NLP Innovations – what does the healthcare industry see as knowledge gap in the NLP research

Moderator:

Hongfang Liu, Mayo Clinic

Panelists:

Eric Brown, IBM Watson

Richard Wolniewicz, 3M Health Information Systems

Brian Hazlehurst, Kaiser Permanente Northwest Center for Health Research

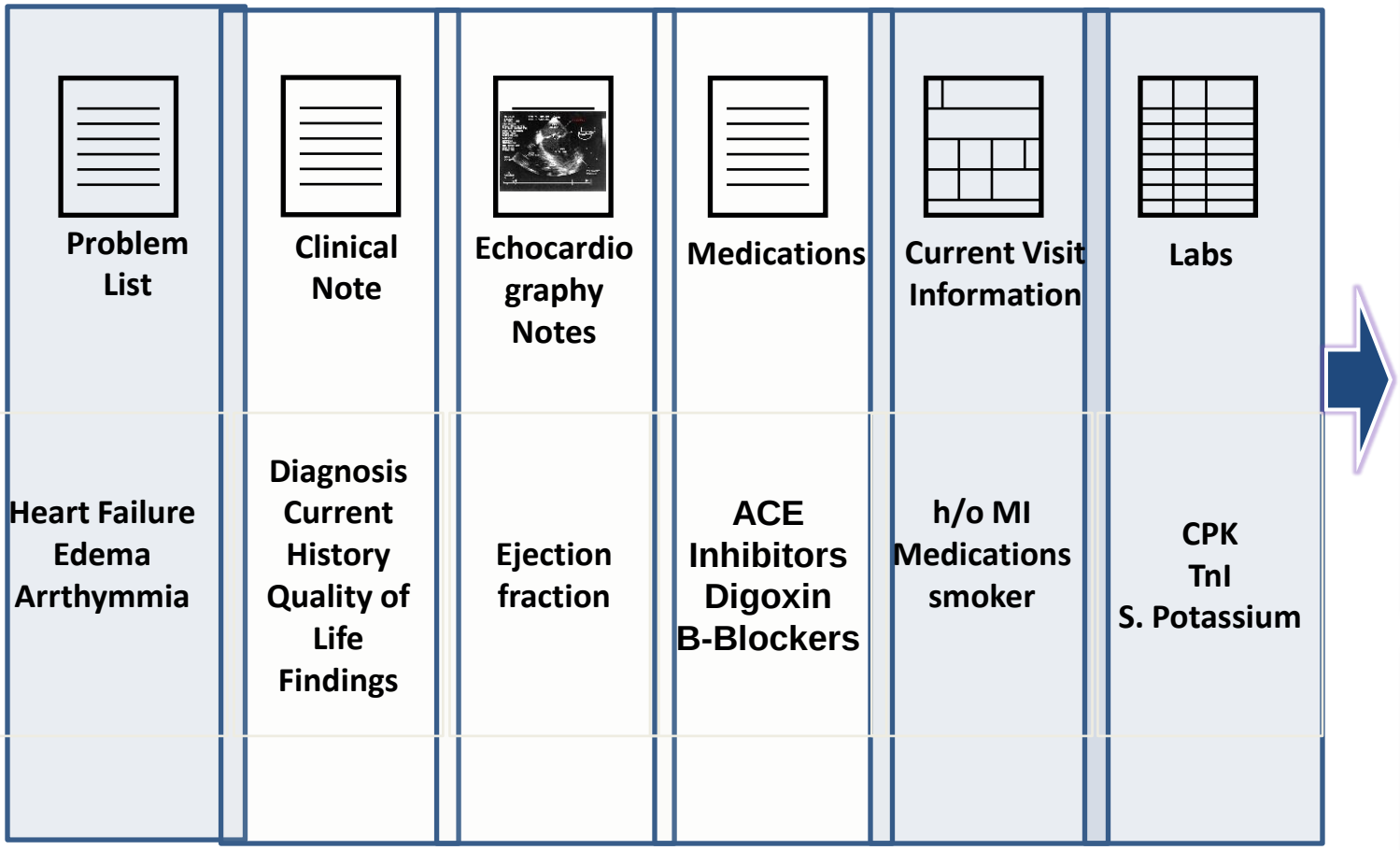
Scott DuVall, VA Salk Lake City VAMC

Ruth Reeves, VA Tennessee Valley VAMC

Dialogues in a clinical setting

- From researchers:
  - Great, you can do NLP, right? *I want to retrieve the number of patients newly diagnosis of diabetes with GI symptoms. I need that number for my submission next week.*
- From quality improvement analysts:
  - NSQIP – 30-day re-admission? *Can you help me obtain the number of re-admissions from EMR? VA did that.*
- From clinicians:
  - Documentation support and computer-assisted coding
  - Clinical decision support
  - Outcomes.
- From leaders:
  - We have established partnerships with XXX, YYY and ZZZ companies to solve the above problems.

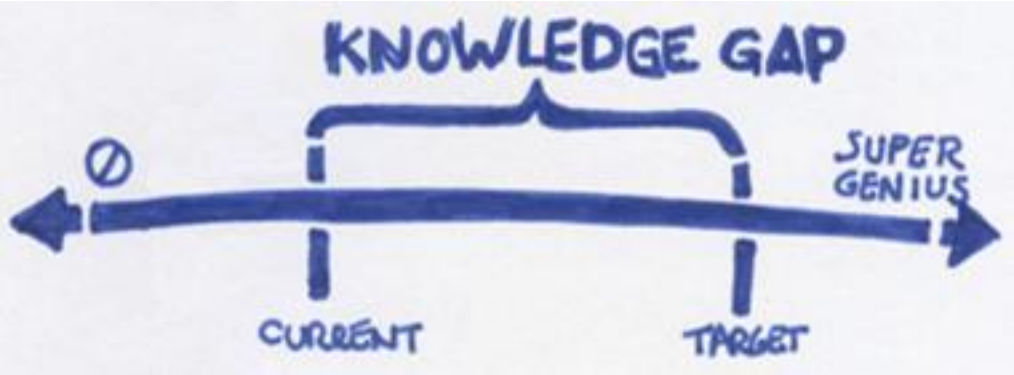
NLP use case: Heart Failure



Quality reporting
e.g. how many HF admissions

Decision support
e.g. put diagnosis on problem list and provide educational material to HF patients

Clinical Research
e.g. identify HF cohort for survey


KNOWLEDGE GAP
∅
SUPER GENIUS
CURRENT
TARGET

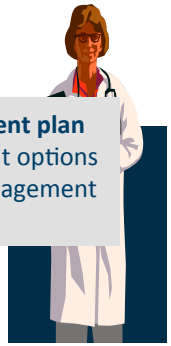
IBM **Research**

IBM Watson NLP in Healthcare

Eric Brown, PhD
IBM Watson Group
Yorktown Heights, NY

IBM **Confidential**

Physician's Workflow



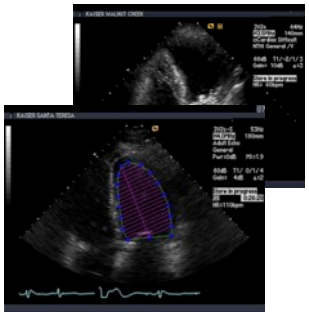
Get an overview of the patient
- What are the patient's active problems?
- What are possible concerns that need to be addressed?

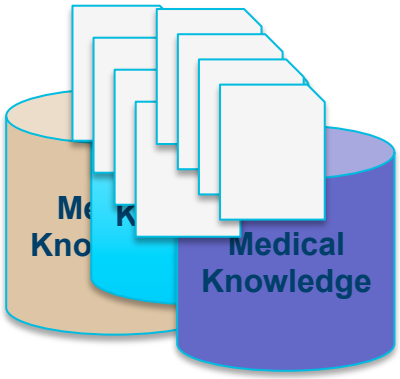
Review the patient's history
- Chief compliant or reason for the visit?
- Past medical history?
- Current medications?
- Most recent lab results?

Determine diagnostic workup
- Explore diagnostic tests and screening / diagnostic guidelines
- Apply knowledge to make diagnosis

Determine treatment plan
- Explore treatment options and possible management guidelines

Visit Preparation → **History and Physical** → **Assessment** → **Plan**

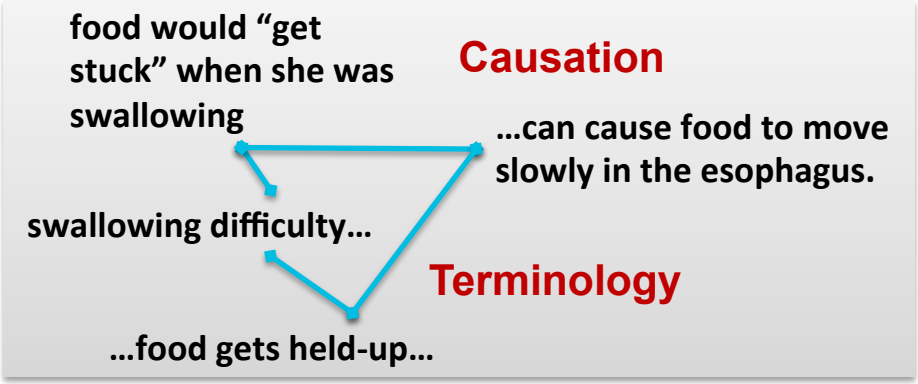


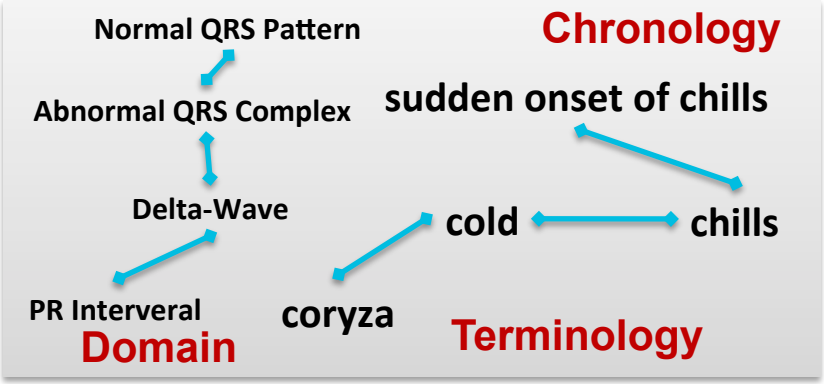


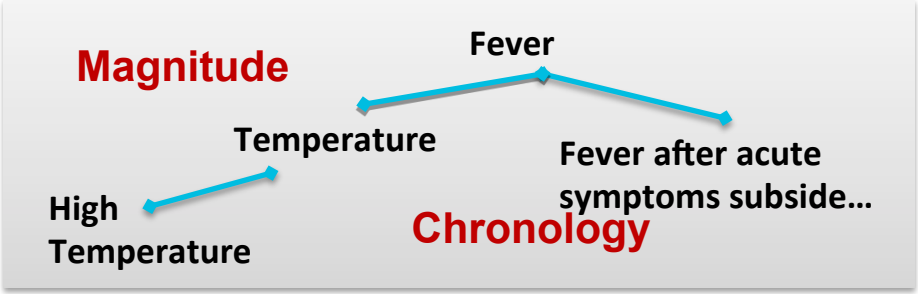
Performance Metrics Evaluation, Research, Cohort Analysis

