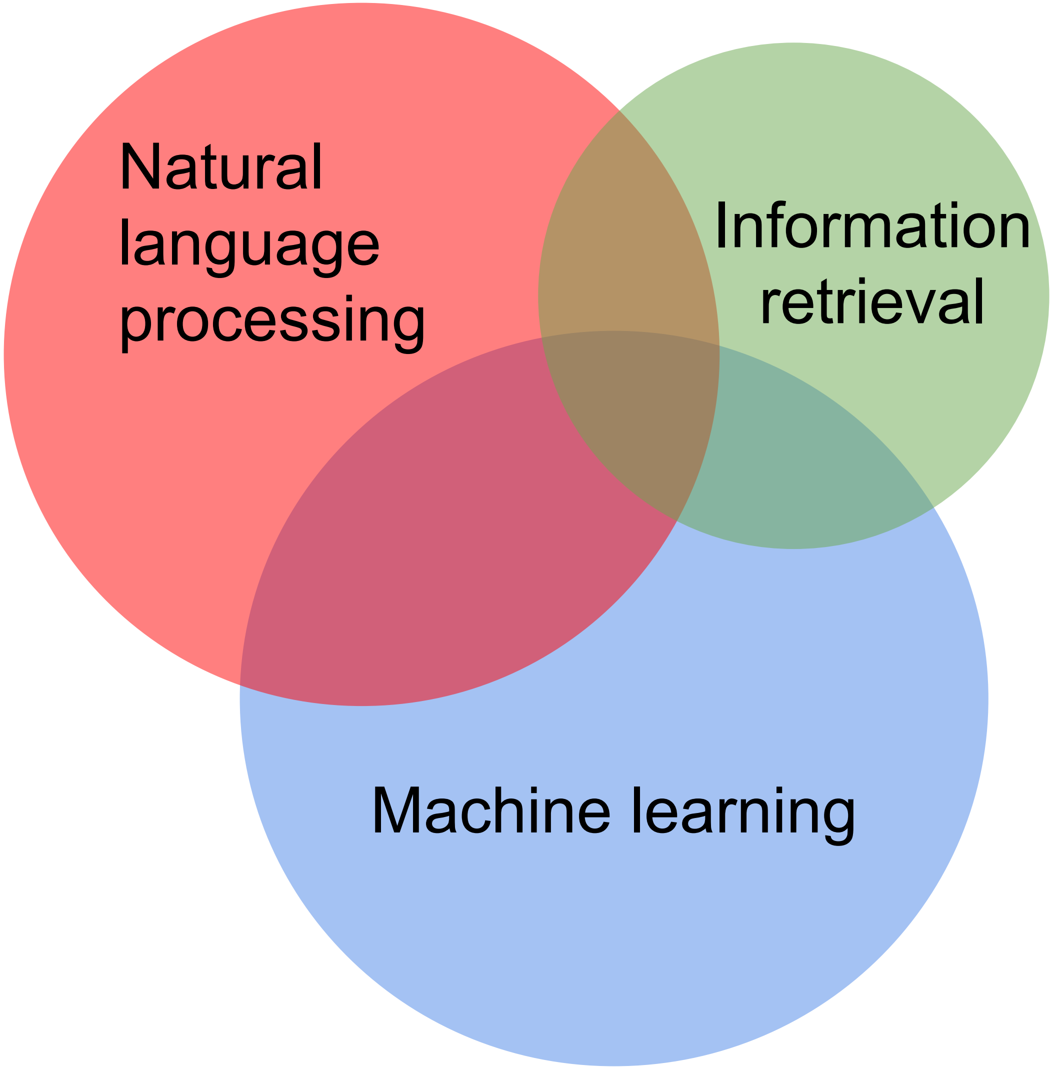
Introduction to Information Retrieval & Natural Language Processing

Speaker: Johann Chu

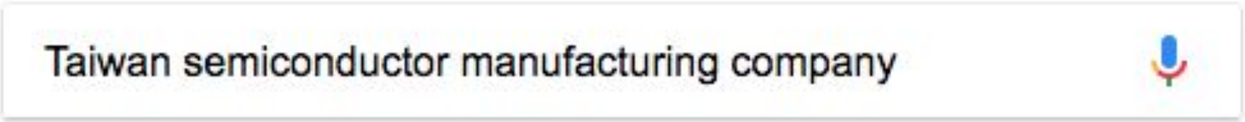
CloudMile

Relation between ML & IR & NLP



Let's imagine this scenario

- Imagine you are running a news agency and one day you want to find archived news article in your database regarding Taiwan Semiconductor Manufacturing Company.



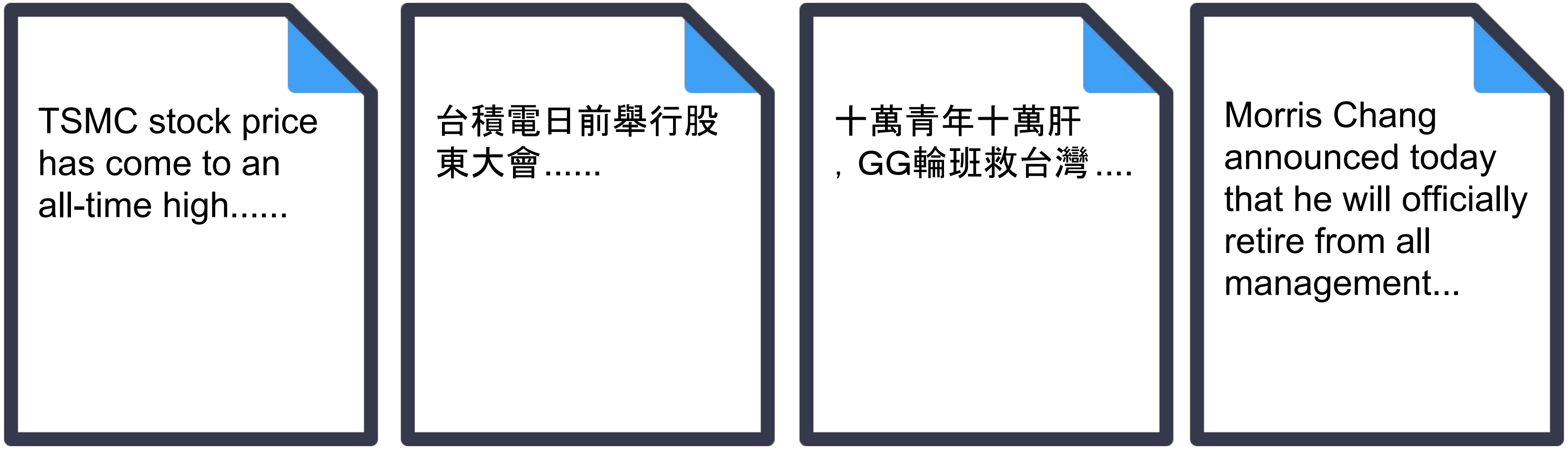
Taiwan semiconductor manufacturing company

SEARCH

- What would happen if we simply do a linear scan through the documents?

What's wrong with "finding" the key term directly?

- A lot of articles refer to them as "TSMC", "台積電", "台積", etc. If you use the company's full name as the query, you might miss other articles talking about the same company under a different term.
- You would probably also want to see articles talking about the founder of TSMC, Dr. Morris Chang, which might not be retrieved either under such query.



TSMC stock price has come to an all-time high......

台積電日前舉行股東大會......

十萬青年十萬肝，GG輪班救台灣....

Morris Chang announced today that he will officially retire from all management...

- An intelligent query system (or search engine) should be smart enough to give you the relevant results, while minimizing the entities that might seem relevant but actually aren't.

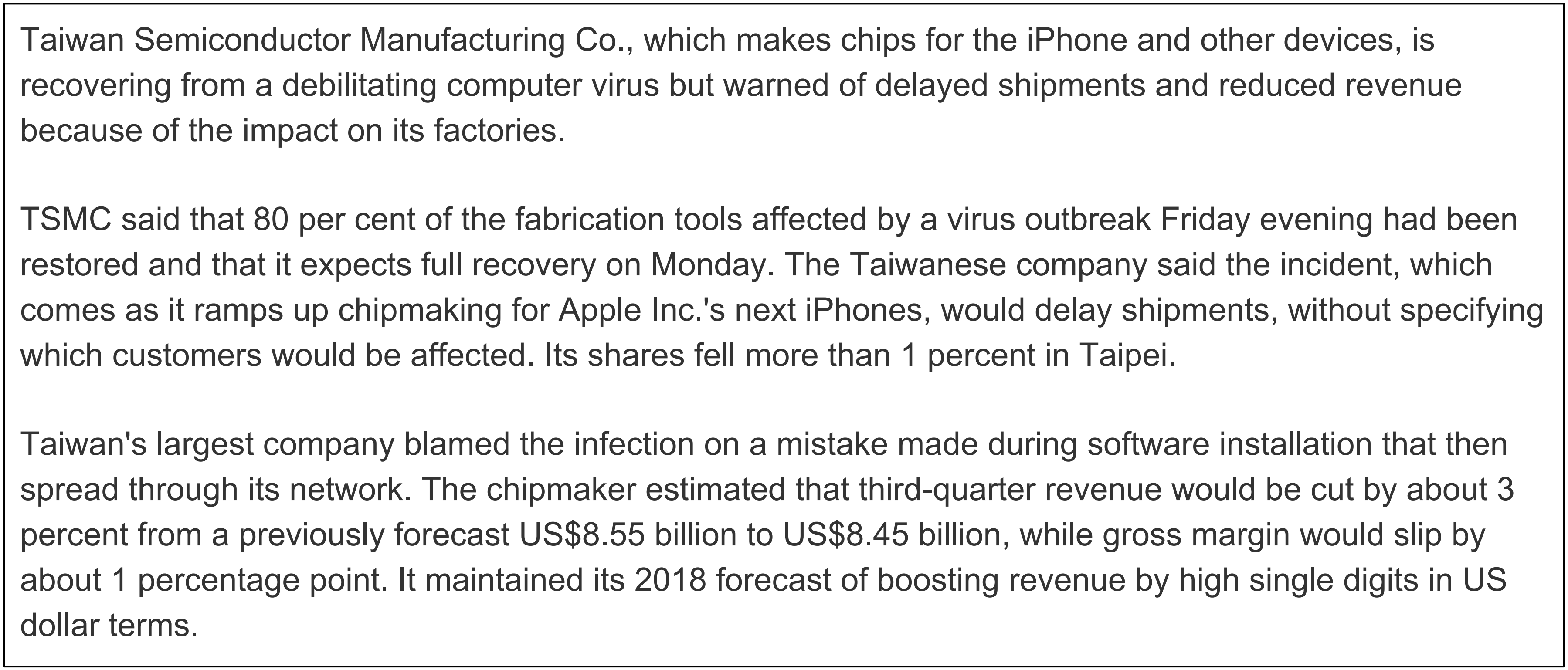
- Information retrieval is the act of finding essential information from a collection of information resource.

- Search engine is the perfect realization of information retrieval in our daily life.

Ok, so how do we fix the problem?

- To address the query issue of previous example, the first step is to index an article after it is placed in the database.

