

The background features a dark blue gradient with faint, light blue concentric circles and degree markings (40, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260) on the left side, suggesting a technical or scientific theme.


多媒體系統

MULTIMEDIA SYSTEM

張家瑋 博士

助理教授

國立臺中科技大學資訊工程系



FOUNDATIONS OF NATURAL LANGUAGE PROCESSING

自然語言處理的原理與應用

自然語言處理的主要範疇

- 機器翻譯 (Machine Translation)
- 自然語言理解/語意分析 (Natural Language Understanding / Semantic Analysis)
 1. 問答系統 (Question Answering)
 2. 萃取式摘要 (Extractive Summarization)
 3. 文件分類 (Text Categorization)
- 自然語言生成 (Natural Language Generation)
 1. 進階問答系統 (Advanced Question Answering)
 2. 抽象式摘要 (Abstractive Summarization)
 3. 聊天機器人 (Chatbot)
- 語法分析 (Syntactic Parsing)
 1. 中文斷詞 (Chinese word segmentation)
 2. 詞性標註 (Part-of-speech Tagging)
 3. 實體辨識 (Named Entity Recognition)
 4. 詞彙依存 (Typed Dependencies)
 5. 文法樹 (Parse Tree)
- 語音辨識 (Speech Recognition)
- 文字轉語音 (Text to Speech)
- 語音轉文字 (Speech to Text)



機器翻譯

MACHINE TRANSLATION

GOOGLE 翻譯



翻譯 關閉即時翻譯 

英文 中文 日文 偵測語言 ▾

↔ 中文(繁體) 英文 中文(簡體) ▾ 翻譯

My dog also likes eating sausage.

33/5000

我的狗也喜歡吃香腸。

提出修改建議

Wǒ de gǒu yě xǐhuān chī xiāngcháng.

BING 翻譯

The screenshot displays the Bing Translator web interface. On the left, the source language is set to '英文 (已偵測)' (English (Detected)). The input text is 'My dog also likes eating sausage.' Below the text is a green 'G' icon and a character count of '33/5000'. On the right, the target language is set to '繁體中文' (Traditional Chinese). The translated text is '我的狗也喜歡吃香腸。'. Below the translation is the pinyin 'wǒ de gǒu yě xǐ huān chī xiāng cháng.' The interface also includes buttons for voice input, a close button, and a swap button between the two text areas.

英文 (已偵測) 繁體中文 英文 義大利文

My dog also likes eating sausage.

我的狗也喜歡吃香腸。

wǒ de gǒu yě xǐ huān chī xiāng cháng.

有道翻译

检测到：英语 » 中文

翻译

人工翻译

划词

My dog also likes eating sausage.

×

G

33/5000

我的狗也喜欢吃香肠。

☆

修改翻译结果

平行語料

Quiero ir a la playa más bonita.

I want to go to the beach more pretty.

We just replace each Spanish word with the matching English word.

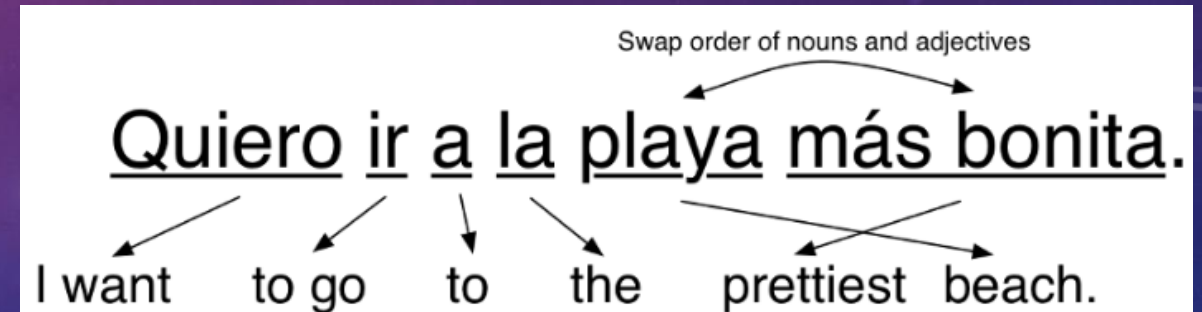


Swap order of nouns and adjectives

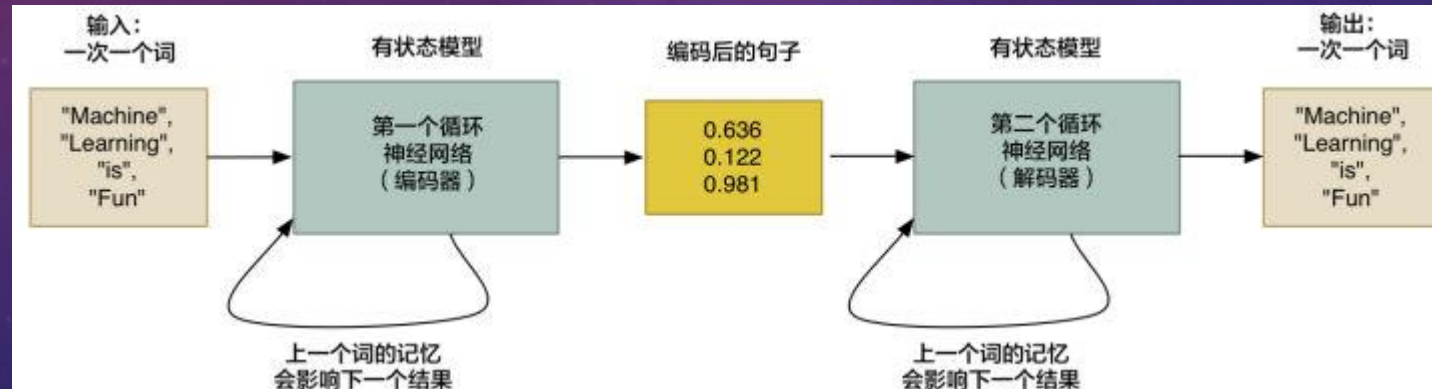
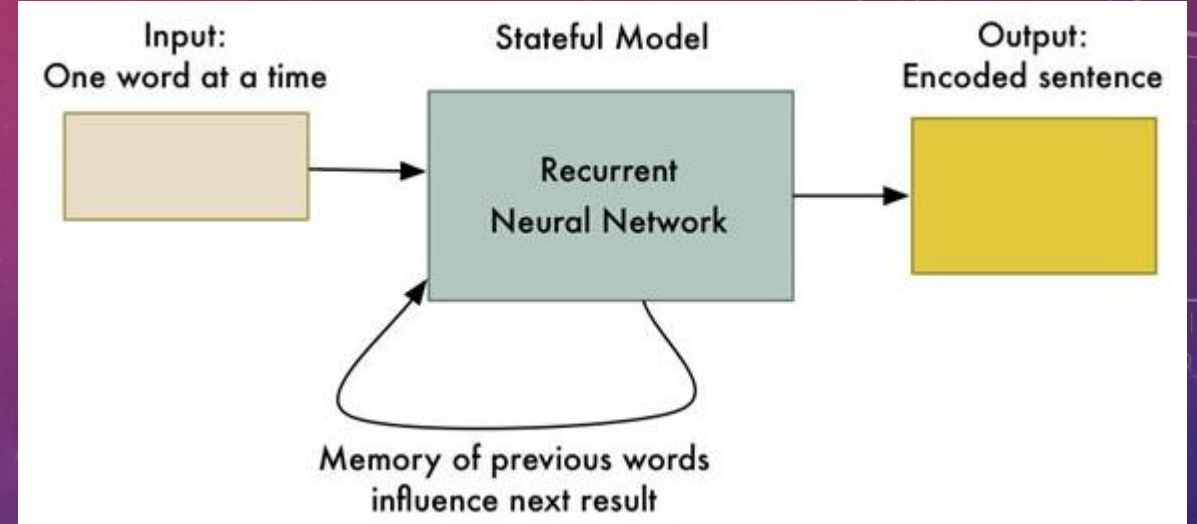
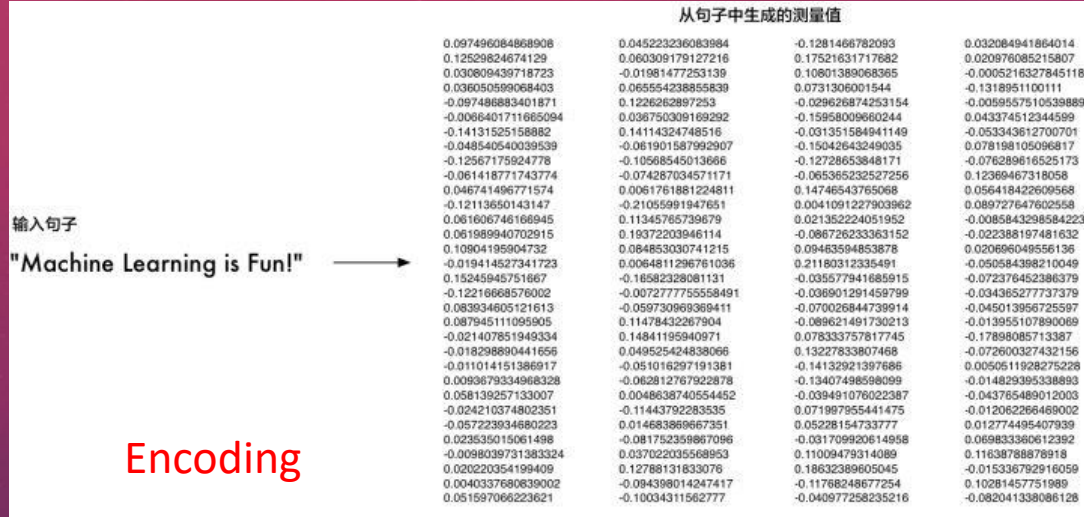
Quiero ir a la playa más bonita.

