GRAPH+AI WORLD

organized by TigerGraph

Accelerate AI with Graph Algorithms

Beyond Context: Answering Deeper Questions by Combining Spark NLP and Graph Database Analytics

Abhishek Mehta - Director Sales Engineering (TigerGraph)

Christian Kasim Loan - Senior Data Scientist (John Snow Labs)

Today's Presenter : Christian Kasim Loan

Christian Kasim Loan

Senior Data Scientist

- Distributed AI Lab(DAI), Daimler-Lab, CKL-IT (Consulting Company) Founder
- 10+ years , Architected and implemented various cloud agnostic big data systems and frameworks
- Creator of the NLU library
- Email: christian@johnsnowlabs.com

John Snow LABS

Leader in AI & Healthcare with NLP, OCR and AI Plattform

Founded in 2015, fully remote

- Customers like Intel, Johnson & Johnson, Roche, and Kaiser Permanente
- Most used NLP library and Leader in AI & Healthcare, won various awards
- **Cloud Agnostic Service, on-prem, Python / Java / Scala / R API's**
- **Key Terms**: State of the art NLP & NLU, Spark, Big Data, Healthcare AI solutions

Today's Presenter : Abhishek Mehta

Abhishek Mehta

Director of Field Engineering

- McKinsey, Bloomberg, Cisco & Dabizmo (NLP Startup) Founder
- 15+ years designing and implementing complex analytics solutions for Fortune 100 companies
- Patents in NLP spanning Conceptunary Ontology Design, Language Pattern Recognition, and Conversion
- Email: abhi@tigergraph.com

TigerGraph

Native Graph with MPP Architecture

Founded in 2012, Redwood City, CA

- World's top 7/10 banks, biggest healthcare company, biggest BioTech Company, biggest utilities company as customers
- Raised 105 Million $ as Series C in Feb 2021
- **Available on-prem, DBaaS, on AWS, GCP, Azure**
- **Key Terms**: OLAP + OLTP,Distributed Graph, ACID Compliant, Terabytes of scale

Graph NLU - Business Context

Gartner expected global revenue of BI to be $22.8 billion by 2020, and Reuters foresees additional growth to $29.48 billion by 2022.

In 2020, The global text analytics market was valued at USD 5.46 billion ; 20% CAGR

80% of Business Data is unstructured

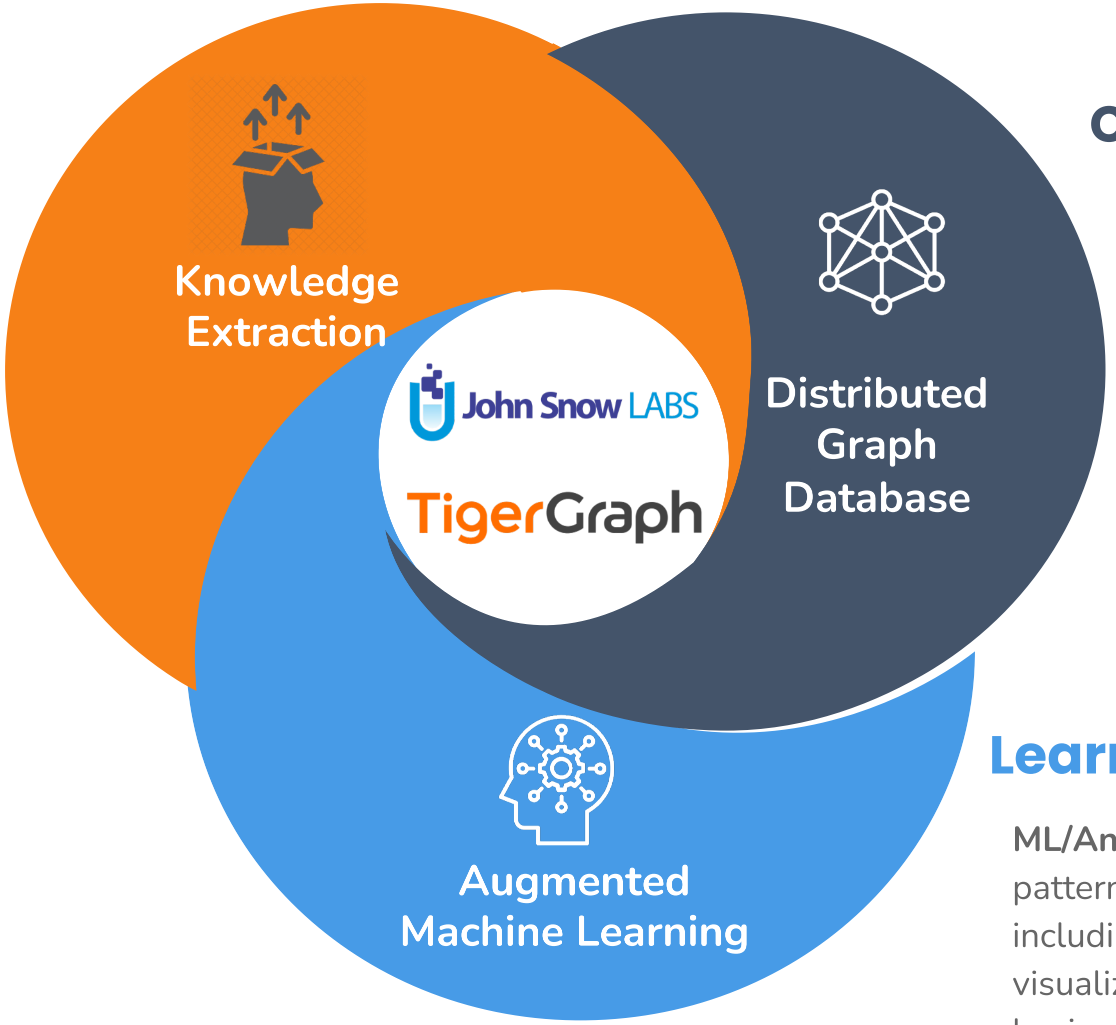
Graph NLU - Combining Spark NLP & TigerGraph

Knowledge Extraction/NLU

360 Applications : PII, EHR, Similarity Detection, Appointments, Health Plan Details, Visits, Tests, Communications, Patient Journey

Rx/Dx Data : Prescriptions, Prescriber Notes, Generics, Detecting Fraud e.g Opioid Fraud

Life Sciences: Find CoMorbidities Genes, path between Disease X & Protein Y.



Connected Data

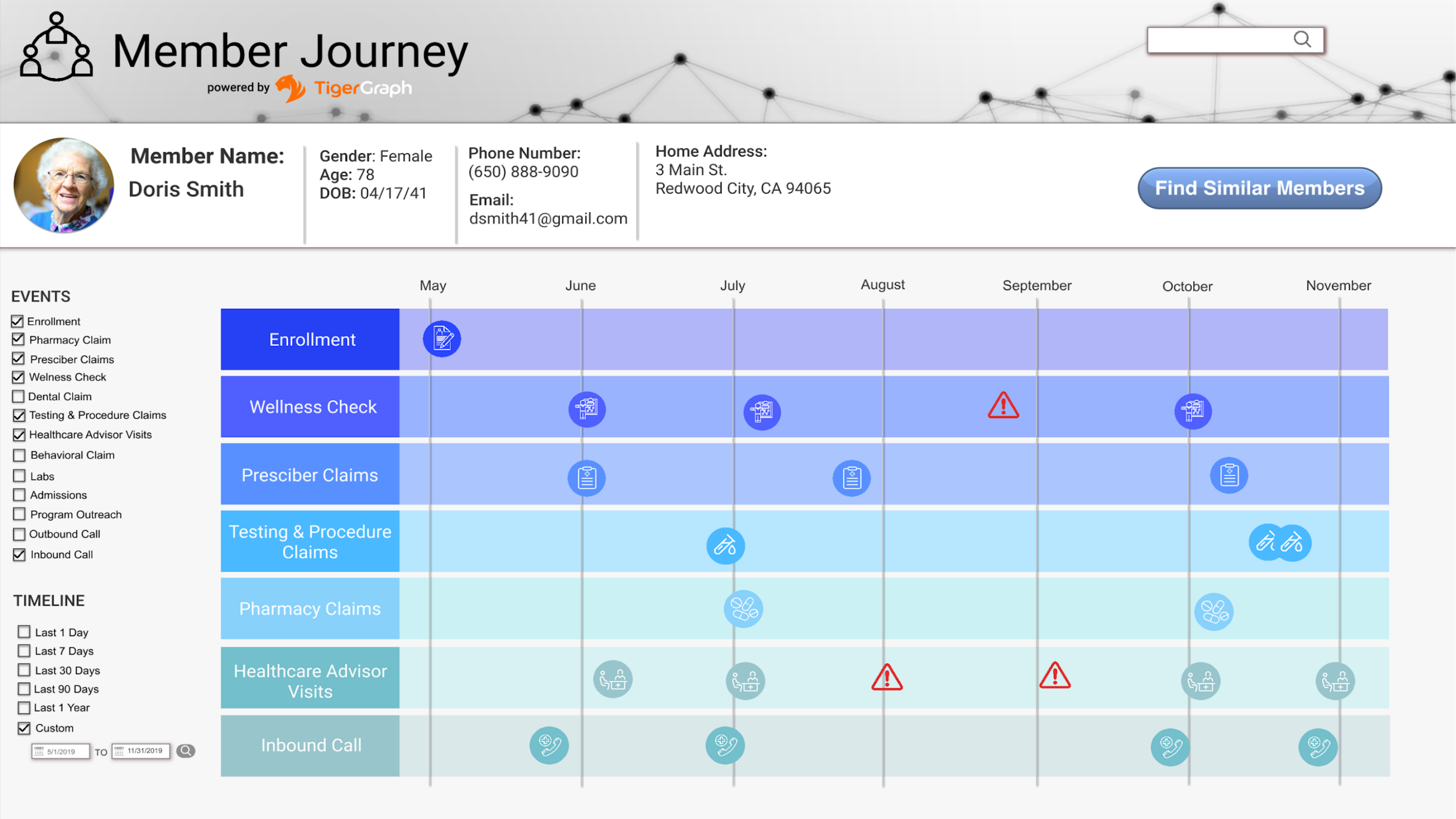
Graph Data Connectivity : Compliance, Regulations, Security

Scale: Friction-free scale up from GB to TB to PB with **lowest cost of ownership**

Speed: MPP Architecture, Ad Hoc Analysis

Learn From Connected Data

ML/Analytics: Deep Link Multi-Hop Analytics, pattern recognition,, data mining techniques including link and association analysis, visualization, and predictive analytics oe as basic as Frequency distributions


Member Journey
powered by TigerGraph
Member Name:
Doris Smith
Gender: Female
Age: 78
DOB: 04/17/41
Phone Number:
(650) 888-9090
Email:
dsmith41@gmail.com
Home Address:
3 Main St.
Redwood City, CA 94065
Find Similar Members
EVENTS
Enrollment
Pharmacy Claim
Presciber Claims
Welness Check
Dental Claim
Testing & Procedure Claims
Healthcare Advisor Visits
Behavioral Claim
Labs
Admissions
Program Outreach
Outbound Call
Inbound Call
TIMELINE
Last 1 Day
Last 7 Days
Last 30 Days
Last 90 Days
Last 1 Year
Custom
5/1/2019
TO
11/31/2019
May
June
July
August
September
October
November
Enrollment
Wellness Check
Presciber Claims
Testing & Procedure Claims
Pharmacy Claims
Healthcare Advisor Visits
Inbound Call

Clinical Named Entity Recognition (NER)

A 28-year-old female with a history of gestational diabetes mellitus diagnosed eight years prior to presentation and subsequent type two diabetes mellitus (T2DM), one prior episode of HTG-induced pancreatitis three years prior to presentation , associated with an acute hepatitis , and obesity with a body mass index (BMI) of 33.5 kg/m2 , presented with a one-week history of polyuria , polydipsia , poor appetite , and vomiting . Two weeks prior to presentation , she was treated with a five-day course of amoxicillin for a respiratory tract infection . She was on metformin , glipizide , and dapagliflozin for T2DM and atorvastatin and gemfibrozil for HTG . She had been on dapagliflozin for six months at the time of presentation . Physical examination on presentation was significant for dry oral mucosa ; significantly , her abdominal examination was benign with no tenderness , guarding , or rigidity . Pertinent laboratory findings on admission were : serum glucose 111 mg/dl , bicarbonate 18 mmol/l , anion gap 20 , creatinine 0.4 mg/dL , triglycerides 508 mg/dL , total cholesterol 122 mg/dL , glycated hemoglobin (HbA1c) 10% , and venous pH 7.27 . Serum lipase was normal at 43 U/L . Serum acetone levels could not be assessed as blood samples kept hemolyzing due to significant lipemia . The patient was initially admitted for starvation ketosis , as she reported poor oral intake for three days prior to admission . However , serum chemistry obtained six hours after presentation revealed her glucose was 186 mg/dL , the anion gap was still elevated at 21 , serum bicarbonate was 16 mmol/L , triglyceride level peaked at 2050 mg/dL , and lipase was 52 U/L . The β-hydroxybutyrate level was obtained and found to be elevated at 5.29 mmol/L - the original sample was centrifuged and the chylomicron layer removed prior to analysis due to interference from turbidity caused by lipemia again .

Color codes:PROBLEM, TREATMENT, TEST,

The patient was prescribed 1 capsule of Advil for 5 days . He was seen by the endocrinology service and she was discharged on 40 units of insulin glargine at night , 12 units of insulin lispro with meals , and metformin 1000 mg two times a day . It was determined that all SGLT2 inhibitors should be discontinued indefinitely fro 3 months .

Color codes:FREQUENCY, DOSAGE, DURATION, DRUG, FORM, STRENGTH,

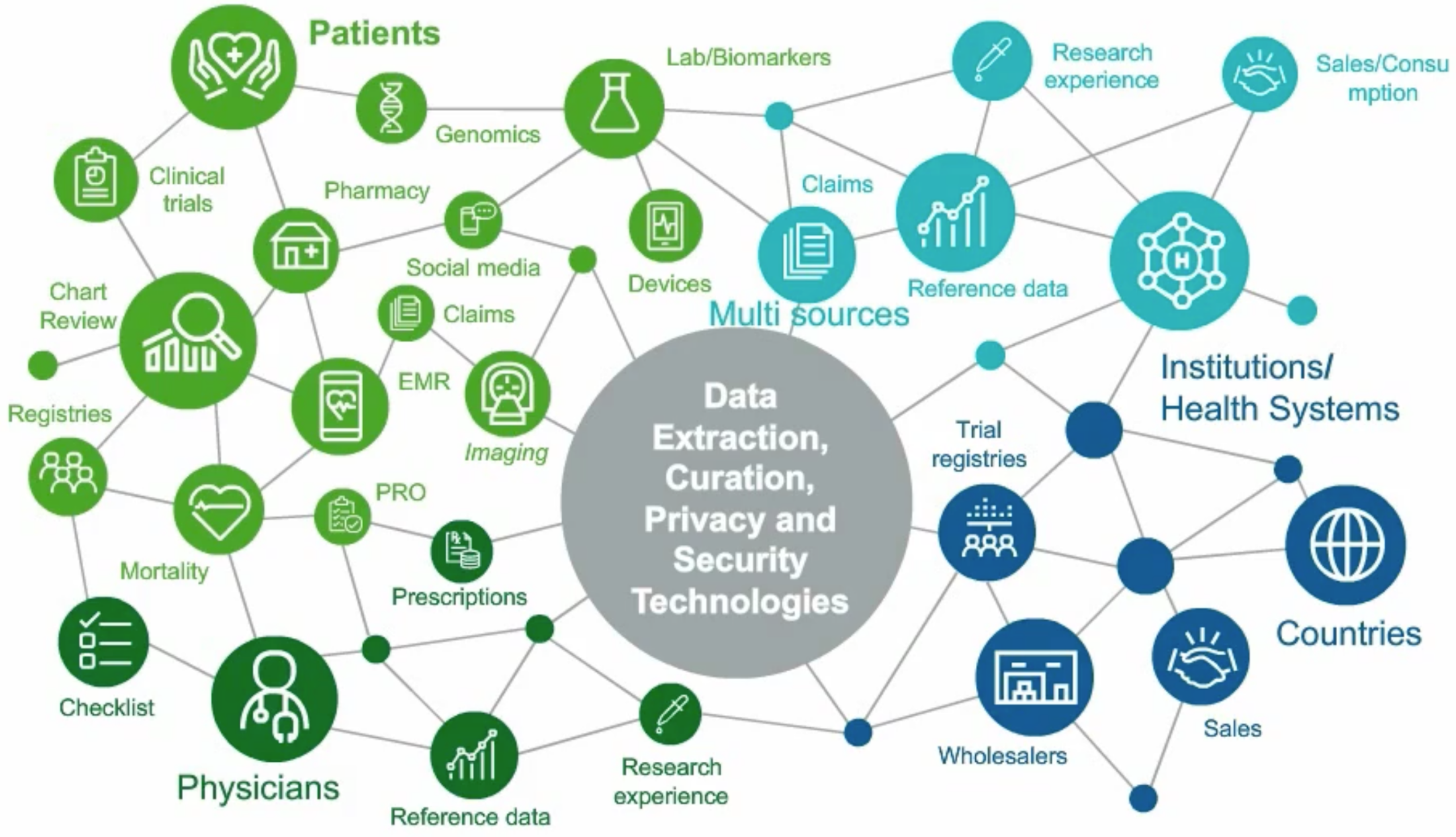
Label	Concept	Description
DOSAGE	1-2, sliding scale, taper, bolus, thirty (30) ml	The total amount of a drug administered
DRUG	aspirin, lisinopril, prednisone, vitamin b, flagyl	Generic or brand name of the medication
DURATION	for 3 days, 7 days, chronic, x5 days, for five more days	The length of time that the drug was prescribed for
FORM	tablet, capsule, solution, puff, adhesive patch, disk with device	A particular configuration of the drug which it is marketed for use
FREQUENCY	once a day, b.i.d., prn, q6h, hs, every six (6) hours as needed	The dosage regimen at which the medication should be administered
ROUTE	iv, p.o. (by mouth), gtt, nasal canula, injection,	The path by which the drug is taken into the body
STRENGTH	5mg, 100 unit/ml, 50mg/2ml, 0.05%, 25-50mg	The amount of drug in a given dosage

A . Record date : 2093-01-13 , David Hale , M.D . , Name : Hendrickson , Ora MR . # 7194334 Date : 01/13/93 PCP : Oliveira , 25 years-old , Record date : 2079-11-09 . Cocke County Baptist Hospital . 0295 Keats Street