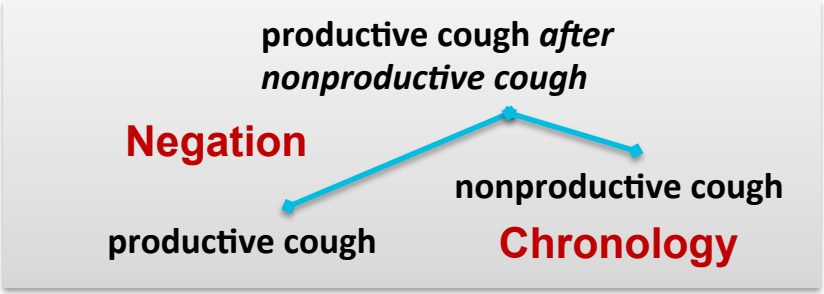
© 2015 International Business Machines Corporation

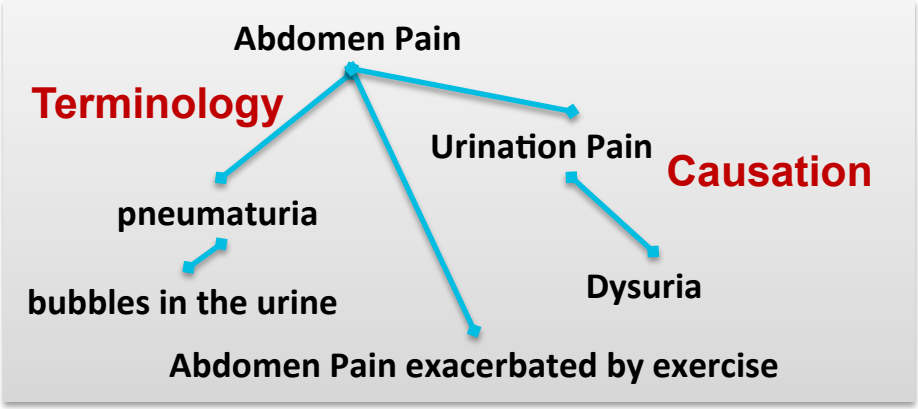
Natural Language Understanding is Foundational

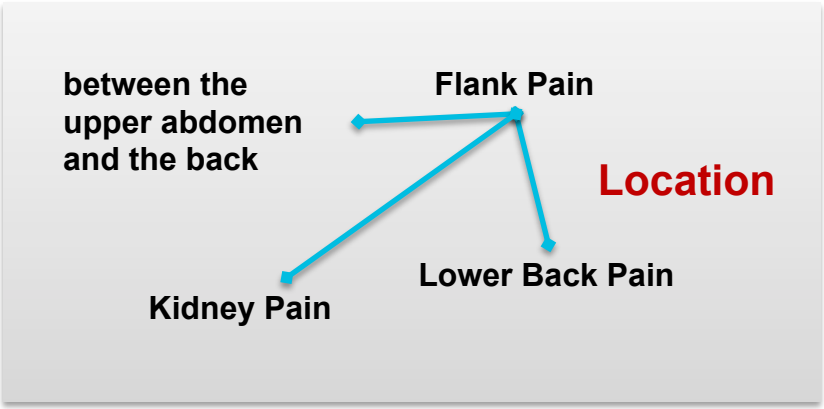






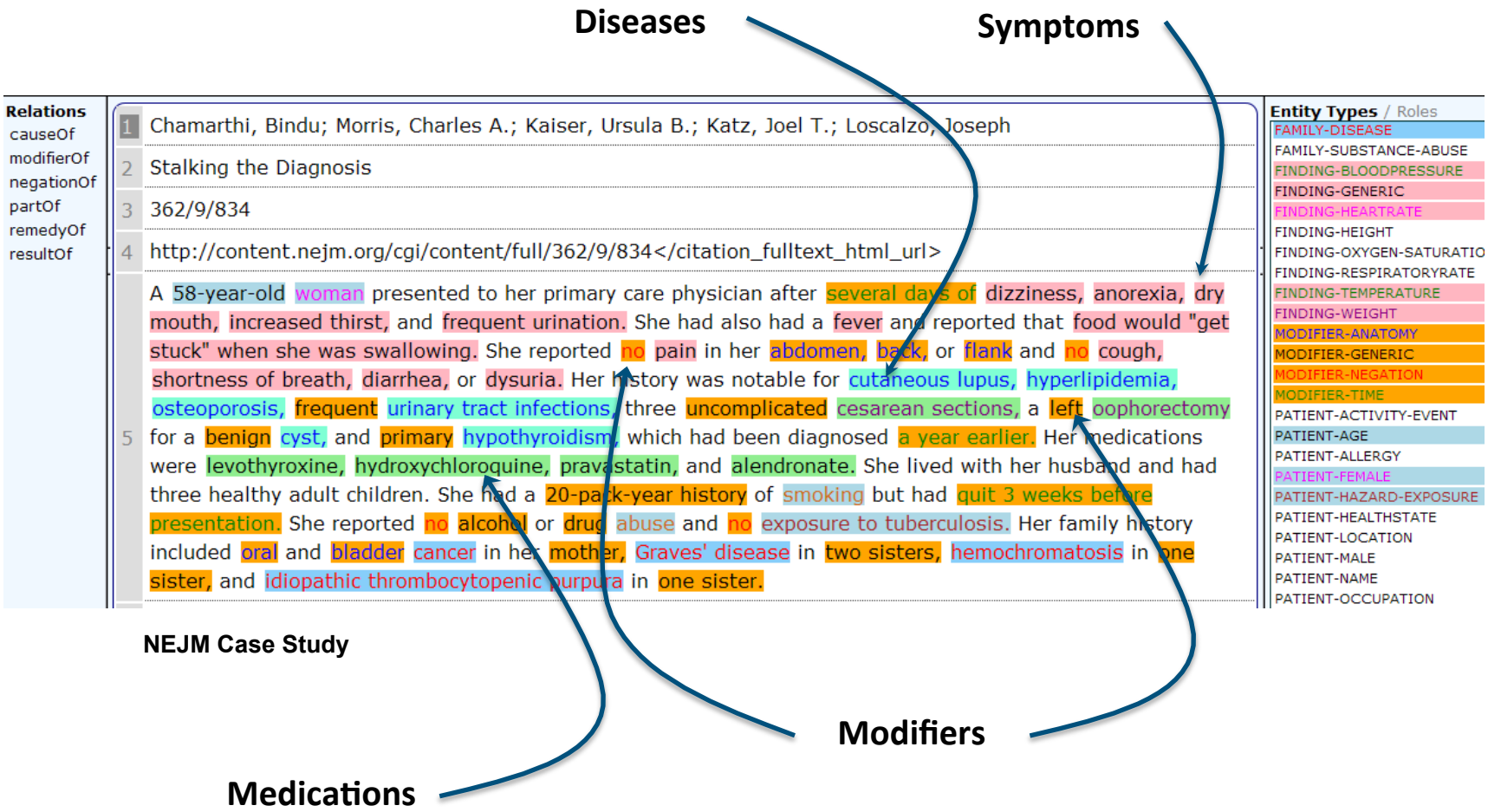






© 2015 International Business Machines Corporation

Text Analytics for Automatic Semantic Annotation



NEJM Case Study

© 2015 International Business Machines Corporation

Typical Discourse Analysis Challenges

Co-Reference

Anaphora

Aortic stenosis.With aortic stenosis, the murmur is systolic, beginning after S1 and ending at or before aortic valve closure. It's harsh and grating, medium-pitched, and crescendo-decrescendo

No Co-referent term

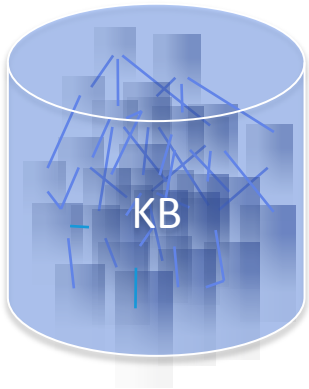
The clinical presentation of the patient with constriction resembles that of the individual with tamponade except for normal pulse pressure and lack of pulsus paradoxus in constriction.
Inspiratory increase in jugular venous pressure (Kussmaul's sign) is occasionally seen…

Discourse Segmentation (Confusors)

Collagenous colitis and lymphocytic colitis are distinguished by the presence or absence of a thickened subepithelial collagen layer. The cause of microscopic colitis syndrome is uncertain

