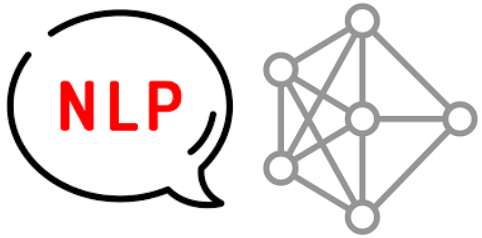
CHULA ENGINEERING
Foundation toward Innovation

COMPUTER



211594: Introduction to Natural Language Processing (NLP)

Peerapon Vateekul & Ekapol Chuangsuwanich
Department of Computer Engineering,
Faculty of Engineering, Chulalongkorn University

Peerapon Vateekul, Ph.D.

Ekapol Chuangsuwanich, Ph.D.

Can Udomcharoenchaikit
P'Can (TA)

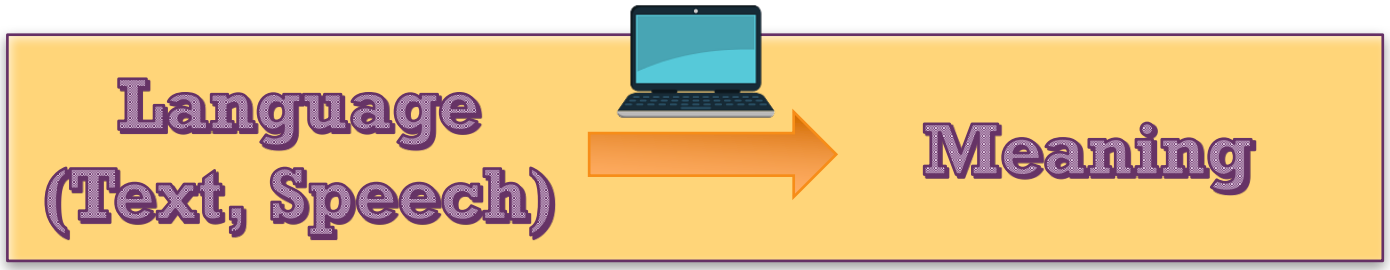
Nattachai Tretasayuth
P'Boss (TA)

Outlines

- What is NLP?
  - Definition
  - Levels of understanding in NLP
  - NLP today
- Why NLP is hard?
  - Ambiguity
  - Issues in Thai NLP
- NLP & Text mining
  - NLP pipeline (English & Thai)
  - History of NLP techniques
  - Case Study: Determiner placement
- Deep Learning
  - Definition
  - Reasons for exploring Deep Learning
- NLP Tools
- Course Logistics
- Google Cloud Demo

What is NLP?

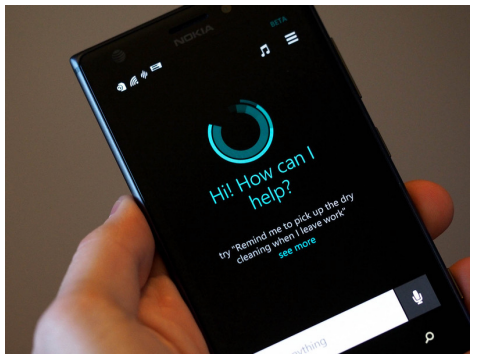
Natural Language Processing (NLP)



- Subfield of AI
- GOAL:
  - Bridge the gap between how people **communicate** and what machines **understand** in order to perform useful tasks, e.g.
    - Making appointments, buying things, question answering, etc.

NLP (Interpretation)

I want to eat Japanese food.



May I order Yayoi for you?

AI/ML (Generation)

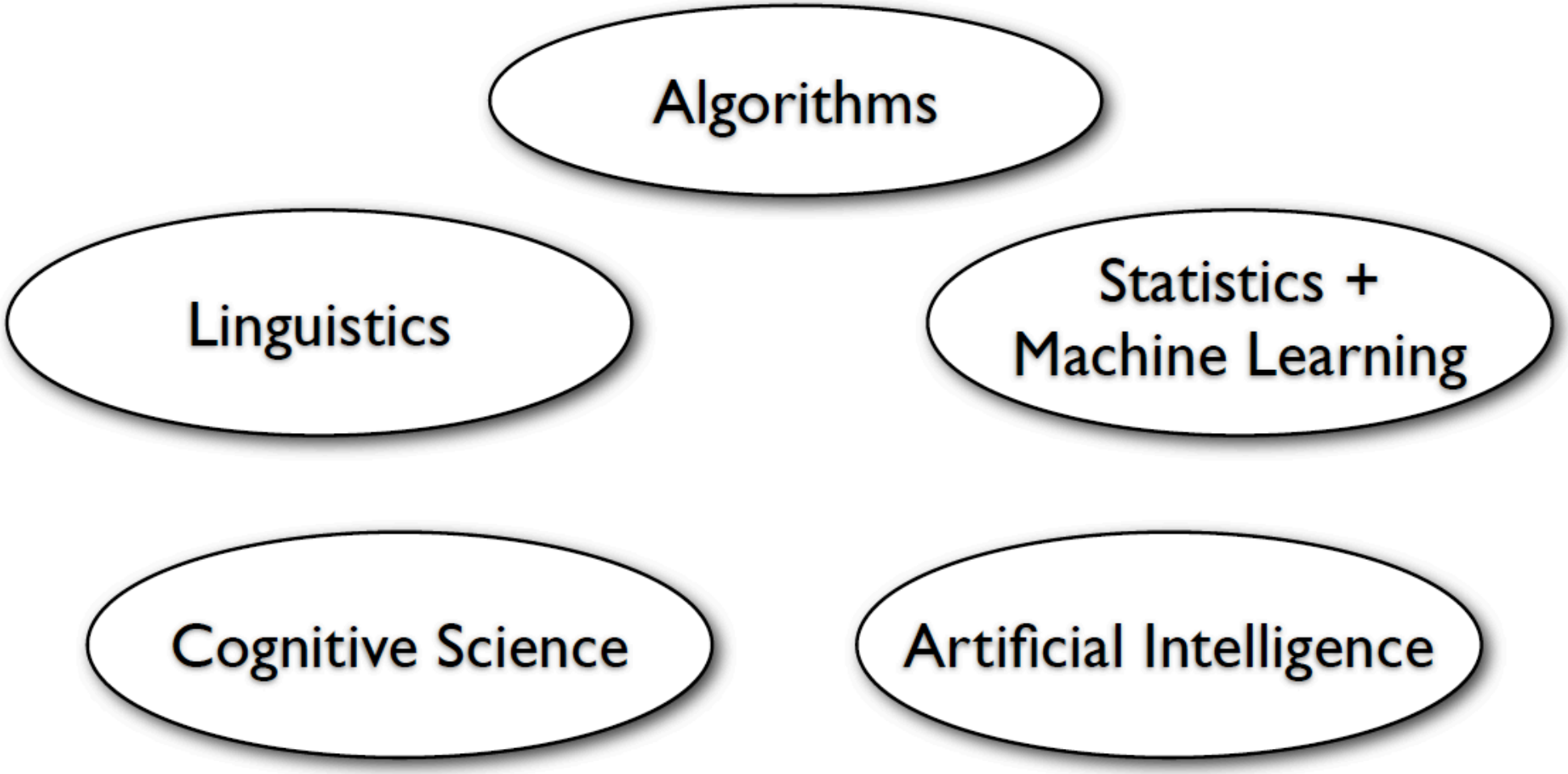
NLP Goal

Goal: intelligent processing of human languages

- Not just string matching

Ref: Prof. Regina Barzilay, NLP @MIT

NLP is interdisciplinary



Ref: Prof. Brendan O'Connor, CS 585 Intro to NLP, @UMass

Level of understanding in NLP

https://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_natural_language_processing.htm

Lexical Analysis:

Text→ Paragraphs, Sentences, and Words

Syntactic Analysis (Parsing):

Grammar/Relationship between words

Semantic Analysis:

Exact meaning of the sentence

Discourse Integration:

Meaning of the sentence **based on the previous sentence (pronouns)**

Pragmatic Analysis:

Actual Meaning based on **the context** and real-world knowledge

Discourse

Semantics

Syntax: Constituents

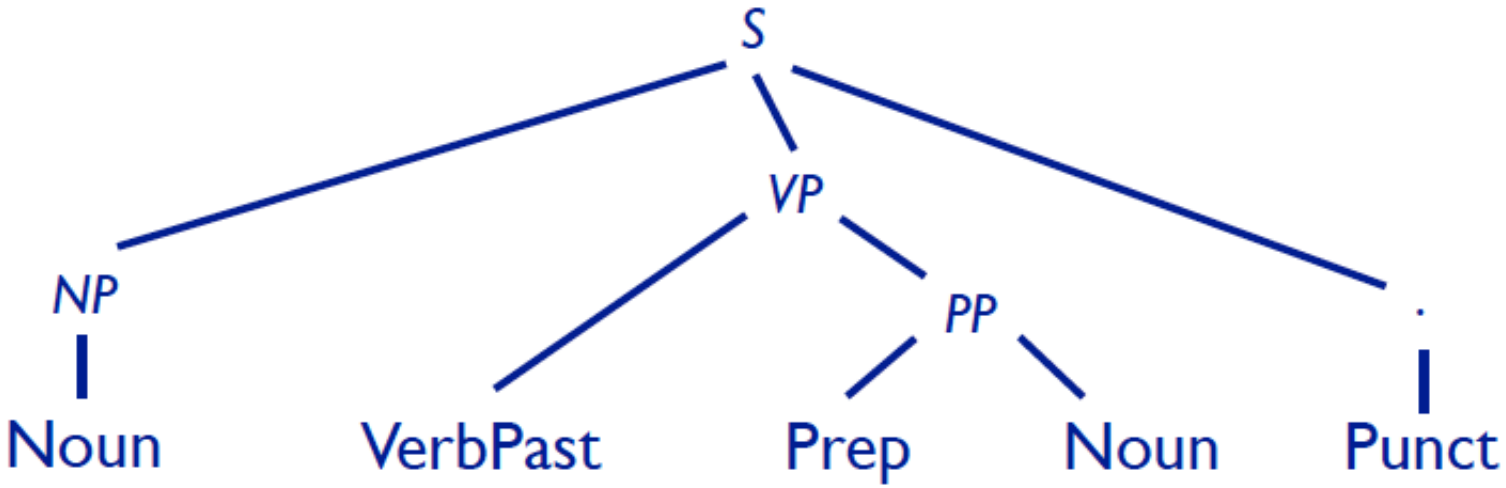
Syntax: Part of Speech

Words

Morphology

Characters

CommunicationEvent(e)
Agent(e, Alice)
Recipient(e, Bob)
SpeakerContext(s)
TemporalBefore(e, s)



