Mahidol University
Faculty of Medicine Ramathibodi Hospital
Department of Clinical Epidemiology and Biostatistics

Introduction to Text Analysis using Natural Language Processing (NLP)

Understanding the Basics and Applications in Healthcare

21 August 2024 12:00 – 13:00

Dr. Wanchana Ponthongmak, Ph.D.

Wisdom of the Land

Dr. Wanchana Ponthongmak, Ph.D.
Position : Lecturer
Email : wanchana.pon@mahidol.edu
Office Location : 4th Floor, Sukho Place Building, Sukhothai Road. Dusit, Bangkok 10300, Thailand.

- **Academic Qualifications**

2018 - 2022	Ph.D. Data Science for Healthcare, Mahidol University
2010 - 2014	M.Sc. (Health Informatics), Mahidol University
2006 - 2009	B.Sc. (Public Health), Mahidol University
2000 - 2005	Boonwattana school

- **Current & previous positions**

2023-Present	Lecturer, CEB, Faculty of Medicine Ramathibodi Hospital
2020-2023	Research Assistant, CEB, Faculty of Medicine Ramathibodi Hospital
2013-2014	Secretariat, Asia eHealth Information Network
2012-2018	Research Assistant, Thai Health Information Standards Development Center

Area of Interests

- Artificial Intelligence (AI)
- Machine Learning (ML)
- Deep Learning (DL)
- Big Data
- Natural Language Processing (NLP)

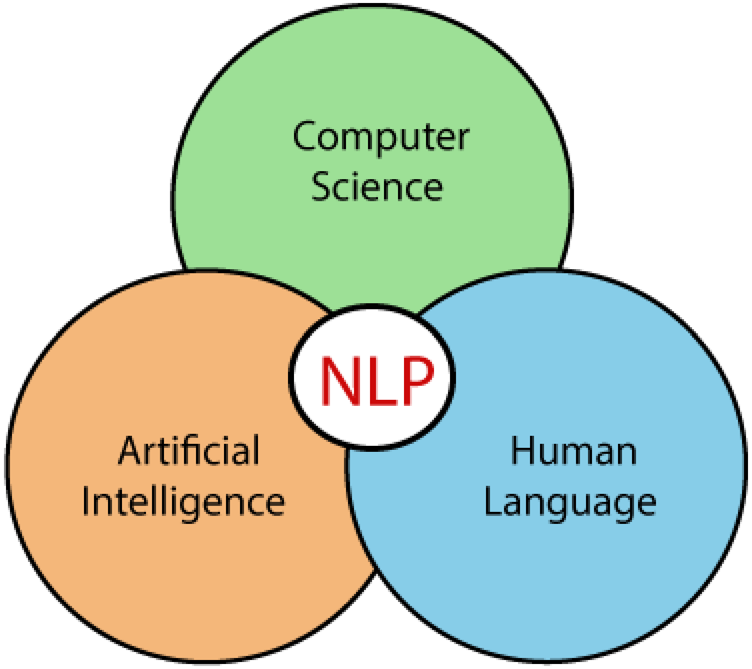
Outlines

- What is NLP and how does it work?
- Common NLP techniques
- Applications and use cases utilizing NLP in healthcare

What is NLP?

"A field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between computers and human (natural) languages, and, in particular, concerned with programming computers to fruitfully process large natural language corpora."

Jurafsky, D., & Martin, J. H. (2009).



"A collection of methods used to process, analyze, and understand natural languages by leveraging computational techniques"

Manning, C. D., & Schütze, H. (1999)

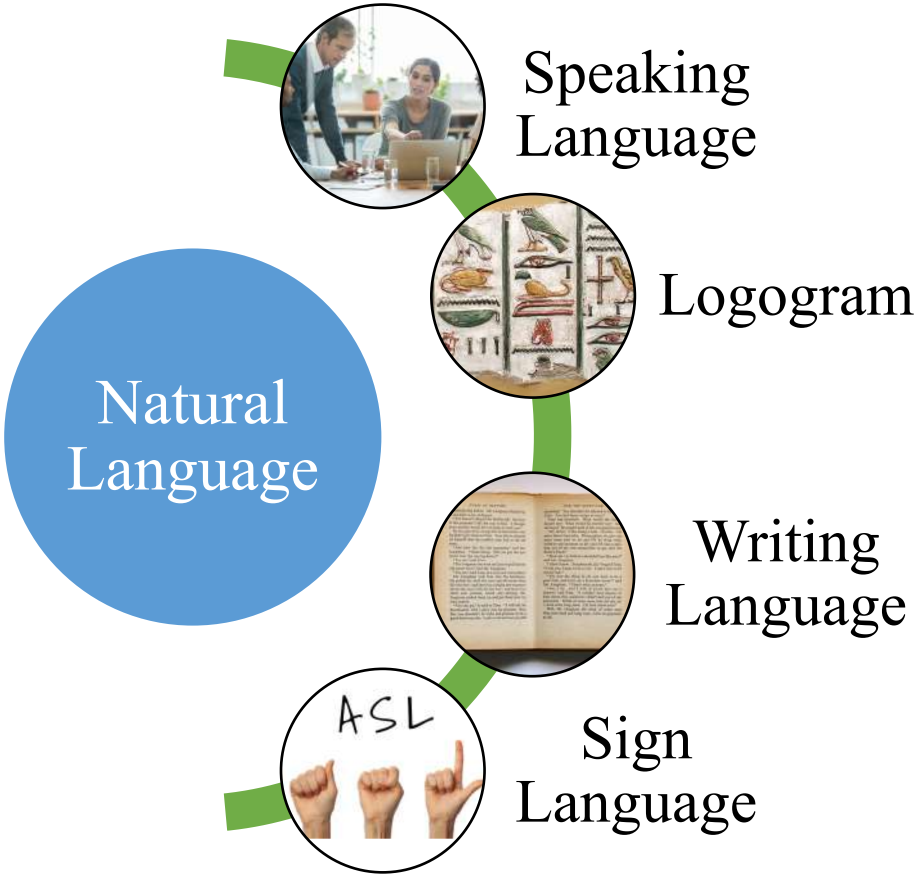
What is natural language?

"A language that has developed in a community and has been passed down through generations through social interaction. It is acquired by individuals naturally as part of their development, without conscious planning or premeditation."

David Crystal, 2010

"Any human language that has evolved naturally through use and social interaction, rather than being artificially created or constructed.

ChatGPT4-o, 2024



Natural Language

Speaking Language

Logogram

Writing Language

Sign Language

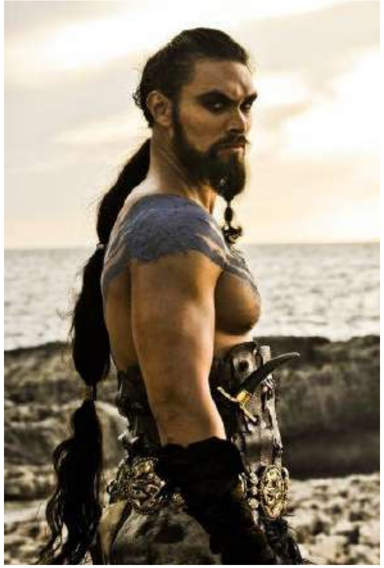
Natural Language

Any language evolved naturally in **humans** through use and repetition without conscious planning and premeditation.

How about these languages?

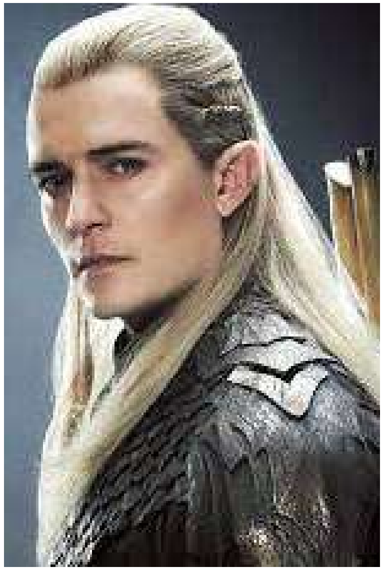
Klingon

▶



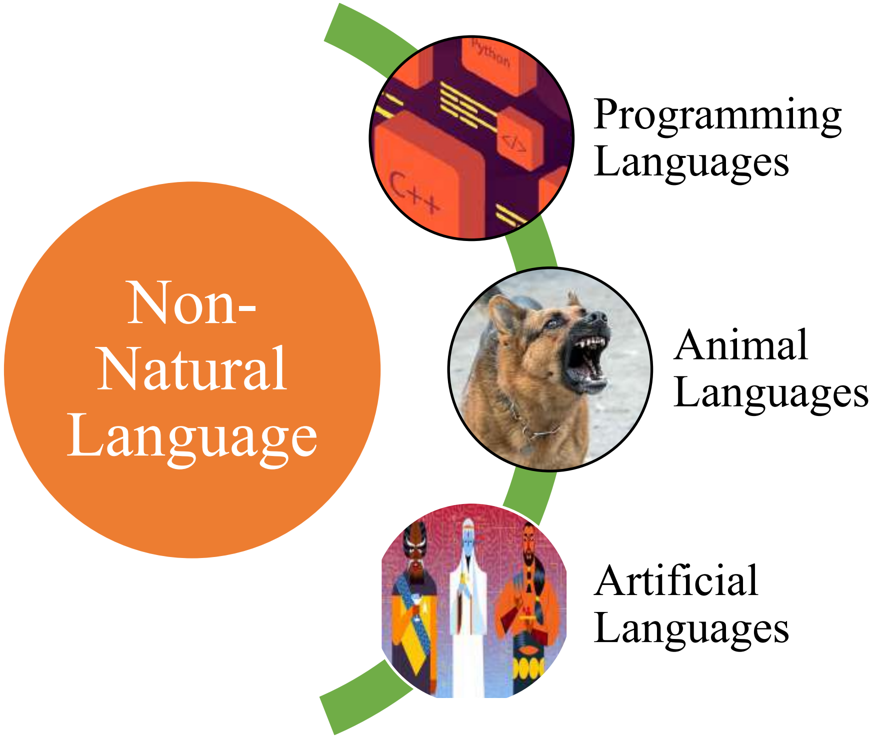
Dothraki

▶



Elvish

▶



Non-Natural Language

Programming Languages

Animal Languages

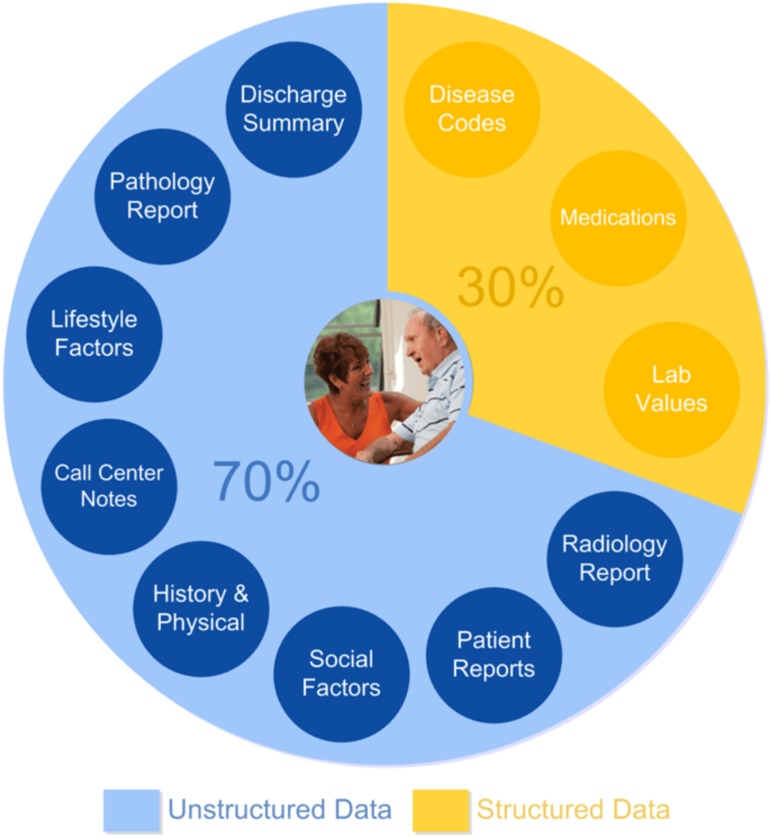
Artificial Languages

Non-Natural Language

Any language evolved in non-humans through usage.

