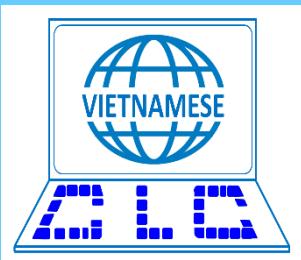


University of Science
Computational Linguistics Center
Natural Language Processing

Section 0:
Natural Languages Processing



Lecturer: Assoc.Prof. Dr. Dinh Dien

ddien@fit.hcmus.edu.vn

Lecturer

Đinh Điện 丁田

Dinh Dien
Динх Диэн

딘 디엔

ディンディエン



Teaching Assistant: Dr. Buu Long – Dr. Tung – Dr. Vinh

Natural Language Processing (NLP)

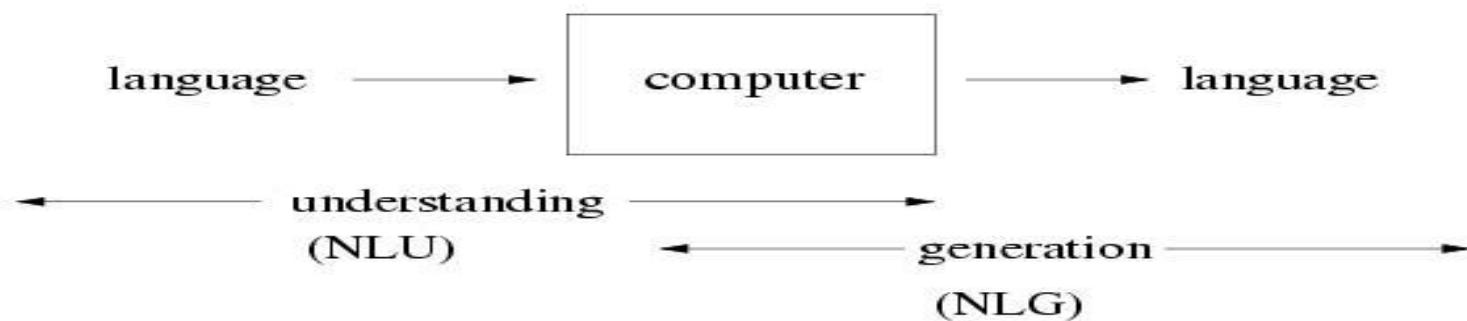
1. What is NLP ?
2. Why NLP is hard?
3. NLP applications
4. What will this course be about?
5. Textbooks - Website - dataset
6. Grading:
 - Mid-term: Exercise/Seminar (50%): Corpus
 - Project (Model): oral examination (50%)
 - Bonus: unlimited

What is NLP?

Using computer (Artificial Intelligence) to deal with human languages.

What is Natural Language Processing?

computers using natural language as input and/or output



Why NLP is hard?

- Reason (1) – human language is **ambiguous**:
- Ex1 (pronoun resolution):
 - Jack drank the wine on the table. *It* was red and round.
 - Jack saw Sam at the party. *He* went back to the bar to get another drink.
 - Jack saw Sam at the party. *He* clearly had drunk too much.
- Ex2: PrePosition Attachment:
 - I ate **the bread with** pecans.
 - I **ate the bread with** fingers.

Why NLP is hard?

- Reason (2) – requires reasoning beyond what is explicitly mentioned (*A,B*) , and some of the reasoning requires world knowledge (*C*).
- *Ex: I couldn't submit my homework because my horse ate it.*

Implies that...

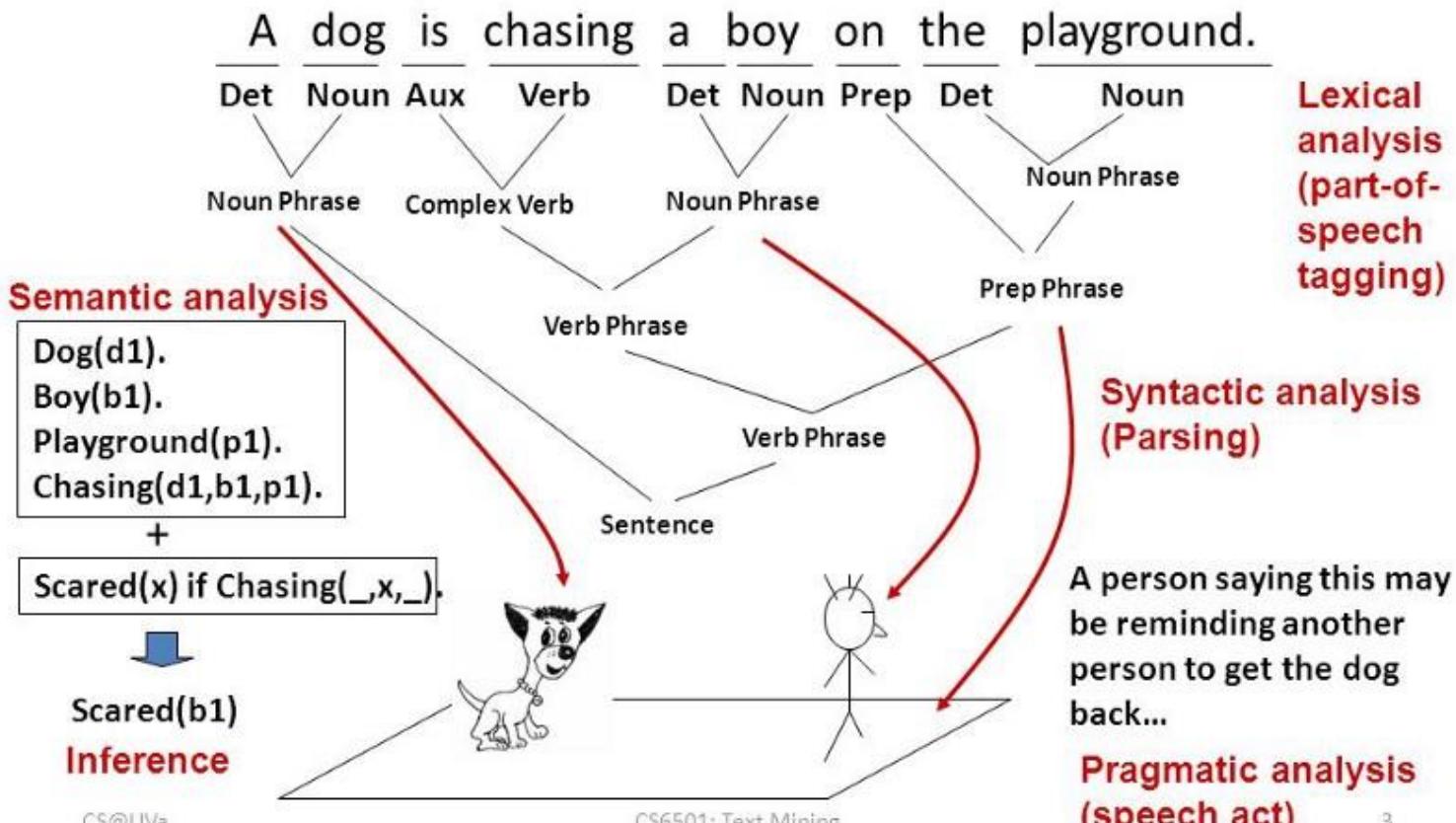
- *A. I have a horse.*
- *B. I did my homework.*
- *C. My homework was done on a soft object (such as papers) as opposed to a hard/heavy object (such as a computer). – it's more likely that my horse ate papers than a computer.*

NLP APPLICATIONS

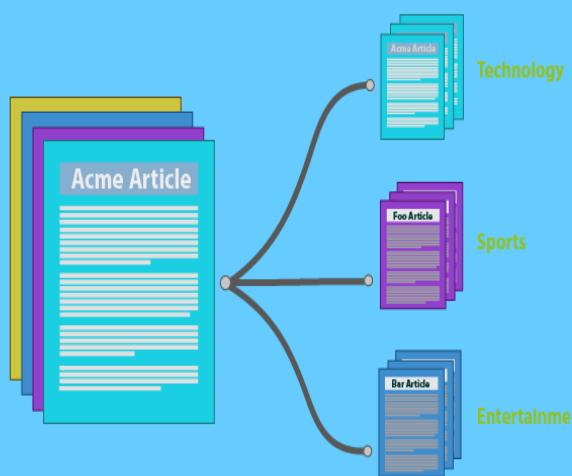
1. Linguistics

analysis:

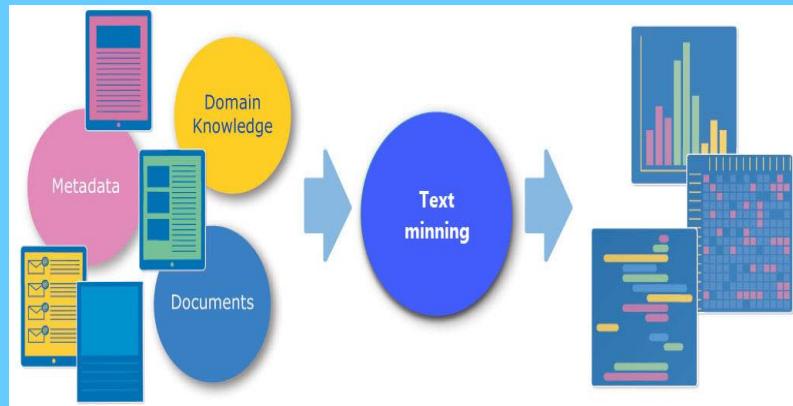
An example of NLP



2. Text classification:

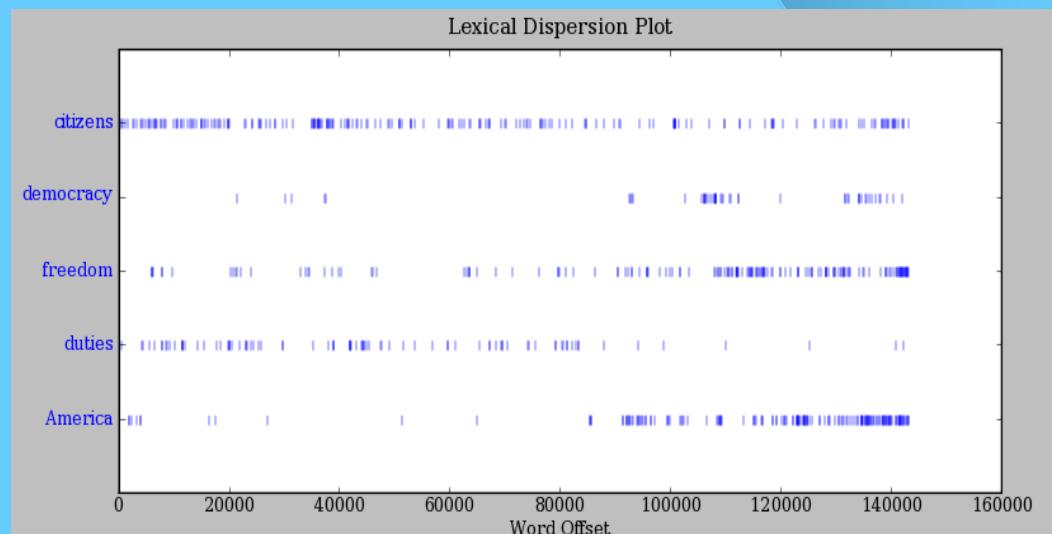
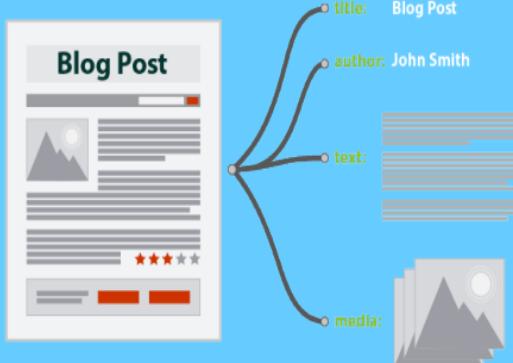


Web Mining



After 9-11: “We”> “I”

3. Text mining:



Information Extraction & Sentiment Analysis



Attributes:
zoom
affordability
size and weight
flash
ease of use

Size and weight

- ✓ nice and compact to carry!
- ✓ since the camera is small and light, I won't need to carry around those heavy, bulky professional cameras either!
- ✗ the camera feels flimsy, is plastic and very light in weight
- ✗ you have to be very delicate in the handling of this camera



