Natural Language Processing Supporting Clinical Decision Support

Applications for Enhancing Clinical Decision Making NIH Worksop; Bethesda, MD, April 24, 2012

Stephane M. Meystre, MD, PhD Department of Biomedical Informatics, University of Utah stephane.meystre@hsc.utah.edu



© Stephane Meystre

Fast growing electronic clinical data

Rapid adoption of Electronic Health Records (EHR). Increasing requirements for electronic documentation of healthcare.

Exponentially growing quantities of electronic investigation results information (imaging, genetic testing, etc.)

So what's the problem?

Mostly unstructured information, like **narrative text**, dictated and transcribed or typed in. Most structured information coded with administrative and reimbursement-oriented terminologies.



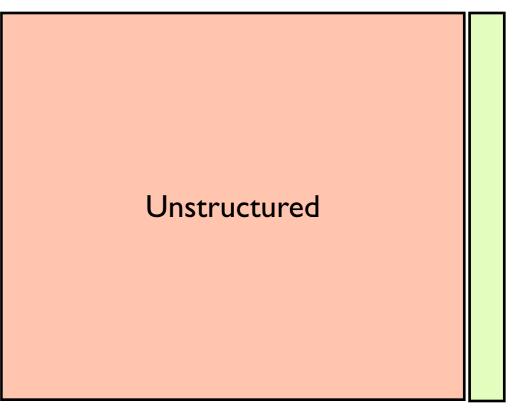
What does the EHR contain?

- Documents
 - History and Physicals
 - Clinical notes, Consult notes
 - Operative reports
 - Surgical pathology reports
 - Progress notes, Letters
 - Orders
 - Discharge summaries
- Imaging / Radiology
- Prescriptions (pharmacy; CPOE)
- Laboratory results
- Administrative information



What does the EHR contain?

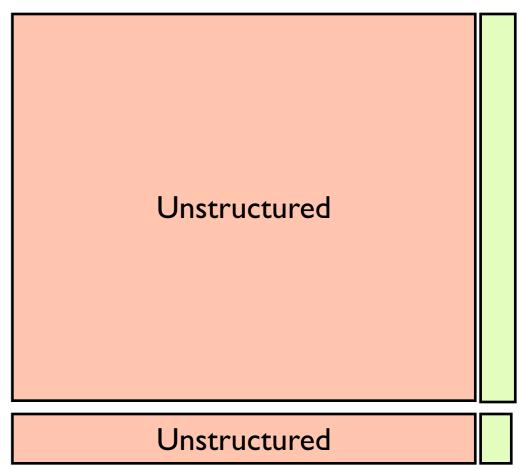
- History and Physicals
- Clinical notes, Consult notes
- Operative reports
- Surgical pathology reports
- Progress notes, Letters
- Orders
- Discharge summaries
- Imaging / Radiology
- Prescriptions (pharmacy; CPOE)
- Laboratory results
- Administrative information





What does the EHR contain?

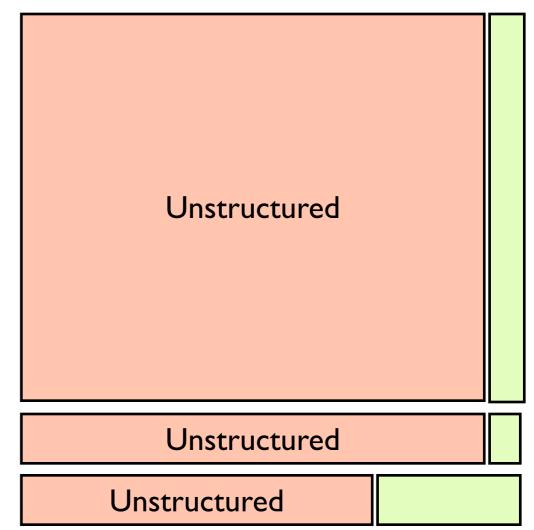
- History and Physicals
- Clinical notes, Consult notes
- Operative reports
- Surgical pathology reports
- Progress notes, Letters
- Orders
- Discharge summaries
- Imaging / Radiology
- Prescriptions (pharmacy; CPOE)
- Laboratory results
- Administrative information





What does the EHR contain?

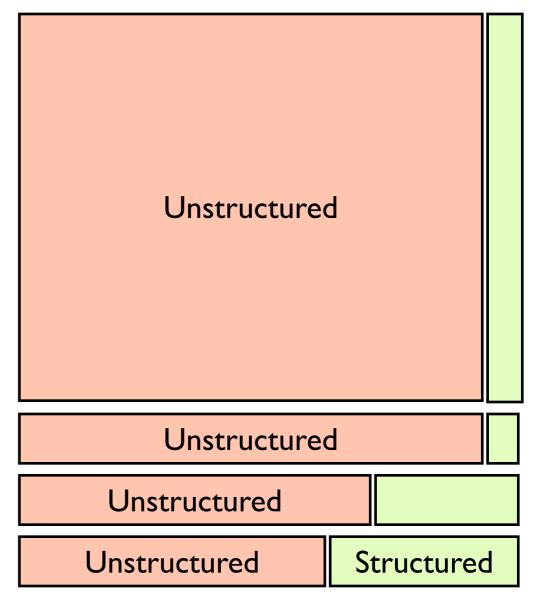
- History and Physicals
- Clinical notes, Consult notes
- Operative reports
- Surgical pathology reports
- Progress notes, Letters
- Orders
- Discharge summaries
- Imaging / Radiology
- Prescriptions (pharmacy; CPOE)
- Laboratory results
- Administrative information





What does the EHR contain?