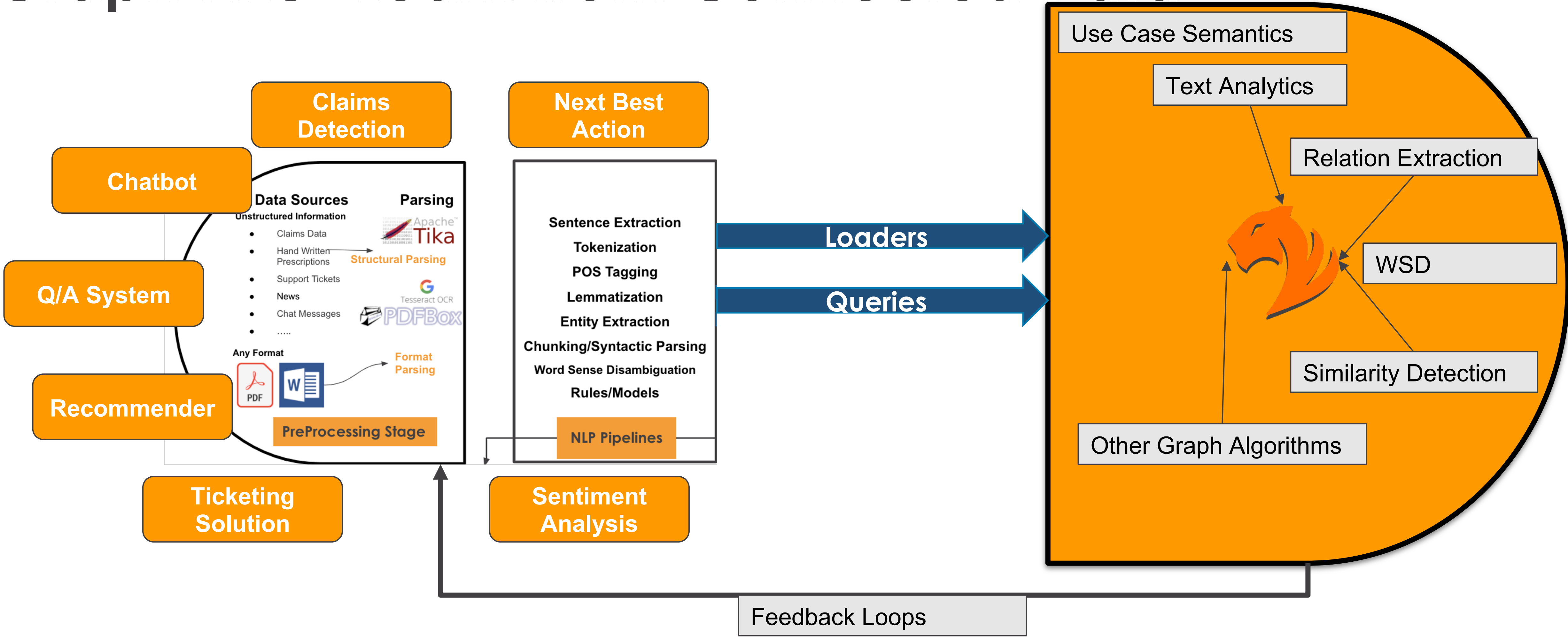
Color codes:STREET, DOCTOR, AGE, HOSPITAL, PATIENT, DATE, MEDICALRECORD,

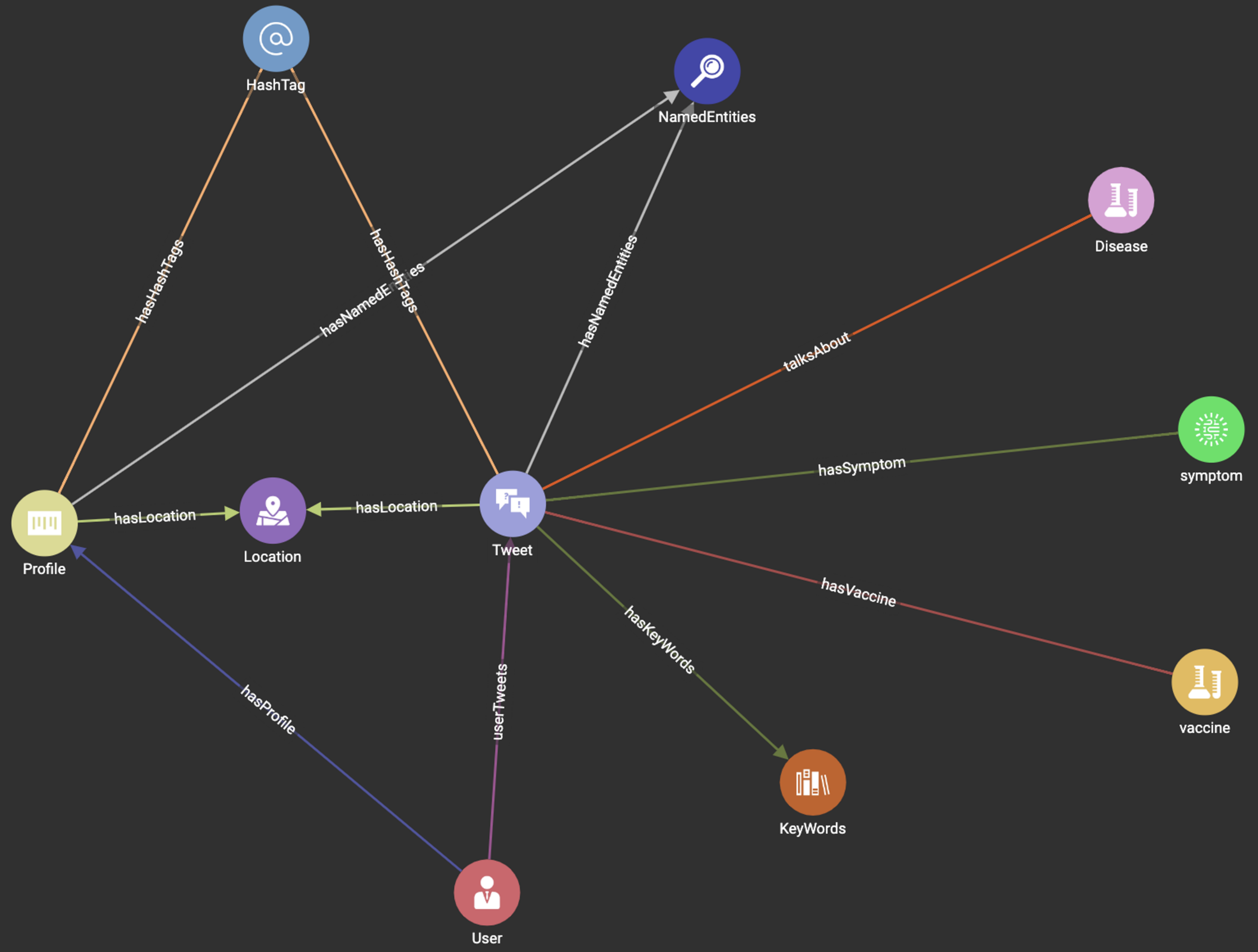
Pre & Post NLU - Connected Data Challenge

1. **Data silos**
2. Highly variable data
3. **Data formats**
4. Terminology
5. Intercompany trust
6. **Data privacy**
7. Complex rules
8. Chatbots and search
9. **Machine learning & explainability**
10. **Scale**



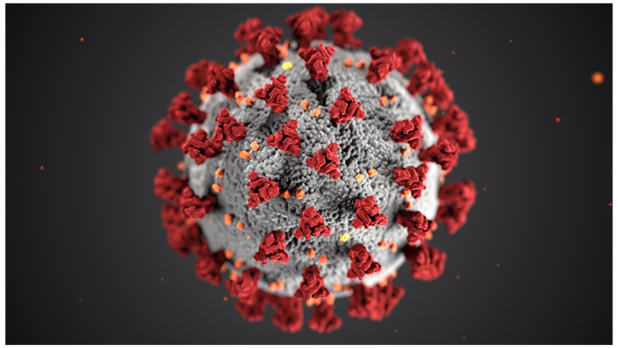
Graph NLU- Learn from Connected Data




HashTag
NamedEntities
Disease
symptom
vaccine
KeyWords
User
Profile
Location
Tweet
hasHashTags
hasNamedEntities
hasHashTags
hasNamedEntities
talksAbout
hasSymptom
hasLocation
hasLocation
hasVaccine
hasKeyWords
userTweets
hasProfile

UseCase

The COVID dataset



- https://www.kaggle.com/smid80/coronavirus-covid19-tweets-late-april
- https://www.kaggle.com/smid80/coronavirus-covid19-tweets-early-april
- https://www.kaggle.com/gpreda/pfizer-vaccine-tweets
- https://www.kaggle.com/gpreda/all-covid19-vaccines-tweets
- 5.6 Million Tweets collected

user_name	user_location	user_created	user_followers	user_friends	user_favourites	user_verified	date	text	hashtags	source	is_retweet	tweet_id
IMSS_SanLuis	na	2017-05-04T22:00:38Z	1008	41	300	False	2020-03-29T00:00:00Z	Ante cualquier enfermedad respiratoria, no te ...	['#PrevenciónCoronavirus', '#Coronavirus', '#C...	TweetDeck	False	0
intrac_ccs	na	2019-05-08T01:21:16Z	90	316	1030	False	2020-03-29T00:00:00Z	#ATENCIÓN En el Terminal Nuevo Circo se implem...	['#ATENCIÓN', '#Coronavirus', '#28Marzo']	TweetDeck	False	1
rlieving	na	2009-10-08T21:06:08Z	136	457	604	False	2020-03-29T00:00:00Z	"People are just storing up. They are staying ...	['#minneapolis', '#mn', '#covid19', '#coronavi...	TweetDeck	False	2
Tu_IMSS_Coah	na	2017-01-05T18:17:00Z	1549	170	1827	False	2020-03-29T00:00:00Z	Si empezaste a trabajar, necesitas dar de alta...	['#IMSS,', '#SanaDistancia', '#QuédateEnCasa',...	TweetDeck	False	3
Tabasco_IMSS	na	2016-10-19T22:05:03Z	868	125	723	False	2020-03-29T00:00:00Z	Una sociedad informada está mejor preparada an...	['#Coronavirus', '#COVID19']	TweetDeck	False	4
SSalud_mx	na	2010-04-12T16:53:45Z	812318	212	3954	True	2020-03-29T00:00:00Z	¡#Infórmate! #ConferenciaDePrensa sobre el #Co...	['#Infórmate!', '#ConferenciaDePrensa', '#Coro...	TweetDeck	False	5
AmerMedicalAssn	na	2009-03-31T17:50:31Z	714952	6877	2894	True	2020-03-29T00:00:00Z	.@PatriceHarrisMD spoke with @YahooFinance abo...	['#COVID19', '#pandemic']	Sprinklr	False	6
CGTNOfficial	na	2013-01-24T03:18:59Z	14040072	55	65	True	2020-03-29T00:00:00Z	First medical team aiding #Wuhan in fight agai...	['#Wuhan', '#COVID19', '#CoronavirusOutbreak']	Twitter Media Studio	False	7
Alaraby_Sport	na	2014-06-05T09:50:31Z	36953	1003	36	True	2020-03-29T00:00:00Z	هكذا ساهم نجم كرة القدم العالمية والفرنسية، كي...	['#معاً_نعزل_كورونا', 'كورونا', '#COVID19']	TweetDeck	False	8
OnTopMag	na	2010-01-27T05:23:15Z	5042	5389	2658	False	2020-03-29T00:00:00Z	.@KathyGriffin: @realDonaldTrump Is 'Lying' Ab...	['#Coronavirus', '#covid19', '#lgbt']	Twitter for Advertisers	False	9
ContraReplicaMX	na	2018-09-19T19:40:04Z	13287	2559	5671	False	2020-03-29T00:00:00Z	A pesar de la contingencia sanitaria provocada...	['#Covid19,']	TweetDeck	False	10
SSC_Pue	na	2010-03-10T20:04:51Z	297013	223	1726	False	2020-03-29T00:00:00Z	Ya sea a pie, en vehículo y hasta por espacio ...	['#COVID19.', '#QuédateEnCasa.']	TweetDeck	False	11
uri_911	na	2020-03-17T13:09:13Z	66	74	441	False	2020-03-29T00:00:00Z	#VEN911Oficial | #28Mar Es muy importante que ...	['#VEN911Oficial', '#28Mar', '#Covid19']	TweetDeck	False	12
SecAytoPue	na	2018-02-08T19:51:52Z	2251	788	312	False	2020-03-29T00:00:00Z	¿Qué es el coronavirus🤔, y cuáles son sus prin...	['#COVID19']	TweetDeck	False	13
livemint	na	2008-11-27T09:07:38Z	1862858	127	474	True	2020-03-29T00:00:00Z	#CoronaUpdate | Johns Hopkins University has s...	['#CoronaUpdate', '#Covid19']	TweetDeck	False	14
DiarioLibre	na	2009-04-23T15:23:32Z	1185042	23738	321	True	2020-03-29T00:00:00Z	| #Coronavirus | EEUU aprueba test de coronavi...	['#Coronavirus', '#DL', '#DiarioLibre', '#Actu...	TweetDeck	False	15
lahoraecuador	na	2010-07-16T13:33:27Z	534729	1696	2384	False	2020-03-29T00:00:00Z	Debido a la emergencia sanitaria que vive el p...	['#Ecuador', '#Covid19']	TweetDeck	False	16
ABSCBNNews	na	2008-08-16T10:09:33Z	6767144	1075	1073	True	2020-03-29T00:00:00Z	Singapore donates 40,000 test kits to the Phil...	['#COVID19']	TweetDeck	False	17
dailyaajupdates	na	2016-12-05T08:39:18Z	1325	32	18	False	2020-03-29T00:00:00Z	سعودی حکام نے امسال حج کے حوالے سے افواہوں کو ...	['#DailyAAJUpdates', '#dailyaaj', '#COVID2019'...	TweetDeck	False	18
RadioNLNews	na	2010-07-27T16:17:02Z	6929	2137	498	False	2020-03-29T00:00:00Z	It's been a remarkable week for bold policy an...	['#COVID19', '#bcpoli', '#canpoli', '#Kamloops']	TweetDeck	False	19
ElSoldeSinaloa_	na	2014-05-25T04:53:23Z	2794	458	187	False	2020-03-29T00:00:00Z	#PorSiNoLoViste\nSe diseñó una estrategia para...	['#PorSiNoLoViste', '#UAS', '#Casa', '#Clases'...	TweetDeck	False	20
techreview_es	na	2009-01-26T22:41:17Z	27514	265	13589	True	2020-03-29T00:00:00Z	#LoMásLeídoMarzo | Esta 'app' del MIT te avisa...	['#LoMásLeídoMarzo', '#coronavirus', '#COVID19...	TweetDeck	False	21
imssjalcontigo	na	2014-06-05T13:46:44Z	3869	634	1530	False	2020-03-29T00:00:00Z	#PrevenciónCoronavirus | ¿Sabías que al estorn...	['#PrevenciónCoronavirus', '#EnfermedadesRespi...	TweetDeck	False	22
alaraby_ar	na	2014-03-10T11:35:38Z	936679	21	101	True	2020-03-29T00:00:00Z	مصر | بعد أن بقي ملقىً في الشارع لساعات أمام# ...	['#معاً_نعزل_كورونا', '#كورونا', 'مصر', '#COV...	TweetDeck	False	23
889Noticias	na	2009-05-06T21:09:11Z	262891	164	476	True	2020-03-29T00:00:00Z	El Secretario Nacional de la @ONU_es anunció l...	['#NuevaYork', '#COVID19', '#EstadosUnidos']	TweetDeck	False	24
SomosLJA	na	2009-05-02T20:35:21Z	20129	927	2307	False	2020-03-29T00:00:00Z	#LoMásVistoEnLJA | TREINTAÑEROS Y VEINTEAÑEROS...	['#LoMásVistoEnLJA', '#COVID19']	TweetDeck	False	25

Introducing Spark NLP

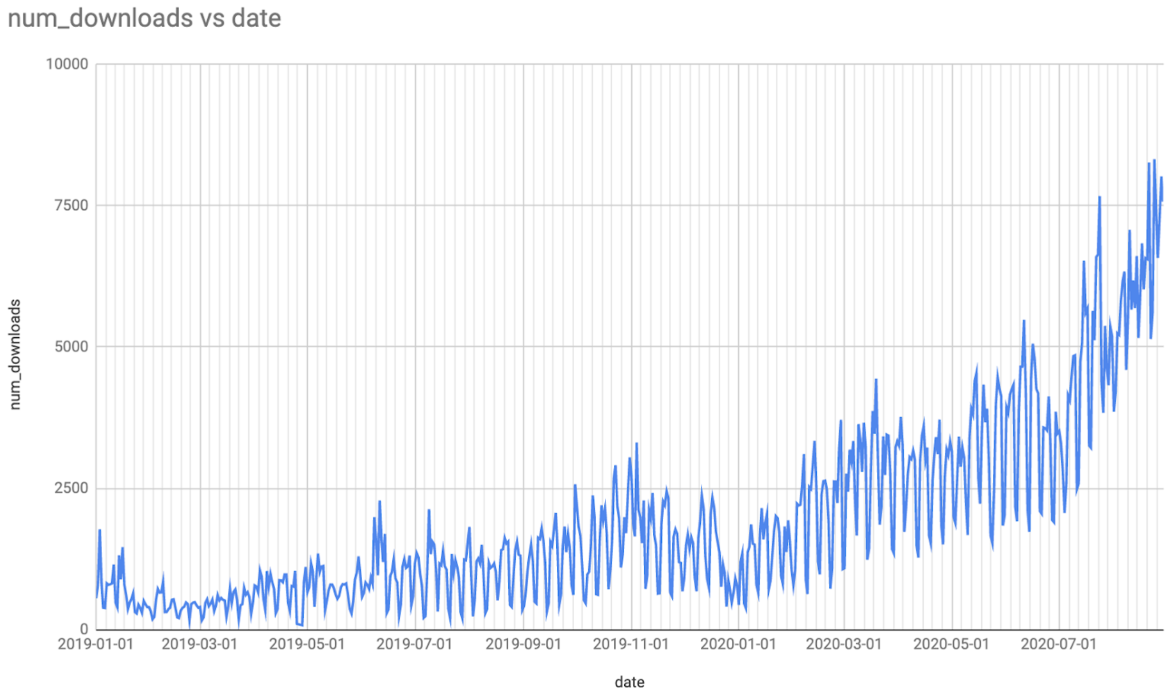
State of the art NLP:

1. **Accuracy**
2. **Speed**
3. **Scalability**

Open-Source Python, Java & Scala Libraries
200+ Pre-Trained Models & Pipelines
Vibrant: 26 new releases in 2018, 28 in 2019