- But how should we index the articles? (e.g. Should we count all words? If not, what words should we use?)



Taiwan Semiconductor Manufacturing Co., which makes chips for the iPhone and other devices, is recovering from a debilitating computer virus but warned of delayed shipments and reduced revenue because of the impact on its factories.

TSMC said that 80 per cent of the fabrication tools affected by a virus outbreak Friday evening had been restored and that it expects full recovery on Monday. The Taiwanese company said the incident, which comes as it ramps up chipmaking for Apple Inc.'s next iPhones, would delay shipments, without specifying which customers would be affected. Its shares fell more than 1 percent in Taipei.

Taiwan's largest company blamed the infection on a mistake made during software installation that then spread through its network. The chipmaker estimated that third-quarter revenue would be cut by about 3 percent from a previously forecast US$8.55 billion to US$8.45 billion, while gross margin would slip by about 1 percentage point. It maintained its 2018 forecast of boosting revenue by high single digits in US dollar terms.

Taiwan, of, which, makes, largest, company, percent, shipment, customers,......

v.s.

TSMC, semiconductor, iPhone, factories, fabrication, chipmaking, forecast,......

- It is certainly unfeasible for us to keep a record for *all* words in the corpus.

- For an article like what you just saw, marking the occurrence of common terms such as "which", "makes", "largest" does not help much for us to find information of interest.

- We have to select the terms that can serve as better identifiers for the documents.

Taiwan Semiconductor Manufacturing Co., which makes chips for the iPhone and other devices, is recovering from a debilitating computer virus but warned of delayed shipments and reduced revenue because of the impact on its factories.

TSMC said that 80 per cent of the fabrication tools affected by a virus outbreak Friday evening had been restored and that it expects full recovery on Monday. The Taiwanese company said the incident, which comes as it ramps up chipmaking for Apple Inc.'s next iPhones, would delay shipments, without specifying which customers would be affected. Its shares fell more than 1 percent in Taipei.

Taiwan's largest company blamed the infection on a mistake made during software installation that then spread through its network. The chipmaker estimated that third-quarter revenue would be cut by about 3 percent from a previously forecast US$8.55 billion to US$8.45 billion, while gross margin would slip by about 1 percentage point. It maintained its 2018 forecast of boosting revenue by high single digits in US dollar terms.

Taiwan, of, which, makes, largest, company, percent, shipment, customers,......

v.s.

TSMC, semiconductor, iPhone, factories, fabrication, chipmaking, forecast,......

But you cannot do this for a million articles...

China's deputy trade negotiator on Tuesday acknowledged the challenge in resuming negotiations with the US, saying: "How could you negotiate with someone when he puts a knife on your neck?" Wang Shouwen made the......

Fan Bingbing, one of China's most famous actresses, known internationally for films such as "X-Men: Days of Future Past", has not been seen in public since June. Though the details of her disappearance remain unknown, it......

An Indonesian teen ha been rescued after drifting at sea for 49 days on a floating fish trap and is back safely with his family, according to the country's foreign affairs ministry.19-year-old Aldi Novel Adilang had been working as a lamp......

China and Sweden are locked in an escalating diplomatic feud after a minor dispute over tourists outstaying their welcome in a hostel devolved into broad accusations of racism and official travel warnings.The crisis......

The founders of Instagram have resigned from the business they started eight years ago in San Francisco and built into a global phenomenon used by a billion people. Kevin Systrom and Mike Krieger founded the photo-sharing app in.......

Google CEO Sundar Pichai responded to reports that some staff had discussed tweaking search results to show a pro-immigration bias. "We do not bias our products to favor any political agenda," he wrote in an email to Google ......

Ebola virus disease ha sickened 150 people and caused the deaths of at least 69 people in the Democratic Republic of Congo's northeastern region, the nation's Ministry of Health reported Monday. Of these 150 cases, ...

A Louisiana soldier wa sentenced to 11 years in prison Monday for constructing and detonating a bomb last year near the Fort Polk Army post. Ryan Keith Taylor, 24, of New Llano, Louisiana, pleaded guilty in June to ......

New Zealand Prime Minister Jacinda Ardern, who gave birth while in office, has made history by bringing her three-month-old daughter into the United Nations assembly hall. Ardern, 38, was photographed kissing and coddling Neve ......

- In certain task we would like to (automatically) find the important words that can be used to serve as indicators (among tens of thousands of documents). This task is commonly called keyword extraction.

I know! We can just count the number of times a word shows up in a document, and that number will tell us how important the word is!

What Peppa Pig means: Term Frequency (TF)

- It is very tempting and intuitive to use the raw count of different words in a document; the more academic way of saying raw count is term frequency.

- We can use distinct terms and their counts to model a document; nevertheless, this method disregards all ordering between words. It resembles having an article printed on paper and have all the words cut out into tiny pieces and throw them into a bag.

- This scheme is called the bag-of-words model.

Taiwan Semiconductor Manufacturing Co., which makes chips for the iPhone and other devices, is recovering from a debilitating computer virus but warned of delayed shipments and reduced revenue because of the impact on its factories.

TSMC said that 80 percent of the fabrication tools affected by a virus outbreak Friday evening had been restored and that it expects full recovery on Monday. The Taiwanese company said the incident, which comes as it ramps up chipmaking for Apple Inc.'s next iPhones, would delay shipments, without specifying which customers would be affected. Its shares fell more than 1 percent in Taipei.

Taiwan's largest company blamed the infection on a mistake made during software installation that then spread through its network. The chipmaker estimated that third-quarter revenue would be cut by about 3 percent from a previously forecast US$8.55 billion to US$8.45 billion, while gross margin would slip by about 1 percentage point. It maintained its 2018 forecast of boosting revenue by high single digits in US dollar terms.

the	6
and	3
of	3
in	2
by	2
Taiwan	2
which	
makes	
largest	
company	
percent	
shipment	
fabrication	1
TSMC	1
Semiconductor	1
Manufacturing	1
......	

Can you see the problem here?

- If we only use raw count to serve as the gauge of importance, common words would have significant importance. (E.g. if, because, of, which)

- Nevertheless, these common words would frequently show up in all types of documents, making them inefficient identifiers.

- We need some method to “bring down” the importance of words that occur too frequently everywhere.

Introducing: Inverse Document Frequency (IDF)

- We assume that a good identifier word should exist in a small number of articles only among a document collection. Under such assumption, we have a new term weighting metric, inverse document frequency, defined as:

$$\mathrm{idf}_t = \log \frac{N}{\mathrm{df}_t}$$

where N = total number of documents in the collection

df_t = number of articles having the term *t*

Combine the two: TF-IDF

- By multiplying the term frequency with its inverse document frequency, we can obtain a more objective weighting of terms in a corpus. The combined weighting is called TF-IDF, defined as:

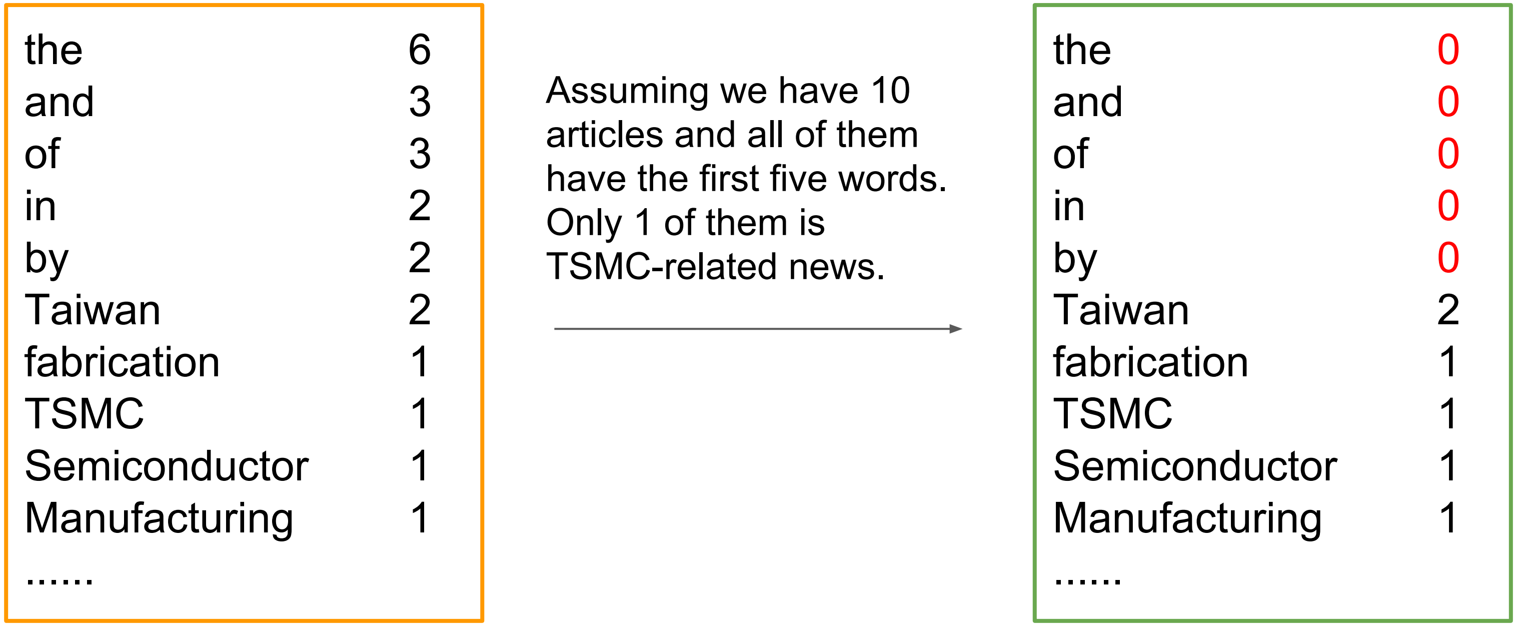
$$\text{tf-idf}_{t,d} = \text{tf}_{t,d} * \text{idf}_{t}$$

- So how does it work?

Taiwan Semiconductor Manufacturing Co., which makes chips for the iPhone and other devices, is recovering from a debilitating computer virus but warned of delayed shipments and reduced revenue because of the impact on its factories.

TSMC said that 80 percent of the fabrication tools affected by a virus outbreak Friday evening had been restored and that it expects full recovery on Monday. The Taiwanese company said the incident, which comes as it ramps up chipmaking for Apple Inc.'s next iPhones, would delay shipments, without specifying which customers would be affected. Its shares fell more than 1 percent in Taipei.

Taiwan's largest company blamed the infection on a mistake made during software installation that then spread through its network. The chipmaker estimated that third-quarter revenue would be cut by about 3 percent from a previously forecast US$8.55 billion to US$8.45 billion, while gross margin would slip by about 1 percentage point. It maintained its 2018 forecast of boosting revenue by high single digits in US dollar terms.



the	6
and	3
of	3
in	2
by	2
Taiwan	2
fabrication	1
TSMC	1
Semiconductor	1
Manufacturing	1
......	