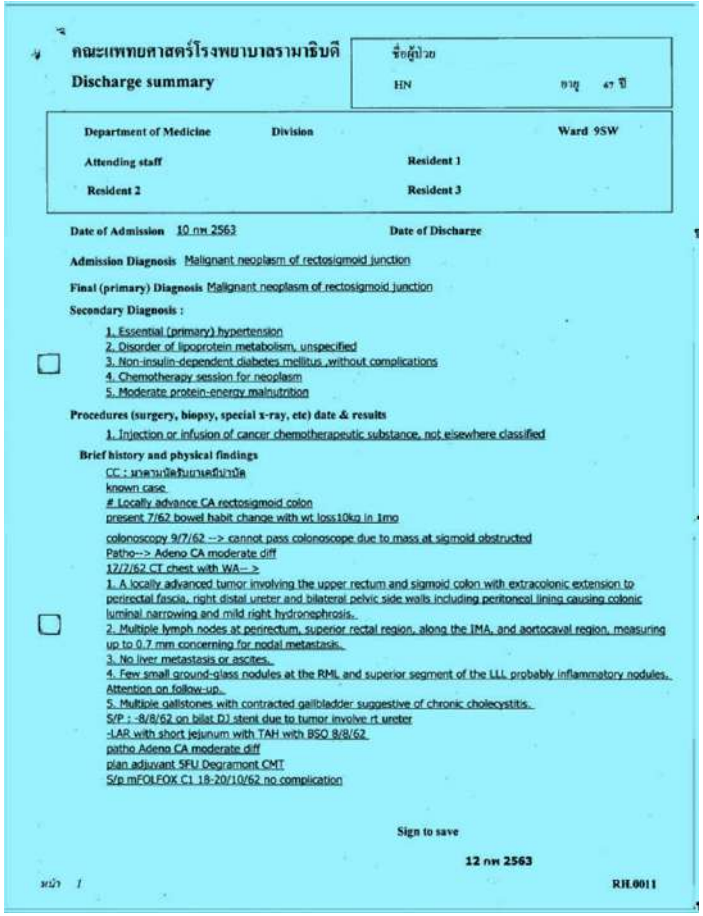
Why do we care about NLP?

- ~70% of data in hospitals are unstructured data.
- Text data are an extremely rich source of information.
- But extracting insights from them can be hard and time-consuming due to its unstructured nature.



https://www.altexsoft.com/blog/nlp-healthcare/

Unstructured data



คณะแพทยศาสตร์โรงพยาบาลรามาธิบดี
Discharge summary

ชื่อผู้ป่วย
HN อายุ 47 ปี

Department of Medicine Division Ward 9SW
Attending staff Resident 1
Resident 2 Resident 3

Date of Admission 10 กพ 2563 Date of Discharge

Admission Diagnosis Malignant neoplasm of rectosigmoid junction

Final (primary) Diagnosis Malignant neoplasm of rectosigmoid junction

Secondary Diagnosis :

□
1. Essential (primary) hypertension
2. Disorder of lipoprotein metabolism, unspecified
3. Non-insulin-dependent diabetes mellitus ,without complications
4. Chemotherapy session for neoplasm
5. Moderate protein-energy malnutrition

Procedures (surgery, biopsy, special x-ray, etc) date & results
1. Injection or infusion of cancer chemotherapeutic substance, not elsewhere classified

Brief history and physical findings
CC : มาตามนัดรับยาเคมีบำบัด
known case
Locally advance CA rectosigmoid colon
present 7/62 bowel habit change with wt loss10kg in 1mo

colonoscopy 9/7/62 --> cannot pass colonoscope due to mass at sigmoid obstructed
Patho--> Adeno CA moderate diff
17/7/62 CT chest with WA-- >
□
1. A locally advanced tumor involving the upper rectum and sigmoid colon with extracolonic extension to perirectal fascia, right distal ureter and bilateral pelvic side walls including peritoneal lining causing colonic luminal narrowing and mild right hydronephrosis.
2. Multiple lymph nodes at perirectum, superior rectal region, along the IMA, and aortocaval region, measuring up to 0.7 mm concerning for nodal metastasis.
3. No liver metastasis or ascites.
4. Few small ground-glass nodules at the RML and superior segment of the LLL probably inflammatory nodules. Attention on follow-up.
5. Multiple gallstones with contracted gallbladder suggestive of chronic cholecystitis.
S/P : -8/8/62 on bilat DJ stent due to tumor involve rt ureter
-LAR with short jejunum with TAH with BSO 8/8/62
patho Adeno CA moderate diff
plan adjuvant 5FU Degramont CMT
S/p mFOLFOX C1 18-20/10/62 no complication

Sign to save

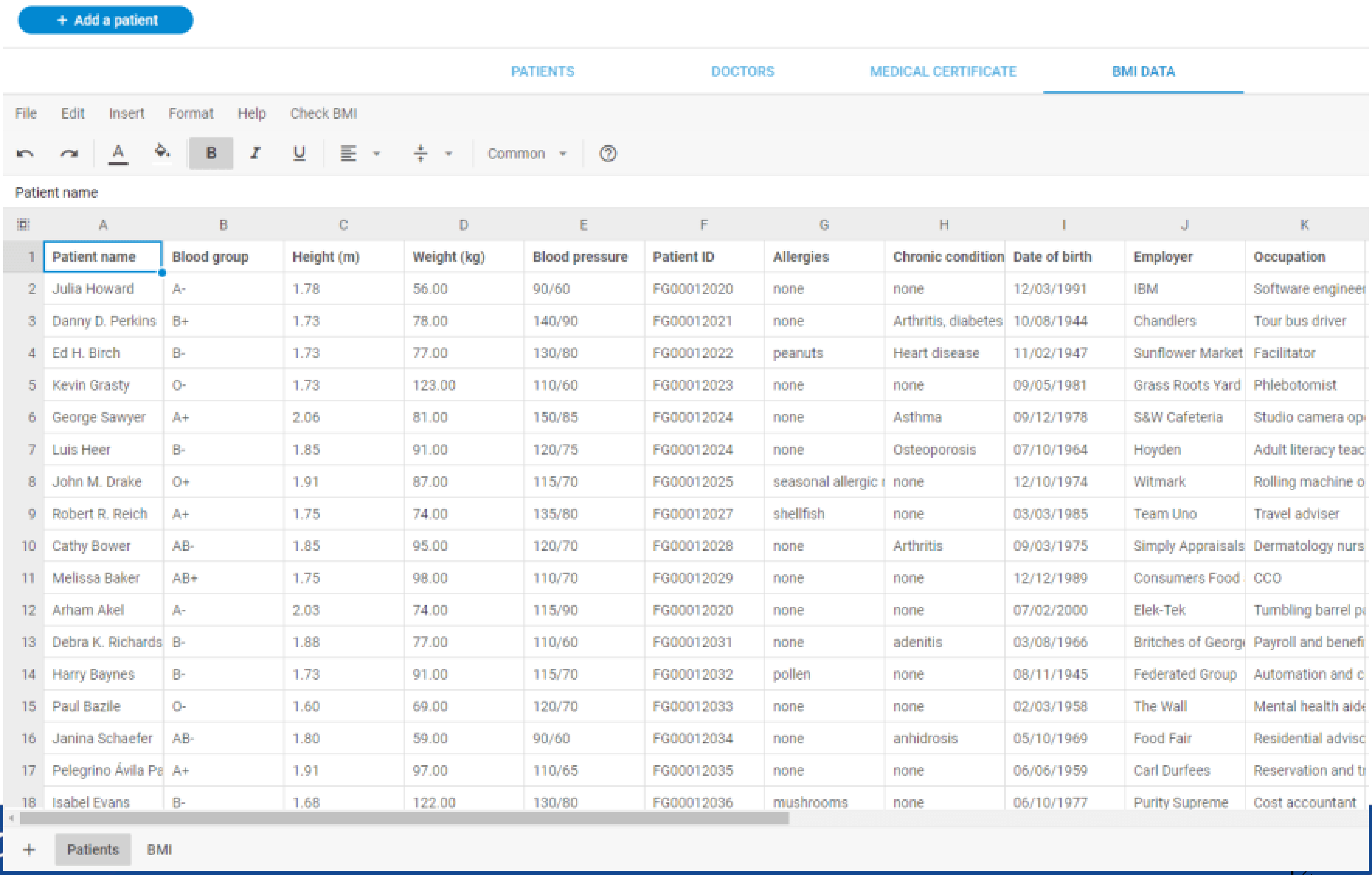
12 กพ 2563

หน้า 1 RH.0011

Textual form

Voice form
(wave)

Structured data



	A	B	C	D	E	F	G	H	I	J	K
1	Patient name	Blood group	Height (m)	Weight (kg)	Blood pressure	Patient ID	Allergies	Chronic condition	Date of birth	Employer	Occupation
2	Julia Howard	A-	1.78	56.00	90/60	FG00012020	none	none	12/03/1991	IBM	Software engineer
3	Danny D. Perkins	B+	1.73	78.00	140/90	FG00012021	none	Arthritis, diabetes	10/08/1944	Chandlers	Tour bus driver
4	Ed H. Birch	B-	1.73	77.00	130/80	FG00012022	peanuts	Heart disease	11/02/1947	Sunflower Market	Facilitator
5	Kevin Grasty	O-	1.73	123.00	110/60	FG00012023	none	none	09/05/1981	Grass Roots Yard	Phlebotomist
6	George Sawyer	A+	2.06	81.00	150/85	FG00012024	none	Asthma	09/12/1978	S&W Cafeteria	Studio camera op
7	Luis Heer	B-	1.85	91.00	120/75	FG00012024	none	Osteoporosis	07/10/1964	Hoyden	Adult literacy teac
8	John M. Drake	O+	1.91	87.00	115/70	FG00012025	seasonal allergic r	none	12/10/1974	Witmark	Rolling machine o
9	Robert R. Reich	A+	1.75	74.00	135/80	FG00012027	shellfish	none	03/03/1985	Team Uno	Travel adviser
10	Cathy Bower	AB-	1.85	95.00	120/70	FG00012028	none	Arthritis	09/03/1975	Simply Appraisals	Dermatology nurs
11	Melissa Baker	AB+	1.75	98.00	110/70	FG00012029	none	none	12/12/1989	Consumers Food	CCO
12	Arham Akel	A-	2.03	74.00	115/90	FG00012020	none	none	07/02/2000	Elek-Tek	Tumbling barrel p
13	Debra K. Richards	B-	1.88	77.00	110/60	FG00012031	none	adenitis	03/08/1966	Britches of Georg	Payroll and benefi
14	Harry Baynes	B-	1.73	91.00	115/70	FG00012032	pollen	none	08/11/1945	Federated Group	Automation and c
15	Paul Bazile	O-	1.60	69.00	120/70	FG00012033	none	none	02/03/1958	The Wall	Mental health aide
16	Janina Schaefer	AB-	1.80	59.00	90/60	FG00012034	none	anhidrosis	05/10/1969	Food Fair	Residential advisc
17	Pelegrino Ávila Pa	A+	1.91	97.00	110/65	FG00012035	none	none	06/06/1959	Carl Durfees	Reservation and ti
18	Isabel Evans	B-	1.68	122.00	130/80	FG00012036	mushrooms	none	06/10/1977	Purity Supreme	Cost accountant

