Flexible NLP for Varied Applications and Data Sources

David Milward, PhD

CTO, Linguamatics

THE QUEEN'S AWARDS FOR ENTERPRISE: 2014

Linguamatics

Overview

- Using NLP to
  - find, normalize and summarize information
- Applications: bench to bedside
  - genotype-phenotype relationships
  - drug repositioning using clinical trial AE reports
  - regulatory QA
- Agile NLP
  - interactive, data-driven approach
- Adverse Event Mining
  - scientific articles, drug labels, Twitter

Find information however it is expressed

Different word, same meaning

cyclosporine

ciclosporin

Neoral

Sandimmune

Different expression, same meaning

Non-smoker

Does not smoke

Does not drink or smoke

Denies tobacco use

NLP

Different grammar, same meaning

5mg/kg of cyclosporine per day

5mg/kg per diem of cyclosporine

cyclosporine 5mg/kg per day

Same word, different context

Diagnosed with diabetes

Family history of diabetes

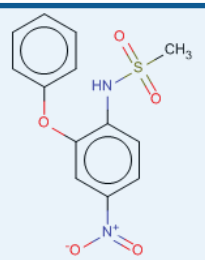
No family history of diabetes

Linguamatics

Represent information in a standard format

Category	Text	Normalized Value
Diseases	breast cancer	Breast Neoplasm
	carcinoma of the breast	
Genes	Raf-1	RAF1
	Raf I	
Dates	27th Feb 2014	20140227
	2014/02/27	
Measurements	0.2g	200 mg
	Two hundred milligrams	
Mutations	Val 158 Met	V158M
	Val by Met at codon 158	

nimesulide, a selective COX2 inhibitor, …



inhibits →

Entrez Gene ID: 5743

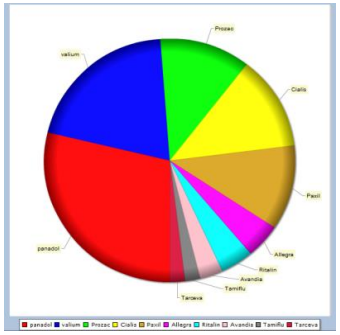
Summarize information for fast review

Identify → Extract → Synthesize → Analyze

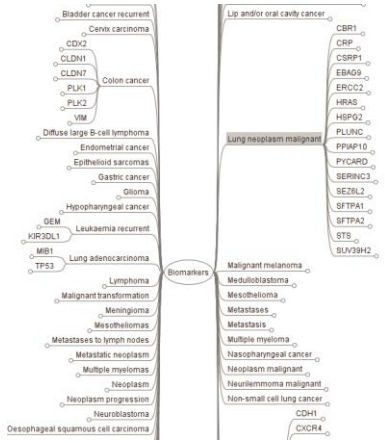
Clustered results

Gene/Protein	Biomarker	Disease	Doc	Hit
TRIM25	as a new potential biomarker for	Breast cancer	1 16144914	1 Estrogen-responsive finger protein as a new potential biomarker for breast cancer.
	is a significant prognostic factor in	Breast cancer	1 16144914	1 CONCLUSIONS: Our data suggest that Efp immunoreactivity is a significant prognostic factor in breast cancer patients.
	was significantly correlated with	Breast cancer	1 16144914	1 Moreover, Efp immunoreactivity was significantly correlated with poor prognosis of breast cancer patients, and multivariate analyses of disease-free survival and overall survival for 151 breast cancer patients showed that Efp immunoreactivity was the independent marker.
BCL2	is a prognostic marker in	Breast cancer	1 16638854	1 Bcl-2 is a prognostic marker in breast cancer independently of the Nottingham Prognostic Index.
	is an independent predictor of	Breast cancer	1 16638854	1 CONCLUSION: Bcl-2 is an independent predictor of breast cancer outcome and seems to be useful as a prognostic adjunct to the NPI, particularly in the first 5 years after diagnosis.
PCNA	as molecular marker in	Breast cancer	1 16637514	1 Cyclin E as molecular marker in the management of breast cancer: a review.
	to be a promising prognostic indicator in	Breast cancer	1 16637514	1 Recent studies found cyclin E to be a promising prognostic indicator in breast cancer and examined its potential as a target for therapy.
NCOR1	is an independent prognostic factor for	Breast cancer	1 16019133	2 NCOR1 mRNA is an independent prognostic factor for breast cancer.
APOE	was associated with	Breast cancer	1 15830139	1 We found risk for breast cancer was associated with the APOE genotype (xi(2) = 8.652, p = 0.013).
BRCA1	was elevated in	Breast cancer	1 16914727	1 Moreover, BRCA1 expression was elevated in HER4-postive human breast cancer specimens.
CDH1	was lower in	Breast cancer	1 15742334	1 RESULTS: E-cadherin mRNA expression was lower in breast carcinoma (P = 0.001), whereas Snail expression was higher (P = 0.003).
CLU	was found to be a prognostic factor for	Breast cancer	1 17121433	1 Cytoplasmic CLU expression was found to be a prognostic factor for human breast cancer.
CTCFL	can be a valuable early blood marker of	Breast cancer	1 17062669	1 Detection of BORIS in a high proportion of patients with various types of breast tumors indicates that BORIS can be a valuable early blood marker of breast cancer.