Term Frequency (TF)

Assuming we have 10 articles and all of them have the first five words. Only 1 of them is TSMC-related news.

the	0
and	0
of	0
in	0
by	0
Taiwan	2
fabrication	1
TSMC	1
Semiconductor	1
Manufacturing	1
......	

TF-IDF

What we can do with TF-IDF

- By using TF-IDF, we can not only effectively extract keywords out of documents, but also use them to represent the documents.

Taiwan Semiconductor Manufacturing Co., which makes chips for the iPhone and other devices, is recovering from a debilitating computer virus but warned of delayed shipments and reduced revenue because of the impact on its factories.

TSMC said that 80 percent of the fabrication tools affected by a virus outbreak Friday evening had been restored and that it expects full recovery on Monday. The Taiwanese company said the incident, which comes as it ramps up chipmaking for Apple Inc.'s next iPhones, would delay shipments, without specifying which customers would be affected. Its shares fell more than 1 percent in Taipei.

Taiwan's largest company blamed the infection on a mistake made during software installation that then spread through its network. The chipmaker estimated that third-quarter revenue would be cut by about 3 percent from a previously forecast US$8.55 billion to US$8.45 billion, while gross margin would slip by about 1 percentage point. It maintained its 2018 forecast of boosting revenue by high single digits in US dollar terms.

the	0
and	0
of	0
in	0
by	0
Taiwan	2
fabrication	1
TSMC	1
Semiconductor	1
Manufacturing	1
......	

We can call this vector space model!

Day2釣出前度新歡 楊丞琳哭掀舊情

【陳薈涵、蔡維歆／台北報導】楊丞琳昨在台北小巨蛋舉辦第2場「青春住了誰」演唱會，演唱嘉賓則邀來綜藝天王吳宗憲（憲哥）合唱《屋頂》，憲哥一登台就讓她淚崩，她虧自己像神經病，憲哥也跟著落淚，她笑說：「我13歲參加第一個電視節目比賽是《超級新人王》，憲哥是這個節目的主持人，是20年前的事了。」男友李榮浩前天還在上海出席活動，昨排除萬難現身，戴著口罩在包廂低調欣賞，快閃來台力挺愛人，而舊愛小鬼也受邀出席。

楊丞琳	13.21
憲哥	13.26
李榮浩	6.25
小鬼	6.10
巨蛋	4.08
出席	2.56
主持人	1.85
電視	1.70

We can even compare the similarity between the documents!

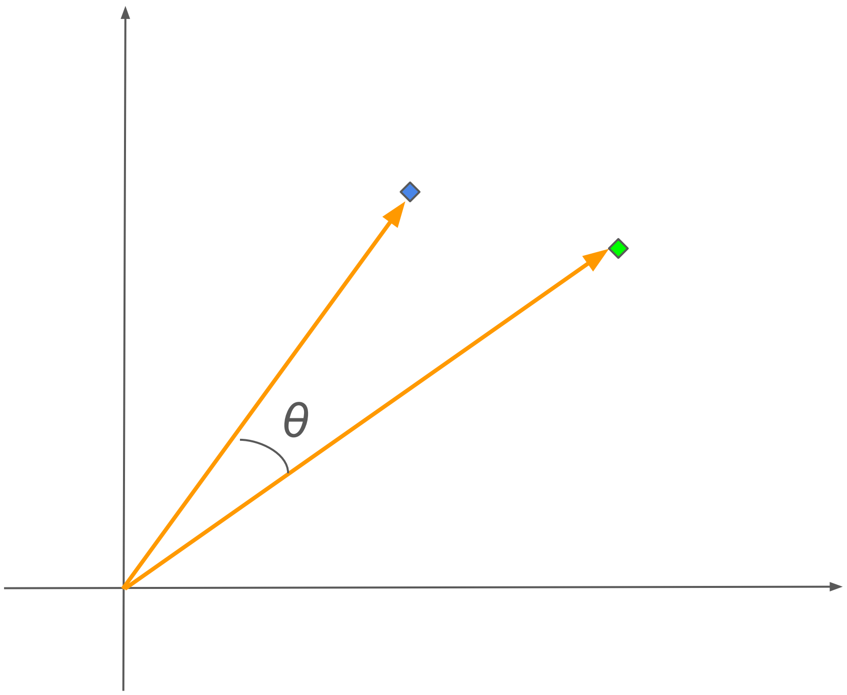
李玖哲淚崩！突爆結婚超過3年　老婆相馬茜不離不棄

（新增動新聞）金曲歌王李玖哲今在記者會上宣布已與相馬茜結婚！他倆2004年因學中文相識，多次被拍到親密同進同出，直到2010年承認戀情。李玖哲事後平訪談到老婆淚流滿面，他邊哭邊感謝相馬茜這幾年即便他遭遇低潮也不離不棄，更感性說：「希望我們愛到老，就算死掉，下輩子也能見到她。」李玖哲7年未發片，今盛大舉辦發片記者會《Will You Remember》，老闆黃立成（麻吉大哥）身兼主持人，他笑問：「媒體很關心你有沒有女朋友？」李玖哲羞答：「大哥你幹嘛問我『前女友』的事？」他接著笑說：「大家可以叫她李小茜，因為她已經是李太太了。」引來現場一陣驚呼。黃立成笑說：「剛剛才知道李玖哲結婚，只知道他辦了場小婚禮，但沒邀我去，很久沒看到相馬茜，她目前住日本。」

相馬茜	55.89
李玖哲	42.18
老婆	19.65
結婚	17.06
發片	13.36
認戀情	6.91
腫胖	4.78
大哥	4.53
麻吉	2.99
前女友	2.38

Measuring Similarity

- For data represented in the form of vector, we can calculate the angle between vectors and use that as a metric for similarity.



Cosine similarity: angle between vectors

$$\cos(\theta) = \frac{\mathbf{A}\cdot\mathbf{B}}{\|\mathbf{A}\|_2\|\mathbf{B}\|_2} = \frac{\sum\limits_{i=1}^{n} A_i B_i}{\sqrt{\sum\limits_{i=1}^{n} A_i^2}\sqrt{\sum\limits_{i=1}^{n} B_i^2}}$$

老闆娘這麼辣可以嗎？林又立捧自創品牌視角太邪惡

遐想	3.33
閒到	4.30
凍齡	1.53
迷航	3.27
入伍	2.72
豁出去	3.43
瑞莎	3.52
化身	2.32
墨鏡	2.71
underline	1.18
日前	0.94
微微	2.98
漂亮	1.98
溫翠蘋	3.60
比基尼	2.58
近期	1.47
還有	0.97
度假	2.19
曬出	3.02

主播兒子成家 《101次求婚》女星當性感婆婆

私會	3.45
圈外	2.87
求婚	2.39
度假	1.52
美女	2.02
兒子	1.40
拍攝	3.65
漂亮	0.02
林心如	2.62
日本	1.07
凍齡	2.88
電視台	2.06
持續	1.12
比基尼	0.46
女友	2.99
同年	2.10
新居	3.57
週刊	6.75
保養	4.15

$$\text{similarity} = \frac{1.53 * 2.88 + 1.98 * 0.02 + 2.58 * 0.46 + 2.19 * 1.52}{(3.33^2+4.30^2+1.53^2+......+2.19^2+3.02^2)^{\wedge 0.5}\ (3.45^2+2.87^2+2.39^2+......+6.75^2+4.15^2)^{\wedge 0.5}}$$

= You do the math ;)

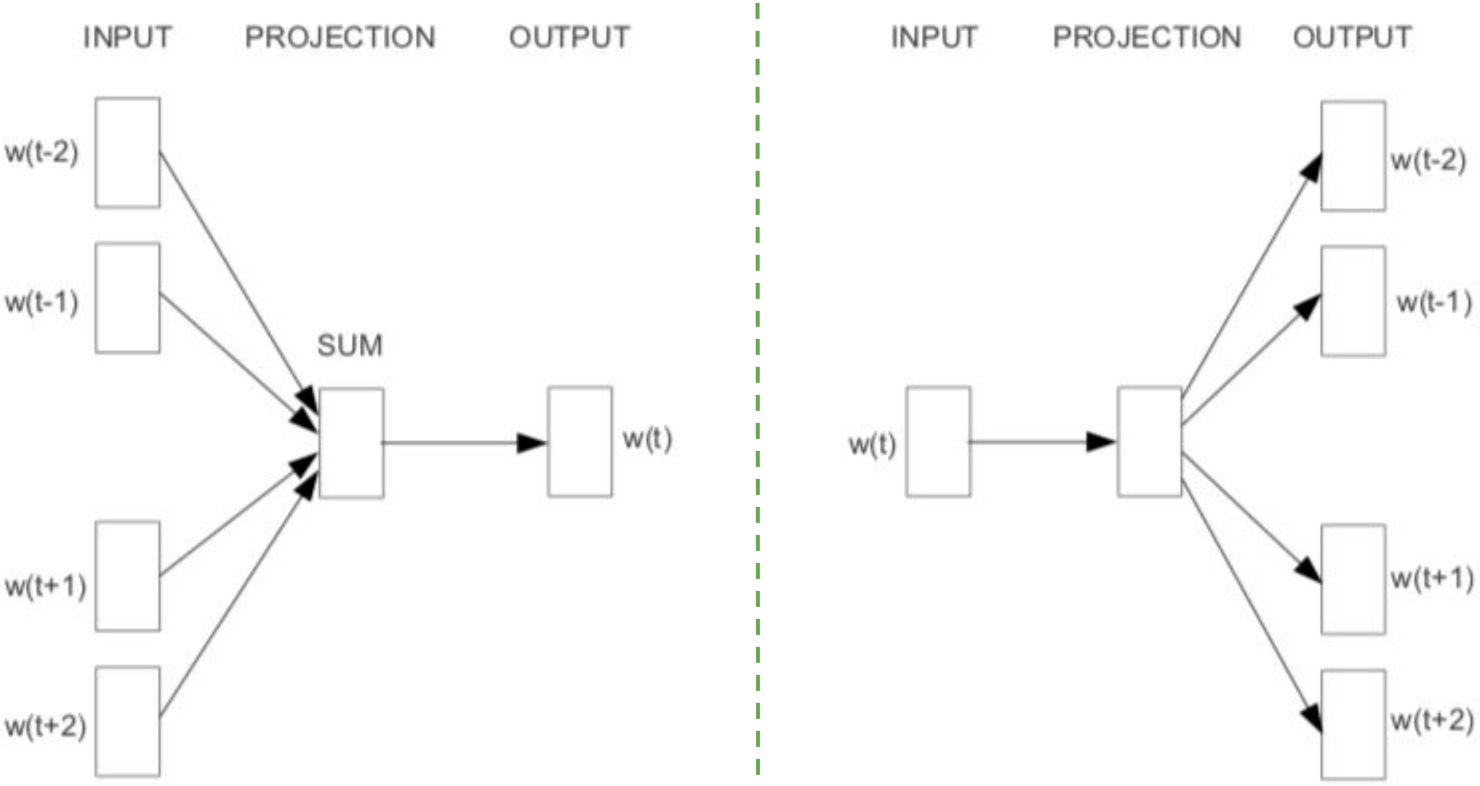
Is that the only way to vectorize the document?