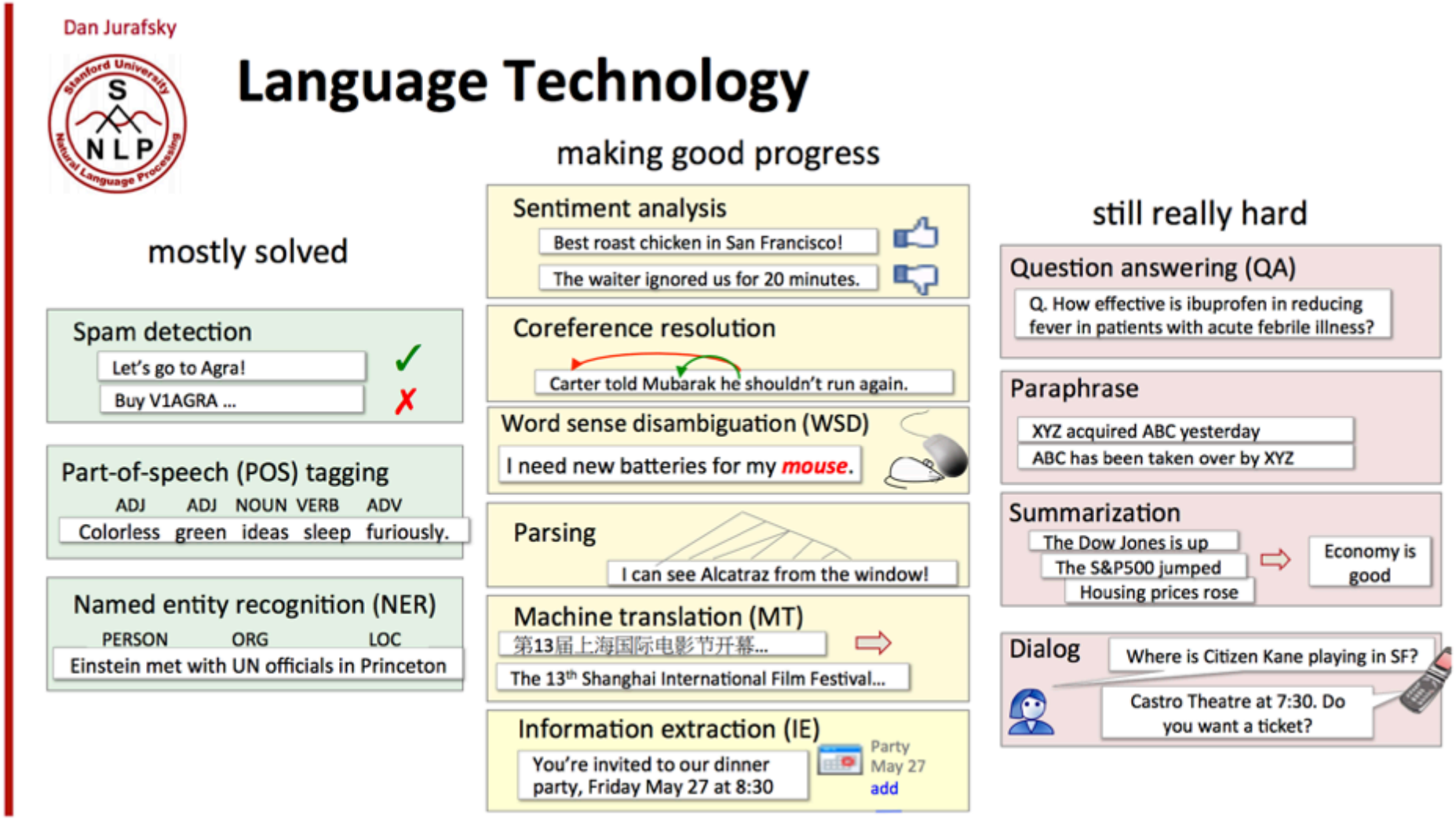
Alice talked to Bob .

talk -ed [VerbPast]

Alice talked to Bob.

Ref: Prof. Brendan O'Connor, CS 585 Intro to NLP, @UMass

NLP today: Technology



Ref: Prof. Regina Barzilay, NLP @MIT

NLP today: Machine Translation (MT)

Google

google translate

All Images Maps News Videos More Settings Tools

About 1,180,000,000 results (0.39 seconds)

English

As the new year gets underway, expert commentators give their view on what 2018 holds in store.

Here are three big themes to watch out for over the next 12 months.

Can the stock market rally go on? The new year has begun with stock markets in the UK and US hitting new record highs.

The Dow Jones Industrial Average rose above 25,000 points for the first time this week, while the broader S&P 500 is also at historic highs.

Thai

เป็นปีใหม่ที่กำลังได้รับการแสดงความคิดเห็นของผู้เชี่ยวชาญให้มุมมองของพวกเขาเกี่ยวกับสิ่งที่ 2018 เก็บในร้าน

ต่อไปนี้เป็นหัวข้อใหญ่สามข้อที่ควรระวังในช่วง 12 เดือนข้างหน้า

การชุมนุมตลาดหุ้นสามารถดำเนินต่อไปได้หรือไม่?
ปีใหม่เริ่มมีตลาดหุ้นในสหราชอาณาจักรและสหรัฐฯพุ่งสูงเป็นประวัติการณ์

ดัชนีเฉลี่ยอุตสาหกรรมดาวโจนส์ปรับตัวสูงขึ้นกว่า 25,000 จุดเป็นครั้งแรกในสัปดาห์นี้ขณะที่ดัชนี S & P 500 ที่ใหญ่ขึ้นก็อยู่ในระดับสูงเป็นประวัติการณ์

Markets, Brexit and Bitcoin: 2018's themes

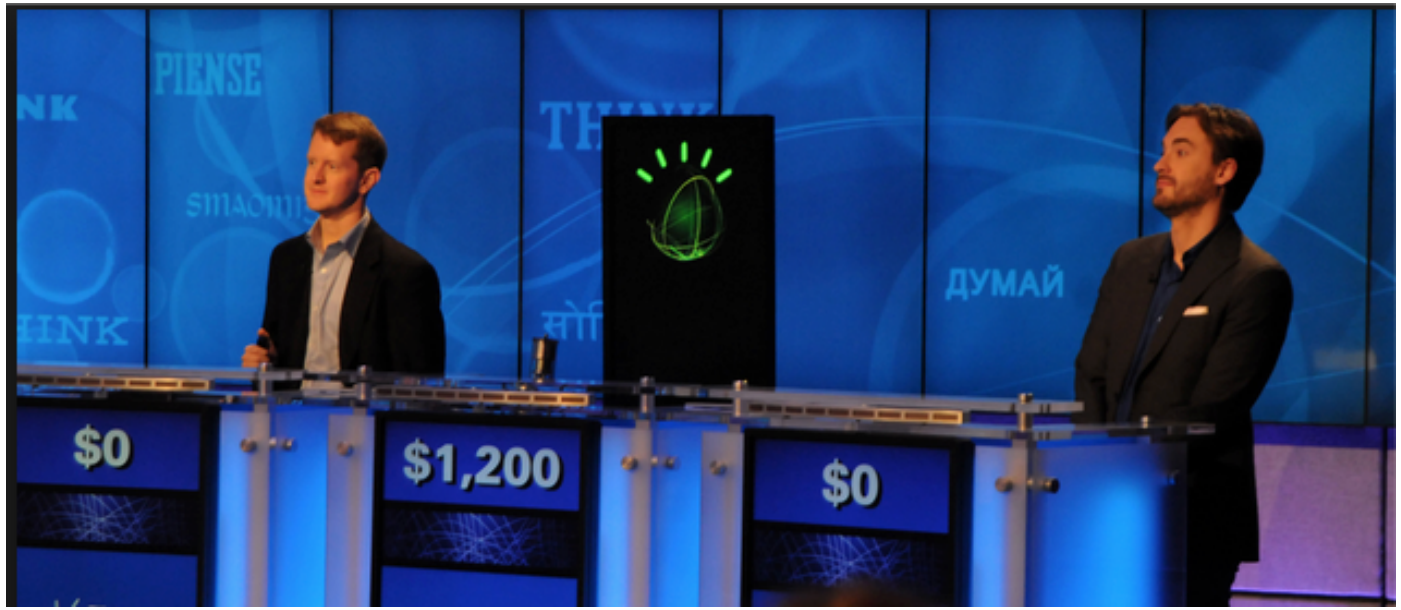
By Chris Johnston
Business reporter

5 January 2018 Share

As the new year gets underway, expert commentators give their view on what 2018 holds in store.

http://www.bbc.com/news/business-42581934

NLP today: Question Answering (QA)



AlphaGo
2015

IBM Watson wowed the tech industry and a corner of U.S. pop culture with its 2011 win against two of Jeopardy's greatest champions. Here's how IBM pulled it off and a look at what Watson's real career is going to be.

https://www.techrepublic.com/article/ibm-watson-the-inside-story-of-how-the-jeopardy-winning-supercomputer-was-born-and-what-it-wants-to-do-next/

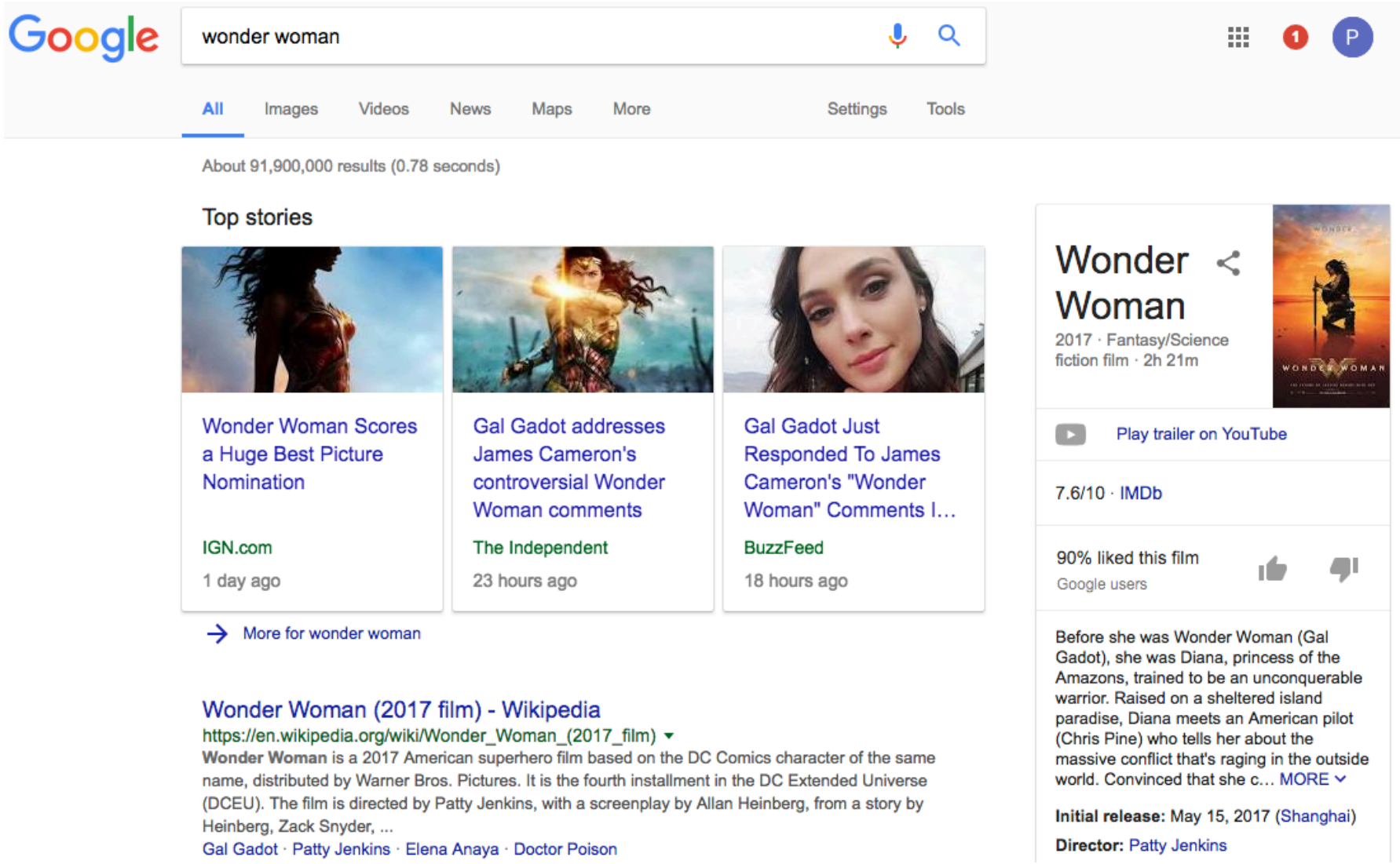
Ref: Prof. Regina Barzilay, NLP @MIT

NLP today: Question Answering (QA) (cont.)