ChatGPT: <https://duongqdaynong.tphcm.gov.vn/trang-chu#>

DU

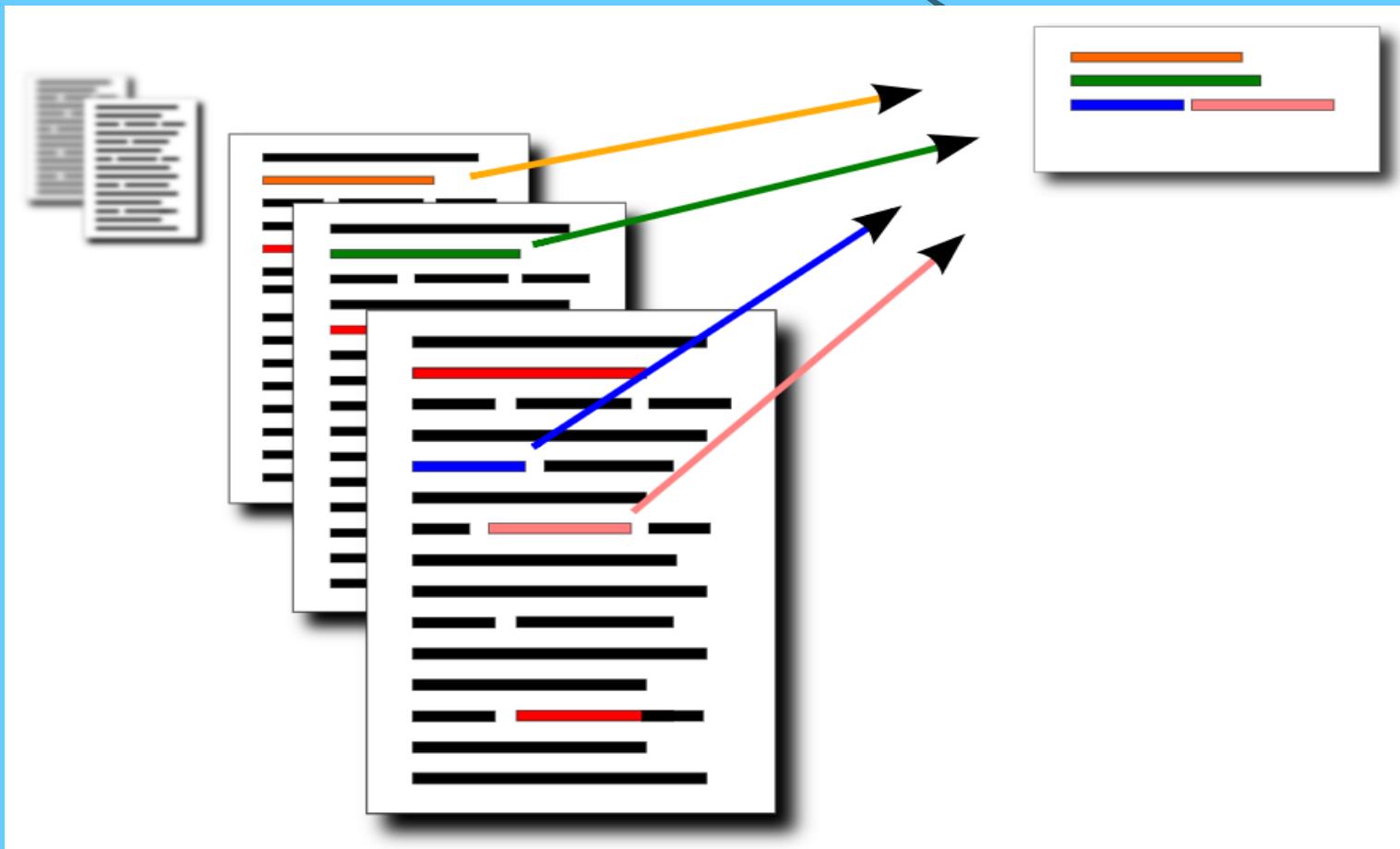
Hãy phân loại các đánh giá sau là tích cực hay tiêu cực:

1. Cán bộ trực ban giao tiếp rất khó chịu, giọng điệu hằn học. Gần 14h chiều ngày thứ ba nhưng bàn trực ban không có ai, 3-4 người dân phải ngồi đợi.
2. Tại địa chỉ số 215/36C, bãi xe Nam, Nguyễn Xí, Phường 13, Quận Bình Thạnh, nhà lấn chiếm hết hẻm làm bãi giữ xe, xây nhà và trổ cửa trái phép lấn ra sông Vầm Thuật.
3. Việc đốt (nhóm) bếp than vào mỗi sáng sớm từ 3 đến 4 giờ sáng của hộ gia đình 144/10 đường Lê Văn Thọ, phường 11, quận Gò Vấp rất độc hại. Gây ảnh hưởng âm thầm đến sức khỏe của rất nhiều người sinh sống xung quanh.
4. Về việc gần tết, các cán bộ không hề lơ là trong công việc và luôn có mặt ở cơ quan trong giờ làm việc hành chánh để hỗ trợ người dân.
5. 133 Nguyễn Tiểu La, Phường 5 Quận 10, nhà cho thuê để các bao xi măng, bao cát, đá lấn chiếm lòng lề đường mấy năm nay. Người dân có liên hệ UBND Phường nhưng không được giải quyết triệt để.



1. Tiêu cực: Cán bộ trực ban không được hỗ trợ và cách hành xử của họ rất khó chịu.
2. Tiêu cực: Ngôi nhà lấn chiếm hẻm, làm bãi giữ xe, xây nhà và trổ cửa trái phép, lấn ra sông Vầm Thuật.
3. Tiêu cực: Hộ gia đình đốt (nhóm) bếp than vào mỗi sáng sớm từ 3 đến 4 giờ sáng, gây ảnh hưởng đến sức khỏe của nhiều người.
4. Tích cực: Các cán bộ không lơ là trong công việc, luôn có mặt ở cơ quan và hỗ trợ người dân trong giờ làm việc hành chánh.
5. Tiêu cực: Nhà cho thuê bị lấn chiếm lòng lề đường mấy năm nay bởi các bao xi măng, bao cát, đá và việc này chưa được giải quyết triệt để bởi UBND Phường.

4. Text summarization:



ChatGPT: text summarization

DU

Hãy tóm tắt Nghị định số 72/2013/NĐ-CP ngày 15/7/2013 về Quản lý, cung cấp, sử dụng dịch vụ Internet và thông tin trên mạng;

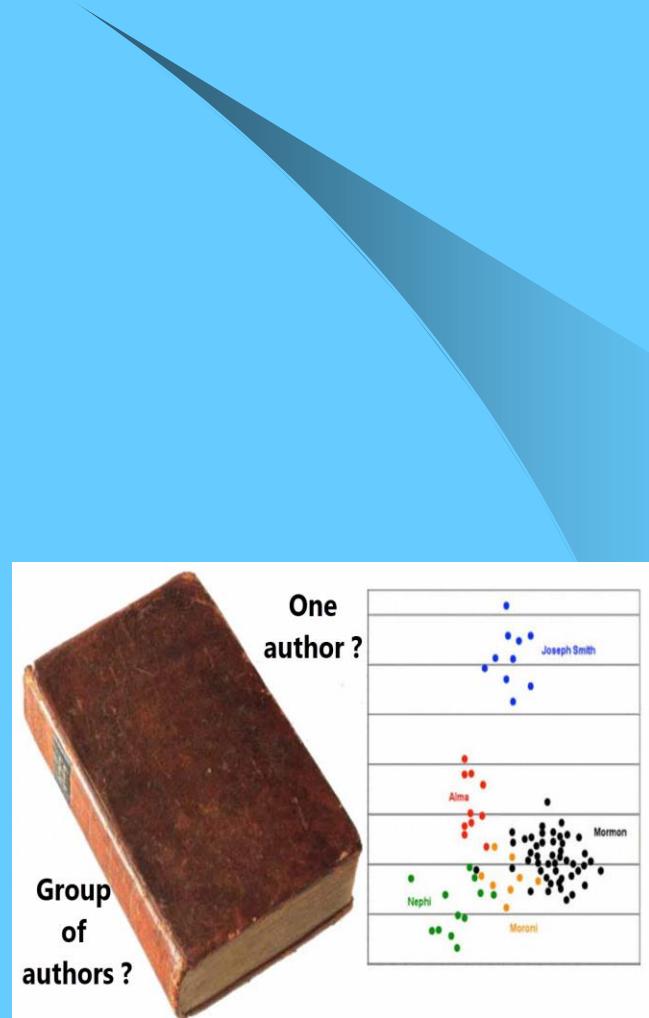
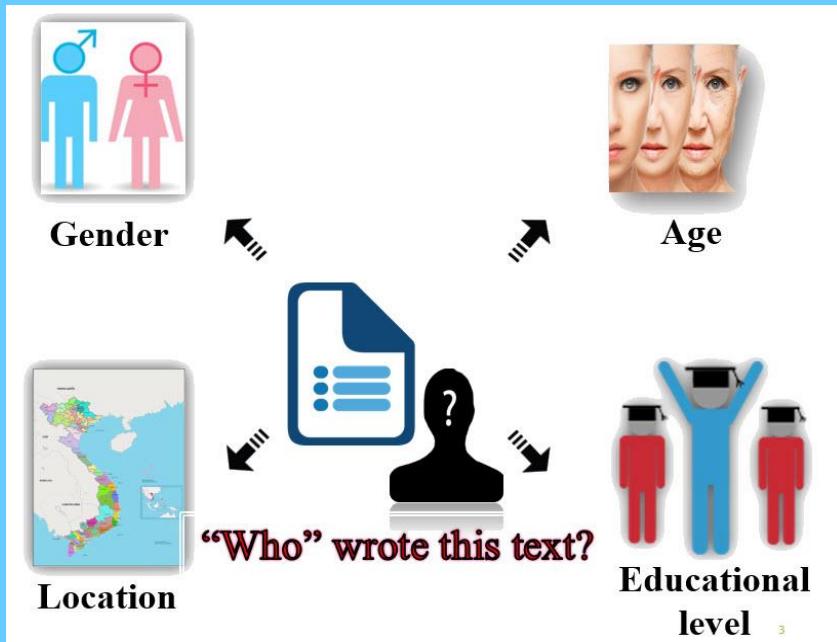


Nghị định số 72/2013/NĐ-CP được ban hành ngày 15/7/2013 về Quản lý, cung cấp, sử dụng dịch vụ Internet và thông tin trên mạng có các điểm chính như sau:

1. Đối tượng áp dụng: các tổ chức, cá nhân cung cấp, sử dụng dịch vụ Internet và thông tin trên mạng.
2. Quy định về việc sử dụng Internet và thông tin trên mạng, bao gồm quản lý, bảo mật, và đảm bảo an toàn thông tin.
3. Quy định về cung cấp dịch vụ Internet, quyền và nghĩa vụ của nhà cung cấp, cơ chế quản lý và giám sát hoạt động của nhà cung cấp dịch vụ Internet.
4. Quy định về trách nhiệm của các tổ chức, cá nhân sử dụng dịch vụ Internet và thông tin trên mạng, bao gồm việc bảo vệ quyền lợi của người dùng và trách nhiệm trong việc cung cấp thông tin trên mạng.
5. Quy định về kiểm tra, giám sát hoạt động sử dụng Internet và thông tin trên mạng.
6. Quy định về xử lý các hành vi vi phạm pháp luật khi sử dụng Internet và thông tin trên mạng.

Nghị định này nhằm đảm bảo an toàn, bảo mật và phát triển bền vững của dịch vụ Internet và thông tin trên mạng, đồng thời đảm bảo quyền lợi và trách nhiệm của các tổ chức, cá nhân sử dụng dịch vụ này.

5. Text stylometry:



Văn phong khác biệt tố cáo vụ ngộ sát, làm giả thư tuyệt mệnh

Qua phân tích cách dùng từ "and", "but", "hopefully", "truly" trong thư tuyệt mệnh, các chuyên gia tại Mỹ xác định nạn nhân không phải là người viết.

- Thói quen lạ của chú chó tố cáo ông chủ vứt xác bốn cô gái bán dâm

Vào buổi sáng năm 1992, khoa cấp cứu một bệnh viện tại Mỹ nhận được cuộc gọi khẩn cấp từ người sống tại căn hộ ở Bắc Carolina. Khi đến nơi, các nhân viên y tế thấy một thanh niên đã tử vong.

Nạn nhân được xác định là Michael Hunter, 23 tuổi, vừa tốt nghiệp đại học và đang làm lập trình viên. Bạn cùng phòng khai với cảnh sát rằng sáng hôm ấy, khi đánh thức Michael Hunter dậy để đi làm thì thấy anh ta bất tỉnh từ bao giờ.

Michael Hunter không có thương tích khả nghi nào trên cơ thể. Xét nghiệm máu cho kết quả dương tính với một loại thuốc gây tê với nồng độ gây chết người. Thông thường, loại thuốc này được sử dụng trong một số trường hợp khẩn cấp để làm ổn định nhịp tim. Tuy nhiên, nhân viên y tế khẳng định khi đến nơi thì thấy nạn nhân đã tử vong và họ không hề tiêm bất cứ thuốc gì.

Cái chết của Michael Hunter làm gia đình anh suy sụp. Cha của anh vì quá đau buồn đã rơi vào cơn trầm cảm kéo dài và tự tử sau đó.



FBI Profiler Says Linguistic Work Was Pivotal In Capture Of Unabomber

August 22, 2017 · 12:18 PM ET

Heard on Fresh Air

DAVE DAVIES

FRESH AIR



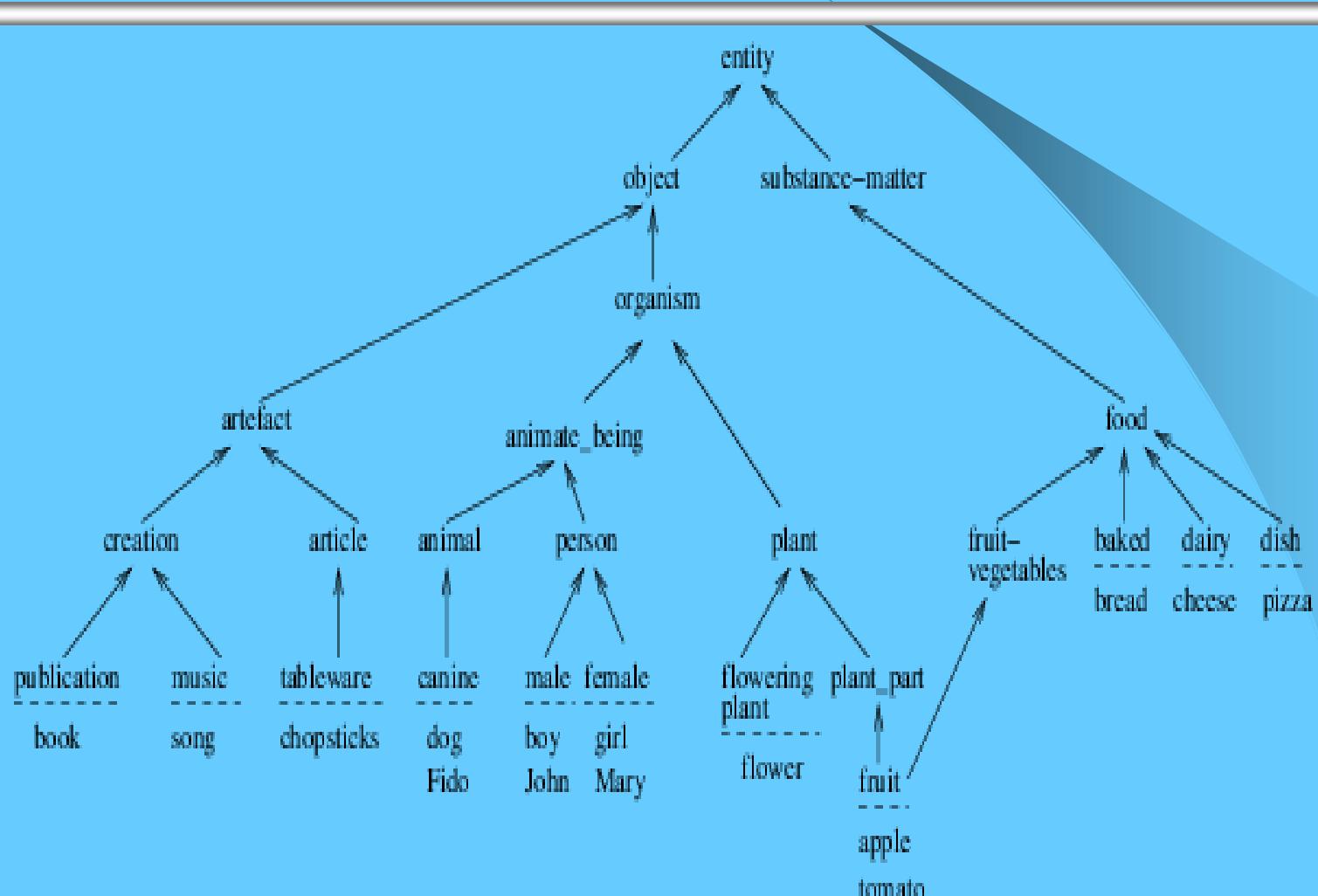
Ted Kaczynski is flanked by federal agents as he is led to a vehicle. Kaczynski is now serving a life sentence in prison for the

his victims. In 1995, he sent a sprawling, 35,000-word "manifesto" to *The New York Times* and *The Washington Post*, in which he explained why he believed technology to be evil and how society should disband the technological system and live in agrarian tribes.