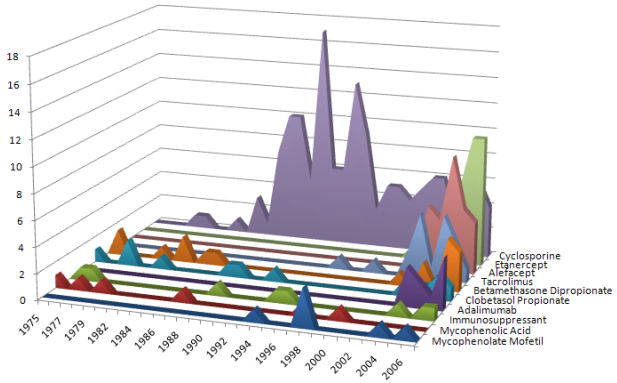
Charts for drill down



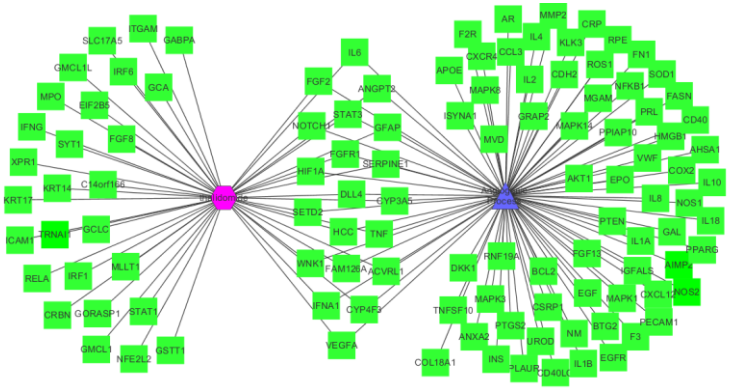
Mind maps with clustering



Trending over time



Indirect Relationships



Linguamatics

Feed information to Machine Learning

- NLP can turn unstructured text at large scale into features to drive predictive models
- Examples using Linguamatics I2E include:
  - Kaiser Permanente for pneumonia prediction
  - Roche to predict success or failure of target-indication pairs
  - Top-10 pharma to categorize call center transcriptions (Voice of the Customer)
    - patient reported outcomes
    - side effects
    - drug interactions

Applications: Bench to Bedside

Linguamatics

From Bench to Bedside: Insight Needed

Business critical questions

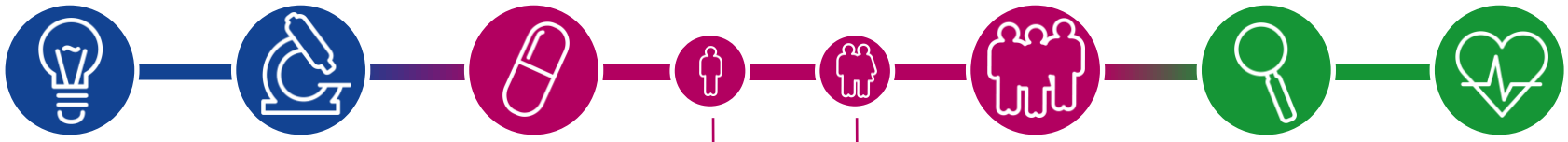
What targets are involved in bone cancer?

What companies are patenting a particular technology?

What are the safety risks of my drug?

Where can I site my Phase 1, Phase 3 clinical study?

What are the clinical risks for my patients?



Idea

Basic research

Clinical trials

Phase 1

Phase 2

Phase 3

Regulatory approval

Patient care

Discovery | Development | Delivery

Linguamatics

Genotype-Phenotype association in Hunter Syndrome (Shire)

- Rare X-linked recessive disorder
- Spectrum of clinical severity (mild to severe); main difference is progressive development of neurodegeneration in the severe form
- Structured databases lack broad phenotypic association data
- Sparse data, needs high recall across full text papers
- Extraction of patient mutations matched or bettered genetic databases
- Enables focused precision medicine approach for patient care

Hit
... and R48P, L196S, Q531X (mild phenotype).
Patients with R88C and H138R mutations displayed a severe phenotype.
In contrast, the attenuated phenotype reported in the patient carrying the E177X mutation (26) is ...
This nonsense mutation is associated with a very mild phenotype (patient 56, aged ...
... mutations present correlation with the attenuated form (c.1122C>T), while a greater ...
... mutations whereas the p.Ser142Phe and p.Ile360Tyrfs*31 mutations caused the severe disease manifestation.
A deletion involving exons 2-4 in the iduronate-2-sulfatase gene of a patient with intermediate Hunter syndrome
Mutation R468W of the iduronate-2-sulfatase gene in mild Hunter syndrome (mucopolysaccharidosis type II) ...
... mutations in exon 9 had mild disease (P469H; Y523C; R468W, ...
... C (1992) Mutation R468W of the iduronate-2-sulfatase gene in mild Hunter syndrome (mucopolysaccaridosis type II) ...
The A346D mutation was associated with the mild phenotype, all others with the ...
... nonsense mutations (Q80X; Q389X) in patients with severe Hunter syndrome (mucopolysaccharidosis type II)...