- Converting lexical resource to vector space model has long been an active field of research. Apart from the document-to-vector conversion you just saw, we can also turn words into vectors. These word vectors are also called word embedding.

- For the past few years, Word2Vec has been the most popular model in NLP, developed by Dr. Tomas Mikolov when he was working in Google Brain.

(Source: https://research.fb.com/people/mikolov-tomas/)

How Word2Vec works

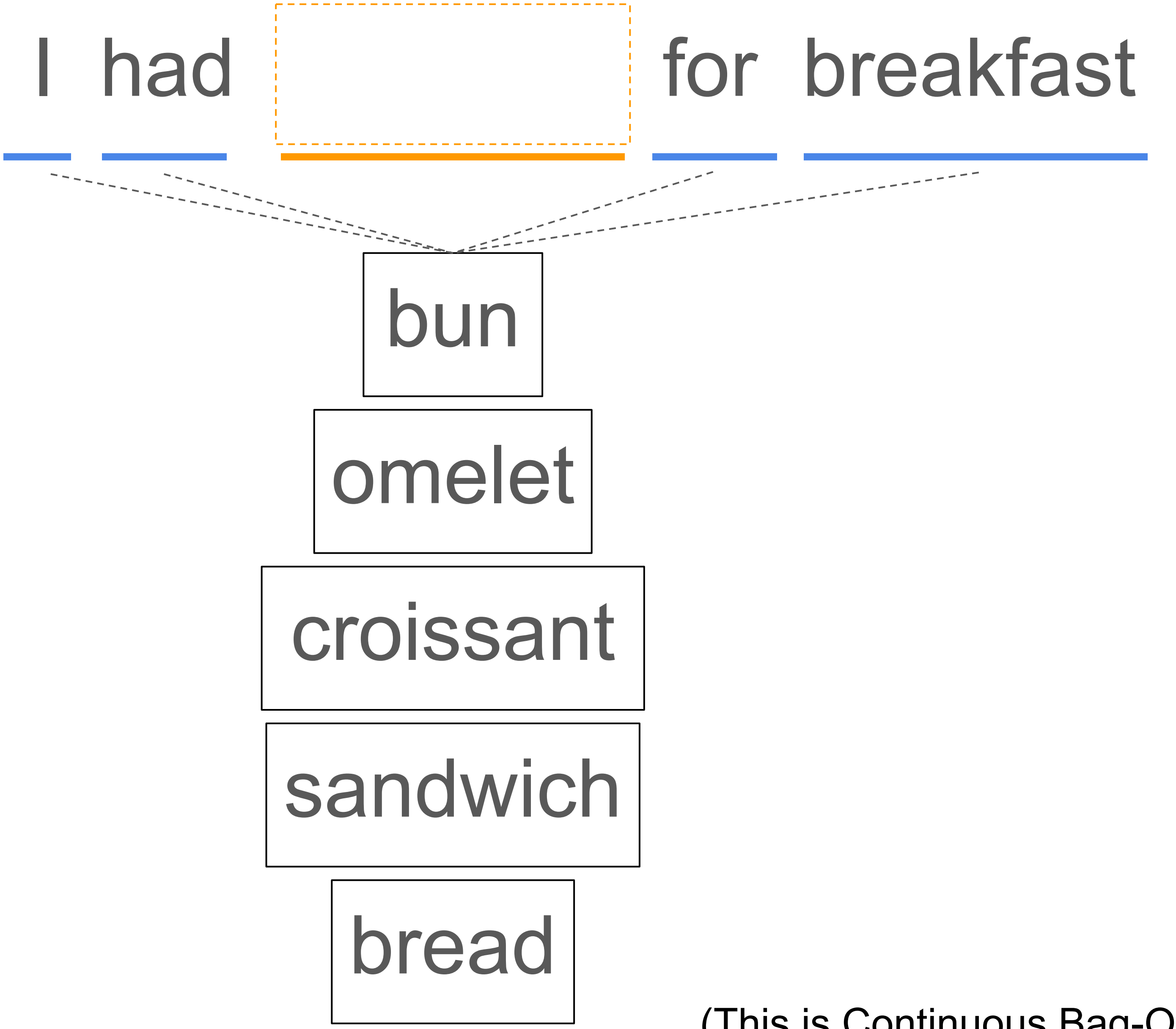


CBOW

(Continuous Bag-Of-Words)

Skip-gram

(Reference: T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space, 2003)



I had for breakfast

bun

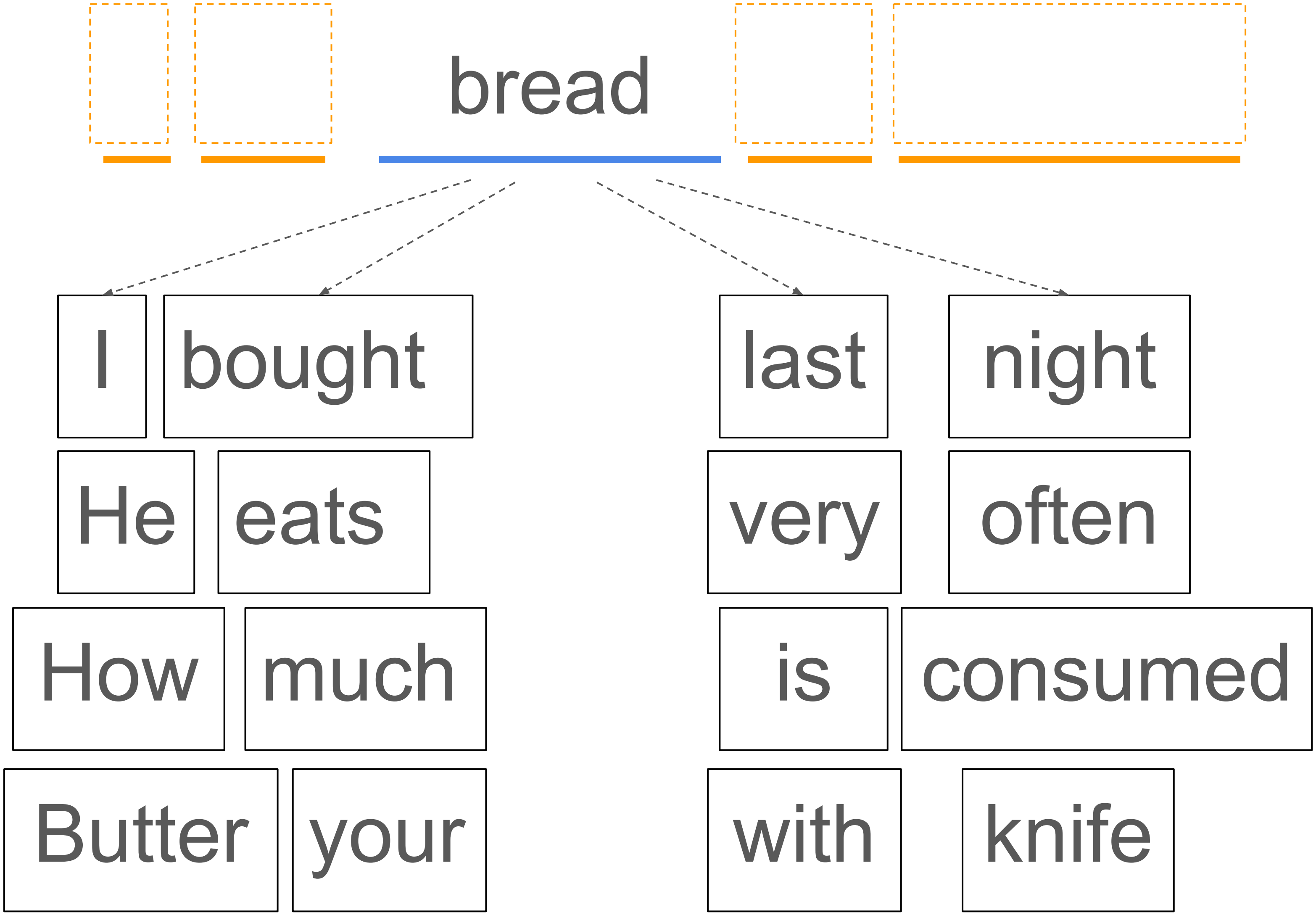
omelet

croissant

sandwich

bread

(This is Continuous Bag-Of-Words.)



bread

I bought last night

He eats very often

How much is consumed

Butter your with knife

(This is Skip-Gram.)

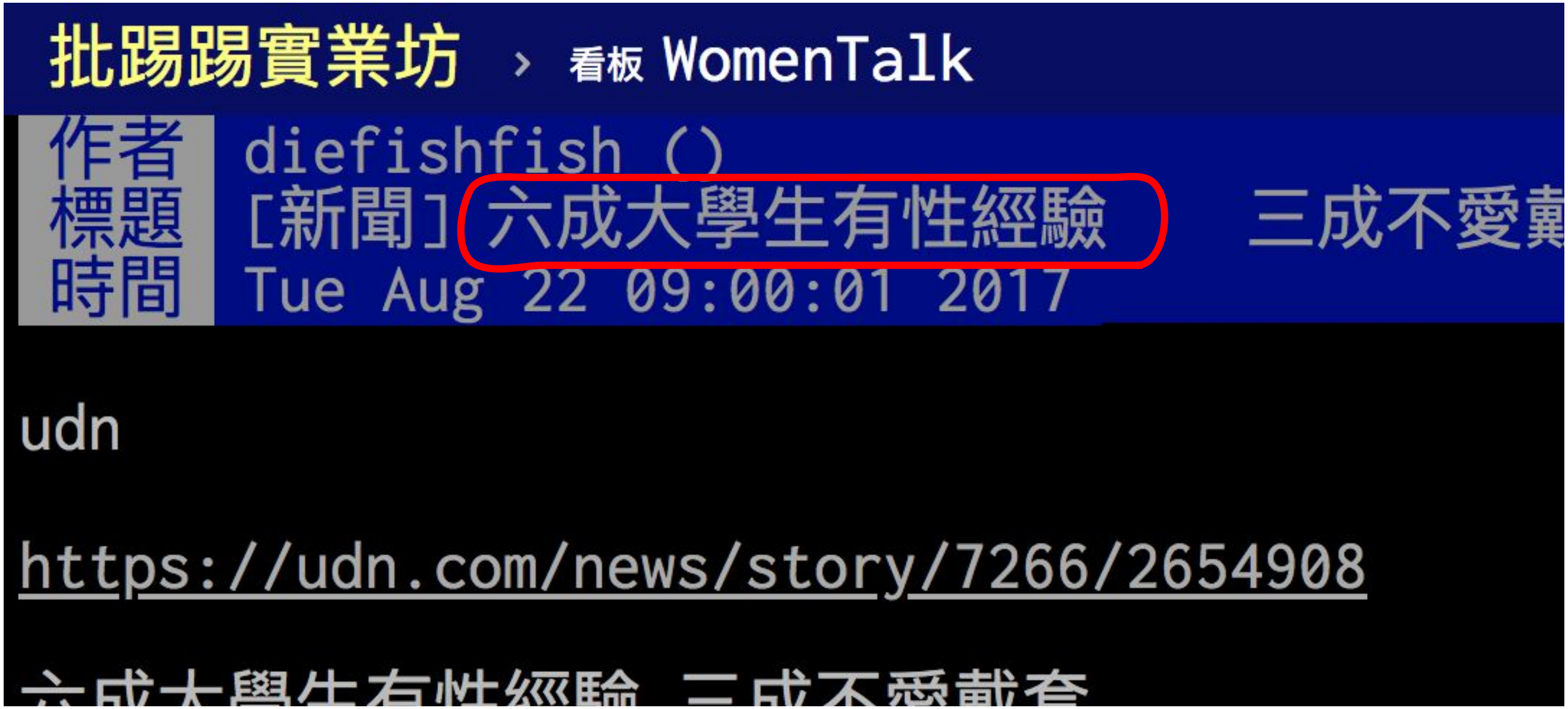
- Word2Vec assigns dimension values to words while ensuring that words occur in similar scenario often are "closer" in the vector space.