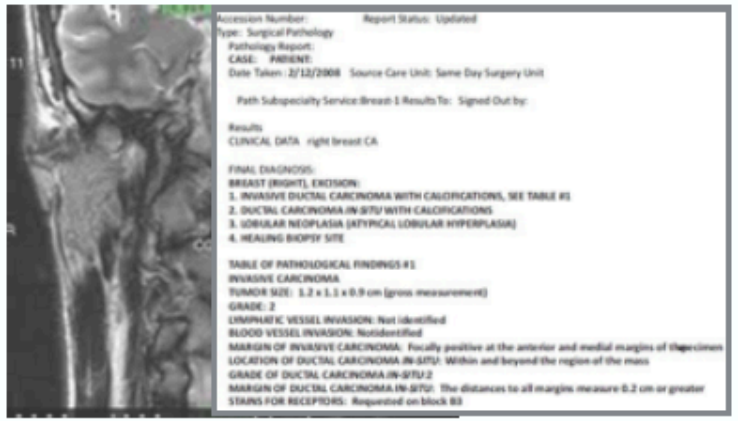
ปริศนาฟ้าแลบ | ตั๊ก, พายไก่, โรเบิร์ต, เป๊ล | 29 พ.ย. 60 Full HD

https://www.youtube.com/watch?v=CAJAQUao7HU

NLP today: Search/Summarization



NLP today: Information Extraction (IE)
Data science perspective on clinical research



Abstract clinical records into a database

ID	AGE	RACE	STUDY	PROC	BIRTHS	MA_AGE	ASSESS	DENSITY	FINDING	FINDING T
9527	78	2	6/12/06	BIDXU-L	0		P	3	CALCS	N
32875	56	1	7/11/06	BIDXB-B	0		N	3		
2247	72	1	4/12/06	BIDXU-R	0		N	3		
45521	61	1	3/30/06	BIDXB-B	0		B	3	CALCS	S
48987	41	1	4/5/06	BIDXB-B	0		P	3	CALCS	N
4179	67	1	5/12/06	BIDXB-B	0		P	2	CALCS	N
26300	59	1	3/31/06	BIDXU-L	0		N	3		
67960	64	1	4/7/06	BIDXU-R	0		P	3	MASS	O
43283	61	W	7/21/06	BIDXB-B	0		B	3		
43319	51	1	4/7/06	BIDXB-B	0		N	3		

Pathology Report: REMOVED_ACCESSION_ID
ACCESSIONED ON: REMOVED_DATE
CLINICAL DATA: Carcinoma **right breast**.
*** FINAL DIAGNOSIS ***
LYMPH NODE (SENTINEL), EXCISION
(REMOVED_CASE_ID): METASTATIC CARCINOMA IN 1 OF 1 LYMPH NODE.
NOTE: The metastatic deposit spans 0.19cm and is identified on H&E and cytokeratin immunostains. A second cytokeratin-positive but cauterized focus likely also represents metastatic tumor (<0.1cm). There is **no evidence of extranodal extension**.
BREAST (RIGHT), EXCISIONAL BIOPSY
(REMOVED_ACCESSION_ID : REMOVED_CASE_ID -B): **INVASIVE DUCTAL CARCINOMA** (SEE TABLE #1). **DUCTAL CARCINOMA IN-SITU**, GRADE 1. **ATYPICAL DUCTAL HYPERPLASIA**. **LOBULAR NEOPLASIA (ATYPICAL LOBULAR HYPERPLASIA)**.
TABLE OF PATHOLOGICAL FINDINGS #1

Name	Extraction
Breast Side	Right
Ductal Carcinoma in Situ	Present
Invasive Lobular Carcinoma	Absent
Invasive Ductal Carcinoma	Present
Cancer	Present
Lobular Carcinoma in Situ	Absent
Atypical Ductal Hyperplasia	Present
Atypical Lobular Hyperplasia	Present
Lobular Neoplasia	Present
Flat Epithelial Atypia	Absent
Blunt Adenosis	Absent
Atypia	Present
Positive Lymph Nodes	Present
Extracapsular Axillary Nodal Extension	Absent
Isolated Cancer Cells in Lymph Nodes	Absent
Lymphovascular Invasion	Absent
Blood Vessel Invasion	Absent
Estrogen Receptor Status	Positive
Progesterone Receptor Status	Positive
HER 2 (FISH) Status	Unknown

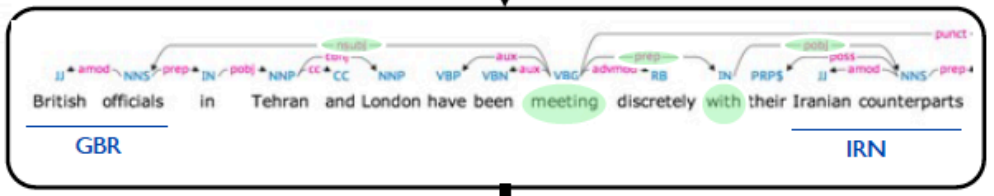
Parsing pathology reports into database

Ref: Prof. Regina Barzilay, NLP @MIT

NLP today: Trend analysis

Data: news articles

Dependency parsing to identify events

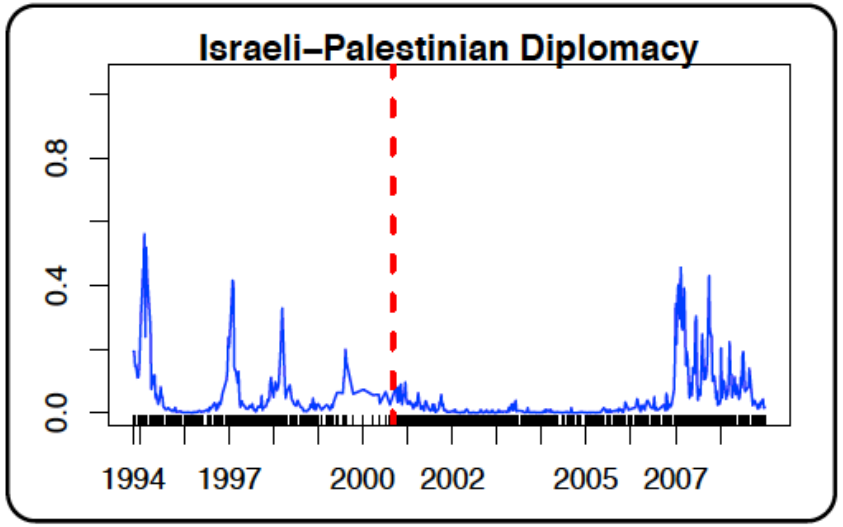


Machine learning from text:

(1) **Event class dictionaries**

"diplomacy"	arrive in, visit, meet with, travel to, leave, hold with, meet, meet in, fly to, be in, arrive for talk with, say in, arrive with, head to, hold in, due in, leave for, make to, arrive to,
"verbal conflict"	accuse, blame, say, break with, sever with, blame on, warn, call, attack, rule with, charge, say←ccomp come from, say ←ccomp, suspect, slam, accuse government ←poss,
"material conflict"	kill in, have troops in, die in, be in, wound in, have soldier in, hold in, kill in attack in, remain in, detain in, have in, capture in, stay in, about ←pobj troops in, kill, have troops

(2) **Political dynamics**



Ref: Prof. Brendan O'Connor, CS 585 Intro to NLP, @UMass

Hathaway Phenomenon

BERKSHIRE HATHAWAY INC.
3555 Farnam Street
Omaha, NE 68131
Official Home Page

- A Message From Warren E. Buffett
- Annual & Interim Reports
  Updated November 3, 2017
- Special Letters From Warren & Charlie RE:Past, Present and Future
- Link to SEC Filings
- Links to Berkshire Subsidiary Companies
- Corporate Governance
- Owner's Manual
- Letters from Warren E. Buffett Regarding Pledges to Make Gifts of Berkshire Stock
- News Releases
  Updated November 3, 2017
- Warren Buffett's Letters to Berkshire Shareholders
  Updated February 25, 2017
- Charlie Munger's Letters to Wesco Shareholders
- Annual Meeting Information
- Celebrating 50 Years of a Profitable Partnership
  (A commemorative book first sold at the 2015 Annual Meeting and now for sale on eBay.)
- Comparative Rights and Relative Prices of Class A and B
- Berkshire Activewear

A couple weeks ago, Huffington Post blogger Dan Mirvish noted a funny trend: when Anne Hathaway was in the news, Warren Buffett's Berkshire Hathaway's shares went up. He pointed to six dates going back to 2008 to show the correlation. Mirvish then suggested a mechanism to explain the trend: "automated, robotic trading programming are picking up the same chatter on the Internet about 'Hathaway' as the IMDb's StarMeter, and they're applying it to the stock market."

NLP is hard!

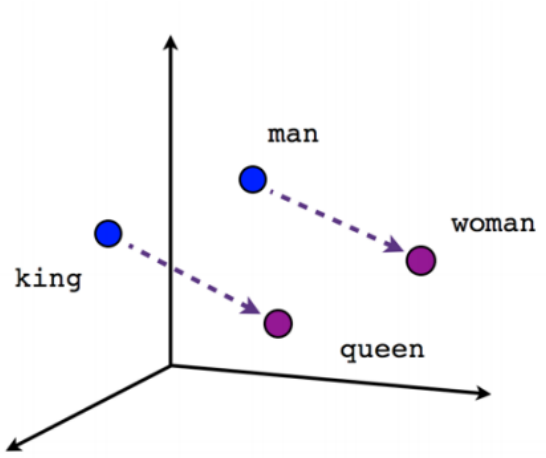
Word-level ambiguity

Ref: Prof. Regina Barzilay, NLP @MIT

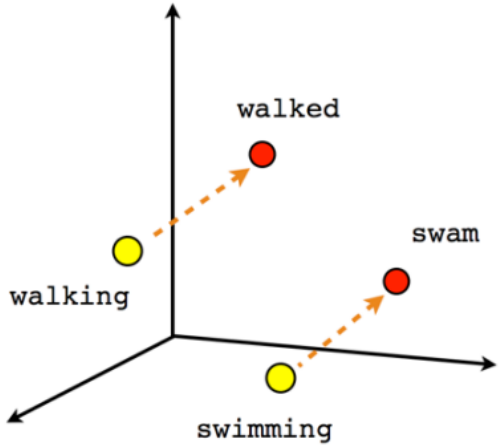
Why NLP is hard?