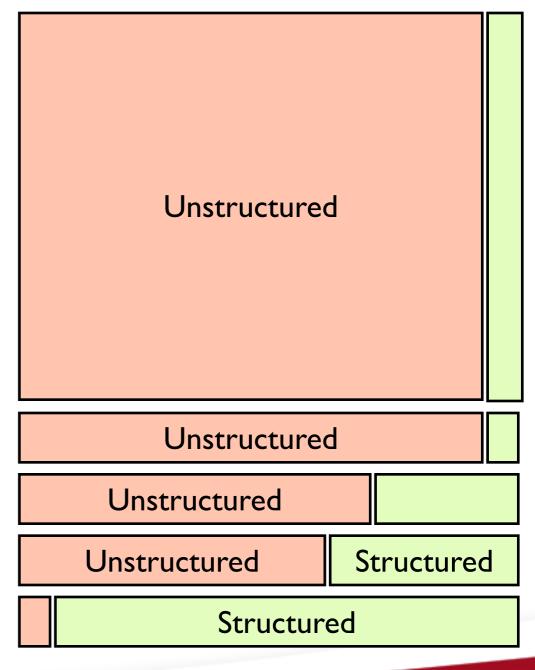
- History and Physicals
- Clinical notes, Consult notes
- Operative reports
- Surgical pathology reports
- Progress notes, Letters
- Orders
- Discharge summaries
- Imaging / Radiology
- Prescriptions (pharmacy; CPOE)
- Laboratory results
- Administrative information





What does the EHR contain?

- History and Physicals
- Clinical notes, Consult notes
- Operative reports
- Surgical pathology reports
- Progress notes, Letters
- Orders
- Discharge summaries
- Imaging / Radiology
- Prescriptions (pharmacy; CPOE)
- Laboratory results
- Administrative information





CDS information needs

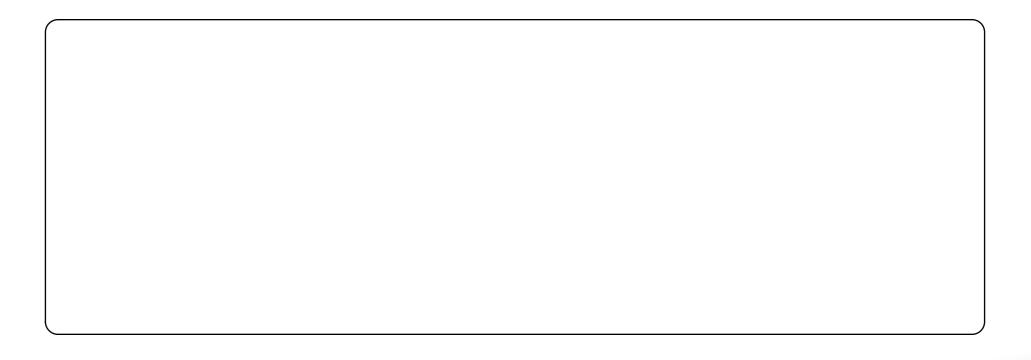
Structured (i.e., using some data model) and **coded** (i.e., labeled with some standard terminology) clinical information

...but most EHR content is narrative text, unstructured, and is therefore inaccessible for CDS.

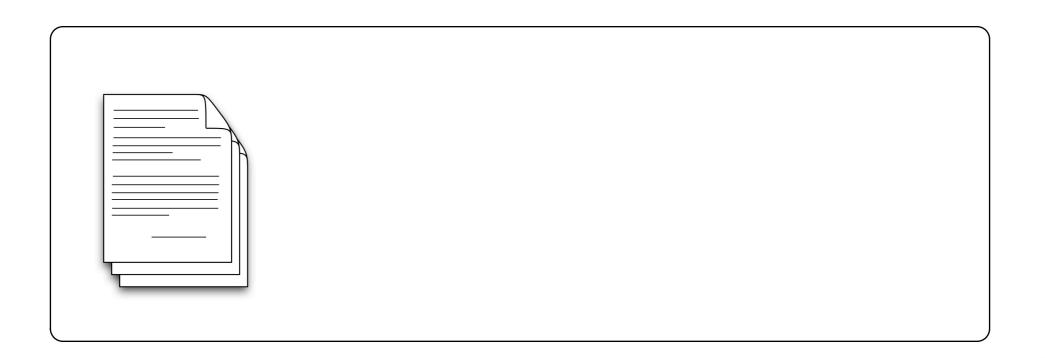
Detailed clinical information, at various levels of granularity

...but most terminologies used were created for mortality and morbidity public health statistics, and for reimbursement (i.e., ICD-9-CM, CPT-4), and don't allow for detailed, clinical care-oriented coding.