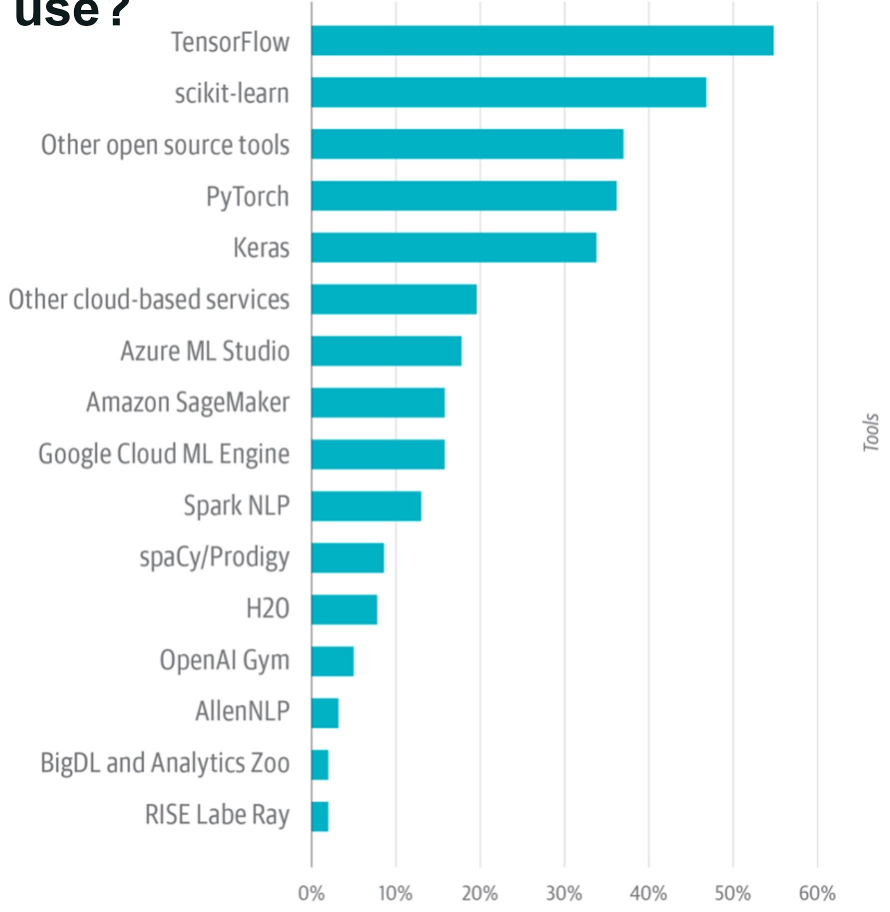
Daily ~ 20K
Monthly ~ 600K

PyPI link	https://pypi.org/project/spark-nlp
Total downloads	3,976,595
Total downloads - 30 days	656,474
Total downloads - 7 days	152,742

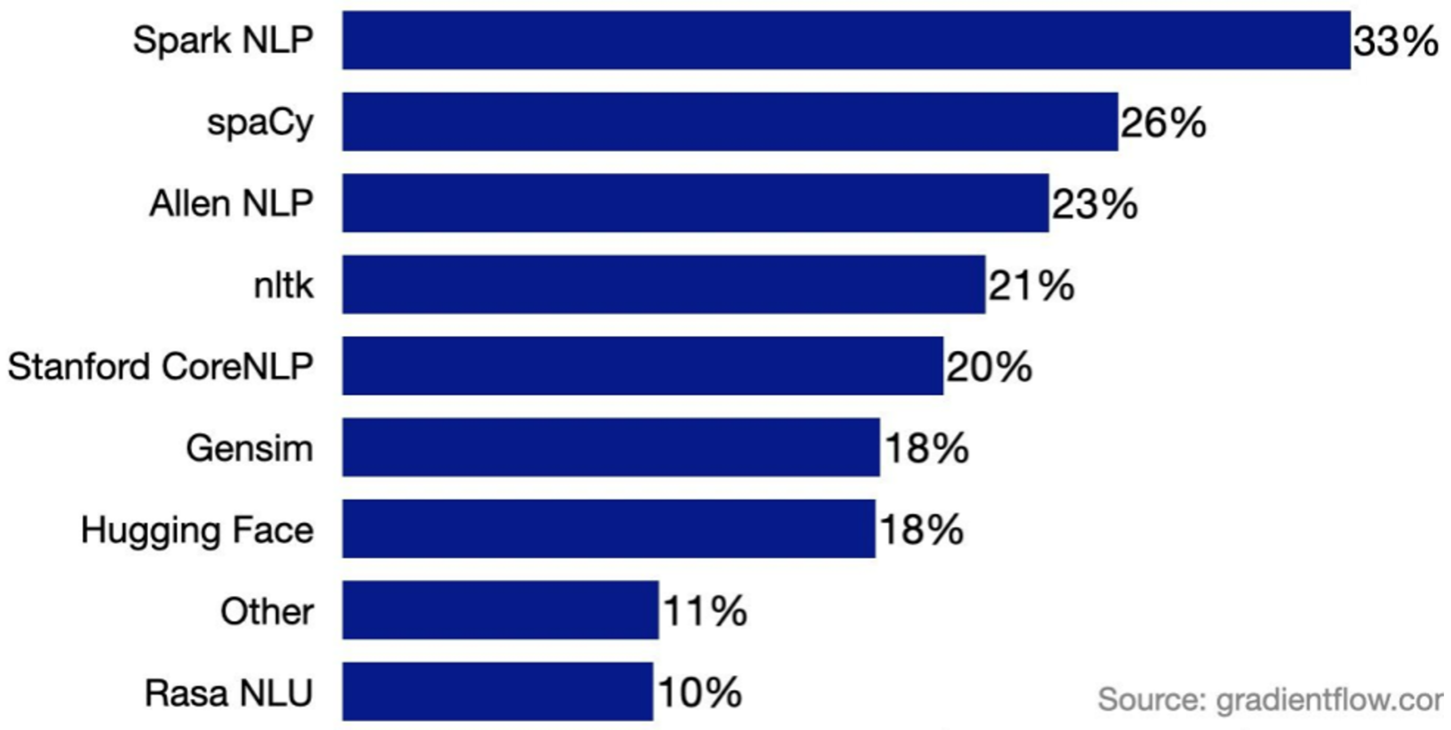


Spark NLP in Industry

Which of the following AI tools do you use?



Which NLP libraries does your organization use?



NLP Industry Survey by Gradient Flow,

an independent data science research & insights company, September 2020

Biomedical Named Entity Recognition at Scale

Veysel Kocaman
John Snow Labs Inc.
16192 Coastal Highway
Lewes, DE , USA 19958
veysel@johnsnowlabs.com

***Abstract*—Named entity recognition (NER) is a widely applicable natural language processing task and building block of question answering, topic modeling, information retrieval, etc. In the medical domain, NER plays a crucial role by extracting meaningful chunks from clinical notes and reports, which are then fed to downstream tasks like assertion status detection, entity resolution, relation extraction, and de-identification. Reimplementing a Bi-LSTM-CNN-Char deep learning architecture on top of Apache Spark, we present a single trainable NER model that obtains new state-of-the-art results on seven public biomedical benchmarks without using heavy contextual embeddings like BERT. This includes improving BC4CHEMD to 93.72% (4.1% gain), Species800 to 80.91% (4.6% gain), and JNLPBA to 81.29% (5.2% gain). In addition, this model is freely available within a production-grade code base as part of the open-source Spark NLP library; can scale up for training and inference in any Spark cluster; has GPU support and libraries for popular programming languages such as Python, R, Scala and Java; and can be extended to support other human languages with no code changes.**

I. Introduction

Electronic health records (EHRs) are the primary source of information for clinicians tracking the care of their patients. Information fed into these systems may be found in structured fields for which values are inputted electronically (e.g. laboratory test orders or results) [1] but most of the time information in these records is unstructured making it largely inaccessible

Spark NLP: Natural Language Understanding at Scale

Veysel Kocaman, David Talby
John Snow Labs Inc.
16192 Coastal Highway
Lewes, DE , USA 19958
eysel, david}@johnsnowlabs.com

Accurate Clinical and Biomedical Named Entity Recognition at Scale

Anonymous NAACL-HLT 2021 submission

Abstract

Named entity recognition (NER) is one of the most important building blocks of NLP tasks in the medical domain by extracting meaningful chunks from clinical notes and reports, which are then fed to downstream tasks like assertion status detection, entity resolution, relation extraction, and de-identification. Due to the growing volume of healthcare data in unstructured format, an increasingly important challenge is providing high accuracy implementations of state-of-the-art deep learning (DL) algorithms at scale. In this study, we introduce a production-grade clinical and biomedical NER algorithm based on a modified BiLSTM-CNN-Char DL architecture built on top of Apache Spark. This algorithm establishes new state-of-the-art accuracy on 7 of 8 well-known biomedical NER benchmarks and 3 clinical concept extraction challenges: 2010 i2b2/VA clinical concept extraction, 2014 n2c2 de-identification, and 2018 n2c2 medication extraction. Moreover, clinical NER models trained using this implemen-

Improving Clinical Document Understanding on COVID-19 Research with Spark NLP

Veysel Kocaman, David Talby
John Snow Labs Inc.
16192 Coastal Highway
Lewes, DE , USA 19958
{veysel, david}@johnsnowlabs.com

Abstract

Following the global COVID-19 pandemic, the number of scientific papers studying the virus has grown massively, leading to increased interest in automated literate review. We present a clinical text mining system that improves on previous efforts in three ways. First, it can recognize over 100 different entity types including social determinants of health, anatomy, risk factors, and adverse events in addition to other commonly used clinical and biomedical entities. Second, the text processing pipeline includes assertion status detection, to distinguish between clinical facts that are present, absent, conditional, or about someone other than the patient. Third, the deep learning models used are more accurate than previously available, leveraging an integrated pipeline of state-of-the-art pre-trained named entity recognition models, and improving on the previous best performing benchmarks for assertion status detection. We illustrate extracting trends and insights - e.g. most frequent disorders and symptoms, and most common vital signs and EKG findings – from the COVID-19 Open Research Dataset (CORD-19). The system is built using the Spark NLP library which natively supports scaling to use distributed clusters, leveraging GPU's, configurable and reusable NLP pipelines, healthcare-specific embeddings, and the ability to train models to support new entity types or human languages with no code changes.

be found in structured fields for which values are inputted electronically (e.g. laboratory test orders or results) (Liede et al. 2015) but most of the time information in these records is unstructured making it largely inaccessible for statistical analysis (Murdoch and Detsky 2013). These records include information such as the reason for administering drugs, previous disorders of the patient or the outcome of past treatments, and they are the largest source of empirical data in biomedical research, allowing for major scientific findings in highly relevant disorders such as cancer and Alzheimer's disease (Perera et al. 2014).

A primary building block in such text mining systems is named entity recognition (NER) - which is regarded as a critical precursor for question answering, topic modelling, information retrieval, etc (Yadav and Bethard 2019). In the medical domain, NER recognizes the first meaningful chunks out of a clinical note, which are then fed down the processing pipeline as an input to subsequent downstream tasks such as clinical assertion status detection (Uzuner et al. 2011), clinical entity resolution (Tzitzivacos 2007) and de-identification of sensitive data (Uzuner, Luo, and Szolovits 2007) (see Figure 1). However, segmentation of clinical and drug entities is considered to be a difficult task in biomedical NER systems because of complex orthographic structures of named entities

Peer-reviewed conference papers on Spark NLP NER

TRUSTED BY

Microsoft
Google
amazon
intel

IBM
databricks
verizon
indeed

Capital One
viacom
McKESSON
MERCK

Roche
selectdata
UiPath Reboot Work.
ASCO CANCER·LINQ DISCOVERY

cnrs
Imperial College London
Georgia Tech
STANFORD UNIVERSITY

AI
ARTIFICIAL INTELLIGENCE
EXCELLENCE
AWARD
2020

“John Snow Labs wins our best AI product or service award thanks to **exceptional success turning AI research into real & dependable systems for a global community**.”

“An open source project, tool, or contribution that **significantly advances the state of data science** is recognized with this award.”

HEALTHCARE ANALYTICS
PROVIDER OF THE YEAR 2020
Healthcare Tech OUTLOOK

“By all accounts, John Snow Labs has created **the most accurate software in history** to extract facts from unstructured text.”

OFFICIALLY SUPPORTED RUNTIMES

Java
python™
Scala
R
NVIDIA®
intel® XEON® inside™
APACHE Spark™
databricks®
CLOUDERA
Azure
aws

Spark NLP & NLU

- A single **unified** library for all your **NLP/NLU** needs
- 1000+ Models,
- 200+ Languages
- 1 Line of code
- Active **community** on **Slack** and **GitHub**

NLP Feature	NLU / Spark NLP	spaCy	NLTK	CoreNLP	Hugging Face
Tokenization	**Yes**	Yes	Yes	Yes	Yes
Sentence segmentation	**Yes**	Yes	Yes	Yes	**No**
Steeming	**Yes**	Yes	Yes	Yes	**No**
Lemmatization	**Yes**	Yes	Yes	Yes	**No**
POS tagging	**Yes**	Yes	Yes	Yes	**No**
Entity recognition	**Yes**	Yes	Yes	Yes	Yes
Dep parser	**Yes**	Yes	Yes	Yes	**No**
Text matcher	**Yes**	Yes	**No**	**No**	**No**
Date matcher	**Yes**	**No**	**No**	**No**	**No**
Sentiment detector	**Yes**	**No**	Yes	Yes	Yes
Text classification	**Yes**	Yes	Yes	**No**	Yes
Spell checker	**Yes**	**No**	**No**	**No**	**No**
Language detector	**Yes**	**No**	**No**	**No**	**No**
Keyword extraction	**Yes**	**No**	**No**	**No**	**No**
Pretrained models	**Yes**	Yes	Yes	Yes	Yes
Trainable models	**Yes**	Yes	Yes	Yes	Yes

200+ Supported Languages

How does it work?

```
model= nlu.load(model)
```

- Returns a nlu pipeline object

```
model.predict(data)
```

- Returns a pandas DF

EMOTION DETECTION

```
nlu.load('emotion').predict('I love NLU!')
```

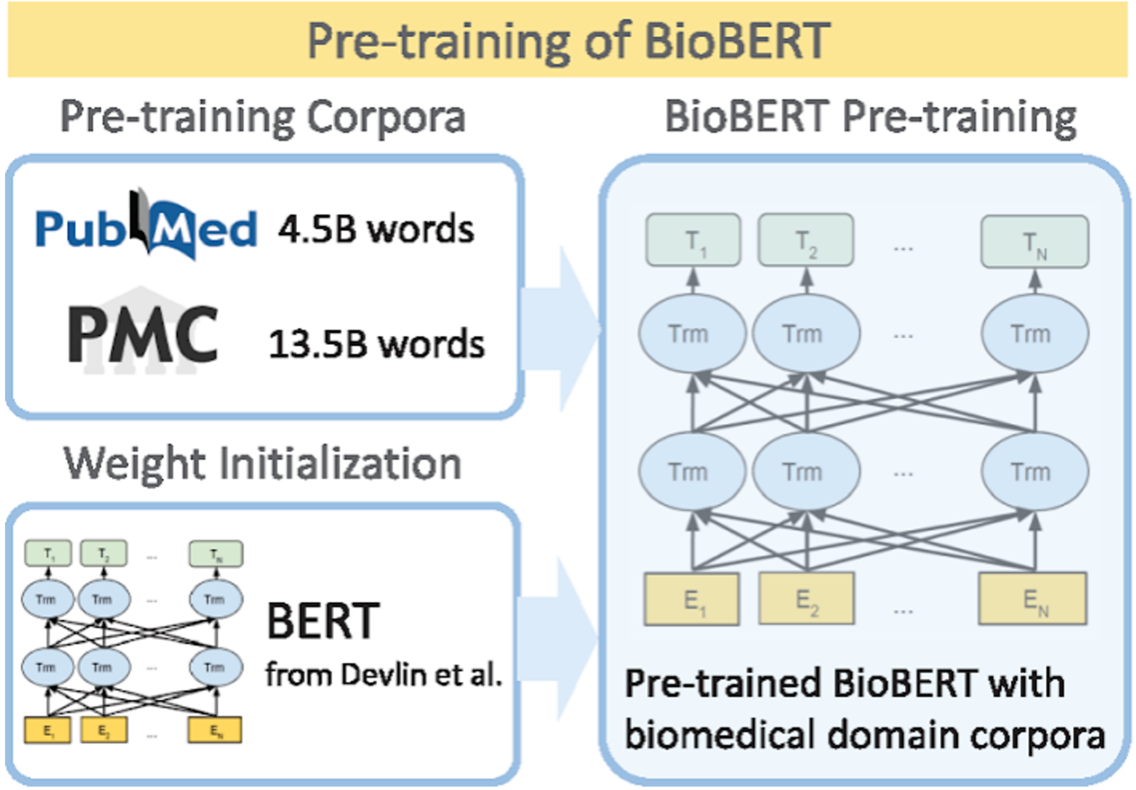
sentence_embeddings	category_sentence	category_surprise	category_sadness	category_joy	category_fear	sentence	category	id
[0.027570432052016258, -0.052647676318883896, ...]	0	0.012899903	0.0015578865	0.9760173	0.0095249	I love NLU!	joy	1

Clinical Word Embeddings

Clinical Glove (200d)	ICDO Glove (200d) (512, 1024)	Bio BERT	Clinical BERT
PubMed + PMC	PubMed + ICD10 UMLS + MIMIC III	Pubmed + PMC	Fine tuned Pubmed + PMC + Discharge summaries

PubMed abstracts and PMC full-text articles

https://www.nlm.nih.gov/bsd/difference.html



Clinical Named Entity Recognition

Pretrained NER Models in Spark NLP

PROBLEM × TEST × TREATMENT × DURATION × EVIDENTIAL × TREATMENT × FREQUENCY ×
OCCURRENCE × TEST × TIME × PROBLEM × DATE × CLINICAL_DEPT × Disease ×
DrugChem × Immaterial_anatomic... × Developing_anatomic... × Pathological_formation × Organ ×
Organism_subdivision × Cellular_component × Multi × Tissue × Anatomical_system ×
Organism_substance × Cell × DURATION × ROUTE × FREQUENCY × DOSAGE × DRUG ×
FORM × STRENGTH × Immaterial_anatomic... × Developing_anatomic... × Pathological_formation ×
Cancer × Organism × Organ × Organism_subdivision × Cellular_component × Amino_acid ×
Multi × Tissue × Anatomical_system × Gene_or_gene_product × Simple_chemical ×
Organism_substance × Cell × Weight × Drug_Name × Negation × Procedure ×
Causative_Agents_(Vi... × O2_Saturation × Route × Temperature × Procedure_Name ×
Substance_Name × Symptom_Name × Respiratory_Rate × Dosage × Name × Gender ×
Pulse_Rate × Lab_Result × Lab_Name × Maybe × Allergenic_substance × Age × Frequency ×
Diagnosis × Modifier × Section_Name × Blood_Pressure × MEDICATION × CAD ×
HYPERLIPIDEMIA × FAMILY_HIST × DIABETES × SMOKER × OBESE × PHI ×
HYPERTENSION × RNA × cell_type × protein × cell_line × DNA × CHEM ×
Organization × Body_System × Professional_or_Occu... × Clinical_Attribute × Indicator,_Reagent,_... ×
Organic_Chemical × Anatomical_Structure × Organism_Attribute × Food × Body_Part,_Organ,_or... ×
Biologic_Function × Medical_Device × Tissue × Disease_or_Syndrome × Chemical ×
Neoplastic_Process × Health_Care_Activity × Body_Location_or_Re... × Qualitative_Concept ×
Injury_or_Poisoning × Population_Group × Geographic_Area × Manufactured_Object × Mental_Process ×
Group × Daily_or_Recreational... × Therapeutic_or_Preve... × Research_Activity × Cell ×
Pathologic_Function × Mammal × Quantitative_Concept × Spatial_Concept × Pharmacologic_Subst... ×
Diagnostic_Procedure × Eukaryote × Cell_Component × Prokaryote × Molecular_Biology_R... ×
Substance × Mental_or_Behavioral... × Molecular_Function × Fungus × Virus ×
Laboratory_Procedure × Nucleotide_Sequence × Body_Substance × Plant × Amino_Acid,_Peptide... ×
Genetic_Function × Nucleic_Acid,_Nucle... × Biomedical_or_Denta... × Gene_or_Genome ×
Sign_or_Symptom × HP × GO × HP × GENE ×

The patient was prescribed 1 capsule of Advil for 5 days . He was seen by the endocrinology service and she was discharged on 40 units of insulin glargine at night , 12 units of insulin lispro with meals , and metformin 1000 mg two times a day . It was determined that all SGLT2 inhibitors should be discontinued indefinitely fro 3 months .