© 2015 International Business Machines Corporation

Leveraging Existing Medical Resources

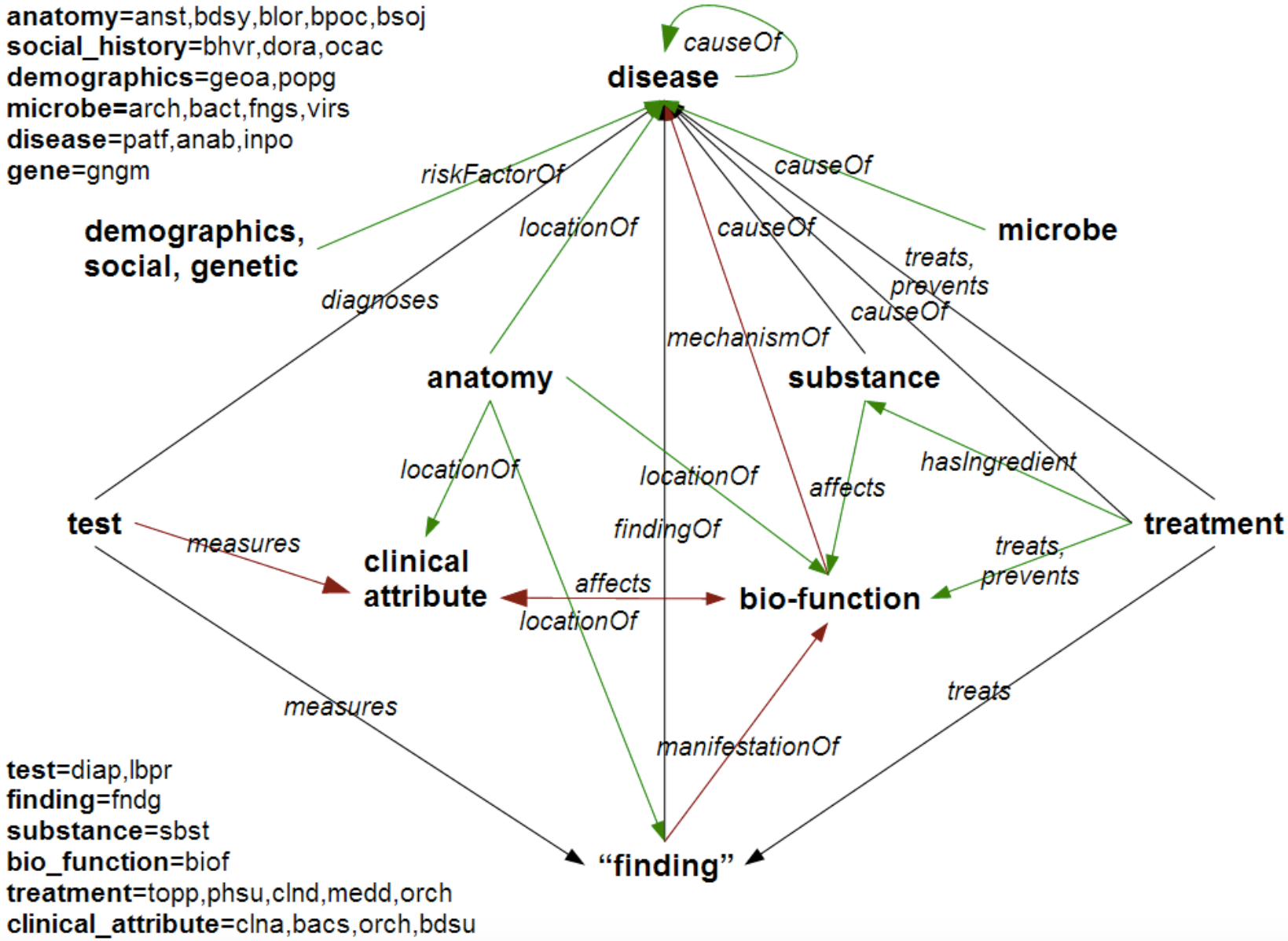


- UMLS (Unified Medical Language System) from NLM
  - ~100 sources, sort of merged
  - ~3M unique concept identifiers (not unique concepts), organized in a type hierarchy
    - activities, anatomy, chemicals/drugs, devices, disorders, genetics, organisms, physiology, procedures, ...
  - ~350 relation types; ~30M unique relation instances
    - diagnoses, treats, finding_site_of, has_causative_agent, contraindicates, …

- Sample Uses of UMLS
  - Type Coercion: does a candidate answer match the type the question is seeking
  - Candidate generation
  - Term matching
  - Clinical factor identification
  - Relation generation in inference graph

© 2015 International Business Machines Corporation

Domain Model – key entities and relationships



© 2015 International Business Machines Corporation

Individual EMR = plain text + semi-structured data

EMR
(up to 50MBs per patient)

Encounter (Clinical) Notes

Medications

Lab Results

NOTES

td65 year old male

Wants to discuss the results of the MRI - has anterospondylolisthesis. GD 1 - reviewed images with him in detail.
Also results of his labs.

He was unable to get Januvia - he is not on it.
Cost is also very expensive
His DM control is improved -
Will hold off starting it at this point
Has lost about 6-8 lb since last time as he is doing all the cooking and thus cooking better.

His wife needs knee surgery and bariatric surgery
He has a history of repaired TOF - saw Dr.**NAME[ZZZ] in 2010 - next in 2013

has trace Microalbuminuria on labs which is new from last time
Will try and improve his diet and try and lose 8 more lb
His BP is excellent.
Will start low dose lisinopril

738.4 Spondylolisthesis of lumbosacral region (primary encounter diagnosis)
Comment: see HPI
Plan: CONSULT TO PHYSICAL THERAPY

244.9 HYPOTHYROIDISM NOS
Comment:
TSH (uU/mL)
Date1
Date Value Range plain Status
**DATE[Feb 02 2012] 3.940 0.400-5.500 (uU/mL) Final

Plan: well controlled

746.89 VSD
Comment: has loud holosystolic murmur heard over entire precordium
Plan: stable. f'u with Dr. **NAME[ZZZ] next year

272.4 HYPERLIPIDEMIA **INITIALS
Comment: Last Lipid Panel
CHOL 134 **DATE[Jan 22 2013]
HDL 42 **DATE[Jan 22 2013]
LDL 72 **DATE[Jan 22 2013]
TG 98 **DATE[Jan 22 2013]

MEDICATIONS

ENCOUNTER ID	ORDERING TIME	ORDER CLASS	MEDICATION	ORDER STATUS	ORDERING MODE	DOSAGE	STRENGTH	FORM	ROUTE	INSTRUCTIONS	REFILLS	START DATE	END DATE	DISCONTINUE TIME	REASON FOR DISCONTINUE	GENERIC NAME	PHARM CLASS	PHARM SUB CLASS	
4088-117	2000022202400		CIPRO 500 MG TABLET	SENT	Outpatient	1.0	500 mg	tablet	ORAL	1.0 tablet Oral bid	0	20001007	20000221			ciprofloxacin 500 mg tablet	QUINOLONES	Fluoroquinolone Antibiotics	ANTIBIO
4088-117	2000011717000			SENT	Outpatient	0.0				0.0 ointment Topical, dermal bid	0	20001017	20000129	20010129104000	Other				
4088-117	2000110109000			SENT	Outpatient	1.0				1.0 tablet Oral qd	0	20001017	20001103						
4088-117	2000110109000			SENT	Outpatient	1.0				1.0 tablet Oral qd	0	20001103	20000129	20010129104000	Other				
4088-117	2000022202400		AMOXICILLIN 500 MG CAPSULE	SENT	Outpatient	1.0	500 mg	capsule	ORAL	1.0 capsule Oral bid	0	20010221	20000129	20010129104000	Other	amoxicillin 500 mg capsule	PENICILLINS	Aminopenicillin Antibiotic	ANTIBIO
4088-112	2000220414000	PRINT RX	SYNTHROID 150 MCG TABLET	SENT	Outpatient		150 mcg	tablet	ORAL	Take one(1) tablet daily.	0	20021204	20000129	20010129104700	Other	levothyroxine 150 mcg tablet	THYROID HORMONES	Thyroid Hormones - Synthetic T4 (Thyroxine)	THYROI
4088-112	2000220414150	HISTORICAL MED	ACYCLOVIR 400 MG TABLET	SENT	Outpatient		400 mg	tablet	ORAL	qd for 14 days.	0	20021204	20021204			acyclovir 400 mg tablet	ANTIVIRALS, GENERAL	Herpes Antiviral Agent - Purine Analogs	ANTIVI
4088-112	2000220414000	PRINT RX	ACYCLOVIR 400 MG TABLET	SENT	Outpatient		400 mg	tablet	ORAL	qd for 14 days.	1	20021204	20040429	20040429141000		acyclovir 400 mg tablet	ANTIVIRALS, GENERAL	Herpes Antiviral Agent - Purine Analogs	ANTIVI
4088-88	2000012015000	PRINT RX	GLUCOPHAGE 500 MG TABLET	SENT	Outpatient		500 mg	tablet	ORAL	Take one (1) tablet twice daily.	0	20030129	20000129	20030129201400		metformin 500 mg tablet	ANTIHYPERGLYCEMIC, BIGUANIDE TYPE	Insulin Response Enhancers - Biguanides	ANTIHY
4088-88	2000012015400	PRINT RX	SYNTHROID 175 MCG TABLET	SENT	Outpatient		175 mcg	tablet	ORAL	Take one(1) tablet daily.	0	20030129	20000901			levothyroxine 175 mcg tablet	THYROID HORMONES	Thyroid Hormones - Synthetic T4 (Thyroxine)	THYROI
4088-88	2000000015000	NORMAL	SYNTHROID 175 MCG TABLET	SENT	Outpatient		175 mcg	tablet	ORAL	Take one(1) tablet daily.	0	20030927	20040330			levothyroxine 175 mcg tablet	THYROID HORMONES	Thyroid Hormones - Synthetic T4 (Thyroxine)	THYROI
4088-85	2000030040900	NORMAL	SYNTHROID 175 MCG TABLET	SENT	Outpatient		175 mcg	tablet	ORAL	Take one(1) tablet daily.	0	20040330	20040331			levothyroxine 175 mcg tablet	THYROID HORMONES	Thyroid Hormones - Synthetic T4 (Thyroxine)	THYROI
4088-84	2000031024000	NORMAL	SYNTHROID 175 MCG TABLET	SENT	Outpatient		175 mcg	tablet	ORAL	Take one(1) tablet daily.	0	20040331	20040102	20040102122000		levothyroxine 175 mcg tablet	THYROID HORMONES	Thyroid Hormones - Synthetic T4 (Thyroxine)	THYROI
4088-82	2000082000400	PRINT RX	AMOXICILLIN 500 MG CAPSULE	SENT	Outpatient		500 mg	capsule	ORAL	Take one(1) tablet three times daily for 7 days.	0	20040829	20000823	20050823202700		amoxicillin 500 mg capsule	PENICILLINS	Aminopenicillin Antibiotic	ANTIBIO
4088-82	2000082000400	PRINT RX	VALTREX 500 MG TABLET	SENT	Outpatient		500 mg	tablet	ORAL	Take one(1) tablet twice(2) times daily X 14	1	20040829	20000816	20050816231000		valacyclovir 500 mg tablet	ANTIVIRALS, GENERAL	Herpes Antiviral Agent - Purine Analogs	ANTIVI
4088-82	2000082000200	PRINT RX	METFORMIN 500 MG TABLET	SENT	Outpatient		500 mg	tablet	ORAL	Take one(1) tablet twice(2) times daily.	0	20040829	20040102	20040102121000		metformin 500 mg tablet	ANTIHYPERGLYCEMIC, BIGUANIDE TYPE	Insulin Response Enhancers - Biguanides	ANTIHY
4088-71	2000031202000	PRINT RX	METFORMIN 1,000 MG TABLET	SENT	Outpatient		1,000 mg	tablet	ORAL	Take one(1) tablet twice(2) times daily.	0	20041012	20050422	20050422202000		metformin 1,000 mg tablet	ANTIHYPERGLYCEMIC, BIGUANIDE TYPE	Insulin Response Enhancers - Biguanides	ANTIHY
4088-71	2000031202000	PRINT RX	SYNTHROID 175 MCG TABLET	SENT	Outpatient		175 mcg	tablet	ORAL	Take one(1) tablet daily.	0	20041012	20050823	20050823202700		levothyroxine 175 mcg tablet	THYROID HORMONES	Thyroid Hormones - Synthetic T4 (Thyroxine)	THYROI
4088-59	2000021212000	PRINT RX	LIPITOR 10 MG TABLET	SENT	Outpatient		10 mg	tablet	ORAL	Take one(1) tablet daily.	0	20050123	20050816	20050816231000		atorvastatin 10 mg tablet	ANTIHYPERLIPIDEMIC - HMG-COA REDUCTASE INHIBITORS	Antihyperlipidemic - HMG-CoA Reductase Inhibitors (statins)	CARDIC