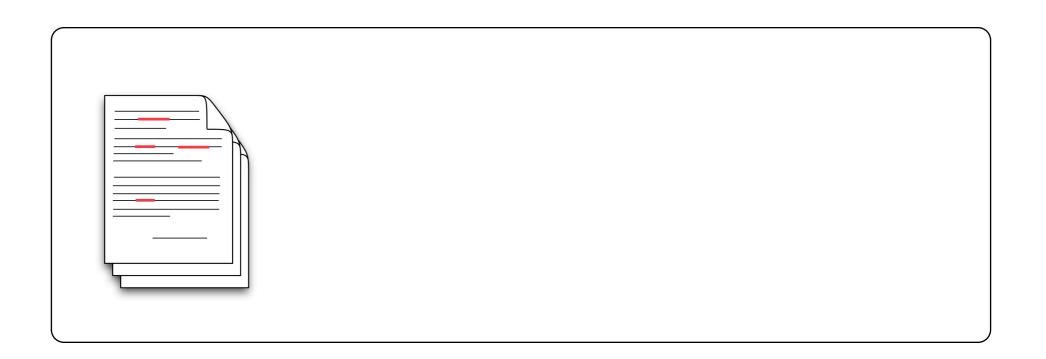




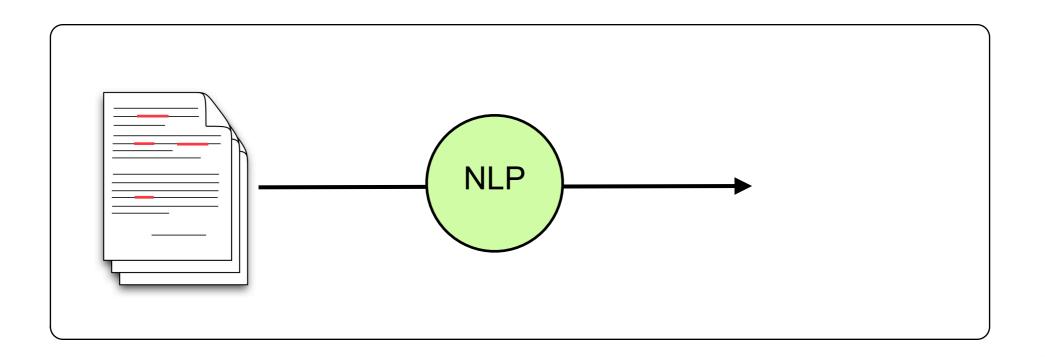




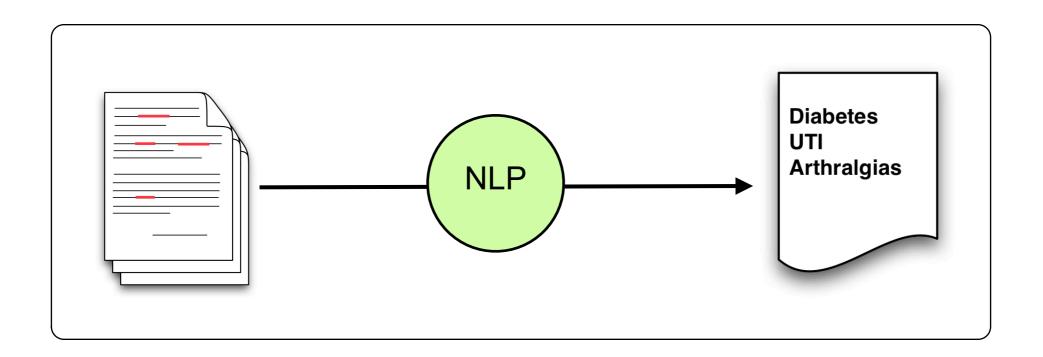




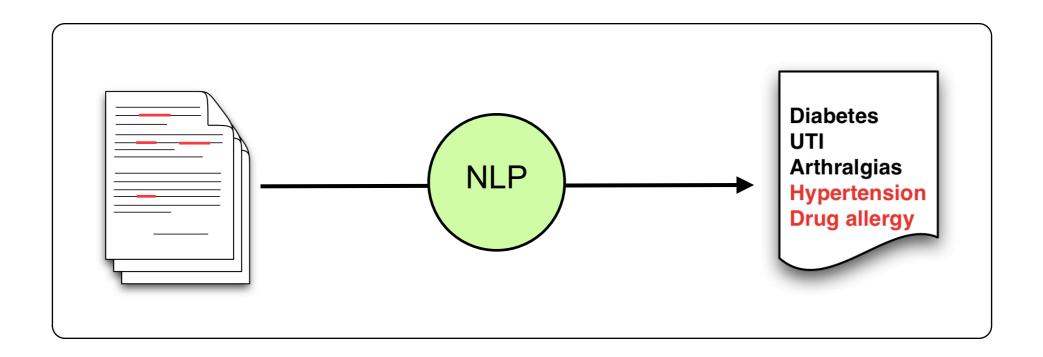














Clinical information extraction

Information Extraction (IE) involves extracting predefined types of information from text (e.g., people, diseases, symptoms, medication doses).

The development of IE applications in the biomedical domain has been far more advanced than in the clinical domain, mostly for two reasons:

-Extremely limited availability of clinical text because of patient confidentiality and privacy reasons.

-Characteristics of clinical text and related difficulties to analyze it.



Clinical Information Extraction

The Automated Problem List

What was the problem?

An electronic problem list was already available at IHC, but

- it was often incomplete, inaccurate, not timely...not used!
- the problem list was becoming a central component of the EHR and was to be used by many applications (CPOE, documentation, knowledge access, etc.).



Clinical Information Extraction

The Automated Problem List

What was the problem?

An electronic problem list was already available at IHC, but

- it was often incomplete, inaccurate, not timely...not used!
- the problem list was becoming a central component of the EHR and was to be used by many applications (CPOE, documentation, knowledge access, etc.).
 - Need for a problem list of good quality (complete, accurate, timely, and coded)