I want to go to the prettiest beach.

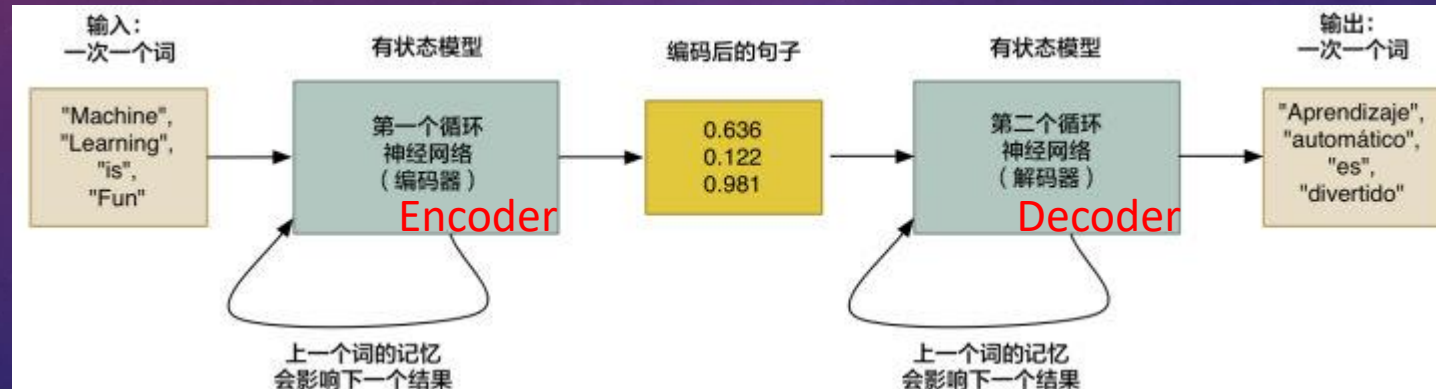
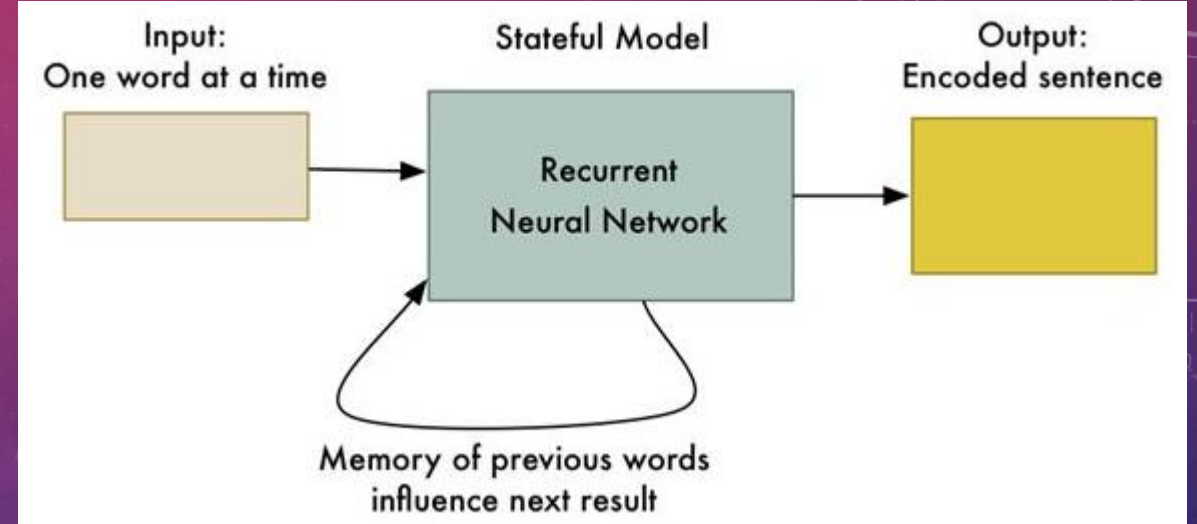
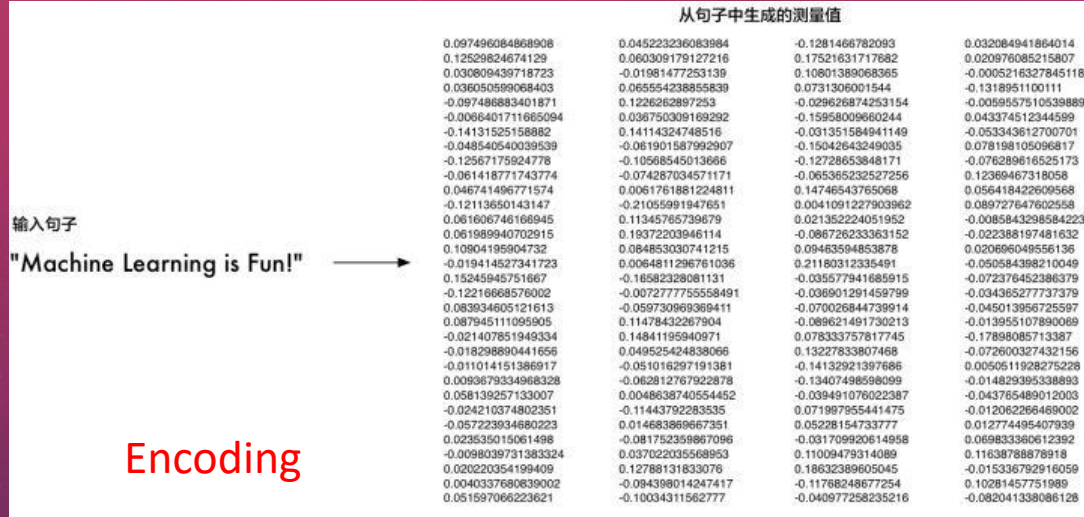
統計式機器翻譯之原理



深度學習於機器翻譯之原理



深度學習於機器翻譯之原理





自然語言理解

NATURAL LANGUAGE UNDERSTANDING

WORD-SENSE DISAMBIGUATION

- Ambiguity: a word or phrase with multiple meanings.
 1. "procure" (I will get the drinks)
 2. "become" (she got scared)
 3. "have" (I have got three dollars)
 4. "understand" (I get it)

