



# Actionable Social Media Insights with Big Data Analytics

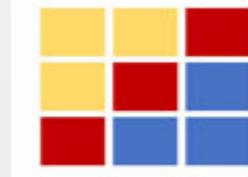
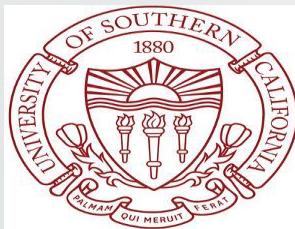
Ken Liu

2022-Oct

*(Content is reserved to be used by this course only)*

# My Background

- Study Computer Science
- iBanks Algo trade technologist
- Startup on Quant Trading, Big Data and AI



LinkedIn: <https://www.linkedin.com/in/ken-liu-76b11575/>

# My Work



你們聽過 [大數定律] 嗎？

*"Have you heard of Law of Big Numbers"*

# Course Outline – Day 1

## Day 1

- Case Studies: Social listening for company/product brandings, KOL and micro influencer identification, Machine Learning algorithm for product cross selling and recommendation
- Social Media data collection, pre-processing and visualization
- Special handling for Chinese and Cantonese languages
- Massive dataset organization such as indexing, keywords searching and ranking for relevance

# Intro to Social Media

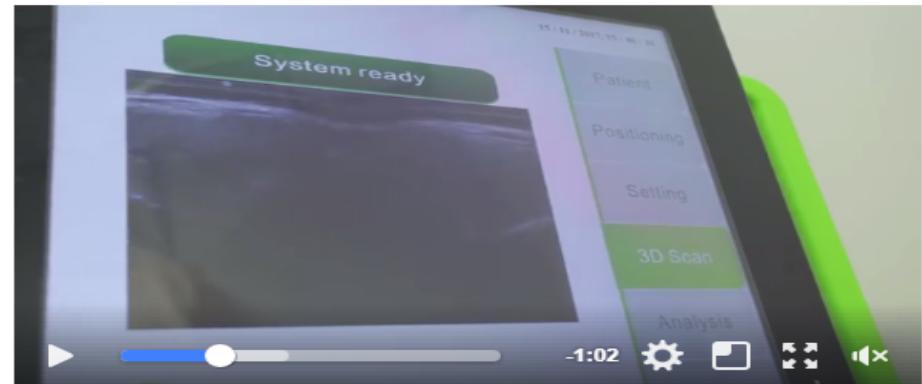
- Social Media Big Data Analytics
  - Fan Pages (*1K +*)
  - Posts (*2.7M +*)
  - Comments (*35M +*)
  - Reactions - Like, Love, Haha, Angry, Wow, Sad (*50M +*)
  - Shares (*5M +*)
- Co-founder's Project



The Hong Kong Polytechnic University  
Sponsored · 0

【理大八十周年校慶開放日】想知自己和家人健康嗎? #理大 在醫療方面的科研一直為人類的健康福祉帶來影響。開放日將展示多個醫療科研項目的成果,並有健康講座、免費健康檢查評估,千萬不要錯過!另外,歡迎大小朋友參加STEAM工作坊,由理大支持的初創企業主辦,邊玩邊學習科學知識!

最佳留言比賽:  
立即瀏覽開放日網頁 [www.polyu.edu.hk/tc/openday](http://www.polyu.edu.hk/tc/openday), 超過80個精彩活動讓大家參加。留言分享你最想參加的一個講座或工作坊及原因。我們將會選出兩個最佳留言,得獎者將分別獲得唯港薈港幣300元餐飲禮券! 2017年11月30日(中午12:00)截止。[... See more](#)



Yuen Peter and 588 others    64 Comments 92 Shares 54k Views

[Like](#) [Comment](#) [Share](#)

Chronological ▾

[View previous comments](#)

Tom Tong 我想參加的講座是“近視成因你要知”，因為我有家族遺傳性的深近視，怕深近視會引致視網膜退化或破裂，更會有青光眼等情況，這個講座會使我了解怎樣控制近視及其科技的最新發展。

# Social Listening

- Public images of Cathay Pacific
- 2017/07/12

Dataact 數據富 Browse Social Media Content									
Browse Content		Correlations		Fan Pages		Trend	Sentiment	Comparison	Admin
<b>Total Hits: 77</b>									
Date	Page	Like	User	Name	Pic	Post	Comment		
2017-06-15 17:27:52	立場新聞	17	<a href="#">10155132993124744</a>	Ten Boom		<a href="#">Post</a>	賣了給中資的，鮮有管理得宜。國泰可算是墮落的例子。		
2017-06-09 10:47:09	蘋果香港新聞	0	<a href="#">1211113258957780</a>	Samuel Chiu Wing Hung		<a href="#">Post</a>	【管理不善 國泰CEO調職】 <a href="http://orientaldaily.on.cc/cnt/news/20170413/00176_014.html">http://orientaldaily.on.cc/cnt/news/20170413/00176_014.html</a>		
2017-06-09 09:45:26	蘋果香港新聞	17	<a href="#">1546327955386411</a>	Seraph Wong		<a href="#">Post</a>	因管理不善而在全球最佳航空公司排名榜中節節下跌的國泰航空		
2017-06-08 12:53:35	蘋果日報	0	<a href="#">1211113258957780</a>	Samuel Chiu Wing Hung		<a href="#">Post</a>	管理層做錯低層承擔，中國人特色... 【管理不善 國泰CEO調職】 <a href="http://orientaldaily.on.cc/cnt/news/20170413/00176_014.html">http://orientaldaily.on.cc/cnt/news/20170413/00176_014.html</a>		
2017-06-06 10:19:24	蘋果香港新聞	2	<a href="#">10210291579178671</a>	Brian Siu		<a href="#">Post</a>	因管理不善而於全球最佳航空公司排名中節節下跌的國泰航空，其全資附屬公司國泰港龍再發生事故！一睇個開頭都知邊份報紙！		
2017-06-05 17:45:00	立場 - 蘋果空姐	33	<a href="#">10155217113114548</a>	MicMic Kwok		<a href="#">Post</a>	國泰管理層問題		

# Social Listening

- Impact of Hato Typhoon to Macau Casinos ([Wynn](#))
- 2017/08/30

**Dataact 數據富**  
Browse Social Media Content

Date	Page	Like	Love	Haha	Wow	Sad	Angry	Shares	Message	Message Ext	Story	Caption	Name	Description	Description Ext	
2017-08-28 13:37:06	<a href="#">澳門巴士/公共交通迷 Macau Buses and Public Transport Enthusiastic</a>	39	0	3	0	10	0	2	【天鴿風災】新福利屬下 AA 旅遊有限公司位於北安的車廠亦在今次風災中倒塌，壓著兩部分別是永利皇宮及銀河的發財巴。 ===== 28/8 #天鴿風災 #十號風球 #颱風 #災難 #澳門加油 #北安車廠 #AA #新福利							

Total Hits: 1

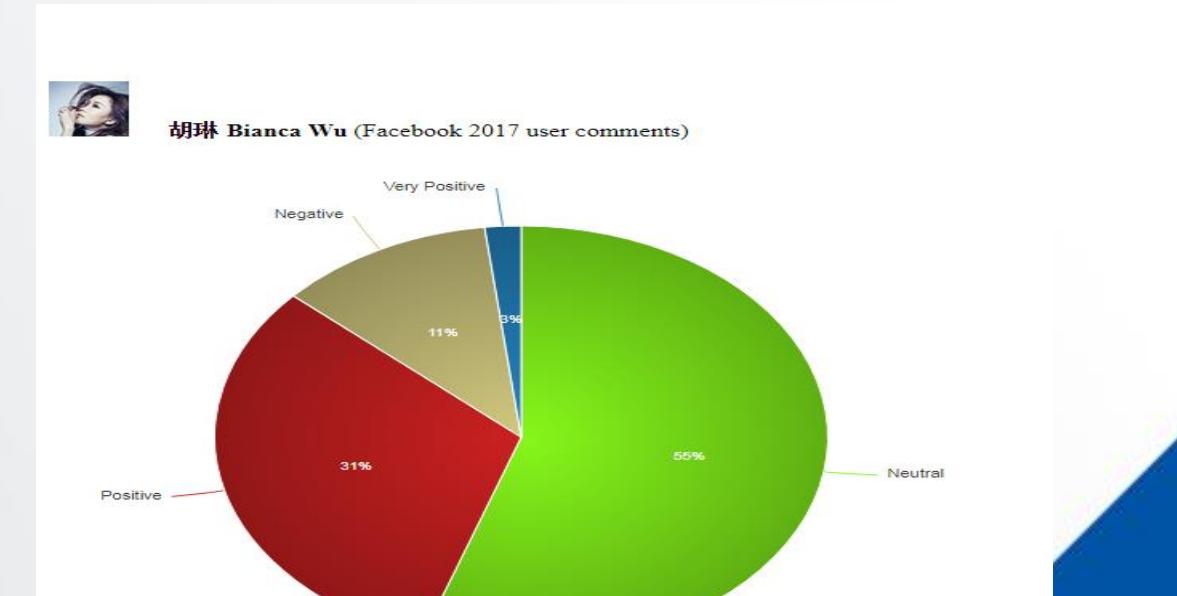
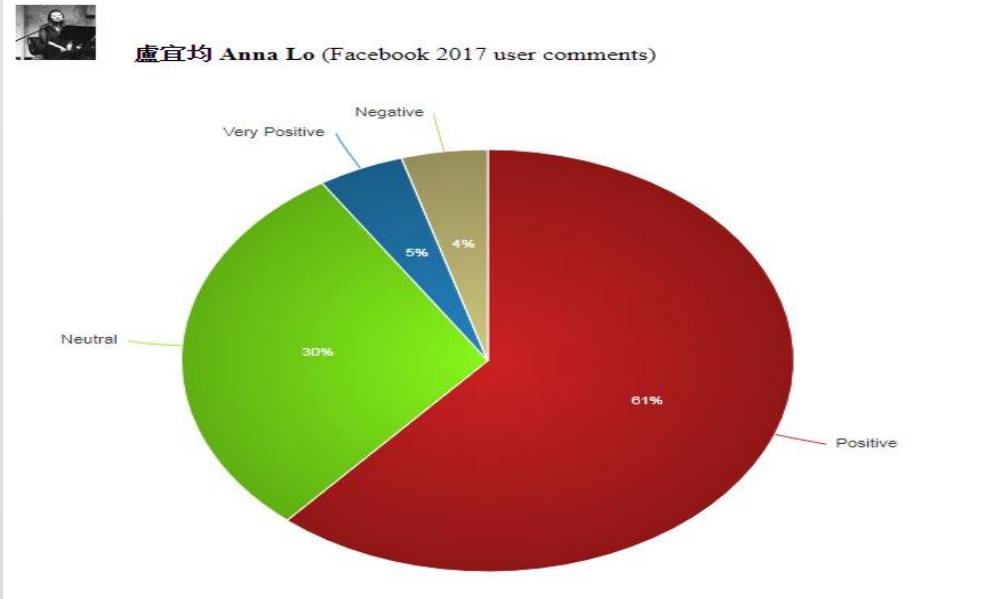
Timeline Photos

Previous Page | Next Page



# Users Sentiment

**Sentiment:** Based on Facebook [user comments](#) on the Fan Pages



# Content Farm

Generate **large amounts of textual content** to satisfy algorithms for **maximal retrieval** by automated **search engines** or to generate advertising revenue through attracting reader **page views**



# Amazon: Books Recommendation

amazon.com

WELCOME BOOKS MUSIC VIDEO TOYS & GAMES ELECTRONICS e-CARDS AUCTIONS zSHOPS

BOOK SEARCH BROWSE SUBJECTS BESTSELLERS FEATURED IN THE MEDIA AWARD WINNERS COMPUTERS & INTERNET KIDS BUSINESS & INVESTING

**Book Information**

**Hearts in Atlantis**  
by Stephen King

**List Price:** \$20.00  
**Our Price:** \$14.00  
**You Save:** \$14.00 (50%)

**Availability:** Usually ships within 24 hours.

**Hardcover** - 523 pages (September 14, 1999)  
Scribner; ISBN: 0684863515 ; Dimensions (in inches): 1.73 x 9.59 x 6.53

**Other Editions:** [Audio Cassette](#) (Unabridged), [Audio CD](#) (Unabridged), [Large Print](#)

**Amazon.com Sales Rank:** 37  
**Avg. Customer Review:** ★★★★☆  
Number of Reviews: 107

[Write an online review](#) and share your thoughts with other readers!