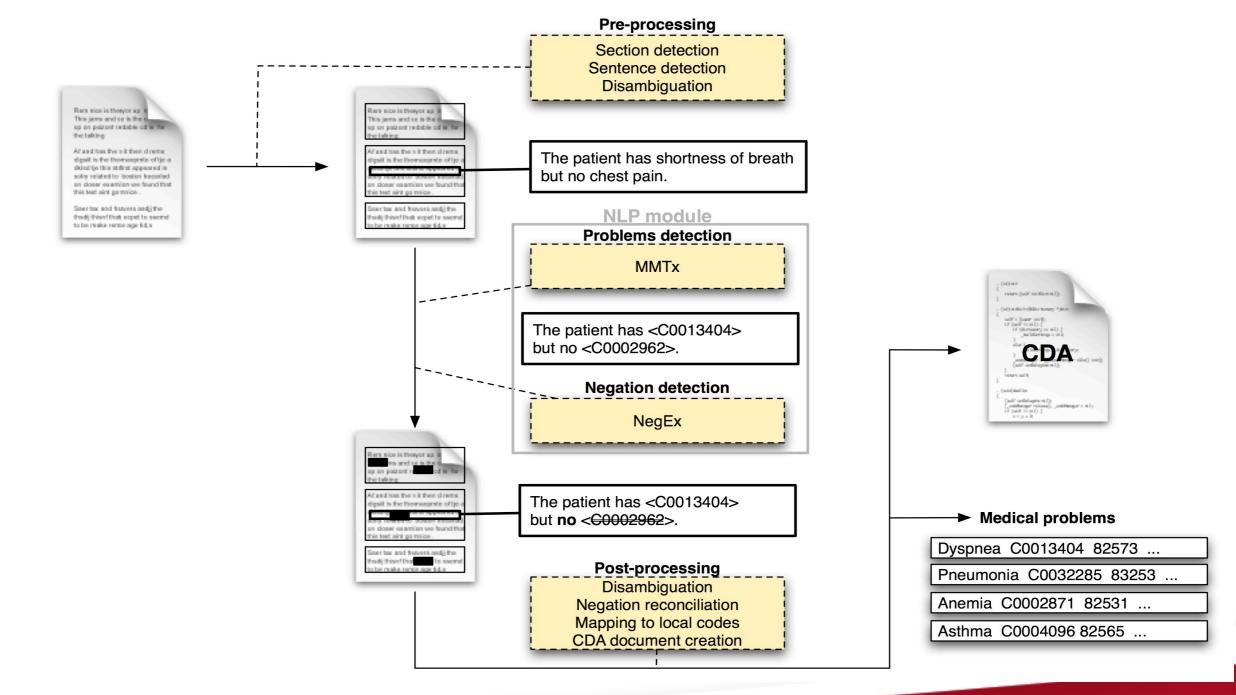
cal desktop	T, ALE	Х		Room: N	1901 Age: 21	Y Sex: M MM	II: 545777393 MRN:				의	ptions
Patient Prob	lems		Filter:	Active An	nd Proposed Pr					Refresh	Prefere	nces j
					olved; E = Er		1					
	s Rvw		Problem	1	Туре	Onset Dt	Noted Dt -	Rvw Dt	Clinician	Body Syste		POC
ogy A.▼ ms		Oyspnea			Diag	ter Brahlana wa	10/29/04	ally as tracted	MEYSTRE, STE	Enc: #0-LDS H		S Hospit
es ations	-	O Deels est	-			Its: Problem wa	as originally automatic 10/29/04	any extracted	MEYSTRE, STE	Enc: #0-LDS I		S Hospit
igns		Back pai	1		Diag	te: Brahlam wa	as originally automatic:	ally extracted		Enc: #0-LDS H		5 HUSPIL
Weight	-	 Arrhythm 	1			Its: Frubierri wa	10/29/04	any extracted	MEYSTRE, STE	Enc: HO-LDS I		S Hospit
raphics A -		• Armyunn	la		Diag	ter Decklass		alle au drug a band		En 401.001		5 Huspit
nce					Commer	nts: Problem wa	as originally automatic	ally extracted	from text	Enc: #0-LDS H	Hospital	
Tandem tt tion ∀iew tids												
t tion ∀iew												

OF UTAH

SCHOOL OF MEDICINE

 $\boldsymbol{\Theta}$

Information Extraction application components





Statu	olems	EX Room: N Filter: Active A	nd Proposed Prc 💌	1				Refresh	Options Logout Preferences Print
		= Active; I = Inactive; R = Res			ad:				
	is Rvw		Type	Onset Dt	Noted Dt 🗸	Rvw Dt	Clinician	Body Syst	em POC
A.		Oyspnea	Diag		10/29/04		MEYSTRE, STE		LDS Hospit
			Comments	Problem was	s originally automatic	ally extracted	from text	Enc: #0-LDS	Hospital
A.		Back pain	Diag		10/29/04		MEYSTRE, STE		LDS Hospit
			Comments	: Problem was	s originally automatic	ally extracted	from text	Enc: #0-LDS	Hospital
A -		Arrhythmia	Diag		10/29/04		MEYSTRE, STE		LDS Hospit
			Comments	: Problem was	s originally automatic	ally extracted	from text	Enc: #0-LDS	Hospital
P		Pulmonary embolus	Diag		10/29/04		MEYSTRE, STE		LDS Hospit
ís		source	Comments	: Problem was	s originally automatic	ally extracted	from text	Enc: #0-LDS	Hospital
P		Deep vein thrombosis	Diag		10/29/04		MEYSTRE, STE		LDS Hospit
		(source)	Comments	: Problem was	s originally automatic	ally extracted	from text	Enc: #0-LDS	Hospital
P 💌		Pulmonary Edema	Diag		10/29/04		MEYSTRE, STE		LDS Hospit
v		(source)	Comments	: Problem was	s originally automatic	ally extracted	from text	Enc: #0-LDS	Hospital
P-		Heart failure	Diag		10/29/04		MEYSTRE, STE		LDS Hospit
		source	Comments	: Problem was	s originally automatic	ally extracted	from text	Enc: #0-LDS	Hospital
P 💌		Heart Block	Diag		10/29/04		MEYSTRE, STE		LDS Hospit
		(source)	Comments	: Problem was	s originally automatic	ally extracted	from text	Enc: #0-LDS	Hospital
Pv		Aortic valve insufficiency	Diag		10/29/04		MEYSTRE, STE		LDS Hospit
		source	Comments	: Problem was	s originally automatic	ally extracted	from text	Enc: #0-LDS	Hospital
P 💌		Mitral insufficiency	Diag		10/29/04		MEYSTRE, STE		LDS Hospit
		(source)	Comments	: Problem was	s originally automatic	ally extracted		Enc: #0-LDS	
P 💌		Mitral stenosis	Diag		10/29/04		MEYSTRE, STE		LDS Hospit
		source	Comments	: Problem was	s originally automatic	ally extracted	from text	Enc: #0-LDS	Hospital