WORDNET

WordNet Search - 3.1

- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

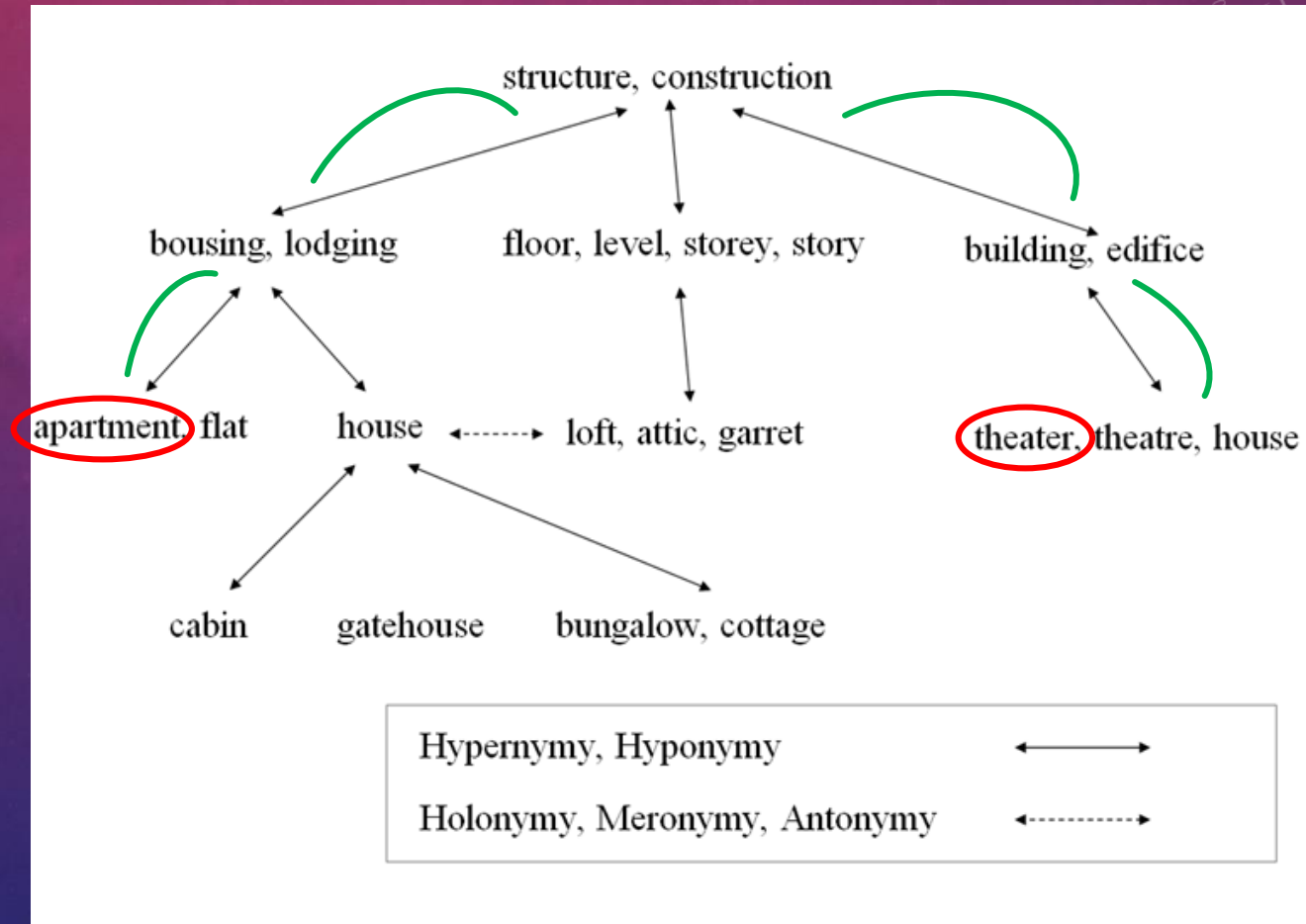
Noun

- [S:](#) [\(n\)](#) **apple** (fruit with red or yellow or green skin and sweet to tart crisp whitish flesh)
- [S:](#) [\(n\)](#) **apple**, [orchard apple tree](#), [Malus pumila](#) (native Eurasian tree widely cultivated in many varieties for its firm rounded edible fruits)

<http://wordnetweb.princeton.edu/perl/webwn>

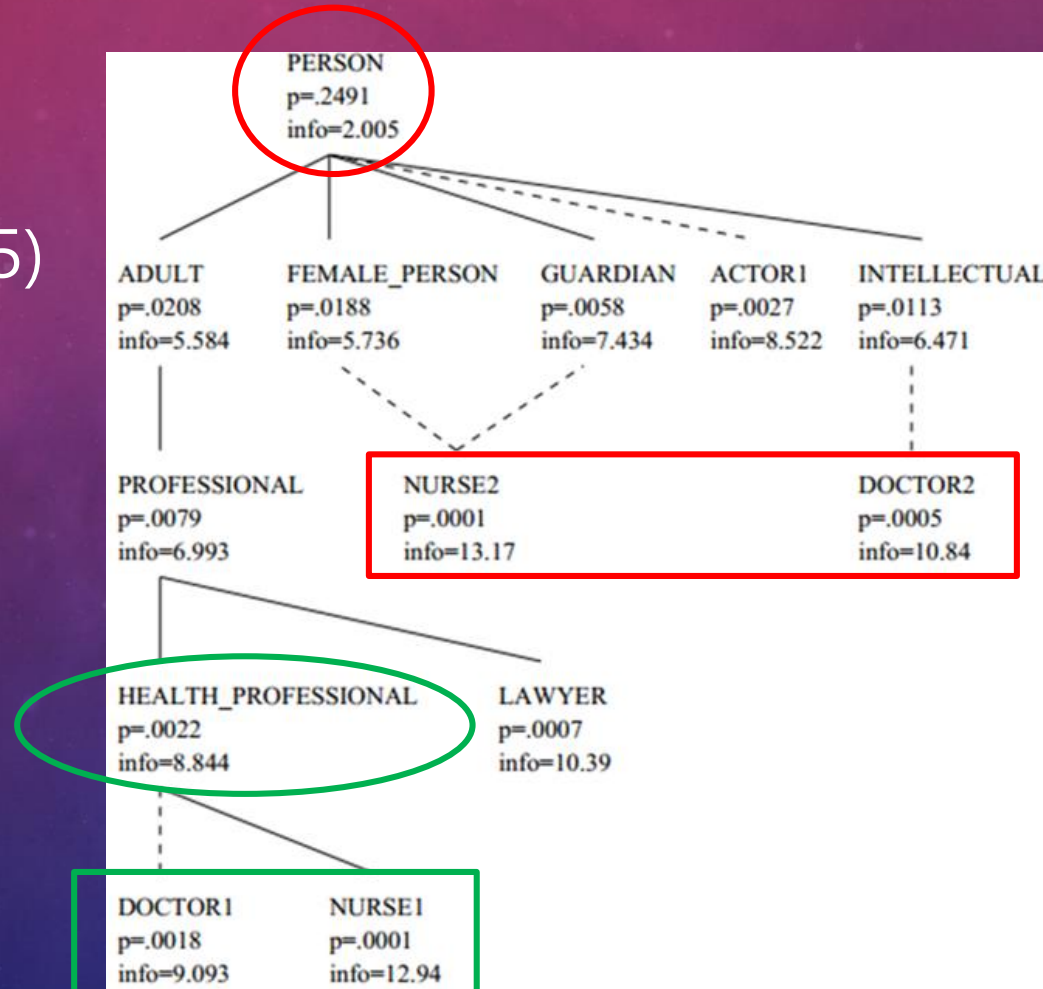
WORD-SENSE DISAMBIGUATION

- Distance-based: PATH (Rada, Mili, Bicknell, & Blettner, 1989)

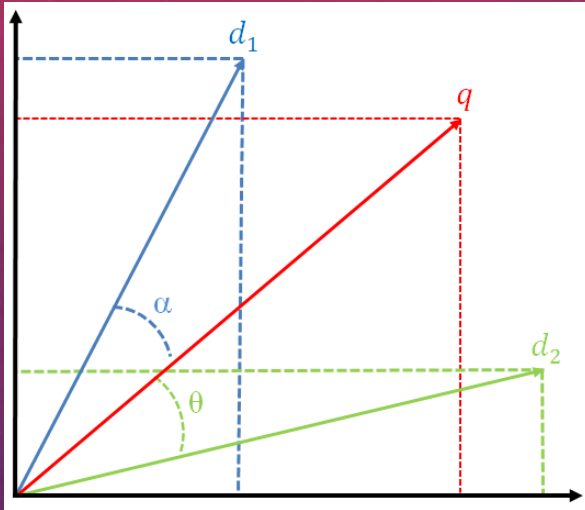


WORD-SENSE DISAMBIGUATION

- Information Content-based:
RES (Resnik, 1995)



WORD-SENSE DISAMBIGUATION



- Gloss-based: VECTOR (Patwardhan, 2003)