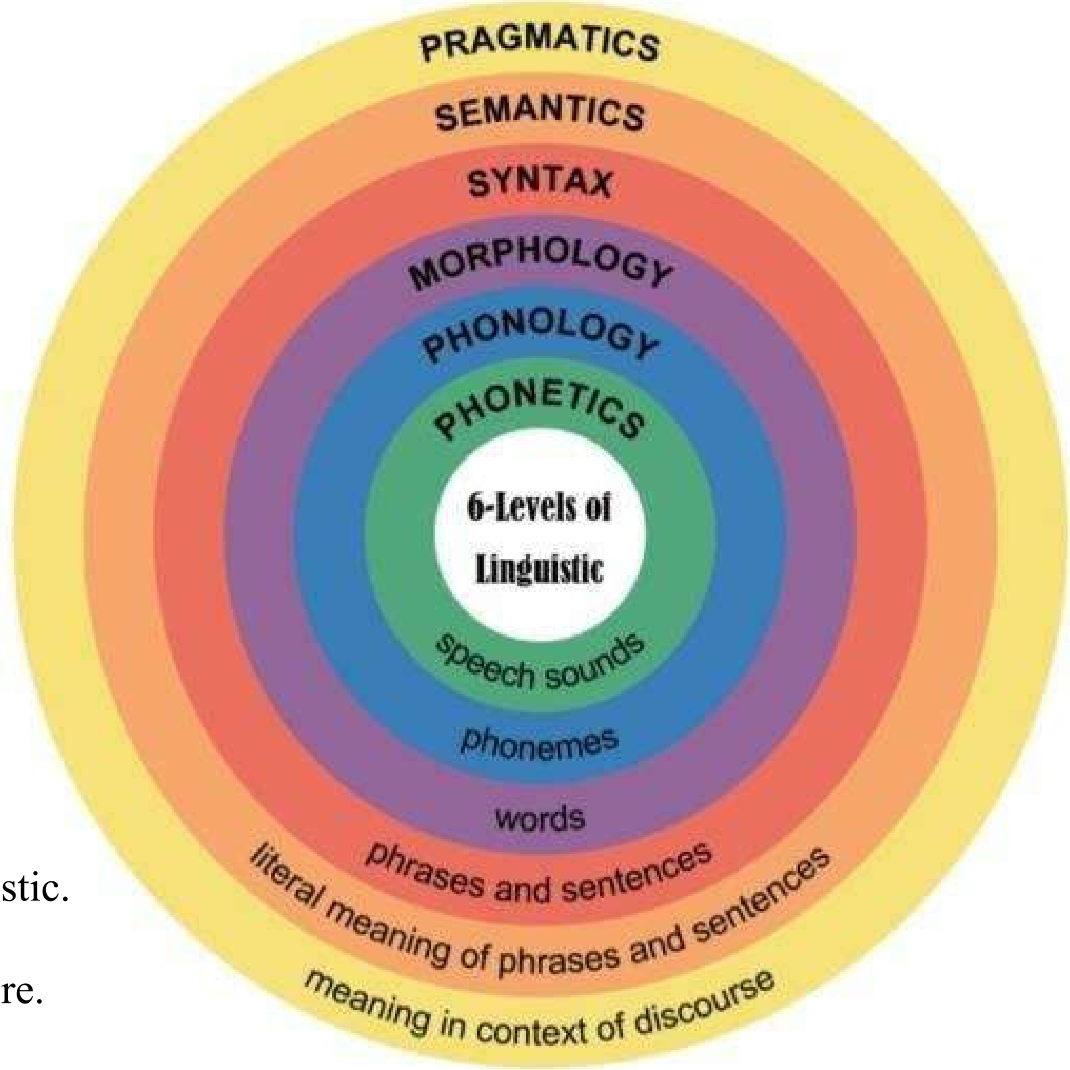
Real-world applications of NLP

alexa

IBM Tone analyser

Google Cloud Speech API

How does NLP work?

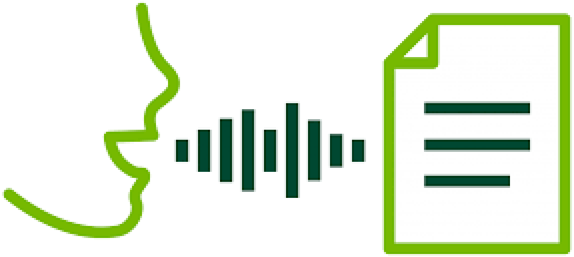


- NLP processes classified by level of linguistic.
- Involves several processes: tokenization, parsing, stemming, lemmatization, and more.

- Utilizes algorithms to extract meaning from text.
- Machine learning models play a crucial role in improving NLP accuracy

Phonetics, Phonology

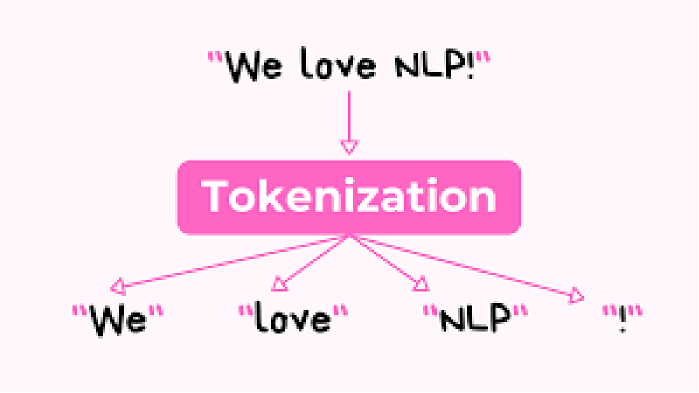
- Speech Recognition



- Pronunciation Modeling
  - Cardiology → kar dee ALL oh jee
  - Gastrohepatic → GAS troh heh PAT ik

Word & Morphology

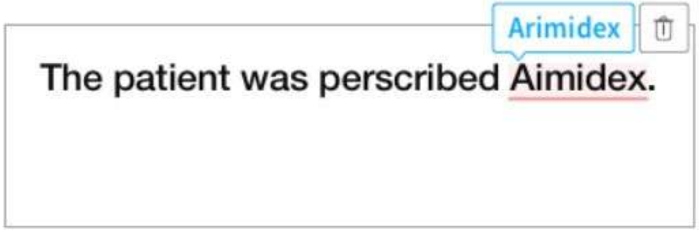
- Word
  - Tokenization
  - Spelling correction
- Morphology
  - Lemmatization / Stemming
  - Morphological segmentation



"We love NLP!"

Tokenization

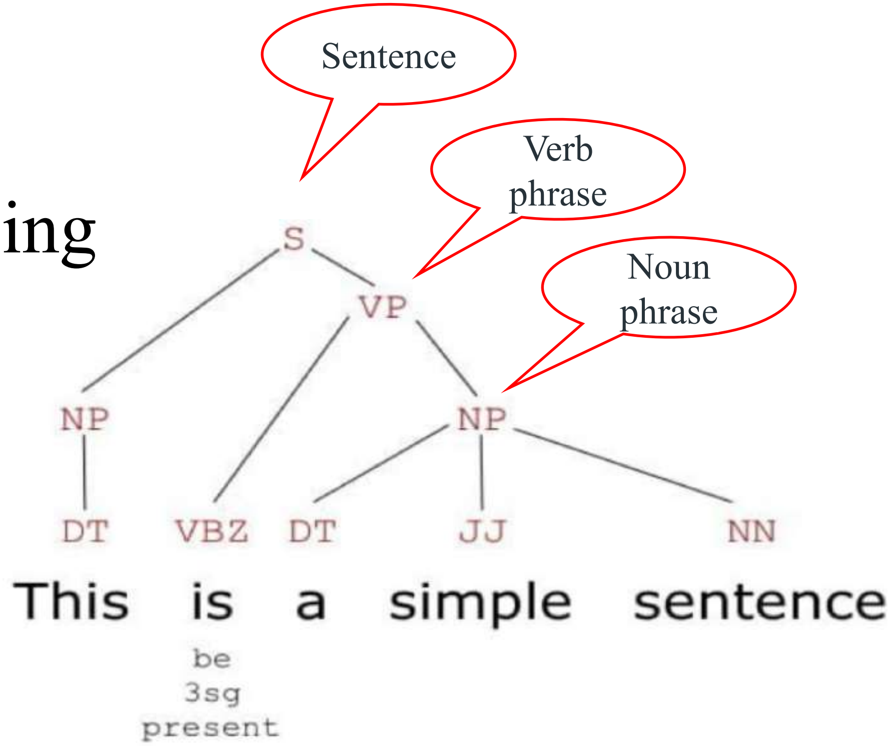
"We" "love" "NLP" "!"

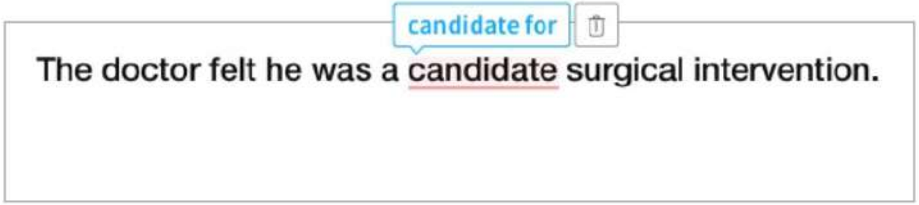


The patient was perscribed Aimidex.

Syntax

- Part of speech (POS) tagging
- Syntactic parsing
- Grammar checking

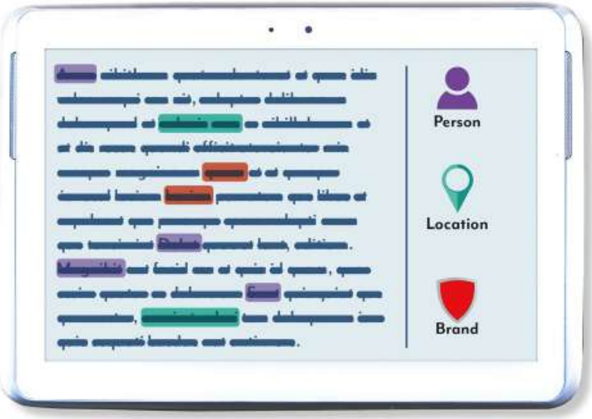


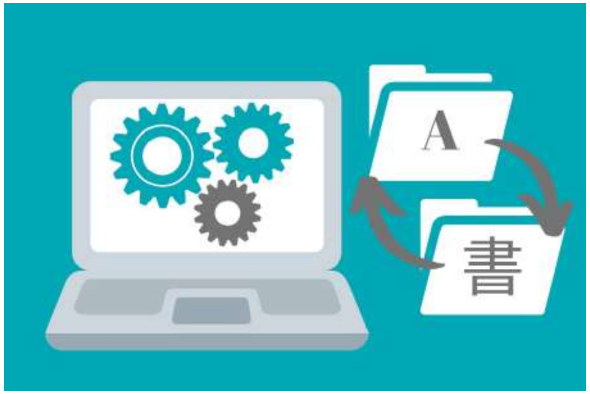


The doctor felt he was a candidate surgical intervention.

Semantics

- Named entity recognition (NER)
- Machine translation

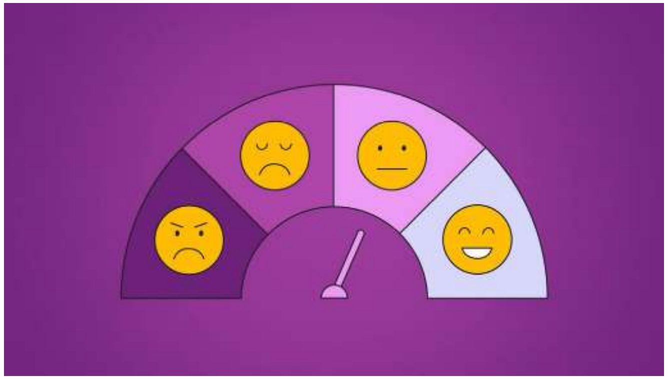


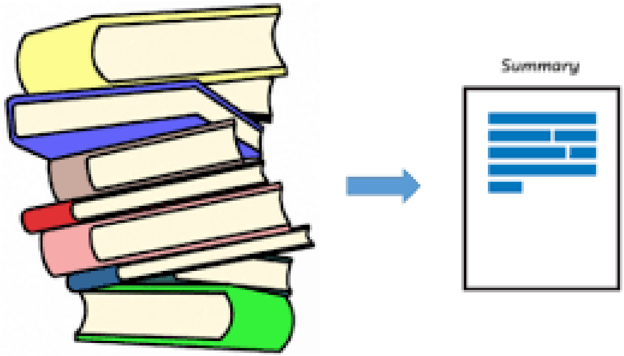


https://www.amperetranslations.com/blog/machine-translation-for-business-content-a-quick-guide/

Pragmatics

- Sentiment analysis
- Text summarization





https://sproutsocial.com/insights/sentiment-analysis/
https://www.amperetranslations.com/blog/machine-translation-for-business-content-a-quick-guide/

Common NLP techniques

- Text Preprocessing
- Tokenization
- POS Tagging
- NER
- Sentiment Analysis
- Text Classification
- Machine Translation
- Text Summarization
- etc.