Ex-Math-Prof.
UC Berkeley

Fitzgerald says the Unabomber's writings were a "pivotal factor" in cracking the case. He and his colleagues used them to help pinpoint the age and geographic origin of their suspect — evidence that helped lead to the April 6, 1996, arrest of Ted Kaczynski.

6. Text similarity:



Plagiarism detection

noplag

Title: Health Vision(1)
Author: Aleks B

100%
Similarity
43 Matches
en Language

You have not seen your eye doctor for more than a year.

What eye problems your eye doctor is looking for?

- ? Nearsightedness, farsightedness or astigmatism. These conditions are corrected with eyeglasses, contact lenses or surgery.
- ? Amblyopia and strabismus. Amblyopia occurs when eyes are misaligned. Strabismus is another word for crossed eyes.
- ? Focusing problems and ability of your eyes to work together.
- ? Any problems with eye tearing.
- ? Eye diseases such as glaucoma and diabetic retinopathy which have no clear symptoms at early stages. In most cases, early detection can reduce risk for vision loss.
- ? Age-related conditions. For example, cataracts occur mostly at the age of 65 and older.

What can you do to protect your eyes?

- ? Have a healthy diet, rich in fruits and vegetables.
- ? Take care of your health in general.
- ? Maintain a healthy weight.
- ? Quit smoking.
- ? Remember to give your eyes a rest when working at the computer.
- ? Do not forget to blink.
- ? Keep your eyes safe when playing sports or doing any potentially eye-dangerous activity.
- ? Protect your eyes from ultraviolet rays with sunglasses.
- ? Know your family's eye health history. Many eye diseases and conditions are hereditary.
- ? Visit your eye doctor once a year. Conducting regular eye exams will help preserve your vision and reduce risk of serious eye and vision problems.

Originality report Powered by Noplag.com

Page 1

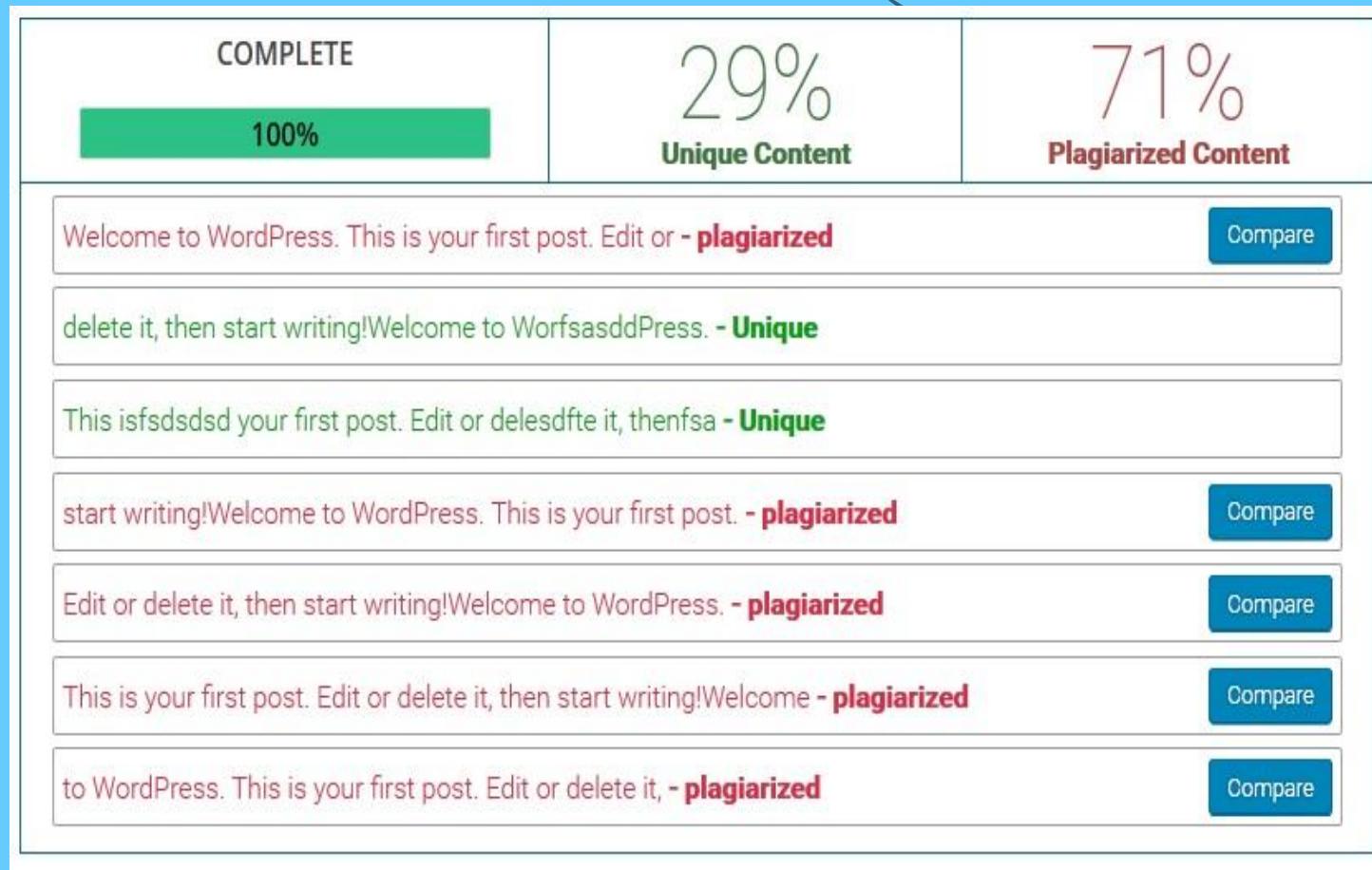
Match Overview

Rank	Type	Similarity
1	My Library	W 364 100%
2	Web	W 345 95.40%
3	Web	W 345 95.40%
4	Web	W 345 95.40%
5	Web	W 345 95.40%
6	Web	W 39 10.93%
7	Web	W 33 9.68%
8	Web	W 27 7.91%
9	Web	W 23 6.69%

ID 468042 Checked on 04 Nov 2016 5:42 PM

Words: 364 Pages: 1/1

Cross-Lingual Plagiarism detection



7. Text readability:

Earthquake in Indonesia – level 1



02-10-2018 07:00

Level 1

Level 2

Level 3

Sulawesi is an island in Indonesia. An **earthquake** hits near it. The earthquake makes a **tsunami**. It is 3 metres tall.

The tsunami moves into two cities. Around 600,000 people live there. More than 832 people die. Hospitals, hotels, a shopping centre, and thousands of homes are **destroyed**.

Difficult words: **earthquake** (when the ground moves), **tsunami** (a big wave started by an earthquake), **destroy** (break completely).

Earthquake in Indonesia – level 2



02-10-2018 07:00

Level 1

Level 2

Level 3

A 7.5-magnitude earthquake hit near the Indonesian island of Sulawesi which triggered a 3-metre tsunami that smashed into two cities on the coast. These cities are home to 600,000 people.

The tsunami killed more than 832 people and destroyed hospitals, hotels, a shopping centre, and thousands of homes. The event affected the lives of as many as 1.6 million people.

Difficult words: **magnitude** (the size of power of something), **trigger** (start suddenly), **smash** (move into with a lot of force).

Earthquake in Indonesia – level 3



02-10-2018 07:00

Level 1

Level 2

Level 3

A 7.5-magnitude earthquake hit near the Indonesian island of Sulawesi, triggering a 3-metre tsunami, which smashed into two cities on the coast.

Palu and Donggala are the cities affected the worst, and they are home to over 600,000 people. At least 832 people have been confirmed dead, thousands of homes collapsed, along with hospitals, hotels, and a shopping centre. The disaster affected as many as 1.6 million people, according to Red Cross estimates.

Difficult words: trigger (start), estimate (a careful guess based on data).

Linguistic features of Text readability

Word popularity: word usage frequency

Syntactic structure: complexity of parsing tree

Text organization: text coherence

Text readability <> comprehensibility

Writer (encoder) <> Reader (decoder)



Ex: A top-3,000 wordlist in English has been used in all definitions/explanations in the Oxford OALD8,

e.g. **phil·an·throp·ist** /fɪ'lænθrəpɪst/ noun a rich person who helps the poor and those in need, especially by giving money •nhà từ thiện, mạnh thường quân

Whilst, in an existing Vietnamese dictionary: the definition of the word “đường” (sugar) is “một hợp chất kết tinh...” (“hợp chất” = compound, “kết tinh” = crystallize”).

Ex: “tòa” = “kiến trúc đơn nguyên trong xây dựng”

Should not use difficult words: “gà qué” (35.216), “con ngóe” (23.670), ... in grade-1 textbooks.

MS word\proof reading: available for English

HubSpot Blogs - Marketing scanned on 14 Apr 2015 | Run new scan | New folder | More ▾

Summary Clear Language Links Spelling Bad Language Good Language Discovery Activity Discussions

HUBSPOT BLOGS - MARKETING Clarity Grader Report

Url Scanned: <http://blog.hubspot.com/marketing>

The Clarity Grader report analyzes this site for **clear, transparent** language.
We also check for **consistent language** using customizable bad and good language dictionaries.

1 PAGES SCANNED ON 14 APRIL 2015

[Tweet Report](#) [Email Report](#) [PDF this Report](#)



Clear Language

Long Sentences 71 Sentences 25.27%	Average Sentence Length 14	Passive Language 9 Sentences 3.20%	Readability 62
---	--------------------------------------	---	--------------------------

Aim for 5% or lower

Long sentences exceed 20 words. At 25.27% your content is 5.1 times the recommended level of 5%. The message is likely buried in complex statements and run on sentences. Split the long sentences or use lists.

Aim for 10 or lower

The average sentence length is fair at 14. For web copy you should aim for 10 or less. You may be burying certain key messages.

Aim for 5% or lower

The passive voice % is good at 3.20%, Well done! Your text is punchy and active. This means readers can easily absorb your message and follow instructions.

Aim for at least 60

Great. Your **readability** score is above 60. Your message is clear and readers can easily follow instructional text.

8. Text translation:

8a. Machine Translation

The screenshot shows the Al Jazeera English website interface. At the top, there's a banner with Arabic text "الجزيرة نت" and "أخبار مميزة". Below it, a large headline in Arabic reads "اشترى الآن الأخبار الساخنة أينما تكون" (Buy hot news anywhere). The main content area features several news stories with thumbnail images and Arabic subtitles. One story discusses a raid in Gaza, another about Bashir meeting Fraser, and others about Rumsfeld, Cheney, and a suicide attack in Afghanistan. The bottom of the page includes a footer with links like "frontiers@aljazeera.net", "الفضائية الفرنسية", and "الفضائية الفرنسية".

Killing Palestinians and wounding nine in the raids Sector

Nine Palestinians were wounded among civilians in an Israeli air raid in the neighborhood result in the Gaza Strip. This comes immediately after the killing of two prominent Al-Aqsa Martyrs Brigades in the Israeli occupying forces carried out air and infantry forces in the Balata camp in the West Bank.



Bashir meets Fraser, the Security Council will not impose forces Darfur

Is scheduled to meet with Sudanese President Omar al-Bashir Jenday Fraser Assistant Minister for Foreign Affairs of the American attempt to persuade officials in Khartoum, Sudanese Darfur deployment of the nationalities. For his part, US Ambassador to the United Nations that it has no intention of the Security Council to impose its forces in the province.



Rumsfield and Cheney insist on keeping the American forces in Iraq

Called American Defense Minister Donald Rumsfield Americans to show patience on Iraq. I take Vice President Dick Cheney calls Democrats withdrawal of American forces from Iraq link and the possibility of early withdrawal of attacks inside the United States.



Killing civilians and wounding officer suicide attack in Afghanistan

The international force to help establish security (ISAF) killed civilians and the wounding of an officer in an attack against Afghan forces convoy south Atlantic Afghanistan. In the capital Kabul, a hand grenade exploded at the passage of manufacture French patrol was not reported injuries or damage.