Cute	Cunning
1. attractive especially by means of smallness or prettiness or quaintness	1. attractive especially by means of smallness or prettiness or quaintness
2. obviously contrived to charm	2. marked by skill in deception
	3. showing inventiveness and skill

廣義知網知識本體

TopNode

- entity|事物
 - event|事件
 - object|物體 [事物，客體]
 - thing|萬物 [東西，萬物，萬有，東東]
 - physical|物質 [物質，實體，物體，物產]
 - animate|生物
 - inanimate|無生物
 - NaturalThing|天然物 [自然，大自然，造化，自然物]
 - artifact|人工物 [貨，物品，製品，成品，消費品，物件]
 - clothing|衣物 [服裝，衣服，服飾，衣物，衣褲，衣裳，衣衫]
 - edible|食物
 - medicine|藥物
 - addictive|嗜好物
 - building|建築物 [建設，建築，建築物，建物，地上物]
 - house|房屋 [房屋，房，房子，房舍，厝，屋子，屋舍]
 - facilities|設施 [設備，設施]
 - bridge|橋樑 [橋，橋樑，梁]
 - route|道路
 - StageSettings|佈景
 - BulletinBoard|看板 [看板，佈告欄，公佈欄]
 - trap|陷阱
 - counter|櫃臺 [服務台，櫃臺]
 - ThrottleValve|閘門 [閘門，壩門]
 - trough|槽
 - ASwing|鞦韆
 - console|控制台 [儀表板]
 - railings|欄杆
 - StrategicBorder|關隘
 - fence|籬笆
 - platform|台 [台，平台，臺，平臺]
 - airport|機場 [機場，航空站，航站，飛機場]
 - reservoir|水庫 [水庫，水壩，壩]
 - wharf|碼頭 [碼頭，船塢]
 - embankment|堤防 [堤防，堤，河堤，防波堤，護岸，堤岸]
 - shed|棚子
 - MilitaryCamp|軍營 [軍營，營房]
 - sentry|崗哨 [檢查站，哨]
 - church|教堂 [教堂，禮拜堂]
 - port|港口 [港，港口，口岸，港埠，港灣，埠，避風港]
 - factory|工廠 [工廠，廠，廠房，工場，製造廠，廠家，作坊]
 - farm|農場 [農場，農園，農莊]
 - station|車站
 - hospital|醫院 [醫院，醫療院，病院]
 - museum|博物館 [博物館，博物院，文物館]
 - restaurant|餐廳 [餐廳，餐館，酒樓，酒家，食堂，啤酒屋，飯館，館子，飯廳]
 - school|學校
 - college|學院 [學院]

VECTOR REPRESENTATION

	w_1	w_2	w_3	w_{n-1}	w_n	label
D_1	0.11	0.23	0	0.57	0	0
D_2	0	0	0	0.29	0.7	1
D_3	0	0.81	0.44	0	0	0
D_4	0	0.37	0	0	0.16	1
..
D_k	1

TF-IDF

- TF: term frequency:
$$\text{tf}_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$
 - IDF: inverse document frequency:
$$\text{idf}_i = \log \frac{|D|}{|\{j : t_i \in d_j\}|}$$
- where:
- $|D|$: total number of documents in the corpus
 - $|\{j : t_i \in d_j\}|$: number of documents where term t_i appears

Then:

- $$\text{tfidf}_{i,j} = \text{tf}_{i,j} \times \text{idf}_i$$

Document 1		Document 2	
Term	Term Count	Term	Term Count
this	1	this	1
is	1	is	1
a	2	another	2
sample	1	example	3

- The calculation of tf-idf for the term "this" is performed as follows:

$$\begin{aligned} \text{tf}(\text{"this"}, d_1) &= \frac{1}{5} = 0.2 \\ \text{tf}(\text{"this"}, d_2) &= \frac{1}{7} \approx 0.14 \end{aligned}$$

$$\text{idf}(\text{"this"}, D) = \log\left(\frac{2}{2}\right) = 0$$

- So tf-idf is zero for the word "this", which implies that the word is not very informative as it appears in all documents.

$$\begin{aligned} \text{tfidf}(\text{"this"}, d_1) &= 0.2 \times 0 = 0 \\ \text{tfidf}(\text{"this"}, d_2) &= 0.14 \times 0 = 0 \end{aligned}$$

Document 1		Document 2	
Term	Term Count	Term	Term Count
this	1	this	1
is	1	is	1
a	2	another	2
sample	1	example	3

- A slightly more interesting example arises from the word "example", which occurs three times only in the second document:

$$\text{tf}(\text{"example"}, d_1) = \frac{0}{5} = 0$$

$$\text{tf}(\text{"example"}, d_2) = \frac{3}{7} \approx 0.429$$

$$\text{idf}(\text{"example"}, D) = \log\left(\frac{2}{1}\right) = 0.301$$

$$\begin{aligned} \text{tfidf}(\text{"example"}, d_1) &= \text{tf}(\text{"example"}, d_1) \times \text{idf}(\text{"example"}, D) = 0 \times 0.301 = 0 \\ \text{tfidf}(\text{"example"}, d_2) &= \text{tf}(\text{"example"}, d_2) \times \text{idf}(\text{"example"}, D) = 0.429 \times 0.301 \approx 0.13 \end{aligned}$$

潛藏語意分析(LSA)

- 奇異值分解
 - Singular Value Decomposition (SVD)

Index Words	Titles								
	T1	T2	T3	T4	T5	T6	T7	T8	T9
book			1	1					
dads						1			1
dummies		1						1	
estate							1		1
guide	1					1			
investing	1	1	1	1	1	1	1	1	1
market	1		1						
real							1		1
rich						2			1
stock	1		1					1	
value				1	1				

=

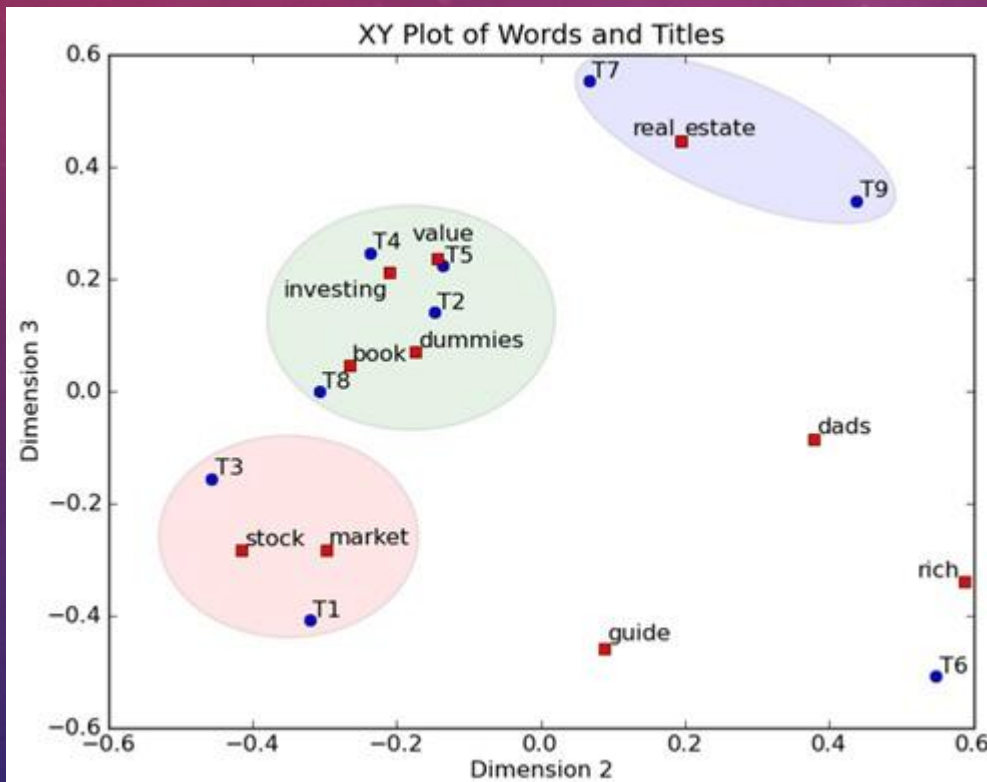
book	0.15	-0.27	0.04
dads	0.24	0.38	-0.09
dummies	0.13	-0.17	0.07
estate	0.18	0.19	0.45
guide	0.22	0.09	-0.46
investing	0.74	-0.21	0.21
market	0.18	-0.30	-0.28
real	0.18	0.19	0.45
rich	0.36	0.59	-0.34
stock	0.25	-0.42	-0.28
value	0.12	-0.14	0.23