Text Preprocessing

- **Definition**
    - Cleaning and transforming raw text into a usable format.
- **Common Techniques**
    - Lowercasing
    - Removing Punctuation (. , ? ! : ; “ ” ’ — - ( ) [ ] … / ‘ ’ { } | < > _ ~)
    - Removing StopWords (a, an, the, and, in, of, to, is, on, that, with, for, as, by)
    - Stemming (e.g., "prescribing" -> "prescrib")
    - Lemmatization (e.g., "diagnosed" -> "diagnosis")
- **Importance**
    - Enhances performance → improves the accuracy by reduce noise
    - Normalizing textual data → ensures consistency in text analysis

https://www.dreamstime.com/illustration/data-preprocessing.html

Tokenization

- **Definition**
  - Process of splitting text into individual words or phrases (tokens)
- **Common Techniques**
  - Word Tokenization
  - Sub-word Tokenization
  - Sentence Tokenization

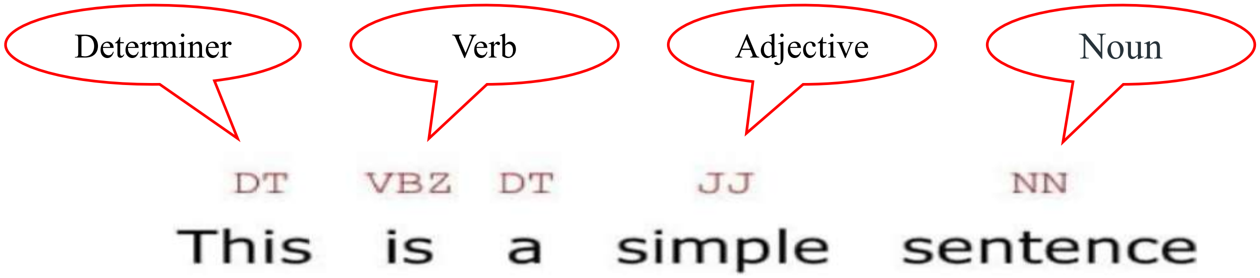
“Patient shows symptoms of fever and cough”

["Patient", "shows", "symptoms", "of", "fever", "and", "cough"]

https://medium.com/@zakhtar2020/nlp-preprocessing-tokenization-44ef71eead10

POS Tagging

- **Definition**
  - Assigning parts of speech to each word in a text
    - e.g., noun, verb, adjective.

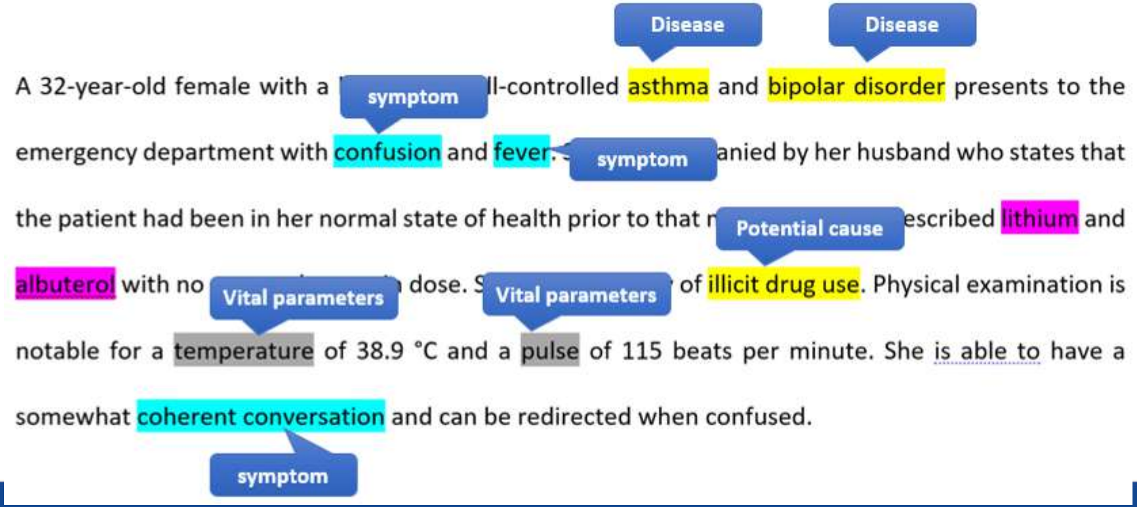


- **Importance**
  - Grammatical Structure → Understand sentence syntactic structure

https://byteiota.com/pos-tagging/

NER

- **Definition**
  - Identifying and classifying entities in text
    - e.g., people, locations, organizations
    - e.g., diseases, medications, procedures, medical terms

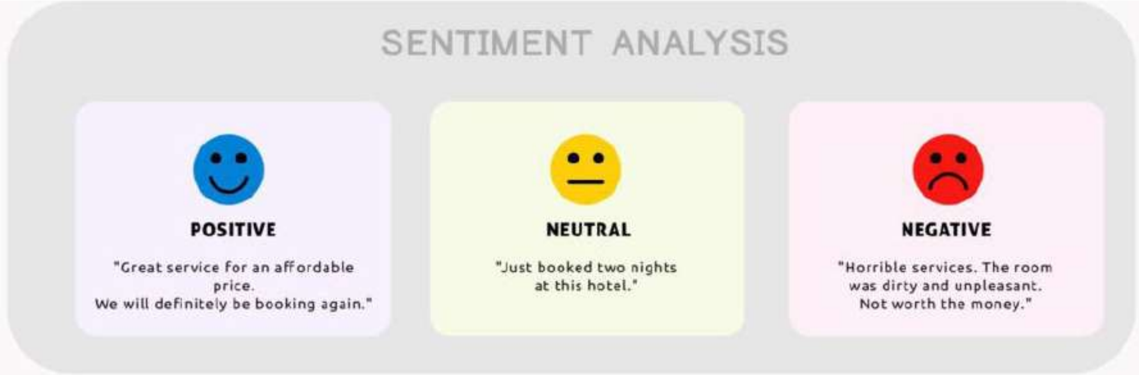


A 32-year-old female with a ll-controlled asthma and bipolar disorder presents to the emergency department with confusion and fever. anied by her husband who states that the patient had been in her normal state of health prior to that r escribed lithium and albuterol with no dose. S of illicit drug use. Physical examination is notable for a temperature of 38.9 °C and a pulse of 115 beats per minute. She is able to have a somewhat coherent conversation and can be redirected when confused.

https://sigma.ai/es/named-entity-recognition/

Sentiment Analysis

- **Definition**
  - Determining the emotional tone of a text
    - e.g., positive, negative, neutral
- **Example of tasks**
  - Analyzing patient feedback to understand patient satisfaction
  - To find support evidence for sentiment
    - "The treatment was excellent, but the wait time was too long."



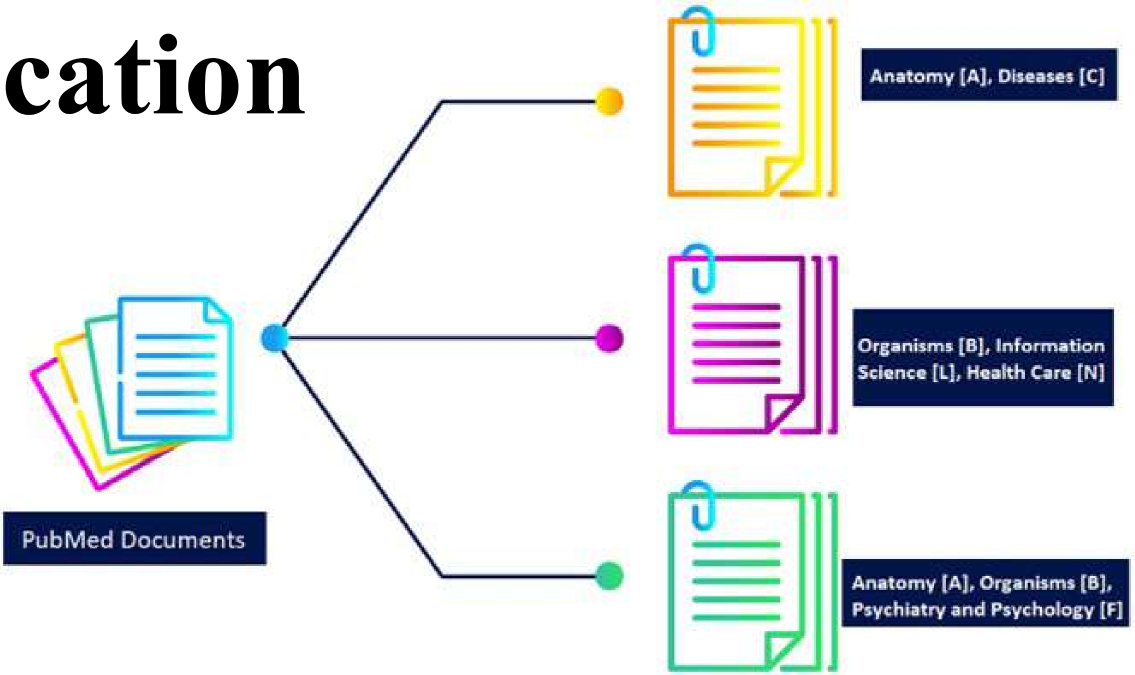
"Great service for an affordable price. We will definitely be booking again."

"Just booked two nights at this hotel."

"Horrible services. The room was dirty and unpleasant. Not worth the money."

https://www.expressanalytics.com/blog/social-media-sentiment-analysis/

Text Classification

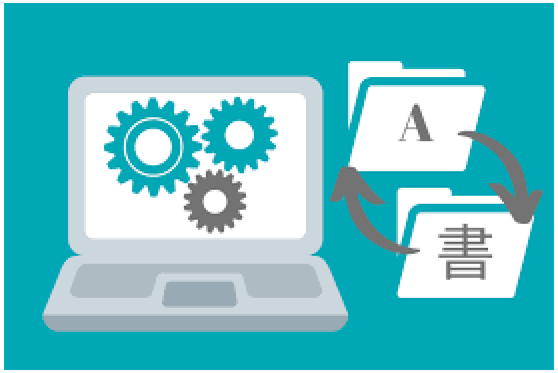

PubMed Documents
Anatomy [A], Diseases [C]
Organisms [B], Information Science [L], Health Care [N]
Anatomy [A], Organisms [B], Psychiatry and Psychology [F]

- **Definition**
  - Categorizing text into predefined classes
- **Common Techniques**
  - Supervised Learning
    - Naive Bayes, SVM, neural networks, etc.
  - Unsupervised Learning
    - Clustering similar texts without labeled data

https://www.kaggle.com/datasets/owaiskhan9654/pubmed-multilabel-text-classification