ChatGPT translation

DD

Hãy dịch câu sau sang tiếng Anh, Pháp, Hoa, Nhật, Hàn, Đức, Nga, Latin và Quốc tế ngữ: "Chúng tôi đang học một ngôn ngữ"



ChatGPT translation

DD

Hãy dịch câu sau sang tiếng Anh, Pháp, Hoa, Nhật, Hàn, Đức, Nga, Latin và Esperanto: "Chúng tôi đang học một ngôn ngữ"



- Tiếng Anh: "We are learning a language"
- Tiếng Pháp: "Nous apprenons une langue"
- Tiếng Hoa: "我们正在学习一门语言" (Wǒmen zhèngzài xuéxí yī mén yǔyán)
- Tiếng Nhật: "私たちは言語を学んでいます" (Watashitachi wa gengo o manande imasu)
- Tiếng Hàn: "우리는 언어를 배우고 있습니다" (Urineun eoneoreul baeugo issseubnida)
- Tiếng Đức: "Wir lernen eine Sprache"
- Tiếng Nga: "Мы изучаем язык" (My izuchayem yazyk)
- Tiếng Latin: "Nos discimus linguam"
- Esperanto: "Ni lernas lingvon"

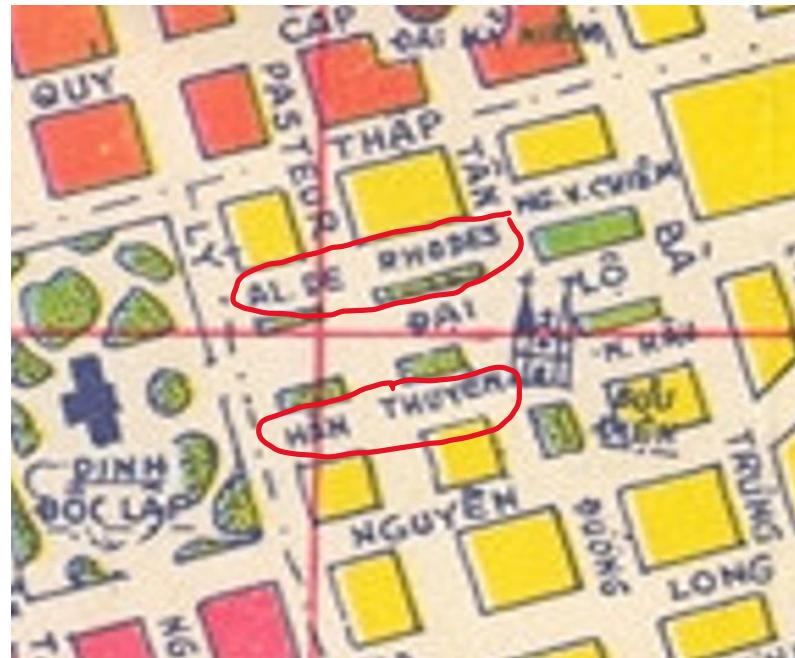


ChatGPT: cannot process Nôm-scripts

- “nǎm” (5): 酉 = 南 (sound /nán/ “nam”) + 五 (meaning: *ngũ*)
- “nǎm” (year): 酉 = 南 (sound /nán/) + 年 (meaning: *niên*)
- “中” => “đúng, trong, trung, trúng, truồng”
- 穀 = “trăm/trặn” ;



Hàn
Thuyên



Alexandre de
Rhodes

Sino-Nôm text decoding



Vua Quang Trung (ảnh minh họa).

車: xa (xe)
心: tâm (tim)
惠: huệ (ơn)

車 心 折 軸 多 田 鼠

XÂ
TÂM
CHIẾT
TRỰC,
ĐÀ
DIỀN
THỦ



Theo sách Hoàng Lê nhất thống chí ghi lại thì khi vua Quang Trung giả làm sứ giả sang Bắc Kinh (Trung Quốc) gặp vua Càn Long, được vua Càn Long tặng cho chiếc áo, có thêu 7 chữ: Xa tâm chiết trực, đà điền thủ. Nghĩa là: Bụng xe gãy trực, nhiều chuột đồng.

Theo phép chiết tự, chữ "xa" và chữ "tâm" ghép lại thành chữ "Huệ" là tên của Nguyễn Huệ; "chuột" nghĩa là năm Tý (Nhâm Tý 1792). Ý của dòng chữ trên áo là Nguyễn Huệ sẽ chết vào năm Tý. Giả thiết này ý nói rằng, vua Quang Trung chết do áo bị yểm bùa?



季 群 底 固 故 駁

Trời còn để có hôm nay,

散 霜 頭 午 援 遽 駁 季

Tan sương đầu **ngõ**, vén mây giữa trời.

nomfoundation.org/nom-tools/Nom-Lookup-Tool/Nom-Lookup-Tool?uiLang=vn



喃 產 遺 存 保 會

English

Thông tin về Hội ▼ Chữ Nôm ▼ Sách Nôm ▼ Từ điển Nôm ▼ Dự án Nôm ▼ Phòng chữ Nôm ▼ Liên hệ

Another Nôm Lookup Tool
based on Unicode

- Quốc Ngữ hoặc Hán-Nôm Mã Unicode hoặc TCVN (dùng hệ hex) Tiếng anh
 Bắc kinh Quảng đông Thương hiệt
 Bộ thủ.Tổng nét

TRA CỨU CHỮ NÔM

Tra cứu chữ Nôm

Câu chuyện về tra cứu Nôm

午

GO

Quốc Ngữ	Hán-Nôm	Context	Ref.	Tiếng anh
ngõ	午	cửa ngõ	btcn	gateway
ngọ	午	giờ ngọ	vhn	midday, noon
ngọ	午	ngó ý	btcn	to express a wish

Automatically translating Sino-Nôm into National Scripts

Truyện Kiều

慕辭沖 埽 得 些

Trăm năm trong cõi người ta

字 才 字 命 簿 羅 恕 饒

Chữ tài chữ mệnh khéo là ghét nhau

浪辭嘉靖朝明

Rằng: Năm Gia Tịnh triều Minh

眾方滂朗台京凭傍

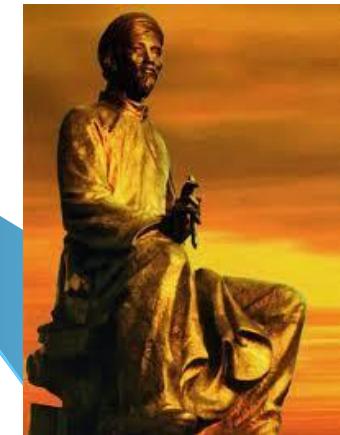
Bốn phương phảng lặng hai kinh vững vàng

固茹員外戶王

Có nhà viên ngoại họ Vương

家資擬拱常常塙中

Gia tư nghỉ cũng thường thường bậc trung



Nguyen Du
(1766-1820)

- *nghĩ* (think)
- *nghỉ* (he)

別 身 典 跳 落 類

金: kim

Biết thân đến bước lạc lõai

重: trọng

蕊 桃 他 披 朱 得 情 鍾

Nhị đào thà bỉ cho người tình chung

Theo Đại Việt Sử ký Toàn thư, tục truyền rằng vua Lý Thái Tổ khi mới được thiêng hạ, xá giá về Cổ Pháp, tại chùa làng Phù Đổng, có thần nhân đề thơ ở cột chùa rằng: "Nhất bát công đức thủy/ Tùy duyên hóa thế gian/ Quang quang trùng chiếu chúc/ Một ảnh nhật đăng san". Nghĩa là: "Một bát nước công đức của Phật, theo cơ duyên sinh thành thế gian. Vầng vặc hai lần được soi sáng; mặt trời gác núi là hết bóng". Sư Vạn Hạnh trụ trì chùa ấy đem bài thơ dâng lên vua, Lý Thái Tổ xem xong rồi nói: "Việc của thần nhân thị không thể hiểu được". Đến khi họ Lý mất ngôi mới cho bài thơ ấy là nghiệm, có người làm phép chiết tự cho rằng: tên ông là Sảm (昌), theo Hán tự thì bên trên là chữ nhật (mặt trời), bên dưới là chữ sơn (núi), chữ "Sảm" nghĩa là "mặt trời gác núi"; theo nghĩa đó mà suy thì đến đời Lý Sảm, ngôi vị của nhà Lý sẽ mất.

- Lý Thái Tổ, tên húy Lý Công Uẩn.
- Lý Thái Tông, tên húy Lý Phật Mã.
- Lý Thánh Tông, tên húy Lý Nhật Tôn.
- Lý Nhân Tông, tên húy Lý Càn Đức.
- Lý Thần Tông, tên húy Lý Dương Hoán.
- Lý Anh Tông, tên húy Lý Thiên Tộ.
- Lý Cao Tông, tên húy Lý Long Cán.
- Lý Huệ Tông, tên húy Lý Hạo Sảm.

昌: sảm

Lý Huệ Tông (chữ Hán: 李惠宗; tháng 7 năm 1194 – 3 tháng 9 năm 1226) là vị hoàng đế thứ tám của nhà Lý, cai trị từ năm 1210 đến năm 1224. Ông tên thật là Lý Sảm (李昌) hay Lý Hạo Sảm, sinh tại kinh đô Thăng Long, Đại Việt.



Nghiệm thu hệ thống tự động chuyển văn bản chữ Nôm sang chữ Quốc ngữ

① 14/02/2023 - 19:33

Chia sẻ **Ý KIẾN BẠN ĐỌC(2)**

Xem nhiều

Thị trường sinh 10



PNO - Các nhà khoa học của TPHCM đã xây dựng thành công hệ thống tự động chuyển văn bản chữ Nôm sang chữ Quốc ngữ.



- Chữ quốc ngữ 130 năm thăng trầm
- Lại bàn về việc cải tiến chữ quốc ngữ
- Bài 1: "Ông tổ thơ Nôm" đúng nhìn "cha đẻ chữ Quốc ngữ"

GSTS Võ Văn Sen

Công việc này có ý nghĩa rất lớn-Phải nói là hết sức vĩ đại!!! Cảm tạ nhóm tác giả!!!

Trả lời 1 thích 0 không thích

Nguyễn Ngọc Lan

Khâm phục
Tâm rất sáng, công rất lớn...

Trả lời 1 thích 0 không thích



Hội đồng nghiệm thu nhiệm vụ khoa học công nghệ

Sở Khoa học và Công nghệ TPHCM vừa tổ chức Hội đồng nghiệm thu nhiệm vụ khoa học và công nghệ "Xây dựng hệ thống chuyển tự từ văn bản chữ Nôm sang chữ Quốc ngữ". Đây là nhiệm



học sinh thiếu sóch
giáo khoa mỹ thuật



Tư vấn nguyện vọng
tuyển sinh lớp 10:
Đừng gieo căng thẳng
đôi bên



"Có thể học được kỹ
năng kiến tạo hạnh
phúc không?"

Tuyển sinh



Hà Nội chốt lịch thi vào
lớp 10 THPT, sớm một
tuần so với năm ngoái



Đầu tháng 3 sẽ công
bố mẫu đề thi tốt
nghiệp THPT



Hà Nội chốt thi 3 môn
vào lớp 10 công lập

Website: <https://tools.clc.hcmus.edu.vn>



https://tools.clc.hcmus.edu.vn

諺 嘴
chữ Nôm

CLC - Chuyển tự chữ Nôm

fit@hcmus



字

VĂN BẢN

TÀI LIỆU

HÌNH ẢNH

LỊCH SỬ

ĐÃ LƯU

Vietnamese

English

HÁN - NÔM

QUỐC NGỮ

嚮台閒事懶委
委箕包扒臥固身
扒風塵沛風塵
朱清高買特份清高
固兜偏為臥茚
笄才笄命灑澑奇卽
固才麻悞之才
笄才連貝笄灾沒韻
包芒惄業匱身
拱臂責吝委斯委賒
善根於在悉些
笄心算買朋匱笄才

ngẫm thay muôn sự bởi trời
trời kia đã bắt làm người có thân
bắt phong trần phải phong trần
cho thanh cao mới được phân thanh cao
có đâu thiên vị người nào
chữ tài chữ mệnh dôi dào cả hai
có tài mà cậy chi tài
chữ tài liền với chữ tai một vần
đã mang lấy nghiệp vào thân
cũng đừng trách lắn trói gần trót xa
thiện căn ở tại lòng ta
chữ tâm kia mới bằng ba chữ tài

206 / 500



Công trình này được thực hiện bởi một số giảng viên Khoa Công nghệ Thông tin và Trung tâm Ngôn ngữ học Tính toán
thuộc Trường ĐH Khoa học Tự nhiên - ĐHQG - HCM trong khuôn khổ đề tài nghiên cứu khoa học của Sở Khoa học & Công nghệ TpHCM



錦哈閉事在丕
扒風塵沛風塵
固亮偏為馴市
固才麻悞之才
奄芒祿業包身
善根於在惠些
丕圭扣捨容魄
丕箕伐扒夕馴固身
朱清高買特分清高
笄才笄命灑灑奇仁
笄才連貝笄灾汝韻
拱停責客丕貯丕賒
笄心笄買朋芒笄才
模濫拱特汝辨穀更

HÁN - NÔM

囉哈閒事在歪
歪箕也扒臥固身
扒風塵沛風塵
朱清高買特分清高
固兜爲趴節
筍才筍命瀧濶奇仁
固才麻悞之才
筍才連貝筍灾沒韻
也芒裯業趴身
拱停責吝歪斯歪賒
善根於在惹些
筍心箕買朋匪筍才

金雲超傳卷完		齋吟事在云 扒風塵沐風塵 固壳偏為馴市 固才麻悞之才 毫芒祿業鉤身 善根於在惡些 逐主招捨容魄	至箕色扒夕馴固身 朱清高買持分清高 芥才葑命雄灑奇仁 芥才連貝芥灰汝韻 拱停責客至斯至除 芥心箕買明三芥才 模盪摸特沒辨教更
--------	--	---	--

QUỐC NGŨ

X ngầm hay muôn sự tại trời
trời kia đã bắt làm người có thân
bắt phong trần phải phong trần
cho thanh cao mới được phần thanh cao
có đâu vì người nào
chữ tài chữ mệnh dồi dào cả hai
có tài mà cậy chi tài
chữ tài liền với chữ tai một vần
đã mang lấy nghiệp vào thân
cũng đừng trách lẩn trót gần trót xa
thiện căn ở tại lòng ta
chữ tâm kia mới bằng ba chữ tài

Outdoor-Image Translation



HÁN - NÔM

亭福清

福生重厚有财有土有人民

清化巍我乃聖乃神乃文武



QUỐC NGỮ



đinh phúc thanh



phúc sinh trọng hậu hữu tài hữu thổ hữu nhân dân
thanh hoá nguy ngã nãi thánh nãi thần nãi văn võ

Dịch nghĩa (thử nghiệm):

đinh phúc thanh

phúc đức trọng hậu đãi có tài có đất có người dân

thanh hoá nguy ta lại là thánh bèn thần bèn văn võ

VTV3

VTV.vn

06:40



Sino-Nôm text interpretation

汝等行看取敗虛
如何逆虜來侵犯
截然定分在天書
南國山河南帝居

Nam quốc sơn hà

Nam quốc sơn hà Nam đế cư,
Tiết nhiên định phận tại thiên thưⁱ.
Như hà nghịch lỗ lai xâm phạm,
Nhữ đẳng hành khan thủ bại hưng.