Why NLP is hard?

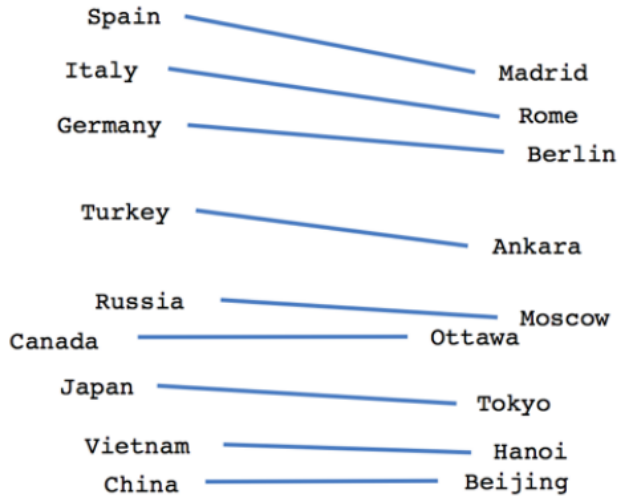
- Complexity in representing, learning and using linguistic/situational/world/visual knowledge



Male-Female



Verb tense



Country-Capital

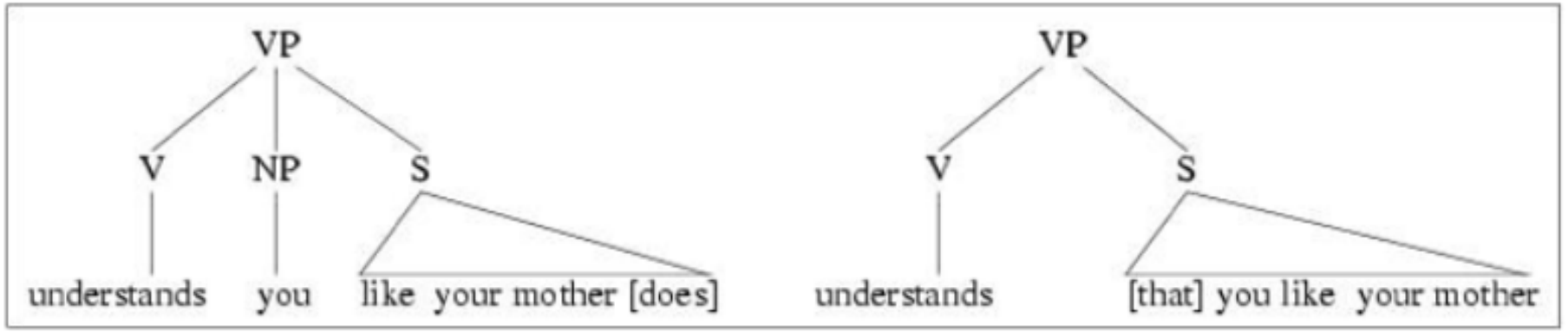
Ref: Prof. Christopher Manning, CS224N/Ling284, 2017

Why NLP is hard? (cont.)

- Human languages are ambiguous (unlike programming and other formal languages).
  - Some parts can be ignored
- Human languages are interpretation depends on real world, common sense, and contextual knowledge (pragmatic analysis)

At last, a computer understands you like your mother"

Ambiguity at syntactic level: Different structures lead to different interpretations



The Pope's baby steps on gays. [Ref: Prof. Christopher Manning, CS224N/Ling284, 2017]

Ref: Prof. Christopher Manning, CS224N/Ling284, 2017 Ref: Prof. Regina Barzilay, NLP @MIT

Issues in Thai NLP

- Word segmentation
  - No word delimiters
  - ฉัน|นำ|ดอก|ไม้|ไป|ไหว้|ศาล|พระ|ภูมิ|ที่|โรง|เรียน|ประจำ|
  - ฉัน|นำ|ดอกไม้|ไป|ไหว้|ศาลพระภูมิ|ที่|โรงเรียน|ประจำ|
  - ฉัน|นำ|ดอก|ไม้|ไป|ไหว้|ศาล|พระภูมิ|ที่|โรง|เรียน|ประจำ|
  - ฉัน|นำ|ดอกไม้|ไป|ไหว้|ศาลพระภูมิ|ที่|โรงเรียนประจำ|

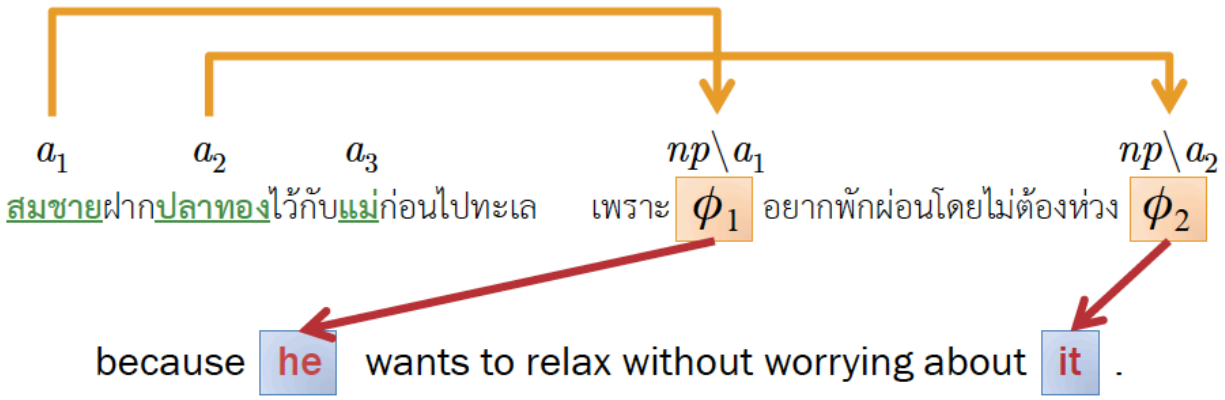
- Sentence segmentation
  - No sentence boundary markers

อย่างไรก็ตาม อดีตประธาน ทปอ. กล่าวว่า มีการทักท้วงเรื่องนี้มาตลอดว่า มีช่วงเวลาว่างนานขนาดนี้ ทำไมถึงยังต้องมีการจัดสอบนอกเหนือจากนี้อีก เพราะการสอบล่วงล้ำไปในเวลาระหว่างเรียนมัธยมนั้นกระทบกับเรื่องอื่นๆ โดยเฉพาะการเรียนในชั้นเป็นวงจรลูกโซ่ แนวโน้มที่เข้ามาแก้เรื่องนี้ เป็นความคิดที่ดี แต่ยังไม่เห็นเรื่องใช้ผลการเรียนในชั้นมาเป็นองค์ประกอบรับตรง ซึ่งอาจทำให้เด็กไม่สนใจห้องเรียน และมุ่งกวดวิชา ทำให้การสอบเข้าอุดมศึกษาตกเป็นจำเลยข้อหาทำลายระบบการศึกษาขั้นพื้นฐาน วนไปสู่ปัญหาเก่าๆ ได้

Ref: Introduction to Thai NLP (Prachya Boonkwan), 2017

Issues in Thai NLP (cont.)

- Syntax ambiguity
  - Pronouns and some constituents can be omitted as long as they can be implied from the context



- Nostalgic Thai slangs

Ref: Introduction to Thai NLP (Prachya Boonkwan), 2017

Ref: Introduction to Thai NLP (Prachya Boonkwan), 2017

+

NLP & Text Mining

NLP & Text Mining

- NLP: Language → Meaning
- Text Mining
  - Text mining, which is sometimes referred to "text analytics" is one way to make qualitative or "unstructured data" usable by a computer.
  - Convert from unstructured to structured data
  - **NLP** techniques are the building blocks for text mining tasks

Comment	Good	Like	Hate	#
Tweet1	7	8	0	😄
Tweet2	1	0	10	😡
Tweet3	2	9	1	😄

NLP pipeline example (English)

Tokenization

- Input: Mr.Smith goes to Washington
- Output: [Mr.Smith, goes, to, Washington]

Part of Speech tagging

- Input: [Mr.Smith,goes,to,Washington]
- Output:[(Mr.Smith,**NNP**), (goes,**VBZ**), (to,**TO**), (Washington,**NNP**)]

NER

- Input:[(Mr.Smith,**NNP**), (goes,**VBZ**), (to,**TO**), (Washington,**NNP**)]
- Output:[(Mr.Smith,NNP,**PER**), (goes,VBZ,**O**), (to,TO,**O**), (Washington,NNP,**LOC**)]

Application

- e.g. Word Cloud (Named Entity Only)

PENN Part Of Speech Tags

- NNP – proper noun
- VBZ - Verb, 3rd person singular present
- TO –to

Ref: https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html

Named Entity Tags

- PER –Person
- LOC – Location
- ORG – Organization
- O – Other

NLP pipeline example (Thai)

Tokenization

- Input:ขสมก.เล็งจัดหารถ
- Output: ขสมก.,เล็ง,จัดหา,รถ

Part of Speech tagging

- Input:[ขสมก.,เล็ง,จัดหา,รถ]
- Output:[(ขสมก.,NR),(เล็ง,VV),(จัดหา,VV),(รถ,NN)]

PENN Part Of Speech Tags
- NR – proper noun
- VV - Main verbs in clauses, verb-form
- NN – Non-proper noun

Ref: BEST2010 dataset

NER

- Input:[(ขสมก.,NR),(เล็ง,VV),(จัดหา,VV),(รถ,NN)]
- Output:[(ขสมก.,NR,ORG),(เล็ง,VV,O),(จัดหา,VV,O),(รถ,NN,O)]

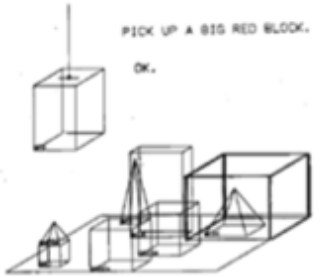
Named Entity Tags
- PER –Person
- LOC – Location
- ORG – Organization
- O – Other

Application

- e.g. Word Cloud (Named Entity Only)

History of NLP techniques

- 1) Symbolic approach:
  - encode all the required information into computer
  - In 1960s and 1970s, Noam Chomsky (eminent linguist) believed that that statistical techniques would never be sufficient to gain a deep understanding on human language.
  - This led to the dominance of **knowledge-based approaches**, requiring humans experts to encode knowledge into computers.
  - Disadvantage: It is required substantial human effort.