Drug Repositioning: ClinicalTrials.gov (Lilly)

- Clinical trials report Adverse Events that occur while a patient is taking a drug vs. a placebo
- If Adverse Events are fewer taking the drug vs. the placebo, the drug may be stopping the disease occurring
- Information extracted into Excel using Linguamatics I2E
  - Combination of use of terminologies, and structured fields in the document
  - 100K serious AEs classified as cancer
- Odds ratio and z-score calculated
  - Excel output loaded into Megaputer's PolyAnalyst
  - Results thresholded by number of AEs
- Results suggests existing drugs that could be used as novel treatments for cancer
- Opportunity to do the same on other data sources e.g. FDA AERS data, EHRs

Drug Repositioning: ClinicalTrials.gov

Serious Adverse Events

	Phyloquinone	Placebo
Total, serious adverse events		
Cancer † 1 [4]		
# participants affected / at risk	3/217 (1.38%)	11/223 (4.93%)
# events	3	12

1	Serious Event Subtit	Study Arm	Num. of Ptnts w SAE	Number of Patients	Doc
89392	Cancer	Phyloquinone	3	217	NCT00150969
89396	Cancer	Placebo	11	223	NCT00150969

Drug	Serious adverse event	*Ds*	*Dn*	*Cs*	*Cn*	Control	SE	OR	Lower limit	Upper limit	*z*	Clinical Trials.gov ID
V501	Cervical dysplasia	20	480	46	468	Placebo	0.28	0.40	0.23	0.69	−3.33	NCT00378560
Clopidogrel/ Telmisartan	Colon cancer	4	5,000	14	5,023	Clopidogrel/ Placebo	0.57	0.29	0.09	0.87	−2.20	NCT00153062

Regulatory Examples

- Extraction of values for IDMP (Identification of Medicinal Products)
  - Converting information within drug submission documents into structured data
  - Required in all the European languages
- Regulatory submission documents QA/QC
  - Varied documents (Office, PDF etc.)
  - Check summaries vs. source documents
  - Check information within tables vs. text
  - Check formatting, calculations, thresholds
  - MedDRA coding/code checking

Table: Most Frequently Reported Medical Conditions (≥5% in Any Treatment Group)

Study	2000 Pooled Studies		2003 Pooled Study	
Total Number Subjects	Rx N=997	Pbo N=927	Rx N=1021	Pbo N=956
Number (%) of Subjects				
Cardiac disorders	70 (7.0)	32 (3..5)	108 (10.6)	101 (10.6)
Angina pectoris	4 (0.4)	5 (0.5)	74 (7.2)	71 (7.4)
Dyspepsia	174 (17.5)	120 (12.9)	3 (0.3)	2 (0.2)
GERD	83 (8.3)	52 (5.6)	30 (2.9)	27 (2.8%)
Metabolic / nutritional disorders	253 (25.4)	165 (17.8)	194 (19.0)	212 (22.2)
Dyslipedaemia	1 (0.1)	0 (0)	15 (1.5)	19 (2.0)
Hypercholesterolaemia	65 (6.5)	50 (5.4)	88 (8.6)	103 (10.8)
Hyperlipidaemia	147 (14.7)	79 (8.5)	56 (5.5)	66 (6.9)
Osteoarthritis	102 (10.2)	57 (6.6)	12 (1.2)	11 (1.2)
Nervous system disorders	628 (63.0)	409 (44.1)	28 (2.7)	19 (2.0)
Headache	413 (41.4)	280 (30.2)	9 (0.9)	7 (0.7)
Psychiatric disorders	137 (13.7)	81 (8.7)	14 (1.4)	15 (1.6)
Insomnia	84 (8.4)	47 (5.1)	9 (0.9)	8 (0.8)

Commonly reported conditions included Seasonal allergies, Back pain, and Hypercholesterolaemia. The majority of AEs were considered treatment related in all cohorts and the relationship between treatment groups and between cohorts was similar to that observed for all-causality AEs. Permanent discontinuations were reported at higher rates in the Rx groups than in the placebo groups in the 3 pooled cohorts. The majority of AEs leading to permanent discontinuation were considered treatment related in both treatment groups in all cohorts. The single most frequently reported event was headache, which was reported in approximately 40% of Rx subjects and 20% of placebo subjects in the 2000 Pooled cohort. Other AEs reported across all cohorts at rates greater in Rx subjects than placebo subjects included Seasonal allergies and Insomnia (2000 8.4% vs 5.4%, 2003 0.9% vs 0.8%, 2006 14.0% vs 10.1%; Rx vs placebo respectively).

Key

Incorrect formatting: doubled period, incorrect number of decimal places, addition of percent sign

Incorrect calculation: number of patients divided by total number does not agree with percent term

Incorrect threshold: presence of row does not agree with table title

Text-Table inconsistency: numbers in the table do not agree with numbers in the accompanying text