© 2015 International Business Machines Corporation

Medical Concepts Annotations – A Simpler View

Note-00020.xmi

HPI: **AGE[in 50s] yr old man here for f/u hypothyroid, diabetes mellitus.
c/o feeling sluggish, otherwise wel l. able to perform rather strenuous physical activities.
TTE done.
Eye exam done- reportedly fine.
Meds as of **DATE[Jan 29 2003]:
SYNTHROID 150MCG TABLET Take one(1) tablet daily.
Physical Exam:
Gen: alert , oriented, NAD, obese
HEENT: atraumatic, normocephalic, PERRL, EOMI, sclera clear, conjunctiva pink , fundi crisp, TM's clear, nares normal, moist mucous membranes,
oropharynx clear without erythema or exudate.
Neck: supple, no LAD, no carotid bruits, no JVD, GIOTER
CV: S1, S2 regular, 3/6 SEM at **NAME[ZZZ] SB
Resp: Clear to auscultation bilaterally, anterior and posterior, no wheezes, rales or rhonchi
Abd : +BS, soft, nontender, OBESE, no hepatosplenomegaly
Ext: no cyanosis, no clubbing, no edema, warm a
Pulses: radial 2/2, brachial 2/2, axillary 2/2, femo /2, posterior tibial 2/2
Skin: no rashes
Neuro: nonfocal
244.9 HYPOTHYROIDISM NOS
Note : pt is subtherapeutic on sythroid, will increase to 175 mcg. follow up TSH in 3 months.
Plan: SYNTH ROID 175MCG TABLET

250.0 DIABETES MELLITUS UNCOMP
Note: Newly diagnosed DM type 2. Will sta rt glucophage.
Plan: Glucophage 500mg po bid
746.89 VSD
Note: TTE- preliminary results show: normal LV/RV function, mild RVH, small residual VSD with left to right flow, RVOT with subvalvular and valv
ular PS, trivial TR, trivial **NAME[YYY,
XXX] has not seen peds cards since 1996. although well should f/u.
Pl an: CONSULT TO CARDIOLOGY - f/u with Dr. **NAME[WWW] [peds cards]

272.4 HYPERLIPIDEMIA **INITIALS
Not e: LDL 109, should start statin, but pt wants to try diet and excersize. will f/u in 3 months and re check lipid profile.
Plan: lipid profile in 3 months.
**NAME[VVV UUU], MD

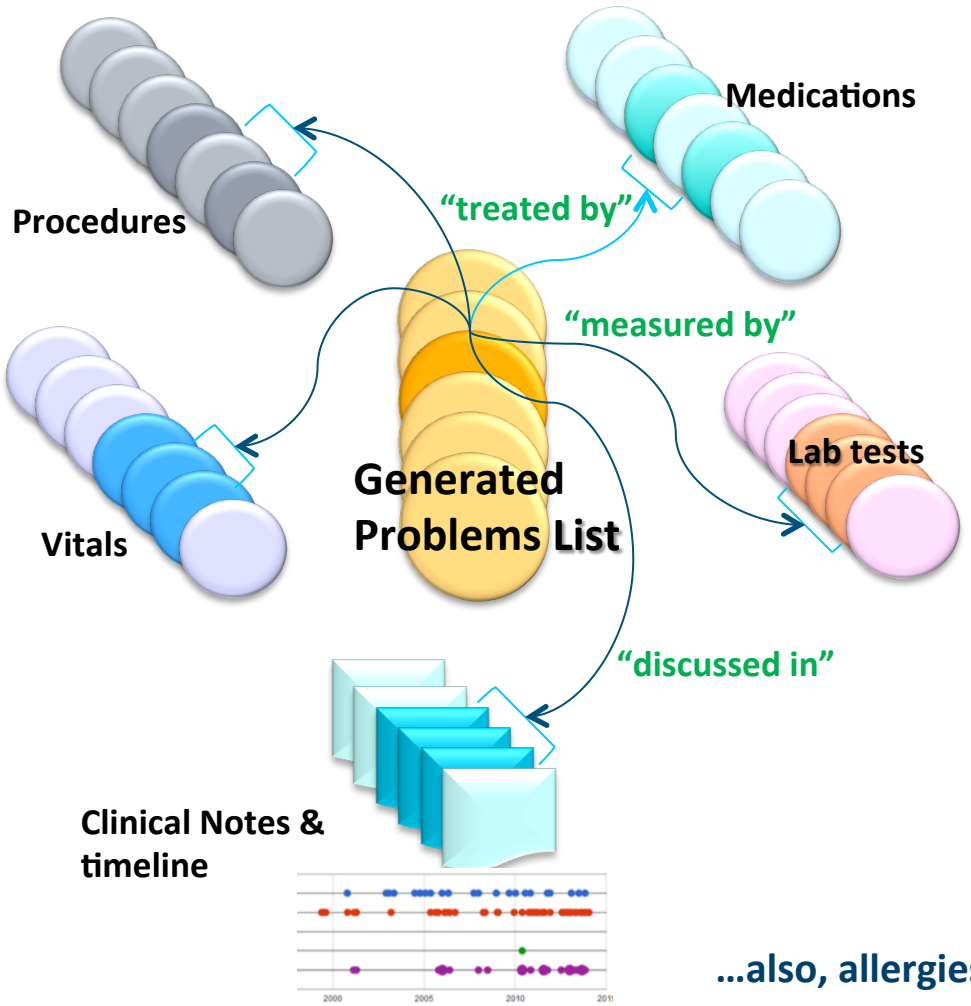
Examined patient and confirmed key findings on history and examination and **NAME[SSS M] as noted by Dr. **NAME[TTT UUU] above and addended by me [MY COMMENTS].
Summary of my findings and i mpressions are as follows:
DM - is well controlled with diet alone. He has been losing weight and has lost about 15-20 lb over the last 2 years per the patient. by our EMR he

"disease or syndrome"
CUI = C0011849

"sign or symptom"
CUI = C0014743

© 2015 International Business Machines Corporation

Problem-Oriented Patient Record Summary



Goals

- Generate medical problems list automatically
- Relate medications, labs, procedures, and clinical notes to medical problems
- Organize lists in a clinical order
- Enable one/two click access to raw data such as Notes, labs over a time line, medication history,…

…also, allergies, social history, and demography

© 2015 International Business Machines Corporation

Problem List Challenge

- Medical problem list is a patient's diagnosed diseases and significant not-yet diagnosed symptoms that require care and management
- Maintaining an accurate PL is challenge because it requires:
  - Broad and high level of medical expertise
  - Significant time
- Our assessment of entered problem list based on a gold standard indicates the challenge:

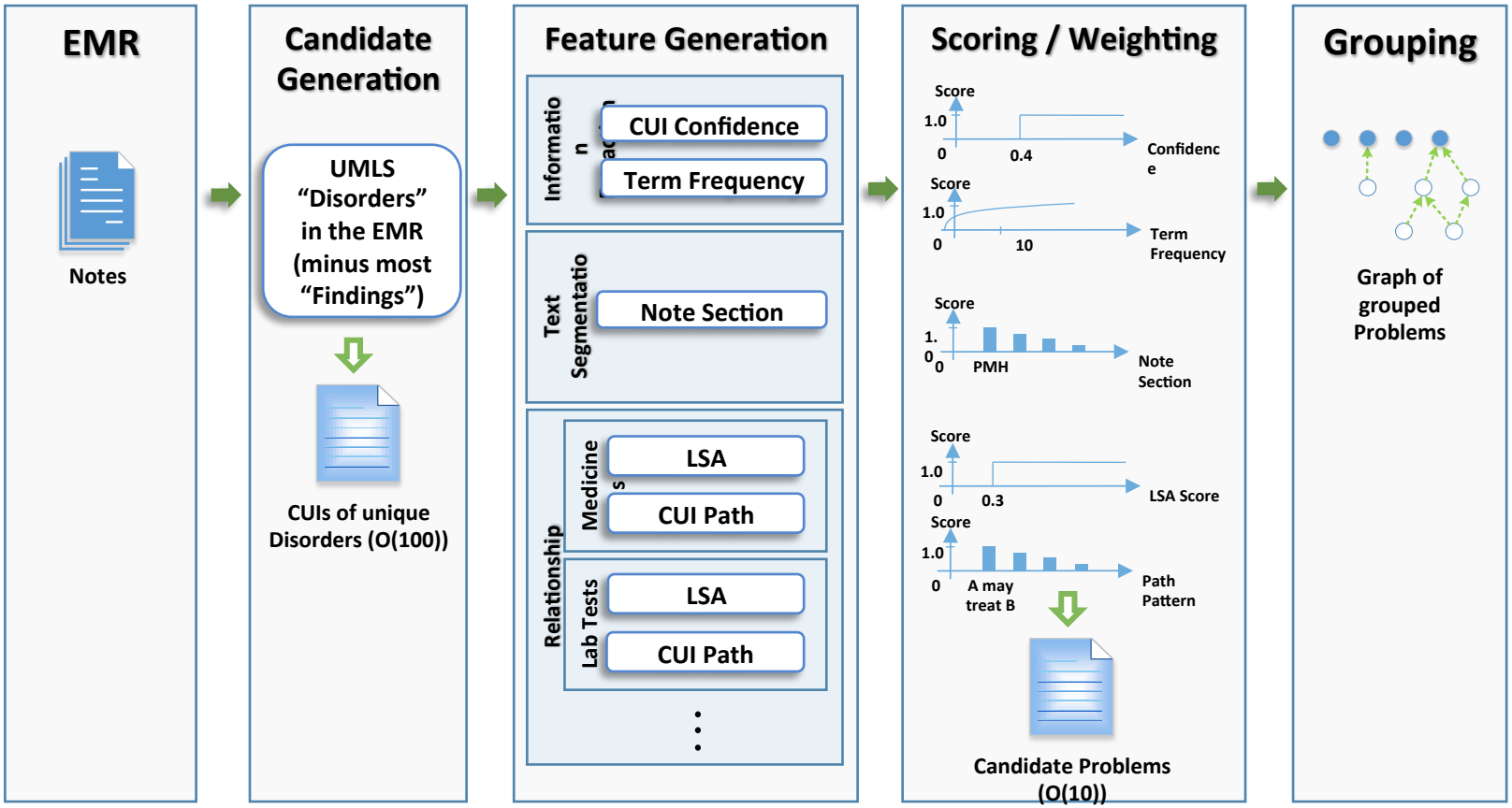
Entered Problem List Accuracy

Recall (Sensitivity) = 0.55

Precision (Positive Predictive Value) = 0.28

© 2015 International Business Machines Corporation

EMRA Problem List Generation



Entered Problem List Accuracy
Recall (Sensitivity) = 0.55
Precision (Positive Predictive Value) = 0.28

EMRA Problem List Accuracy:
Recall (Sensitivity) = 0.84
Precision (Positive Predictive Rate) = 0.53

© 2015 International Business Machines Corporation

Physician opens the EMR and sees a summary of the record…

Automatically generated problems grouped by clinical relevance

Medications grouped by clinical relevance

Automatically categorized encounters

EMR: …D19_patientID4068 (70 year old Male)

Problems

Name	Date
metabolic syndrome x	10/12/2004
diabetes mellitus	12/04/2002
dyslipidemias	01/29/2003
obesity	01/29/2003
microalbuminuria	10/12/2004
hypothyroidism	12/04/2002
thyroid diseases	05/11/2005
sleep apnea syndromes	07/31/2012
allergic rhinitis (disorder)	07/31/2012

Medications

Name	Prescribed
lisinopril	01/29/2013
glipizide	07/25/2007
metformin	06/29/2004
freestyle lancets	05/11/2005
freestyle system kit	05/11/2005
lipase-protease-amylase	02/17/2010
lactated ringers iv	05/22/2010
cyclobenzaprine	10/30/2010
fluocinonide	04/12/2013
clindamycin	10/30/2013
valacyclovir	07/26/2011

…r Encounter Instance Timeline

Wed Oct/30/2013
One note.