Color codes:FREQUENCY, DOSAGE, DURATION, DRUG, FORM, STRENGTH,

Posology NER

No findings in urinary system , skin color is normal , brain CT and cranial checks are clear . Swollen fingers and eyes . Extensive stage small cell lung cancer . Chemotherapy with carboplatin and etoposide . Left scapular pain status post CT scan of the thorax .

Color codes:Organ, Organism_subdivision, Organism_substance, Pathological_formation, Anatomical_system,

Anatomy NER

A . Record date : 2093-01-13 , David Hale , M.D . , Name : Hendrickson , Ora MR . # 7194334 Date : 01/13/93 PCP : Oliveira , 25 years-old , Record date : 2079-11-09 . Cocke County Baptist Hospital . 0295 Keats Street

Color codes:STREET, DOCTOR, AGE, HOSPITAL, PATIENT, DATE, MEDICALRECORD,

PHI NER

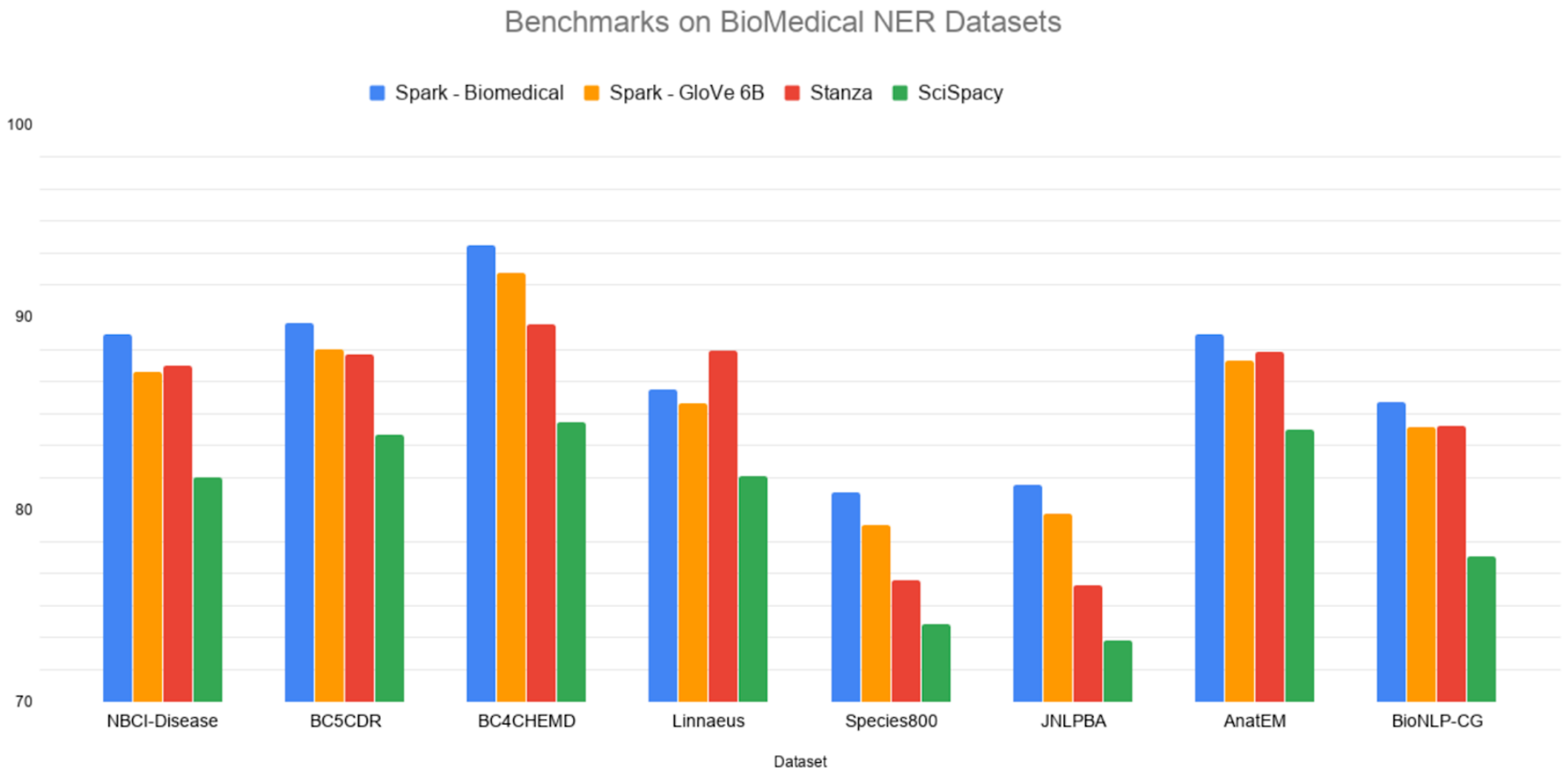
NER_demo.py

```
# Extract Various entities from the medical domain
nlu.load('med_ner.ade')   # Drug Adverse Events
nlu.load('med_ner.anatomy')
nlu.load('med_ner.aspect_sentiment')
nlu.load('med_ner.bacterial_species')
nlu.load('med_ner.bionlp')
nlu.load('med_ner.cancer')
nlu.load('med_ner.cellular')
nlu.load('med_ner.chemicals')
nlu.load('med_ner.chemprot')
nlu.load('med_ner.clinical')
nlu.load('med_ner.diseases')
nlu.load('med_ner.drugs')
nlu.load('med_ner.events_healthcre')
nlu.load('med_ner.human_phenotype')
nlu.load('med_ner.measurements')
nlu.load('med_ner.medmentions')
nlu.load('med_ner.posology')
nlu.load('med_ner.radiology')
nlu.load('med_ner.risk_factors')
nlu.load('med_ner.i2b2')
nlu.load('med_ner.tumour')
```

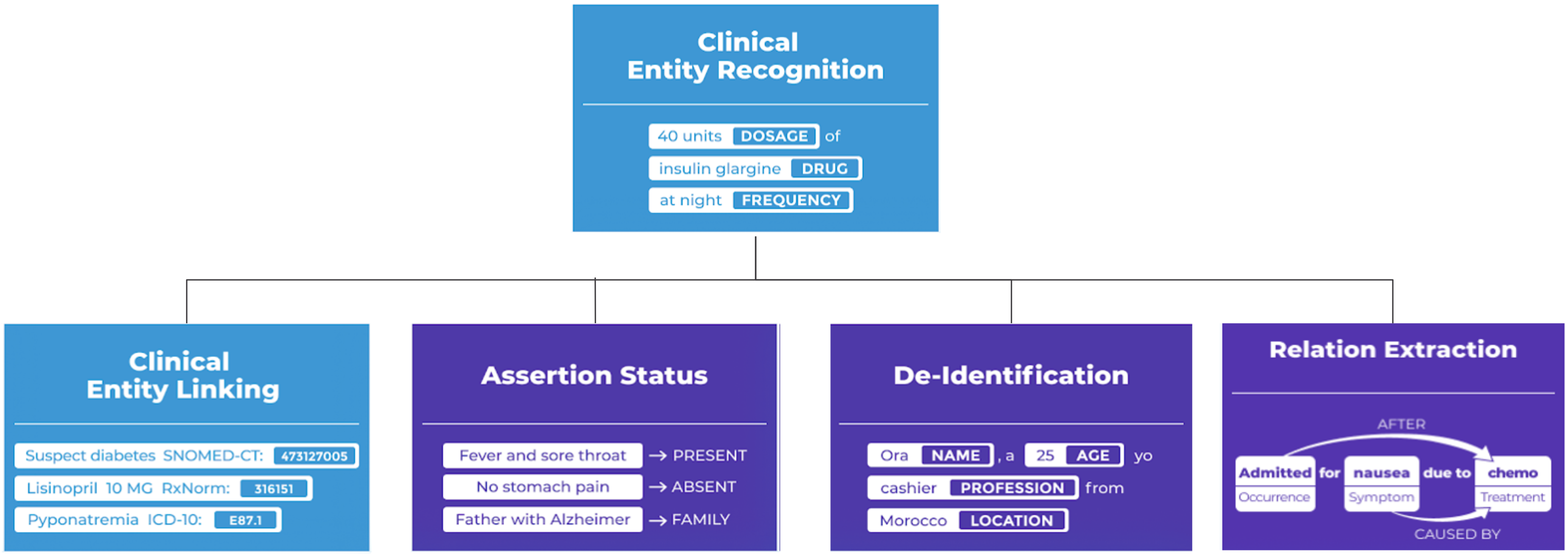
Spark NLP NerDL

The best NER score in production

Dataset	Entities	Spark - Biomedical	Spark - GloVe 6B	Stanza	SciSpacy
NBCI-Disease	Disease	**89.13**	87.19	87.49	81.65
BC5CDR	Chemical, Disease	**89.73**	88.32	88.08	83.92
BC4CHEMD	Chemical	**93.72**	92.32	89.65	84.55
Linnaeus	Species	86.26	85.51	**88.27**	81.74
Species800	Species	**80.91**	79.22	76.35	74.06
JNLPBA	5 types in cellular	**81.29**	79.78	76.09	73.21
AnatEM	Anatomy	**89.13**	87.74	88.18	84.14
BioNLP13-CG	16 types in Cancer Genetics	**85.58**	84.3	84.34	77.6

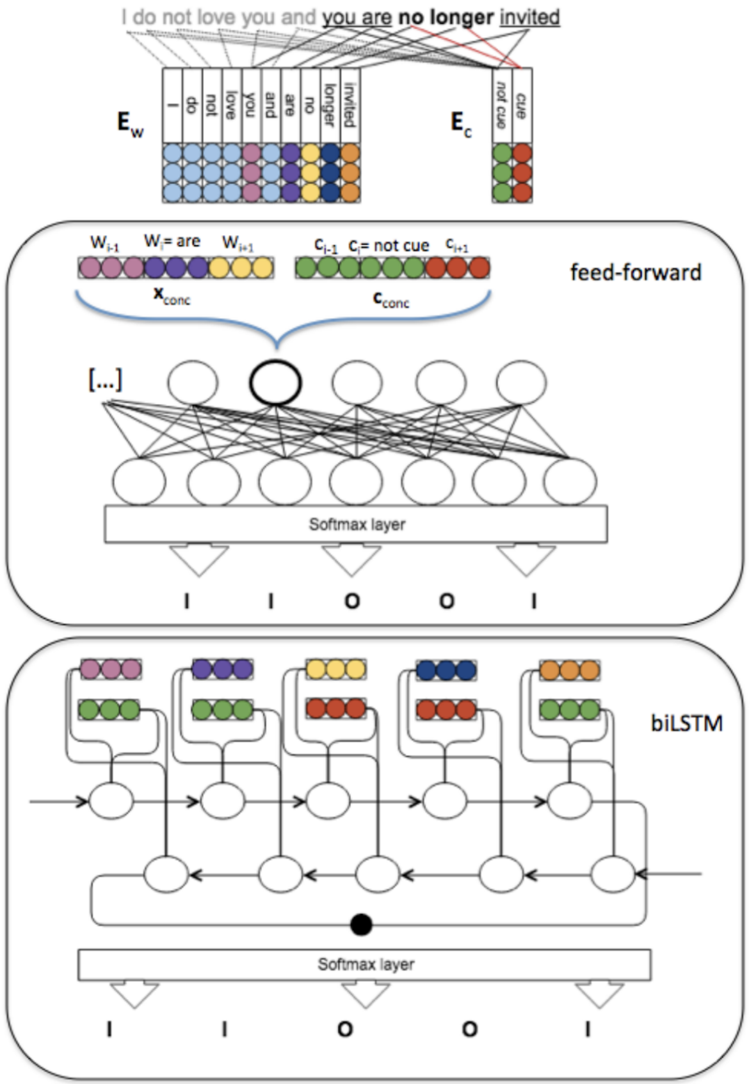


NER in Healthcare



Clinical Assertion Model

Prescribing sick days due to diagnosis of **influenza**.	*Present*
41 yo man with CRFs of DM Type II, high cholesterol, smoking history, family hx, HTN p/w episodes of **atypical CP x 1 week**, with rest and exertion.	*Conditional*
Jane's **RIDT** came back clean.	*Absent*
Jane is at risk for **flu** if she's not vaccinated.	*Hypothetical*
There was a **dense hemianopsia** on the left side.	*Present*



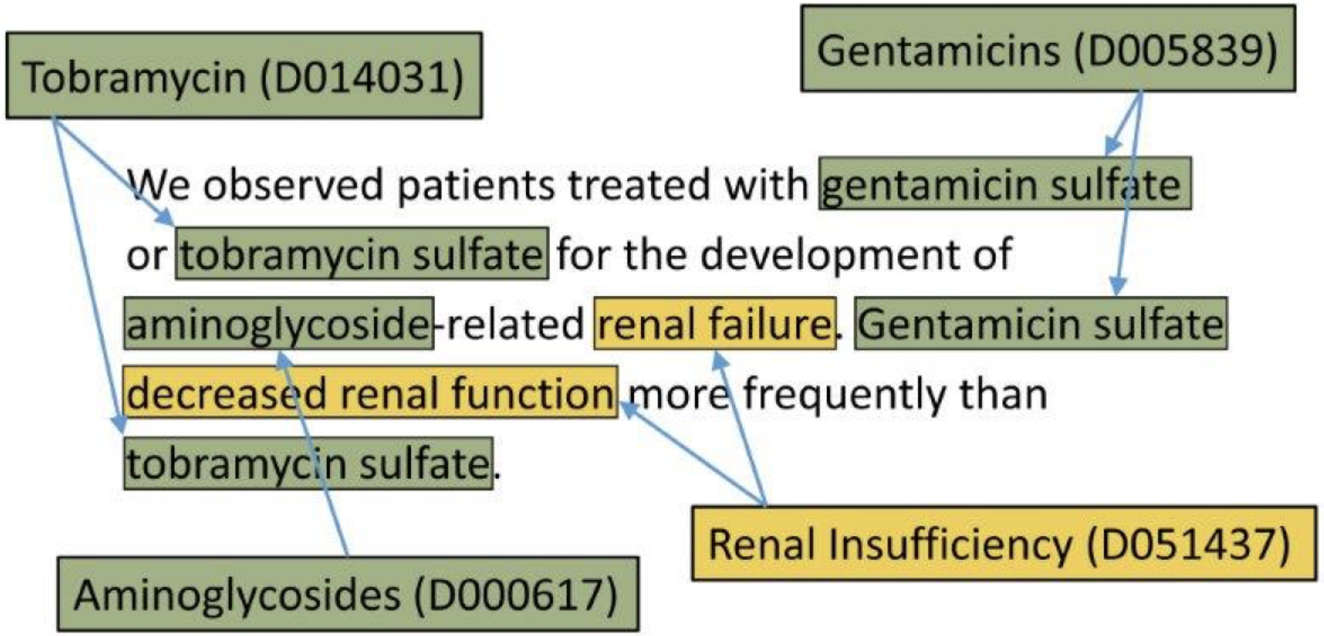
F-Score	Dataset	Task
94.17%	4th i2b2/VA	Disease & problem norm.

"Neural Networks For Negation Scope Detection", Fancellu et al., In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016.

scope of negation: given a negative instance, to identify which tokens are affected by negation assert_demo.py

```
# Assert statuses of detected entities
nlu.load('<medical_ner_model> assert').predict('The patient has no cancer')
```

Entity Resolution



"CNN-based ranking for biomedical entity normalization".

Li et al., *BMC Bioinformatics*, October 2017.