3.91	0	0
0	2.61	0
0	0	2.00

T1	T2	T3	T4	T5	T6	T7	T8	T9
0.35	0.22	0.34	0.26	0.22	0.49	0.28	0.29	0.44
-0.32	-0.15	-0.46	-0.24	-0.14	0.55	0.07	-0.31	0.44
-0.41	0.14	-0.16	0.25	0.22	-0.51	0.55	0.00	0.34

潛藏語意分析(LSA)

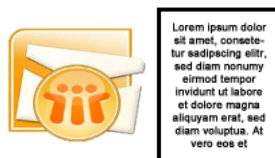
- 文件分類/主題探勘
- 語意分析



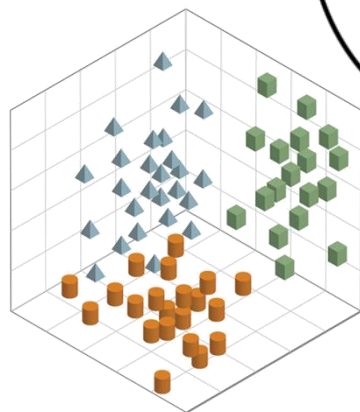
Index Words	Titles								
	T1	T2	T3	T4	T5	T6	T7	T8	T9
book			1	1					
dads						1			1
dummies		1						1	
estate							1		1
guide	1					1			
investing	1	1	1	1	1	1	1	1	1
market	1		1						
real							1		1
rich						2			1
stock	1		1					1	
value				1	1				

文字檔案

Input:
one document



word
vectors



word2vec

將被拆解成多個字元

Model:



vector space

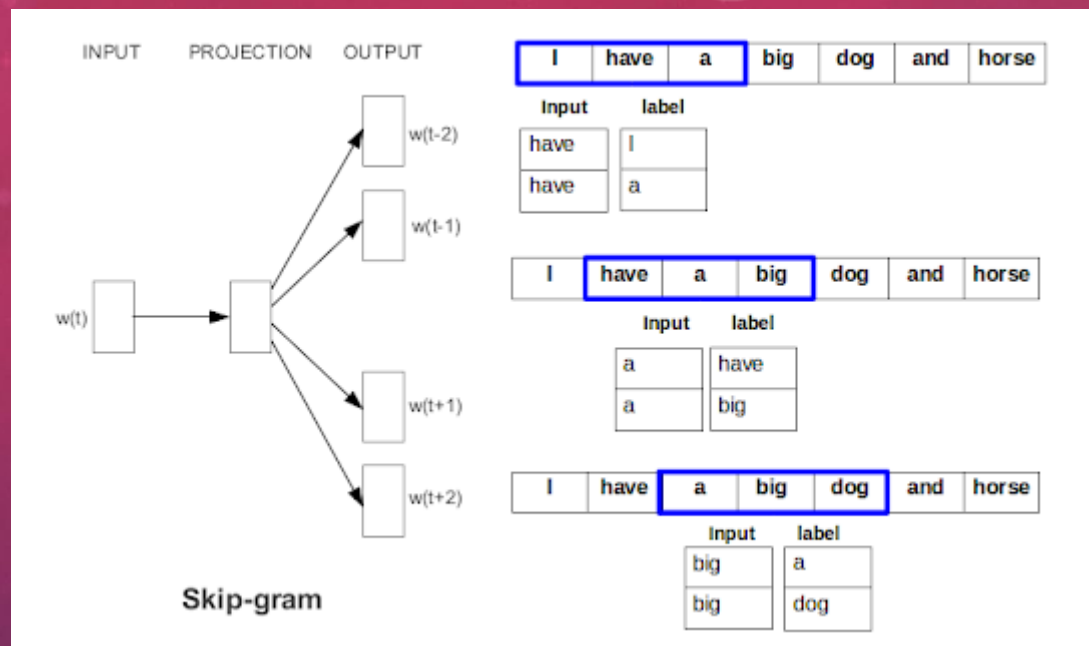
解析成多元維度的向量

透過向量比對
找出相似的資料

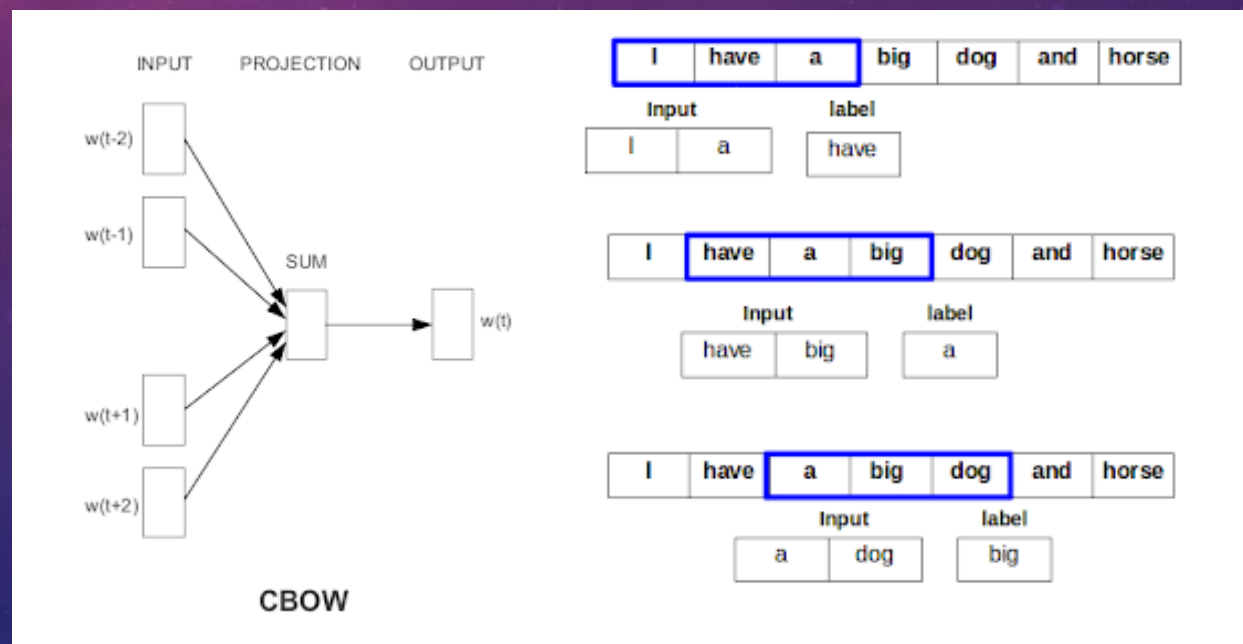
most_similar('france'):