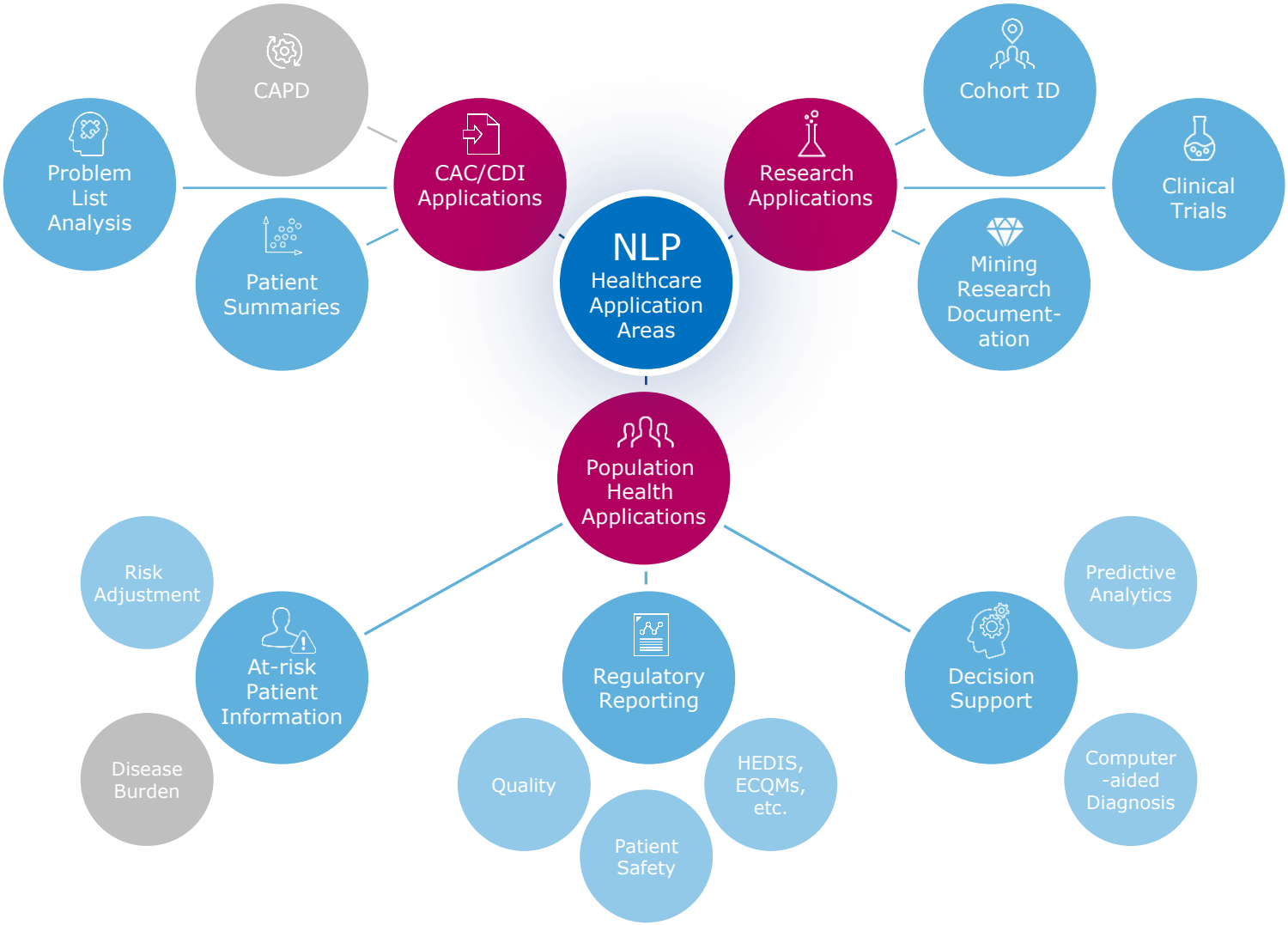
Sample table and text highlighting, to show inconsistencies between data. The highlight colour makes it easy for the reviewer to rapidly assess where there are errors and what type of errors, and can then correct these appropriately.

Coding Consistency Checking

- Take a term and output the standardized name from the coding scheme
- Compare the name to the name provided for the code (in brackets)
- Showing all examples here, but could restrict to those where there is an inconsistency, marked with red

Allergic arthritis	716.2	Allergic arthritis	716.2	ICD9	1	287.0 287.0... 999.51 999.59... Excludes: allergic arthritis NOS (716.2) 716.2 716.2 allergic arthritis NOS
Amebic liver abscess	006.3	Amebic liver abscess	006.3	ICD9	1	572... 572.0... Excludes: amebic liver abscess (006.3) 006.3 006.3 amebic liver abscess
Amyloidosis	277.3	Other general diseases with articular involvement	713.7	ICD9	1	272.0 272.9... 270.2 270.2... Excludes: arthropathy associated with: amyloidosis (713.7) 713.7 713.7... 712.1 712.9... 713.5 713.5... 274.00 274.03 ...
Anal and rectal polyp	569.0	Benign neoplasm of rectum and anal canal	211.4	ICD9	1	569... 569.0... Excludes: adenomatous anal and rectal polyp (211.4) 211.4 211.4 ...
Anemia in neoplastic disease	285.22	Anemia in neoplastic disease	285.22	ICD9	1	285.29... 285.3... Excludes: anemia due to drug NEC - code to type of anemia anemia in neoplastic disease (285.22) 285.22 285.22... 284.89 284.89 ...
Anencephalus	740.0	Anencephalus	740.0	ICD9	1	Excludes: acrocephalosyndactyly [Apert's syndrome] (755.55) 755.55 755.55... 524.0 524.9... anencephalus (740.0) 740.0 740.0... 742.0 742.0... 742.3 742.3... 742.1 742.1 ...