Primary Care
Emergency Medicine
Specialties
Nursing
Other

Procedures

Name	Order Date
CERVICAL AP/LAT + OBL/A21	06/27/1997
CHEST PA / LATERAL	11/10/1999
CHEST PA AND LATERAL	03/12/2003
COLONOSCOPY W/BX	02/17/2006
COLONOSCOPY, GI	03/22/2006
COMPLETE ECG	02/16/2011
CT ORBITS WWO CONTRAST	08/13/2013
DILATED FUNDUS EXAM	01/14/2014
DOPPLER ECHO HEART,COMPLETE	12/04/2002
ECG COMPLETE W	01/18/2011

Laboratory Test Values

Name	Value	Date
albumin/creat rnd ur		
creatinine, ur random (ucrr)	16.4	08/02/2013
albumin random urine		
albumin, urine random	11.0	08/02/2013
basic metabolic pnl		
chloride	98	08/02/2013
glucose	151	08/02/2013
bun	17	08/02/2013
anion gap	11	08/02/2013
calcium	10.0	08/02/2013
co2	28	08/02/2013
creatinine	0.79	08/02/2013

Vitals

Name	Value	Date
BLOOD PRESSURE	95/68	08/06/2013
HEIGHT (cms)	172	07/26/2011
PULSE	89	08/06/2013
PULSE OXIMETRY	97	02/16/2011
TEMPERATURE	97.7	06/28/2013
WEIGHT (kgs)	81.33	08/06/2013

Social History

Alchohol	Drug	Tobacco	Date
Not recorded	Not recorded	Not recorded	12/04/2…
Not recorded	Not recorded	Not recorded	12/04/2…

Allergies

Name	Date
Erythromycin	10/17/2000

IBM Confidential

Version 2014-03-20

© 2015 International Business Machines Corporation

EMRA Usage

A closer look at the left side of the screenshot…

Problems

Name	Date
metabolic syndrome x	10/12/2004
diabetes mellitus	12/04/2002
dyslipidemias	01/29/2003
obesity	01/29/2003
microalbuminuria	10/12/2004
hypothyroidism	12/04/2002
thyroid diseases	05/11/2005
sleep apnea syndromes	07/31/2012
allergic rhinitis (disorder)	07/31/2012

Medications

Name	Prescribed
lisinopril	01/29/2013
glipizide	07/25/2007
metformin	06/29/2004
freestyle lancets	05/11/2005
freestyle system kit	05/11/2005
lipase-protease-amylase	02/17/2010
lactated ringers iv	05/22/2010
cyclobenzaprine	10/30/2010
fluocinonide	04/12/2013
clindamycin	10/30/2013
valacyclovir	07/26/2011

Procedures

Name	Order Date
CERVICAL AP/LAT + OBL/A21	06/27/1997
CHEST PA / LATERAL	11/10/1999
CHEST PA AND LATERAL	03/12/2003
COLONOSCOPY W/BX	02/17/2006
COLONOSCOPY, GI	03/22/2006
COMPLETE ECG	02/16/2011
CT ORBITS WWO CONTRAST	08/13/2013
DILATED FUNDUS EXAM	01/14/2014
DOPPLER ECHO HEART,COMPLETE	12/04/2002
ECG COMPLETE W	01/18/2011

Laboratory Test Values

Name	Value	Date
albumin/creat rnd ur		
creatinine, ur random (ucrr)	16.4	08/02/2013
albumin random urine		
albumin, urine random	11.0	08/02/2013
basic metabolic pnl		
chloride	98	08/02/2013
glucose	151	08/02/2013
bun	17	08/02/2013
anion gap	11	08/02/2013
calcium	10.0	08/02/2013
co2	28	08/02/2013
creatinine	0.79	08/02/2013

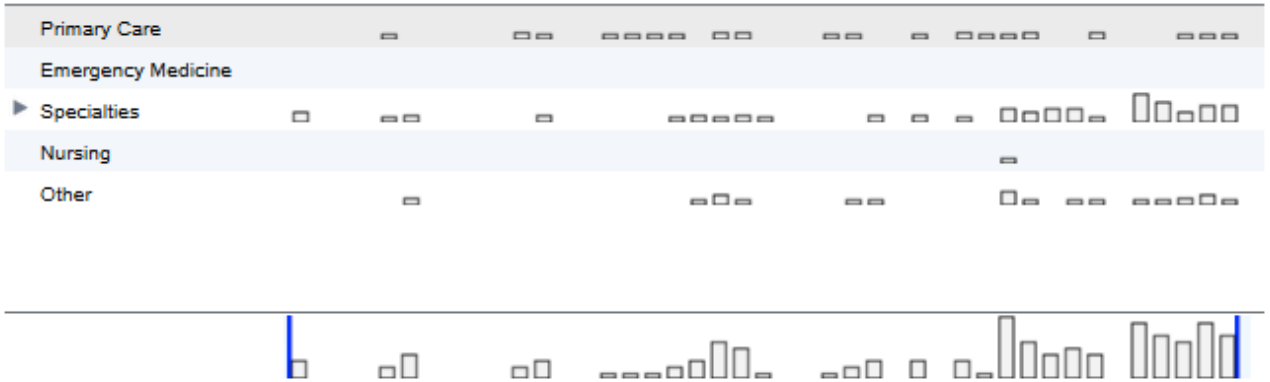
© 2015 International Business Machines Corporation

EMRA Usage

A closer look at the right side of the screenshot…

Clinical Encounter Instance Timeline

Wed Oct/30/2013
One note.



Vitals

Name	Value	Date
BLOOD PRESSURE	95/68	08/06/2013
HEIGHT (cms)	172	07/26/2011
PULSE	89	08/06/2013
PULSE OXIMETRY	97	02/16/2011
TEMPERATURE	97.7	06/28/2013
WEIGHT (kgs)	81.33	08/06/2013

Social History

Alchohol	Drug	Tobacco	Date
Not recorded	Not recorded	Not recorded	12/04/2008
Not recorded	Not recorded	Not recorded	12/04/2008

Allergies

Name	Date
Erythromycin	10/17/2000

© 2015 International Business Machines Corporation

EMRA Usage

As the Physician selects Diabetes Mellitus, screen changes to show related active medications, labs, related Notes,…

When a problem is selected related labs, meds, notes are shown

Current and related meds are highlighted

Relevant labs are highlighted

Labs show elevated glucose and A1C among the others…

Name	Date
esophagitis	07/31/2012
pancreatitis, chronic	05/25/2006
diabetic retinopathy	07/31/2012
diabetes mellitus, non-insulin-dependent	08/23/2005
hyperlipidemia	01/29/2003
obesity	12/04/2002
sinusitis	02/21/2001
allergic rhinitis (disorder)	10/07/2011
cyst	05/25/1999

Medications

Name	Prescribed
glipizide	07/25/2007
metformin	06/29/2004
triamcinolone acetonide	05/14/2013
desonide	09/30/2010
clobetasol	09/30/2010
fluocinonide	04/12/2013
valacyclovir	07/26/2011
clindamycin	10/30/2013
lisinopril	01/29/2013
cyclobenzaprine	10/30/2010
blood sugar diagnostic strips	11/29/2011

Clinical Encounter Instance Timeline

Sat Mar/24/2007 - Wed Sep/11/2013

21 problem notes out of 79 notes.

Primary Care
Emergency Medicine
Specialties
Nursing
Other

Procedures

Name	Order Date
LUMBAR SPINE AP & LAT	10/25/1996
HIP MINIMUM 2 VIEWS	10/25/1996
CHEST PA / LATERAL	11/10/1999
CERVICAL AP/LAT + OBL/A21	06/27/1997
MR UPPER EXTREM W/O CONTRAST	08/31/1999
COMPLETE ECG	02/16/2011
DOPPLER ECHO HEART,COMPLETE	12/04/2002
PEDIATRIC ECHO TYPE 1	02/16/2011
CHEST PA AND LATERAL	
COLONOSCOPY, GI	
C. DIFFICILE	

Laboratory Test Values

Name	Value	Date
basic metabolic pnl		
glucose	151	08/02/201
anion gap	11	08/0 013
hgb a1c		
hemoglobin a1c	6.9	08/03/2013
lipid panel basic		
fasting time		08/02/2013
ldl hdl rati	1.52	08/02/2013
ldl ch	67	08/02/2013
		05/27/2006

Vitals

Name	Value	Date
BLOOD PRESSURE	95/68	08/06/2013
HEIGHT (cms)	172	07/26/2011
PULSE	89	08/06/2013
PULSE OXIMETRY	97	02/16/2011
TEMPERATURE	97.7	06/28/2013
WEIGHT (kgs)	81.33	08/06/2013

Social History

Alchohol	Drug	Tobacco	Date
Not recorded	Not recorded	Not recorded	12/04/2
Not recorded	Not recorded	Not recorded	12/04/2

Allergies

Name	Date
Erythromycin	10/17/2000

© 2015 International Business Machines Corporation

THANK YOU

© 2015 International Business Machines Corporation

3M Science. Applied to Life.™

NLP Research & Application

Richard Wolniewicz, Division Scientist, Health Information Systems Division

September, 2015

1: NLP Adds Value Within a Clinical Workflow

- Application Value Generally Delivered by:
  1. **Eliminating a human workflow** entirely, or
  2. **Increasing human productivity or effectiveness** in a workflow
- **Eliminating Human Workflow**
  - Confidence model is essential
- **Increasing Human Effectiveness**
  - Workflow-specific
  - Explicability of NLP recommendations
  - Reduction in noise (false positives) – is F-measure a good score?

2: N ≥ 100,000

- Real-world **Data** is **Big**
  - Does the result which held at **1,000** samples hold at **1,000,000** samples?
  - Does the algorithm add computational complexity? If so, is it worth it?
- Real-world **Targets** are **Big**
  - There are >100,000 ICD-10 codes. Does an approach which works for 100 codes really work on 100,000?
  - Do precision and recall fool us?