**Customers who bought this book also bought:**

- [The Girl Who Loved Tom Gordon](#); Stephen King
- [Bag of Bones](#); Stephen King
- [Storm of the Century \(Screenplay\)](#); Stephen King(Introduction)
- [Mr. X](#); Peter Straub

[Click here for more suggestions...](#)

**Add to Shopping Cart**  
(you can always remove it later)

**Shopping with us is 100% safe.  
Guaranteed.**

**Add to my Wish List**  
(We'll set one up for you)  
[View my Wish List](#)

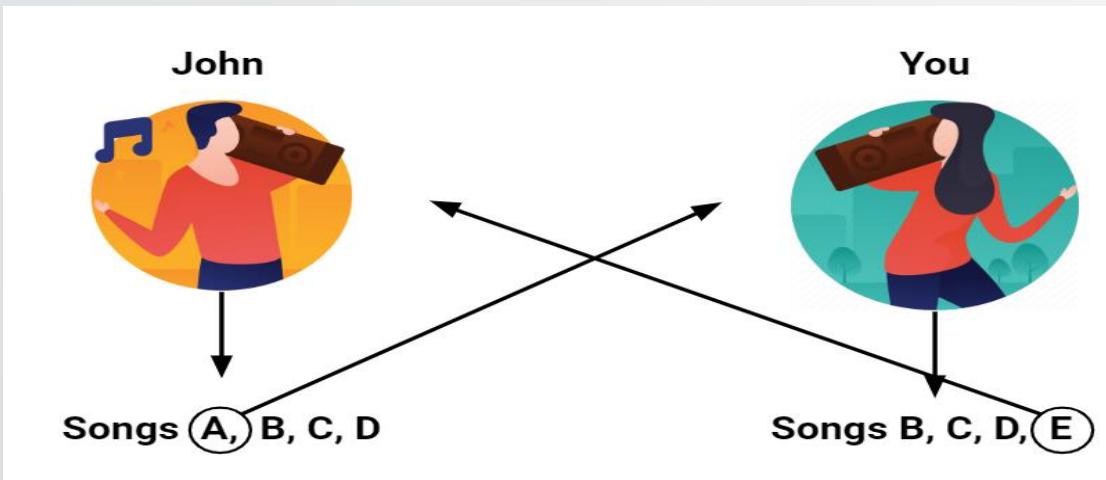
Books

All search: [Books](#), [Music](#), [Video](#), [Toys](#), or [Electronics](#)

# Spotify: Songs Recommendation

**Discover Weekly playlist:** Spotify finds songs that fit your profile but you haven't listened to

How Does Spotify Know You So Well?

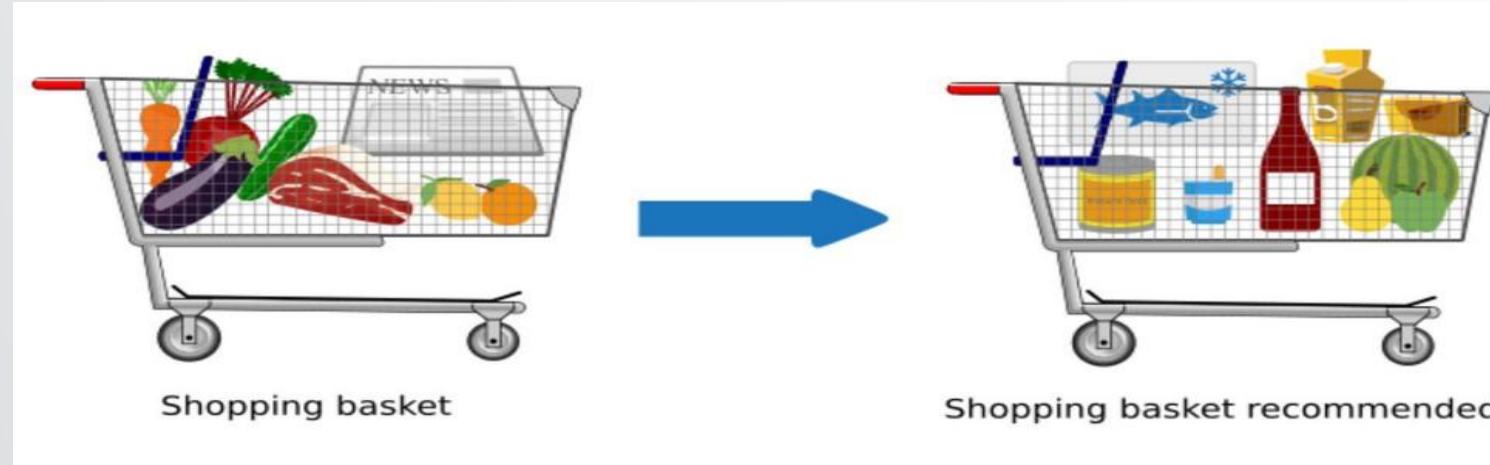


Collaborative Filtering



# Walmart Market Basket

**Market Basket:** identifies the **strength of association** between pairs of products purchased together and identify patterns of **co-occurrence**.



- YouTube reference: [Apriori FP-Growth](#)

# Walmart Market Basket



Supermarket Diaper & Beer Story

# Product Cross-selling Case Study



Datatact Example on Beauty Fan Pages

# Micro Influencer: Crowd Opinion Leader

Big companies may use **famous people (KOLs)** to promote their brands. But they are usually too **expensive** for small companies.



But effectiveness is not guaranteed !



MTR derailed

# Micro Influencer: Crowd Opinion Leader

- In Social Media (like FB), friends **like to tag each** other for certain brands and products
  - A word cloud from the 200 Beauty related FB Fan Pages shows that many **comments are friends “names”**
  - The details reveal a particular user tagging her friend frequently

Total Hits: 347							
Date	Page	Like	User	Name	Pic	Post	Comment
2017-04-18 08:17:02	<a href="#">BeautyExchange.com.hk</a>	1	<a href="#">10210672932272679</a>	<a href="#">Vicky Lau</a>		<a href="#">Post</a>	Candy Choi
2017-04-17 09:53:34	<a href="#">BeautyExchange.com.hk</a>	1	<a href="#">10210672932272679</a>	<a href="#">Vicky Lau</a>		<a href="#">Post</a>	Candy Choi that's true
2017-04-16 09:37:20	<a href="#">BeautyExchange.com.hk</a>	1	<a href="#">10210672932272679</a>	<a href="#">Vicky Lau</a>		<a href="#">Post</a>	Candy Choi agreed
2017-04-14 13:53:05	<a href="#">BeautyExchange.com.hk</a>	1	<a href="#">10210672932272679</a>	<a href="#">Vicky Lau</a>		<a href="#">Post</a>	Candy Choi your friend
2017-04-13 22:56:39	<a href="#">BeautyExchange.com.hk</a>	1	<a href="#">10210672932272679</a>	<a href="#">Vicky Lau</a>		<a href="#">Post</a>	Candy Choi
2017-04-13 17:26:41	<a href="#">BeautyExchange.com.hk</a>	1	<a href="#">10210672932272679</a>	<a href="#">Vicky Lau</a>		<a href="#">Post</a>	Candy Choi
2017-04-13 11:54:16	<a href="#">BeautyExchange.com.hk</a>	1	<a href="#">10210672932272679</a>	<a href="#">Vicky Lau</a>		<a href="#">Post</a>	Candy Choi



# Micro Influencer: Crowd Opinion Leader

## COL Identification

- **Friend List:** number of friends
- **COL Profile:** trace the user actions
- **Activity Heat Map:** based on the activity metrics which include the frequency of tagging friends, reactions (likes, angry, wow) to particular content or comments
- **Friends reaction:** tagged friends like it or reply with positive comments



## Goals

- Match the COL profile to the appropriate brands/companies
- COL **matching fee is based on the effectiveness** (e.g. Feedback Rate, the usage of **discount coupon** issued by the COL)
- Make the COL **exclusively** for a company will incur additional charge

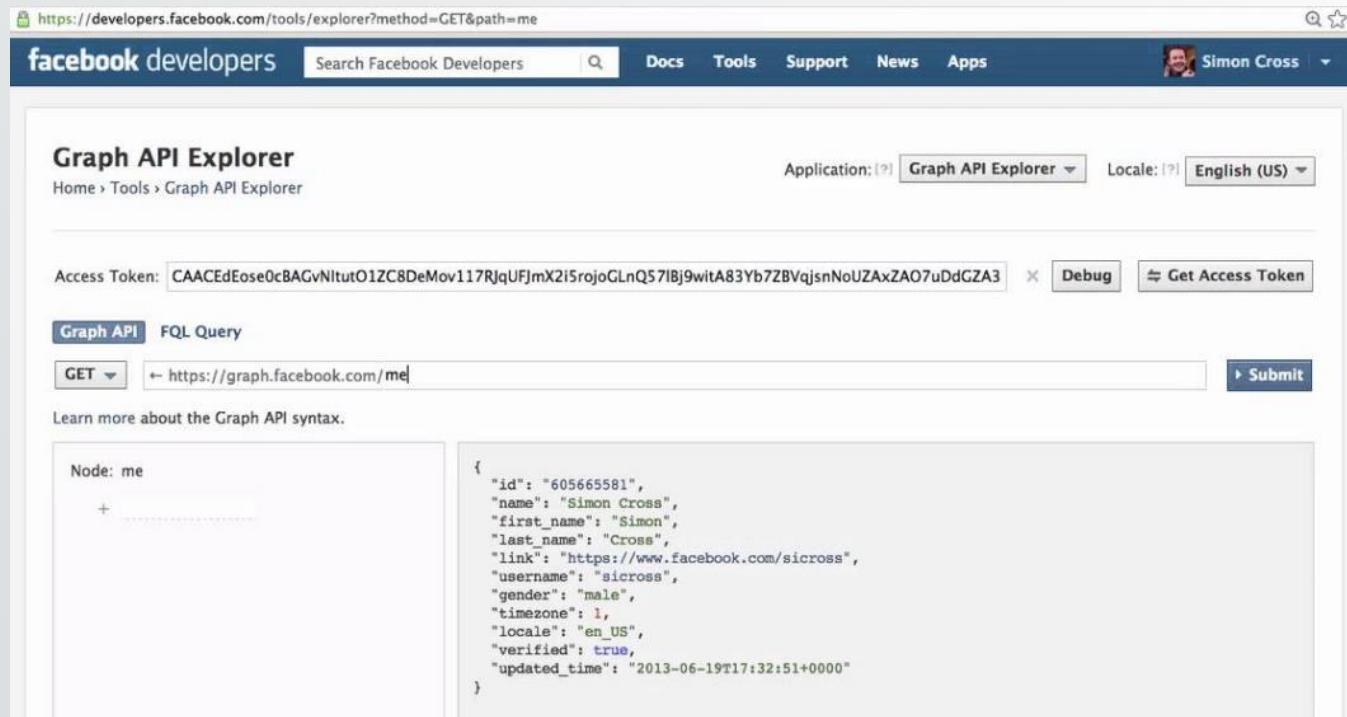
# Data Processing Pipeline

- Extract data from Facebook
- Storage (Database)
- Pre-process:
  - Cleaning
  - Small Caps
  - Tokenize
  - Remove StopWord
  - Word Stemming
- Visualization