- It is essentially a neural network model using negative sampling (noise-contrastive estimation, to be exact) as the loss function to optimize the chance it can guess the right words when given the context.

- You can now use it conveniently with gensim, a python module developed to handle numerous NLP tasks.

But wait, there seems to be a major problem

- For continuous-character language system like Chinese, how do we disassemble the following sentence BEFORE we vectorize them?



批踢踢實業坊 › 看板 WomenTalk

作者 diefishfish ()
標題 [新聞] 六成大學生有性經驗 三成不愛戴
時間 Tue Aug 22 09:00:01 2017

udn

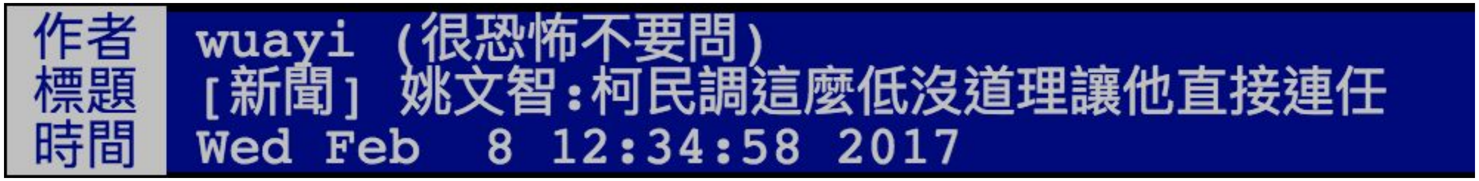
https://udn.com/news/story/7266/2654908

六成大學生有性經驗 三成不愛戴套

Word Segmentation

- Unlike English, Chinese literature do not have spaces between vocabulary to signify the beginning and end of a term. And that's a *huge* deal.

Example:



```
作者  wuayi (很恐怖不要問)
標題  [新聞] 姚文智:柯民調這麼低沒道理讓他直接連任
時間  Wed Feb  8 12:34:58 2017
```

⇨ 柯民調這麼低，沒道理讓他直接連任

⇨ 柯民調這麼低沒道理，讓他直接連任

- To build an automated segmentation system, we first need a collection of textual data manually annotated with parts-of-speech and other information.

E.g.

我	今	天	很	開	心
PRP	ADV-start	ADV-end	ADV	ADJ-start	ADJ-end

- We will then use statistical methods to analyze how often each character exists in each individual parts-of-speech form.

- For the past two decades, most of the segmentation methods make use of hidden Markov Model, Maximum Entropy, or Conditional Random Field on the annotated data to build the model.

- Recent years, people have been implementing neural network to build the segmentation system with good results.

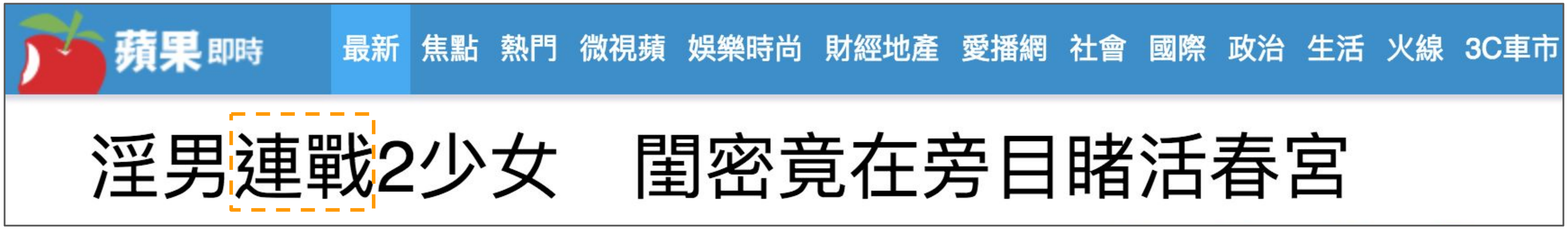
- Available Chinese segmentation system includes CKIP from Academia Sinica and Jieba from Mainland China.

Lab time

(Source: *Breaking Bad*, AMC)

POS tagging

- The segmentation task mentioned in the previous slides brought up a very important NLP issue: POS tagging. It is, like its name, used to tag raw corpus with parts-of-speech for each individual term.



淫男連戰2少女　閨密竟在旁目睹活春宮

ADV　V　?

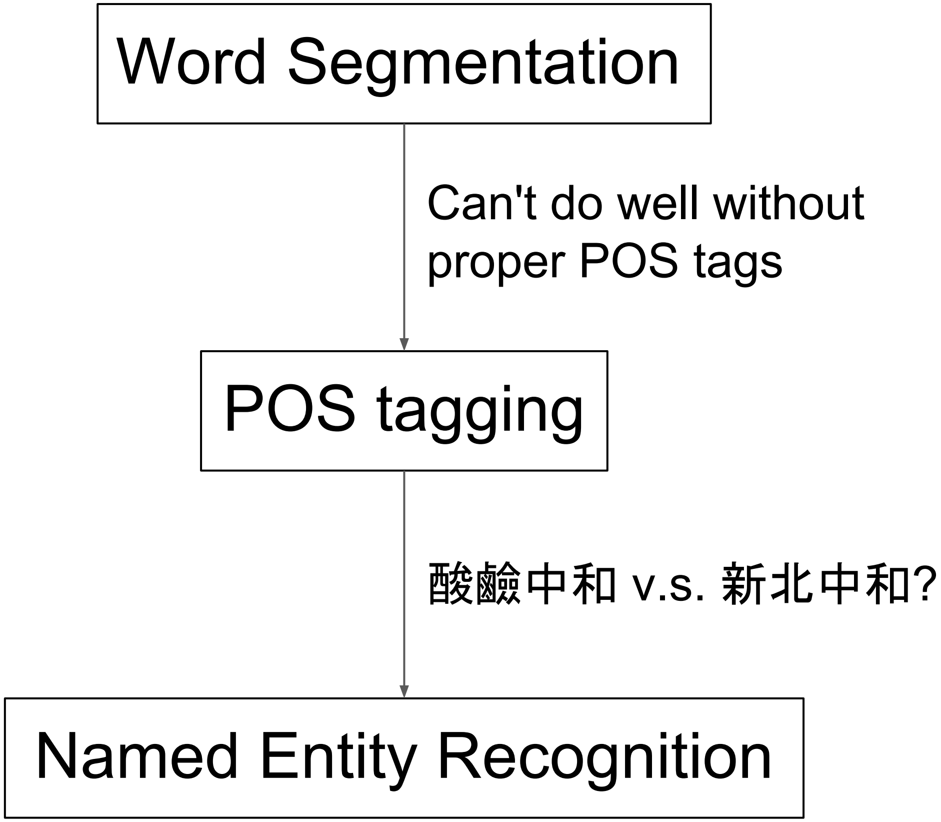
NN-start　NN-stop　?

Named Entity Recognition (NER)

- POS tagging is not easy; sometimes we even have to distinguish non-general names apart from the ordinary words in a context to deliver good result. This non-general noun identification is called named entity recognition. Named entity encompasses three types of nouns: person, location, and organization.

柯	文	哲	找	北	農	總	經	理	去	市	政	府
PER-START	PER-MID	PER-END	V	ORG-START	ORG-END	NN-START	NN-MID	NN-END	V	LOC-START	LOC-MID	LOC-END

- After going through these tasks, you can see that it is not easy to conduct one task without using another.



Word Segmentation

Can't do well without proper POS tags

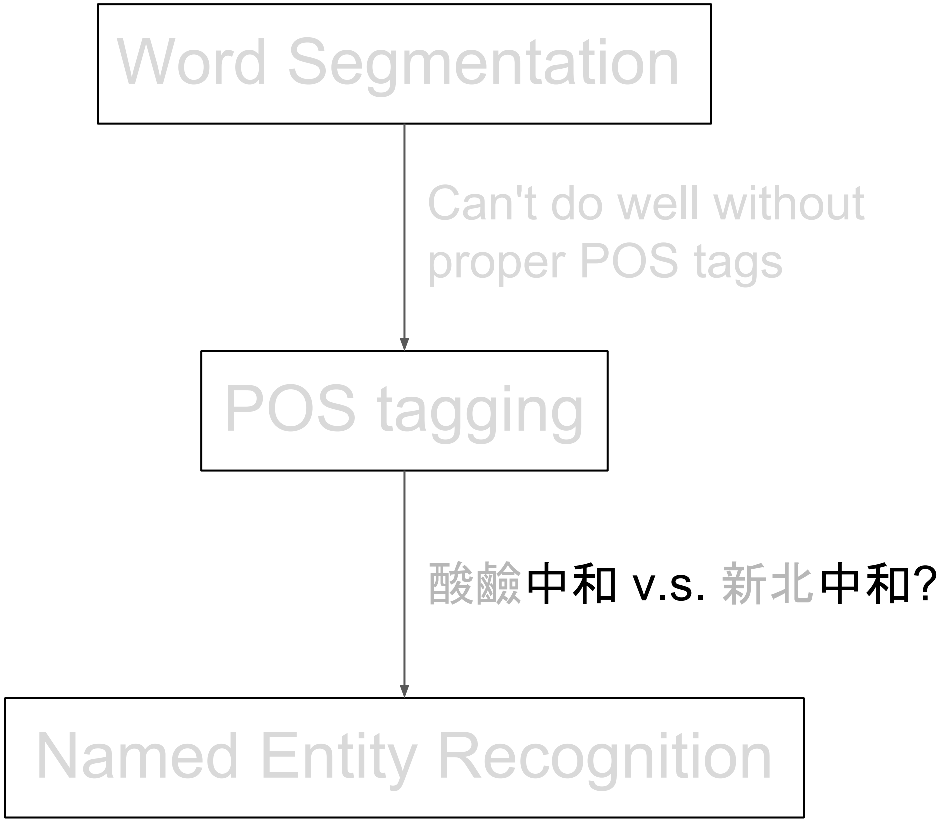
POS tagging

酸鹼中和 v.s. 新北中和?

Named Entity Recognition

Did you notice something?

- After going through these tasks, you can see that it is not easy to conduct one task without using another.



Word Segmentation

Can't do well without proper POS tags

POS tagging

酸鹼中和 v.s. 新北中和?