F-Score	Dataset	Task
90.30%	ShARe / CLEF	Disease & problem norm.
92.29%	NCBI	Disease norm. in literature

```
+---------+----------------------------------------------------
|codes    |description
+---------+----------------------------------------------------
|17473003 |Cecotomy
|17473003 |Cecotomy (procedure)
|304587000|Excision of colonic pouch
|304587000|Excision of colonic pouch (procedure)
|87279008 |Excision of lesion of colon
|174117007|Excision of lesion of colon NEC
|174117007|Excision of lesion of colon NEC (procedure)
|87279008 |Excision of lesion of colon (procedure)
|276190007|Ileocolic resection
|276190007|Ileocolic resection (procedure)
|43075005 |Partial resection of colon
|43075005 |Partial resection of colon (procedure)
|428305005|History of partial resection of colon (situation)
|428305005|History of partial resection of colon
|444165004|Partial resection of colon and resection of terminal
|738552004|Partial resection of colon with stoma (procedure)
|738552004|Partial resection of colon with stoma
|84952009 |Resection of colon for interposition
|84952009 |Resection of colon for interposition (procedure)
|445884009|Wedge resection of colon
+---------+----------------------------------------------------
only showing top 20 rows
```

Assigns a **ICD10** (International Classification of Diseases version 10) code to chunks identified as "PROBLEMS" by the NER Clinical Model

Entity Resolution - RxNorm

the patient was prescribed 1 capsule **DRUG** of advil **DRUG** for 5 days **DURATION** . he was seen by the endocrinology service and she was discharged on 40 units **DRUG** of insulin glargine **DRUG** at night **FREQUENCY** , 12 units **DRUG** of insulin lispro **DRUG** with meals **FREQUENCY** , and metformin 1000 mg **DRUG** two times a day **FREQUENCY** . it was determined that all sglt2 inhibitors **DRUG** should be discontinued indefinitely .

advil : DRUG

	rxnorm_code	description	distance
0	153010	advil	0
1	669348	advate	0.0417

insulin lispro : DRUG

	rxnorm_code	description	distance
0	86009	insulin lispro	0
1	1157461	insulin lispro injectable product	0.0743

insulin glargine : DRUG

	rxnorm_code	description	distance
0	274783	insulin glargine	0.0000
1	1157459	insulin glargine injectable product	0.0653

metformin 1000 mg : DRUG

	rxnorm_code	description	distance
0	316255	metformin 1000 mg	0.0000
1	860995	metformin hydrochloride 1000 mg	0.0445

Entity Resolution - Snomed / ICD-10

a 28-year-old female with a history of gestational diabetes mellitus **PROBLEM** diagnosed eight years prior to presentation and subsequent type two diabetes mellitus **PROBLEM** (t2dm **PROBLEM**), one prior episode of htg-induced pancreatitis **PROBLEM** three years prior to presentation , associated with an acute hepatitis **PROBLEM** , and obesity **PROBLEM** with a body mass index **PROBLEM** (bmi) of 33.5 kg/m2 , presented with a one-week history of polyuria **PROBLEM** , polydipsia **PROBLEM** , poor appetite **PROBLEM** , and vomiting **PROBLEM** .

gestational diabetes mellitus : PROBLEM

	snomed_code	description	distance	athena_concept_id	domain_id	concept_class_id	ICD10CM_mapping
0	11687002	gestational diabetes mellitus	0.0000	4024659	Condition	Clinical Finding	O24.429, O24.439, O24.414, O24.419, O24.4, O24.410
1	40791000119105	postpartum gestational diabetes mellitus	0.0423	45757789	Condition	Clinical Finding	O24.4, O24.439

obesity : PROBLEM

	snomed_code	description	distance	athena_concept_id	domain_id	concept_class_id	ICD10CM_mapping
0	414916001	obesity	0.0000	433736	Condition	Clinical Finding	E66.9
1	414915002	obese	0.0264	4215968	Observation	Clinical Finding	Z68.41, E66.9, E66.8

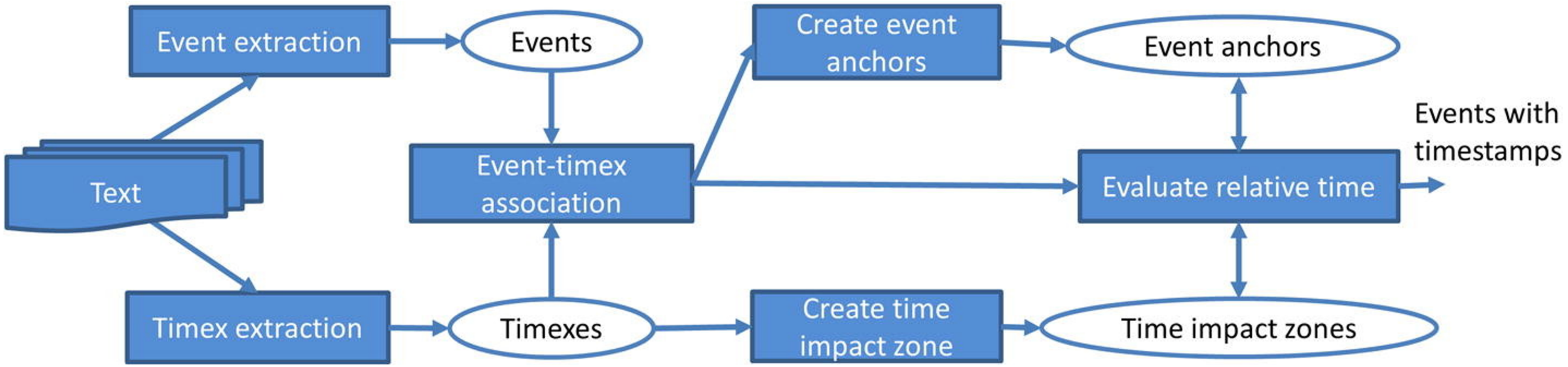
resolve_demo.py

```
## Resolve entities to various International billable codes and standards
nlu.load('<medical_ner_model>  resolve.cpt')
nlu.load('<medical_ner_model>  resolve.hcc')
nlu.load('<medical_ner_model>  resolve.10cm')
nlu.load('<medical_ner_model>  resolve.10pcs')
nlu.load('<medical_ner_model>  resolve.icdo')
nlu.load('<medical_ner_model>  resolve.rxciui')
nlu.load('<medical_ner_model>  resolve.rxnorm')
nlu.load('<medical_ner_model>  resolve.snomed')
nlu.load('<medical_ner_model>  resolve_chunk.athena')
```

Relation Extraction

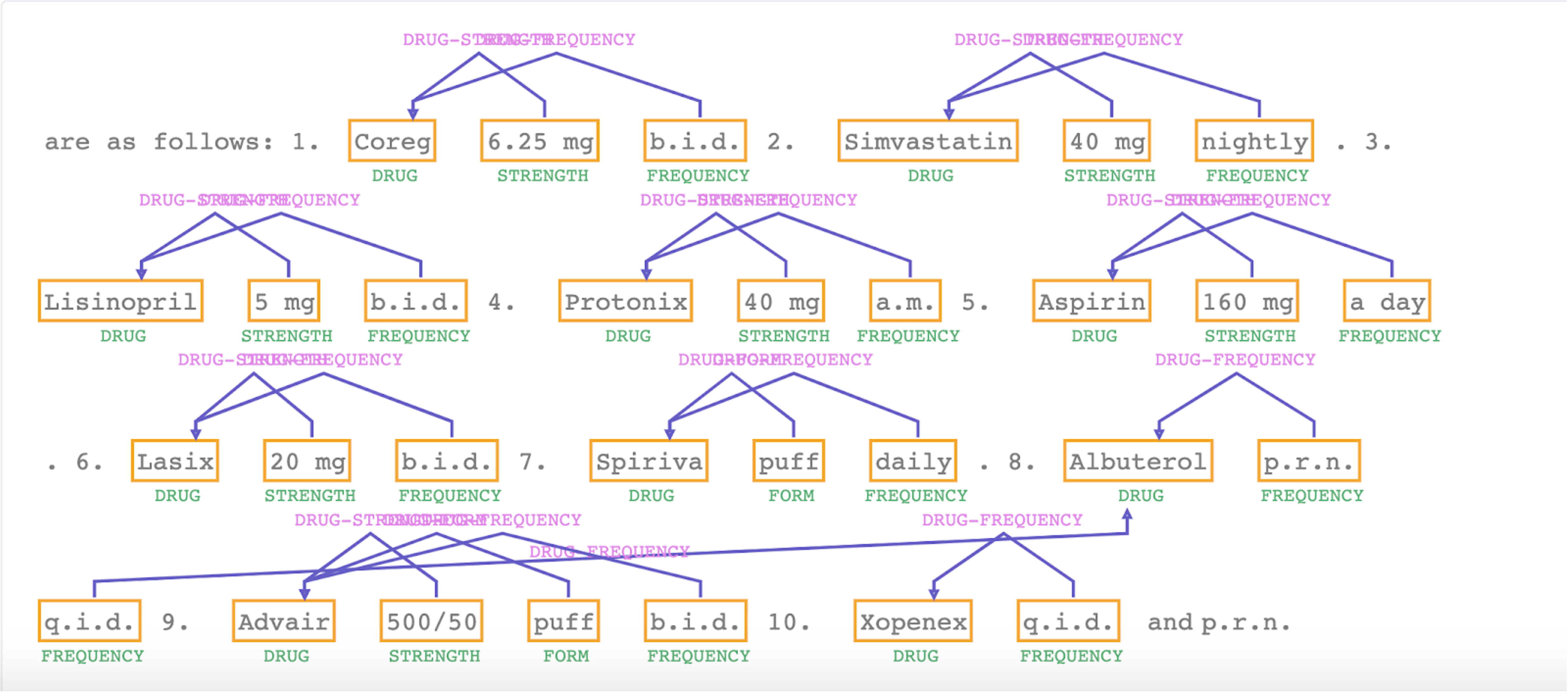
On 5/21/99, the infant received her 1st dose of vaccine A and her 2nd injection of vaccine B. The infant began vomiting and having diarrhea 5 days later. She was taken to the local ER where evaluation was ""non-diagnostic""...

Feature	Type	Date
Vaccine A	Vaccine	1999-05-21
Vaccine B	Vaccine	1999-05-21
Vomiting	Symptom	1999-05-26
Diarrhea	Symptom	1999-05-26
...	...	...



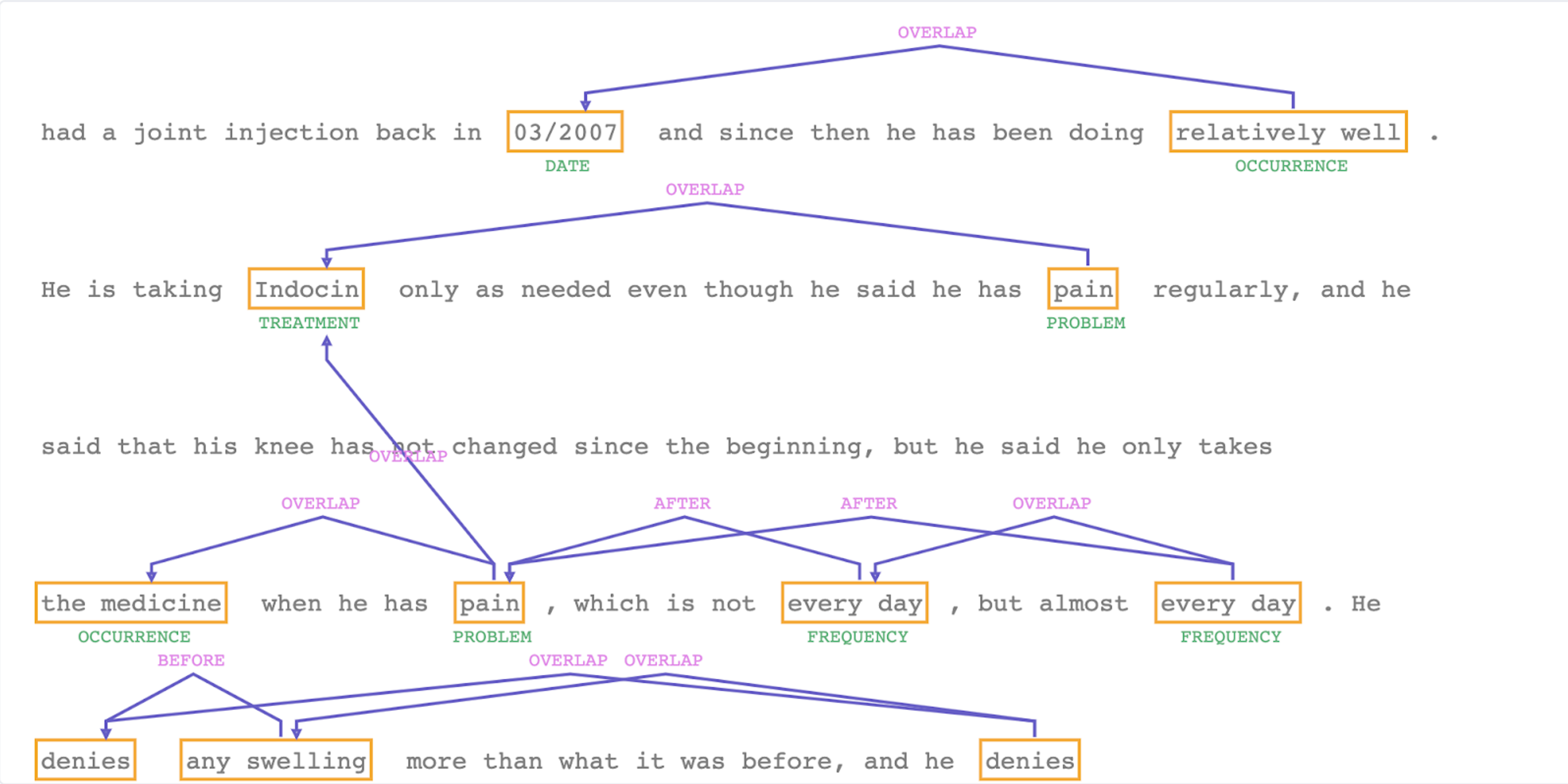
Relation Extraction

Posology



Relation Extraction

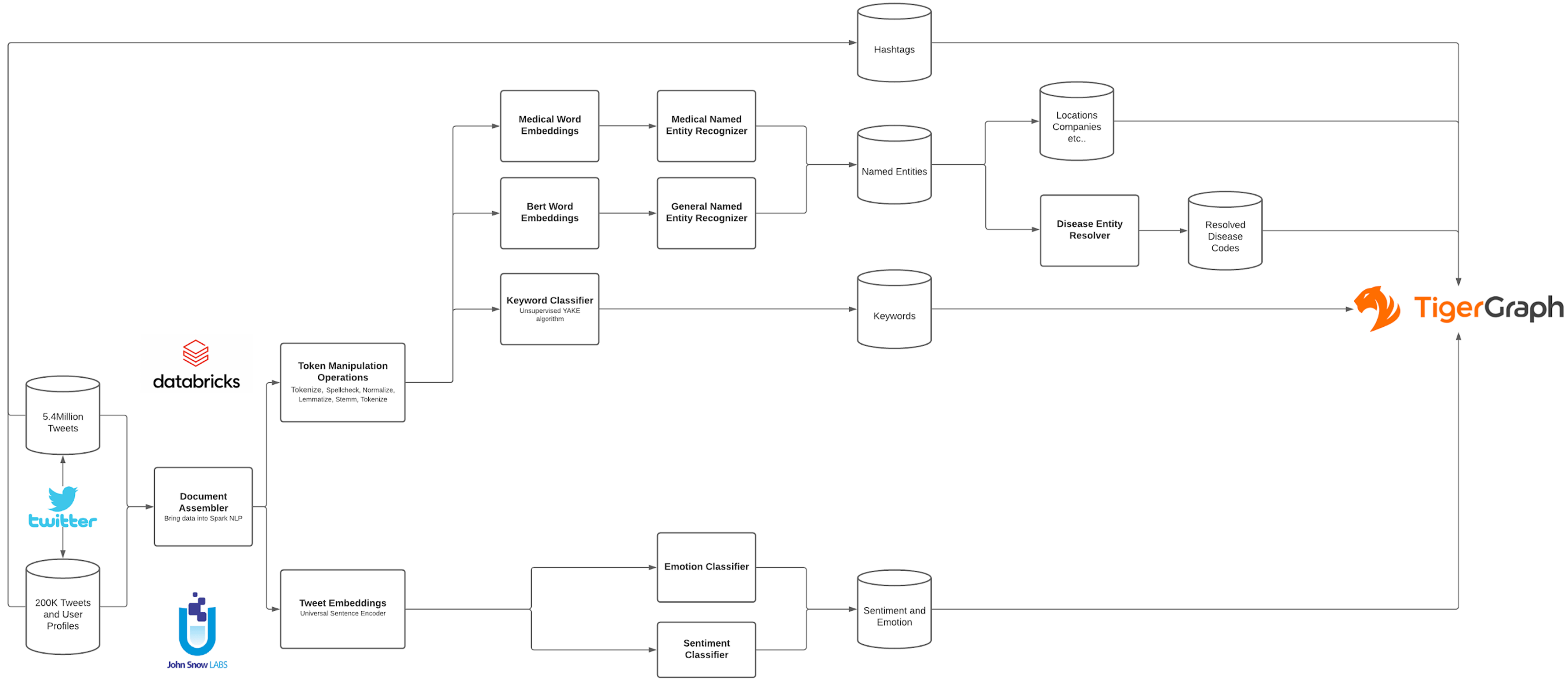
Temporal Events



relation_demo.py

```
# Extract relation between detected entities
nlu.load('<medical_ner_model>  relation.bodypart')
nlu.load('<medical_ner_model>  relation.chemprot')
nlu.load('<medical_ner_model>  relation.clinical')
nlu.load('<medical_ner_model>  relation.date')
nlu.load('<medical_ner_model>  relation.drug_drug_interaction')
nlu.load('<medical_ner_model>  relation.humen_phenotype_gene')
nlu.load('<medical_ner_model>  relation.temporal_events')
```

The NLU COVID data extraction for Graph NLU



Demo part 1

https://github.com/JohnSnowLabs/nlu/tree/master/examples/we
binars_conferences_etc/graph_ai_summit

Demo - Deep Analytics with graph

Spark NLP and NLU : Apache License 2.0

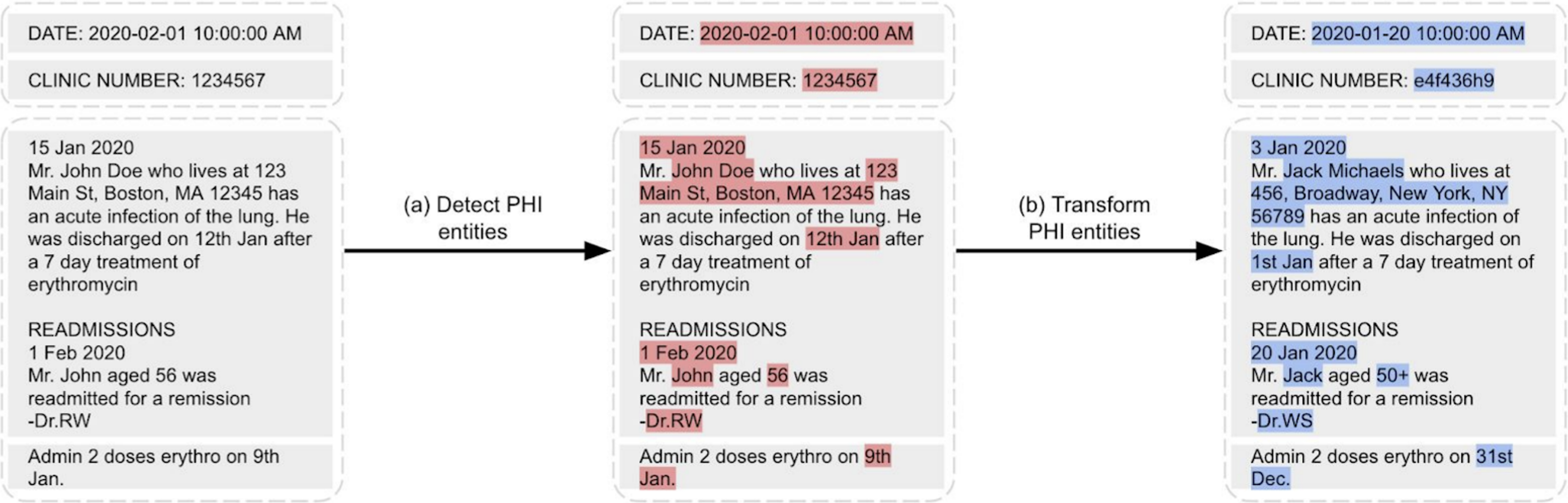
```
# Multiple binary sentiment classifiers trained on various datasets
nlu.load('classify.sentiment').predict('I love NLU and Python WebDev Conf 2021!')
nlu.load('classify.sentiment.imdb').predict('The Matrix was a pretty good movie')
nlu.load('classify.sentiment.twitter').predict('@elonmusk Tesla stock price is too high imo')

# Translate between 200 languages
nlu.load('en.translate_to.zh').predict('NLU can translate between 200 languages!')

# Spellchecking
nlu.load('spell').predict('I liek to live dangertus!')

# Extract Named Entities
nlu.load('ner').predict('Donald Trump and John Biden dont share many oppinions')

# Unsupervised Keyword Extraction
nlu.load('yake').predict('Weights extract keywords withouth requiring weights!')

# Over 50+ classifiers on various problems
nlu.load('classify.emotion').predict('He was suprised by the diversity of NLU')
nlu.load('classify.spam').predict('Hello you are the heir to a 100 Million fortune!')
nlu.load('classify.fakenews').predict('Unicorns landed on mars!')
nlu.load('classify.sarcasm').predict('love the teachers who give exams the day after halloween')
nlu.load('en.classify.question').predict('How expensive is the Watch?')
nlu.load('en.classify.toxic').predict('You are to stupid')
nlu.load('classify.cyberbullying').predict('Women belong in the kitchen!') #sorry

# Get BERTology and Transformer Embeddings for Sentences and Words
nlu.load('bert').predict('BERTolgy Word embeddings!')
nlu.load('bert elmo albert glove').predict('Multiple BERTolgy Word embeddings!')
nlu.load('embed_sentence.bert ').predict('BERTolgy Sentence embeddings!')

# Text cleaning and Pre-Processing
nlu.load('lemmatize').predict('Get me the lemmatized version of a string')
nlu.load('normalize').predict('Get me the lemmatized version of a string')
nlu.load('clean').predict('Get me the lemmatized version of a string')

# Grammatical Parts of Speech
nlu.load('pos').predict('Extract Parts of Speech')
```

- Tokenization
- Sentence Detector
- Stop Words Removal
- Normalizer
- Stemmer
- Lemmatizer
- NGrams
- Regex Matching
- Text Matching
- Chunking
- Date Matcher
- **Part-of-speech** tagging
- **Dependency** parsing
- **Sentiment** Detection (ML models)
- **Spell Checker** (ML and DL models)
- Word **Embeddings**

- **BERT** Embeddings
- **ELMO** Embeddings
- **ALBERT** Embeddings
- **XLNet** Embeddings
- **Universal Sentence Encoder**
- **BERT** Sentence Embeddings
- Sentence Embeddings
- Chunk Embeddings
- Unsupervised **keywords extraction**
- **Language Detection** & Identification
- **Multi-class** Text Classification
- **Multi-label** Text Classification
- Multi-class **Sentiment Analysis**
- **Named entity recognition**
- Easy **TensorFlow** integration
- Full integration with Spark ML functions
- **+250 pre-trained models** in 46 language
- **+90 pre-trained pipelines** in 13 language

De-Identification

* Identifies potential pieces of content with personal information about patients and remove them by replacing with semantic tags.



