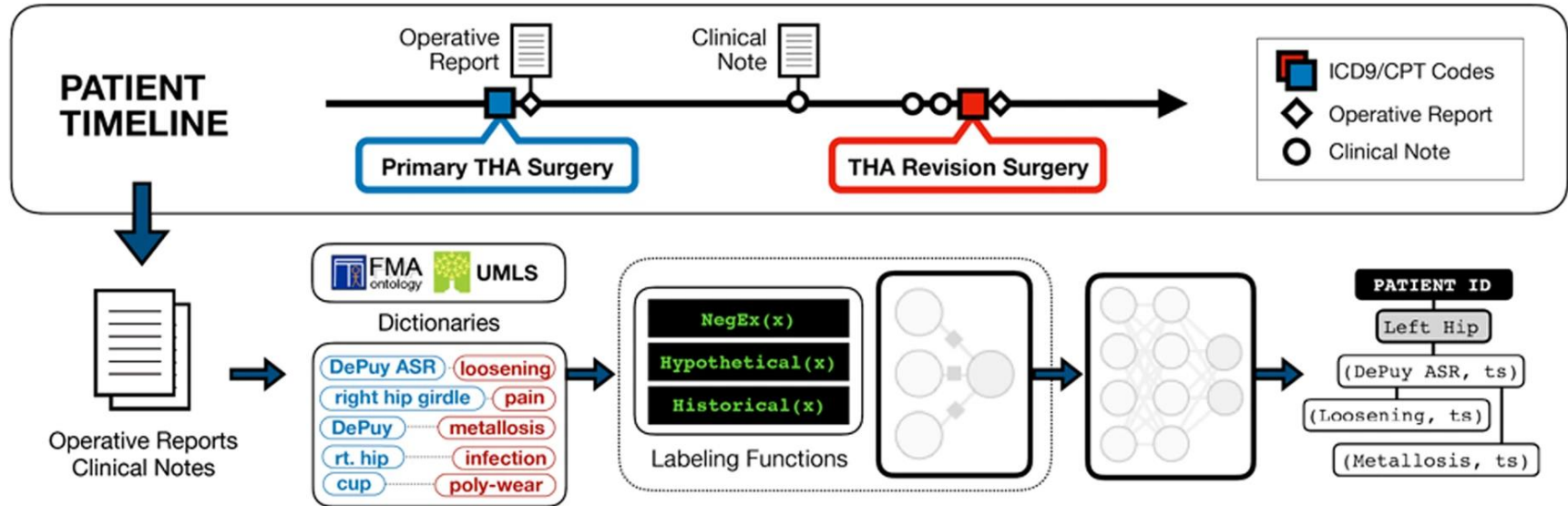
Weakly supervised symptom tagger
+2.3 F1 points over baseline

Enabled Symptom Data Sharing

Bill & Melinda
Gates Foundation

CMU Delphi Group's
COVIDCAST

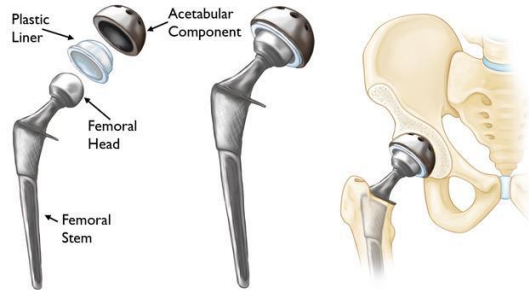
Medical device surveillance with EHRs



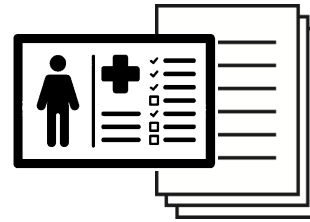
Callahan and Fries et al, *npj Digital Medicine* volume 2, Article number: 94 (2019)

Extracting Structured Information from Notes

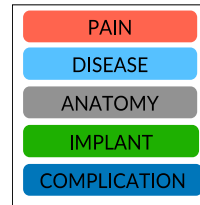
Medical Device Surveillance



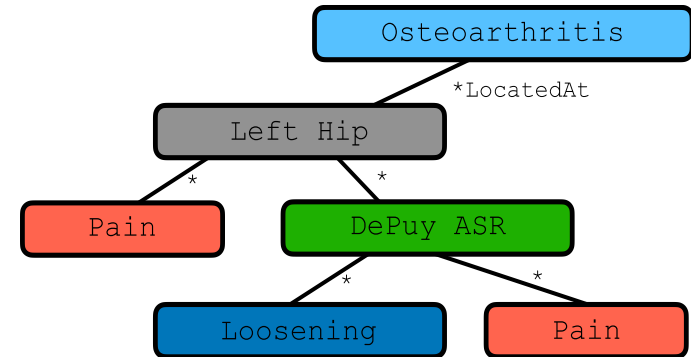
2018 global investigation on the medical device industry