Named Entity Recognition

Word Sense Disambiguation

- A word, regardless of language, tends to have multiple meanings. For instance:

"Curry"

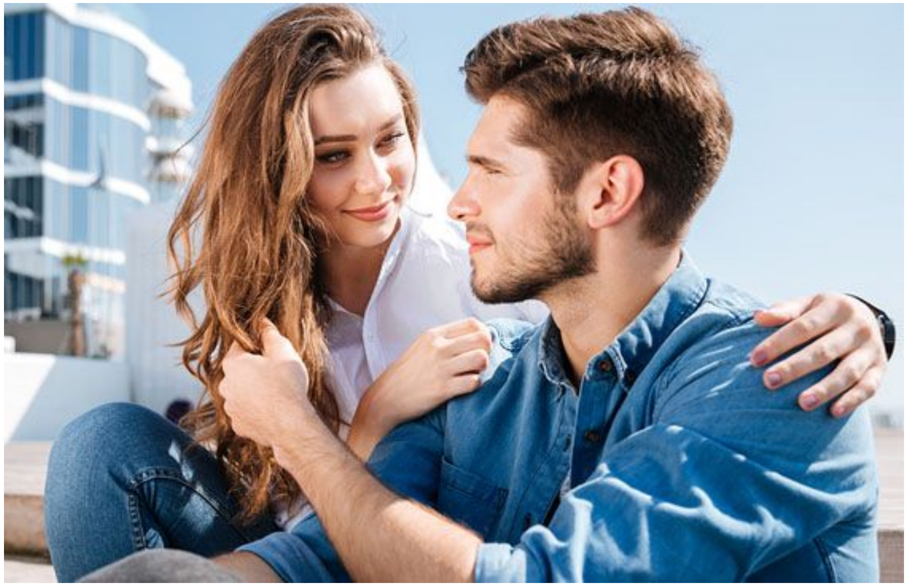
Warriors, three pointers, NBA, MVP, Oracle Center, Draymond Green

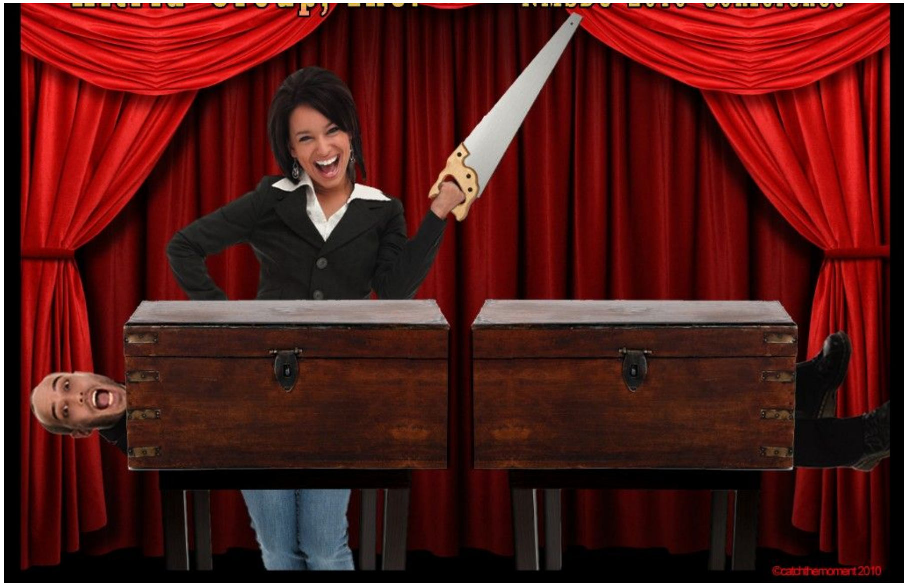
Pork, chicken, beef, Japanese, Indian, Thai, green, spicy

That's why machine translation is hard.

- Why is it hard? Let's take a look at the example below:

Input: I saw my husband.



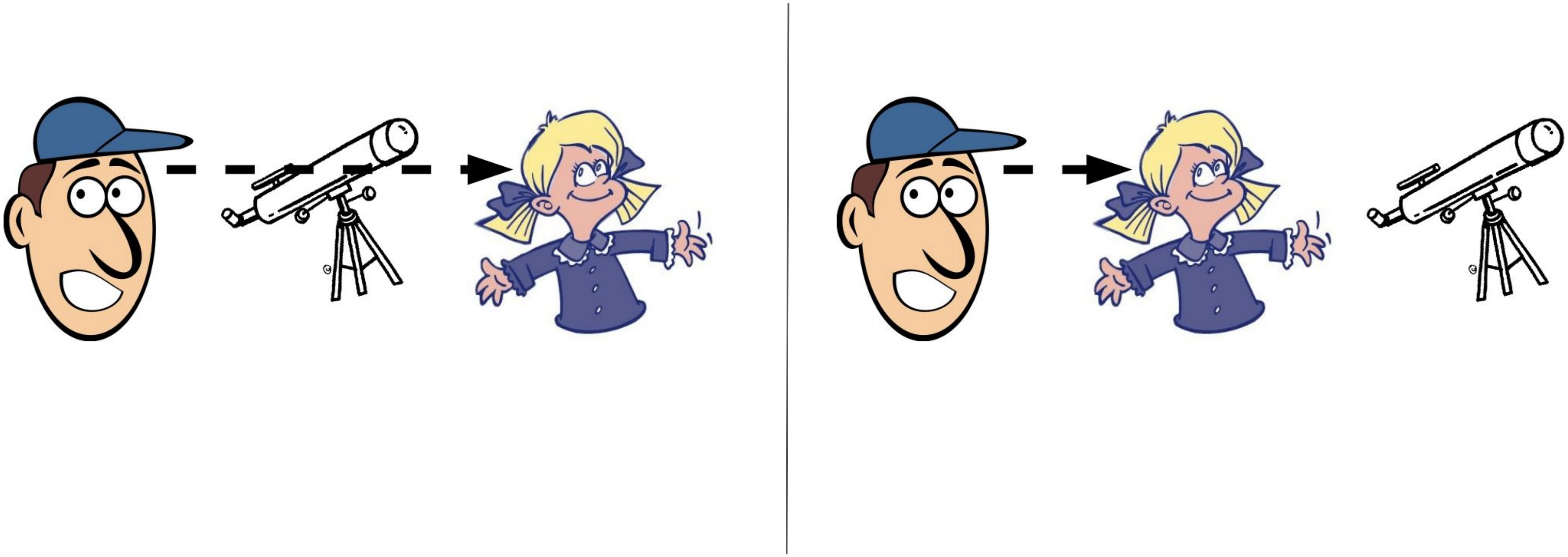


There are also other kinds of ambiguity...

- Consider this sentence. What does it mean?

E.g.

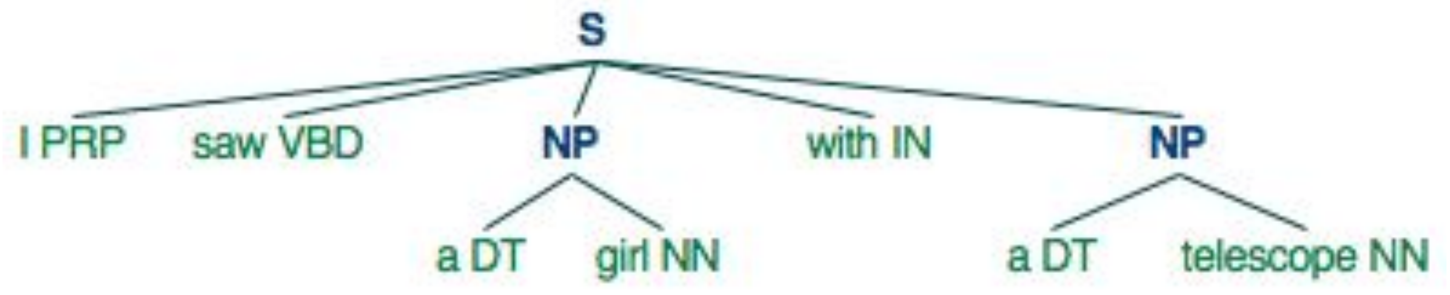
I saw a girl with a telescope



Parsing

- To better understand the structure of a sentence, the first step would be performing parsing on text.

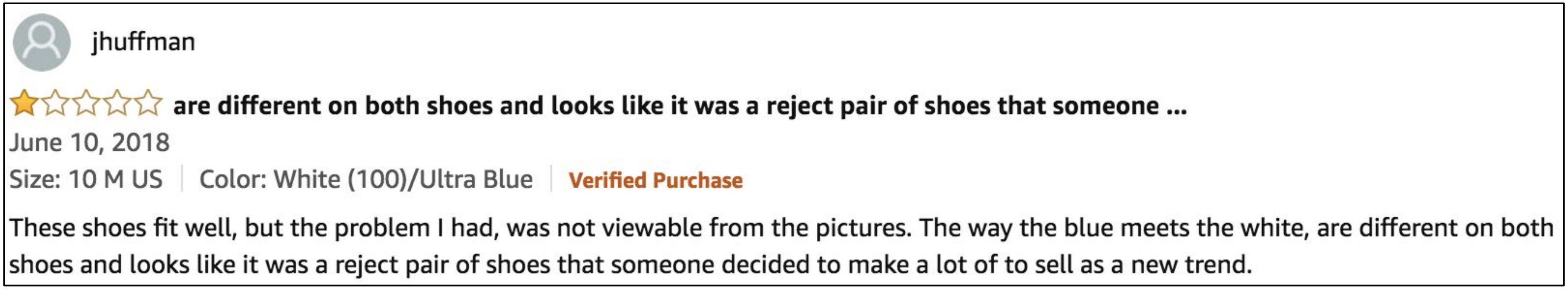
- A natural language parser is a program that works out the grammatical structure of sentences.



- For convenience you can use NLTK with python; you can also try Stanford CoreNLP package for more powerful features.

Some other things we can do with NLP...

- Imagine you are hosting an e-commerce website and you have tons of user comments about products every day. How can you quickly find the negative comments to grasp the quality of products?