Dịch nghĩa

Núi sông nước Nam thì vua Nam ở,
Cương giới đã ghi rành rành ở trong sách trời.
Cớ sao lũ giặc bạo ngược kia dám tới xâm phạm?
Chúng bay hãy chờ xem, thế nào cũng chuốc lấy bại vong.

Ancient text translation



fit@hcmus



字 VĂN BẢN

TÀI LIỆU

HÌNH ẢNH

LỊCH SỬ

ĐÃ LƯU

Vietnamese

English

HÁN - NÔM



QUỐC NGỮ

南國山河南帝居，
截然定分在天書。
如何逆虜來侵犯，
汝等行看取敗虛

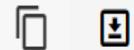
X nam quốc sơn hà nam đế cư ,
tiệt nhiên định phận tại thiên thư



như hà nghịch lỗ lai xâm phạm ,
nhữ đẳng hành khan thủ bại hư

Dịch nghĩa (thử nghiệm):
sông núi nước nam đế ở phía nam
cắt ngang nhiên định chia ở trên trời thư
như thế nào giặc đến xâm lăng
các ngươi những đi xem lấy bại hư

34 / 500



Ancient text translation & interpretation

HÁN - NÔM



QUỐC NGỮ

獄中無酒亦無花
對此良宵奈若何
人向窗前看明月
月從窗隙看詩家



ngục trung vô tửu diệc vô hoa
đối thử lương tiêu nại nhược hè
nhân hướng song tiền khan minh nguyệt
nguyệt tòng song khích khán thi gia



Dịch nghĩa (thử nghiệm):
trong tù không rượu cũng không hoa
trước cảnh đẹp đêm nay biết làm thế nào
người hướng ra trước song ngắm trăng sáng
từ ngoài khe cửa trăng ngắm nhà thơ

31 / 500



Ancient word/phrase explanation



Etymology (từ nguyên):

Cảnh nào cảnh chằng đeo sầu,
Người buồn, cảnh có vui đâu bao giờ!

Đòi phen nét vẽ, câu thơ,
Cung cầm trong nguyệt, nước cờ dưới hoa,

Vui là vui gượng kẹo mà,

Ai **tri âm** đó, mặn mà với ai?

Tri âm (知音) vốn là một từ ghép tiếng Hán được sử dụng rộng rãi trong tiếng Việt. Trong đó, từ “tri” có nghĩa là biết, còn từ “âm” mang nghĩa là âm điệu.

Du Bá Nha (Xuân Thu Chiến Quốc), người nước Sở, làm quan nước Tấn, có tài năng đánh đàn nổi tiếng. Vào ngày trên đường công tác trở về, Bá Nha nổi hứng muốn dừng lại ngắm trăng, đánh đàn.

Không ngờ lại gặp được Chung Tử Kỳ, một ẩn sĩ làm nghề tiêu phu sống bên sông, nhưng có khả năng cảm nhận được thấu đáo ý nghĩa âm thanh của mọi tiếng đàn Bá Nha đàn. Cả hai nhanh chóng bị ẩn tượng và khâm phục bởi tâm hồn và chí khí của người kia và kết thành bạn **tri âm** với nhau. Mùa thu năm sau, Bá Nha quay lại bến sông ngày ấy để tìm bạn mình thì hay tin Chung Tử Kỳ đã mất. Bá Nha đến mộ bạn mình thắp hương, gảy đàn rồi đập vỡ cây đàn, thề rằng từ giờ sẽ không bao giờ gảy đàn nữa vì không còn ai có thể cảm được tiếng đàn của ông nữa.

Classical reference (điển tích):

Đêm đêm **Hàn thực** ngày ngày Nguyên tiêu.

Muôn nghìn người thấy cũng yêu,
Xôn xao oanh yến rập rìu trúc mai.

Từ “hàn thực” là một từ dùng để chỉ một ngày Tết truyền thống của người Việt Nam và Trung Quốc, vào ngày mồng 3 tháng 3 âm lịch. Theo ý nghĩa mặt chữ, “hàn thực” có nghĩa là “thức ăn lạnh”, vì vào ngày này, người ta sẽ ăn đồ ăn nguội đã nấu sẵn.

Đời Xuân Thu, vua Tấn Văn Công nước Tấn gặp loạn nên phải lưu vong và có hiền sĩ Giới Tử Thôi theo phò. Trên đường đi, lương thực cạn kiệt, Giới Tử Thôi phải lén cắt một miếng thịt đùi mình nấu lên dâng vua. Vua ăn xong, hỏi ra mới biết đem lòng cảm kích vô cùng. Về sau, Tấn Văn Công dành lại được ngôi vương, phong thưởng cho những người có công nhưng lại quên Giới Tử Thôi. Giới Tử Thôi cũng không oán giận gì, nghĩ mình theo vua phò vua là chuyên nên làm. Ông đưa mẹ vào núi Điền Sơn ở ẩn. Tấn Văn Công về sau nhớ ra, cho người đi tìm Tử Thôi. Nhưng vì là người không tham danh vọng, Tử Thôi nhất quyết không quay về lĩnh thưởng, Tấn Văn Công ra lệnh đốt rừng (để ép Tử Thôi quay về). Không ngờ Tử Thôi quyết chí, hai mẹ con cùng chịu chết cháy trong rừng. Vua đau lòng, thương xót nên đã lập miếu thờ và hạ lệnh trong dân gian phải kiêng đốt lửa ba ngày, chỉ ăn đồ nguội nấu sẵn tưởng niệm Giới Tử Thôi.

Interpretation: knowledge of etymology, classical reference

幸遇 Hạnh ngộ May gặp	Trần Minh Tông	平生曾蓄一張琴 埋沒塵埃歲月深 幸遇伯牙高妙手 更教山水嗣徽音	bình sinh tàng súc nhất trương cầm mai một tràn ai tuế nguyệt thâm hạnh ngộ Bá Nha cao diệu thủ cánh giao sơn thuỷ tự huy âm	đời ta từng có một cây đàn chôn vùi đã lâu ngày trong bụi tràn may gặp được tay cao diệu của Bá Nha lại khiến cho tiếng đàn nổi được lời nước non
罕知音 Hân tri âm ít bạn tri âm	Tịnh Giới thiền sư	此時說道罕知音 只為如斯道喪心 奚似子期多爽慘 聽來一達伯牙琴	thử thời thuyết đạo hân tri âm chỉ vị như tư đạo táng tâm hè tự Tử Ký đa sảng sẩm thính lai nhất đạt Bá Nha cầm	lúc này bàn luận về đạo thì rất ít tri âm chỉ vì ngày nay đạo đã mất tâm mấy ai được như chung Tử Ký là người sành nhạc mới nghe qua đã hiểu thấu tiếng đàn của Bá Nha

Interpretation: differences Chinese and Sino-Vietnamese

除夕其四 Trừ tịch kỳ 4 Đêm trừ tịch kỳ 4	Nguyễn Khuyến	書齊不寐獨沉吟 滴滴銅龍夜轉深 遠近競推除夕鼓 寒喧誰記昔年心 盧陵悶甚憐蟲響 林仲閒多學鶯暗 坐待明朝是元旦 改觀萬像一橫襟	thư trai bất mị độc trầm ngâm trích trích đồng long dạ chuyển thâm viễn cận cạnh thôi trừ tịch cổ hàn huyên thuỳ ký tích niên tâm lư lăng muộn thậm liên trùng hưởng lâm trọng nhàn hạ học tiếng chim non hót toạ đai minh triêu thị nguyên đán cải quan vạn tượng nhất hoành khâm	trong phòng sách một mình không ngủ trầm ngâm đồng hồ nước điểm từng giọt đêm đã khuya xa gần rộn lên tiếng trống thúc giao thừa lúc hàn huyên ai nhớ được nỗi lòng những năm xưa lư lăng buồn quá thương giun dế kêu lâm trọng nhàn hạ học tiếng chim non hót ngồi đợi xem sáng mai là tết nguyên đán chỉ mới xoay ngang vật áo một cái mà mọi cảnh vật đổi thay
--	------------------	--	--	--

白藤江 Bạch Đằng giang Sông Bạch Đằng	Nguyễn Sưởng	鯨觀如山草木春 海潮 洶洶石嶙峋 誰知萬古重興業 半在關河半在人	kinh quán sơn hà thảo mộc xuân hải triều hùng hùng thạch lân tuân thuỷ tri vạn cổ trùng hưng nghiệp bán tại quan hà bán tại nhân	mồ chôn quân thù cao như núi cây cỏ xanh tươi nước thuỷ triều ngoài biển ầm ầm đá núi lởm chởm mấy ai biết sự nghiệp muôn thuở đời trùng hưng một nửa nhờ địa thế núi sông một nửa do con người
--	-----------------	--	--	---

Historical text mining: event relations

"Thượng Hoàng ban yến cho các quan ở Diên Hiền, yến chưa xong, **bỗng có sao chổi xuất hiện ở phương đông bắc, đuôi dài suốt trời**, Thượng Hoàng ra xem bảo: Ta xem sao chổi này rất sáng, đuôi rất dài, không phải là tai họa của nước ta, lệnh cứ dự xong yến. **Tháng 10 mùa đông năm ấy, vua Tống băng.**"

from Dai Viet Su Ki Toan Thu, during the reign of King Tran Thanh Tong

In 1264, a super comet appeared (the ancients saw it as **a sign of disaster from God**). At that time, a series of major events occurred in the world.

- In China, The emperor of the “Nam Tong” Dynasty, Tong Ly Tong, **died later that year**.
- At the same time, Pope Urban IV fell ill on the day the comet was first seen, and **died at the exact moment it disappeared on October 3, 1264**.

Bấy giờ, ở Thanh Hoa mưa dầm nhiều, nước lụt đến 7 lần, lúa má bị hạn nhiều, dân đói to.

Họ Mạc sách phong Vũ Thị Hoành là con gái của thư vệ sự vệ Cầm y Phú Sơn hầu Vũ Văn Khê làm chính phi.

Tháng 11, sao Chổi mọc, trở thẳng hướng đông nam, tua sáng dài đến 40 trượng, sắc hồng và tia ánh nhau, người người đều kinh hãi.

Tháng 12, ngày mồng 1, sao Chổi hết.

Xuống chiếu đổi năm sau làm Quang Hưng

Historical text mining: event relations

Ngày 25, vùng Thanh Hoa trời không mưa gió mà nước sông Mã bỗng dâng cao, thành Tây Đô bị ngập, ở ngoài sông, nước xoáy, chảy xiết như bão, cây cối lấp cả sông, các nhà ven sông phần nhiều bị trôi ra biển.

Mùa thu, tháng 7, ngày mồng 2, lụt lớn. Đêm ngày Quý Mão, mặt trăng có quầng đỏ, bóng sáng lờ mờ, sắc như phun máu. Các nhà chiêm tinh đoán là tai biến ấy ứng vào điểm bắc phi hậu chét. Ngày 25, lại lụt to. Trong 1 năm, lụt đến 7 lần.

Tháng 8, ngày mồng 5, sấm động trái thì.

Đại hạn.

Ngày 13, Thái quốc phu nhân là Trương Thị Ngọc Lãnh chết.

Ngày 17, dinh Yên Trường cháy lớn. Khi ấy, gió to, lửa mạnh, cháy lan cả phủ dinh, trại quân, giải vũ, phố xá đến vài nghìn nhà. Mây sắc đỏ che kín mặt trời, khói đen đầy trời, từ giờ Ngọ đến giờ Thân lửa mới tắt. Bà phi của Thái Vương là Nguyễn Thị Ngọc Bảo bị chết cháy, Tiết chế Trịnh Tùng ở lánh chỗ khác để tang. Truy tôn là Minh Khang Thái Vương Thái Phi.

Ngày 23, Thái bảo Hùng quận công chết.

Ngày 28, trong dinh Yên Trường nổ cơn lốc lớn đến hai dặm, người ngã, nhà tốc, đá bay cát cuộn, gãy cây sụt ngói.

Tháng 7 , nước sông dâng cao .

Tháng 8 , ngày rằm Nhâm_Thân , đêm có nguyệt_thực .

Tháng 9 , ngày mồng một Đinh Hợi , nhật_thực .

Hà_lệnh cho xứ đầu nguồn ở các huyện lấy 500 cây gỗ làm cọc , chia cắm ở các cửa_biển để phòng thuyền giặc vào cướp và làm kẽ chống giữ .

Tháng 10 , ngày Ất Hợi , cầu_vòng đỏ xuất_hiện ở phuong khôn .

Tháng ấy , Tiết_chế Trịnh_Tùng đem quân đánh phá các huyện Gia_Viễn , Yên_Hoá , Phụng_Hoá thuộc các phủ Trường_Yên , Thiên_Quan , đánh dẹp đến chợ_Rịa , sai làm cầu_phao , cǎ phá dinh của Tân quận_công của họ_Mạc .

Historical text mining: event relations

Mậu Dần, [Thái Bình] năm thứ 9 [978], (Tổng Thái Bình Hưng Quốc năm thứ 3). **Mùa xuân, tháng giêng, động đất.**

Mậu Tuất, /Úng Thiên/ năm thứ 5 [998], (Tổng Chân Tông Hằng, Hàm Bình năm thứ 1).
Mùa xuân, tháng 3, động đất 3 ngày.

Quý Ty, [Sùng Hưng Đại Bảo] năm thứ 5 [1053], (Tổng Hoàng Hựu năm thứ 5). **Mùa xuân, tháng giêng, ngày mùng 5, động đất 3 lần. Mùng 10, có mây không có mưa, rồng vàng hiện ở gác Đoan Minh. Bây tôi chúc mừng, duy có nhà sư Pháp Ngũ nói: "Rồng bay trên trời, nay lại hiện ở dưới là điềm không lành".**

Mậu Dần, /Hội Phong/ năm thứ 7 [1098], (Tổng Nguyên Phù năm thứ 1). **Mùa Thu, tháng 8, động đất.**

Đinh Hợi, /Long Phù/ năm thứ 7 [1107], (Tổng Đại Quan năm thứ 1). **Mùa hạ, động đất.**
Tháng 2, 1137 châu Nghệ An động đất, nước sông đỏ như máu. Công Bình sai Nội nhân hỏa đầu Đặng Khánh Hương về Kinh sư đem việc ấy tâu lên.....