# Data Extraction

Meta (facebook) developer account: <https://developers.facebook.com/>

- [How To Set up Facebook Developer Account & Create Facebook App ID](#)
- [Getting Started with the Graph API](#)



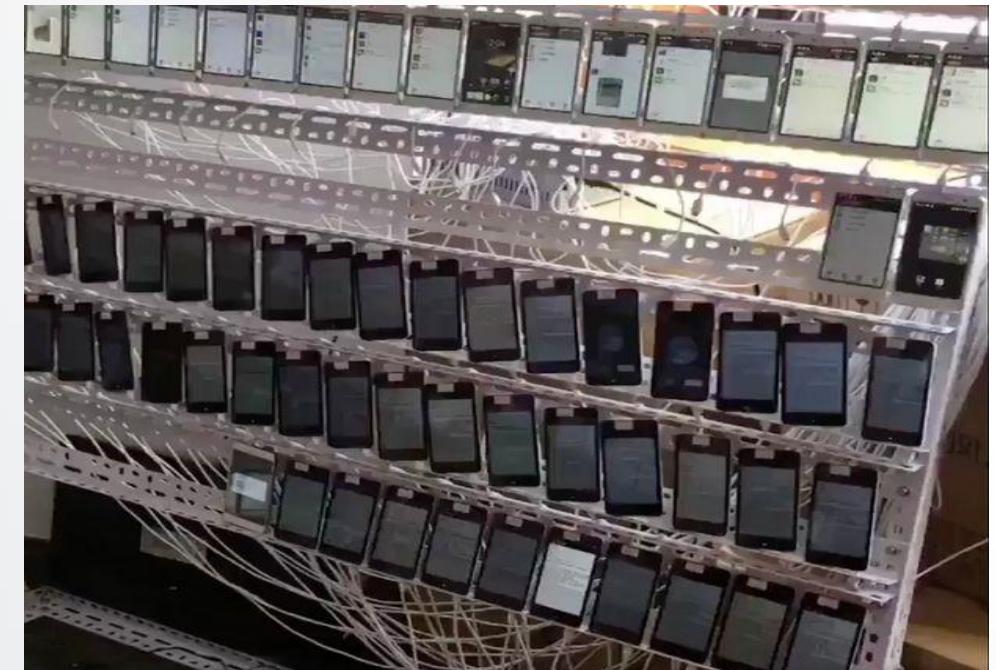
The screenshot shows the Facebook Graph API Explorer interface. At the top, it displays the URL <https://developers.facebook.com/tools/explorer?method=GET&path=me>. The page title is "Graph API Explorer". The "Application" dropdown is set to "Graph API Explorer" and the "Locale" dropdown is set to "English (US)". The "Access Token" input field contains a long string of characters: CAACEdEose0cBAGvNltutO1ZC8DeMov117RjqUFJmX2iSrojoGLnQ57IBj9witA83Yb7ZBVqjsnNoUZAxZA07uDdGZA3. Below this, there are buttons for "Debug" and "Get Access Token". The "Graph API" tab is selected, and the "FQL Query" tab is also visible. Under "Graph API", a "GET" button is selected, and the URL <https://graph.facebook.com/me> is entered into the query field. A "Submit" button is located to the right of the URL. On the left, under "Node: me", there is a "+ Add Edge" button. On the right, the JSON response for the "me" node is displayed:

```
{
  "id": "605665581",
  "name": "Simon Cross",
  "first_name": "Simon",
  "last_name": "Cross",
  "link": "https://www.facebook.com/sicross",
  "username": "sicross",
  "gender": "male",
  "timezone": 1,
  "locale": "en_US",
  "verified": true,
  "updated_time": "2013-06-19T17:32:51+0000"
}
```

# Data Extraction

- Manage Facebook account
- Disconnect and retry
- Throttling
- Privacy limitation (only **public fan pages**)
- Web Scraping

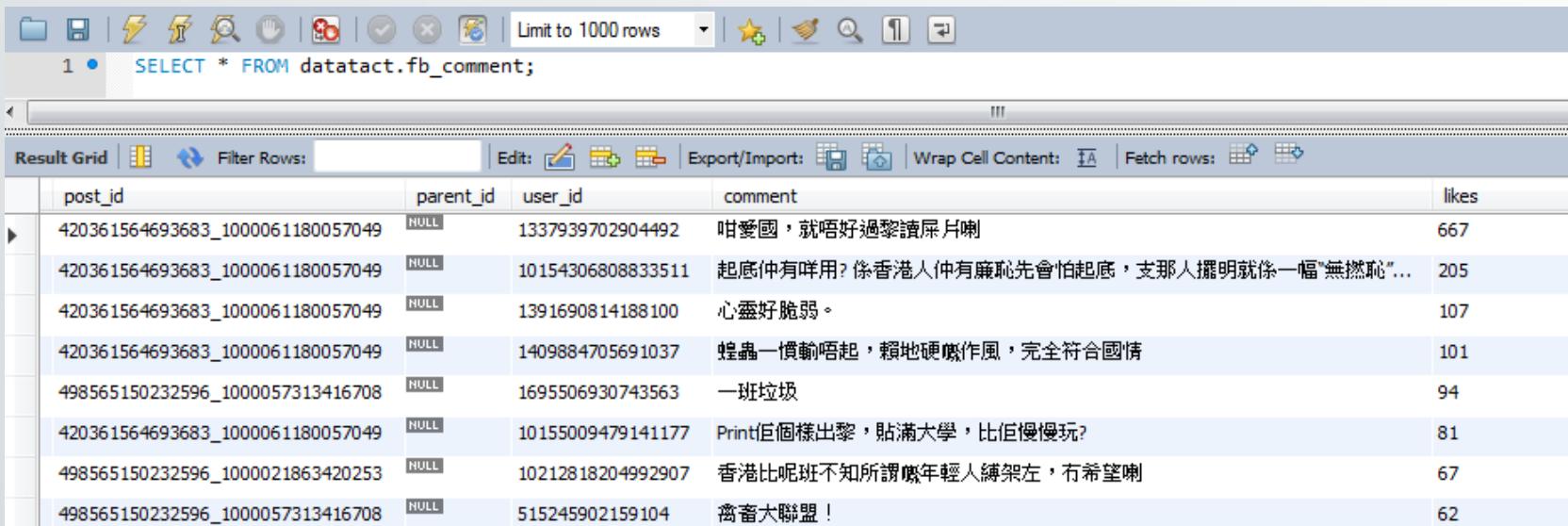
[Cambridge Analytica Scandal: Facebook Allowed Firm To Exploit Data To Target American Voters](#)



# Data Storage

## (Relational) Database

- Durable
- Handle concurrency (multiple writer)
- Fast retrieval



The screenshot shows a MySQL Workbench interface with a query editor and a result grid. The query is:

```
1 •  SELECT * FROM datatact.fb_comment;
```

The result grid displays the following data:

post_id	parent_id	user_id	comment	likes
420361564693683_1000061180057049	NULL	1337939702904492	咁愛國，就唔好過黎讀屎片喇	667
420361564693683_1000061180057049	NULL	10154306808833511	起底仲有咩用？係香港人仲有廉恥先會怕起底，支那人擺明就係一幅“無廉恥”...	205
420361564693683_1000061180057049	NULL	1391690814188100	心靈好脆弱。	107
420361564693683_1000061180057049	NULL	1409884705691037	蝗蟲一慣輸唔起，賴地硬噏作風，完全符合國情	101
498565150232596_1000057313416708	NULL	1695506930743563	一班垃圾	94
420361564693683_1000061180057049	NULL	10155009479141177	Print佢個樣出黎，貼滿大學，比佢慢慢玩？	81
498565150232596_1000021863420253	NULL	10212818204992907	香港比呢班不知所謂嘅年輕人縛架左，冇希望喇	67
498565150232596_1000057313416708	NULL	515245902159104	畜生大聯盟！	62

# Text Pre-processing

## Cleaning

- Remove symbols (, [ ] \! # etc)
- Remove URLs

## Tokenize (Break the sentence into words)

- English is easy but Chinese is not
- Example:
  - 兒子生性病母倍感安慰
  - 兒子 / 生性病 / 母 / 倍感 / 安慰  
*(Wrong: a mother is happy about his son getting sexual disease)*
  - 兒子 / 生性 / 病母 / 倍感 / 安慰  
*(Correct: a sicken mother is happy about his son being good)*

# Text Pre-processing

## Remove Stop Word

- **English:** a an the
- **Chinese:** 的 了 嘴

## Small Caps

- Convert all English letters to small caps

## Word Stemming

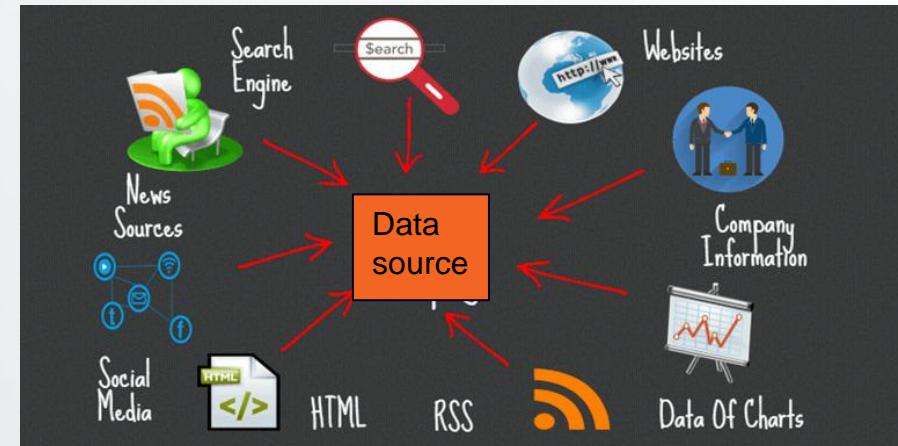
- **English:** stems stemmer stemming stemmed
- **Chinese:** 香港大學, 港大, hku

## Code Example

# Other Data Source

Where do you get the data ?

- Social Media
- Internet communities
- Web Scraping
- Govt / public organizations
- Buy from Data companies
- Generate from your own business



# Searching

**Dataact 數據富**  
**Browse Social Media Content**

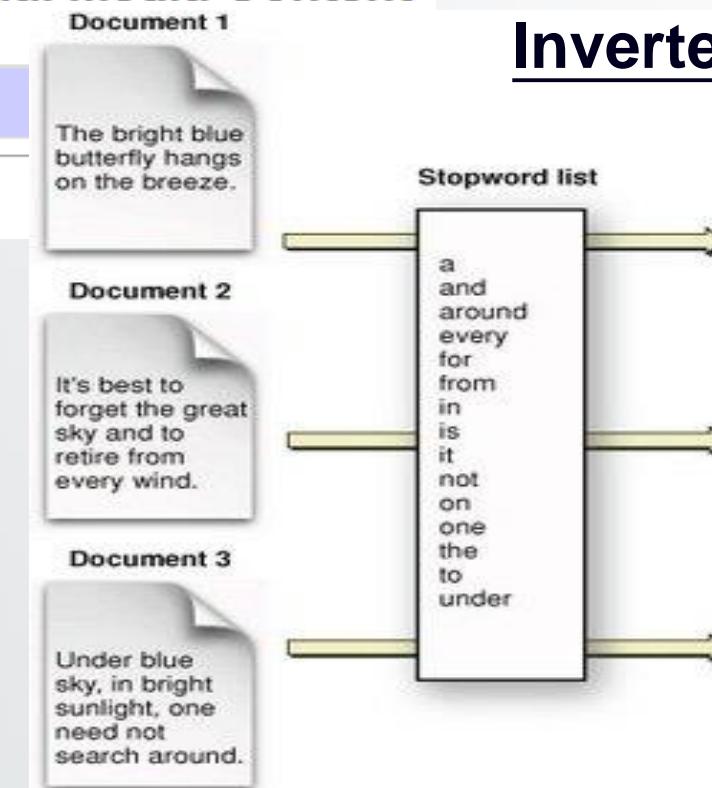
Browse Content      Correlations

Total Hits: 168,928

big data

All Images News Videos

About 7,200,000,000 results (0.70 seconds)