IELP2 clinical desktop	TEST	, ALE	X	Room: N901 A	ge:21	Y Sex:M MMI	:545777393 MRN:						Options	Logou
	Proble	ems	Filter:	Active And Propo	sed Pro	-					<u>R</u> efresh	Prefe	rences	Print
elect Patient ab	Status k	A	Active; I = Inactive;	D = Decelued:	E = E r	Ar D = Dranas	a de							
licro linical Notes	Status				c = cri [ype	Onset Dt	Noted Dt 🗸	Ryw Dt		Clinician	Body Sys	tem	POC	
adiology	A -		Oyspnea	Dia			10/29/04		MEYST	RE, STE	,-,-		DS Hosp	
roblems Ilergies	_			Ce	ommen	ts: Problem was	originally automatic	ally extracted fr	om text	t	Enc: #0-LDS	Hospit	al	
edications	A -		Back pain	Source Doct	ument	Web Page Dia	aloa		×	, STE		L	DS Hosp	it
tal Signs eight/Weight	_							-	_		Enc: #0-LDS	Hospit	al	
emographics	Α.		 Arrhythmia 			Source Do	cument		<u>C</u> lose	, STE		L	DS Hosp	it
CG surance	_			HOSPITAL CO			a and had a paturatio	n of 70% Ho	-		Enc: #0-LDS	Hospit	al	
essage Log	P-		Pulmonary emboli				c and had a saturation nospital where he wa			, STE		L	DS Hosp	it
b Order Entry	_		source	respiratory dist	ress . H	His INR was 1.7	at that time . During h	his hospital stay			Enc: #0-LDS	Hospit	al	
patient Reports ert Review	P -		Deep vein thromb				art failure due to his 3 evidence of pulmon		5	, STE		L	DS Hosp	it
/eb Forms	_		source				al days down to 2 to		s		Enc: #0-LDS	Hospit	al	
ELP/Tandem rders	P		Pulmonary Edema				was 1.5 and the go		se	, STE		L	DS Hosp	it
otText	_		source				of significant weight, pulation but concern				Enc: #0-LDS	Hospit	al	
opulation View RT	P		Heart failure	hypoxic . He was on adequate anticoagulation but concern was raised that he might still have a pulmonary embolus, so he was sent to PE lab, and was found to have chronic DVTs in both legs. He was assumed to be having pulmonary embolus because of this and ultimately I got a spiral CT scan					, STE	LDS Hosp			it	
eb Kids			source							Enc: #0-LDS	Enc: #0-LDS Hospital			
AP Protocol rotocols	P		Heart Block				nbolus His . Cournat		d	, STE		L	DS Hosp	it
0100010			source	with vitamin K a	and his	heparin was cor	ntinued with a goal of	stopping it and			Enc: #0-LDS	Hospit	al	
Resources	P		Aortic valve insuffi				e had increased oxy possible to dischar			, STE	LDS Hospit		it	
nline Tutorials assword	_		source	50 11 14435 111000	IT IL IL VVC	is not going to be	possible to discribit	ge min to nome	40.4C		Enc: #0-LDS	Hospit	al	
D Info	P•		Mitral insufficiency	DISPOSITION						, STE		L	DS Hosp	it
			source				his son and daughte im to St. John's Hos				Enc: #0-LDS	Hospit	al	
omments	P		Mitral stenosis	rehabilitation	Here II			pitarior		, STE		L	DS Hosp	it
	_		source	DISCUADOS		ATIONIC					Enc: #0-LDS	Hospit	al	
		-		One hahy asni			6 liters by nasal can	nula Hewast						
							ischarge from the nu							
				DISCHARGE I	DIACN	OCEC				-				
							vein thrombosis .3.	Inferior vena c	ava					
							5. Sleep apnea							
									_					
	-													
	Save	Canc	el Add New Problem											