De-Identification

* Identifies potential pieces of content with personal information about patients and remove them by replacing with semantic tags.

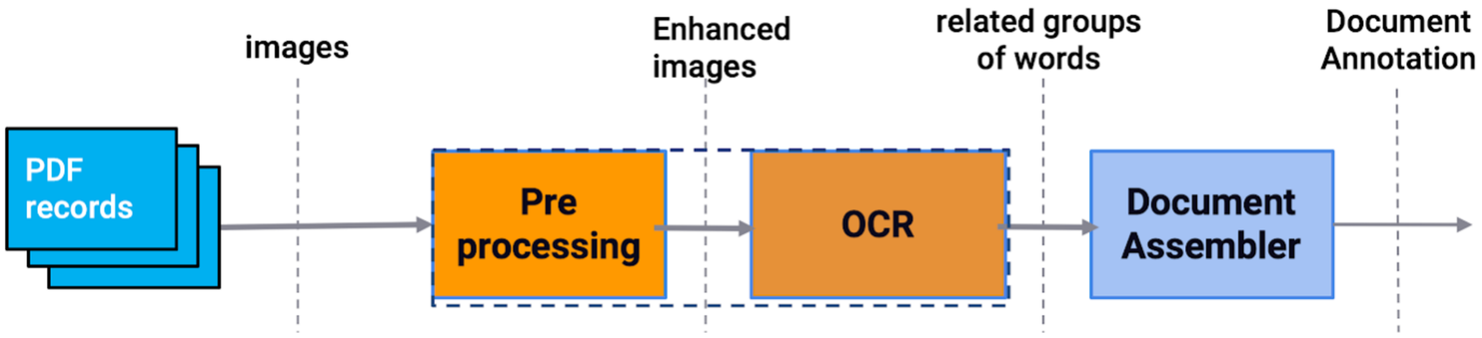
Record date : 2093-01-13 **DATE** , David Hale **DOCTOR** , M.D . , Name : Hendrickson , Ora **PATIENT** MR . # 7194334 **MEDICALRECORD** Date : 01/13/93 **DATE** PCP : Oliveira **DOCTOR** , 25 **AGE** years-old , Record date : 2079-11-09 **DATE** . Cocke County Baptist Hospital **HOSPITAL** . 0295 Keats Street **STREET** . Phone (302) 786-5227 **PHONE** .

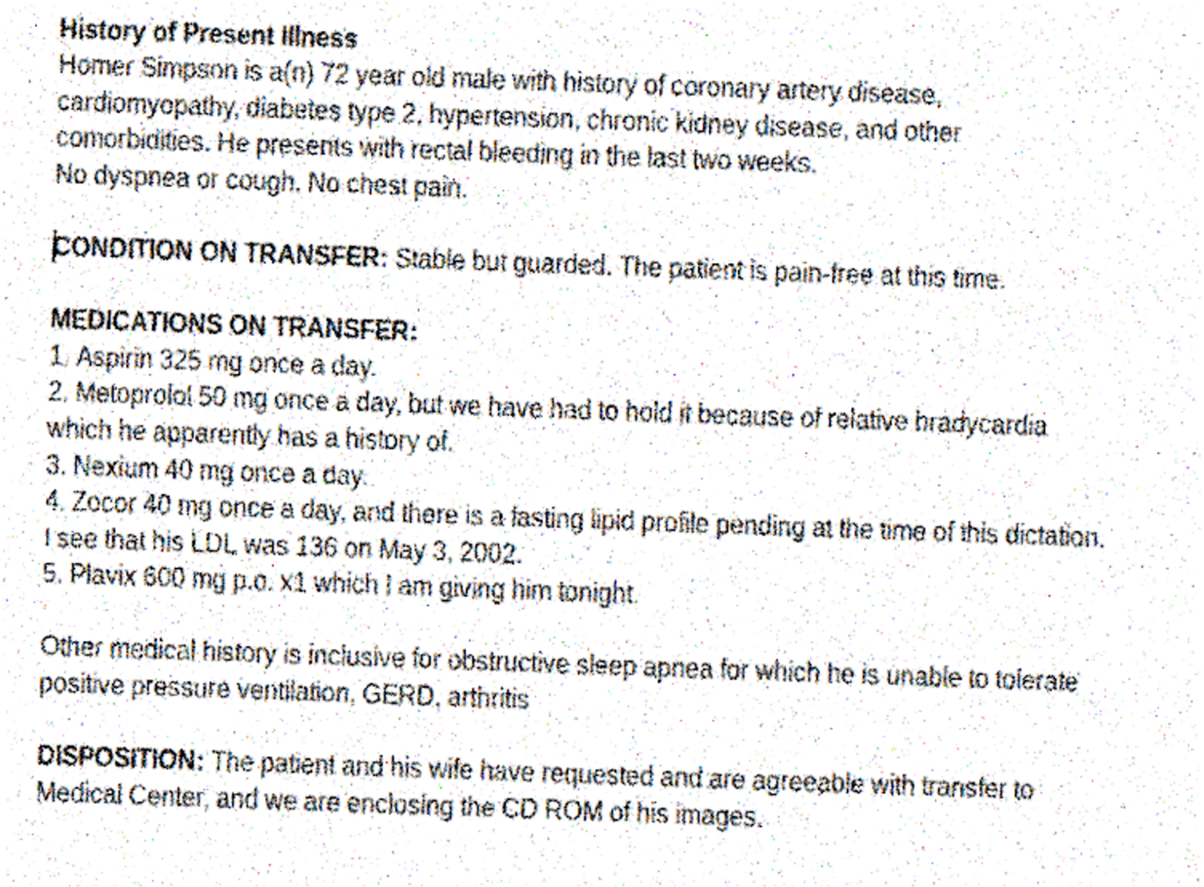
	sentence	deidentified
0	A .	A .
1	Record date : 2093-01-13 , David Hale , M.D .	Record date : <DATE> , <NAME> , M.D .
2	, Name : Hendrickson , Ora MR .	, Name : <NAME> MR .
3	# 7194334 Date : 01/13/93 PCP : Oliveira , 25 years-old , Record date : 2079-11-09 .	# <ID> Date : <DATE> PCP : <NAME> , <AGE> years-old , Record date : <DATE> .
4	Cocke County Baptist Hospital .	<LOCATION> .
5	0295 Keats Street.	<LOCATION>.
6	Phone (302) 786-5227.	Phone <CONTACT>.

deid_demo.py

```
# De-Identify and anonymize detected entities
nlu.load('<medical_ner_model> en.med_ner.deid')
```

Spark OCR




History of Present Illness
Homer Simpson is a(n) 72 year old male with history of coronary artery disease, cardiomyopathy, diabetes type 2, hypertension, chronic kidney disease, and other comorbidities. He presents with rectal bleeding in the last two weeks.
No dyspnea or cough. No chest pain.

CONDITION ON TRANSFER: Stable but guarded. The patient is pain-free at this time.

MEDICATIONS ON TRANSFER:
1. Aspirin 325 mg once a day.
2. Metoprolol 50 mg once a day, but we have had to hold it because of relative bradycardia which he apparently has a history of.
3. Nexium 40 mg once a day.
4. Zocor 40 mg once a day, and there is a fasting lipid profile pending at the time of this dictation. I see that his LDL was 136 on May 3, 2002.
5. Plavix 600 mg p.o. x1 which I am giving him tonight.

Other medical history is inclusive for obstructive sleep apnea for which he is unable to tolerate positive pressure ventilation, GERD, arthritis

DISPOSITION: The patient and his wife have requested and are agreeable with transfer to Medical Center, and we are enclosing the CD ROM of his images.

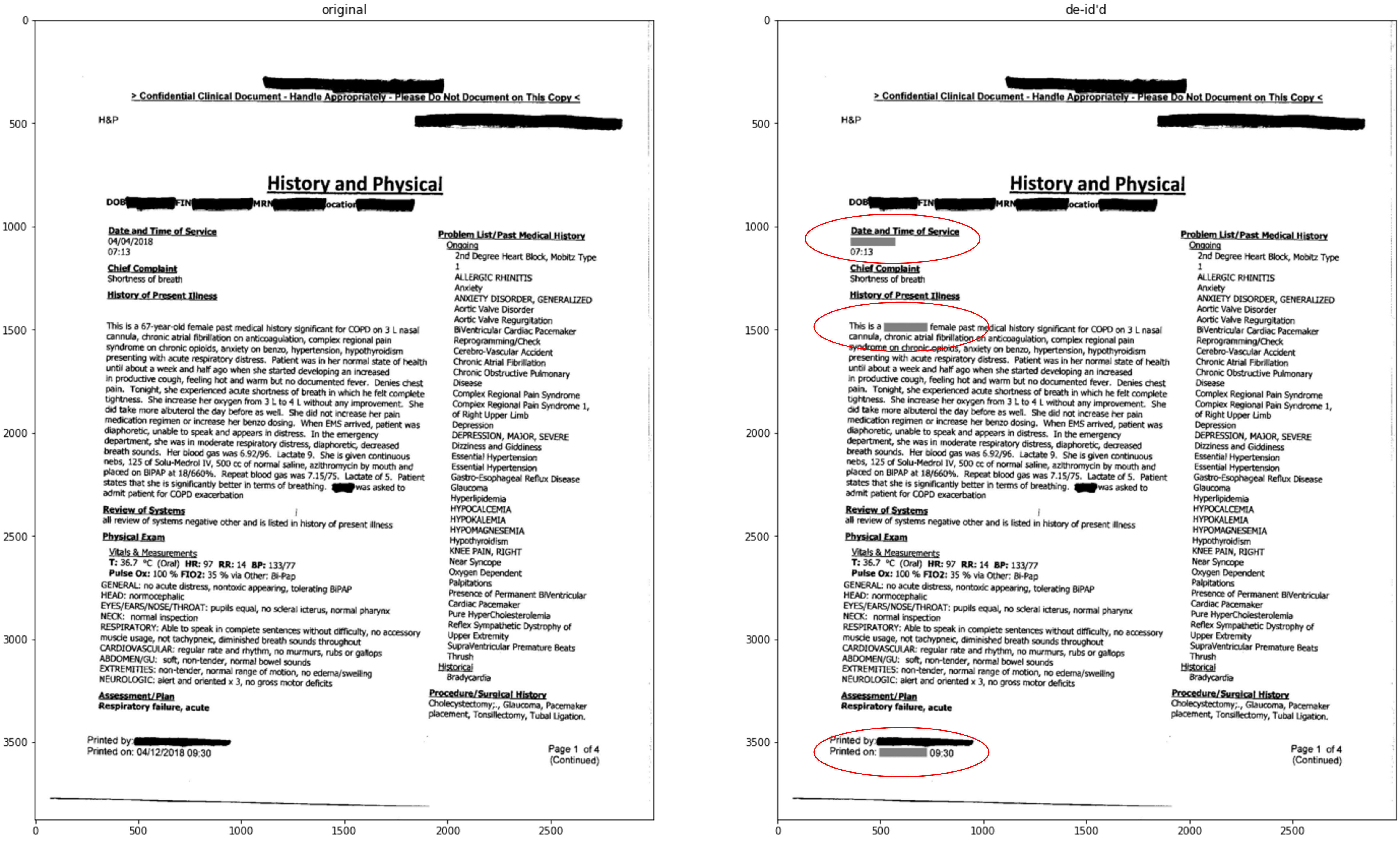
History of Present Illness
Homer Simpson is a(n) 72 year old male with history of coronary artery disease, cardiomyopathy, diabetes type 2, hypertension, chronic kidney disease, and other comorbidities. He presents with rectal bleeding in the last two weeks.
No dyspnea or cough. No chest pain.

CONDITION ON TRANSFER: Stable but guarded. The patient is pain-free at this time.

MEDICATIONS ON TRANSFER:
1. Aspirin 325 mg once a day.
2. Metoprolol 50 mg once a day, but we have had to hold it because of relative bradycardia which he apparently has a history of.
3. Nexium 40 mg once a day.
4. Zocor 40 mg once a day, and there is a fasting lipid profile pending at the time of this dictation. I see that his LDL was 136 on May 3, 2002.
5. Plavix 600 mg p.o. x1 which I am giving him tonight.

Other medical history is inclusive for obstructive sleep apnea for which he is unable to tolerate positive pressure ventilation, GERD, arthritis

DISPOSITION: The patient and his wife have requested and are agreeable with transfer to Medical Center, and we are enclosing the CD ROM of his images.



[redacted]

> Confidential Clinical Document - Handle Appropriately - Please Do Not Document on This Copy <

H&P [redacted]

History and Physical

DOB [redacted] FIN [redacted] MRN [redacted] ocation [redacted]

Date and Time of Service
04/04/2018
07:13

Chief Complaint
Shortness of breath

History of Present Illness

This is a 67-year-old female past medical history significant for COPD on 3 L nasal cannula, chronic atrial fibrillation on anticoagulation, complex regional pain syndrome on chronic opioids, anxiety on benzo, hypertension, hypothyroidism presenting with acute respiratory distress. Patient was in her normal state of health until about a week and half ago when she started developing an increased in productive cough, feeling hot and warm but no documented fever. Denies chest pain. Tonight, she experienced acute shortness of breath in which he felt complete tightness. She increase her oxygen from 3 L to 4 L without any improvement. She did take more albuterol the day before as well. She did not increase her pain medication regimen or increase her benzo dosing. When EMS arrived, patient was diaphoretic, unable to speak and appears in distress. In the emergency department, she was in moderate respiratory distress, diaphoretic, decreased breath sounds. Her blood gas was 6.92/96. Lactate 9. She is given continuous nebs, 125 of Solu-Medrol IV, 500 cc of normal saline, azithromycin by mouth and placed on BIPAP at 18/660%. Repeat blood gas was 7.15/75. Lactate of 5. Patient states that she is significantly better in terms of breathing. [redacted] was asked to admit patient for COPD exacerbation

Review of Systems
all review of systems negative other and is listed in history of present illness

Physical Exam

Vitals & Measurements
T: 36.7 °C (Oral) **HR:** 97 **RR:** 14 **BP:** 133/77
Pulse Ox: 100 % **FIO2:** 35 % via Other: Bi-Pap

GENERAL: no acute distress, nontoxic appearing, tolerating BiPAP
HEAD: normocephalic
EYES/EARS/NOSE/THROAT: pupils equal, no scleral icterus, normal pharynx
NECK: normal inspection
RESPIRATORY: Able to speak in complete sentences without difficulty, no accessory muscle usage, not tachypneic, diminished breath sounds throughout
CARDIOVASCULAR: regular rate and rhythm, no murmurs, rubs or gallops
ABDOMEN/GU: soft, non-tender, normal bowel sounds
EXTREMITIES: non-tender, normal range of motion, no edema/swelling
NEUROLOGIC: alert and oriented x 3, no gross motor deficits

Assessment/Plan
Respiratory failure, acute

Problem List/Past Medical History
Ongoing
2nd Degree Heart Block, Mobitz Type 1
ALLERGIC RHINITIS
Anxiety
ANXIETY DISORDER, GENERALIZED
Aortic Valve Disorder
Aortic Valve Regurgitation
BiVentricular Cardiac Pacemaker Reprogramming/Check
Cerebro-Vascular Accident
Chronic Atrial Fibrillation
Chronic Obstructive Pulmonary Disease
Complex Regional Pain Syndrome
Complex Regional Pain Syndrome 1, of Right Upper Limb
Depression
DEPRESSION, MAJOR, SEVERE
Dizziness and Giddiness
Essential Hypertension
Essential Hypertension
Gastro-Esophageal Reflux Disease
Glaucoma
Hyperlipidemia
HYPOCALCEMIA
HYPOKALEMIA
HYPOMAGNESEMIA
Hypothyroidism
KNEE PAIN, RIGHT
Near Syncope
Oxygen Dependent
Palpitations
Presence of Permanent BiVentricular Cardiac Pacemaker
Pure HyperCholesterolemia
Reflex Sympathetic Dystrophy of Upper Extremity
SupraVentricular Premature Beats
Thrush
Historical
Bradycardia

Procedure/Surgical History
Cholecystectomy;., Glaucoma, Pacemaker placement, Tonsillectomy, Tubal Ligation.