From “Đại Việt Sử Ký Toàn Thư”

Bính ngọ, năm thứ 18 [1666], **mùa xuân, tháng 3**, ở Hồ Xá có động đất.

Át sửu, năm thứ 37 [1685], **mùa hạ, tháng 5**, ở Cam Lộ động đất. Ở Gia Lộc ngoại châu Bố Chính có động đất. Ở hai phủ Thăng Hoa và Quy Ninh động đất.

Tháng 9, Quảng Bình động đất năm 1736. Thanh Hoa và Thanh Bình động đất.
Nghệ An động đất.

From “Đại Nam Thực Lục”

~~Prophephy (sấm): Trạng Trình's oracles~~

"Đầu Thu gà gáy xôn xao

Trăng xưa sáng tõ soi vào Thăng Long"

“Đầu Thu”: Tháng 7 (âm lịch)

古: cỗ

=> Tháng 8 (dương)

月: nguyệt

“gà”: năm Dậu (1945)

胡: Hồ

“Trăng xưa” => “cỗ nguyệt”



Tượng đài Trạng Trình Nguyễn Bỉnh Khiêm.

=> Sự kiện CMT8: Bác Hồ đọc bản tuyên ngôn độc lập

"Biển Đông vạn dặm dang tay giữ

Đất Việt muôn năm vững trị bình

Cự ngao đới sơn

碧浸仙山徹底清 ·
 巨鰐戴得玉壺生 ·
 到頭石有補天力 ·
 著腳潮無卷地聲 ·
 萬里東溟歸把握 ·
 億年南極奠隆平 ·
 我今欲展扶危力 ·
 挽卻關河舊帝城 ·

Bích tẩm tiên sơn triệt đế thanh,
 Cự ngao đới đắc ngọc hồ sinh.
Đáo đầu thạch hữu bổ thiên lực¹,
 Trước cước trào vô quyển địa thanh.
 Vạn lý Đông minh quy bả ác,
 Úc niên Nam cực điện long bình.
 Ngã kim dục triển phù nguy lực,
 Văn khước quan hà cựu đế thành.

Non tiên ngâm tẩm nước trong xanh,
 Bầu ngọc đội nê, ngao lớn sinh.
 Đầu ngọc, vá trời còn sức đá,
 Chân đưa, lăng sóng chằng âm thanh.
 Biển Đông, vạn dặm quơ tay nắm,
 Nam cực, muôn năm vững trị bình.
 Ta muốn phù nguy ra sức giúp,
 Quan hà thu lại cựu kinh thành.



CLC Phiên dịch Hán Nôm

chữ Nôm

[VĂN BẢN](#) [TÀI LIỆU](#) [HÌNH ẢNH](#)

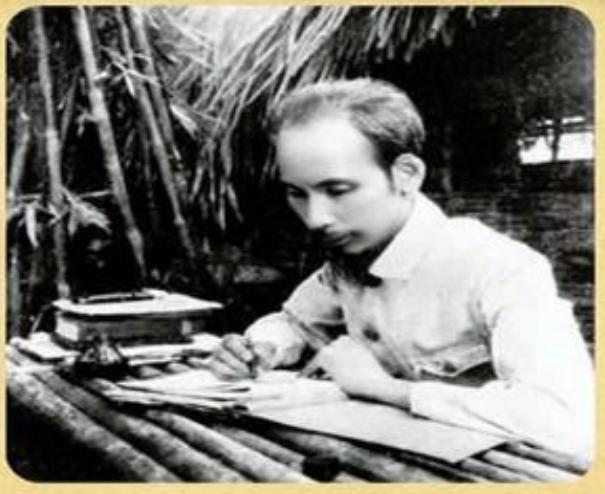
HÁN - NÔM	QUỐC NGỮ
<p>巨鰐戴山 碧浸仙山徹底清， 巨鰐戴得玉壺生。 到頭石有補天力， 著腳潮無卷地聲。 萬里東溟歸把握， 億年南極奠隆平。 我今欲展扶危力， 挽卻關河舊帝城。</p>	<p>cự ngao đới sơn bích tẩm tiên sơn triệt đế thanh , cự ngao đới đắc ngọc hồ sinh đáo đầu thạch hữu bổ thiên lực , trước cước trào vô quyển địa thanh vạn lý đông minh quy bả ác , úc niên nam cực điện long bình ngã kim dục triển phù nguy lực , văn khước quan hà cựu đế thành Dịch nghĩa (thử nghiệm): như con ngao lớn đội núi núi tiên nhuộm biếc cỏi trong suốt như con ngao lớn đội được bầu ngọc mà sinh ra ngoi đầu lên đá có sức vá trời đất chân xuống sống không có tiếng cuồn đất vạn dặm biển đồng qua vào tay nắm úc nắm cõi nam đặt vững cảnh trị bình ta nay muốn thi thổ sứ phò nguy cứu vãn lại quan hà thành cũ của nhà vua</p>



Historical text mining: named entity changes

"Dân ta phải biết sử ta,
cho tương gốc tích nước nhà
Việt Nam"

"Lịch sử nước ta" -
Chủ tịch Hồ Chí Minh



Thăng Long (7/1010)

từ thành Hoa Lư, dời đô ra kinh phủ ở thành Đại La, tạm đỗ thuyền dưới thành, có rồng vàng hiện lên ở thuyền ngự, nhân đó đổi tên thành gọi là thành Thăng Long. Đổi chau Cổ Pháp gọi là phủ Thiên Đức, thành Hoa Lư gọi là phủ Trường

Đông Đô (4/1397)

Lấy Phó tướng Lê Hán Thương coi phủ đô hộ lộ Đông Đô; Thái bảo Trần Hàng coi phủ đô thống lộ Bắc Giang; Trần Nguyên Trứ [Chú giải]



Đông Quan (12/1408)

Vua bao các quân:

"Hãy thừa thế chè tre, đánh cuốn chiếu thằng một mạch, như sét đánh không kịp bịt tai, tiến đánh thành Đông Quan [Chú giải] thì chắc chắn phá được chúng".

Đông Kinh (4/1427)

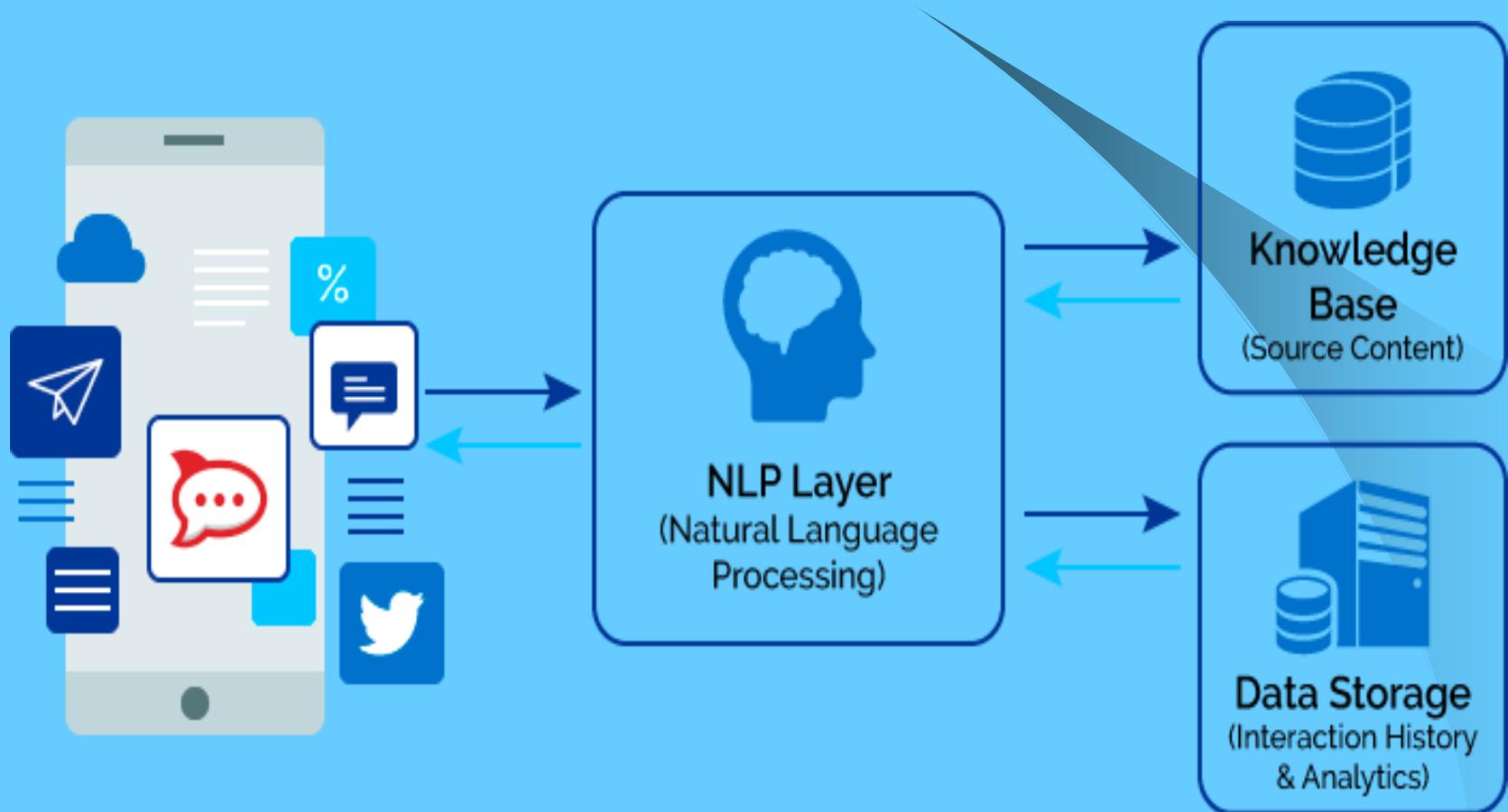
Mùa hạ, tháng 4, vua từ điện tranh ở Bồ Đào yết đóng ở thành Đông Kinh.

Ngày 15, vua lên ngôi ở Đông Kinh, đại xá, đổi niên hiệu là Thuận Thiên, dựng quốc hiệu là Đại Việt,

From Đại Việt sử ký toàn thư



10. Chatbot: ChatGPT et al.



ChatGPT: Giải đề thi môn Văn THPT QG-2023

DU

Chỉ ra những từ ngữ, hình ảnh miêu tả cơn giông mùa hè trong những dòng thơ sau:

Tiếng sấm gõ trên bầu trời thật thấp

Gió từ đất thổi lên rất猛

Cát bay, lá bay, đá bay



Trong đoạn thơ trích dưới đây, có những từ ngữ và hình ảnh miêu tả cơn giông mùa hè như sau:



- **Tiếng sấm gõ trên bầu trời thật thấp:** Hình ảnh của tiếng sấm gõ trên bầu trời thấp tạo ra một âm thanh mạnh mẽ và xa xa, mang đến sự cảm giác của cơn giông đang đến gần.
- **Gió từ đất thổi lên rất猛:** Miêu tả về gió từ đất thổi lên và cảm nhận của người viết khi gió mạnh đến mức rất猛, cho thấy sức mạnh và cường độ của cơn giông.
- **Cát bay, lá bay, đá bay:** Hình ảnh của các vật liệu như cát, lá và đá được cuốn đi trong giông bão, tạo nên cảm giác hỗn loạn và mạnh mẽ của cơn giông mùa hè.

Những hình ảnh này tạo ra một bức tranh sống động về cơn giông mùa hè với âm thanh, sức mạnh và sự hỗn loạn của nó.

~~Không chính xác: ngữ nghĩa từ Hán-Việt~~

DD

Hãy giải thích nghĩa của chữ "giai" trong thành ngữ "bách niên giai lão"



Trong thành ngữ "bách niên giai lão", chữ "giai" có nghĩa là "già nua", "đã già". "Bách niên" nghĩa là "trăm tuổi", "lão" có nghĩa là "người già". Vì vậy, "bách niên giai lão" được dịch là "trăm tuổi đã già nua" hoặc "trăm tuổi đời đã già".

Thành ngữ này thường được dùng để miêu tả một người đã sống đến tuổi rất cao, đánh giá đó là sự kiện đáng kính và đáng quý. Nó thể hiện sự tôn trọng và ngưỡng mộ đối với người già, đồng thời cũng nhắc nhở về sự quan trọng của sức khỏe, thời gian và đời sống. Ngoài ra, thành ngữ này còn có ý nghĩa đề cao lòng kiên trì, nghị lực, sức mạnh và tinh thần của con người, vì để sống đến độ tuổi trăm tuổi, cần phải có những phẩm chất đó.

giai (階): *giai cấp, giai đoạn, ...*

"bậc thềm"

giai (佳): *giai điệu, giai thoại, giai nhân, ...*

"hay/đẹp"

giai (偕): *giai lão, ...*

"cùng nhau"

~~Không chính xác: cả ngũ nghĩa từ thông dụng~~

Model: GPT-4

Talk-t

DD

Hãy giải thích ý nghĩa của chữ "trắc" trong từ "bất trắc"



Trong tiếng Việt, từ "bất trắc" thường được dùng để mô tả một tình huống không ổn định, khó lường, rủi ro hoặc gặp phải những khó khăn và vấn đề. Từ này bao gồm hai chữ "bất" và "trắc".

Chữ "trắc" trong từ "bất trắc" có nghĩa là **không ổn định, không chắc chắn**. Khi kết hợp với chữ "bất", nó nhấn mạnh ý nghĩa của sự không ổn định, rủi ro hoặc tình trạng khó lường. Ví dụ, bạn có thể dùng từ "bất trắc" để mô tả tình hình kinh tế, chính trị, thời tiết, hoặc một mối quan hệ nào đó.

Tóm lại, ý nghĩa của chữ "trắc" trong từ "bất trắc" là sự không ổn định, không chắc chắn, phản ánh một tình huống không tốt hoặc khó khăn.