3: Protected Health Information

- PHI is Fundamental to Clinical NLP Research
- Volumes (and thus Risk Exposures) are Large
- Partitioning is Often Difficult
  - Not all algorithms are the same … open research question
- Algorithm Intermediate Results
  - E.g. are word embeddings PHI?

Top Ten features of text in electronic health records

Brian Hazlehurst, PhD

Kaiser Permanente

Center for Health Research

© 2015, KAISER PERMANENTE CENTER FOR HEALTH RESEARCH

Center for Health Research

10. Physicians write cryptically

- Incomplete sentences
- Lack of grammar
- Shortened words and abbreviations
- Brief statements
- Sparse, ideosyncratic, or no use of punctuation

Fam Hx:Fa-aodm, pgf colon ca, mgm bone marrow ca

© 2015, KAISER PERMANENTE CENTER FOR HEALTH RESEARCH

Center for Health Research

Clinical practice generates documentation variation

9. Documentation events in the EHR are created for many purposes

   administrative, billing, legal, patient care

8. There are many different providers that touch the patient, each with their own protocols for care and documentation

   Specialties, trainees, professional/legal divisions of labor

© 2015, KAISER PERMANENTE CENTER FOR HEALTH RESEARCH

Center for Health Research

Clinical discourse is special

7. Traditions exist for writing notes and documenting care (e.g., SOAP structure for progress notes)
   - However, many EHR implementations don't promote such structure to the writer

6. Concepts discussed are complex (eg., disease, differential diagnosis, intentions for action by multiple parties with distinct roles)
   - Requires deep and situated knowledge to understand correctly

© 2015, KAISER PERMANENTE CENTER FOR HEALTH RESEARCH

Center for Health Research

The EHR is a data aggregator (garbage collector?)

5. Patient is often seen by multiple institutions each with their own EHR implementation
   - Just because note writing has a "place" in the workflow of one does not mean these data are easily imported into the database of the other (the same is true across specialties within an institution)

4. Many practices are simply importing paper-based text (e.g., faxes, scans, legal and form letters) into the EHR
   - Creates a great historical "dossier" of contact with the patient, but a messy clinical record

© 2015, KAISER PERMANENTE CENTER FOR HEALTH RESEARCH

Center for Health Research

EHR software generates documentation noise

3. Unused, partially-used, inconsistently-used templates for text entry

2. UI “features” allow point-click addition of content into notes (e.g., adding current medications list to the progress note)

1. Vendors are trying to keep all customers happy, allowing for near-infinite customizations to accelerate/simplify data entry

© 2015, KAISER PERMANENTE CENTER FOR HEALTH RESEARCH

Center for Health Research

NLP Research Knowledge Gap

Scott L. DuVall

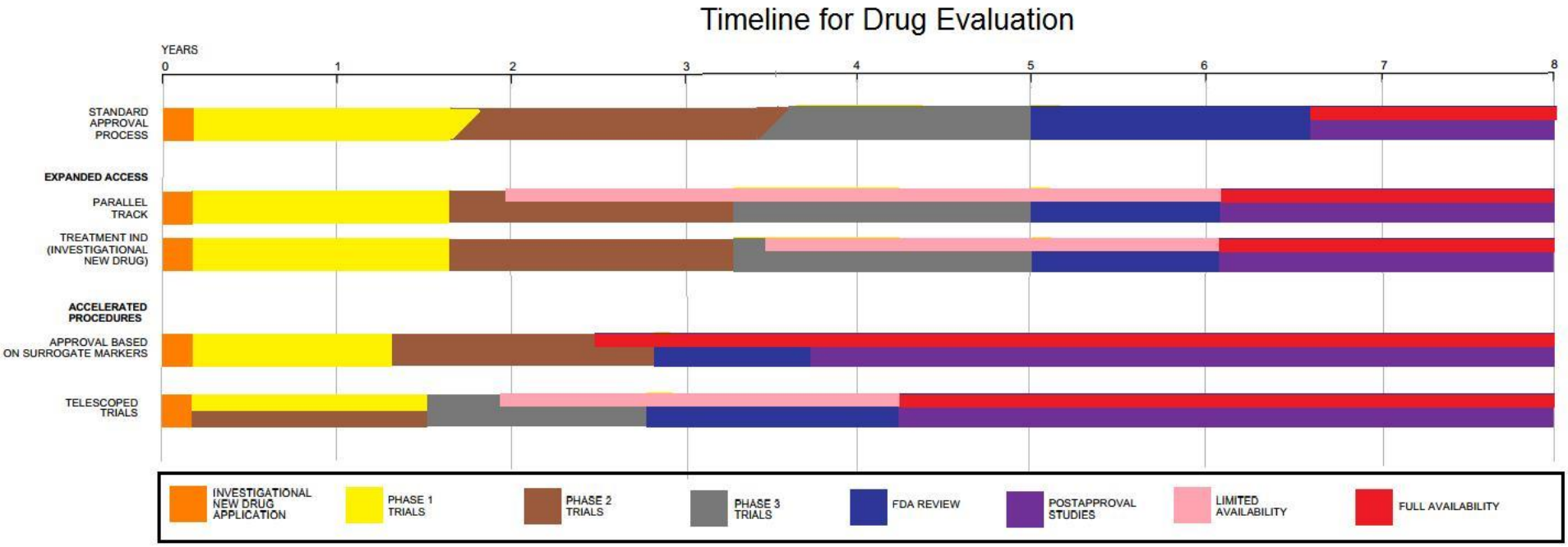
Sep 9 2015

Acknowledgements

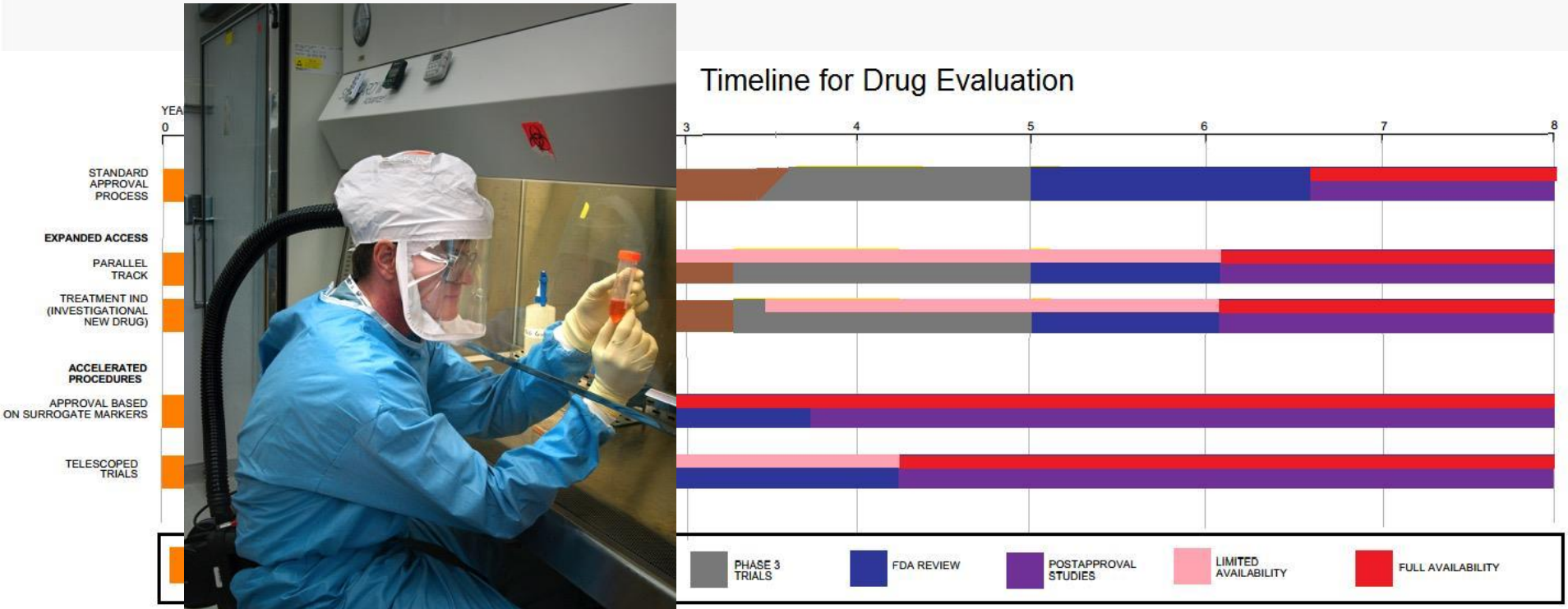
- Resources and Facilities
  - Veterans Affairs Salt Lake City Health Care System
  - Department of Epidemiology, University of Utah
- Funding Support
  - VA Informatics and Computing Infrastructure VA HSR RES 13-457
- Financial Relationships
  - Research Grants from AbbVie Inc, Amgen Inc, Anolinx LLC, AstraZeneca Pharmaceuticals LP, F. Hoffmann-La Roche Ltd, Genentech Inc, Genomic Health, Inc., Merck & Co., Inc., Mylan Specialty LP, PAREXEL International Corporation, and Shire PLC through the University of Utah and Western Institute for Biomedical Research.
  - Federal funding from Centers for Disease Control and Prevention, Department of Defense, Department of Veterans Affairs, Intermountain Healthcare, National Heart, Lung, and Blood Institute, National Institute on Alcohol Abuse and Alcoholism, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institute of General Medical Sciences, National Institute of Standards and Technology, National Library of Medicine, National Science Foundation, and Patient Centered Outcomes Research Institute

Acknowledgements

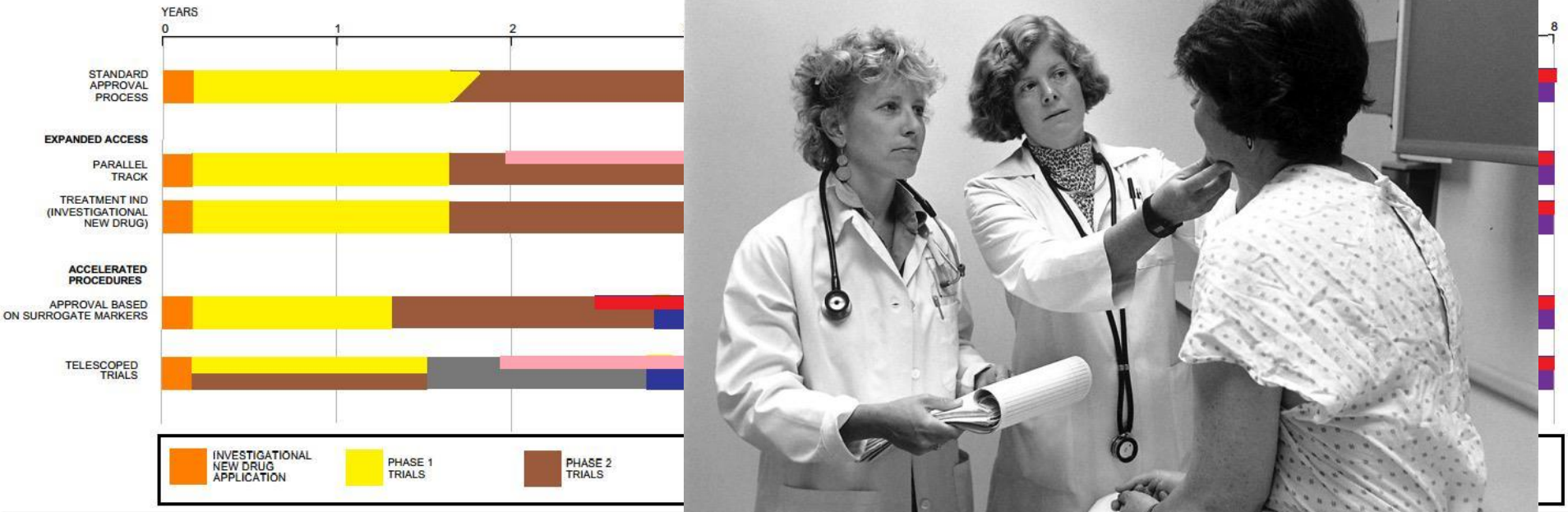
- Jonathan Nebeker
- Olga Patterson
- Patrick Alba
- Lalinda De Silva
- Ryan Cornia
- Brad Adams
- Tom Ginter
- VINCI team