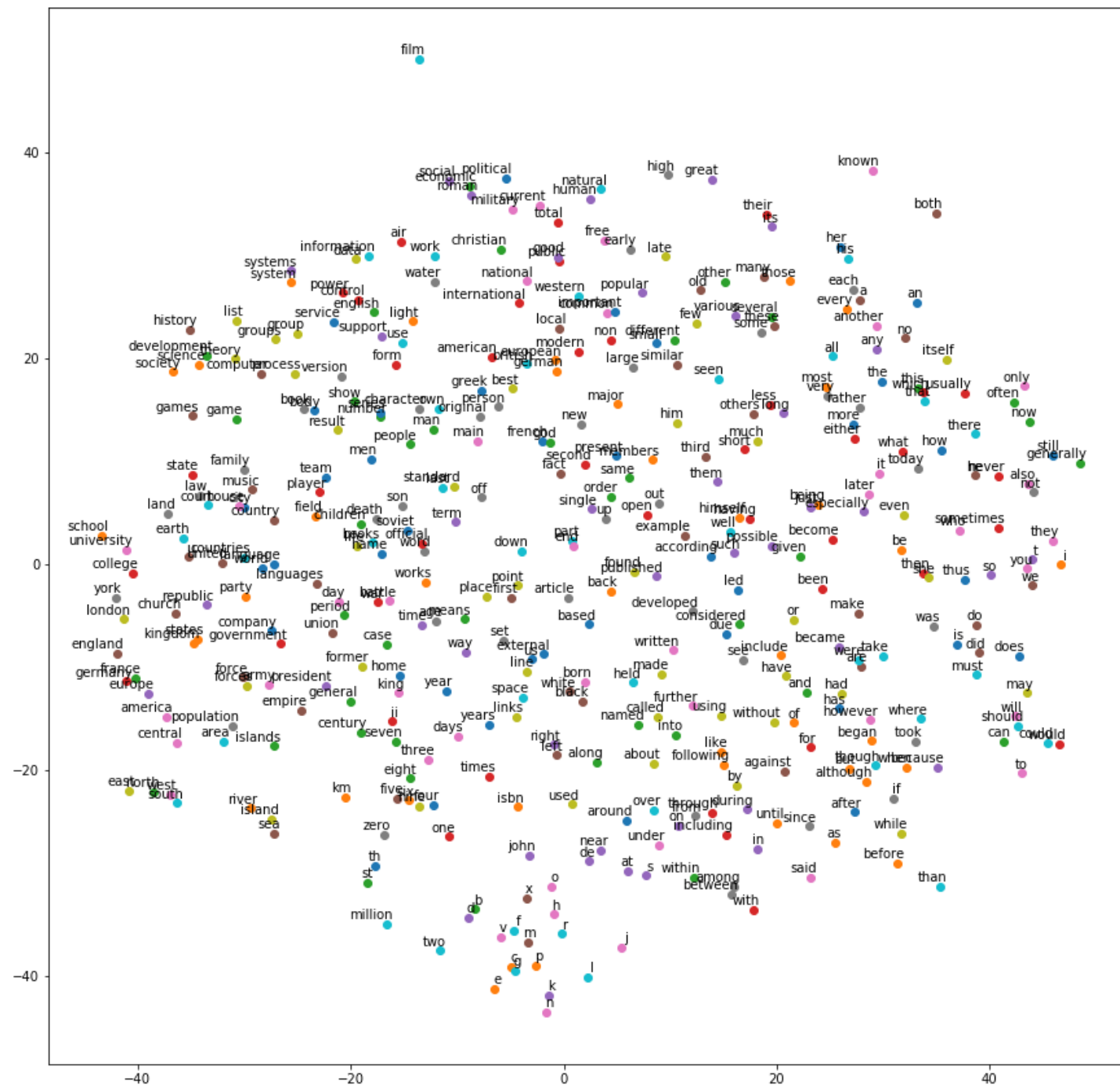
spain	0.678515
belgium	0.665923
netherlands	0.652428
italy	0.633130

highest cosine
distance values
in vector space
of the nearest
words



<http://zongsoftwarenate.blogspot.com/2017/04/word2vec-model-introduction-skip-gram.html>





語法分析

SYNTACTIC PARSING

STANFORD PARSER

Stanford Parser

Please enter a sentence to be parsed:

My dog also likes eating sausage.

Language: English Sample Sentence

Parse

Your query

My dog also likes eating sausage.

Tagging

My/PRP\$ dog/NN also/RB likes/VBZ eating/VBG sausage/NN ./.

Parse

```
(ROOT
 (S
  (NP (PRP$ My) (NN dog))
  (ADVP (RB also))
  (VP (VBZ likes)
   (S
    (VP (VBG eating)
     (NP (NN sausage))))
   (. .)))
```

Universal dependencies

```
nmod:poss(dog-2, My-1)
nsubj(likes-4, dog-2)
advmod(likes-4, also-3)
root(ROOT-0, likes-4)
xcomp(likes-4, eating-5)
dobj(eating-5, sausage-6)
```



The Stanford Natural Language Processing Group

[people](#) [publications](#) [research blog](#) [software](#) [teaching](#) [local](#)

Software > Stanford Parser

The Stanford Parser: A statistical parser

[About](#) | [Citing](#) | [Questions](#) | [Download](#) | [Included Tools](#) | [Extensions](#) | [Release history](#) | [Sample output](#) | [Online](#) | [FAQ](#)

About

A natural language parser is a program that works out the grammatical **structure of sentences**, for instance, which groups of words go together (as "phrases") and which words are the **subject** or **object** of a verb. Probabilistic parsers use knowledge of language gained from hand-parsed sentences to try to produce the *most likely* analysis of new sentences. These statistical parsers still make some mistakes, but commonly work rather well. Their development was one of the biggest breakthroughs in natural language processing in the 1990s. You can try out our parser online.

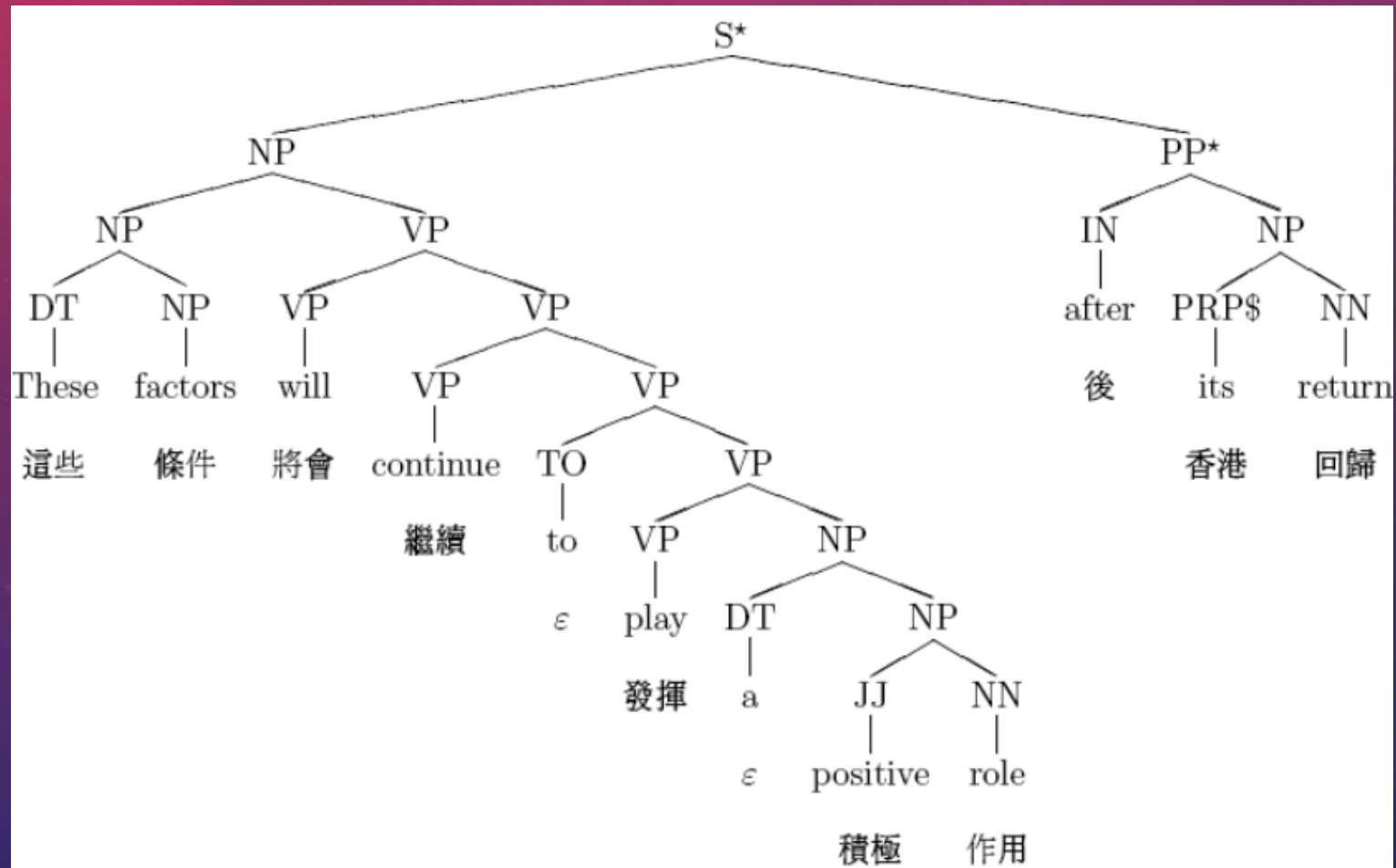
Package contents

This package is a Java implementation of probabilistic natural language parsers, both highly optimized PCFG and lexicalized dependency parsers, and a lexicalized PCFG parser. The original version of this parser was mainly written by Dan Klein, with support code and linguistic grammar development by Christopher Manning. Extensive additional work (internationalization and language-specific modeling, flexible input/output, grammar compaction, lattice parsing, *k*-best parsing, typed dependencies output, user support, etc.) has been done by Roger Levy, Christopher Manning, Teg Grenager, Galen Andrew, Marie-Catherine de Marneffe, Bill MacCartney, Anna Rafferty, Spence Green, Huihsin Tseng, Pi-Chuan Chang, Wolfgang Maier, and Jenny Finkel.