Trắc: đo/lường: **trắc nghiệm, quan trắc, sinh trắc, ...**
“Sông sâu khó đo, lòng người không lường được”

Chưa thể hiểu: nghĩa hàm ý, ẩn dụ, hoán dụ

2. "Thiều quang chín chục đã ngoài sáu mươi": "Thiều quang" có nghĩa là ánh sáng mờ nhạt, chỉ thời gian gần hoàng hôn. Số "chín chục" ở đây ám chỉ tuổi thọ trung bình của con người, còn "sáu mươi" là thời gian mà ánh sáng bắt đầu mờ dần, tượng trưng cho tuổi già. Như vậy, câu thơ này có thể hiểu là thời gian cuộc đời đã đi qua nhanh chóng, giống như cách mà ánh sáng của ngày chuyển mình vào hoàng hôn.

Tổng kết lại, hai câu thơ này mang ý nghĩa về sự vụt qua của thời gian, sự thay đổi từ mùa xuân sang hoàng hôn tượng trưng cho quá trình từ thanh xuân đi vào tuổi già. Đồng thời, chúng cũng nêu lên một cảm giác nuối tiếc về những thời gian đã qua nhưng vẫn phải đón nhận và chấp nhận sự thay đổi của cuộc sống.

*Ngày xuân con én đưa thoi => dệt nên Tấm lụa mùa Xuân (90 ngày)
60 ngày = đã qua 2 tháng => Thanh minh trong tiết Tháng Ba*

Không thể hiểu: Nghĩa hình tượng

2. "Phận liễu sao mà nảy nét ngang."

- "Phận liễu" là số phận tầm thường, không có gì nổi bật, giống như cây liễu nhỏ bé và mong manh.
- "Sao mà" là "làm sao" hoặc "làm cách nào".
- "Nảy nét ngang" là phát triển và đạt được sự kiên cường, mạnh mẽ trong cuộc sống.

Câu thơ này thể hiện sự trăn trở của tác giả về số phận của mình, liệu với một số phận tầm thường như cây liễu, có thể kiên cường và vượt qua khó khăn, thử thách trong cuộc sống hay không.

Nhìn chung, hai câu thơ này diễn đạt nỗi lòng của tác giả về sự trăn trở, tiếc nuối về số phận và duyên nợ trong cuộc đời. Tác giả nhấn mạnh rằng mỗi con người phải đối mặt với những khó khăn và thử thách trong cuộc sống, dù số phận có tốt hay không. Đồng thời, tác giả cũng đặt câu hỏi liệu có thể vượt qua những thử thách đó để trở nên kiên cường và mạnh mẽ hơn.

Đặc thù tiếng Việt: không hiểu chữ Nôm

NG

GPT có thể dịch giúp tôi câu "慕辭韙塲𠵼些" không



Rất vui làm việc cho bạn! Câu "慕辭韙塲𠵼些" trong chữ Nôm có thể dịch như sau: "Hãy cho tôi một vài."

Truyện Kiều: “*Trăm năm trong cõi người ta*”

Kết quả: “*Hãy cho tôi một vài.*” !?

Tiếng Anh: chính xác hơn

DD

Xin cho tôi biết nguồn gốc của từ "breakfast"



Từ "breakfast" được hình thành từ hai từ "break" và "fast". Trong tiếng Anh cổ, "to break" có nghĩa là "phá vỡ, làm gián đoạn", và "fast" có nghĩa là "đói khát" trong thời gian dài, thường là do tôn giáo hay y học. Vì vậy, "breakfast" có nghĩa là bữa ăn đầu tiên trong ngày, giúp phá vỡ trạng thái đói khát sau một đêm ngủ.

break: phá vỡ **fast:** ăn chay/nhịn ăn (do **đêm:** ngủ)

breakfast: bữa đầu tiên trong ngày mà phá chay

Tiếng Anh: hiểu rất tốt

DD

Xin cho biết nguồn gốc của từ "disaster" trong tiếng Anh?



Từ "disaster" trong tiếng Anh có nguồn gốc từ tiếng Ý, từ "disastro". "Disastro" được ghép từ hai phần: tiền tố "dis-" mang nghĩa "không tốt" hay "xấu" và "**astro**" nghĩa là "sao" (trong ngữ cảnh thiên văn học). Trong tiếng Ý, "disastro" được sử dụng để chỉ một sự kiện xấu xảy ra do ảnh hưởng của các vì sao, vì người xưa thường tin rằng các thiên thể có ảnh hưởng đến cuộc sống trên Trái đất.

astro/aster: astronomy, astronaut, asteroid
disappear: biến mất => **disaster**: tai họa

Từ "disastro" sau đó được chuyển sang tiếng Pháp dưới dạng "désastre" và cuối cùng trở thành "disaster" trong tiếng Anh vào khoảng thế kỷ 16. Ngày nay, từ "disaster" được sử dụng để chỉ một sự kiện đột ngột và khủng khiếp gây hậu quả nghiêm trọng về mặt nhân vật, kinh tế, môi trường hoặc xã hội.

Tiếng Anh: kiến thức cập nhật

DD

Xin cho biết ý nghĩa của thương hiệu NVidia?



NVidia là một công ty đa quốc gia của Mỹ chuyên về thiết kế và sản xuất chip đồ họa. NVidia được biết đến rộng rãi với các sản phẩm của mình trong lĩnh vực đồ họa máy tính, đặc biệt là chip đồ họa cho máy tính xách tay và máy tính để bàn. Họ cũng có mặt trong các lĩnh vực khác như trí tuệ nhân tạo, xe tự hành, và máy chủ.

Thương hiệu NVidia được xem như một biểu tượng của công nghệ tiên tiến và chất lượng cao trong ngành công nghệ thông tin. Nó tượng trưng cho sự đổi mới liên tục, độ tin cậy và hiệu suất mạnh mẽ. Công ty đã tạo ra nhiều công nghệ đột phá trong lĩnh vực đồ họa máy tính và trí tuệ nhân tạo, và thường xuyên đẩy giới hạn của những gì có thể thực hiện với công nghệ.

Tên "NVidia" bắt nguồn từ Latinh "*invidia*", có nghĩa là "*ghen tị*". Nó cũng là một sự chơi chữ với từ "*video*", một chỉ dẫn đến trọng tâm của công ty: đồ họa và video.

Tiếng Anh: hiểu rất sâu rộng

DD

Xin cho tôi biết nguồn gốc của từ "Milky Way"



The creation of the Milky Way in Greek and Roman Mythology!

A dark, star-filled night sky with a bright, winding Milky Way galaxy. A small white comet is visible in the lower left corner. The background is a dark blue to black gradient.

Thần Zeus: Vua các vị thần
Hera: Vợ thần Zeus
Hercule: Con rơi của thần Zeus

Tiếng Anh: câu quá khó => cũng dịch sai

Câu lắt léo (nhập nhằng):

“*Xem lại album, mình thấy ông già đi nhanh quá*”

DA

Hãy dịch câu sau đây sang tiếng Anh: Xem lại album, mình thấy ông già đi nhanh quá



Review the album, I see that the old man walks too fast.

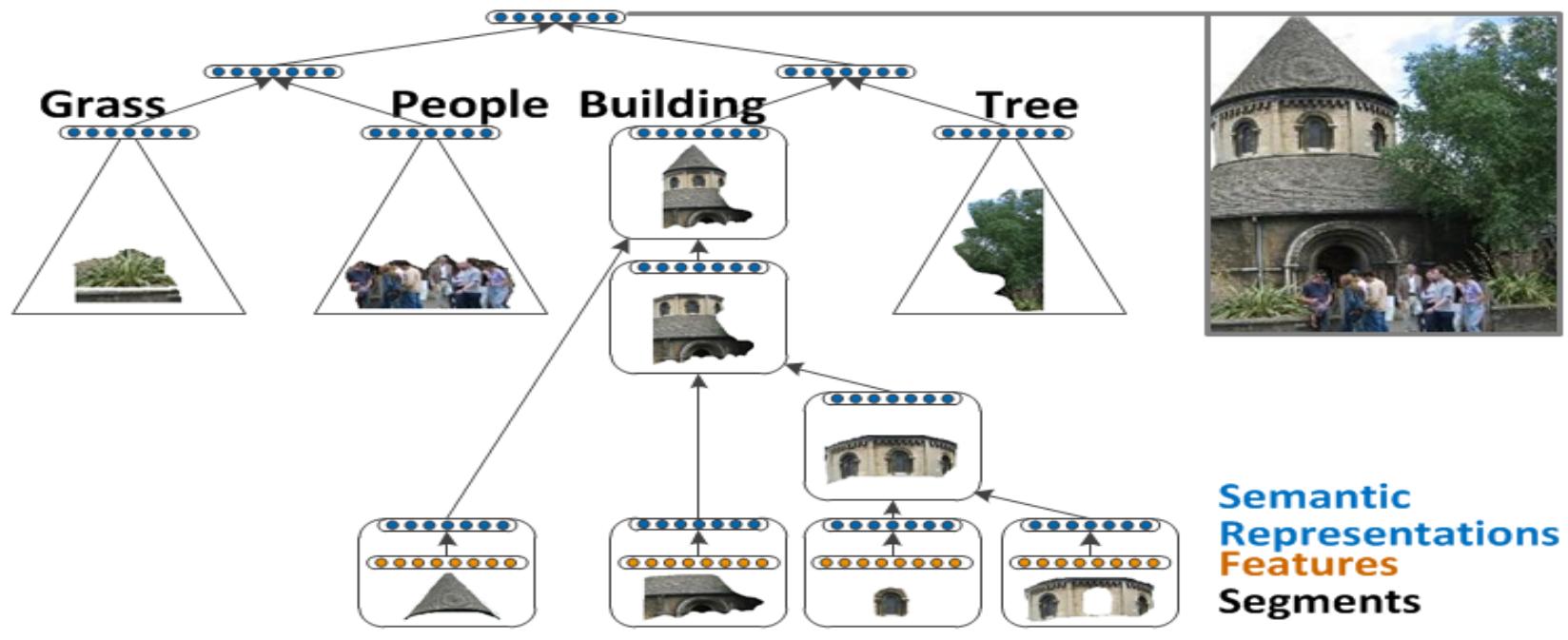
SAI: “*Review the album, I see that the old man walks too fast*”

(ông già / đi nhanh!)

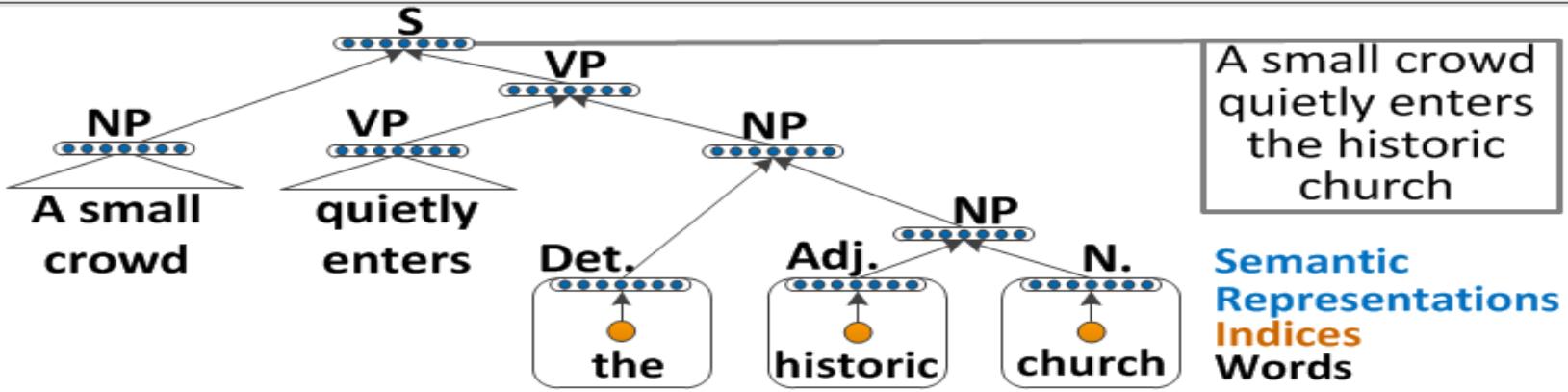
ĐÚNG: “*Review the album, I see that the grandfather gets old too fast*”

(ông / già đi nhanh)