Machine Translation

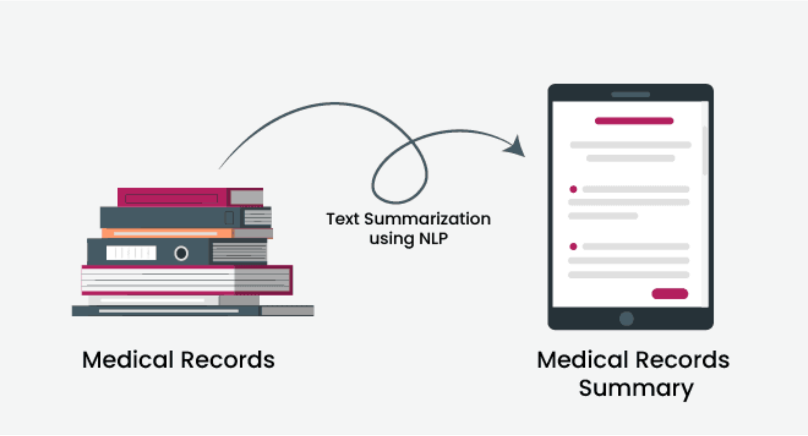
- **Definition**
  - Automatically translating text from one to another language
- **Common Techniques**
  - Rule-Based Machine Translation
  - Statistical Machine Translation
  - Neural Machine Translation



https://www.amperetranslations.com/blog/machine-translation-for-business-content-a-quick-guide/

Text Summarization

- **Definition**
  - Condensing text to its essential points while preserving meaning
- **Common Techniques**
  - Extractive Summarization
  - Abstractive Summarization



https://marutitech.medium.com/ai-powered-medical-records-summarization-a-game-changer-2167929faa42

Applications and use cases utilizing NLP in healthcare

- In Faculty of Medicine, Ramathibodi Hospital
  - ICD-10 classification from discharge summaries
  - AI for literature screening in systematic reviews

ICD-10 classification from discharge summaries

Impacts of assigning the ICD (Benefits)

Population health

- Policy planning
- Health surveillance
- Care monitoring
- Reimbursement
- Healthcare research

Healthcare provider

- Patient data documentation
- Integrated care

Patient

- Quality of care
- Patient safety

www.vectorstock.com

Impacts of assigning the ICD (Burdens)

Increase workload

- Increase workload by coding practice
- Decrease in clinical care productivities

Time consumption

- Coding practice time (charts per hour)
  - 1.43-2.08 (United States)
  - 3.75 (Canada)
  - 3-4 (Thailand)

- Prasanwong C. Medical coding practices in Thailand [Internet]. Health Systems Research Institute; 2002
- Libicki MC, Brahmakulam IT. The costs and benefits of moving to the ICD-10 code sets. Santa Monica, CA: RAND; 2004. 63 p.
- Nachimson S. Documentation, documentation, documentation. The key to ICD-10 readiness. Md Med. 2014;15(1):20.
- พระราชบัญญัติ ระเบียบข้าราชการพลเรือน (ฉบับที่ ๒) พ.ศ. ๒๕๕๘

Impacts of assigning the ICD (Burdens) cont.

Resource consumption

- Costs
  - Hiring for coders (Thailand, 2015)

➢Nurse ≈ ฿20,000 – ฿30,000 // Clerk ≈ ฿15,000

  - Training coders (US)

➢$500 - $1500 per one coder (2004, 2014) [20,000฿ to 50,000฿]

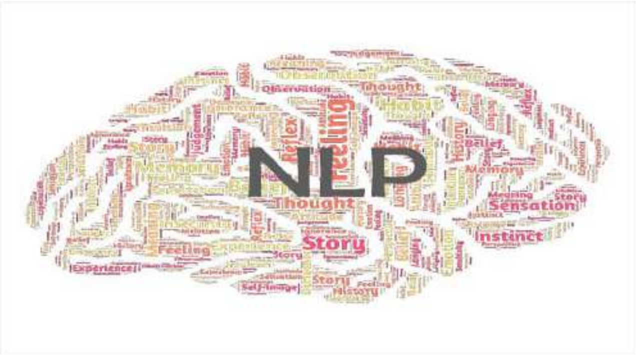
Errors from coding

- 17.1 to 76.9% of errors from manual coding (1988–2005)
- 62.1 to 92.7% of errors for principal diagnosis (2017, Thailand)

- AHIMA. ICD-10-CM Field Testing Project: Report on Findings: Perceptions, Ideas and Recommendations from Coding Professionals Across the Nation. ICD-10-CM Field Testing Project: 2003
- Weems, Shelley; Fenton, Susan H.. "Results from the Veterans Health Administration ICD-10-CM/PCS Coding Pilot Study" Perspectives in Health Information Management (Summer, July 2015).
- Johnson K. Implementation of ICD-10: Experiences and Lessons Learned from a Canadian Hospital. 2004 Oct 15
- Hsia et al. 1988;Fischer et al. 1992; Benesch et al. 1997; Faciszewski, Broste, 1997; Goldstein 1998
- Quan H, Li B, Saunders LD, et al. Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. Health Serv Res. 2008
- Sukanya C. Validity of Principal Diagnoses in Discharge Summaries and ICD-10 Coding Assessments Based on National Health Data of Thailand. Healthc Inform Res. 2017

Manual ICD coding

Automate coding for ICD-10



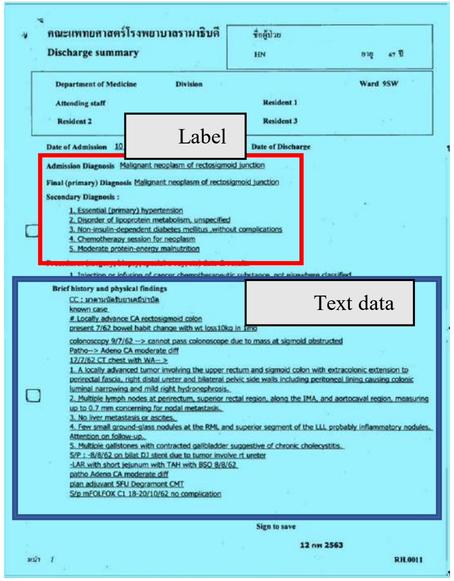
Reduce workload

Reduce time and cost

Reduce errors

https://insightsnlp.com/why-learn-nlp

ICD-10 classification from discharge summary



คณะแพทยศาสตร์โรงพยาบาลรามาธิบดี
Discharge summary

ชื่อผู้ป่วย
HN อายุ 67 ปี

Department of Medicine **Division** **Ward** 9SW
Attending staff **Resident 1**
Resident 2 **Resident 3**

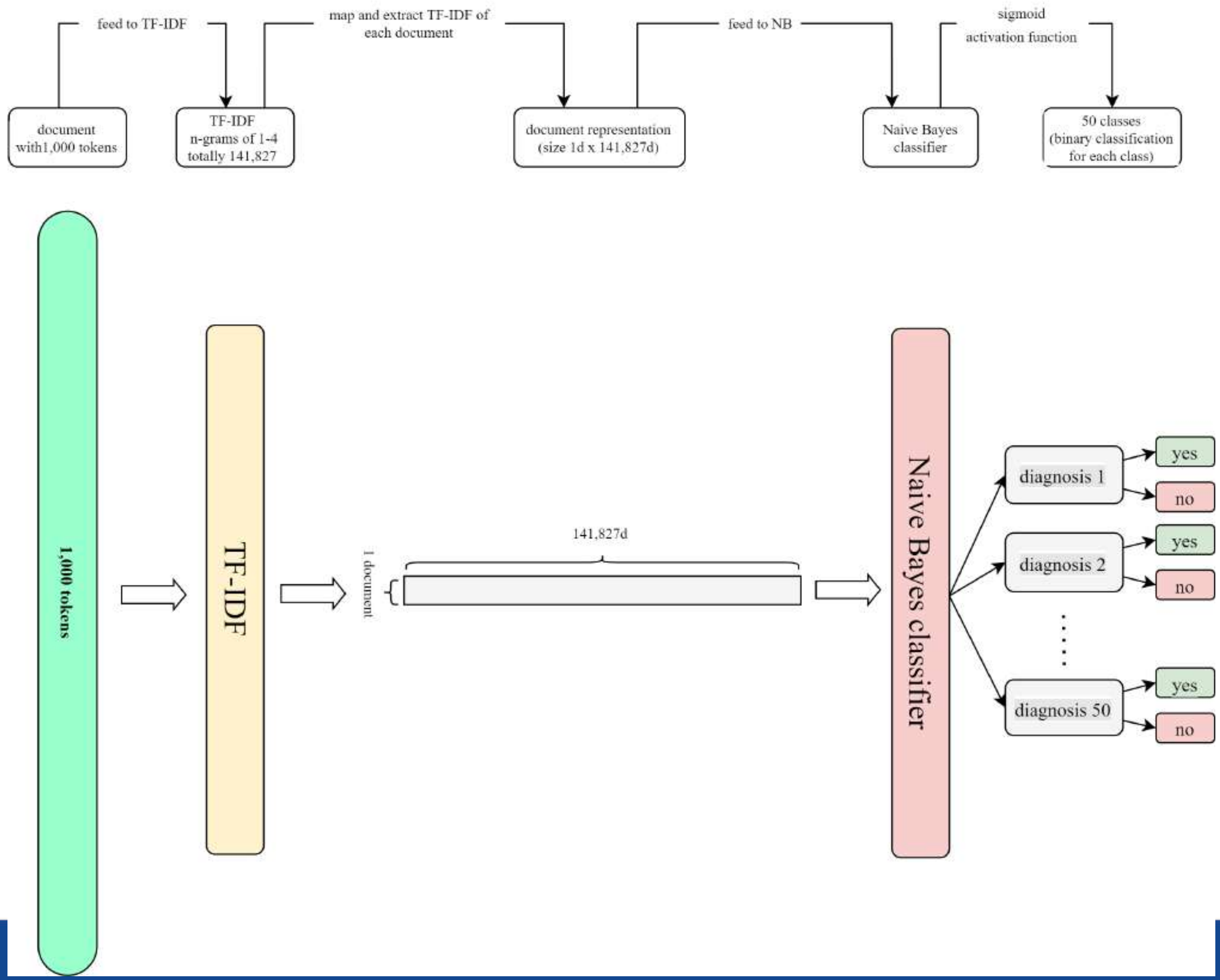
Date of Admission 10 **Date of Discharge**

Admission Diagnosis Malignant neoplasm of rectosigmoid junction
Final (primary) Diagnosis Malignant neoplasm of rectosigmoid junction
Secondary Diagnosis :
1. Essential (primary) hypertension
2. Disorder of lipoprotein metabolism, unspecified
3. Non-insulin-dependent diabetes mellitus ,without complications
4. Chemotherapy session for neoplasm
5. Moderate protein-energy malnutrition