The dialog that was used as a SHRDLU demo:

```
Person: PICK UP A BIG RED BLOCK.
Computer: OK. (does it)
Person: GRASP THE PYRAMID.
Computer: I DON'T UNDERSTAND WHICH PYRAMID YOU MEAN.
Person: FIND A BLOCK WHICH IS TALLER THAN THE ONE YOU ARE HOLDING AND PUT IT INTO THE BOX.
Computer: BY "IT", I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM HOLDING.
Computer: OK. (does it)
Person: WHAT DOES THE BOX CONTAIN?
Computer: THE BLUE PYRAMID AND THE BLUE BLOCK.
Person: WHAT IS THE PYRAMID SUPPORTED BY?
Computer: THE BOX.
Person: HOW MANY BLOCKS ARE NOT IN THE BOX?
```

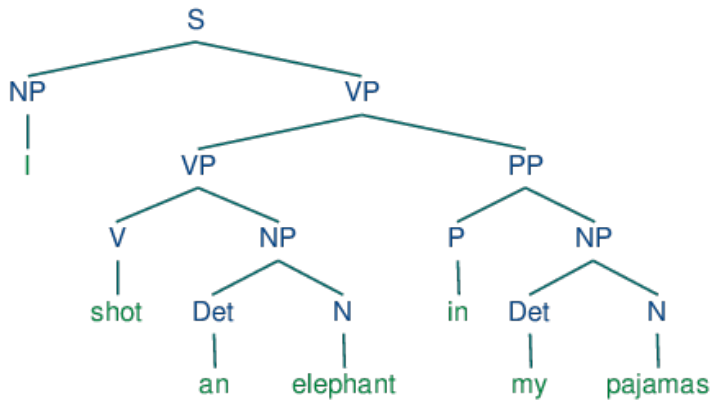
```
(DEFTHEOREM TC-GRASP
        (THCONSE (X Y)   (#GRASP $?X)
                 (THGOAL(#MANIP $?X))
                 (THCOND ((THGOAL (#GRASPING $?X)))
                         ((THGOAL (#GRASPING $_Y))
                          (THGOAL (#GET-RID-OF $?Y)
                                  (THUSE TC-GET-RID-OF))))
                         (T))
                 (THGOAL (#CLEARTOP $?X) (THUSE TC-CLEARTOP))
                 (THSETQ $_Y (TOPCENTER $?X))
                 (THGOAL (#MOVEHAND $?Y)
                         (THUSE TC-MOVEHAND))
                 (THASSERT (#GRASPING $?X))))

(DEFTHEOREM TC-PUT
        (THCONSE (X Y Z)  (#PUT $?X $?Y)
                 (CLEAR $?Y (SIZE $?X) $?X)
                 (SUPPORT $?Y (SIZE $?X) $?X)
                 (THGOAL (#GRASP $?X) (THUSE TC-GRASP))
                 (THSETQ $_Z (TCENT $?Y (SIZE $?X)))
                 (THGOAL (#MOVEHAND $?Z)  (THUSE TC-MOVEHAND))
                 (THGOAL (#UNGRASP)  (THUSE TC-UNGRASP))))
```

The Internals of SHRDLU

Ref: Prof. Regina Barzilay, NLP @MIT

History of NLP techniques (cont.)

- 2) Statistical approach:
  - infer language properties from language samples
  - In 1980s, an empirical revolution took place. Inspired by information theory, it began using **probabilistic approaches** in NLP.
  - Disadvantage: It is required hand-crafted features.
- 2.5) Deep Learning approach:
  - It is a **feature-engineering embedded** neural approach.
  - Since 2010s, it has been gaining a lot of attentions and showing many successes.



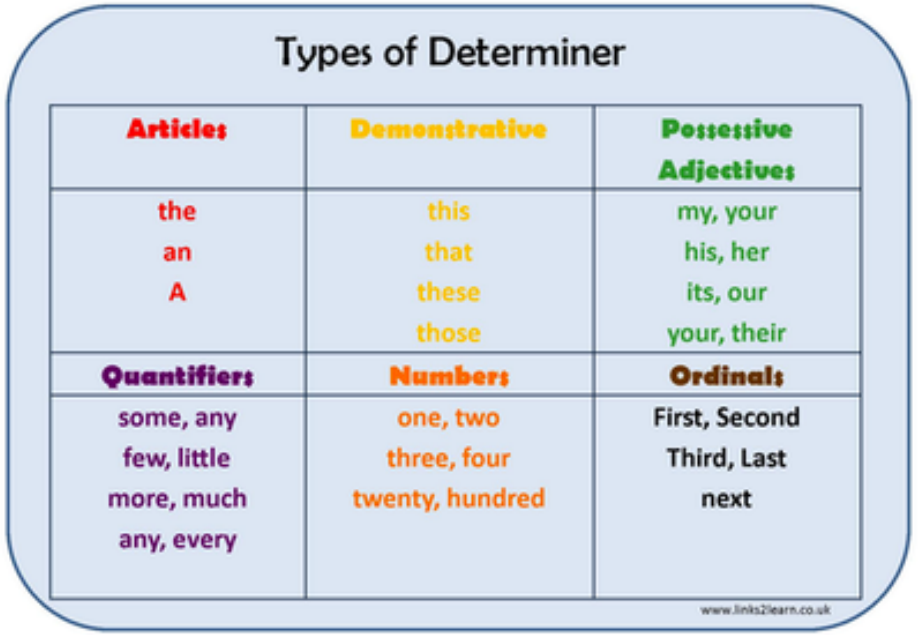
PennTree Bank (1993): one million words from WSJ, manually annotated with syntactic structure

Ref: Prof. Regina Barzilay, NLP @MIT

Case Study: Determiner placement Symbolic vs. statistical approaches

- Goal: It aims to place "the" (determiner).

Scientists in United States have found way of turning lazy monkeys intoworkaholics using gene therapy. Usually monkeys work hard only whenthey know reward is coming, but animals given this treatment did theirbest all time. Researchers at National Institute of Mental Health nearWashington DC, led by Dr Barry Richmond, have now developed genetictreatment which changes their work ethic markedly. "Monkeys underinfluence of treatment don't procrastinate," Dr Richmond says.Treatment consists of anti-sense DNA - mirror image of piece of one ofour genes - and basically prevents that gene from working. But for rest of us, day when such treatments fall into hands of our bosses may be one we would prefer to put off.



Ref: Prof. Regina Barzilay, NLP @MIT

Case Study: Determiner placement (cont.) Symbolic vs. statistical approaches

Symbolic approach

- Determiner placement is largely determined by:
  - Type of noun (countable, uncountable)
  - Uniqueness of reference
  - Information value (given, new)
  - Number (singular, plural)
- However, many exceptions and special cases play a role:
  - The definite article is used with newspaper titles (The Times), but zero article in names of magazines and journals (Time)
- Hard to manually encode this information!

Statistical approach

- Consider it as classification
- Predictions: {-1, +1}
- Features:
  - plural?
  - first appearance in text?
  - head token
  - …

Minnen et al.	83.58%
Turner&Charniak	86.74%
Knight&Chander	78%

"lazy monkeys"

$$\Downarrow$$

$$[1\ 1\ 0\ 0\ 0\ \ldots\ 1]^T$$

-1

"the United States"

$$\Downarrow$$

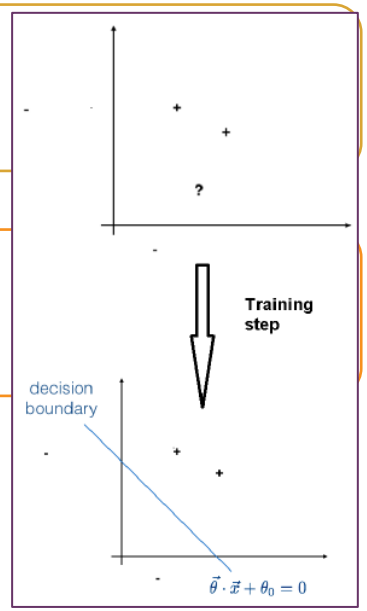
$$[1\ 1\ 0\ 0\ 0\ \ldots\ 0]^T$$

+1

Ref: Prof. Regina Barzilay, NLP @MIT

Limitation of traditional statistical approach

1. • Data collection
   • Obtain the labels {-1, +1}
2. • Feature Extraction
   • Most common features in NLP is word n-grams → encoded as one-hot vector
3. • Training
4. • Testing



- Sparsity:
  - feature vectors are typically high-dimensional and sparse (i.e. most elements are 0).
- Feature engineering:
  - Need experts to manually design features

Map discrete, one-hot vectors into low-dimensional continuous representations.
Self learned features → Deep Learning

pear

$[1\ 0\ 0\ 0\ \ldots\ 0]$

↓

$[0.4\ 0.1\ 0.1]$

apple

$[0\ 0\ 1\ 0\ \ldots\ 0]$

↓

$[0.6\ 0.2\ 0.3]$

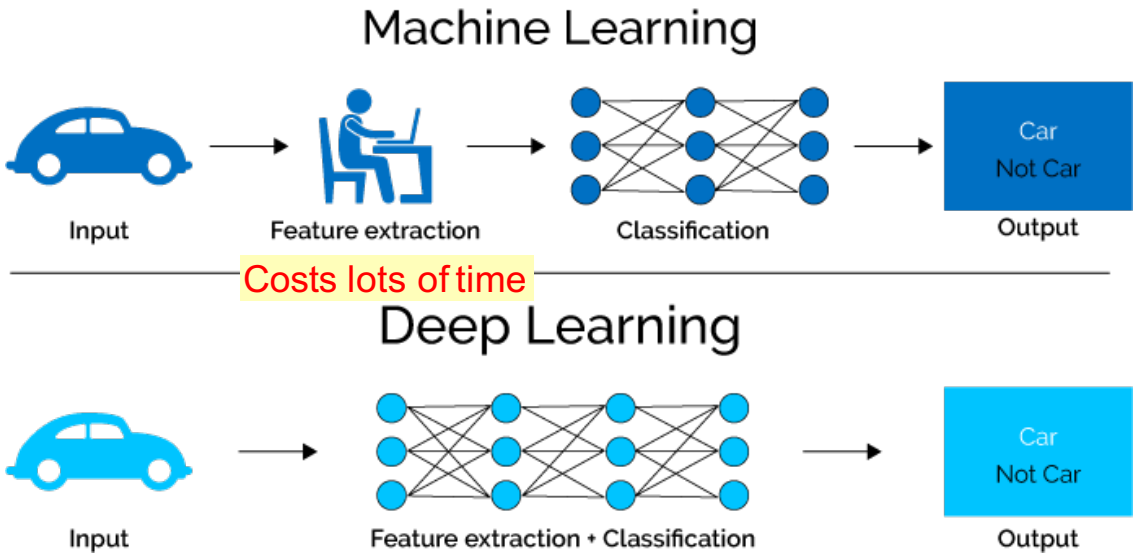
Ref: Prof. Regina Barzilay, NLP @MIT

Deep Learning

What is Deep Learning (DL)?