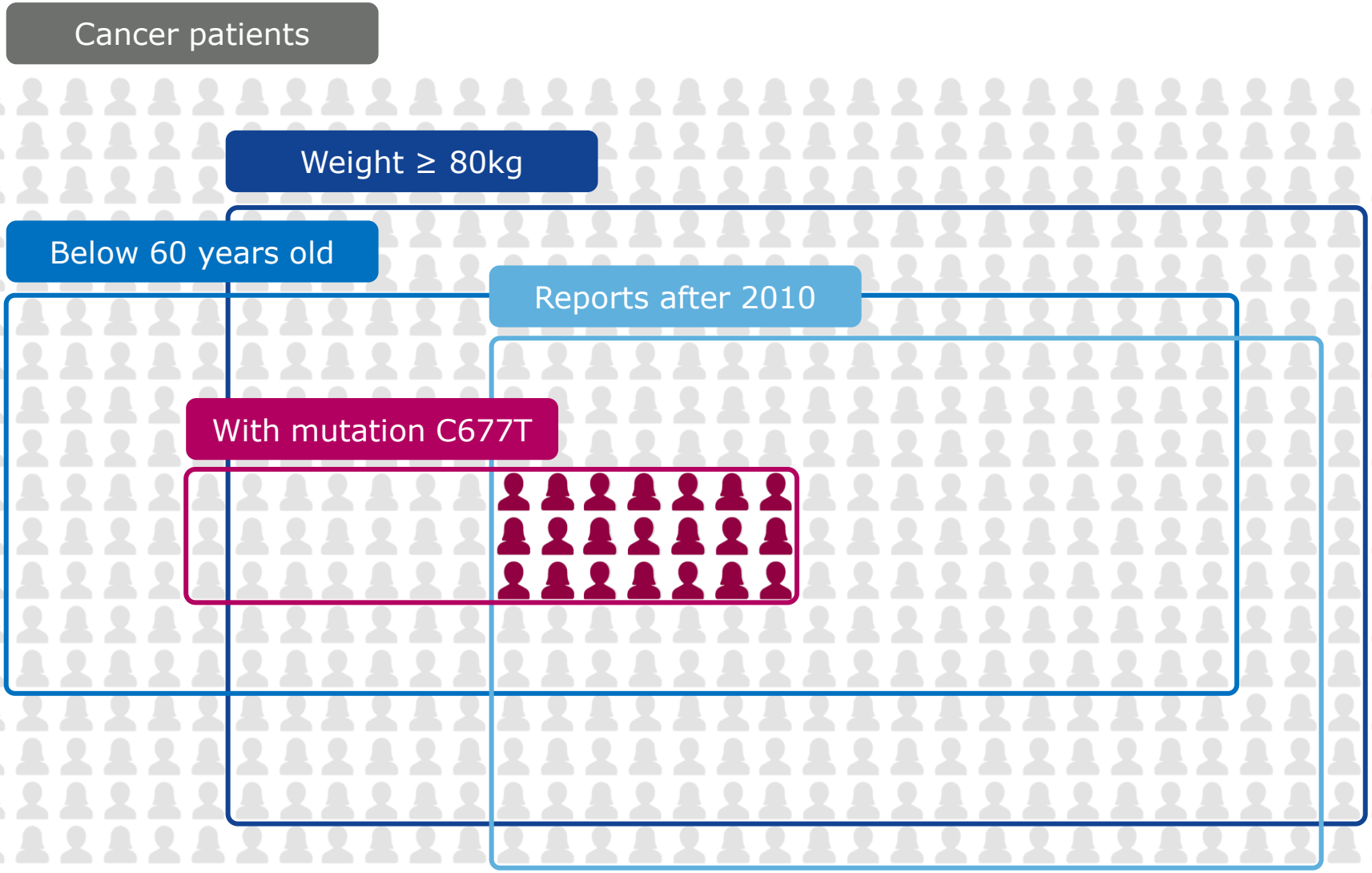
Clinical NLP Applications



Linguamatics

What Do People Look Like in Your Data?

Cohort Selection and Patient Stratification



Linguamatics

IMPROVING CHART REVIEW

MINING PATIENT RECORDS FOR DISEASE COMORBIDITIES

CHALLENGE

Identifying disease comorbidities for study via patient narratives. To find 700 patients with HIV and Hepatitis C manually took 5 medical students 4 months.

SOLUTION

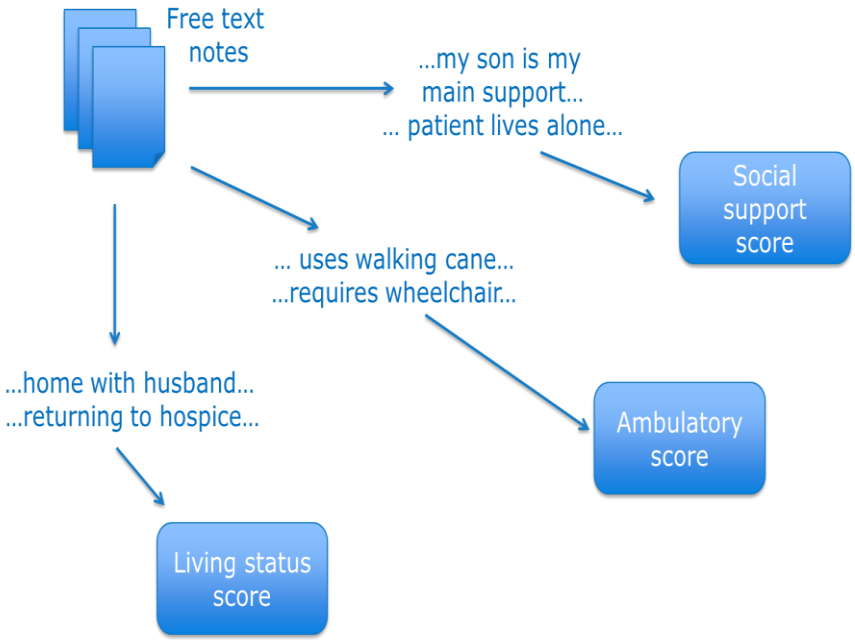
Using Linguamatics I2E queries for disease codes and terminology took less than half a day to identify 1100 patients.

BENEFIT

Patient groups can be quickly identified from both structured and unstructured text. Identifying new disease cohorts is easy and can be quickly iterated to select new groups for study.