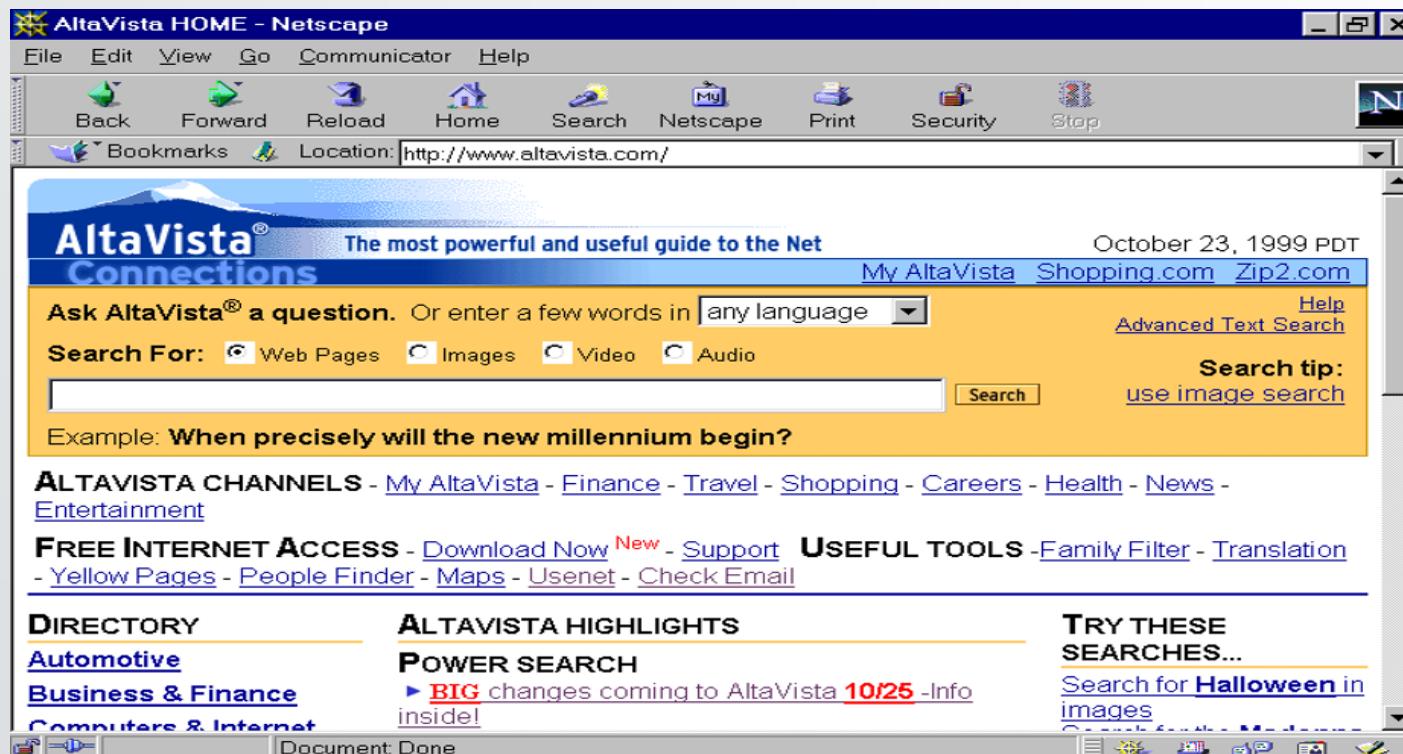


## Inverted Index

Source: Hitachi Vantara Community

# Ranking

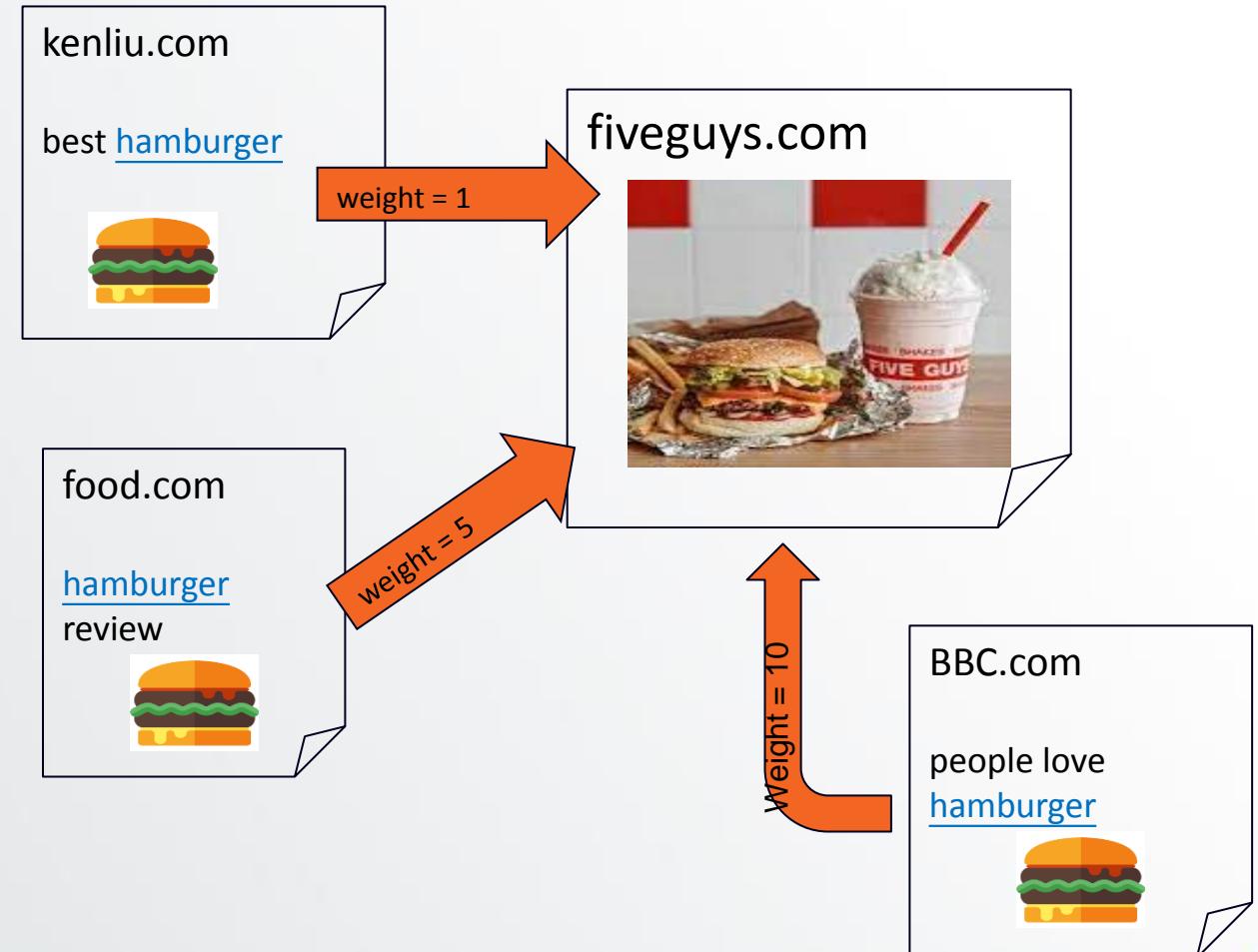
How to rank and display the top 10 from the 1 millions matched results?



AltaVista in 1995: ranking is primarily based on text

# Ranking

- Billion dollars insight: Google realized that relevance **depended on trust, not just text**
- Trust is calculated by **counting “weighted” back-links** Each link is like a vote or endorsement (PageRank Algorithm)

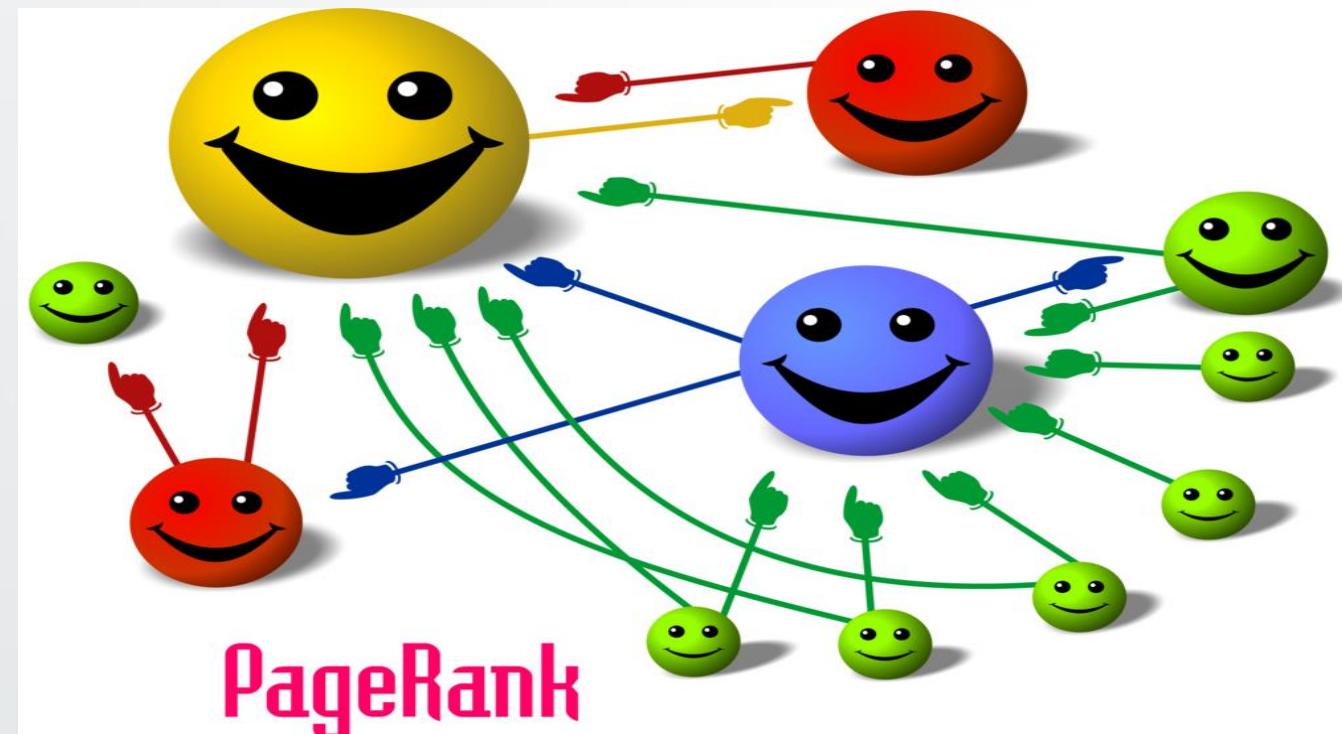


# Google PageRank

What's the biggest challenge to make PageRank accurate?

How many back-links for a particular website in total?