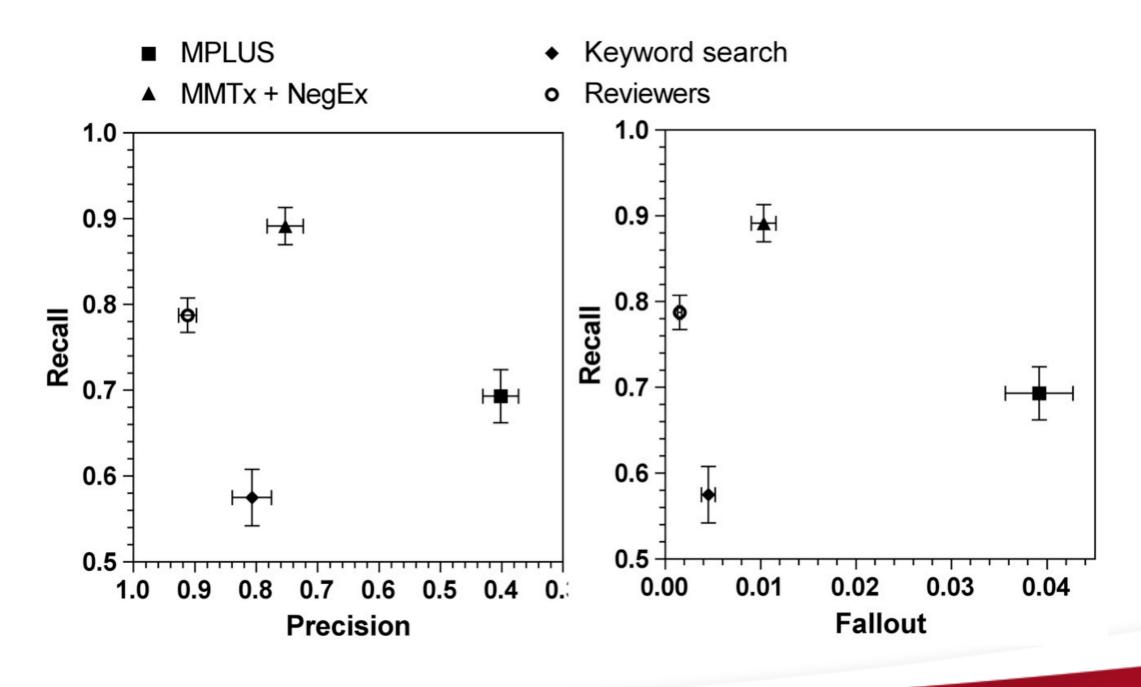
AΗ

NE

SCHOOL OF MEDICI

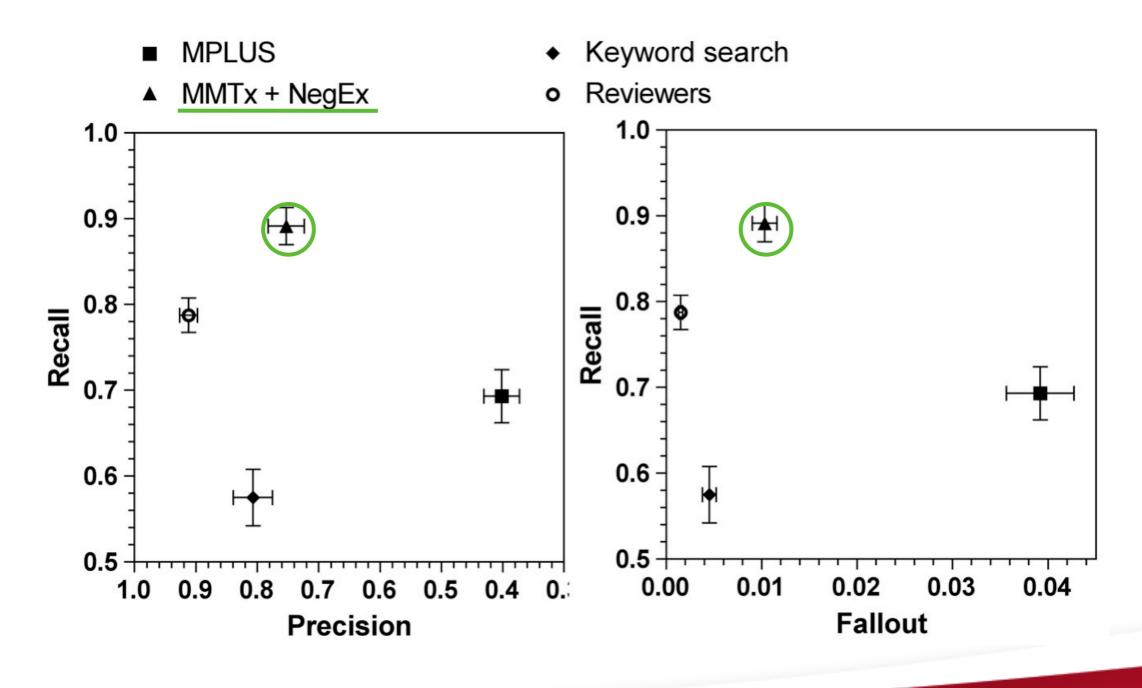
 \odot

Information Extraction application versions comparison



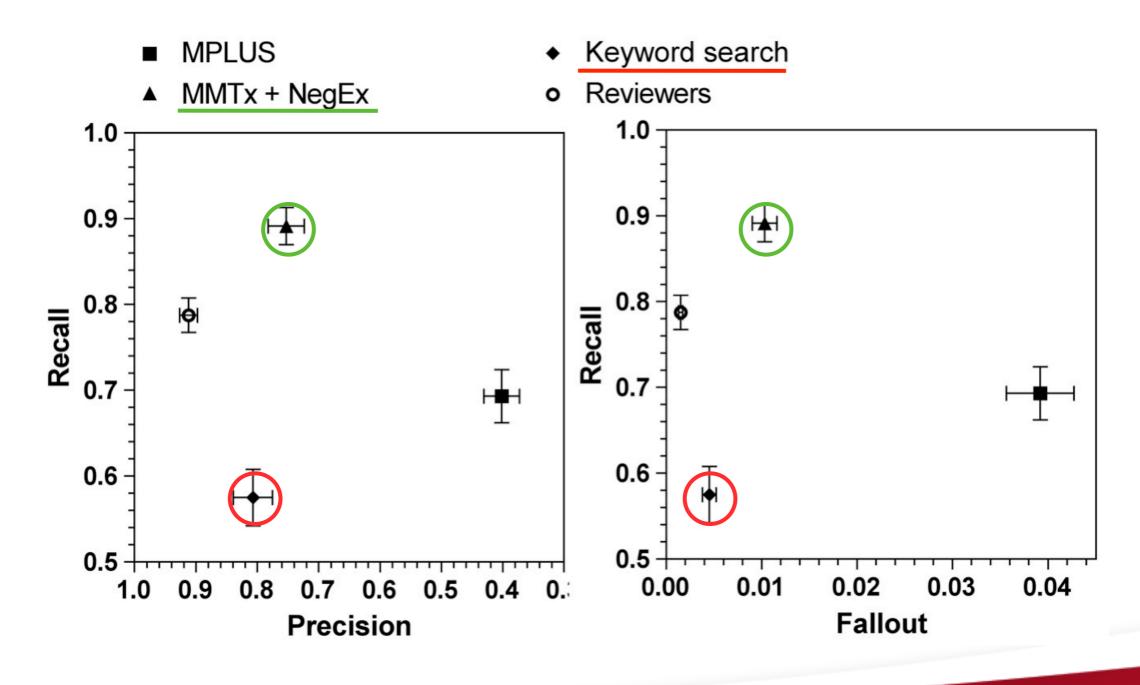


Information Extraction application versions comparison





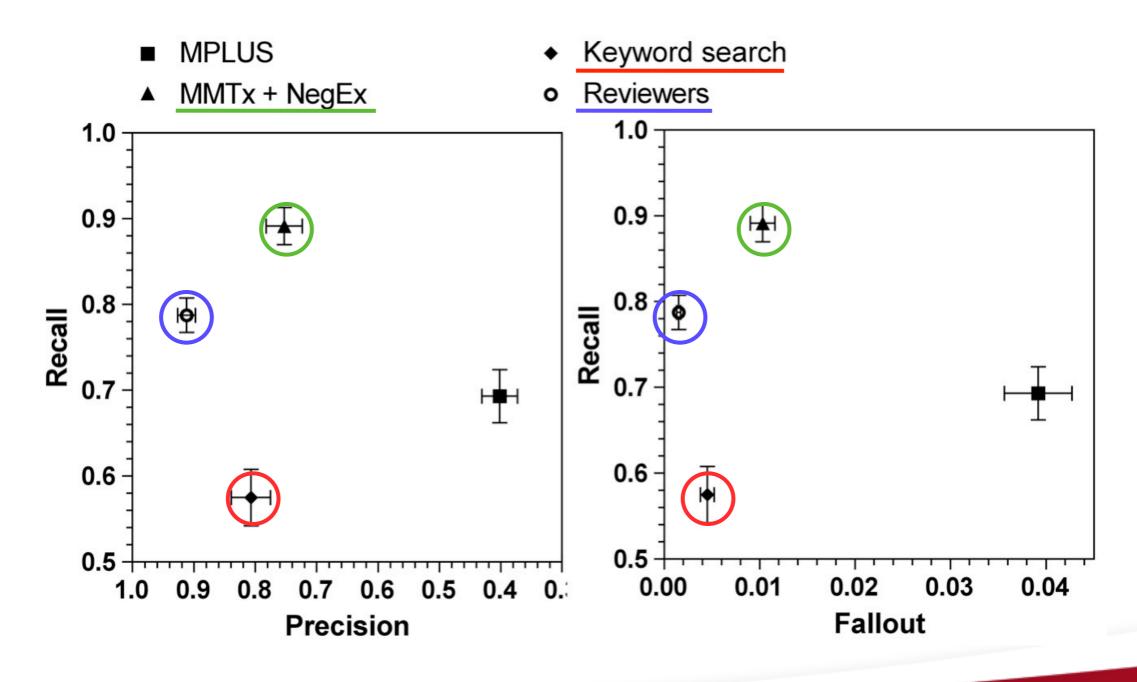
Information Extraction application versions comparison





SCHOOL OF MEDIC

Information Extraction application versions comparison





Study population

- Adult inpatients in the ICU and the Cardiovascular Surgery unit (LDS Hospital, Salt Lake City, UT)
- Control group: physicians had access to the standard electronic problem list
- Intervention group: physicians had access to the APL (i.e. with proposed problems)

	All patients	ICU patients	CVS patients
Initial controls	76	44	32
RCT:Tests	88	54	34
RCT: Controls	83	51	32
TOTAL	247	149	98



APL system RCT results

		Sensitivity	Specificity
	Controls	0.102 (0.069-0.135)	0.998 (0.995-1)
All patients	Tests	0.266 (0.192-0.34)	0.993 (0.988-0.999)
	Tests with proposed probs.	0.815 (0.771-0.859)	0.957 (0.947-0.966)
	Controls	0.089 (0.049-0.129)	0.999 (0.998-1)
ICU patients	Tests	0.41 (0.308-0.512)	0.989 (0.98-0.998)
	Tests with proposed probs.	0.774 (0.714-0.835)	0.963 (0.95-0.976)
	Controls	0.123 (0.063-0.182)	0.995 (0.989-1)
CVS patients	Tests	0.037 (0.013-0.063)	
	Tests with proposed probs.	0.88 (0.823-0.938)	0.947 (0.933-0.96)



APL system RCT results

		Sensitivity	Specificity