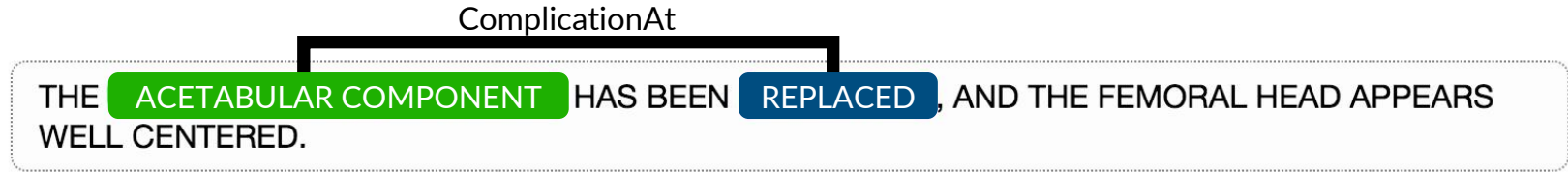
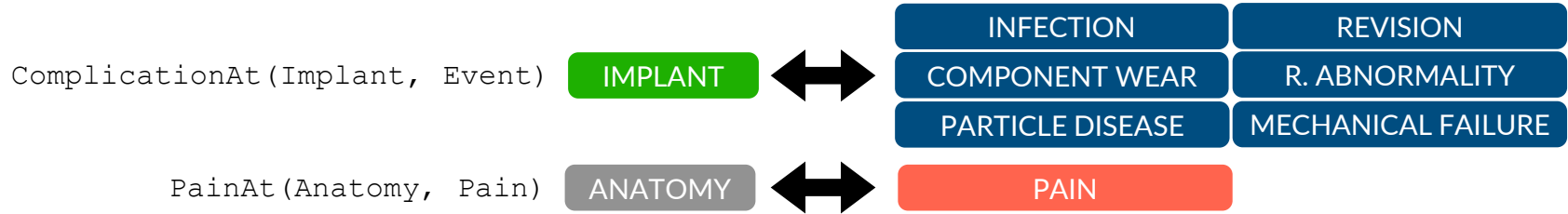
Entities Relations



Entity Types



Implant Outcomes as Relation Extraction



Complication relations identified over sentences

Label Small Gold Standard Dev & Test Sets

Programmatically label the training set

Example labeling function

A. CLINICAL NOTE + MARKUP

HISTORY OF PRESENT ILLNESS:
60 yo male with **infected R hip (MRSA)** s/p previous **hip replacement**.
LTHA **November 2004** demonstrates **component wear**.
No **lucencies** were observed around the **implant**.
Implant is being evaluated for possible **revision**.

PAST MEDICAL HISTORY:
Hx **right Zimmer Biomet hip** **1/1/05** complicated by **infection**.

NOTE DATE: **07/01/2008 06:11 PM**

ENTITIES: **HEADER** **CLINICAL CONCEPT** **DATETIME**
ATTRIBUTES: **HYPOTHETICAL** **HISTORICAL** **NEGATED** **TIME DELTA**

Output from text-processor

B. LABELING FUNCTION DEFINITIONS

```
def LF1_contiguous_entities(c):  
    v = len(between_words(c)) == 0  
    return TRUE if v else ABSTAIN
```

```
def LF2_historical(c):  
    v = has_historical_attrib(c)  
    return FALSE if v else ABSTAIN
```

```
def LF3_reject_section(c):  
    h1 = get_section_header(c)  
    v = h1 in reject_headers  
    return FALSE if v else ABSTAIN
```

```
def LF4_negated(c):  
    v = NegEx.is_negated(c)  
    return FALSE if v else ABSTAIN
```

FALSE: -1 **ABSTAIN:** 0 **TRUE:** 1

Labeling function

CATEGORY	# INSTANCES	PRECISION	RECALL	F1	+/- F1
Revision	63	75.5	58.7	66.1	+16.2%
Component Wear	48	72.9	72.9	72.9	+38.6%
Mechanical Failure	25	91.7	44.0	59.5	+40.3%
Particle Disease	65	97.1	52.3	68.0	+496.5%
Radiographic Abnormality	17	60.0	25.3	44.4	-18.5%
Infection	58	90.7	84.5	87.5	+54.0%
Implant-Complications	276	82.7	62.3	71.1	+53.2%
Pain-Anatomy	236	80.2	82.6	81.4	+12.7%

Big Boost
in Recall

6x

More complication
events than with
coded data alone

Closing thoughts

CLINICAL DATA SOURCES

Stanford
Mt. Sinai
Sutter
Denmark

stride 5
stride 6
OHDSI

text

TEXT PROCESSING

mgrep
Unitex
Lexigram
Deep Dive
CLEVER

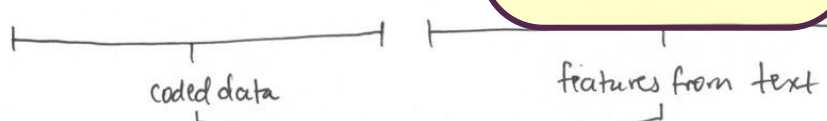
Text processing

KNOWLEDGE GRAPH

terminology
terminology 4

ontology
ontologies

Knowledge Graph handling



queries
search engine

Search

frequencies
co-frequencies
information content
syntactic types
bigrams + trigrams
stride, mayo, medline

Phenotyping Strategy

distance learning for phenotype extraction (CLEVER)

term to concept maps
from knowledge graph

classifiers for phenotypes
- XPRESS
- APHRODITE

generative models

list of patients
data extract

patient feature matrix

Patient feature matrix

features



Acknowledgements and Contact Information

<https://shahlab.stanford.edu/>
email: nigam@stanford.edu