"Drug Evaluation Process" by Kernsters - Graph created based on information provided in Scientific American article, "Faster Evaluation of Vital Drugs". Licensed under CC BY-SA 3.0 via Wikipedia - https://en.wikipedia.org/wiki/File:Drug_Evaluation_Process.jpg#/media/File:Drug_Evaluation_Process.jpg

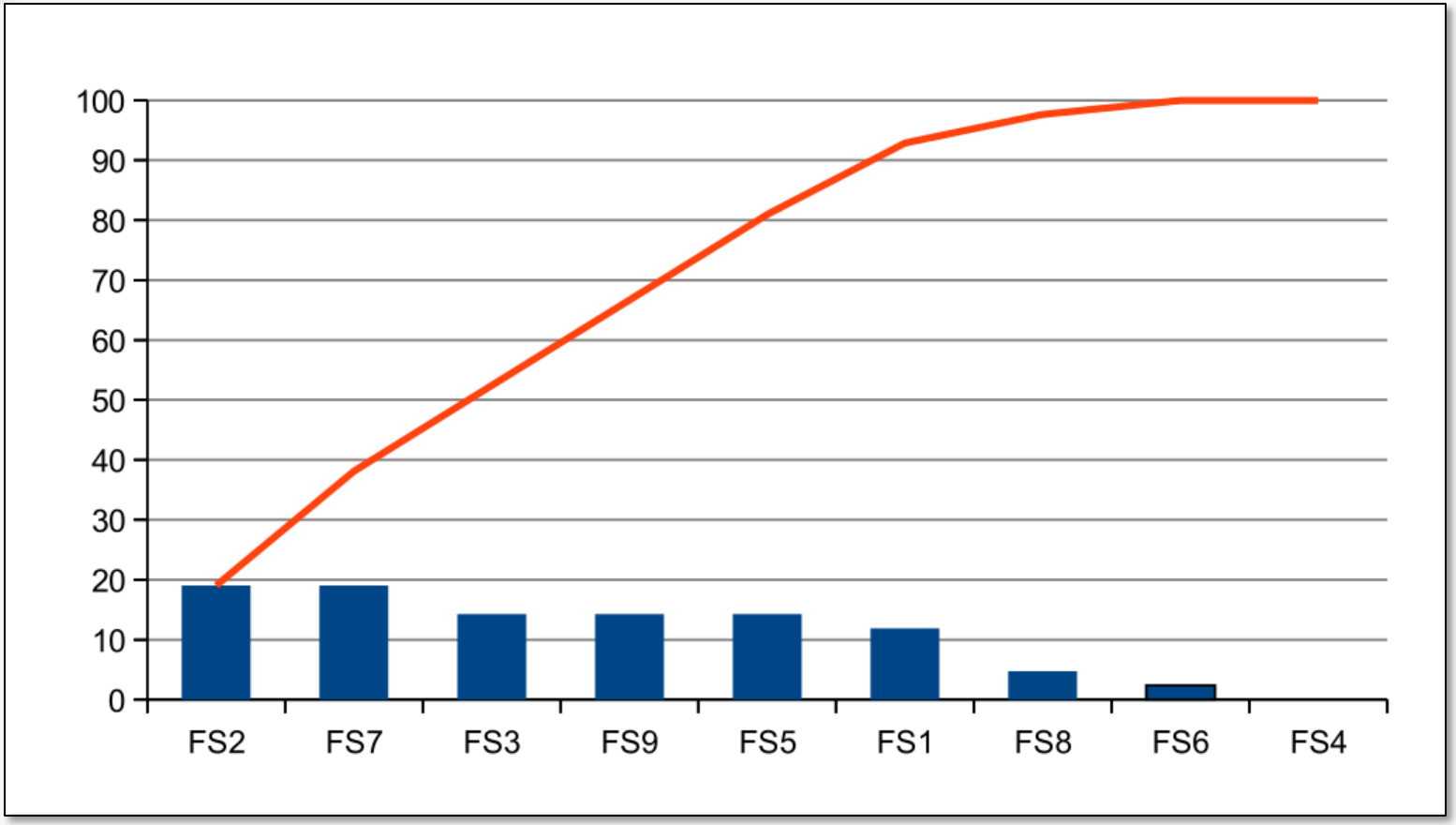


"Influenza virus research" by Photo Credit: James GathanyContent Providers(s): CDC - This media comes from the Centers for Disease Control and Prevention's Public Health Image Library (PHIL), with identification number #7988. Licensed under Public Domain via Commons - https://commons.wikimedia.org/wiki/File:Influenza_virus_research.jpg#/media/File:Influenza_virus_research.jpg



"Seattle physician with patient 1999" by Seattle Municipal Archives from Seattle, WA - Doctors with patient, 1999. Licensed under CC BY 2.0 via Commons - https://commons.wikimedia.org/wiki/File:Seattle_physician_with_patient_1999.jpg#/media/File:Seattle_physician_with_patient_1999.jpg

Pareto Development



By Cdang (Own work) [GFDL (http://www.gnu.org/copyleft/fdl.html) or CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons

Shotgun

- 4 Classes: [25/25 annotations]
  - Anatomy
  - Disorder [2/2]
  - Event [17/17]
  - Finding [6/6]
- 0 Public Attributes:
- Relationships:[0/0]

CC F/U Depression

The patient indicates that his symptoms have improved significantly, but not as much as he expected. He is still sleeping a lot (about 12 hours per day) and finds it hard to concentrate on looking for work. He denies suicidal ideation. His PHQ-9 score is 16 today.

The patient has a history of binge eating episodes. He is an emotional eater and often feels out of control, but continues to eat after job search disappointments. He often binges at night and has done this 3-4 times per week for the past several years.

Rifle

- 6 Classes: [28/38 annotations]
  - Anatomy [6/8]
  - BMD_Concept [3/6]
  - BMD_Measure [5/5]
  - TScore_Concept [3/6]
  - TScore_Measure [7/7]
  - Units [4/6]
  - 0 Public Attributes:
- Relationships:[26/3]
  - hasUnits
  - hasAnatomy
  - hasMeasure

Summary findings for bone mineral density (BMD) are as follows:
L1-L4 BMD (g/cm2) 1.397 +/- 0.010
L1-L4 % Young Adult 113% T = 1.3

Bone mineral density as determined from the left total femur is 0.610 grams per centimeter squared with a T score of -3.1.
---

the AP spine T-score was -2.9 and BMD was 0.922 grams/cm squared.
AP spine T-score is now -2.4 and BMD is 0.843

---
left hip .724 g/cm2, T -1.8. L1-4 T score -2.5, femur total -2.1

Shotgun

- 4 Classes: [12/22 annotations]
  - Anatomy [5/7]
  - Disorder [2/6]
  - Event [4/8]
  - Finding [1/1]
- 0 Public Attributes:
- Relationships:[0/0]

```
Summary findings for bone mineral density (BMD) are as follows:
 L1-L4 BMD (g/cm2)                1.397 +/- 0.010
 L1-L4 % Young Adult                113%          T =  1.3
-------------------

Bone mineral density as determined from the left total femur is 0.610 grams per
centimeter squared with a T score of -3.1.
--------------------

the AP spine T-score was -2.9 and BMD was 0.922 grams/cm squared.
AP spine T-score is now -2.4 and BMD is 0.843

----------------
left hip .724 g/cm2, T -1.8. L1-4 T score -2.5, femur total -2.1
```

Rifle

- 6 Classes: [0/0 annotations]
  - Anatomy
  - BMD_Concept
  - BMD_Measure
  - TScore_Concept
  - TScore_Measure
  - Units
- 0 Public Attributes:
- Relationships:[0/0]

CC F/U Depression

The patient indicates that his symptoms have improved significantly, but not as much as he expected. He is still sleeping a lot (about 12 hours per day) and finds it hard to concentrate on looking for work. He denies suicidal ideation. His PHQ-9 score is 16 today.

The patient has a history of binge eating episodes. He is an emotional eater and often feels out of control, but continues to eat after job search disappointments. He often binges at night

Bridging knowledge gap

- Specialization in early development and translation
- Framework for rapid deployment and processing
- Support of “shotgun” and “rifle” approaches – rapid, modular development

Questions?

Scott.Duvall@va.gov

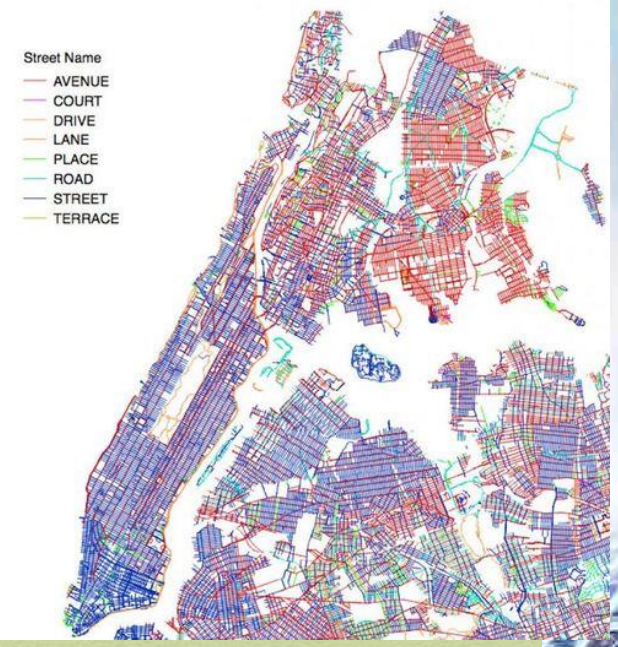
Which Gap?