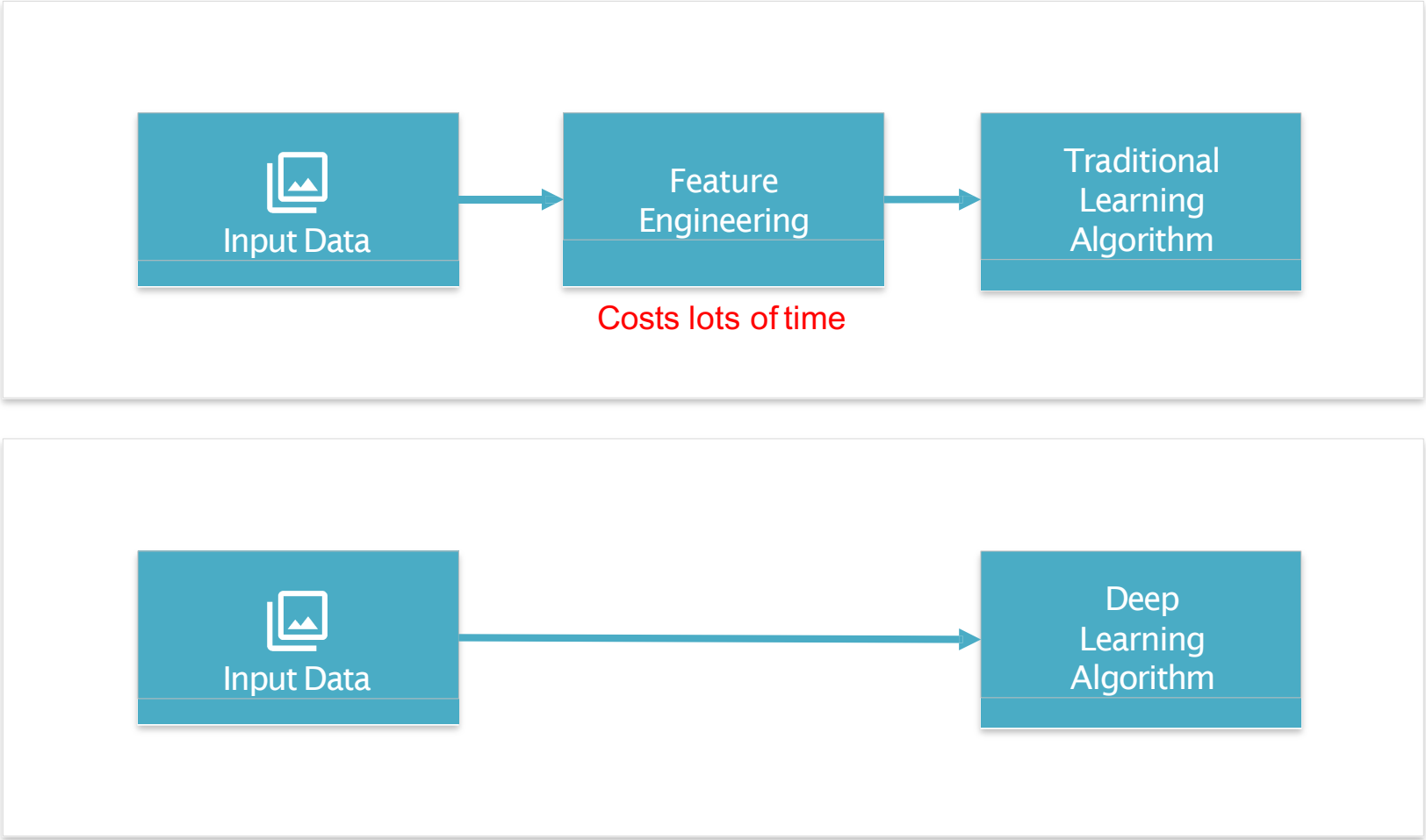
Part of the machine learning field of learning representations of data. Exceptional effective at learning patterns.

Utilizes learning algorithms that derive meaning out of data by using a hierarchy of multiple layers that mimic the neural networks of our brain.



Deep Learning - Basics

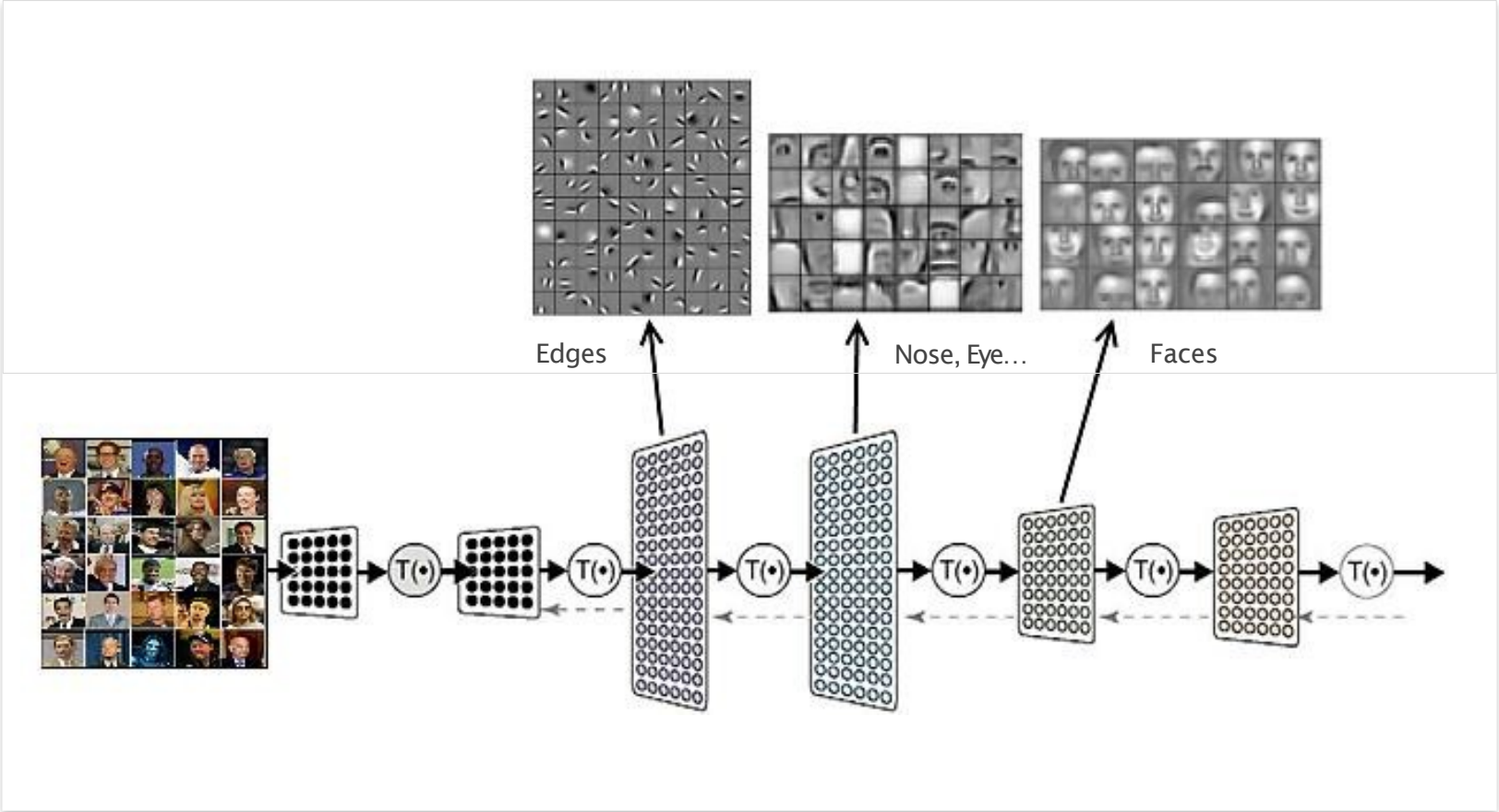
No more feature engineering



Deep Learning – Basics (cont.)

What did it learn?

A deep neural network consists of a hierarchy of layers, whereby each layer transforms the input data into more abstract representations (e.g. edge -> nose -> face). The output layer combines those features to make predictions.

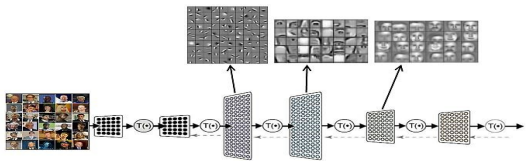


Deep Learning Application

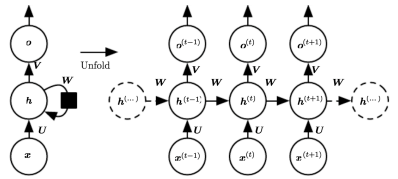
Speech Recognition

Computer Vision

Natural Language Processing



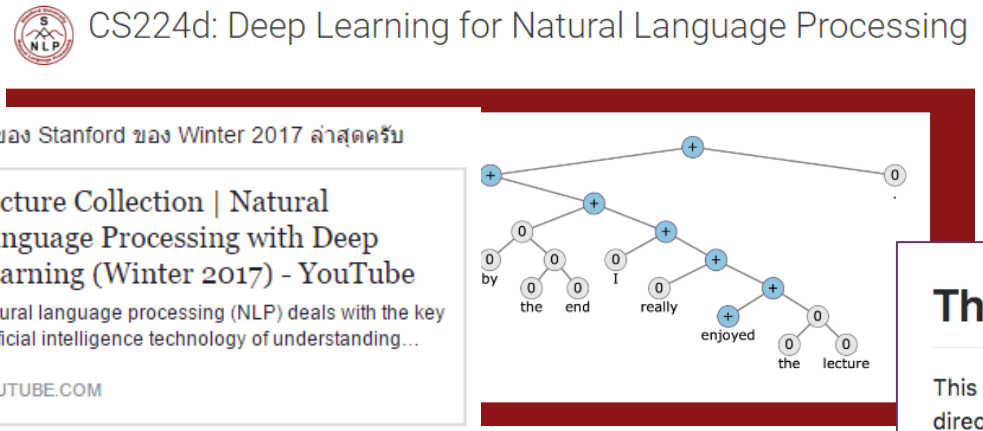
CNN



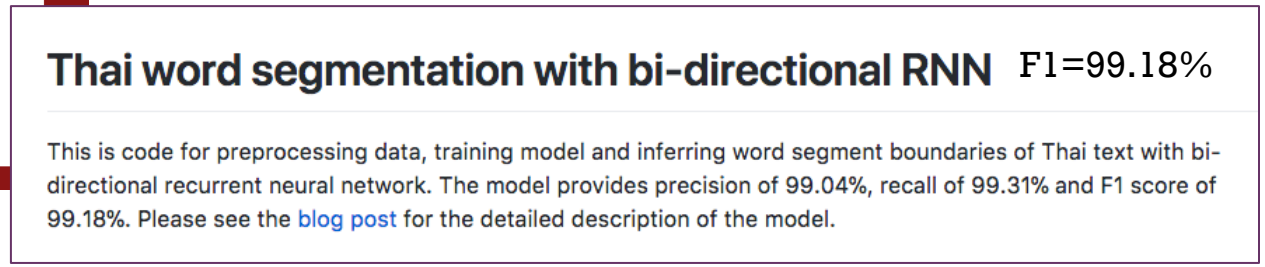
RNN (LSTM)

NLP + Deep Learning = Deep NLP

- Modern NLP techniques are based on deep learning models.
- These models have obtained very high performance across various NLP tasks.
- They often do not require traditional linguistic feature engineering to perform well.