Printed by: [redacted]
Printed on: 04/12/2018 09:30

[redacted]

> Confidential Clinical Document - Handle Appropriately - Please Do Not Document on This Copy <

H&P [redacted]

History and Physical

DOB [redacted] FIN [redacted] MRN [redacted] ocation [redacted]

Date and Time of Service
[redacted]
07:13

Chief Complaint
Shortness of breath

History of Present Illness

This is a [redacted] female past medical history significant for COPD on 3 L nasal cannula, chronic atrial fibrillation on anticoagulation, complex regional pain syndrome on chronic opioids, anxiety on benzo, hypertension, hypothyroidism presenting with acute respiratory distress. Patient was in her normal state of health until about a week and half ago when she started developing an increased in productive cough, feeling hot and warm but no documented fever. Denies chest pain. Tonight, she experienced acute shortness of breath in which he felt complete tightness. She increase her oxygen from 3 L to 4 L without any improvement. She did take more albuterol the day before as well. She did not increase her pain medication regimen or increase her benzo dosing. When EMS arrived, patient was diaphoretic, unable to speak and appears in distress. In the emergency department, she was in moderate respiratory distress, diaphoretic, decreased breath sounds. Her blood gas was 6.92/96. Lactate 9. She is given continuous nebs, 125 of Solu-Medrol IV, 500 cc of normal saline, azithromycin by mouth and placed on BIPAP at 18/660%. Repeat blood gas was 7.15/75. Lactate of 5. Patient states that she is significantly better in terms of breathing. [redacted] was asked to admit patient for COPD exacerbation

Review of Systems
all review of systems negative other and is listed in history of present illness

Physical Exam

Vitals & Measurements
T: 36.7 °C (Oral) **HR:** 97 **RR:** 14 **BP:** 133/77
Pulse Ox: 100 % **FIO2:** 35 % via Other: Bi-Pap

GENERAL: no acute distress, nontoxic appearing, tolerating BiPAP
HEAD: normocephalic
EYES/EARS/NOSE/THROAT: pupils equal, no scleral icterus, normal pharynx
NECK: normal inspection
RESPIRATORY: Able to speak in complete sentences without difficulty, no accessory muscle usage, not tachypneic, diminished breath sounds throughout
CARDIOVASCULAR: regular rate and rhythm, no murmurs, rubs or gallops
ABDOMEN/GU: soft, non-tender, normal bowel sounds
EXTREMITIES: non-tender, normal range of motion, no edema/swelling
NEUROLOGIC: alert and oriented x 3, no gross motor deficits

Assessment/Plan
Respiratory failure, acute

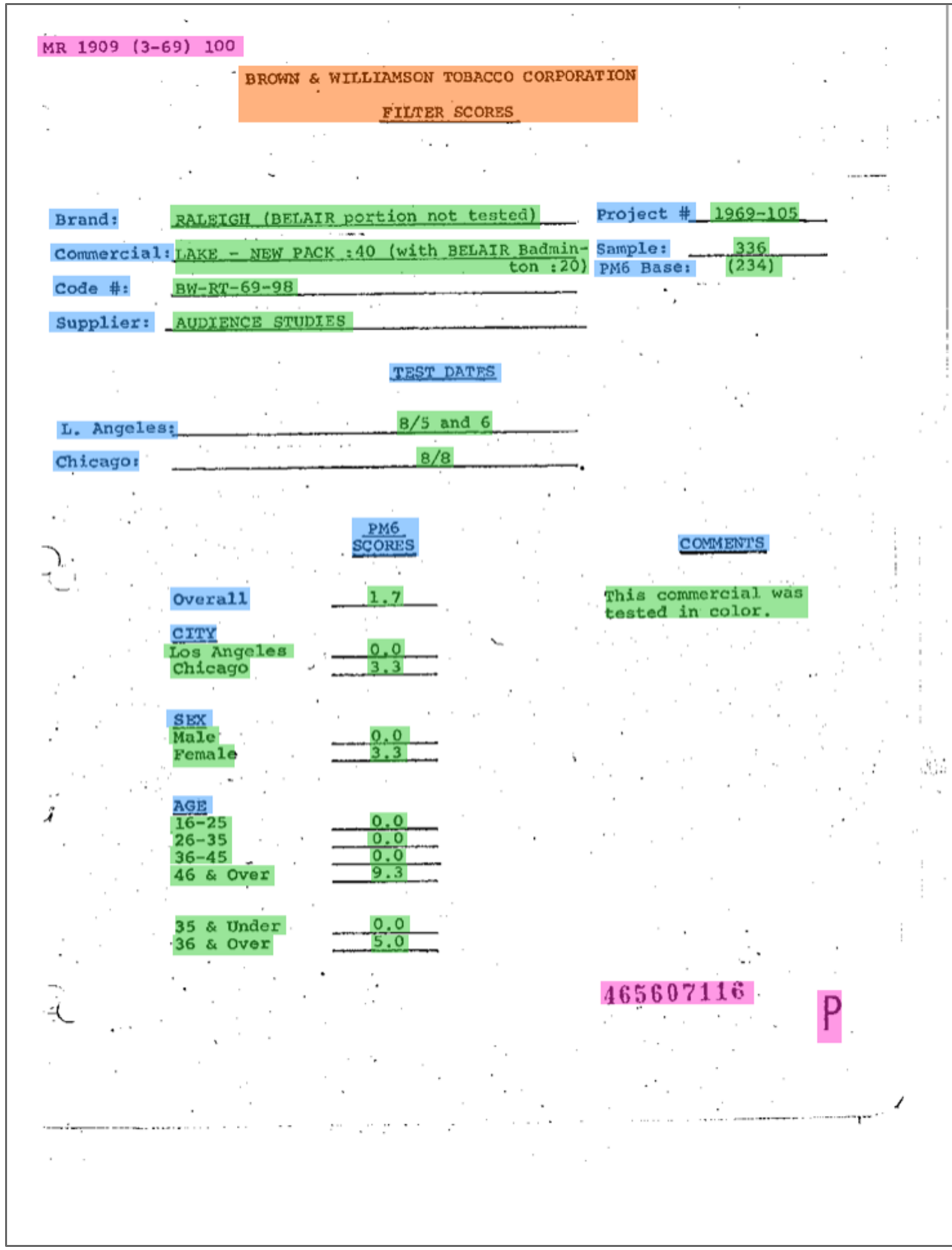
Problem List/Past Medical History
Ongoing
2nd Degree Heart Block, Mobitz Type 1
ALLERGIC RHINITIS
Anxiety
ANXIETY DISORDER, GENERALIZED
Aortic Valve Disorder
Aortic Valve Regurgitation
BiVentricular Cardiac Pacemaker Reprogramming/Check
Cerebro-Vascular Accident
Chronic Atrial Fibrillation
Chronic Obstructive Pulmonary Disease
Complex Regional Pain Syndrome
Complex Regional Pain Syndrome 1, of Right Upper Limb
Depression
DEPRESSION, MAJOR, SEVERE
Dizziness and Giddiness
Essential Hypertension
Essential Hypertension
Gastro-Esophageal Reflux Disease
Glaucoma
Hyperlipidemia
HYPOCALCEMIA
HYPOKALEMIA
HYPOMAGNESEMIA
Hypothyroidism
KNEE PAIN, RIGHT
Near Syncope
Oxygen Dependent
Palpitations
Presence of Permanent BiVentricular Cardiac Pacemaker
Pure HyperCholesterolemia
Reflex Sympathetic Dystrophy of Upper Extremity
SupraVentricular Premature Beats
Thrush
Historical
Bradycardia

Procedure/Surgical History
Cholecystectomy;., Glaucoma, Pacemaker placement, Tonsillectomy, Tubal Ligation.

Printed by: [redacted]
Printed on: [redacted] 09:30

Visual Document Classifier

Visual Document NER



MR 1909 (3-69) 100

BROWN & WILLIAMSON TOBACCO CORPORATION

FILTER SCORES

Brand: RALEIGH (BELAIR portion not tested)

Project #: 1969-105

Commercial: LAKE - NEW PACK :40 (with BELAIR Badminton :20)

Sample: 336

PM6 Base: (234)

Code #: BW-RT-69-98

Supplier: AUDIENCE STUDIES

TEST DATES

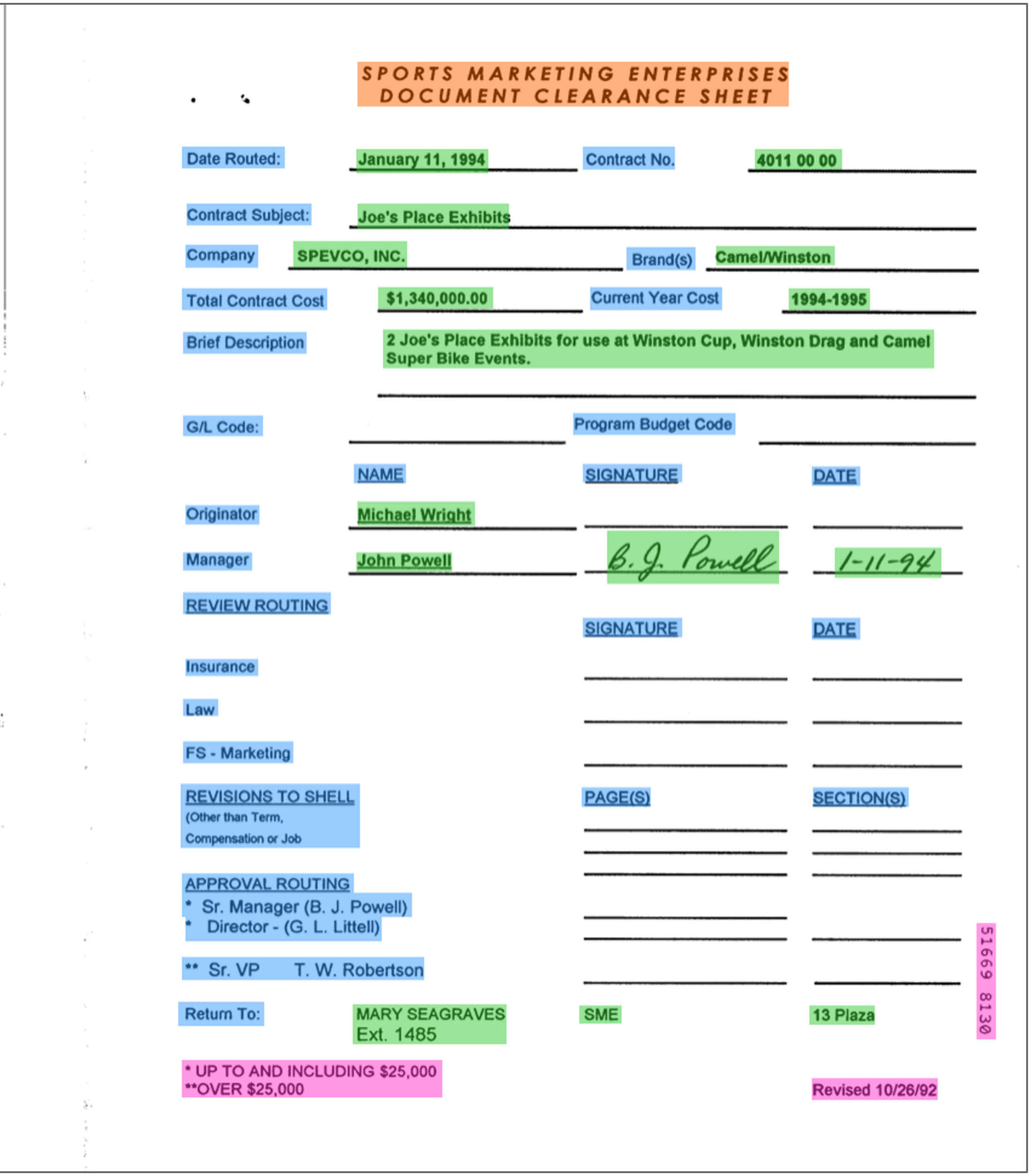
L. Angeles: 8/5 and 6

Chicago: 8/8

	PM6 SCORES	COMMENTS
Overall	1.7	This commercial was tested in color.
CITY		
Los Angeles	0.0	
Chicago	3.3	
SEX		
Male	0.0	
Female	3.3	
AGE		
16-25	0.0	
26-35	0.0	
36-45	0.0	
46 & Over	9.3	
35 & Under	0.0	
36 & Over	5.0	

465607116

P



SPORTS MARKETING ENTERPRISES
DOCUMENT CLEARANCE SHEET

Date Routed: January 11, 1994   Contract No. 4011 00 00

Contract Subject: Joe's Place Exhibits

Company: SPEVCO, INC.   Brand(s): Camel/Winston

Total Contract Cost: $1,340,000.00   Current Year Cost: 1994-1995

Brief Description: 2 Joe's Place Exhibits for use at Winston Cup, Winston Drag and Camel Super Bike Events.

G/L Code: Program Budget Code

	NAME	SIGNATURE	DATE
Originator	Michael Wright		
Manager	John Powell	B. J. Powell	1-11-94

REVIEW ROUTING

	SIGNATURE	DATE
Insurance		
Law		
FS - Marketing		

REVISIONS TO SHELL (Other than Term, Compensation or Job   PAGE(S)   SECTION(S)

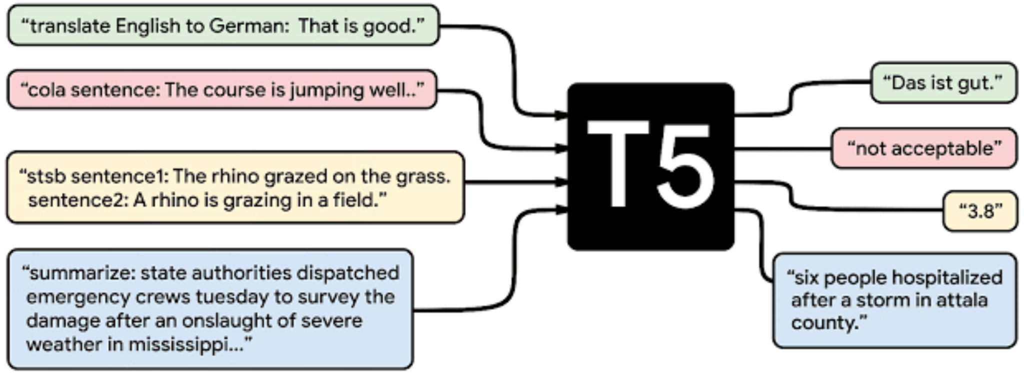
APPROVAL ROUTING
* Sr. Manager (B. J. Powell)
* Director - (G. L. Littell)
** Sr. VP   T. W. Robertson

Return To: MARY SEAGRAVES Ext. 1485   SME   13 Plaza

* UP TO AND INCLUDING $25,000
**OVER $25,000

Revised 10/26/92

51669 8130




```
# Closed book Question Answering
nlu.load('en.t5').predict('what is the capital of Germany?') # >>> Berlin
# Open Book Question answering
nlu.load('en.t5').predict('Who is president of Nigeria?') # >>> Muhammadu Buhari


# Open book Question Answering
context   = 'Peters last week was terrible! He had an accident and broke his leg while skiing!'
question1  = 'Why was peters week so bad?'
question2  = 'How did peter broke his leg?'
nlu.load('answer_question').predict(question1 + context) # >>> broke his leg
nlu.load('answer_question').predict(question2 + context) # >>> skiing



# Big T5 model for Summarization, Sentiment, Text Similarity and other SQUAD/GLUE tasks
pipe = nlu.load('t5')
pipe['t5'].settask('summarize')
pipe.predict(long_text)
```

[Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer]

Every T5 Task with explanation:

Task Name	Explanation
1.CoLA	Classify if a sentence is grammatically correct
2.RTE	Classify whether a statement can be deducted from a sentence
3.MNLI	Classify for a hypothesis and premise whether they contradict or contradict each other or neither of both (3 class).
4.MRPC	Classify whether a pair of sentences is a re-phrasing of each other (semantically equivalent)
5.QNLI	Classify whether the answer to a question can be deducted from an answer candidate.
6.QQP	Classify whether a pair of questions is a re-phrasing of each other (semantically equivalent)
7.SST2	Classify the sentiment of a sentence as positive or negative
8.STSB	Classify the sentiment of a sentence on a scale from 1 to 5 (21 Sentiment classes)
9.CB	Classify for a premise and a hypothesis whether they contradict each other or not (binary).
10.COPA	Classify for a question, premise, and 2 choices which choice the correct choice is (binary).
11.MultiRc	Classify for a question, a paragraph of text, and an answer candidate, if the answer is correct (binary),
12.WiC	Classify for a pair of sentences and a disambigous word if the word has the same meaning in both sentences.
13.WSC/DPR	Predict for an ambiguous pronoun in a sentence what it is referring to.
14.Summarization	Summarize text into a shorter representation.
15.SQuAD	Answer a question for a given context.
16.WMT1.	Translate English to German
17.WMT2.	Translate English to French
18.WMT3.	Translate English to Romanian

1. Text summarization
2. Question answering
3. Translation
4. Sentiment analysis
5. Natural Language inference
6. Coreference resolution
7. Sentence Completion
8. Word sense disambiguation

Google

Translate between **200+ Languages**

With Marian: Fast Neural Machine Translation in C++

MARIANNMT

Fast Neural Machine Translation in C++


```
# Use ISO standards for the languages
nlu.load('<start_language>.translate_to.<target_language>')

#Translate Turkish to English:
nlu.load('tr.translate_to.en')

#Translate English to French:
nlu.load('en.translate_to.fr')

#Translate French to Hebrew
nlu.load('fr.translate_to.he')`

#Translate English to German
nlu.load('en.translate_to.de')`
```

109 Languages supported by **Language-agnostic BERT Sentence Embedding (LABSE)**

Train in **1 Language**, classify in **100 different languages correct**

```
# Binary Class Classifier, 2 classes
nlu.load('xx.embed_sentence.labse train.sentiment').fit(train_df).predict(test_df)

# Multi Class Classifier, N classes
nlu.load('xx.embed_sentence.labse train.classifier').fit(train_df).predict(test_df)

# Multi Class Classifier with multiple labels example (i.e. Hashtags)
# N classes, where one row can be assigned up to N labels
nlu.load('xx.embed_sentence.labse train.multi_classifier').fit(train_df).predict(test_df)
```

ISO	NAME	ISO	NAME	ISO	NAME
af	AFRIKAANS	ht	HAITIAN_CREOLE	pt	PORTUGUESE
am	AMHARIC	hu	HUNGARIAN	ro	ROMANIAN
ar	ARABIC	hy	ARMENIAN	ru	RUSSIAN
as	ASSAMESE	id	INDONESIAN	rw	KINYARWANDA
az	AZERBAIJANI	ig	IGBO	si	SINHALESE
be	BELARUSIAN	is	ICELANDIC	sk	SLOVAK
bg	BULGARIAN	it	ITALIAN	sl	SLOVENIAN
bn	BENGALI	ja	Japanese	sm	SAMOAN
bo	TIBETAN	jv	JAVANESE	sn	SHONA
bs	BOSNIAN	ka	GEORGIAN	so	SOMALI
ca	CATALAN	kk	KAZAKH	sq	ALBANIAN
ceb	CEBUANO	km	KHMER	sr	SERBIAN
co	CORSICAN	kn	KANNADA	st	SESOTHO
cs	CZECH	ko	KOREAN	su	SUNDANESE
cy	WELSH	ku	KURDISH	sv	SWEDISH
da	DANISH	ky	KYRGYZ	sw	SWAHILI
de	GERMAN	la	LATIN	ta	TAMIL
el	GREEK	lb	LUXEMBOURGISH	te	TELUGU
en	ENGLISH	lo	LAOTHIAN	tg	TAJIK
eo	ESPERANTO	lt	LITHUANIAN	th	THAI
es	SPANISH	lv	LATVIAN	tk	TURKMEN
et	ESTONIAN	mg	MALAGASY	tl	TAGALOG
eu	BASQUE	mi	MAORI	tr	TURKISH
fa	PERSIAN	mk	MACEDONIAN	tt	TATAR
fi	FINNISH	ml	MALAYALAM	ug	UIGHUR
fr	FRENCH	mn	MONGOLIAN	uk	UKRAINIAN
fy	FRISIAN	mr	MARATHI	ur	URDU
ga	IRISH	ms	MALAY	uz	UZBEK
gd	SCOTS_GAELIC	mt	MALTESE	vi	VIETNAMESE
gl	GALICIAN	my	BURMESE	wo	WOLOF
gu	GUJARATI	ne	NEPALI	xh	XHOSA
ha	HAUSA	nl	DUTCH	yi	YIDDISH
haw	HAWAIIAN	no	NORWEGIAN	yo	YORUBA
he	HEBREW	ny	NYANJA	zh	Chinese
hi	HINDI	or	ORIYA	zu	ZULU
hmn	HMONG	pa	PUNJABI		
hr	CROATIAN	pl	POLISH		