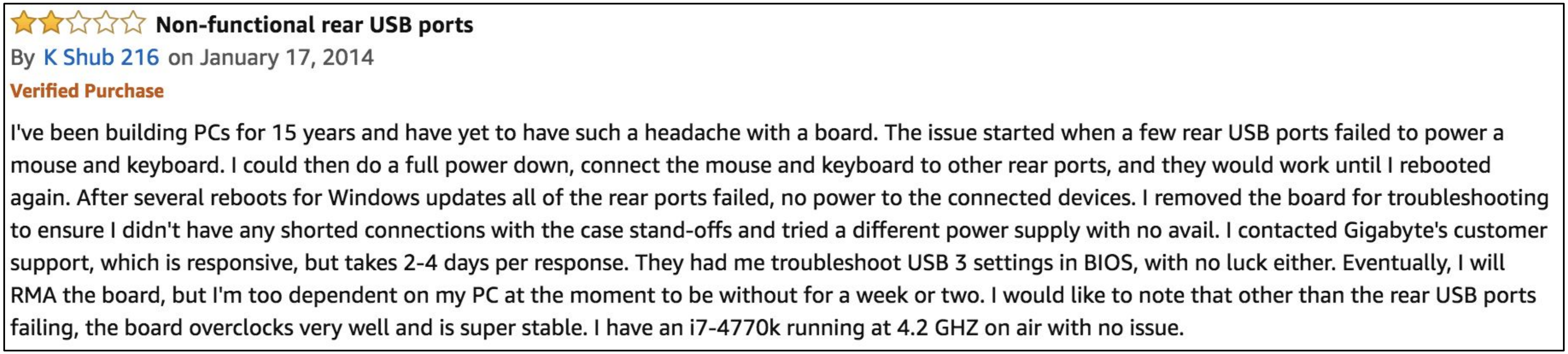
jhuffman

★☆☆☆☆ **are different on both shoes and looks like it was a reject pair of shoes that someone ...**

June 10, 2018

Size: 10 M US | Color: White (100)/Ultra Blue | **Verified Purchase**

These shoes fit well, but the problem I had, was not viewable from the pictures. The way the blue meets the white, are different on both shoes and looks like it was a reject pair of shoes that someone decided to make a lot of to sell as a new trend.



★★☆☆☆ **Non-functional rear USB ports**

By K Shub 216 on January 17, 2014

Verified Purchase

I've been building PCs for 15 years and have yet to have such a headache with a board. The issue started when a few rear USB ports failed to power a mouse and keyboard. I could then do a full power down, connect the mouse and keyboard to other rear ports, and they would work until I rebooted again. After several reboots for Windows updates all of the rear ports failed, no power to the connected devices. I removed the board for troubleshooting to ensure I didn't have any shorted connections with the case stand-offs and tried a different power supply with no avail. I contacted Gigabyte's customer support, which is responsive, but takes 2-4 days per response. They had me troubleshoot USB 3 settings in BIOS, with no luck either. Eventually, I will RMA the board, but I'm too dependent on my PC at the moment to be without for a week or two. I would like to note that other than the rear USB ports failing, the board overclocks very well and is super stable. I have an i7-4770k running at 4.2 GHZ on air with no issue.

Sentiment Analysis

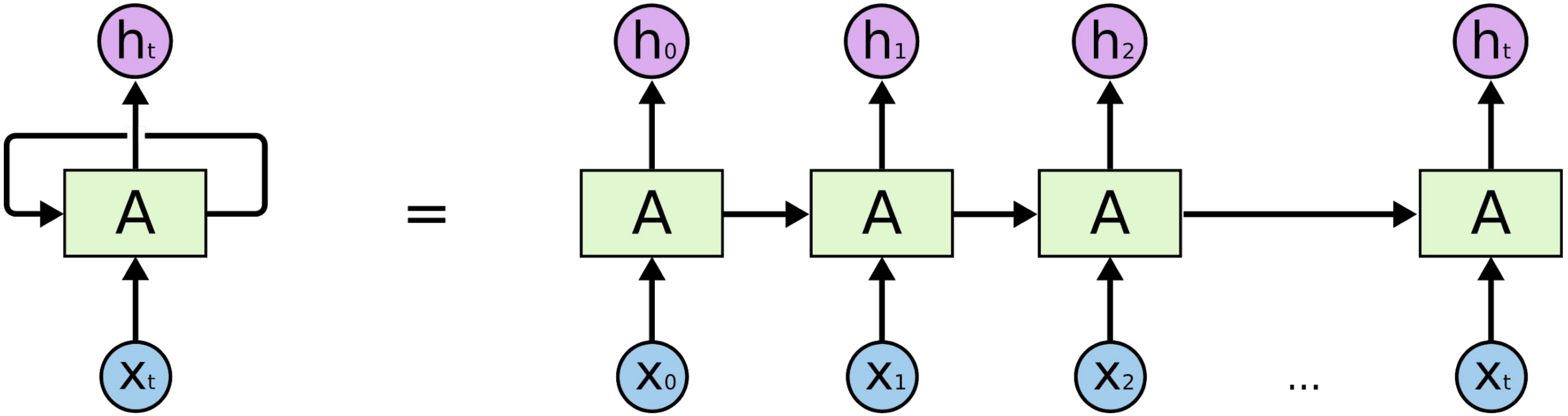
- The art of detecting the emotion of authors from comments is called sentiment analysis. Most of the time it is a binary classification problem (positive v.s. negative), while sometimes people add in neutral as well.

- Most approaches before 2010s were keyword-based (e.g. happy, easy, convenient, frustrated, piss off) or used n-gram. Such approaches are fast, but are easily affected when profanity is used to express strong emotion. (e.g. "This thing is *fxxking* awesome!!!")

- They are also incapable of detecting sarcasm.

- In recent years people are gradually inclined to use neural network for sentiment analysis (and a lot of other stuff).

- In NLP, recurrent neural network (RNN) is often preferred due to the sequential nature of human language and the ability of RNN to take a certain degree of historical data into account during training.



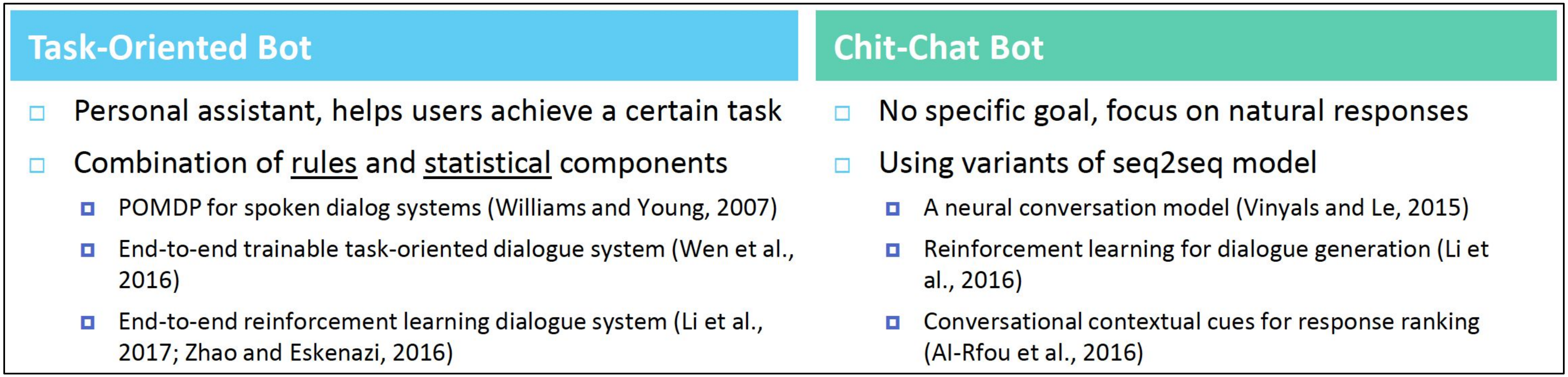
Last but not the least...

- It finally comes to the most interesting and yet challenging application of NLP. This technology used to only exist in sci-fi movies; thanks to the advances in NLP, people are gradually having their hands on it.
  (Not Iron Man's arc reactor.)

(Source: *Iron Man*, 2008, produced by Marvel Studios)

Dialogue Systems

- Dialogue systems, or sometime called chatbot, is the holy grail of NLP research. It is difficult due to the content diversity of human conversation.
- The main branches of dialogue systems are task-oriented bot and chit-chat bot, with the former slightly easier to build than the latter.



Task-Oriented Bot

□ Personal assistant, helps users achieve a certain task

□ Combination of rules and statistical components

- POMDP for spoken dialog systems (Williams and Young, 2007)
- End-to-end trainable task-oriented dialogue system (Wen et al., 2016)
- End-to-end reinforcement learning dialogue system (Li et al., 2017; Zhao and Eskenazi, 2016)

Chit-Chat Bot

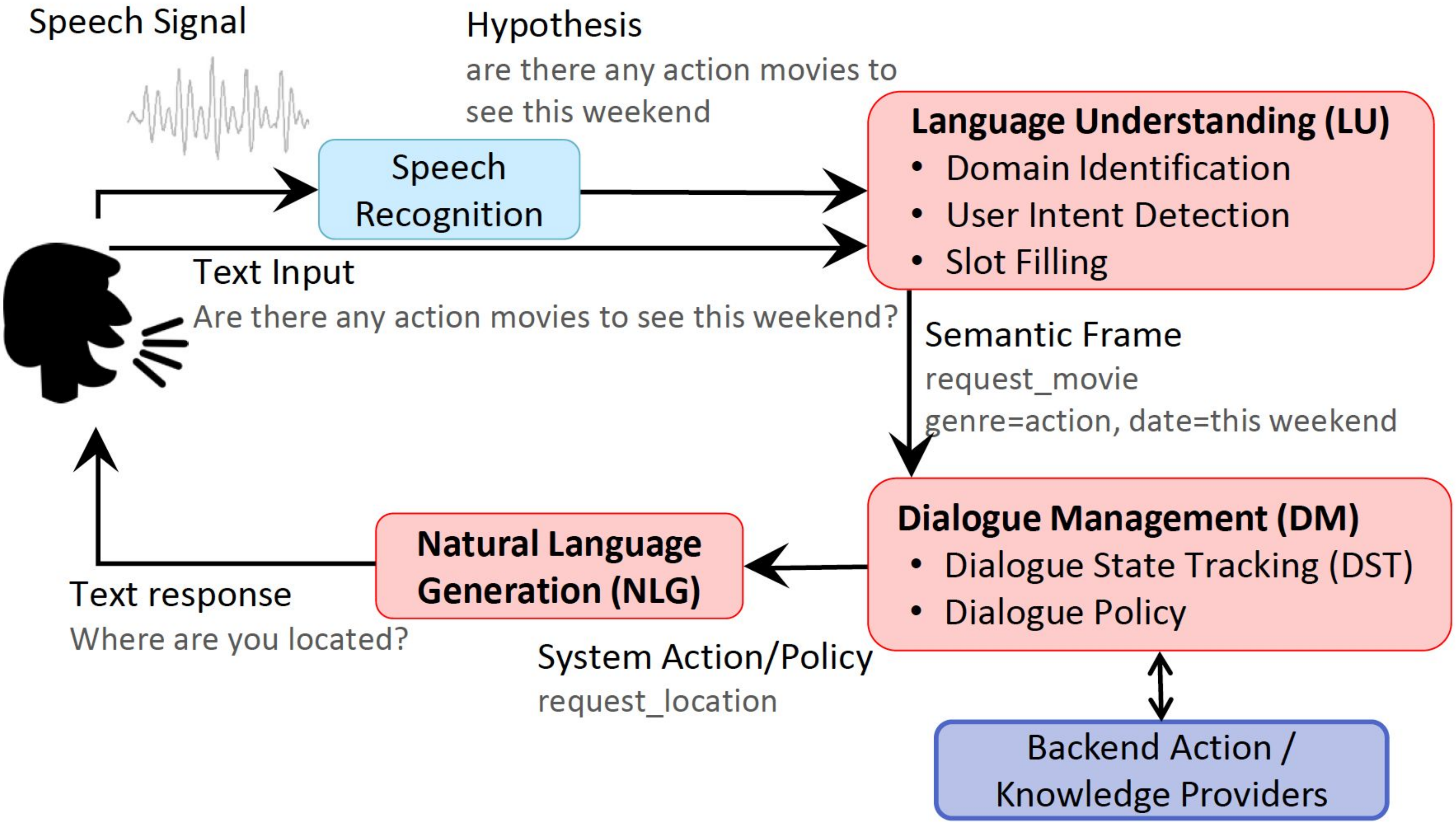
□ No specific goal, focus on natural responses

□ Using variants of seq2seq model

- A neural conversation model (Vinyals and Le, 2015)
- Reinforcement learning for dialogue generation (Li et al., 2016)
- Conversational contextual cues for response ranking (Al-Rfou et al., 2016)

(Reference: "Open-domain neural dialogue systems" slides from Dr. Yun-Nung Chen of NTU CSIE at IJCNLP 2017)

Task-oriented Dialogue System (Young, 2000)



Speech Signal

Hypothesis

are there any action movies to see this weekend

Speech Recognition

Language Understanding (LU)

- Domain Identification
- User Intent Detection
- Slot Filling

Text Input

Are there any action movies to see this weekend?