The lexicalized probabilistic parser implements a factored product model, with separate PCFG phrase structure and lexical dependency experts, whose preferences are combined by efficient exact inference, using an A* algorithm. Or the software can be used simply as an accurate unlexicalized stochastic context-free grammar parser. Either of these yields a good performance statistical parsing system. A GUI is provided for viewing the phrase structure tree output of the parser.

- <http://nlp.stanford.edu:8080/parser/>

PARSE TREE



中研院

傅達仁(Nb) 今(Nd) 將(D) 執行(VC) 安樂死(Na) , (COMMACATEGORY)

卻(D) 突然(D) 爆出(VJ) 自己(Nh) 20(Neu) 年前(Nd) 遭(P) 緯來(Nb) 體育台(Nc) 封殺(VC) , (COMMACATEGORY)

他(Nh) 不(D) 懂(VK) 自己(Nh) 哪裡(D) 得罪(VC) 到(P) 電視台(Nc) 。(PERIODCATEGORY)

實體辨識

傅達仁.PERSON今將執行安樂死，卻突然爆出自己20年前遭緯來體育台封殺，他不懂自己哪裡得罪到電視台。

指代消解

傅達仁今將執行安樂死，**NULL**傅達仁卻突然爆出自己20年前遭緯來體育台封殺，**他**傅達仁不懂自己哪裡得罪到電視台。

☒ show all ☐ show phrase head ☐ show word head

agent(執行_VC2)=傅達仁_NP
time(執行_VC2)=今_Ndabd
time(執行_VC2)=將_Dd
goal(執行_VC2)=安樂死_NP
theme(爆出_VJ3)=傅達仁_Nba
evaluation(爆出_VJ3)=卻_Dbb
time(爆出_VJ3)=突然_Dd
range(爆出_VJ3)=自己20年前遭緯來_NP
complement(爆出_VJ3)=來體育台封殺_VP
experiencer(懂_VK1)=傅達仁_NP
negation(懂_VK1)=不_Dc
goal(懂_VK1)=自己哪裡得罪到電視台_S

agent(執行_VC2)=theme(爆出_VJ3), 1
agent(執行_VC2)=experiencer(懂_VK1), 1
theme(爆出_VJ3)=experiencer(懂_VK1), 1

FUDANNLP

```
D:\fnlp>java -Xmx1024m -Dfile.encoding=UTF-8 -classpath "fnlp-core/target/fnlp-core-2.1-SNAPSHOT.jar;libs/trove-3.1a1.jar;libs/commons-cli-1.4.jar" org.fnlp.nlp.cn.tag.POSTagger -s models/seg.m models/pos.m "周杰伦出生于台湾，生日为79年1月18日，他曾经的绯闻女友是蔡依林。"
```

周杰伦/人名 出生于/动词 台湾/地名， /名词 生日/名词 为/介词 79年/时间短语 1月/时间短语 18日/时间短语， /动词 他/人称代词 曾经/形容词 的/结构助词 绯闻/名词 女友/名词 是/动词 蔡依林/人名 。/标点

http://blog.csdn.net/hhu_lyc

```
D:\fnlp>java -Xmx1024m -Dfile.encoding=UTF-8 -classpath "fnlp-core/target/fnlp-core-2.1-SNAPSHOT.jar;libs/trove-3.1a1.jar;libs/commons-cli-1.4.jar" org.fnlp.nlp.cn.tag.NERTagger -s models/seg.m models/pos.m "詹姆斯·默多克和丽贝卡·布鲁克斯鲁珀特·默多克旗下的美国小报《纽约邮报》的职员被公司律师告知，保存任何也许与电话窃听及贿赂有关的文件。"
```

<美国=地名, 纽约=地名, 詹姆斯·默多克=人名, 鲁珀特·默多克=人名, 丽贝卡·布鲁克斯=人名>

http://blog.csdn.net/hhu_lyc

復旦NLP - 簡體中文剖析工具

https://blog.csdn.net/hhu_lyc/article/details/79179619

以語言學習輔助工具為例

Collocation online suggestion v1.0
英語搭配詞線上檢索系統

[介紹](#) [常用搭配詞查詢](#) [整句搭配詞查詢與推薦](#)

整句搭配詞查詢與推薦

輸入句子：

清除

送出

輸入的句子為
We commonly use a small cell for medical research.

副詞修飾(V/Adv/Adj組合)
commonly + V/Adv/Adj

#	collocation	freq(%)
1	commonly use	46.5
2	commonly used	4.7
3	commonly find	4.4
4	commonly know	3.3
5	commonly employ	2.4
6	commonly refer	2.2
7	commonly observe	1.9
8	commonly report	1.9
9	commonly encounter	1.4
10	commonly available	1.3

commonly與use的搭配字同義組合
commonly + 搭配同義字

#	collocation	freq(%)	
1	commonly use	46.5	✎
2	commonly employ	2.4	✎
3	commonly apply	0.5	✎

同義詞搭配詞組搜尋結果
commonly的同義字 + use 的同義字

#	collocation	count	
1	commonly use	296	✎
2	often use	140	✎
3	frequently use	68	✎
4	commonly employ	15	✎
5	frequently employ	9	✎
6	often employ	6	✎
7	frequently apply	5	✎
8	repeatedly use	5	✎
9	routinely use	5	✎
10	frequently utilize	4	✎
11	routinely employ	3	✎
12	commonly apply	3	✎

查詢總時間:0.52sec

以語言學習輔助工具為例

	computer	data	pinch	result	sugar
aprocot	0	0	1	0	1
pineapple	0	0	1	0	1
digital	2	1	0	1	0
information	1	6	0	4	0

$$P(x = \text{information}, y = \text{data}) = \frac{6}{19} = 0.32$$

$$P(x = \text{information}) = \frac{6 + 4 + 1}{19} = \frac{11}{19} = 0.58$$

$$P(y = \text{data}) = \frac{6 + 1}{19} = \frac{7}{19} = 0.37$$

$$\begin{aligned} & \text{pmi}(x = \text{information}, y = \text{data}) \\ &= \log \frac{P(x = \text{information}, y = \text{data})}{P(x = \text{information}) \times P(y = \text{data})} \\ &= \log 1.49 \\ &= 0.57 \end{aligned}$$

The background is a gradient from deep red at the top to dark blue at the bottom, speckled with white dots resembling stars. Overlaid on the left side are several concentric circular patterns. One large circle has a scale from 140 to 260 in increments of 10, with tick marks. Other circles are smaller and some have dashed lines or arrows, suggesting a sense of motion or cycles.

THINKING TIME

POTENTIAL APPLICATIONS

The background is a gradient from deep red at the top to dark blue at the bottom, speckled with white dots resembling stars. Overlaid on the left side are several concentric circular patterns and arcs in a lighter red/pink color. Some of these arcs have small arrowheads pointing in different directions, suggesting a sense of rotation or movement. The overall aesthetic is futuristic and data-oriented.