Predicting 30-Day Readmissions

- Mining discharge summaries for insights from unstructured data
  - Social determinants, ambulatory status and living location
- I2E used to explore 700,000 patient data set and extract attributes for statistical and Machine Learning modelling
- Project resulted in
  - Well characterized, consistent and well populated data for ML without huge manual curation effort
  - Queries that can be used in real-time to support new predictive models



Population Stratification for Heart Failure

- Identification of risk factors in unstructured patient data relating to Congestive Heart Failure from EHR and nurse notes
- I2E mining large quantities of unstructured text to risk stratify the CHF population
  - 1.5TB of unstructured information, mixed format
  - Results saved into data warehouse

Identifying Missed Heart Failure Diagnosis

- Left Ventricle Ejection Fraction is a measure of heart performance
- Vital data for classification of heart failure
- Used to identify undiagnosed high risk patients for targeted care
- Values drive patient level and population risk models

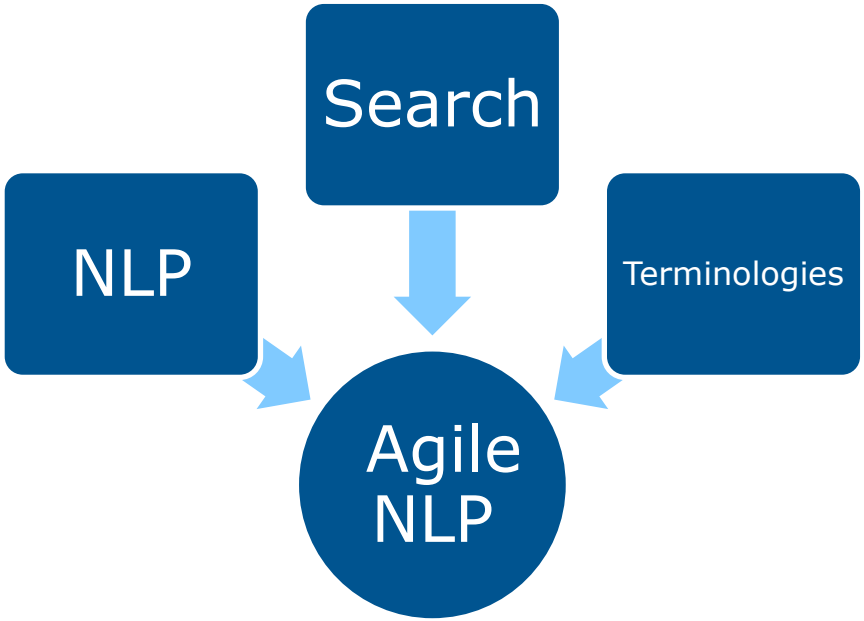
Doc	EKG Type	EKG Value	Hit
M2074		30%	... with ejection fraction of approximately 30 %.
M1727		20 % - 25 %	... heart failure with EF of 20 % - 25 %, hypertension, renal insufficiency...
M1727		20 % - 25 %	... failure with ejection fraction of 20 % - 25 % in December 2005, COPD...
M1924		20 % - 25 %	... with an ejection fraction of 20 % - 25 %.
M2543		35%	... revealed an ejection fraction of 35 %.
M2543		35%	... heart failure, ejection fraction 35 %; hypercholesterolemia; peripheral vascular ...
M933		35%	... with ejection fraction of about 35 % consistent with ischemic cardiomyopathy.
M1091		68%	... motion, ejection fraction of 68 %.
M1091		68%	2. Ejection fraction of 68 % by gated SPECT.
M1035		68%	... left ventricular ejection fraction was 68 %, which is within the ...
M1097		75%	... motion and ejection fraction of 75 %.
M1097		75%	2. Ejection fraction of 75 % by gated SPECT.
M768		75%	... echo imaging shows EF of 75 %, no regional wall motion ...
M1584		77%	... motion with ejection fraction of 77 %.
M1584		77%	... motion with ejection fraction of 77 %.
M2293		77%	... left ventricular ejection fraction of 77 %.
M281	Calculated	69%	Ejection fraction was calculated to be 69 %.
M1798	Calculated	69%	Left ventricular ejection fraction is calculated to be 69 %.
M919	Estimated	40%	... EF was moderately reduced with estimated EF of 40 % with near normal thickening.
M2268	Estimated	40%	Ejection fraction estimated 40 %, 1+ mitral regurgitation (...
M2180	Estimated	55%	... ventricular systolic function with an estimated LVEF of 55 %.
M2192	Estimated	55%	... left ventricular systolic function with estimated left ventricular ejection fraction of 55 %.

Agile NLP

Linguamatics

Agile NLP for Data Scientists/Analysts

- Traditional NLP is powerful, but not accessible to non-experts
  - patterns have to be programmed
  - or machine learned from relevant annotated data
- Linguamatics I2E provides:
  - interactivity and scalability of search
    - quick to develop new extraction patterns
    - accessible to data experts
  - precise, structured results of traditional NLP