Semantic Frame

request_movie

genre=action, date=this weekend

Dialogue Management (DM)

- Dialogue State Tracking (DST)
- Dialogue Policy

Natural Language Generation (NLG)

Text response

Where are you located?

System Action/Policy

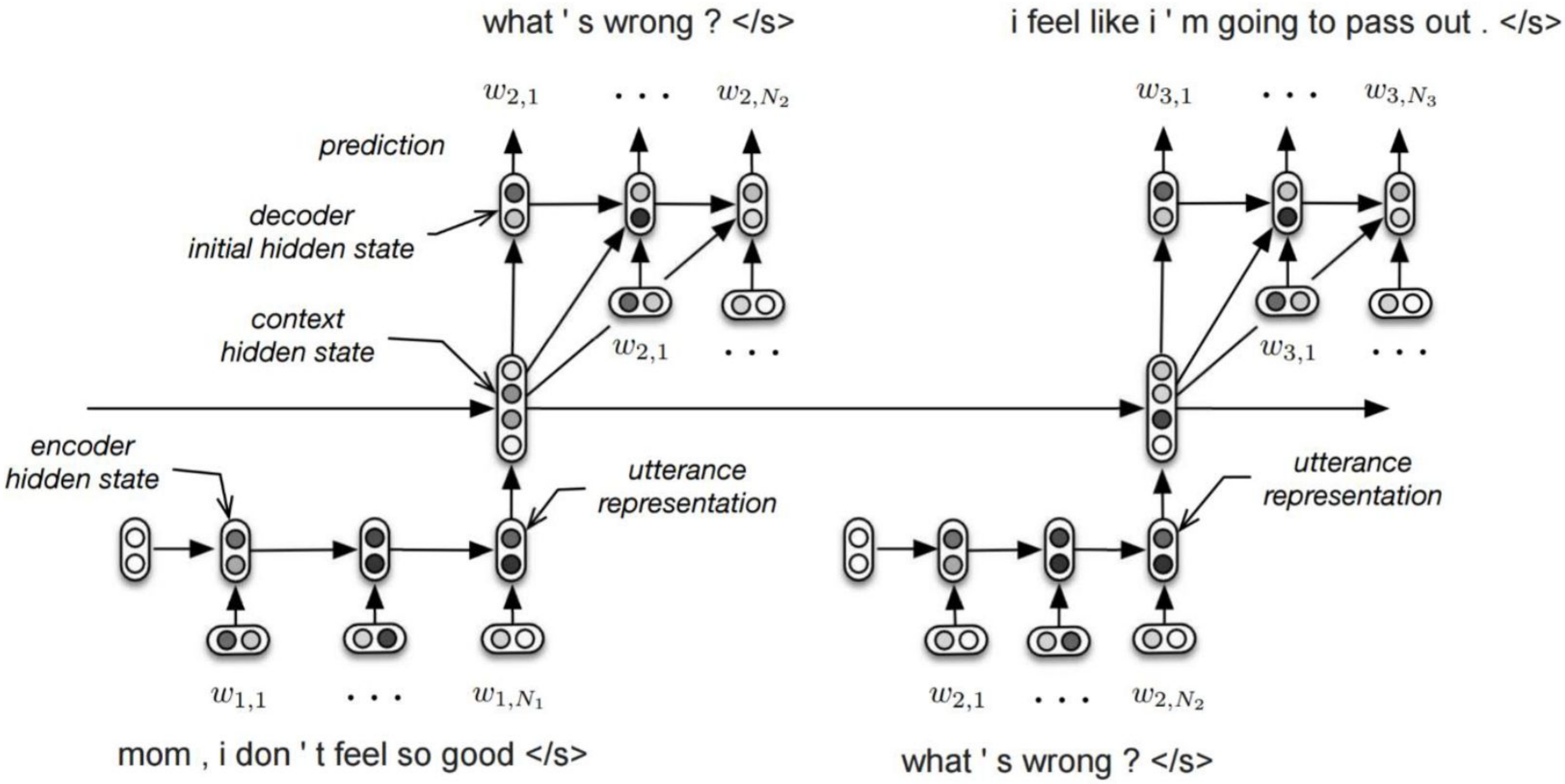
request_location

Backend Action / Knowledge Providers

(Reference: "Open-domain neural dialogue systems" slides from Dr. Yun-Nung Chen of NTU CSIE at IJCNLP 2017)

Chitchat Hierarchical Seq2Seq (Serban et al., 2016)

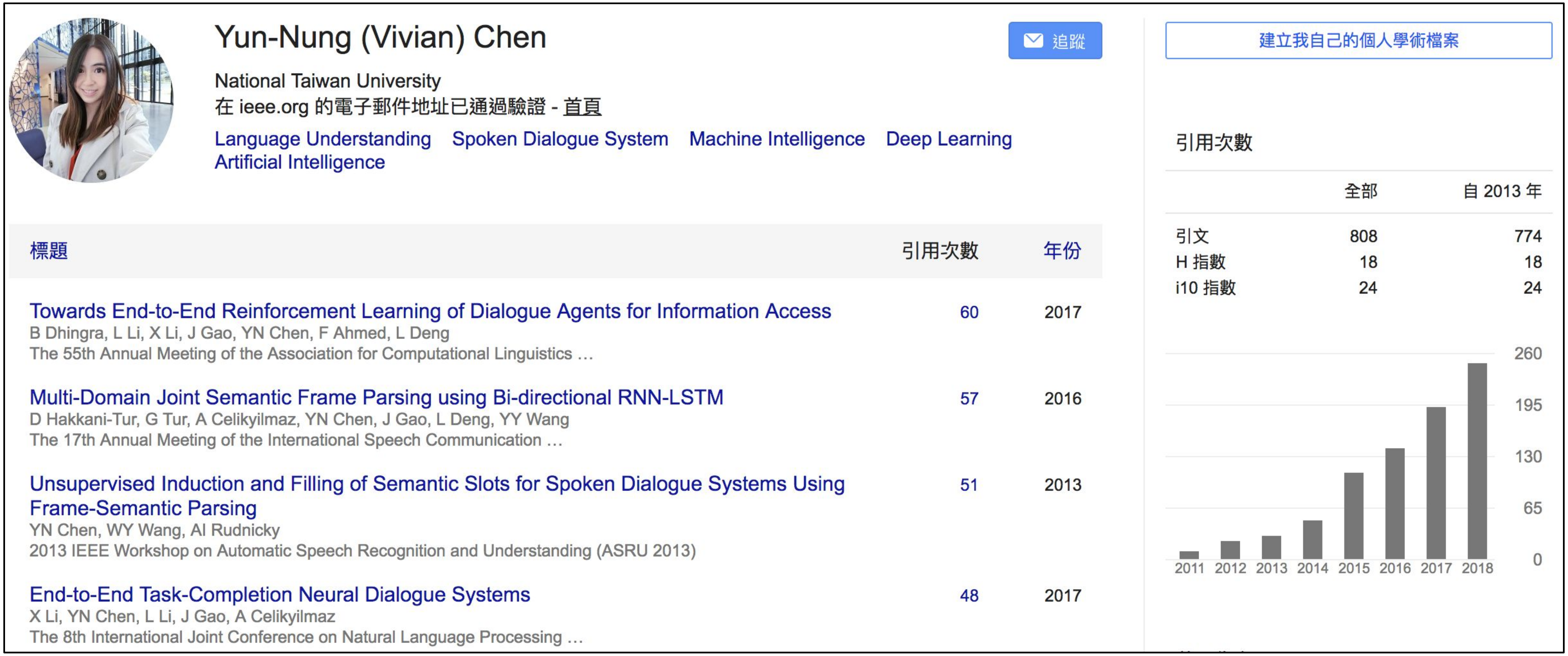
- Learns to generate dialogues from offline dialogs
- No state, action, intent, slot, etc.


what ' s wrong ? </s>
i feel like i ' m going to pass out . </s>
$w_{2,1}$ . . . w_{2,N_2}
$w_{3,1}$ . . . w_{3,N_3}
prediction
decoder initial hidden state
context hidden state
$w_{2,1}$ . . .
$w_{3,1}$ . . .
encoder hidden state
utterance representation
utterance representation
$w_{1,1}$ . . . w_{1,N_1}
$w_{2,1}$ . . . w_{2,N_2}
mom , i don ' t feel so good </s>
what ' s wrong ? </s>

(Reference: "Open-domain neural dialogue systems" slides from Dr. Yun-Nung Chen of NTU CSIE at IJCNLP 2017)

Some trivia you may (or may not) want to know

- Previous slides came from the work of Dr. Yun-Nung Chen (陳縕儂) of National Taiwan University.



Yun-Nung (Vivian) Chen

追蹤

National Taiwan University
在 ieee.org 的電子郵件地址已通過驗證 - 首頁

Language Understanding Spoken Dialogue System Machine Intelligence Deep Learning Artificial Intelligence

標題	引用次數	年份
Towards End-to-End Reinforcement Learning of Dialogue Agents for Information Access B Dhingra, L Li, X Li, J Gao, YN Chen, F Ahmed, L Deng The 55th Annual Meeting of the Association for Computational Linguistics ...	60	2017
Multi-Domain Joint Semantic Frame Parsing using Bi-directional RNN-LSTM D Hakkani-Tur, G Tur, A Celikyilmaz, YN Chen, J Gao, L Deng, YY Wang The 17th Annual Meeting of the International Speech Communication ...	57	2016
Unsupervised Induction and Filling of Semantic Slots for Spoken Dialogue Systems Using Frame-Semantic Parsing YN Chen, WY Wang, AI Rudnicky 2013 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU 2013)	51	2013
End-to-End Task-Completion Neural Dialogue Systems X Li, YN Chen, L Li, J Gao, A Celikyilmaz The 8th International Joint Conference on Natural Language Processing ...	48	2017

建立我自己的個人學術檔案

引用次數

	全部	自 2013 年
引文	808	774
H 指數	18	18
i10 指數	24	24

Any questions?