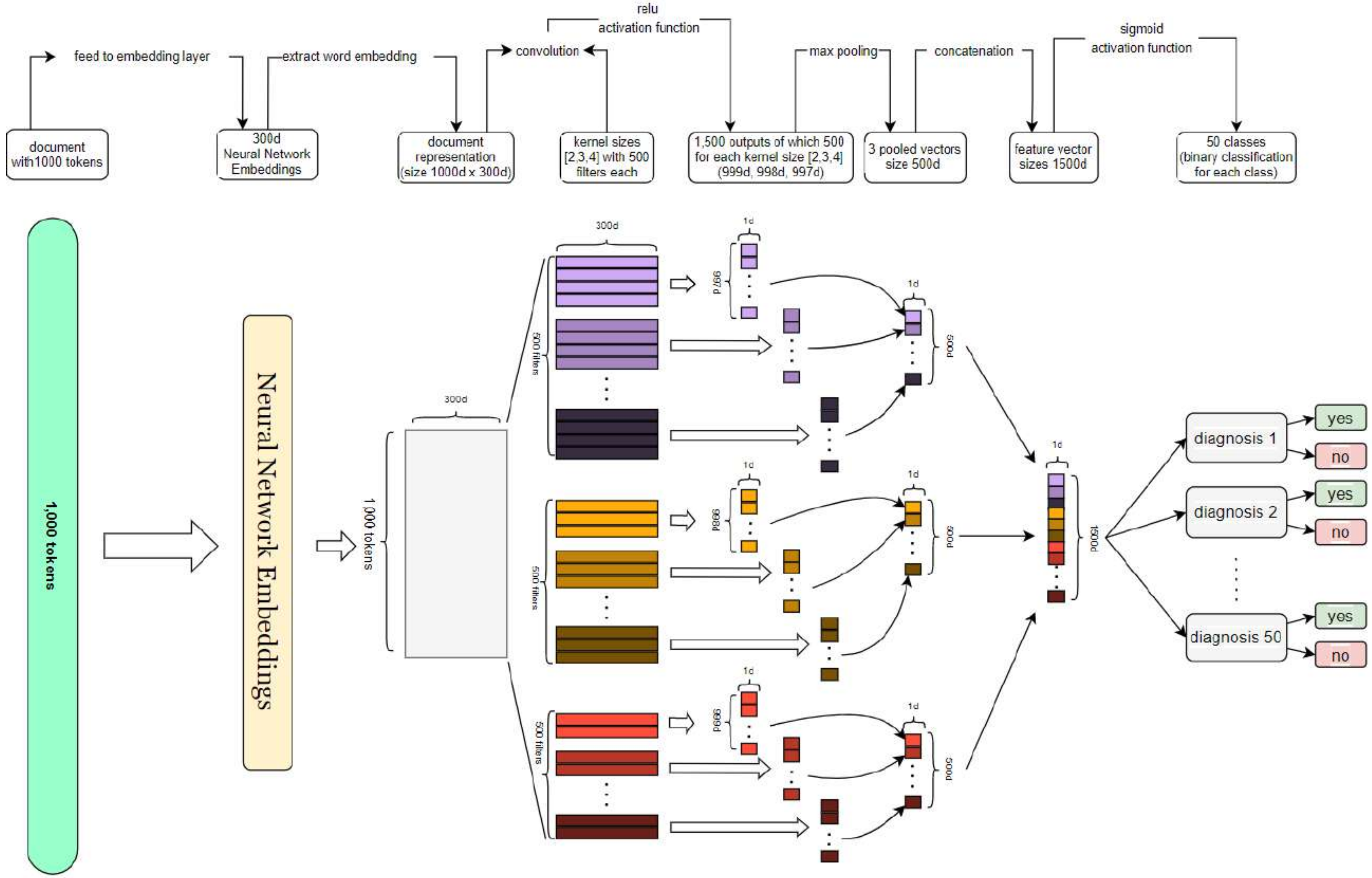
Procedures (surgery, biopsy, special x-ray, etc) date & results
1. Injection or infusion of cancer chemotherapeutic substance, not elsewhere classified

Brief history and physical findings
CC : มาตามนัดรับยาเคมีบำบัด
known case
Locally advance CA rectosigmoid colon
present 7/62 bowel habit change with wt loss10kg in 1mo
colonoscopy 9/7/62 --> cannot pass colonoscope due to mass at sigmoid obstructed
Patho--> Adeno CA moderate diff
12/7/62 CT chest with WA-- >
1. A locally advanced tumor involving the upper rectum and sigmoid colon with extracolonic extension to perirectal fascia, right distal ureter and bilateral pelvic side walls including peritoneal lining causing colonic luminal narrowing and mild right hydronephrosis.
2. Multiple lymph nodes at perirectum, superior rectal region, along the IMA, and aortocaval region, measuring up to 0.7 mm concerning for nodal metastasis.
3. No liver metastasis or ascites.
4. Few small ground-glass nodules at the RML and superior segment of the LLL probably inflammatory nodules. Attention on follow-up.
5. Multiple gallstones with contracted gallbladder suggestive of chronic cholecystitis.
S/P : -8/8/62 on bilat DJ stent due to tumor involve rt ureter
-LAR with short jejunum with TAH with BSO 8/8/62
patho Adeno CA moderate diff
plan adjuvant 5FU Degramont CMT
S/p mFOLFOX C1 18-20/10/62 no complication

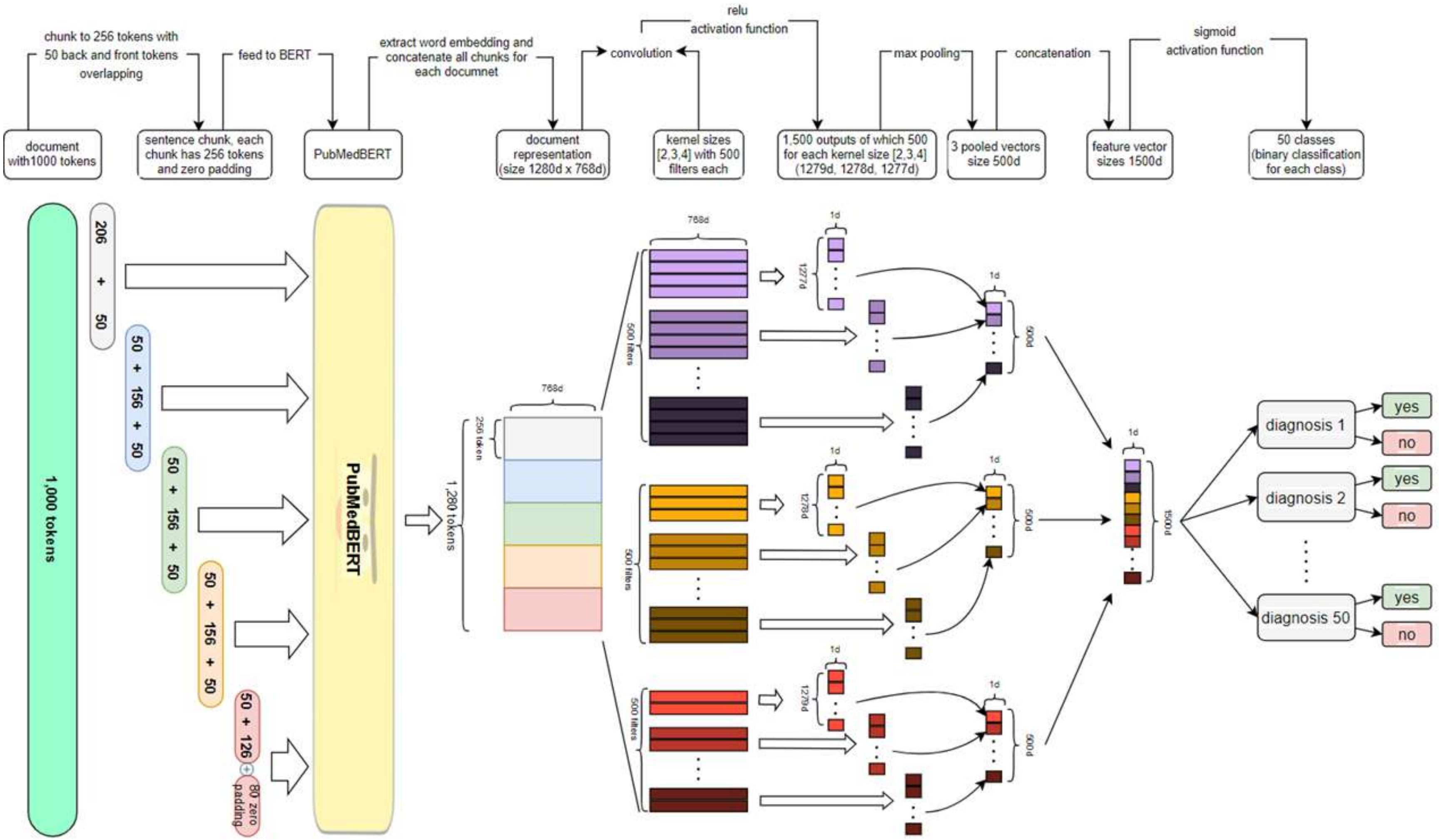
Sign to save
12 กพ 2563
หน้า 1 **RH.0011**

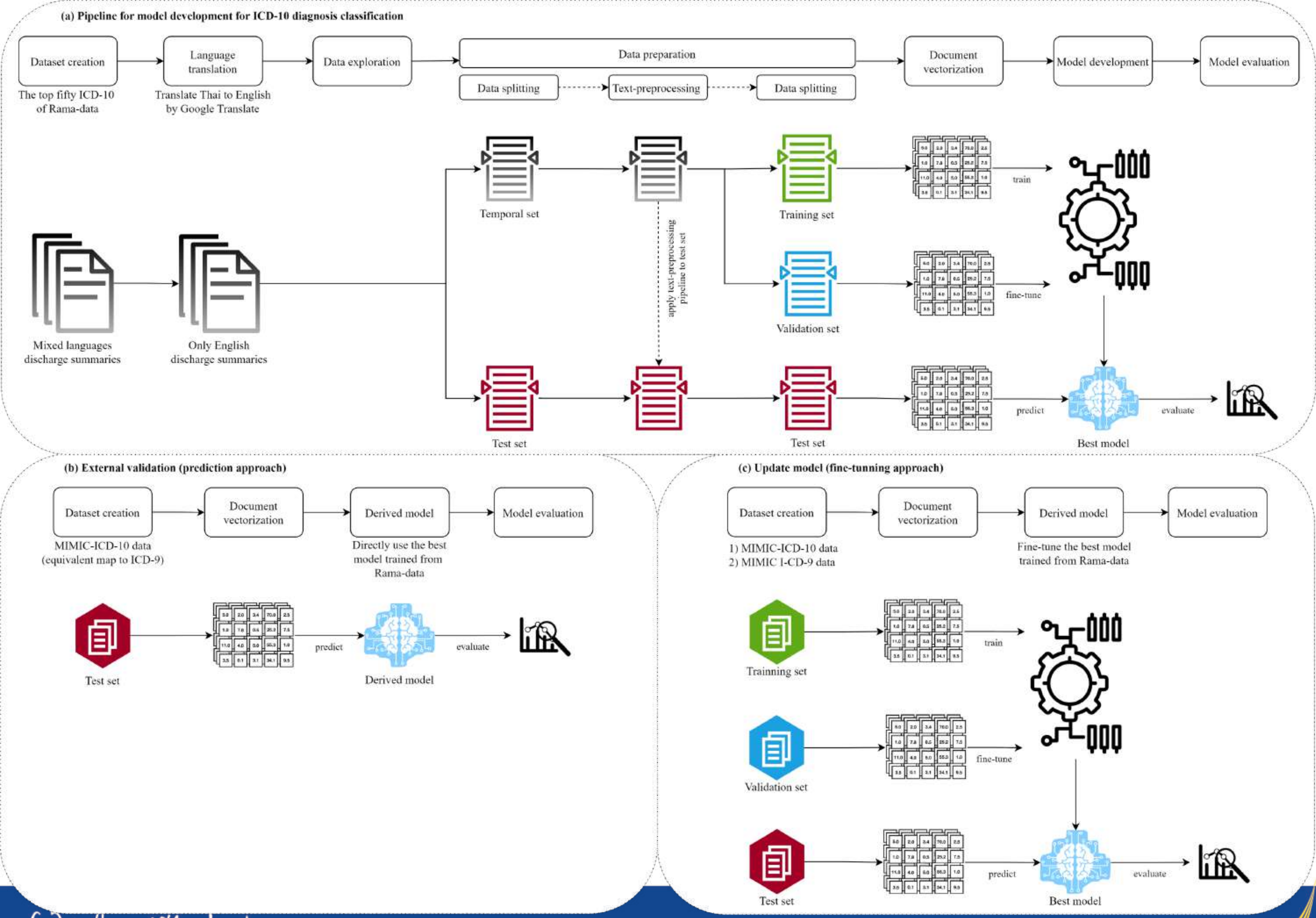

feed to TF-IDF
map and extract TF-IDF of each document
feed to NB
sigmoid activation function
document with1,000 tokens
TF-IDF n-grams of 1-4 totally 141,827
document representation (size 1d x 141,827d)
Naive Bayes classifier
50 classes (binary classification for each class)
1,000 tokens
TF-IDF
1 document
141,827d
Naive Bayes classifier
diagnosis 1
yes
no
diagnosis 2
yes
no
diagnosis 50
yes
no

CNN with neural network embedding



CNN with PubMedBERT


chunk to 256 tokens with 50 back and front tokens overlapping
feed to BERT
extract word embedding and concatenate all chunks for each documnet
convolution
relu activation function
max pooling
concatenation
sigmoid activation function
document with1000 tokens
sentence chunk, each chunk has 256 tokens and zero padding
PubMedBERT
document representation (size 1280d x 768d)
kernel sizes [2,3,4] with 500 filters each
1,500 outputs of which 500 for each kernel size [2,3,4] (1279d, 1278d, 1277d)
3 pooled vectors size 500d
feature vector sizes 1500d
50 classes (binary classification for each class)
1,000 tokens
206 + 50
50 + 156 + 50
50 + 156 + 50
50 + 156 + 50
50 + 126
80 zero padding
PubMedBERT
1,280 tokens
256 token
768d
768d
500 filters
500 filters
500 filters
1d
1277d
1278d
1279d
500d
500d
500d
1500d
diagnosis 1
yes
no
diagnosis 2
yes
no
diagnosis 50
yes
no