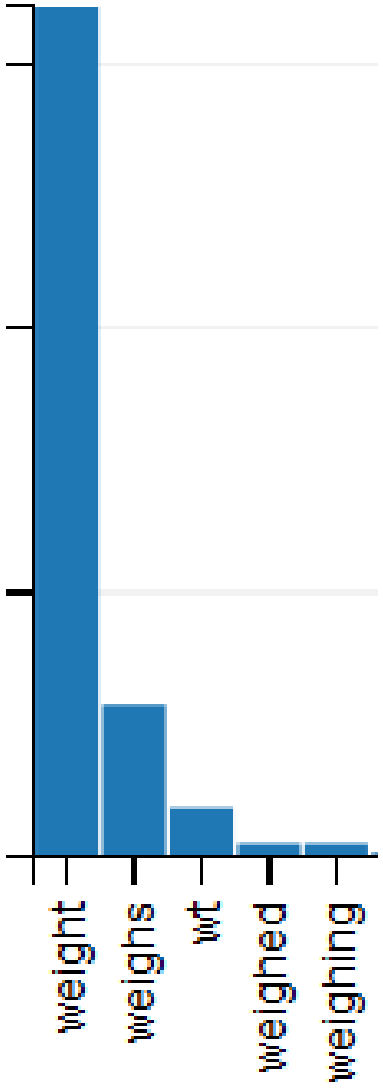
I2B2 2014 Cardiac Risk Factors

- Challenge to extract a fixed set of Cardiac Risk factors
  - Risk factors include:
    - medications, mentions of diabetes, hypertension, hyperlipidaemia, obesity, glucose/LDL/A1C/BMI test results, "cardiac events", family history of coronary artery disease, smoking etc.
  - Each annotation must also be given a temporal relation to the document
    - i.e. the patient had a heart attack BEFORE the day of the report, the patient's LDL was tested DURING the day of the report
- I2E F1-score: 91.7%

Lowest	Mean	Highest	Std. Deviation
35%	81.5%	92.8%	14.5%

Data-Driven Approach

- Interactive development of semantic and syntactic rules
  - similar to refining a keyword search
- Explore millions of documents to see how people express concepts and use different constructions
  - frequency analysis used to prioritize
- Compare results returned by high recall or high precision queries
  - refine precision and recall
  - reduce need for domain expertise



Adverse Event Mining/Coding

Linguamatics

AEs from MEDLINE and Drug Labels

Evidence from unstructured text in external document

Evidence from FDA label

Drug (NCI PT)	MedDRA	Doc Source	FDA Produc		Doc		Hit
Donepezil Hydrochloride	Bradycardia	FDA Labels	Donepezil Hydrochloride	35	9dffbd3e-8046-45ac-b56e-538cde87b59b	2	Donepezil Hydrochloride... heart failure, arteritis, bradycardia, peripheral vascular disease, ...
			donepezil hydrochloride	10	f9bcac55-5b0a-4213-89bd-7d6d30ca76e2	2	donepezil hydrochloride... heart failure, arteritis, bradycardia, peripheral vascular disease, ...
			Donepezil hydrochloride	9	7e79f700-695d-4ab5-b3b3-dfbb[illegible]0853a7	2	Donepezil hydrochloride... heart failure, arteritis, bradycardia, peripheral vascular disease, ...
			DONEPEZIL HYDROCHLORIDE	6	9e4deae[illegible]c73a-4d5f-94bf-a7ca02a815[illegible]	2	DONEPEZIL HYDROCHLORIDE... heart failure, arteritis, bradycardia, peripheral vascular disease, ...
			Aricept	5	d76dc06e-e796-[illegible]4c-87b3-0bf34fce34fc	2	Aricept... heart failure, arteritis, bradycardia, peripheral vascular disease, ...
			ARICEPT	1	E6D28BF2-7C39-256A-[illegible]0-39B48D367EBE	2	ARICEPT... heart failure, arteritis, bradycardia, peripheral vascular disease, ...
		MEDLINE		1	24615803	1	... several case reports suggested that donepezil hydrochloride can induce bradycardia or atrioventricular block, the ...
	Diarrhoea	FDA Labels	Donepezil Hydrochloride	35	ba6ba944-03d3-4ec0-9c6e-15a04044ea40	11	Donepezil Hydrochloride... donepezil hydrochloride are nausea, diarrhea, insomnia, vomiting, ...
	Nausea	FDA Labels	Donepezil Hydrochloride	35	ba6ba944-03d3-4ec0-9c6e-15a04044ea40	11	Donepezil Hydrochloride... studies of donepezil hydrochloride are nausea, diarrhea, insomnia, ...

(A) MEDLINE provides justification for Brachychardia in FDA Label of Donepezil Hydrochloride.