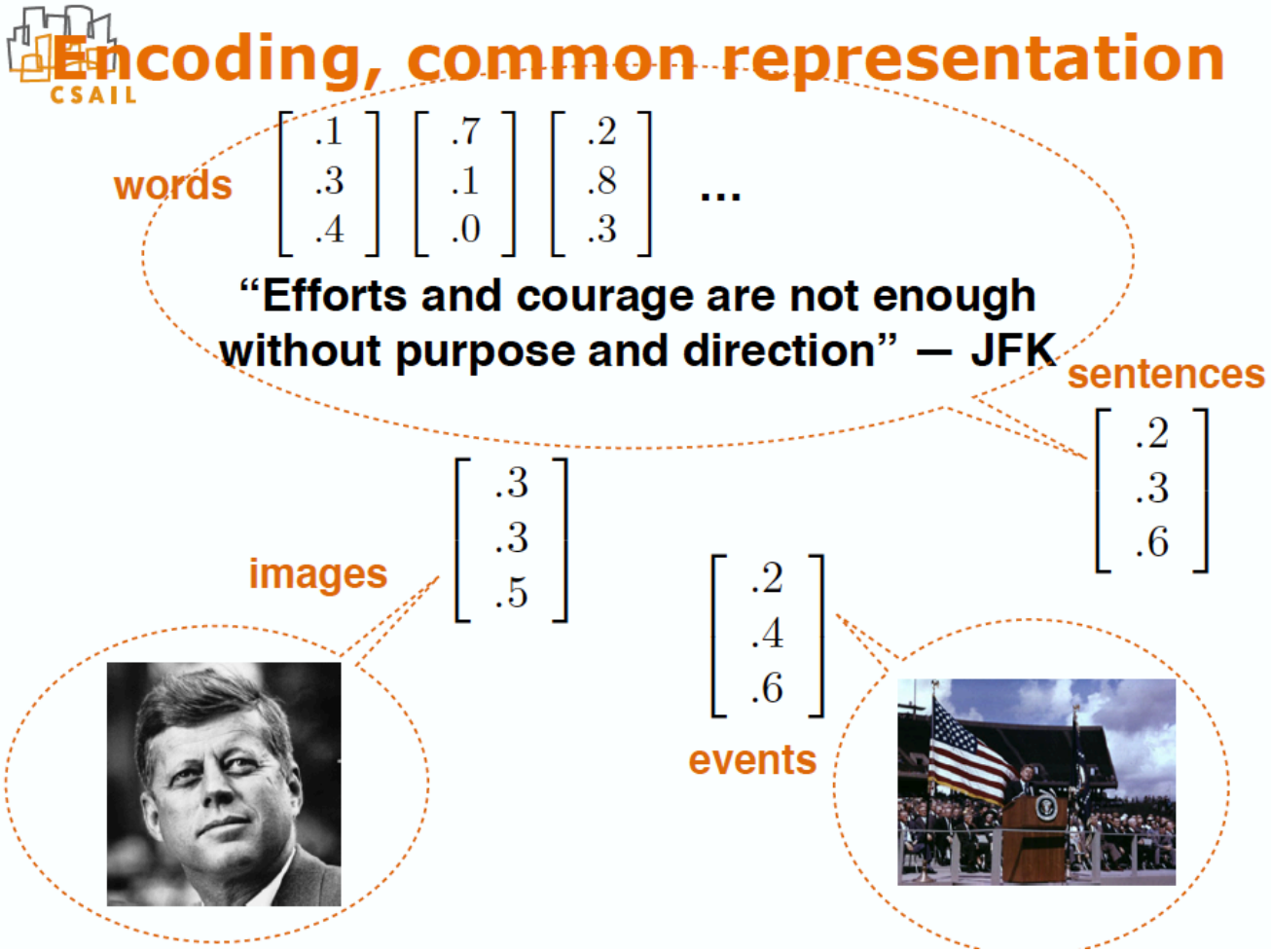
Pucktada Treeratpituk
RNN, F1 = 0.93 on BEST2010



Chula Data Science

Reasons for exploring Deep Learning

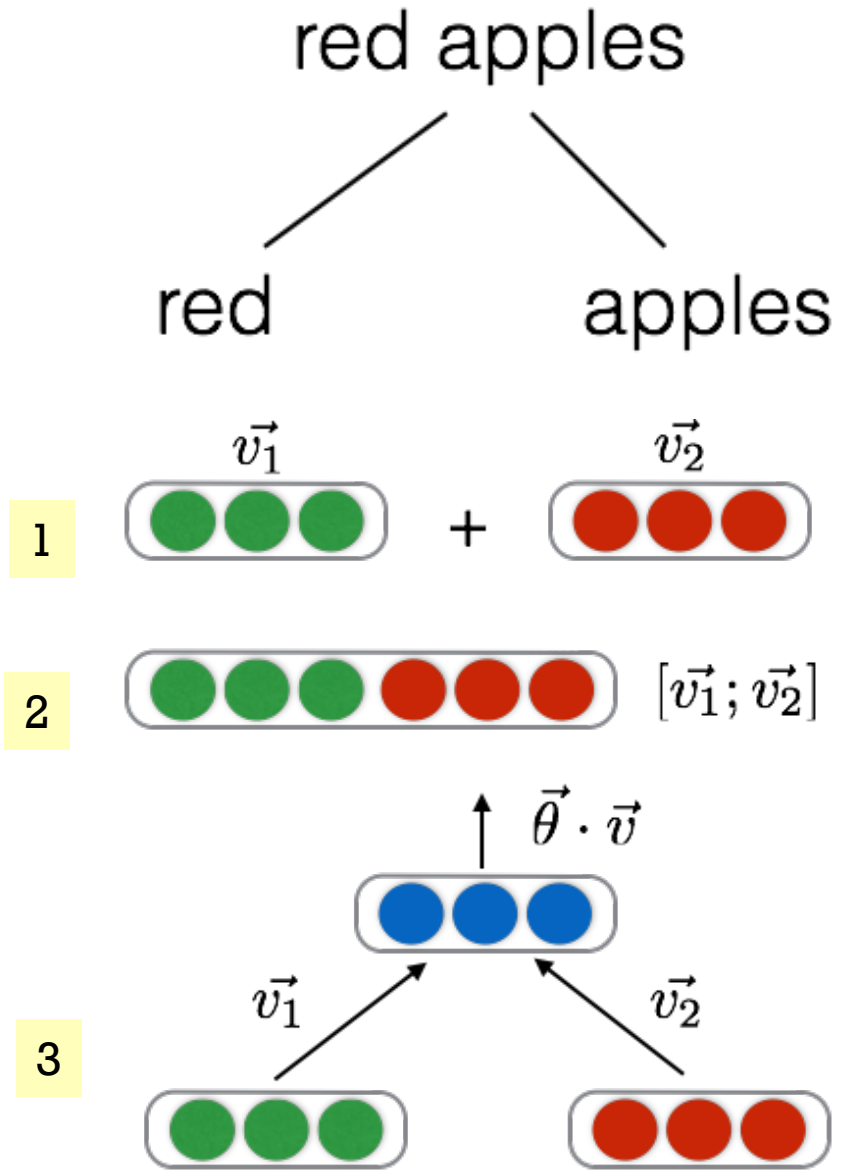
- Learned features are easy to adapt, fast to learn
- Deep learning provides a very flexible? Universal, learnable framework for representing world, visual, and linguistic information



Ref: Prof. Regina Barzilay, NLP @MIT

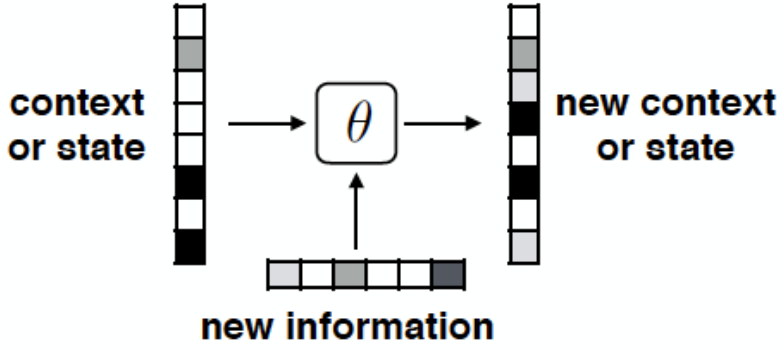
Reasons for exploring Deep Learning (cont.)

- Flexible neural "Lego pieces"
  - Common representation, diversity of architectural choices
- Can represent any levels of NLP
  - Word
  - Phrase
  - Sentence
  - Paragraph (document)

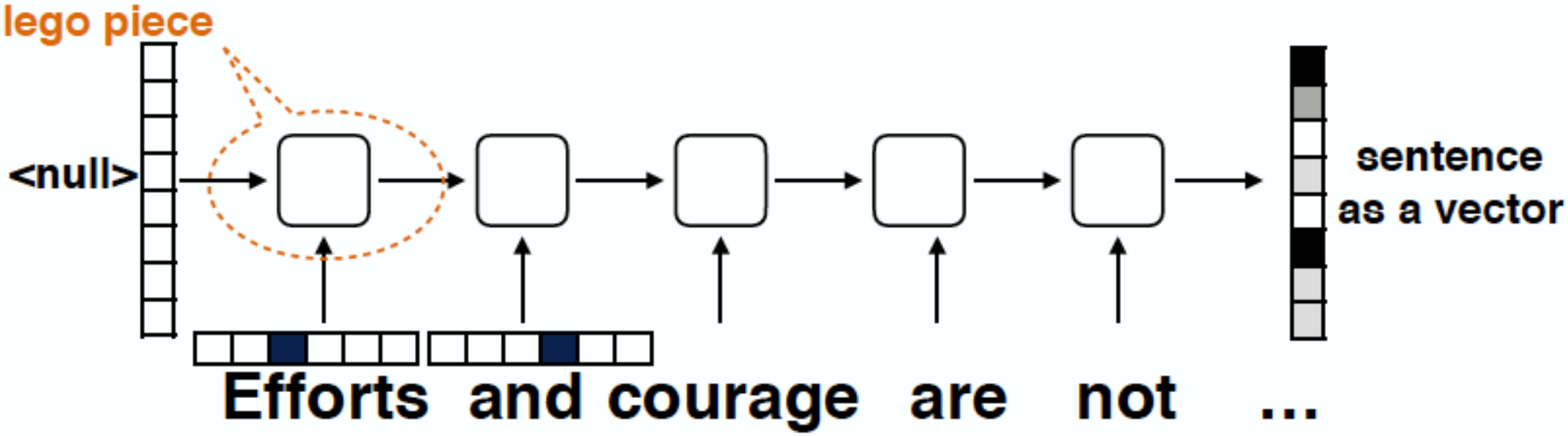


Ref: Prof. Regina Barzilay, NLP @MIT

Reasons for exploring Deep Learning (cont.)
Example of encoding sentences



RNN
(LSTM, GRU)



+

NLP Tools

- http://nltk.org
- Tokenization (Parser)
- Stop words removal
- Stemming
- N-gram
- Part-of-Speech Tagging
- Named Entity Recognition
- Etc.