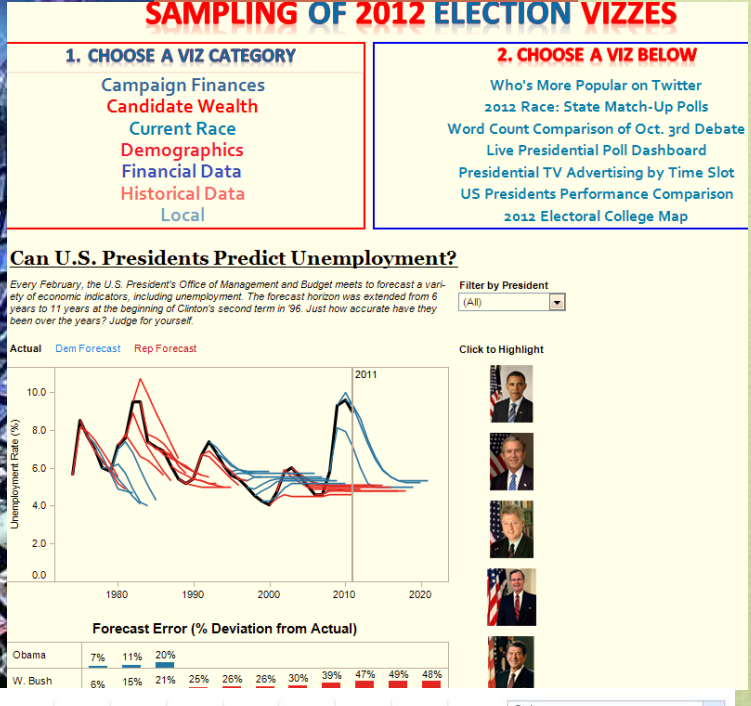
Natural Language Processing in HealthCare

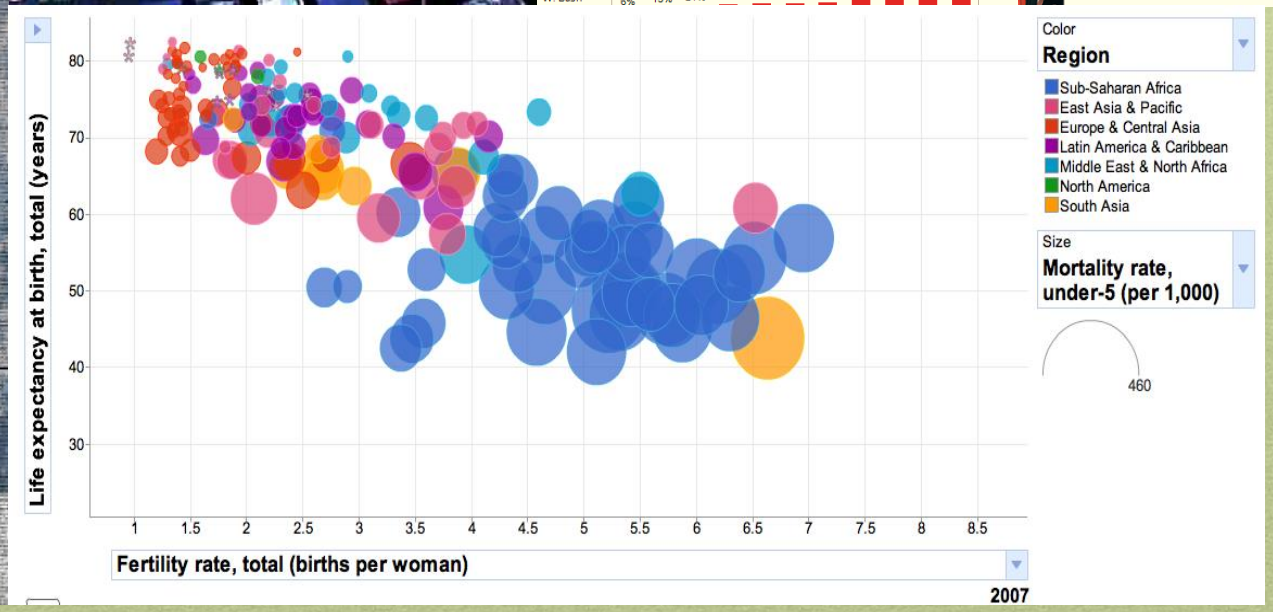
VA | U.S. Department of Veterans Affairs

Ruth M Reeves, PhD

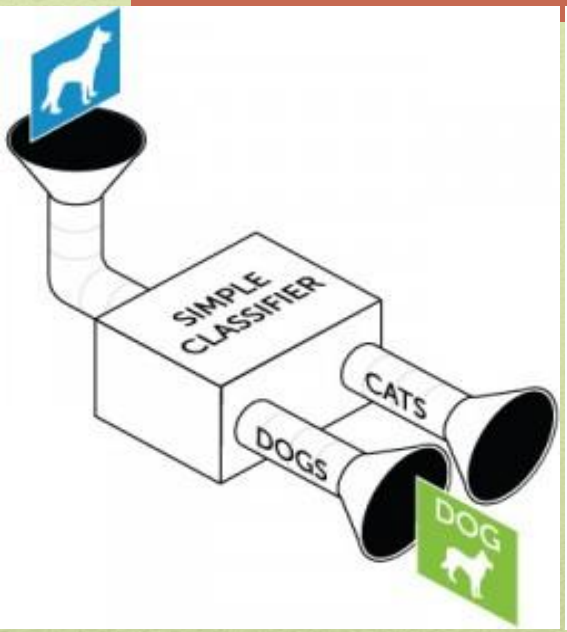
VANDERBILT SCHOOL OF MEDICINE

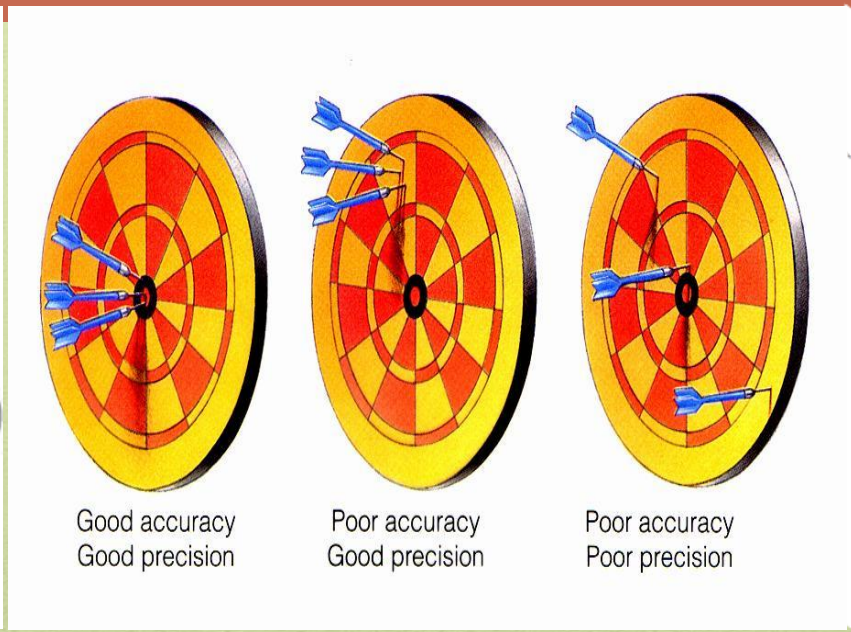


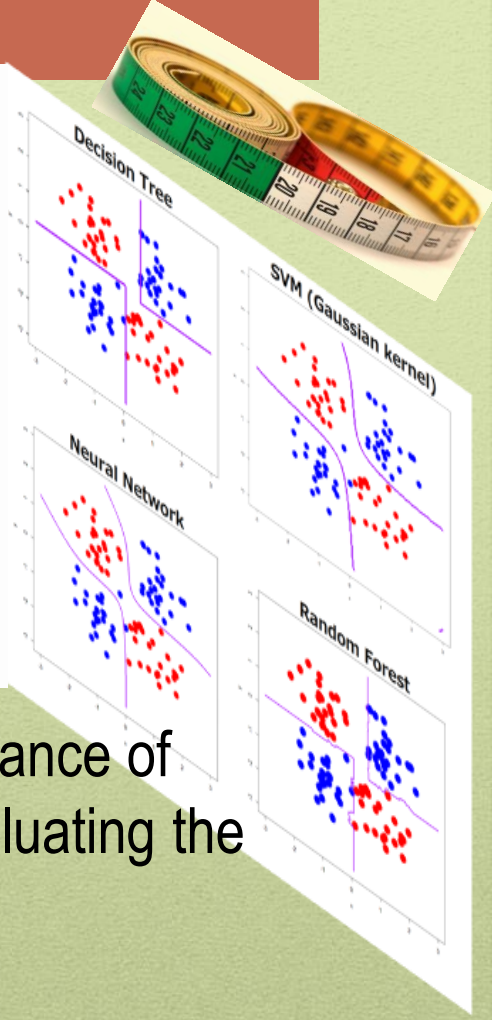




GAP: Shareable Data for Reproducible Results







Evaluation metrics that are good for measuring the performance of information extraction tasks don't necessarily extend to evaluating the contribution of NLP to healthcare related tasks

NLP used for a hypothesis generating task needs to be measured differently from NLP used to discover incidental findings, yet differently again for NLP to identify changes in tumor size over time

GAP: Evaluations that Meet the Information Need

The data model doesn’t necessarily speak to the information model

Modelling Gaps

and just because it has to be said….

The IT cost of classifying clinical information by financial claims data

Cohort definitions based on ICD codes often have to be revalidated against clinical narrative text or external human review.

NLP Supports to the Bridge:

- Event models for typing subparts of trend data
  - **Populate event classes and subclasses with regular ol' information extraction & mapping**
  - **Define semantic & temporal constraints on uniting subparts of clinical events; distinguishing these from unrelated events and their sub-events**
- Semantics of Change & Transition States By Clinical Domain
  - **Align clinical event model to temporal model**
  - **Calibrate probabilities of temporal relations between events**

- Knowledge of temporal relation between events
  - Necessary but not sufficient for causal hypotheses
  - Temporally ordered clusters of symptom events can predict disease progression
- Probabilistic engine for temporal relation assignment
  - Built and under evaluation in Post Traumatic Stress Disorder domain
  - Re-use of infrastructure for tracking pulmonary nodule changes

Temporal Reasoning in Support of Clinical Inferencing

Statistically generated models that underwrite NLP systems are unlikely to be any more useful than butterfly collections for the next use-case down the line, without integration into other existent models and regular upkeep

Infrastructure Maintenance

TAILED JAY
Graphium agamemnon
LEMON EMMIGRANT
Catopsilia Pomona
BLUE BOTTLE
Graphium sarpedon
STRIPED BLUE CROW
Euploca mulciber
RED HELEN
Papilio helenus
BANDED JEZEBEL
Delias descombesi
FIVEBAR SWORDTAIL
Pathysa antiphates
PLAIN TIGER
Danaus chrysippus
MALAYAN ROSE
Pachliopta aristolochiae

THANKS!