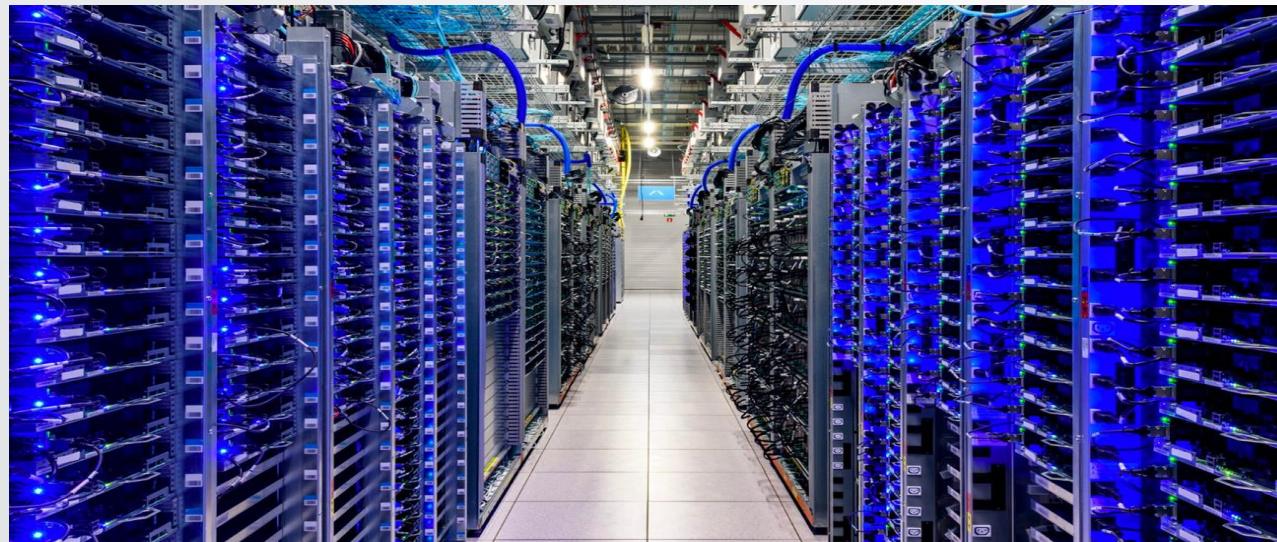
In order to count the total back-links, needs to know all the links from **all other web sites in the world.**



# Google PageRank



Hence Google **stores (caches) entire Web by Crawlers or Spiders** in their **data centers**.



# Social Media Ranking

- Sort by Date

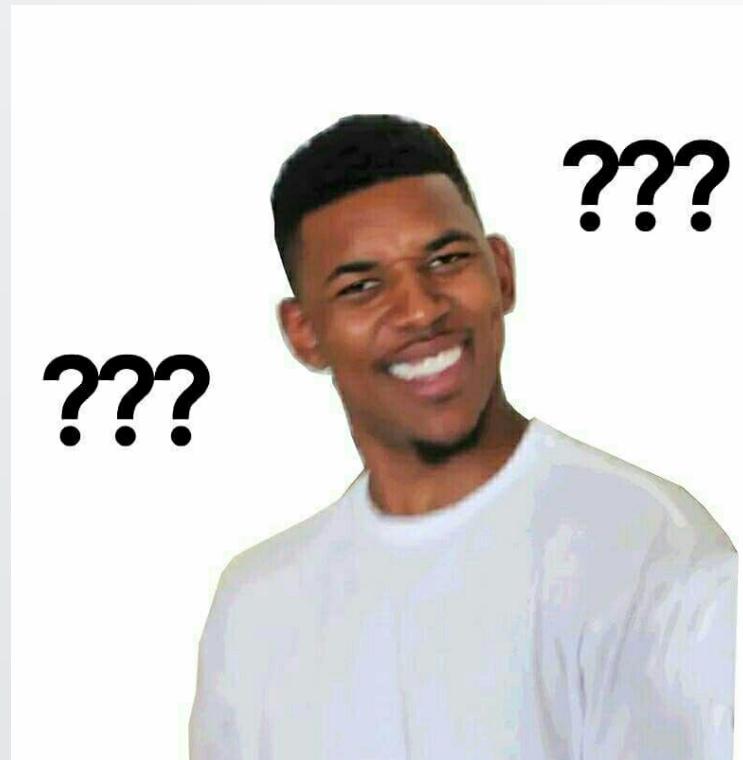
- Sort by



- Sort by Popularity Index

**10\*comment + 5\*Love + 5\*Angry + 4\*Sad + 3\*wow + 3\*haha + 2\*Comment\_like + Page\_like**

# Q & A



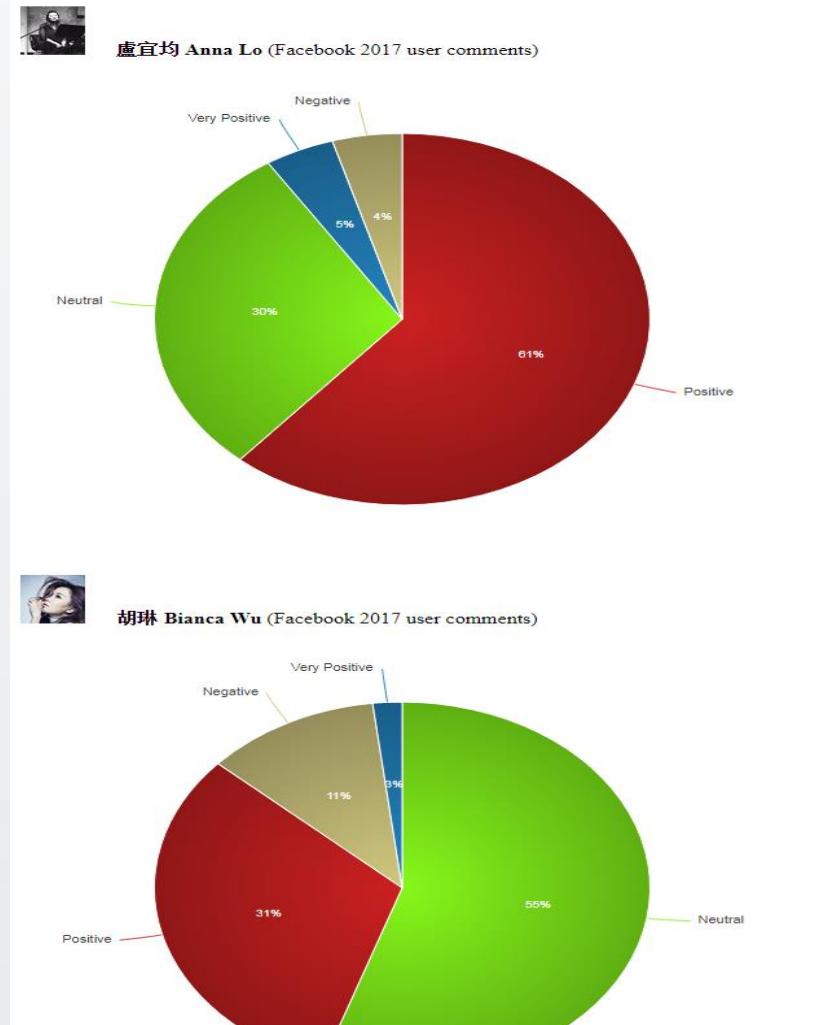
# Course Outline – Day 2

## Day 2

- Case studies: Applications of social sentiment analysis
- Basic building blocks of Natural Language Processing
- Word2Vec embedding to understand language semantics
- Introduction to Deep Learning based image recognitions and case studies

# Customers Sentiments

- Refer to the Excel for detail breakdown
  - [sentimentAnnaLo](#)
  - [sentimentBiancaWu](#)



# Customers Sentiments



- Most of sentiment tools is in English
- Not many for Chinese (especially Cantonese)
- Use [translator to translate Cantonese Chinese to English](#) and then feed to the English sentiment engine

DETECT LANGUAGE CHINESE ENGLISH CHINESE (TRADITIONAL)

頂你 × Like you

Dǐng nǐ

2 / 5000 拼

34

# Social Media Sentiment vs Stock Price



## Social media ramp-and-dump scam

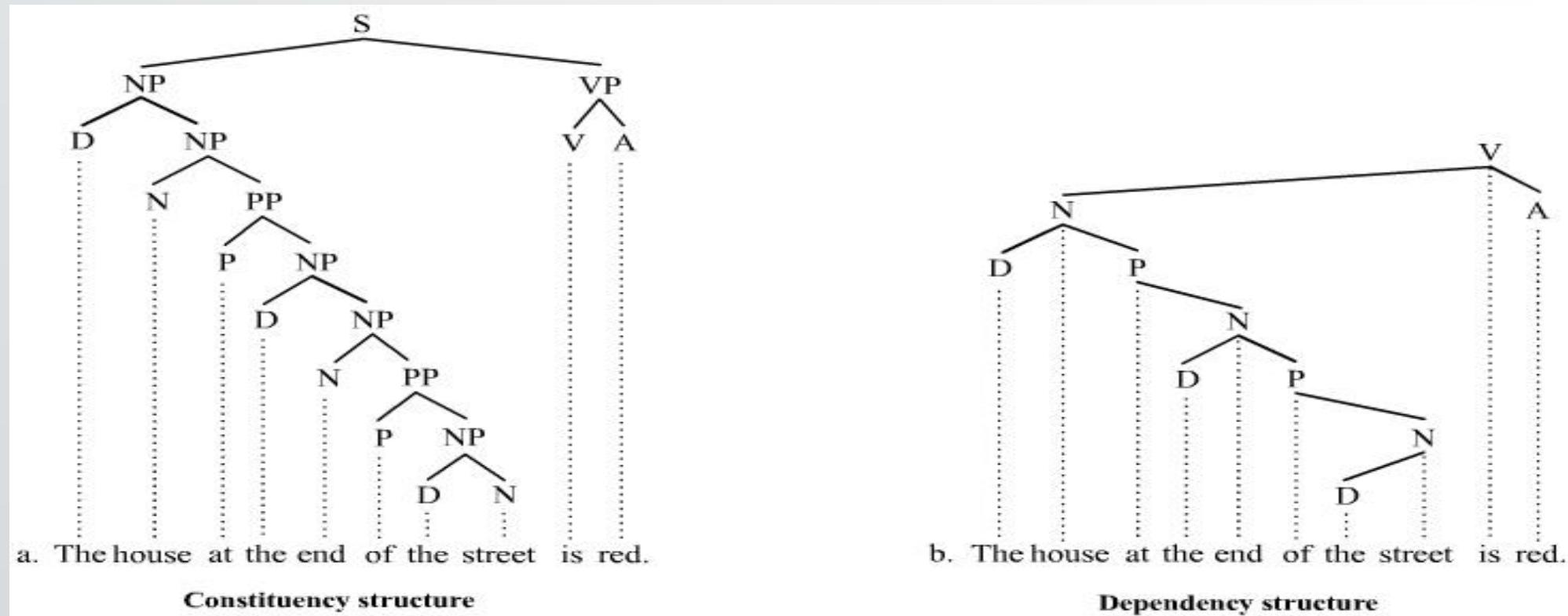
A social media ramp-and-dump scam is a form of stock market manipulation where fraudsters use different means to “ramp” up the share price of a listed company and then induce investors via social media platforms to purchase the shares they “dump” at an artificially high price