	Controls	0.102 (0.069-0.135)	0.998 (0.995-1)
All patients	Tests	0.266 (0.192-0.34)	0.993 (0.988-0.999)
	Tests with proposed probs.	0.815 (0.771-0.859)	0.957 (0.947-0.966)
	Controls	0.089 (0.049-0.129)	0.999 (0.998-1)
ICU patients	Tests	0.41 (0.308-0.512)	0.989 (0.98-0.998)
	Tests with proposed probs.	0.774 (0.714-0.835)	0.963 (0.95-0.976)
	Controls	0.123 (0.063-0.182)	0.995 (0.989-1)
CVS patients	Tests	0.037 (0.013-0.063)	
	Tests with proposed probs.	0.88 (0.823-0.938)	0.947 (0.933-0.96)



APL system RCT results

		Sensitivity	Specificity
	Controls	0.102 (0.069-0.135)	0.998 (0.995-1)
All patients	Tests	0.266 (0.192-0.34)	0.993 (0.988-0.999)
	Tests with proposed probs.	0.815 (0.771-0.859)	0.957 (0.947-0.966)
	Controls	0.089 (0.049-0.129)	0.999 (0.998-1)
ICU patients	Tests	0.41 (0.308-0.512)	0.989 (0.98-0.998)
	Tests with proposed probs.	0.774 (0.714-0.835)	0.963 (0.95-0.976)
	Controls	0.123 (0.063-0.182)	0.995 (0.989-1)
CVS patients	Tests	0.037 (0.013-0.063)	Ι
	Tests with proposed probs.	0.88 (0.823-0.938)	0.947 (0.933-0.96)



Clinical Information Extraction

Textractor (for medications extraction)

Why extract medications?

- CPOE and e-Prescribing systems becoming widely available in the healthcare system.
- Large proportion of medications actually taken by the patient only mentioned in narrative clinical text documents in the EHR:
 - Prescribed in another institution or private practice,
 - Bought over-the-counter,
 - Prescribed before the introduction of CPOE.