OPEN DATASETS

UC IRVINE MACHINE LEARNING REPOSITORY

- <http://archive.ics.uci.edu/ml/datasets.html>

<div>UCI Machine Learning Repository Center for Machine Learning and Intelligent Systems</div> <div>About Citation Policy Donate a Data Set Contact</div> <div>Search</div>							
Browse Through: 426 Data Sets							
<div>Default Task</div> <div>Classification (314)</div> <div>Regression (82)</div> <div>Clustering (72)</div> <div>Other (54)</div> <div>Attribute Type</div> <div>Categorical (37)</div> <div>Numerical (273)</div> <div>Mixed (55)</div> <div>Data Type</div> <div>Multivariate (324)</div> <div>Univariate (19)</div> <div>Sequential (44)</div> <div>Time-Series (79)</div> <div>Text (44)</div> <div>Domain-Theory (23)</div> <div>Other (21)</div> <div>Area</div> <div>Life Sciences (98)</div> <div>Physical Sciences (47)</div> <div>CS/Engineering (148)</div> <div>Social Sciences (24)</div> <div>Business (26)</div> <div>Game (10)</div> <div>Other (69)</div> <div># Attributes</div> <div>Less than 10 (99)</div> <div>10 to 100 (195)</div> <div>Greater than 100 (76)</div> <div># Instances</div> <div>Less than 100 (25)</div> <div>100 to 1000 (149)</div> <div>Greater than 1000 (220)</div> <div>Format Type</div> <div>Matrix (292)</div> <div>Non-Matrix (134)</div>	Name	Data Types	Default Task	Attribute Types	# Instances	# Attributes	Year
	 Abalone	Multivariate	Classification	Categorical, Integer, Real	4177	8	1995
	 Adult	Multivariate	Classification	Categorical, Integer	48842	14	1996
	 Annealing	Multivariate	Classification	Categorical, Integer, Real	798	38	
	 Anonymous Microsoft Web Data		Recommender-Systems	Categorical	37711	294	1998
	 Arrhythmia	Multivariate	Classification	Categorical, Integer, Real	452	279	1998
	 Artificial Characters	Multivariate	Classification	Categorical, Integer, Real	6000	7	1992
	 Audiology (Original)	Multivariate	Classification	Categorical	226		1987
	 Audiology (Standardized)	Multivariate	Classification	Categorical	226	69	1992
	 Auto MPG	Multivariate	Regression	Categorical, Real	398	8	1993
	 Automobile	Multivariate	Regression	Categorical, Integer, Real	205	26	1987
	 Badges	Univariate, Text	Classification		294	1	1994

任務

資料型態







資料筆數

KAGGLE DATASETS

- <https://www.kaggle.com/datasets>

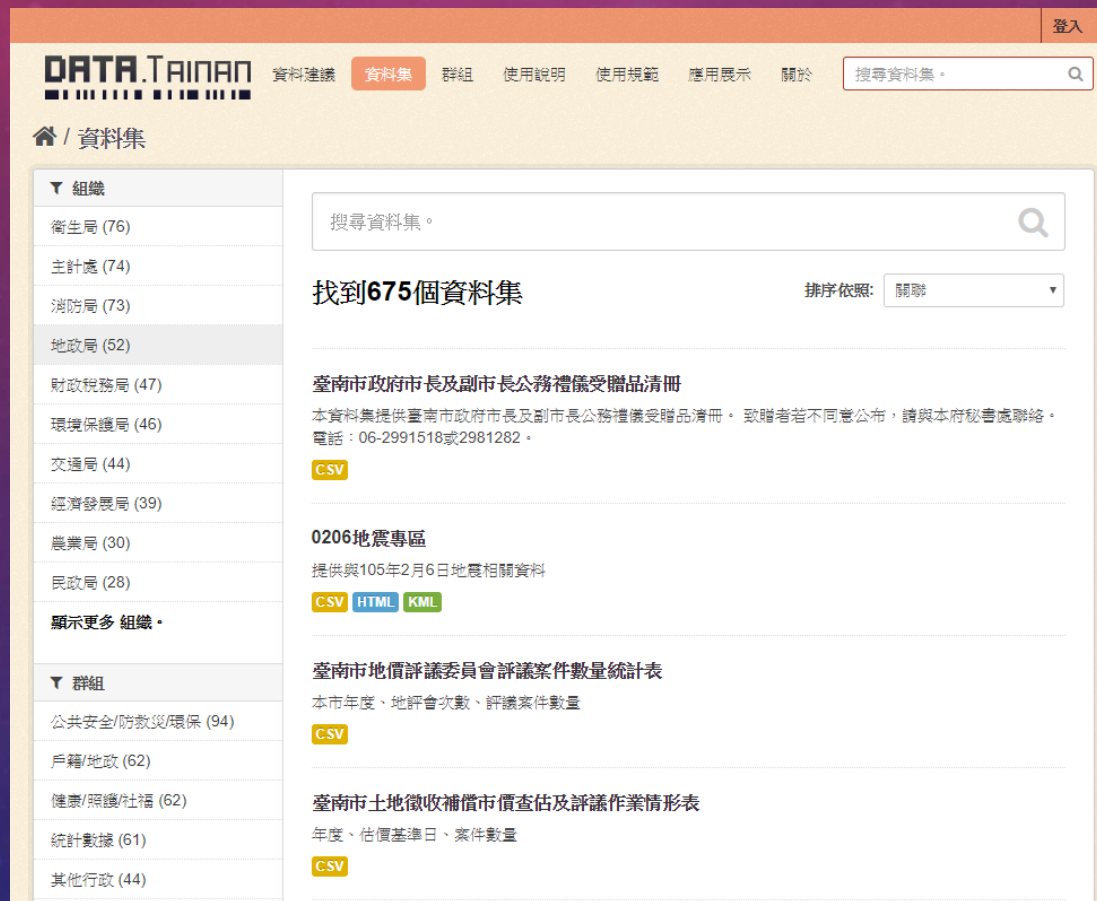
Public Sort by Hotness

12,800 Datasets Sizes File types Licenses Tags

373		Data Science for Good: Kiva Crowdfunding Use Kernels to assess welfare of Kiva borrowers for \$30k in prizes Kiva updated 10 days ago	geography finance lending + 2 more...	CSV 42 MB CC0	106 27 45k
295		Trending YouTube Video Statistics Daily statistics for trending YouTube videos Mitchell J updated 12 hours ago	languages popular culture statistics + 2 more...	CSV 64 MB CC0	27 10 51k
254		Huge Stock Market Dataset Historical daily prices and volumes of all U.S. stocks and ETFs Boris Marjanovic updated 4 months ago	business finance economics artificial intelligen...	Other 245 MB CC0	12 7 28k
154		Bitcoin Blockchain Complete live historical Bitcoin blockchain data Google BigQuery updated a month ago	finance money internet bigquery	BigQuery 821 GB CC0	430 6 31k
70		Historical Air Quality Air Quality Data Collected at Outdoor Monitors Across the US US Environmental Protection Agency updated 3 months ago	pollution bigquery	BigQuery 323 GB CC0	16 1 17k
63		HackerRank Developer Survey 2018 Survey of 25,000 professionals and students on the state of developer skills HackerRank updated 12 days ago	women demographics programming + 2 more...	CSV 5 MB CC4	17 1 6k

臺南市開放資料

- <http://data.tainan.gov.tw/dataset>



The screenshot displays the Tainan Open Data website interface. At the top, there is a navigation bar with the 'DATA.TAINAN' logo, a search bar, and links for '資料建議' (Data Suggestion), '資料集' (Dataset), '群組' (Group), '使用說明' (Usage Guide), '使用規範' (Usage Guidelines), '應用展示' (Application Showcase), and '關於' (About). A '登入' (Login) button is located in the top right corner.

The main content area is titled '資料集' (Dataset). On the left, there is a sidebar with a tree view of categories: '組織' (Organization) and '群組' (Group). Under '組織', various government departments are listed with their respective dataset counts, such as '衛生局 (76)', '主計處 (74)', '消防局 (73)', '地政局 (52)', '財政稅務局 (47)', '環境保護局 (46)', '交通局 (44)', '經濟發展局 (39)', '農業局 (30)', '民政局 (28)', and a link to '顯示更多 組織'. Under '群組', categories like '公共安全/防救災/環保 (94)', '戶籍/地政 (62)', '健康/照護/社福 (62)', '統計數據 (61)', and '其他行政 (44)' are listed.

The main content area features a search bar with the placeholder text '搜尋資料集。' and a magnifying glass icon. Below the search bar, it states '找到675個資料集' (Found 675 datasets) and includes a '排序依照:' (Sort by) dropdown menu set to '關聯' (Relevance).

Three dataset entries are visible:

- 臺南市政府市長及副市長公務禮儀受贈品清冊**
本資料集提供臺南市政府市長及副市長公務禮儀受贈品清冊。致贈者若不同意公布，請與本府秘書處聯絡。電話：06-2991518或2981282。
Format: CSV
- 0206地震專區**
提供與105年2月6日地震相關資料
Formats: CSV, HTML, KML
- 臺南市地價評議委員會評議案件數量統計表**
本市年度、地評會次數、評議案件數量
Format: CSV

At the bottom, another entry is partially visible:

- 臺南市土地徵收補償市價查估及評議作業情形表**
年度、估價基準日、案件數量
Format: CSV



THANK YOU