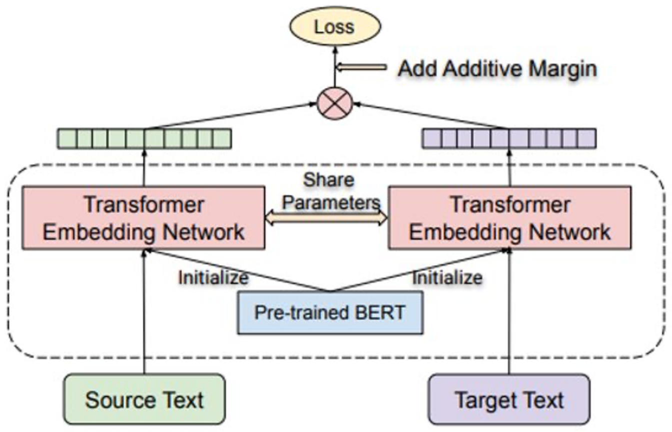
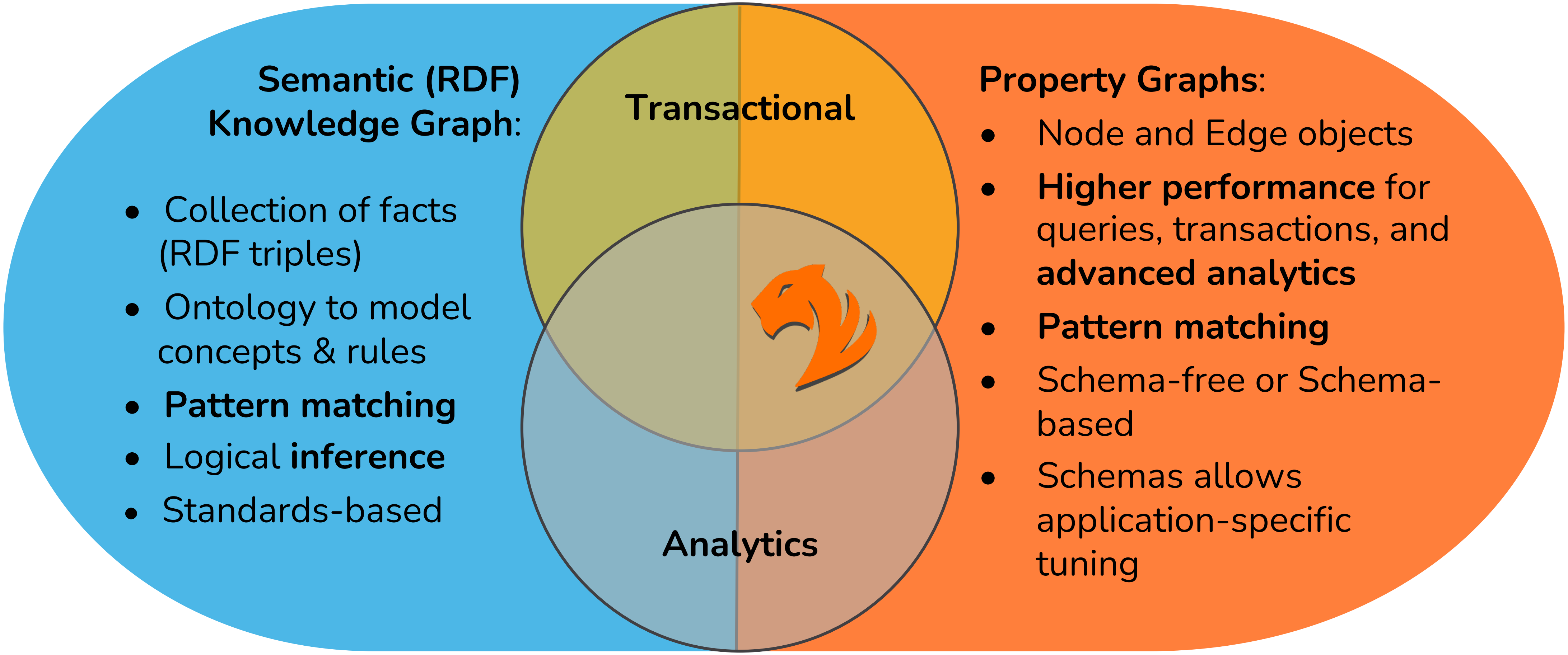


Figure 1: Dual encoder model with BERT based encoding modules.

Demo - Deep Analytics with graph

Types of Graph Databases



TigerGraph is a High-Performance and Scalable Property Graph, for both Analytics & Transactions.

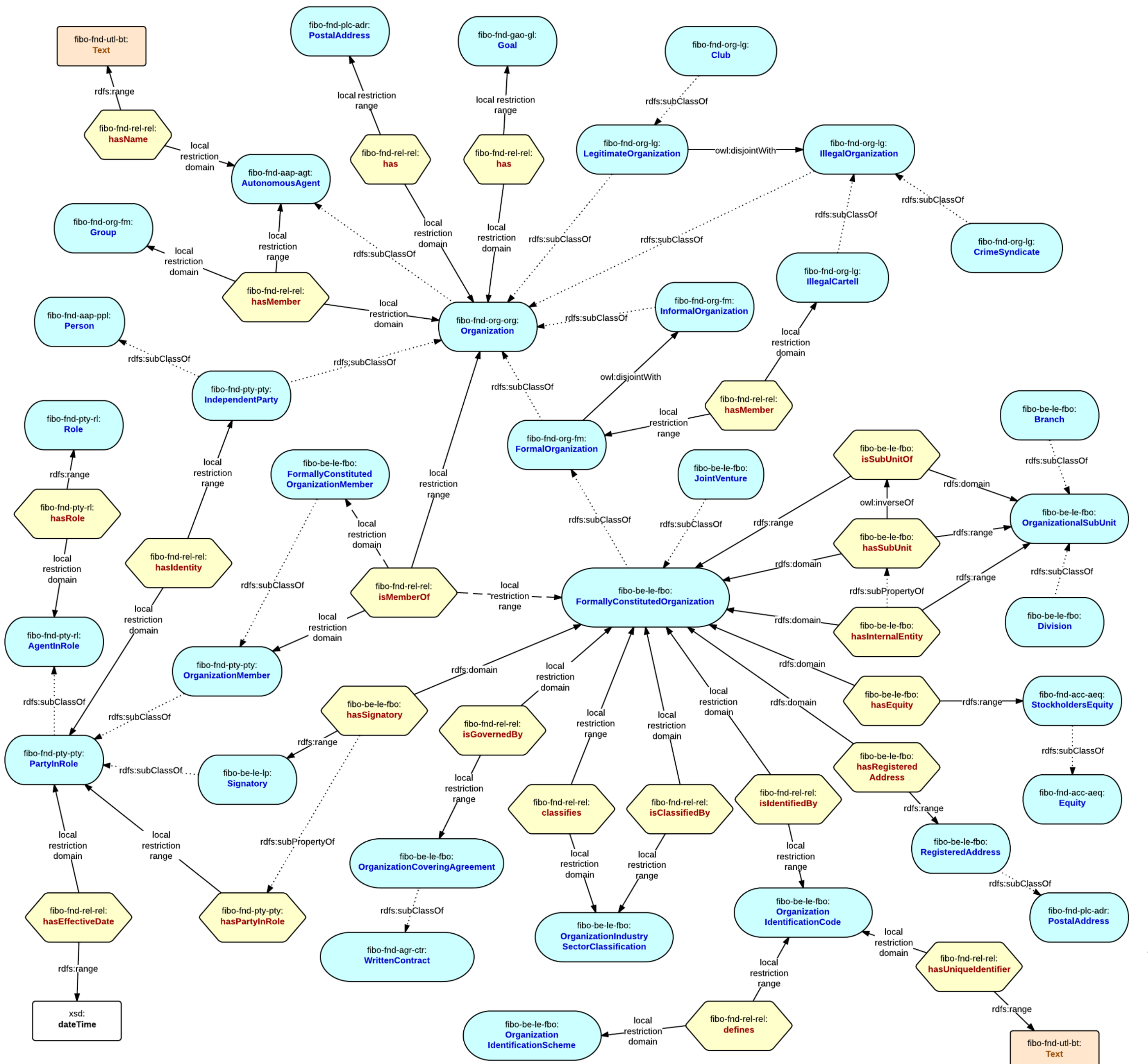
FIBO

The Financial Industry Business Ontology (FIBO) defines the sets of things that are of interest in financial business applications and the ways that those things can relate to one another.

De e

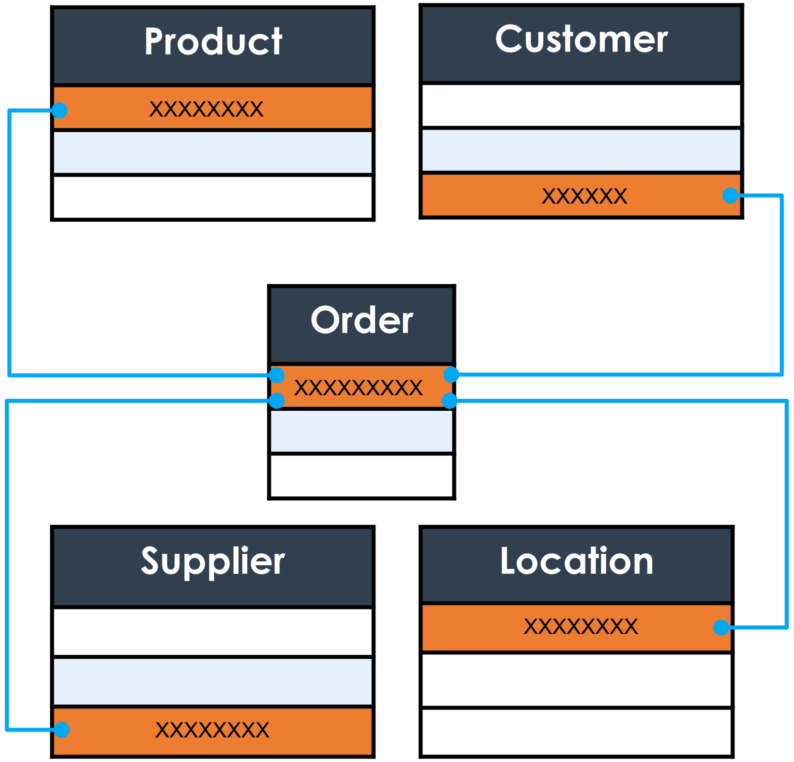
- Business Entities
  - Corporations
  - Functional Entities
  - Government Entities
  - Legal Entities
  - Ownership and Control
  - Partnerships
  - Private Limited Companies
  - Sole Proprietorships
  - Trusts
- Business Process Domain
- Corporate Actions and Events Domain
- Derivatives Domain
- Financial Business and Commerce
- Foundations
- Funds Module
- Indices and Indicators
- Loans
- Market Data Domain
- Securities

FIBO Organization Ontology Model v1.0



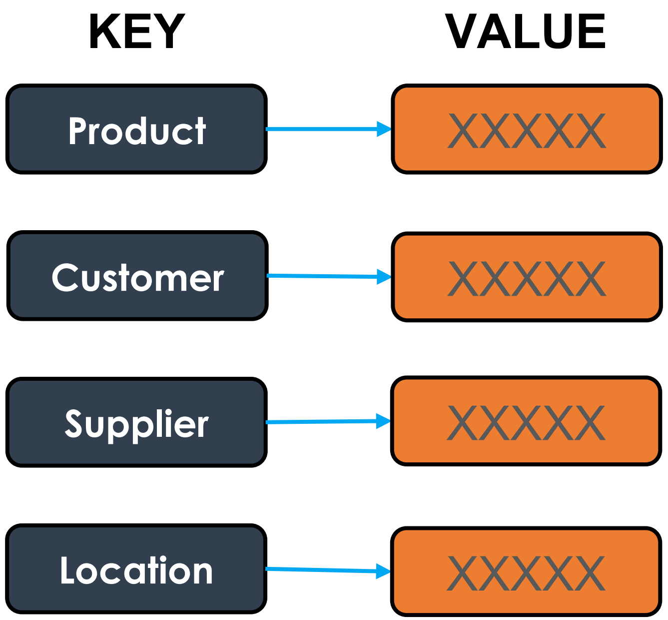
The Evolution of Databases

Relational Database



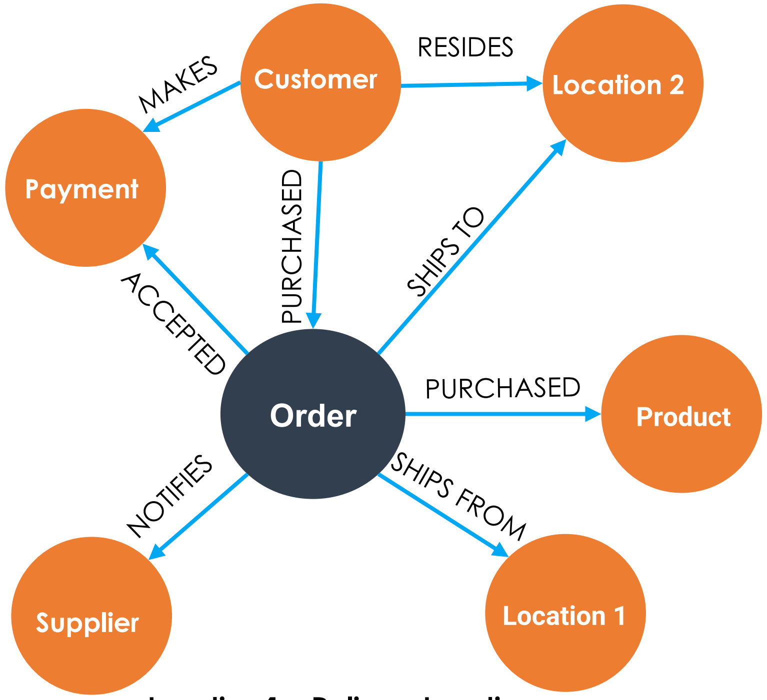
- Rigid schema
- High performance for transactions
- Poor performance for deep analytics

Key-Value Database



- Highly fluid schema/no schema
- High performance for simple transactions
- Poor performance deep analytics

Graph Database



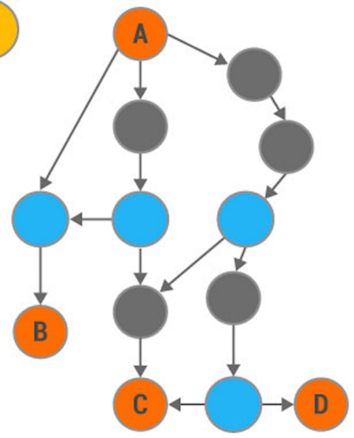
Location 1 = Delivery Location
Location 2 = Warehouse

- Flexible schema
- High performance for complex transactions
- High performance for deep analytics

TigerGraph

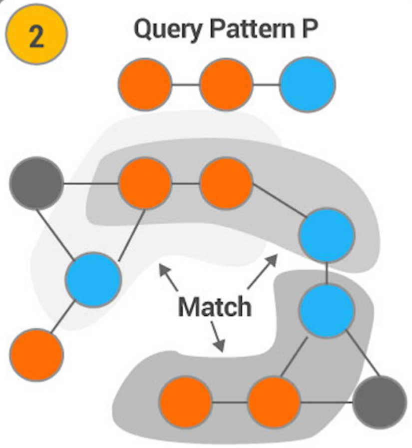
7 Key Data Science Capabilities Powered By a Native Parallel Graph

1 Deep Link Analysis



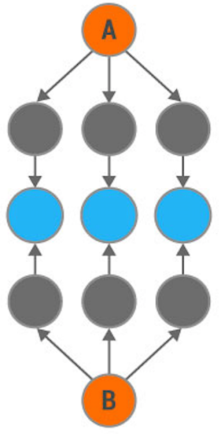
For a set of entities (e.g. customers, accounts, citizens, doctors), show all links or connections

2 Multi-dimensional Entity & Pattern Matching



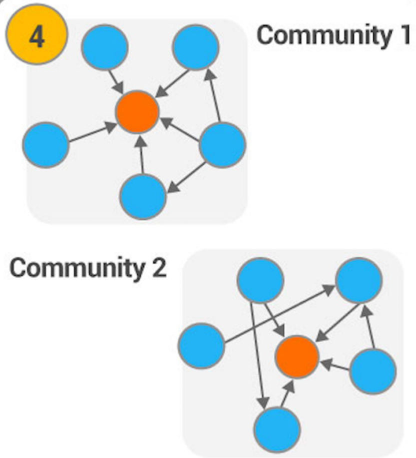
Given a pattern (e.g. connections indicating fraud), find similar patterns in the graph

3 Relational Commonality Discovery & Computation



Given 2 entities (e.g. customers, merchants, doctors), follow their relationship to find commonality

4 Hub & Community Detection



Find most influential members of a group (customers, doctors, citizens) & detect community around them

5	**Geospatial Graph Analysis**	Analyze changes in entities & relationships with location data
6	**Temporal (Time-Series) Graph Analysis**	Analyze changes in entities & relationships over time
7	**Machine Learning Feature Generation & Explainable AI**	Extract graph-based features to feed as training data for machine learning; Power Explainable AI

TigerGraph

Thank You

Christian Kasim Loan

Senior Data Scientist

- Distributed AI Lab(DAI), Daimler-Lab, CKL-IT (Consulting Company) Founder
- 10+ years , Architected and implemented various cloud agnostic big data systems and frameworks
- Creator of the NLU library
- Email: christian@johnsnowlabs.com

Abhishek Mehta

Director of Field Engineering

- McKinsey, Bloomberg, Cisco & Dabizmo (NLP Startup) Founder
- 15+ years designing and implementing complex analytics solutions for Fortune 100 companies
- Patents in NLP spanning Conceptunary Ontology Design, Language Pattern Recognition, and Conversion
- Email: abhi@tigergraph.com