Clinical Information Extraction

Textractor (for medications extraction)

Why extract medications?

- CPOE and e-Prescribing systems becoming widely available in the healthcare system.
- Large proportion of medications actually taken by the patient only mentioned in narrative clinical text documents in the EHR:
 - Prescribed in another institution or private practice,
 - Bought over-the-counter,
 - Prescribed before the introduction of CPOE.
- This is where automated medication information extraction could (and will) help!



2009 i2b2 Medication Extraction Challenge Task :

Identification of medications and related details in discharge summaries:

- Medication name (brand name, generic name, drug class)
- Dosage
- Route
- Frequency of the administration (incl. "as needed")
- Duration of the treatment
- -Reason(s) for the prescription

Corpus

1249 discharge summaries, de-identified and re-identified with realistic surrogates, split in a training corpus (696 documents) and a testing corpus (553 documents).



Textractor

Application details

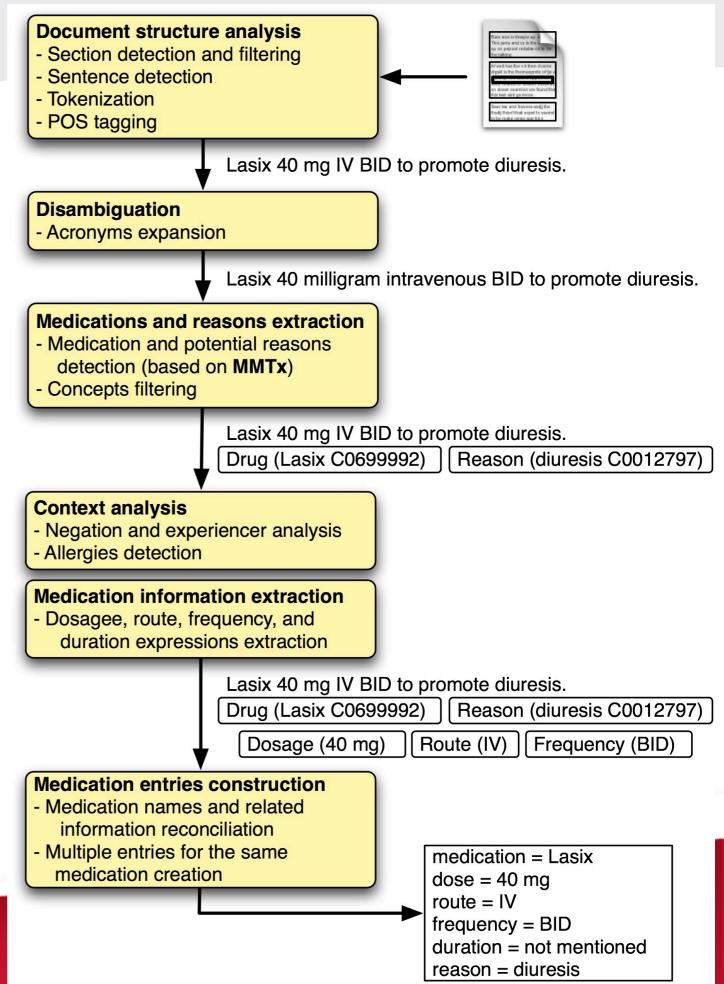
Built on UIMA, with functionalities developed as modules (TAEs), organized in a pipeline.

Several components based on OpenNLP tools, wrapped and retrained.

Uses MMTx (version 2.4.C), wrapped for UIMA in its first version. Dictionary lookup now based on Lucene.

Context analyzer based on the ConText algorithm, with improvements for this task.

Regular expressions used to extract dosage, route, etc.



2009 i2b2 Medication Extraction Challenge Evaluation :

	Exact match				Inexact match			
Fields	Recall	Precision	F ₁ -measure	Recall	Precision	F ₁ -measure		
Medication	0.746	0.772	0.759	0.763	0.784	0.773		
Dose	0.757	0.916	0.829	0.786	0.925	0.850		
Route	0.817	0.920	0.865	0.803	0.926	0.860		
Frequency	0.789	0.892	0.837	0.742	0.924	0.823		
Duration	0.326	0.397	0.358	0.326	0.501	0.395		
Reason	0.169	0.669	0.270	0.148	0.703	0.245		
Overall	0.715	0.832	0.769	0.693	0.839	0.759		



2009 i2b2 Medication Extraction Challenge Evaluation :

Exact match			
Recall	Precision		
0.746	0.772		
0.757	0.916		
0.817	0.920		
0.789	0.892		
0.326	0.397		
0.169	0.669		
0.715	0.832		
	Recall 0.746 0.757 0.817 0.789 0.326 0.169		

Duration and reason for a prescription annotations were difficult to extract, and difficult to annotate by humans!

	Our team's exact match IAA	Experts' Exact F-measure
Duration	0.16	0.47
Reason	0.31	0.40



2009 i2b2 Medication Extraction Challenge Evaluation :

	Exact match			
Fields	Recall	Precision		
Medication	0.746	0.772		
Dose	0.757	0.916		
Route	0.817	0.920		
Frequency	0.789	0.892		
Duration	0.326	0.397		
Reason	0.169	0.669		
Overall	0.715	0.832		

Duration and reason for a prescription annotations were difficult to extract, and difficult to annotate by humans!

	Our team's exact match IAA	Experts' Exact F-measure
Duration	0.16	0.47
Reason	0.31	0.40



Conclusions

For customers of automatically extracted clinical information, or CDS based on this information, trust can be based on several factors:

- Ensuring that there is a **human expert in the loop** for all clinical information generated by NLP and becoming part of the EHR.
- Allowing users to know the origin of the extracted information (and the methods used; and therefore use NLP methods that make it possible).

With trust in the NLP-generated information, even moderate performance is acceptable (moderate sensitivity, but sufficient PPV).