Code Example: Twitter and stock price correlation

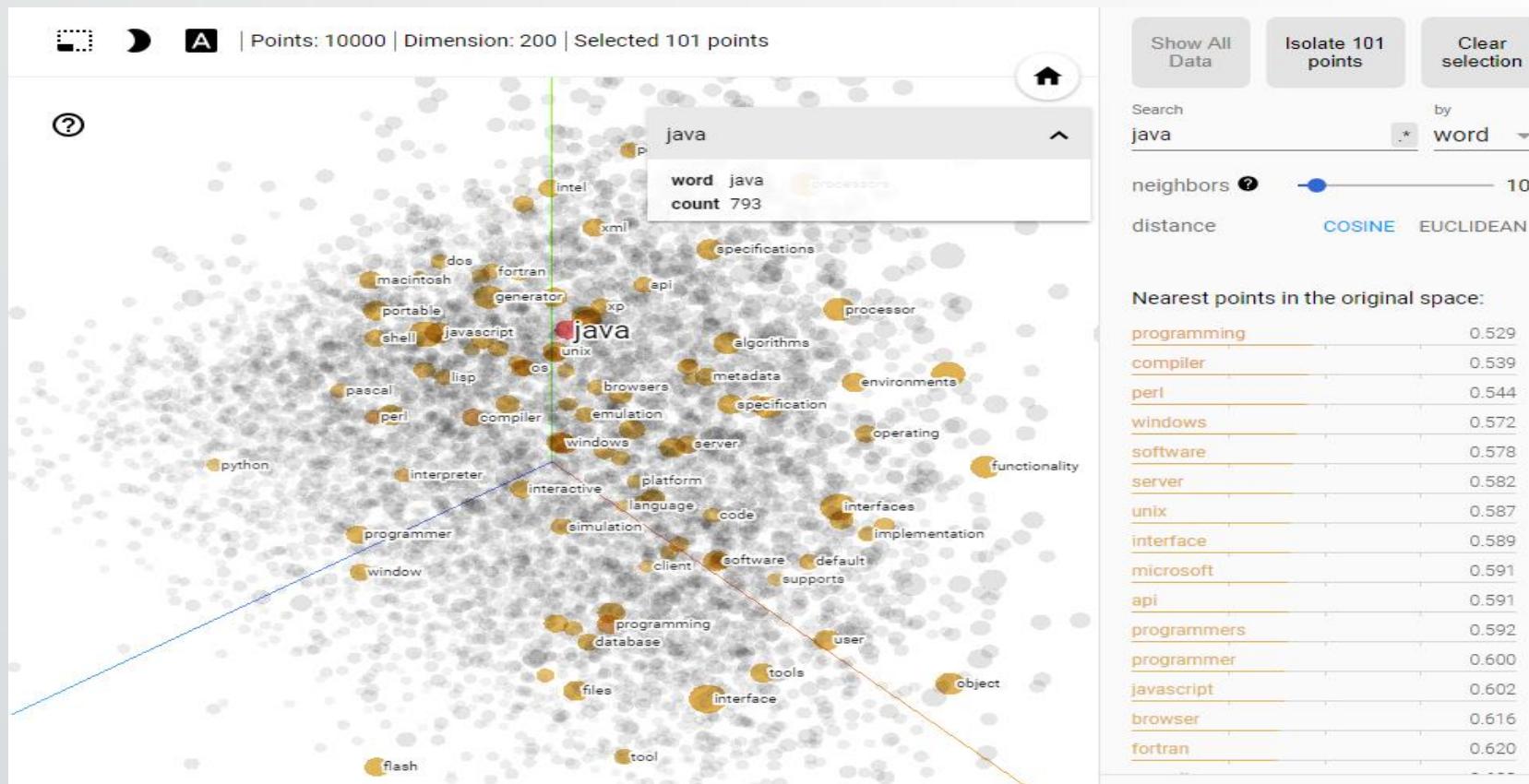
# Natural Language Processing (NLP)

- Treebanks [NLP Demo](#)



Source: Wikipedia

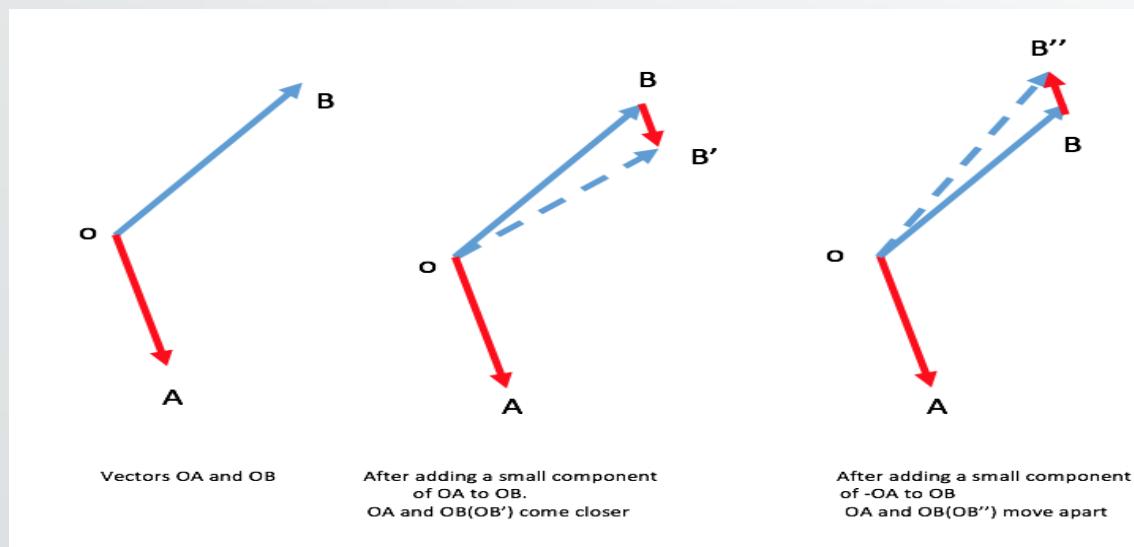
# Word2Vec



Interactive Demo  
Code example of Word2Vec

# Word2Vec – Definition

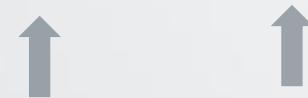
- **Example:** Mary is a very stubborn child. Her pervicacious nature always gets her in trouble: pervicacious is surrounded by stubborn, nature, and trouble.
- **Context:** a fixed number of words **in front of and behind** the word of interest. Learns the meaning of a given word by **looking at its context** and **representing it numerically**.



*“a word is characterized by the company it keeps - John Rupert Firth”*

# Word2Vec – Definition

*I would like to go someplace nearby for lunch*



Input	Output
go	like
go	to
go	someplace
go	nearby

Input	Output
someplace	to
someplace	go
someplace	nearby
someplace	for

- The Illustrated Word2vec

# Word2Vec – Dissimilarity

JPMorgan Equity Research Report 2007 - 17

```
doesn't_match('revenue tax sales income'.split())
Out: 'tax'
```

```
doesn't_match('apple samsung microsoft'.split())
Out: 'microsoft'
```

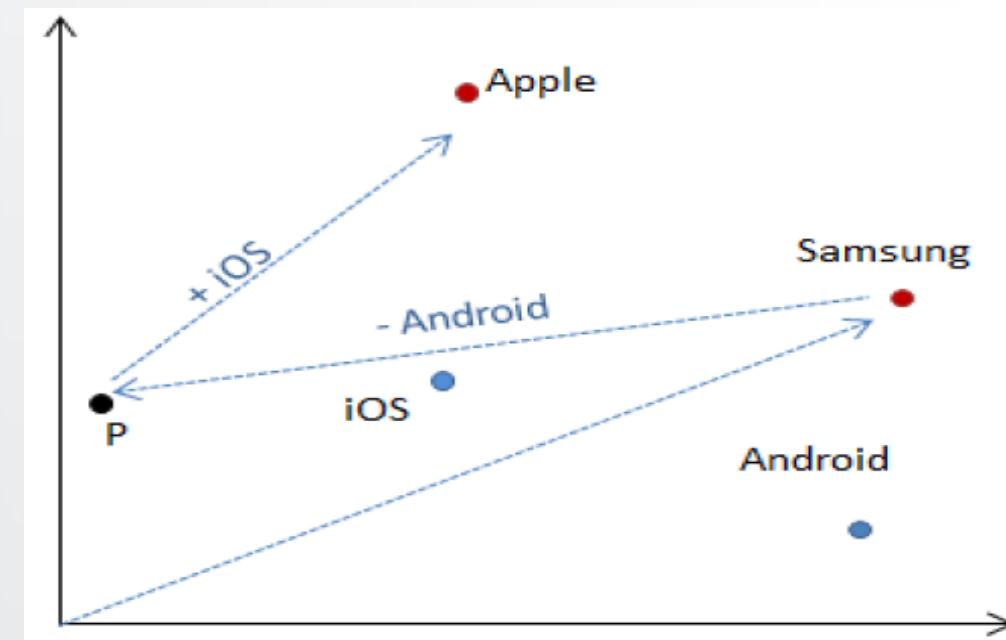
```
doesn't_match('apple ibm microsoft'.split())
Out: 'apple'
```

# Word2Vec – Addition and Subtraction

JPMorgan Equity Research Report 2007 - 17

**samsung + ios – android = ?**

```
most_similar(  
    positive=['samsung', 'ios'],  
    negative=['android'])  
  
[ ('apple', 0.35112401843070984) ]
```



# Word2Vec – Corpus comparison

JPMA	Google News
<b>Driver</b>	
contributor	drivers
drivers	Driver
catalyst	motorist
determinant	driving
culprit	car
beneficiary	motorcyclist
determinants	wheel
differentiator	Drivers
detractor	truck
contributors	vehicle

JPMA	Google News
<b>overweight</b>	
neutral	Overweight
ow	underweight
underweight	obese
uw	morbidly_obese
constructive	obesity
cautious	Underweight
selective	severely_obese
pt	metabolic_syndrome
intact	waist_circumference
bullish	BMI

# Word2Vec – Chinese Language

Tencent AI Lab Embedding Corpus for Chinese Words and Phrases

<b>Input</b>	喀拉喀什河	因吹斯汀	刘德华	自然语言处理
<b>Top similar words</b>	墨玉河 和田河 玉龙喀什河 白玉河 喀什河 叶尔羌河 克里雅河 玛纳斯河	一颗赛艇 因吹斯听 城会玩 厉害了word哥 emmmmm 扎心了老铁 神吐槽 可以说是非常爆笑了	刘天王 周润发 华仔 梁朝伟 张学友 古天乐 张家辉 张国荣	自然语言理解 计算机视觉 自然语言处理技术 深度学习 机器学习 图像识别 语义理解 语音识别

Training Data: News, Webpages, Novels, Wikipedia and Baidu Baike



# Word2Vec – Cultural Vectors

How Does Spotify Know You So Well?

<b>n2 Term</b>	<b>Score</b>
dancing queen	0.0707
mamma mia	0.0622
disco era	0.0346
winner takes	0.0307
chance on	0.0297
swedish pop	0.0296
my my	0.0290
s enduring	0.0287
and gimme	0.0280
enduring appeal	0.0280

<b>np Term</b>	<b>Score</b>
dancing queen	0.0875
mamma mia	0.0553
benny	0.0399
chess	0.0390
its chorus	0.0389
vous	0.0382
the invitations	0.0377
voulez	0.0377
something's	0.0374
priscilla	0.0369

<b>adj Term</b>	<b>Score</b>
perky	0.8157
nonviolent	0.7178
swedish	0.2991
international	0.2010
inner	0.1776
consistent	0.1508
bitter	0.0871
classified	0.0735
junior	0.0664
produced	0.0616



# Word2Vec – Fashions

STITCH FIX: send you clothing and accessories we think you'll like; you keep the items you want and send the others back. We leverage data science to deliver personalization at scale. Word is Worth a thousand of Vectors

```
model.most_similar('ITEM_3469', 'pregnant')
matches = list(filter(lambda x: 'ITEM_' in x[0], matches))

# ['ITEM_13792',
# 'ITEM_11275',
# 'ITEM_11868']
```