Tools for Thai NLP

Open-source

- Word Segmentation
  - CutKum
  - LexTo (NECTEC)
  - SWATH (CMU) – 80's
  - ICU (IBM) – 90's
- POS tagging
  - SWATH (CMU)

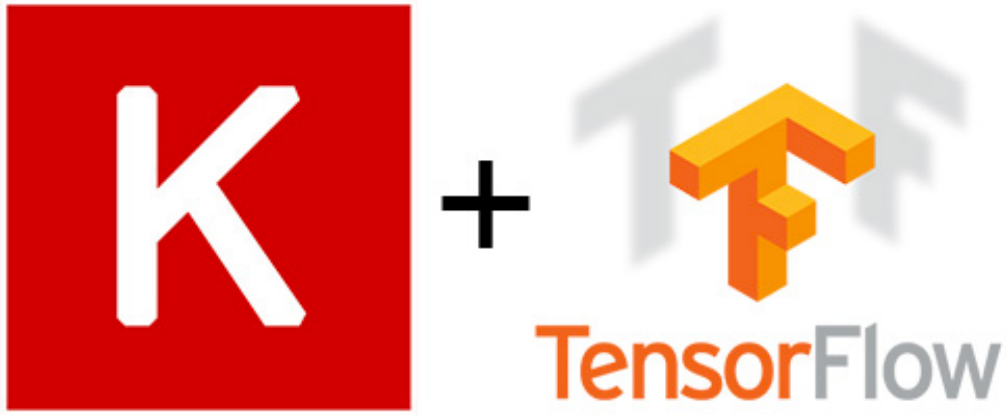
Commercial

- Word Segmentation
  - SegIt (NECTEC)
- POS tagging
  - PosIt (NECTEC)

Ref: Introduction to Thai NLP (Prachya Boonkwan), 2017

Deep Learning tools

- Tensorflow (Google): Python, etc.
- Torch (The Idiap Research Institute):
  - PyTorch: Python
- Theano (University of Montreal): Python
- Keras (François Chollet): Python
- Lasange (Sander Dieleman): Python
- MXNet (Microsoft)



Course Logistics

Class schedule

Word Model

Sequence

Tree

Application

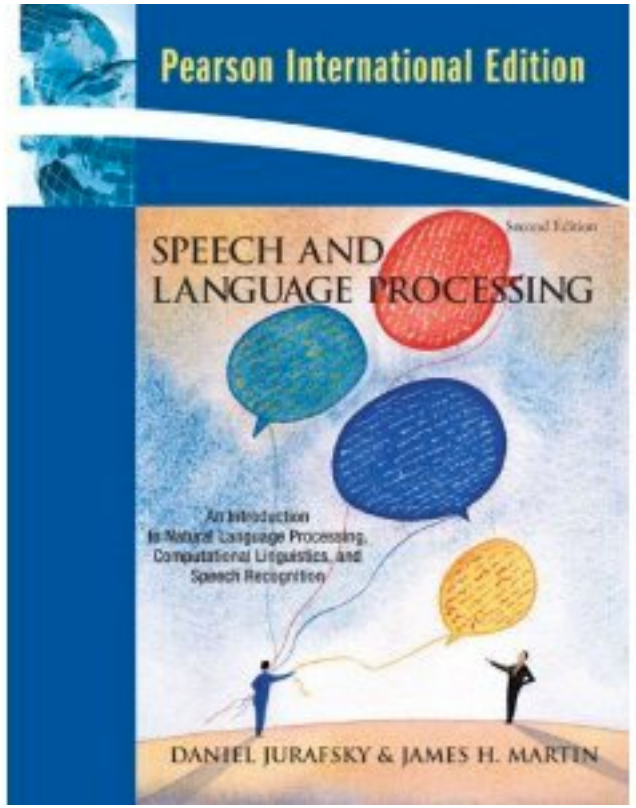
Implementation

Week	Topics (E=Ekapol, P=Peerapon)
1 - 8/1	Intro to NLP (P)
2 - 15/1	Tokenization LexTo, DNN, CNN, LSTM with Keras. (E)
3 - 29/1	Language modeling (P) N-grams, smoothing, Neural LM
4 - 5/2	Representation (P) TF-IDF, word and sentence embeddings, adaptation
5 - 12/2	PoS tagging and information extraction (E) CRF and beamsearch
6 - 19/2	Parsing (E) PCFG, Recursive neural networks
	Midterm week
7 - 26/2	In class midterm
8 - 12/3	Document/sentiment classification (E) LDA, Naive Bayes, EM
9 - 19/3	Question Answering (P&E) Reading comprehension, text generation
10 - 26/3	Chatbots (P)
11 - 2/4	Industry Talk (TBA)
12 - 9/4	Other NLP applications (E) ASR, MT, Summarization
13 - 23/4	Project presentation

Course Grading

- Assignments 30% (5% each with 5% extra)
- Midterm 35%
- Project 35%

Google Cloud Platform



Speech and Language Processing, 2nd Edition 2nd Edition

by Daniel Jurafsky (Author), James H. Martin (Author)

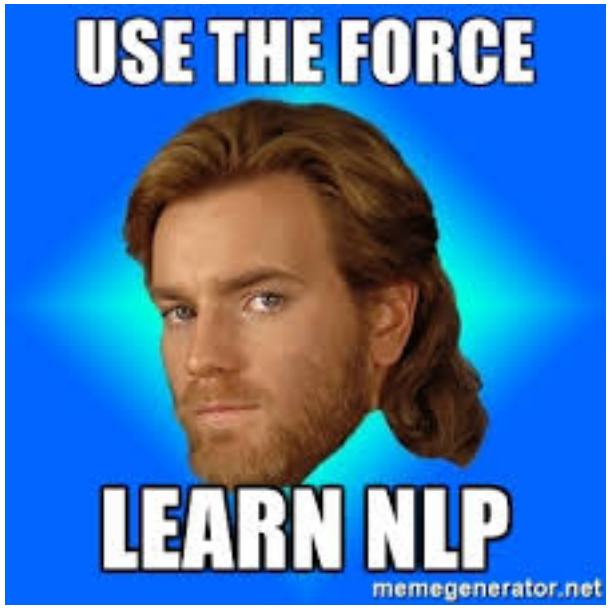
https://nlp.stanford.edu/~manning/xyzzy/JurafskyMartinEd2book.pdf

Google Cloud Demo

NLP

- Neuro-Linguistic Programming
  - Top hit if you google NLP
  - Pseudoscience

- Natural Language Processing



นวัตกรรมเปลี่ยนโลก

'เอไอ-ดิจิทัลทวิน-บล็อกเชน'

บุษกร ภู่แส
กรุงเทพธุรกิจ

สำนักงานคณะกรรมการนโยบายวิทยาศาสตร์ฯ เปิดรายงาน 6 อันดับนวัตกรรมเด่นปี 2561 เผยเทคโนโลยีปัญญาประดิษฐ์ ดิจิทัลทวินมาแรงต่อเนื่อง ตามด้วยบล็อกเชน จีโนมิกส์ โอมิกส์ เทคโนโลยีการพิมพ์สามมิติ และแซร์ริ่งอีโคโนมี ที่ต้องจับตามอง ขณะที่อุปกรณ์ความจำเพื่อเก็บข้อมูลซีดี ดีวีดี แฟลชไดร์ฟ ฮาร์ดดิสก์แบบพกพา กลายเป็นเทคโนโลยีดาวร่วง เหตุผู้ใช้เปลี่ยนไปเก็บข้อมูลในระบบปฏิบัติการคลาวด์แทน

กิติพงค์ พร้อมวงค์ เลขาธิการสำนักงานคณะกรรมการนโยบายวิทยาศาสตร์ เทคโนโลยี และนวัตกรรมแห่งชาติ (สวทน.) กระทรวงวิทยาศาสตร์และเทคโนโลยี กล่าวว่า แนวโน้มเทคโนโลยีและนวัตกรรมที่มีโอกาสพัฒนาและก้าวเข้ามาในปี 2561 จะมุ่งเน้นความเป็นอัจฉริยะและความปลอดภัยในชีวิต ที่เพิ่มมากขึ้น

โลกดิจิทัลจะก้าวเข้าไปเกี่ยวข้องกับวิถีชีวิตและการดำเนินธุรกิจหลายประเภทเพิ่มสูงขึ้น ดังนั้นนวัตกรรมและเทคโนโลยีที่เป็นดาวรุ่ง ที่มีการนำมาประยุกต์ใช้ในชีวิตประจำวันและการดำเนินธุรกิจต่างๆ ในปี 2561 ได้ดังนี้

'เอไอ'มีบทบาทในทุกวงการ

1. กลุ่มอัจฉริยะเทคโนโลยีและนวัตกรรมที่มาแรงและเริ่มมีบทบาทสำคัญในโลกปัจจุบันจนถึงอนาคต คือ เทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence หรือ AI) คือ การใช้เทคโนโลยีคอมพิวเตอร์ร่วมกับเครื่องจักรกลเพื่อการเลียนแบบสติปัญญาและการกระทำของมนุษย์ เช่น การเข้าใจภาษา การวิเคราะห์ด้วยเหตุและผล และการรับรู้สภาพแวดล้อมต่างๆ บริษัทเทคโนโลยีและนวัตกรรมชั้นนำระดับโลกมีการนำเทคโนโลยี เอไอเข้ามาเพื่อพัฒนานวัตกรรมและการบริการเป็นอย่างมาก เช่น กูเกิล ,เฟซบุ๊ค และแอปเปิล ที่นำเอไอ มาใช้งานด้านเพื่อเป็นเครื่องมือเพื่อการตัดสินใจ

มีการคาดการณ์ว่า หากมีการนำเทคโนโลยีเอไอมาใช้งานจำนวนมากจะทำให้การใช้แรงงานมนุษย์ลดลง ตัวอย่างการใช้เทคโนโลยีเอไอแล้วมีผลกระทบต่อการลดจำนวนแรงงานคน เช่น ยานพาหนะที่ขับเคลื่อนได้ด้วยตัวเอง

หนึ่งในตัวอย่างของอุปกรณ์และสิ่งของอัจฉริยะ ที่หลายบริษัทยักษ์ใหญ่ประกาศเปิดตัวและทดลองขับ ทำให้อาชีพคนขับรถที่ใช้แรงงานมนุษย์ลดลง และจากเทคโนโลยีเอไอจะนำไปสู่การพัฒนาเทคโนโลยีที่เกี่ยวข้อง ได้แก่ Machine Learning, Neoro-Linguistic Programming (NLP) ในหลายประเทศแถบอเมริกาเหนือและลาตินอเมริกามีการนำเทคโนโลยีนี้เข้าไปใช้งาน และแก้ไขปัญหาในธุรกิจและอุตสาหกรรมต่างๆ เพิ่มขึ้น

เช่น ในบราซิล บริษัท Progressive Environmental & Agricultural Technologies จำกัด (PEAT) ได้พัฒนาแอพพลิเคชั่น Plantix ที่อาศัยพื้นฐานเทคโนโลยีเอไอ เพื่อการเก็บรวบรวมภาพถ่ายและดิน

http://www.bangkokbiznews.com/news/detail/787146

Computational Linguistics vs NLP

- If you use the term Computational Linguistics (CL)
  - You are probably a linguist
  - You use computers to study and analyze languages
- If you use the term NLP
  - You are probably an engineer/computer scientist
  - You work on applications that involve languages
- Most of the time they're very related (or the same thing)