AEs from MEDLINE vs. Drug Labels

Drug (NCI PT)	MedDRA	Doc Source	FDA Product		Doc		Hit
This page continues an incomplete cluster from the previous page							
▼Galantamine Hydrobromide	Nervous system disorders NEC	▶ FDA Labels	Galantamine hydrobromide	1	241a824a-a025-4d92-b26c-c51e1d9db0f9	▶ 2	Galantamine hydrobromide... Central & peripheral nervous system disorders
	Nightmare	▶ FDA Labels	GALANTAMINE HYDROBROMIDE	1	2bec3ab0-31c5-48d6-b435-f04d1281665b	1	GALANTAMINE HYDROBROMIDE... : Infrequent: apathy, paroniria, paranoid reaction, libido ...
	Nocturia	▶ FDA Labels	GALANTAMINE HYDROBROMIDE	1	2bec3ab0-31c5-48d6-b435-f04d1281665b	1	GALANTAMINE HYDROBROMIDE... cystitis, urinary retention, nocturia, renal calculi
	Oedema peripheral	▶ FDA Labels	GALANTAMINE HYDROBROMIDE	1	2bec3ab0-31c5-48d6-b435-f04d1281665b	1	GALANTAMINE HYDROBROMIDE... injury, back pain, peripheral edema, asthenia, chest pain...
	Oesophageal perforation	▶ FDA Labels	GALANTAMINE HYDROBROMIDE	1	2bec3ab0-31c5-48d6-b435-f04d1281665b	1	GALANTAMINE HYDROBROMIDE... , hiccup; rare: esophageal perforation
	Orthostatic hypotension	▶ FDA Labels	GALANTAMINE HYDROBROMIDE	1	2bec3ab0-31c5-48d6-b435-f04d1281665b	1	GALANTAMINE HYDROBROMIDE... System Disorders: Infrequent: postural hypotension, hypotension, dependent edema...
	Palpitations	▶ FDA Labels	GALANTAMINE HYDROBROMIDE	1	2bec3ab0-31c5-48d6-b435-f04d1281665b	1	GALANTAMINE HYDROBROMIDE... Infrequent: AV block, palpitation, atrial arrhythmias including atrial ...
	Paraesthesia	▶ FDA Labels	GALANTAMINE HYDROBROMIDE	1	2bec3ab0-31c5-48d6-b435-f04d1281665b	1	GALANTAMINE HYDROBROMIDE... , involuntary muscle contractions, paresthesia, ataxia, hypokinesia, ...
	Paranoia	▼FDA Labels	GALANTAMINE HYDROBROMIDE	1	2bec3ab0-31c5-48d6-b435-f04d1281665b	1	GALANTAMINE HYDROBROMIDE... : apathy, paroniria, paranoid reaction, libido increased, delirium...
			Galantamine hydrobromide	1	241a824a-a025-4d92-b26c-c51e1d9db0f9	1	Galantamine hydrobromide... : apathy, paroniria, paranoid reaction, libido increased, delirium...
			RAZADYNE	1	ea6968e3-6383-4ff1-8c81-e4a2fdfe749d	1	RAZADYNE... : apathy, paroniria, paranoid reaction, libido increased, delirium...
	Pemphigoid	MEDLINE		1	19363906	▶ 3	Bullous pemphigoid precipitated by galantamine hydrobromide.

(B) Discovery of an adverse event in an external document which is not listed in any of the FDA labels containing Galantamine Hydrobromide (Pemphigoid is only found in MEDLINE).

Dealing with Varied Context

- Even in a warning section, not everything should be coded as an AE

WARNING: RENAL FAILURE, HEPATIC FAILURE, AND GASTROINTESTINAL HEMORRHAGE

section root-code (**79d04abf-9f97-4bfc-8c93-cd36a9dc8555**)
section code (**34066-1**)
section code-system (**2.16.840.1.113883.6.1**)
section displayName (**BOXED WARNING SECTION**)
ID-code ( )
effective-time (**20130529**)

Renal Failure

- Exjade can cause acute renal failure and death, particularly in patients with comorbidities and those who are in the advanced stages of their hematologic disorders.
- Measure serum creatinine and determine creatinine clearance in duplicate prior to initiation of therapy and monitor renal function at least monthly thereafter. For patients with baseline renal impairment or increased risk of acute renal failure, monitor creatinine weekly for the first month, then at least monthly. Consider dose reduction, interruption, or discontinuation based on increases in serum creatinine [see Dosage and Administration (2.4, 2.5), Warnings and Precautions (5.1)].

Tabular Data, not just Free Text

- Process tables to deal with merged cells, and connect headers with cells
- Allows a single query to access data from differently structured tables

Blood and lymphatic system disorders

Very common:	Neutropenia, thrombocytopenia, anaemia
Common:	Pancytopenia, febrile neutropenia
Uncommon:	Thrombocythaemia, lymphopenia, bone marrow depression, eosinophilia, lymphadenopathy
Rare:	Haemolytic anaemia

System organ class	Very common (>1/10)	Common (>1/100 to <1/10)	Uncommon (>1/1,000 to <1/100)	Rare (>1/10,000 to <1/1,000)	Not Known (can not be estimated from the available data)
Infection and infestations				Conjunctivitis	
Immune system disorders			Allergic oedema and angioedema	Allergic reaction	
Psychiatric disorders			Sleep disorder	Anxiety	

AEs from Social Media

Aripiprazole	other	made me a total zombo.	1	63	1	@birbigs christ, abilify made me a total zombo. ill stick with my effexor ...
Venlafaxine Hydrochloride	other	almost made me kill myself	1	1218	1	@dropkickpikachu yes, seriously effexor almost made me kill myself get off that shit
Topiramate	- weight	Weight	3	123	1	Use of Topamax for Losing Weight in real world, from 21 reports - eHealthMe http://tinyurl.com/ykj6wzm
		weight	1	44	1	Weight Loss Tips - How long does it usually take to lose weight on Topamax? ...
	other	made me nuts	1	192	1	Hmm I thought RSD & topamax made me nuts and it was just husband...
		making me dizzy	1	248	1	Prozac and Topamax making me dizzy.
		really badly affected my memory	1	147	1	@brandibax I got put on Topamax. That was a nightmare to take! It really badly affected my memory. The day I couldn't drive home b/c I had
		worked but made me sleep	1	96	1	@Dcornwe I've had Topamax, it worked but made me sleep.
	positive	effective	3	261	1	Topiramate (Topamax) appears as effective as antipsychotic risperidone in Alzheimer's ...
	negative	bad	1	203	1	... I had to go on topamax but it has bad side affects- mine were stress...

Summary

- NLP is providing access to the approximately 80% of data otherwise trapped in unstructured text
  - key to more effective drug discovery and delivery of better healthcare
- Linguamatics I2E agile NLP provides insights across the bench-to-bedside continuum
  - precise, structured results in the format required
  - interactive and scalable search
- Now completing the loop from the bedside back to the bench