# Generative Models: Text



Talk to Transformer

Chinese Novel Generation

GPT-2 Language Model

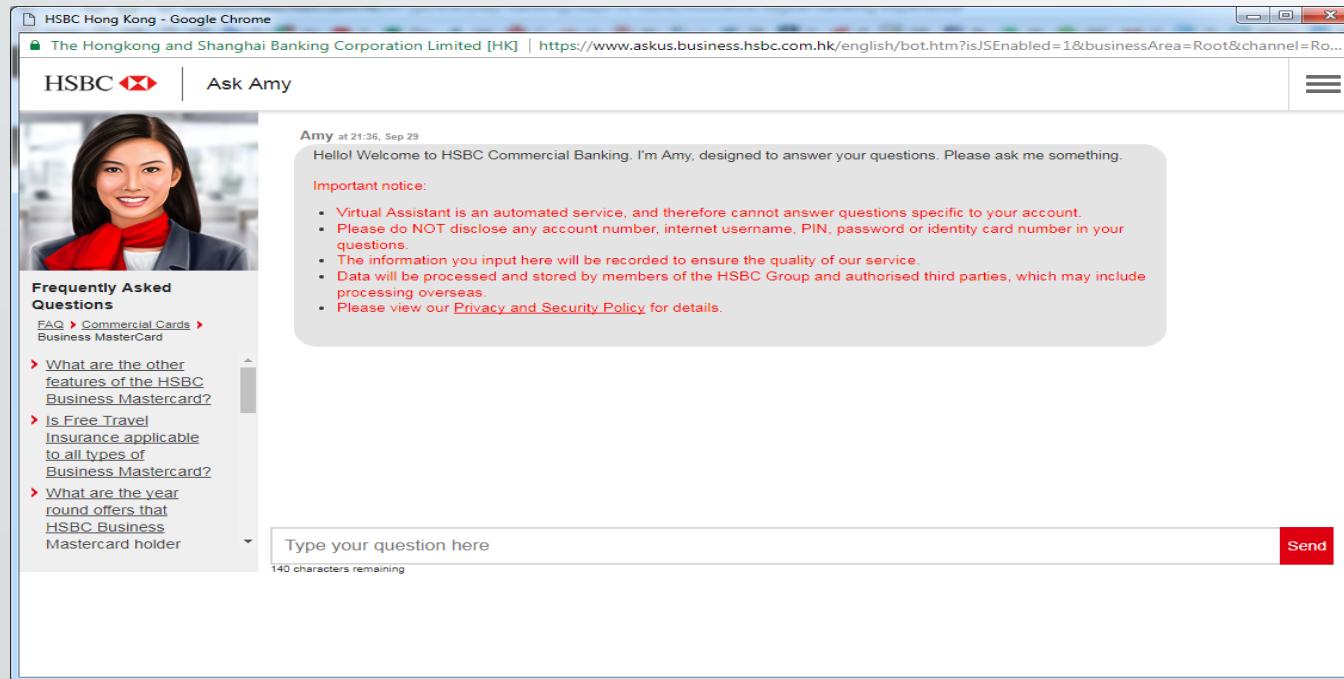
Trained to predict the next word in **40GB** of Internet text. It is a large transformer-based language model with **1.5 billion parameters**, trained on a dataset of **8 million web pages**.

GPT-3

**175 billion parameters**

**Demo:** GPT-3-powered business idea generator

# Example: HSBC Ask Amy Chatbot



The screenshot shows a web browser window for HSBC Hong Kong - Google Chrome. The URL is https://www.askus.business.hsbc.com.hk/english/bot.htm?isJSEnabled=1&businessArea=Root&channel=Ro... The page title is "HSBC Ask Amy". On the left, there's a sidebar with a portrait of a woman, a "Frequently Asked Questions" section, and links for "FAQ", "Commercial Cards", and "Business MasterCard". The main content area features a message from "Amy" at 21:36, Sep 29: "Hello! Welcome to HSBC Commercial Banking. I'm Amy, designed to answer your questions. Please ask me something." Below this is an "Important notice:" box containing several bullet points about data handling and security. At the bottom, there's a text input field with "Type your question here" placeholder text and a "Send" button.



[Give it a try](#)

# Image Recognitions

**Example:** 1.3 billions of faces in China, how can tell a computer to recognize a person?

- # pseudo code

- Color in {brown, yellow, black}
- Has {hair, eye \* 2, nose, mouth,
  - ear \* 2 .....
- Eye: {black, white, brown .....}
- .....

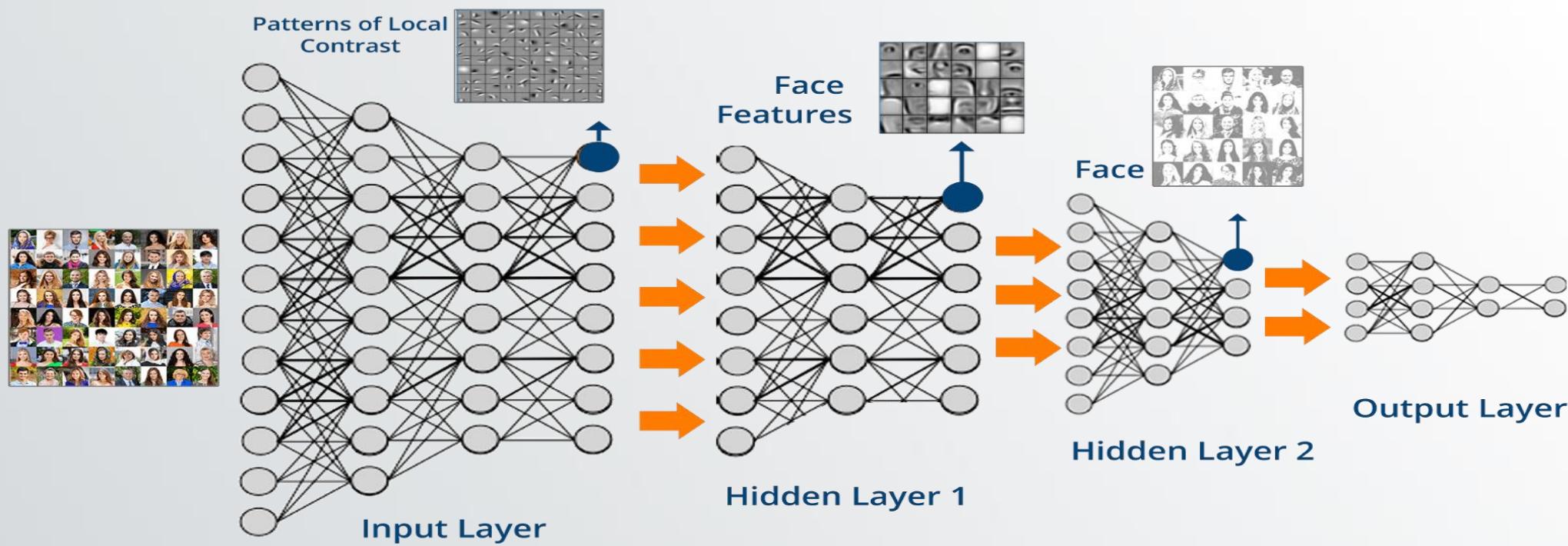


Police arrested a **fugitive** in southeast China after facial recognition technology helped identify him in a **crowd of about 50,000** people attending a **pop concert**.

<https://www.npr.org/sections/therecord/2018/05/23/613692526/thanks-to-ai-a-3rd-person-is-arrested-following-a-pop-superstars-concert>

# Image Recognitions

Learn from data rather than explicitly program the logic: let the computer see many examples (face, name) and learn how to represent it in a Deep Neural Network (brain-like structure)



# Deep Learning - Neural Networks

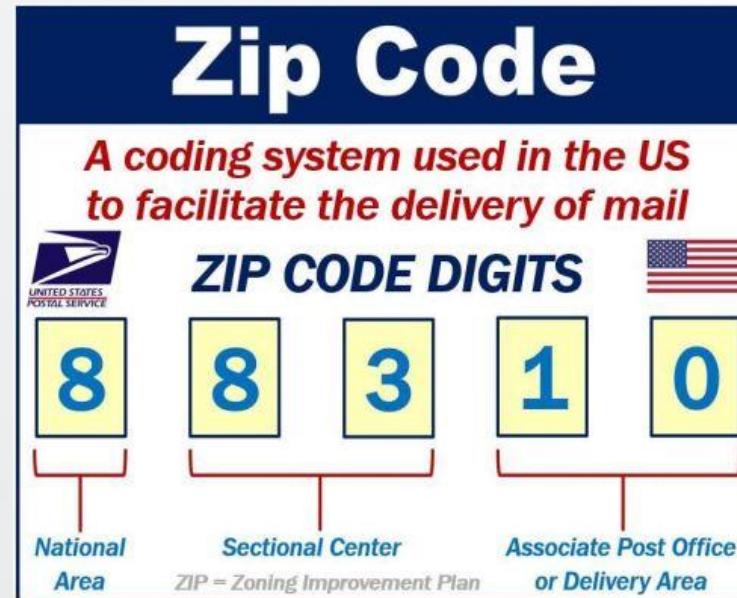
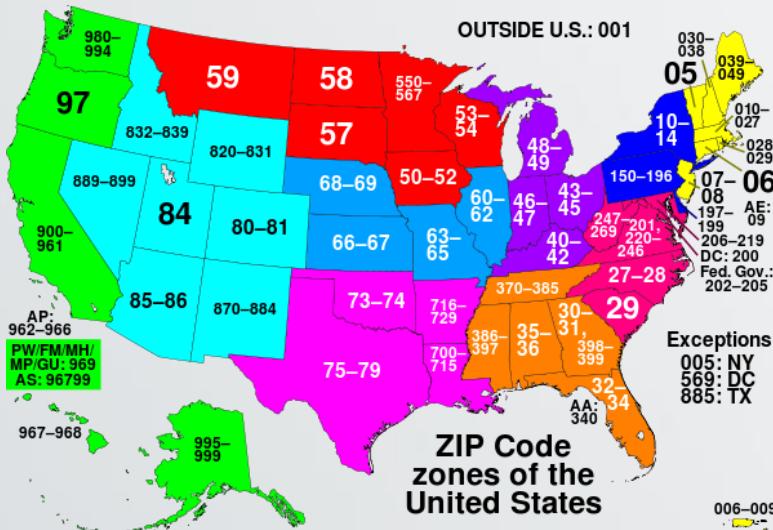
Think of Neurons as a **Lego** building blocks.

They can **connect together** to form many different shapes



# Demo: Hand-written Image Classification

Recognize **hand-written** zip code and route the mails accordingly



# Labelled Data

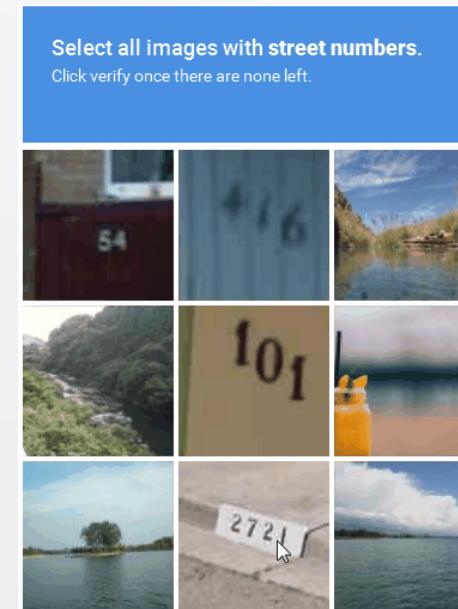
Where to find **labelled dataset**?