(a) Pipeline for model development for ICD-10 diagnosis classification

Dataset creation → Language translation → Data exploration → Data preparation → Document vectorization → Model development → Model evaluation

Data splitting → Text-preprocessing → Data splitting

The top fifty ICD-10 of Rama-data

Translate Thai to English by Google Translate

Mixed languages discharge summaries

Only English discharge summaries

Temporal set

apply text-preprocessing pipeline to test set

Training set

Validation set

Test set

Test set

train

fine-tune

predict

evaluate

Best model

(b) External validation (prediction approach)

Dataset creation → Document vectorization → Derived model → Model evaluation

MIMIC-ICD-10 data (equivalent map to ICD-9)

Directly use the best model trained from Rama-data

Test set

predict

Derived model

evaluate

(c) Update model (fine-tunning approach)

Dataset creation → Document vectorization → Derived model → Model evaluation

1) MIMIC-ICD-10 data
2) MIMIC I-CD-9 data

Fine-tune the best model trained from Rama-data

Trainning set

Validation set

Test set

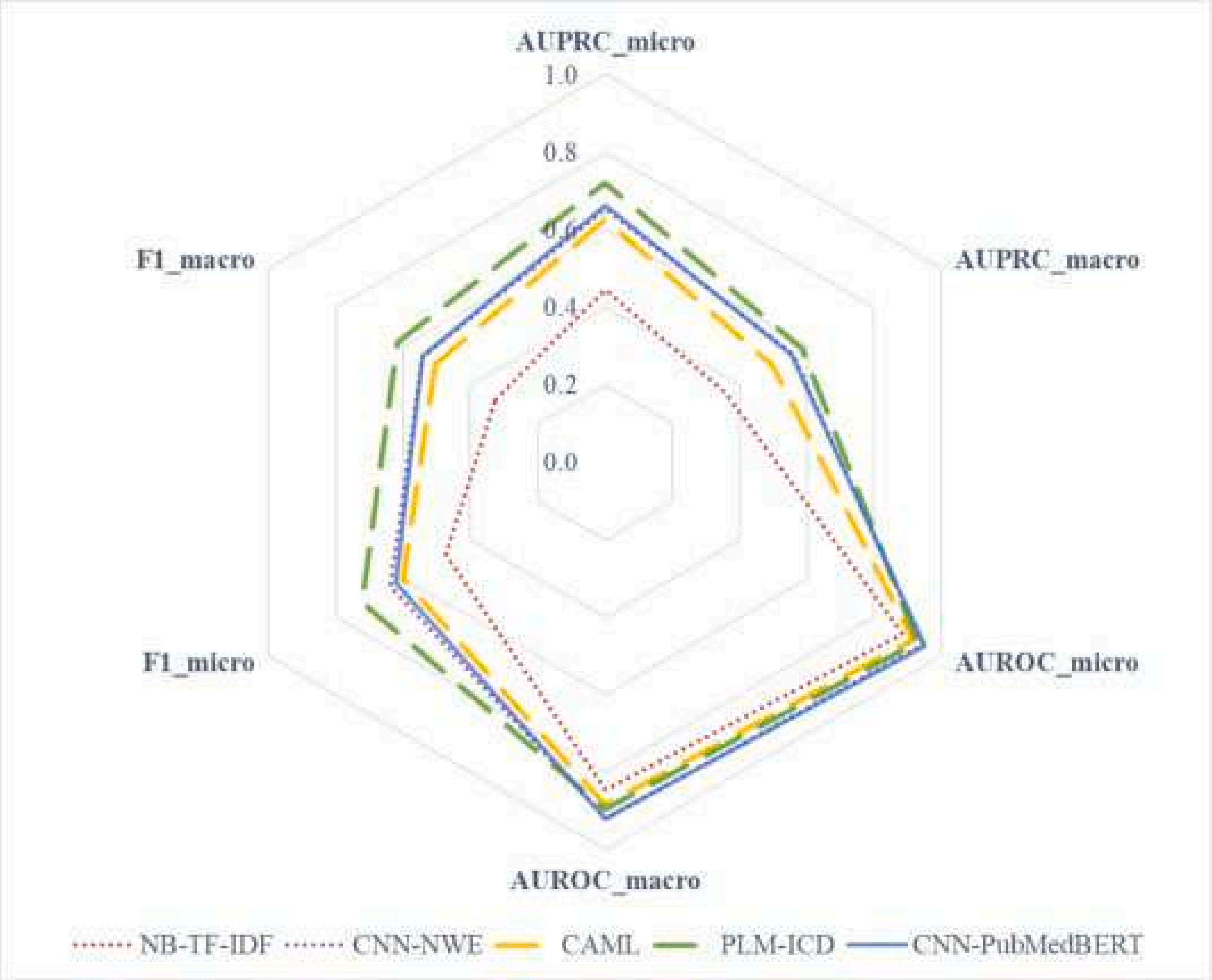
train

fine-tune

predict

evaluate

Best model

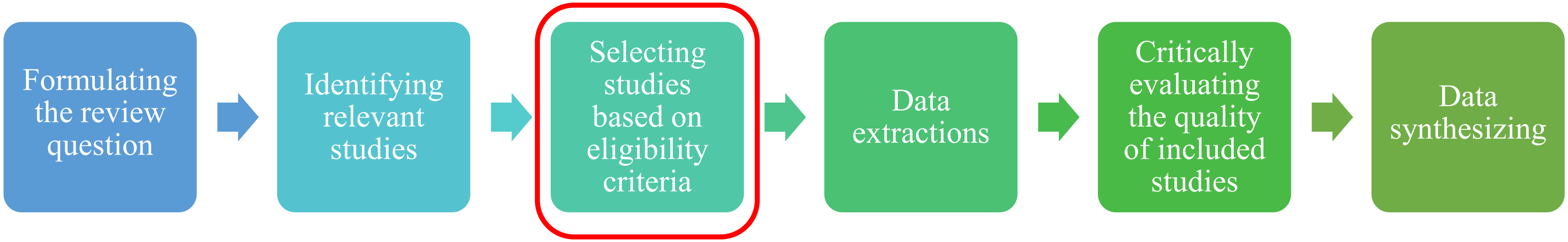

AUPRC_micro
1.0
0.8
0.6
0.4
0.2
0.0
AUPRC_macro
AUROC_micro
AUROC_macro
F1_micro
F1_macro
NB-TF-IDF
CNN-NWE
CAML
PLM-ICD
CNN-PubMedBERT

AI for literature screening in systematic reviews

What is SR?

- Rigorous and comprehensive method to synthesize existing research findings on a specific topic or question.
- Commonly used in healthcare and other fields to inform decision-making, policy development, and further research.

SR processes



Formulating the review question

Identifying relevant studies

Selecting studies based on eligibility criteria

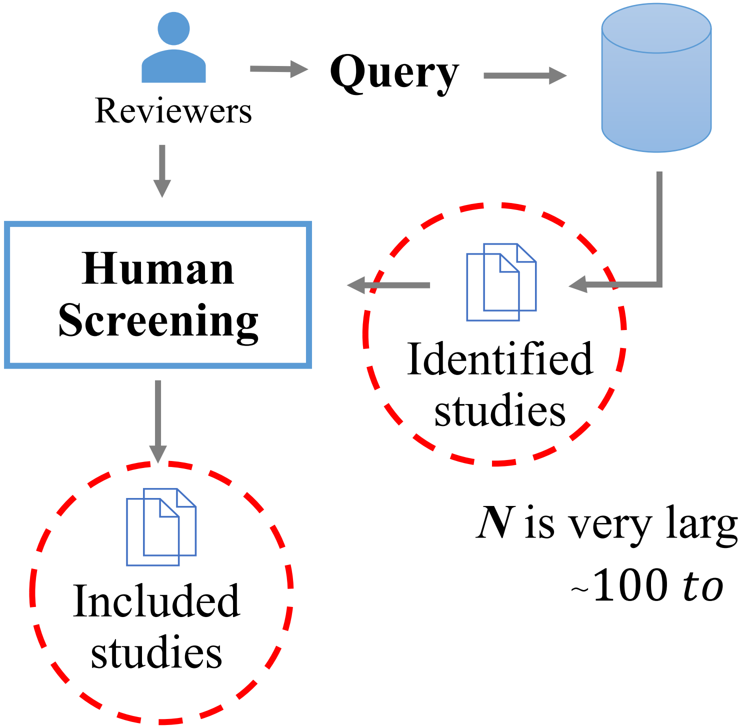
Data extractions

Critically evaluating the quality of included studies

Data synthesizing

The challenges in SR

- Workload



Reviewers

Query

Human Screening

Identified studies

Included studies

n is very small

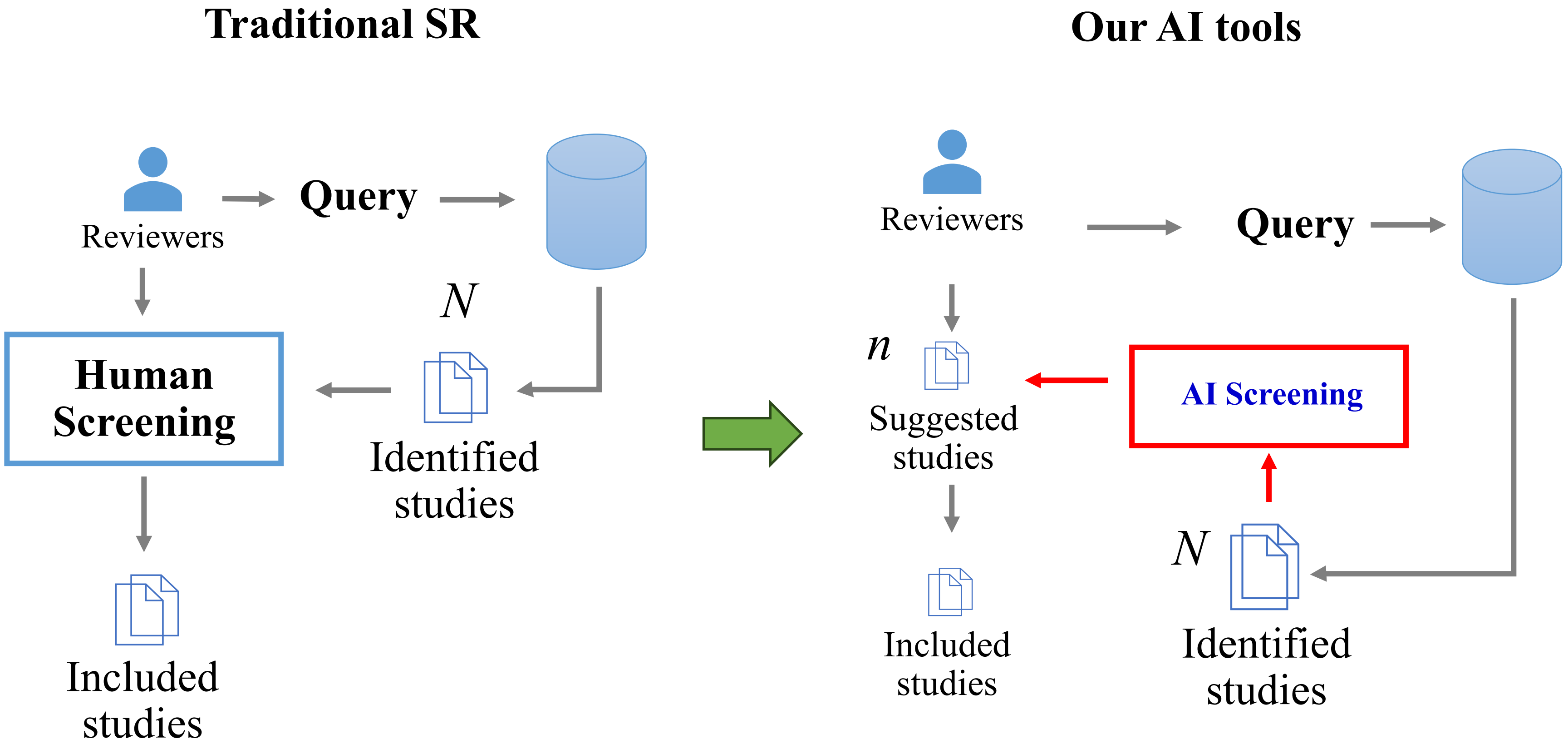
$\sim 0.8\% \ - 27\%$

N is very large:

$\sim 100\ to\ 20K\ studies$

1) Kontonatsios G, et al., 2020

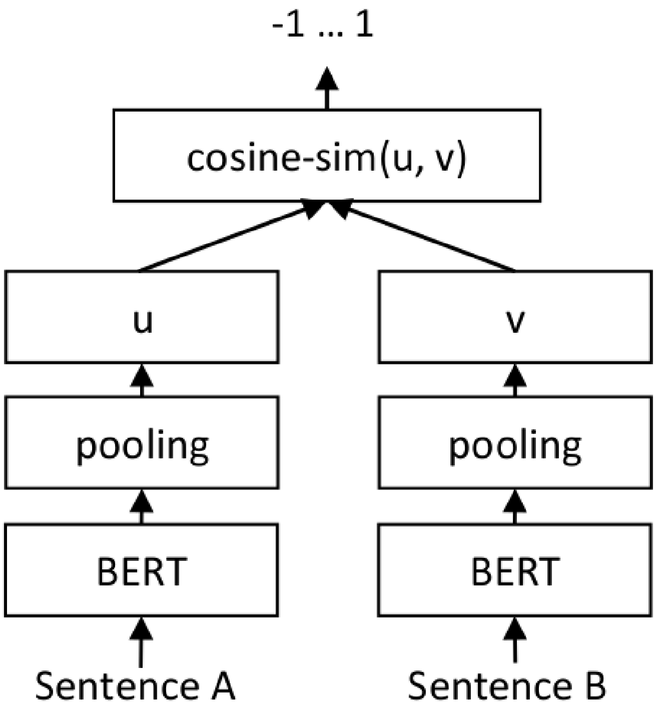
Application of AI in SR



Model development framework

Training framework:	Few-shot Learning
Feature vector representation:	SentenceBERT
Pre-trained:	all-mpnet-base-v2
Loss function:	Cosine similarity

batch_size	8
epochs	1
optimizer_params = {“lr”}	$2e^{-05}$
max_seq_length	384
word_embedding_dimension	768
Layer	● Transformer ● Pooling ● Normalize



-1 … 1

cosine-sim(u, v)

u

v

pooling

pooling

BERT

BERT

Sentence A

Sentence B

Comparison of our tool with existing tools

Performance

Tools	Researchers	Number of SRs	Reduced workload (%)	Sensitivity (%)
EPPI-Reviewer	Tsou A, et al., 2020	3	8.68 – 38.30	100
RobotAnalyst	Reddy SM, et al., 2020	1	30.69	100
Abstrackr	Tsou A, et al., 2020	3	3.99 – 48.41	100
	Gates A, et al., 2018	4	9.50 – 88.40	79 - 96
Rayyan	Valizadeh A, et al., 2022	3	20	87 - 98
DistillerSR	Hamel C, et al., 2020	10	30.00 – 72.50	95
AISR	**This research**	**9**	**51.11 – 97.67**	**100**

Applications and use cases utilizing NLP in healthcare

- In other real-world setting
  - Clinical Documentation Improvement (CDI)
  - Patient Data Extraction from EHRs
  - Predictive Analytics for Patient Outcomes

Clinical Documentation Improvement (CDI)

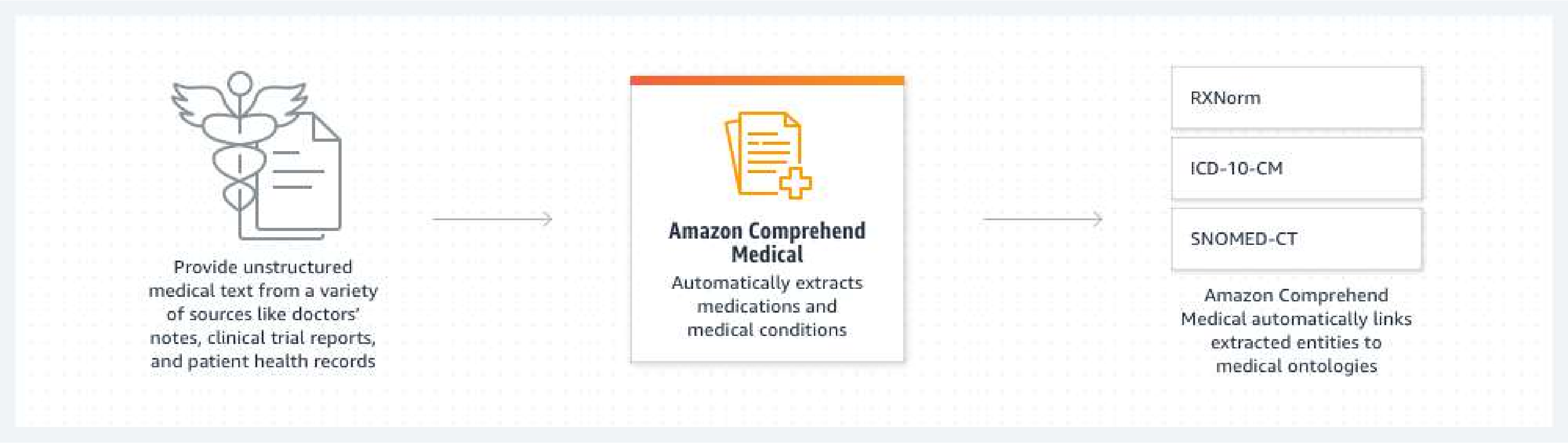
- 3M M*Modal computer-assisted physician documentation (CAPD)
  - Cloud-based model helping enhance clinical documentation by using NLP
    - To identify and correct errors or omissions in patient records.
    - To assign ICD codes

Clinical Documentation Improvement (CDI)

- Dragon Medical One
  - Uses NLP-powered speech recognition to allow clinicians to document patient encounters more accurately and efficiently

Patient Data Extraction from EHRs

- Amazon Comprehend Medical
  - Extracts structured information like medical conditions and treatments from unstructured EHR text



Provide unstructured medical text from a variety of sources like doctors' notes, clinical trial reports, and patient health records

Amazon Comprehend Medical

Automatically extracts medications and medical conditions

RXNorm

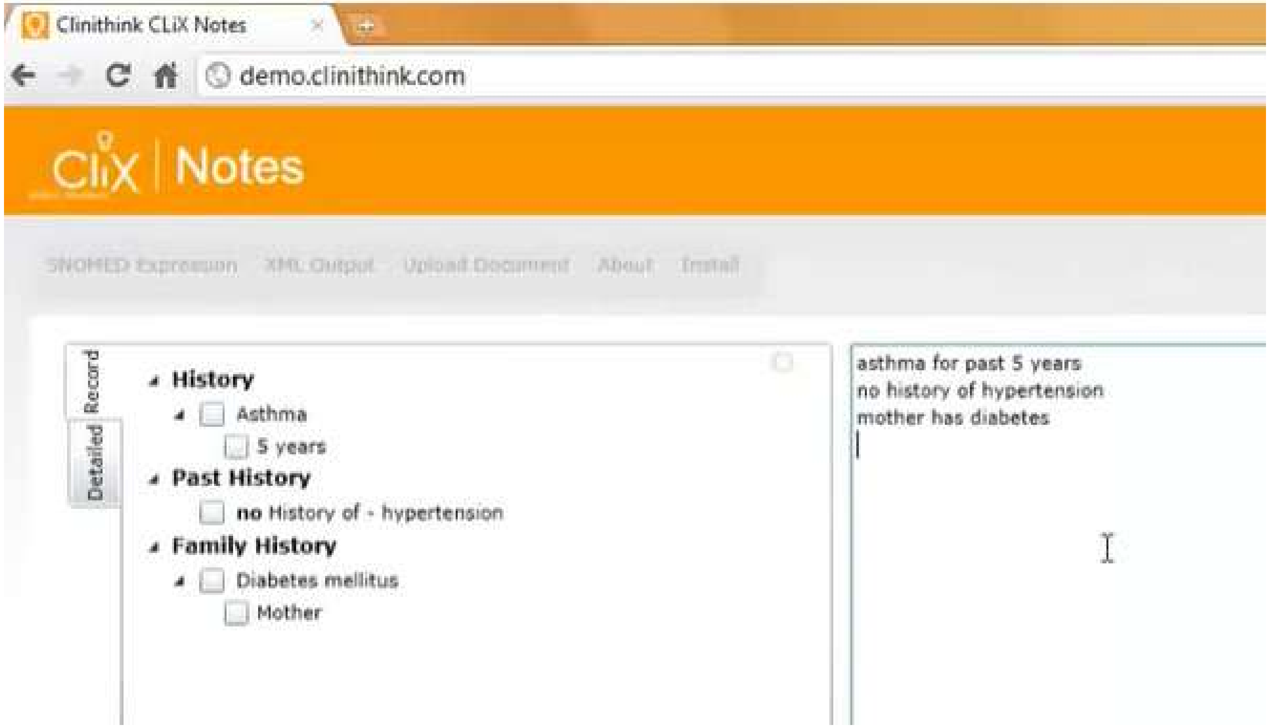
ICD-10-CM

SNOMED-CT

Amazon Comprehend Medical automatically links extracted entities to medical ontologies

Patient Data Extraction from EHRs

- Clinitink
  - NLP technology is used to process and analyze unstructured clinical data.
  - Extract meaningful clinical information, such as diagnoses, symptoms, and procedures, and convert them into structured data.



- History
  - ☐ Asthma
    - ☐ 5 years
- Past History
  - ☐ no History of - hypertension
- Family History
  - ☐ Diabetes mellitus
    - ☐ Mother

asthma for past 5 years
no history of hypertension
mother has diabetes

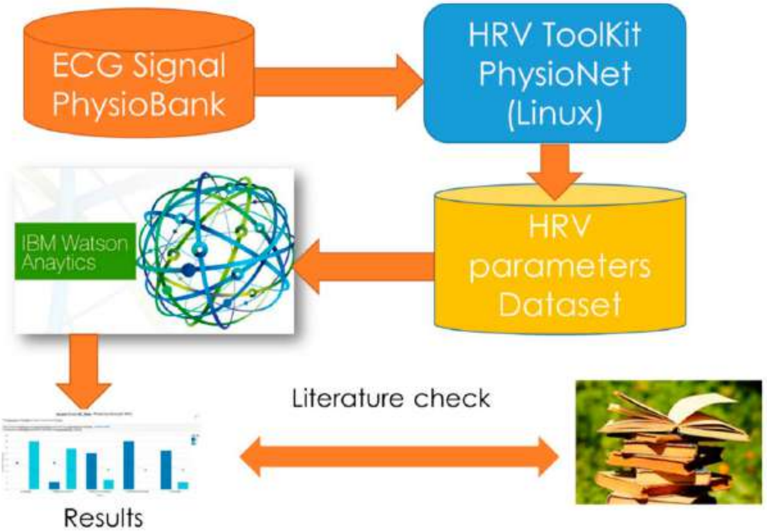
https://www.youtube.com/watch?v=Mb4K7Rjs6yk

Predictive Analytics for Patient Outcomes

- IBM Watson Health
  - Uses NLP to analyze patient records and predict outcomes like readmission risk and diseases.
  - Watson analytics to identify HF patients analyzing only the ECG summary.

Electrocardiogram (ECG)

Heart Rate Variability (HRV)



Q & A

THANK YOU