- Freely source from the web
  - [Kaggle Dataset](#)
  - [Google Dataset Search](#)
  - [AWS Open Data](#)
- Pay \$
  - [Human Data Labeler](#)
  - [Amazon Mechanical Turk](#)
  - [Datafarm.co](#) ← (My friend's startup)



# Labelled Data

Where to find **more labelled dataset?** From **YOU!** (freely)



Google reCAPTCHA

# Labelled Data

WFH (Work from Home Hill)



Tagged your friends in Facebook or IG

# Image Recognitions Use Cases



# Use Cases – Face Recognition

化妝就係你落咗妝，我仲認得你，你落咗妝，我唔認得你，噃啲叫做喬裝。



If you can still unlock Faceld, it's called make-up, otherwise it's a “fake-up”

上咗妝後 **Faceld** 仲認得叫做化妝，認唔到嘅叫做喬裝

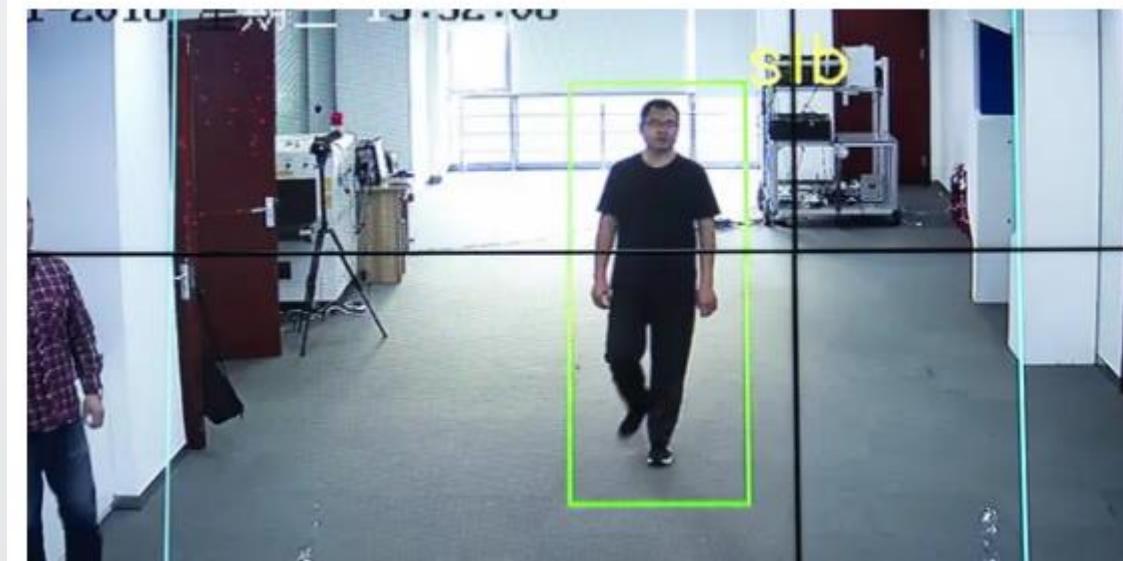


# Gait Recognition

Surveillance camera technology that can recognize you just from how you walk

Where is the labelled training dataset ([Walking Style -> Name](#)) coming from?

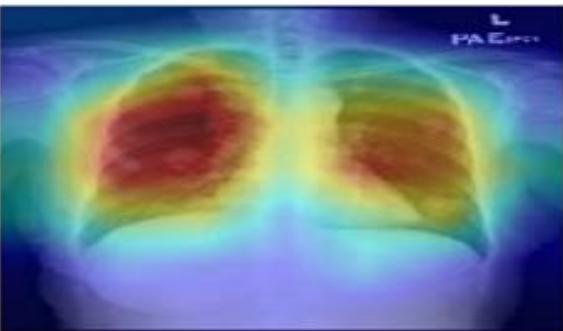
- **Shops:** electronic payment + shops cameras
- **Security check:** in airports, train stations when pass through the security gate
- **Name-registered phone (China):** named-phone GPS + street cameras



Watrix employees demonstrate their firm's gait recognition software at their company's offices in Beijing, October 31, 2018. AP

# Use Cases – Commercial & Medical

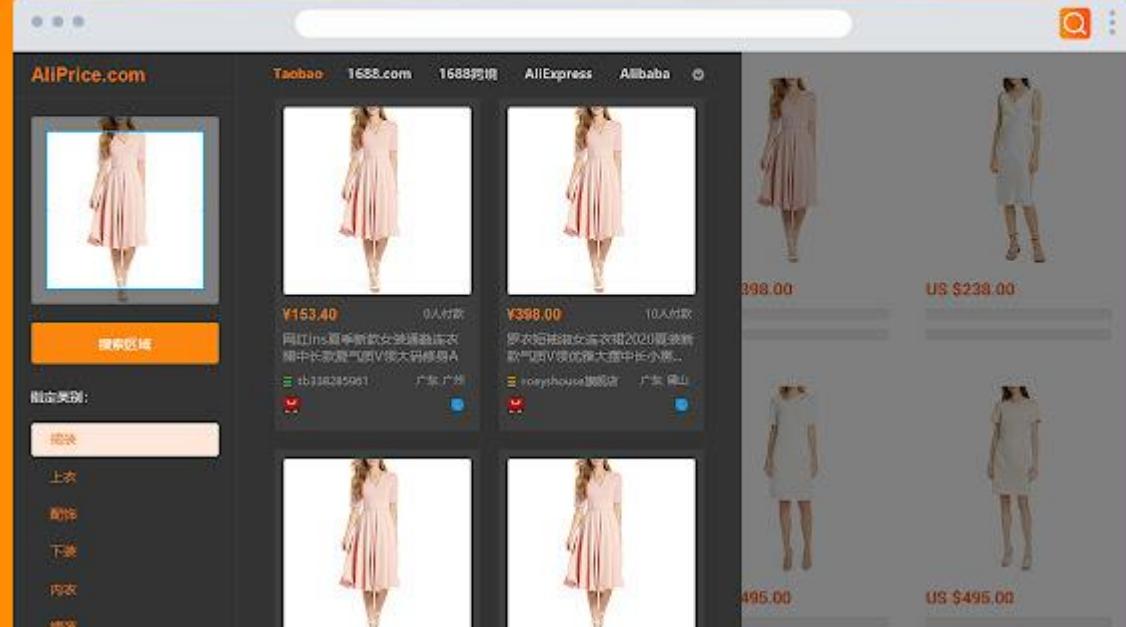
### Auto Identification of Lung Diseases via Image Analytics



- Making use of both public (NIH) and HA data
- Able to automatically identify 14 lung diseases (with AUC=0.75 – 0.95)

### Search for the same products by image

Taobao, Alibaba, AliExpress, 1688.com, 1688 oversea, Pinduoduo, eBay, Shopee, JD

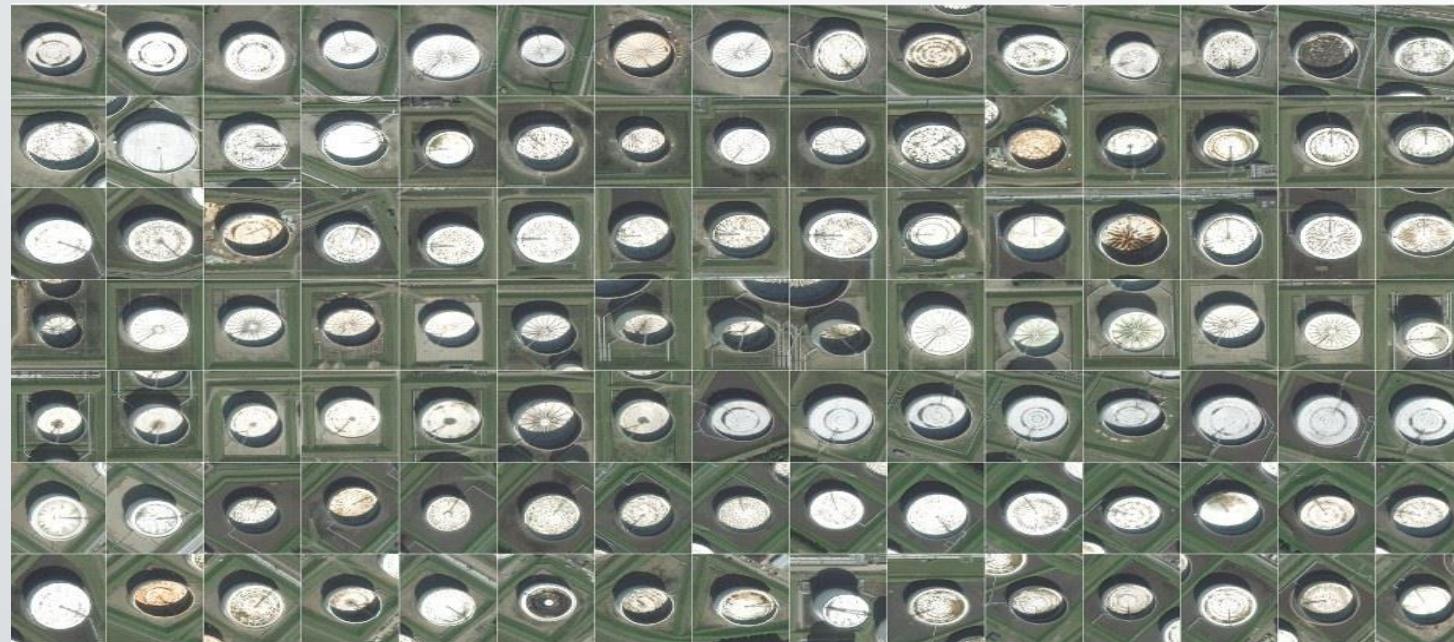


### Other Image Applications

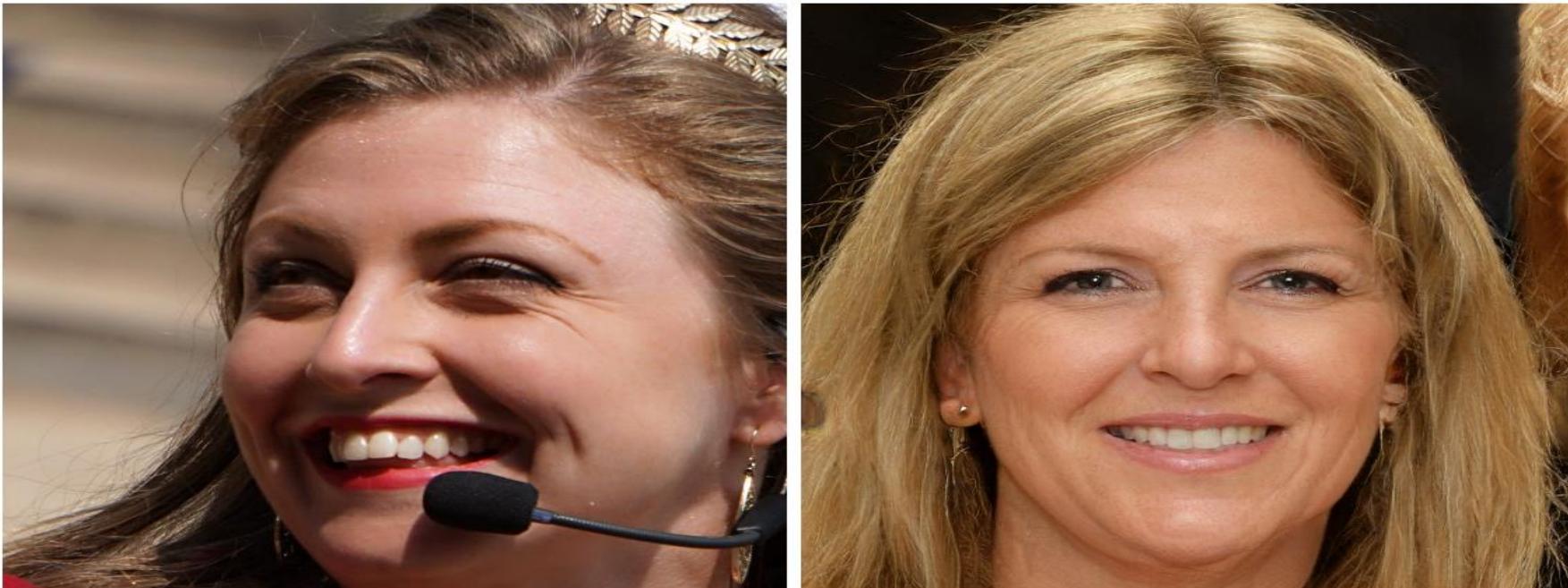
- [Theme park customer tracking via Ali-cloud](#)
- [Face Recognition from Megvii Face++](#)
- [Verifying real identity from Onfido](#)

# Use Cases - Satellite image

Oil reserves are often stored in floating tanks around the world, whose lid rises and falls with increasing or decreasing storage. With the rising and falling height, **the crescent-shaped shadows cast by the sun also change shape**. By **tracking these shape changes (Deep Learning - CNN)**, [Orbital Insights](#) predicts oil inventory levels.



# GANs (Generative Adversarial Networks)



StyleGAN: [Which face is real ??](#)

[AI Creates Fashion Model with Custom Outfits and Poses](#)

[Other GANs Applications](#)

# GANs (Generative Adversarial Networks)



Infernal Affairs 2020 GANs version

**Actors:** Andy Lau & Tony Leung ??



# Q & A