Parsing Natural Scene Images

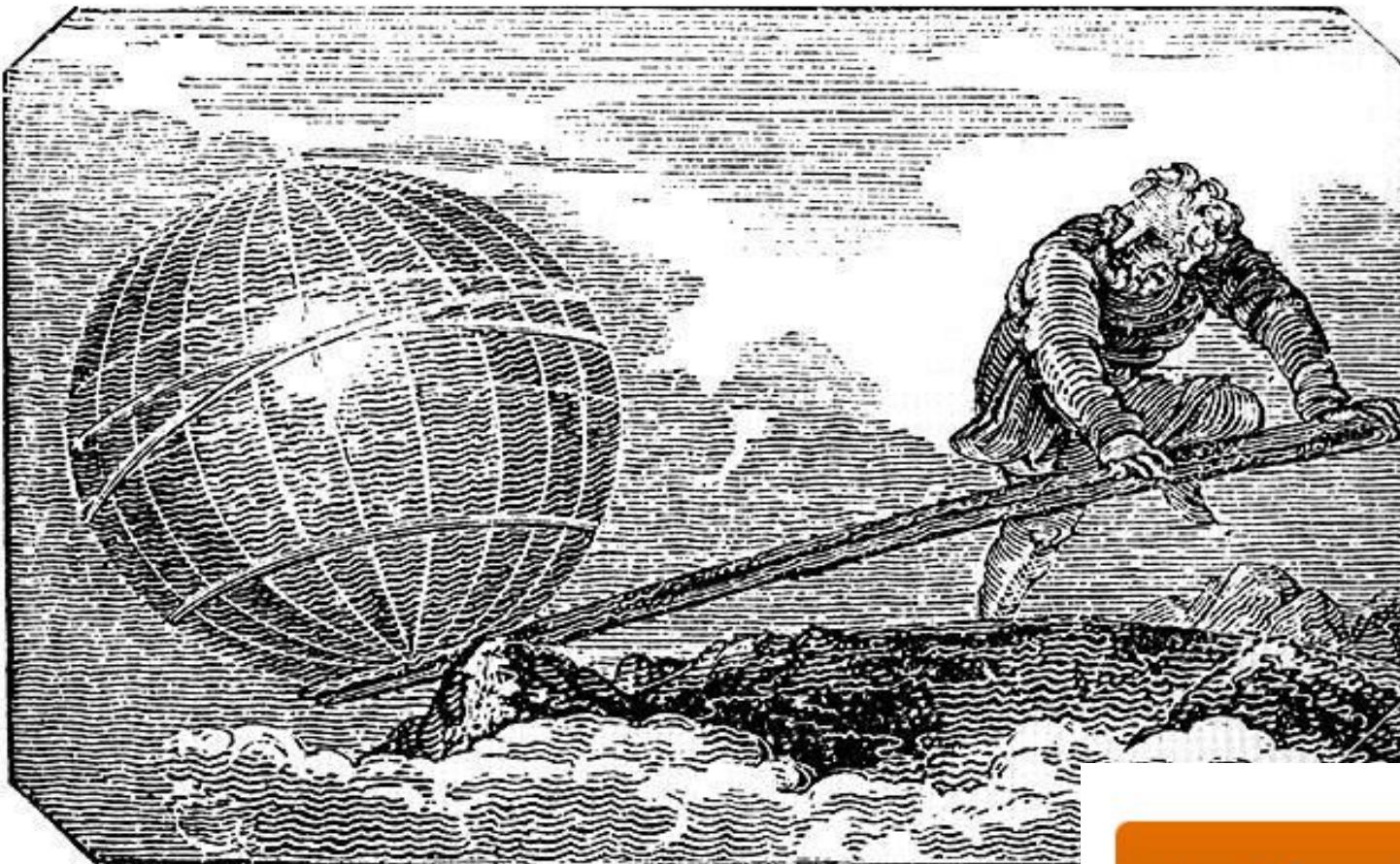


Parsing Natural Language Sentences



Vai trò của kho ngũ liệu trong Xử lý NNTN

= **điểm tựa cho đòn bẩy**



Archimedes ↗

Ancient Greek
mathematician



Computational Models:

Artificial Intelligence: A branch of computer science dealing with the simulation of intelligent behavior

Machine Learning: is a type of artificial intelligence ([AI](#)) that allows software applications to become more accurate at predicting outcomes via [training data](#).

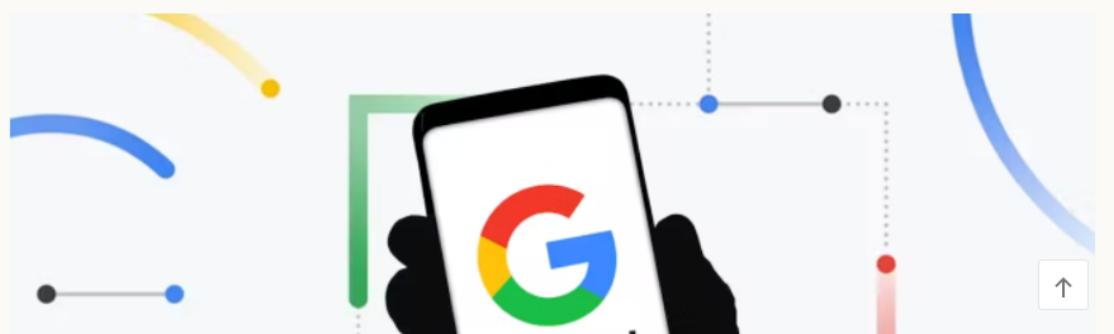
- **Deep Learning:** requires big data
- Computational Linguistics: Data = Corpus
- **Corpus:** 语料库/yǔ liào kù/ “ngữ liệu khô”
- Corpus = Collection of spoken/written text
- Building Corpus: by [native-speaker](#), Master in [Applied Linguistics](#) (Computational Linguistics), Data Science.

Google Bard thử nghiệm vội vã trước khi ra mắt

Google được cho là đã không thực hiện quá trình thử nghiệm Bard, chatbot AI đối đầu ChatGPT, đủ chín chu trước khi công bố.

Theo *Business Insider*, [Google](#) không có thời gian và nhân lực đánh giá và cải thiện chất lượng phản hồi do Bard đưa ra. Thay vào đó, công ty giao nhiệm vụ cho các đối tác thực hiện điều này.

Theo bốn nguồn tin từ nhà thầu của Google cũng như tài liệu nội bộ, việc thử nghiệm Bard diễn ra vội vã do họ không có đủ thời gian để xác minh câu trả lời đúng từ chatbot AI.



Google [ra mắt](#) Bard ngày 6/2. Tuy nhiên, ngay khi vừa trình làng, chatbot được xem là đối thủ của ChatGPT đã [trả lời sai](#) câu hỏi về kiến thức. Điều này khiến Google mất [100 tỷ USD](#) giá trị vốn hóa thị trường sau đó. Trước khi giới thiệu Bard, Google cũng được cho là đã yêu cầu nhân viên toàn thời gian dành từ hai đến bốn giờ để trò chuyện với bot mới. Họ sẽ đặt câu hỏi, gắn cờ các câu trả lời không đáp ứng các tiêu chuẩn về độ chính xác và các yếu tố khác.

Bảo Lâm (theo *Business Insider*)

Corpus:

- PTB (Penn Tree Bank): [Pierre/NNP Vinken/NNP],/, [61/CD years/NNS] old/JJ ,/, will/MD join/VB [the/DT board/NN] as/IN [a/DT nonexecutive/JJ director/NN Nov./NNP 29/CD]./.
- CTB (Chinese Tree Bank): <S ID=12>((IP-HLN (NP-SBJ (NN 外商) (NN 投资) (NN 企业)) (VP (VV 成为) (NP-OBJ (NP (NP-PN (NR 中国)) (NP (NN 外贸)))) (ADJP (JJ 重要)) (NP (NN 增长点))))) </S>
- (VTB: Vietnamese Tree Bank): <SEG id="1">
Nguyên_nhân/Nn/O là/Vc/O bão/Nn/O số/Nn/O 10/An/O
đang/R/O chịu/Vv/O ảnh_hưởng/Nn/O bởi/Cp/O
hệ_thống/Nn/O trực/Nn/O rãnh/Nn/O cao/Aa/O và/Cp/O
sự/Nc/O lôi_kéo/Vv/O từ/Cm/O siêu_bão/Nn/TRM_B
Melor/Nr/TRM_I_ở/Cm/O ngoài/Cm/O khơi/Nn/O
Philippines/Nr/LOC_B ./PU/O</SEG

Corpus:

[Many/JJ styles/NNS]

have/VBP

[perforations/NNS]

and/CC

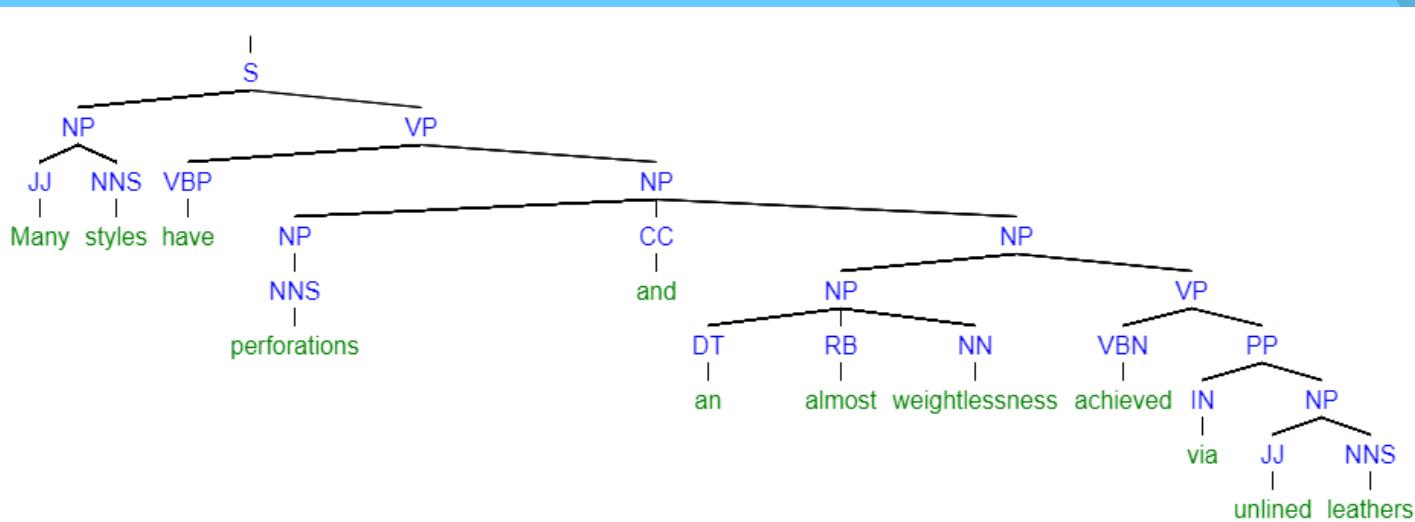
[an/DT almost/RB weightlessness/NN]

achieved/VBN via/IN

[unlined/JJ leathers/NNS]

./.

((S - - -
 (NP (JJ Many) (NNS styles))
 (UP (VBP have)
 (NP
 (NP (NNS perforations))
 (CC and)
 (NP
 (NP (DT an) (RB almost) (NN weightlessness))
 (UP (VBN achieved)
 (PP (IN via)
 (NP (JJ unlined) (NNS leathers)))))))))



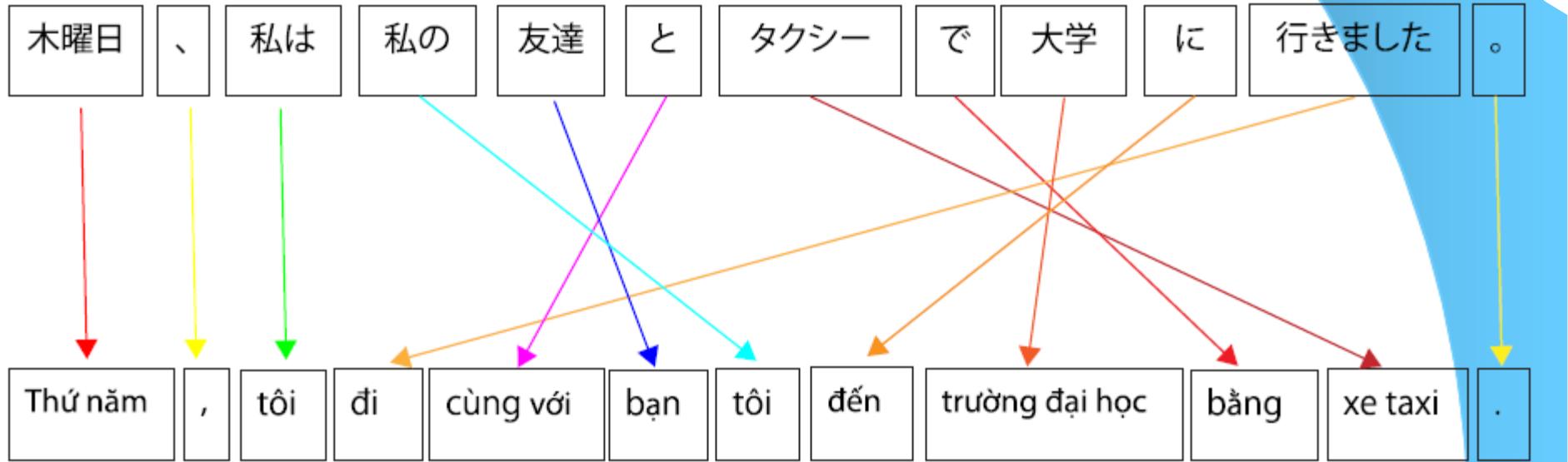
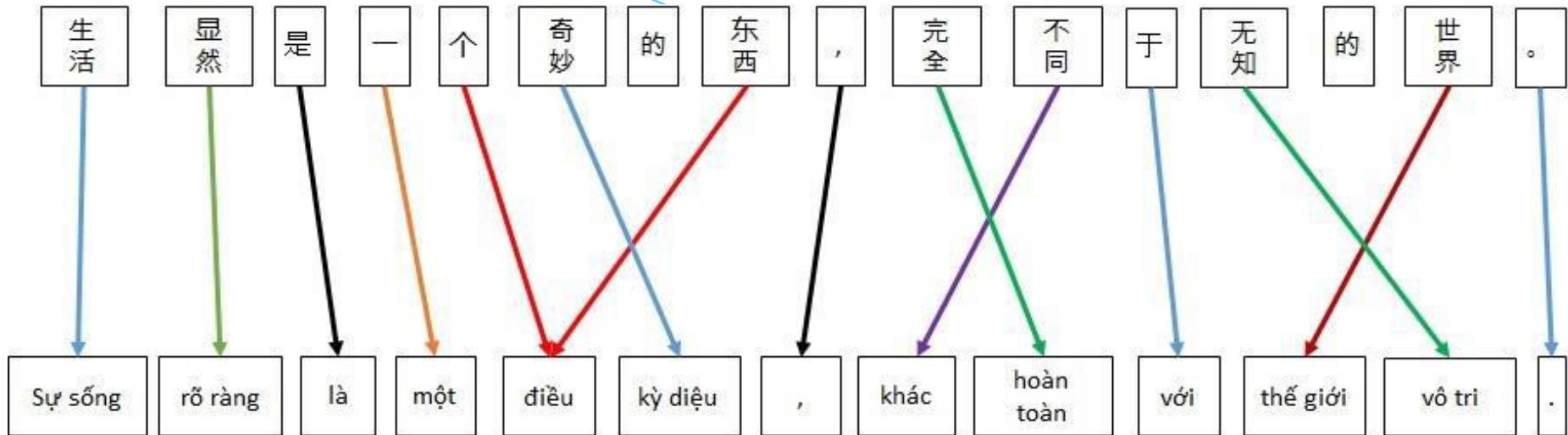
Parallel corpus: paragraph/sentence alignment

Helicopters can rise straight up into the air and can go straight down. They can stand still in the air. Helicopters do not have wings. A huge whirling propeller, called a rotor, on top of a helicopter provides the lift.

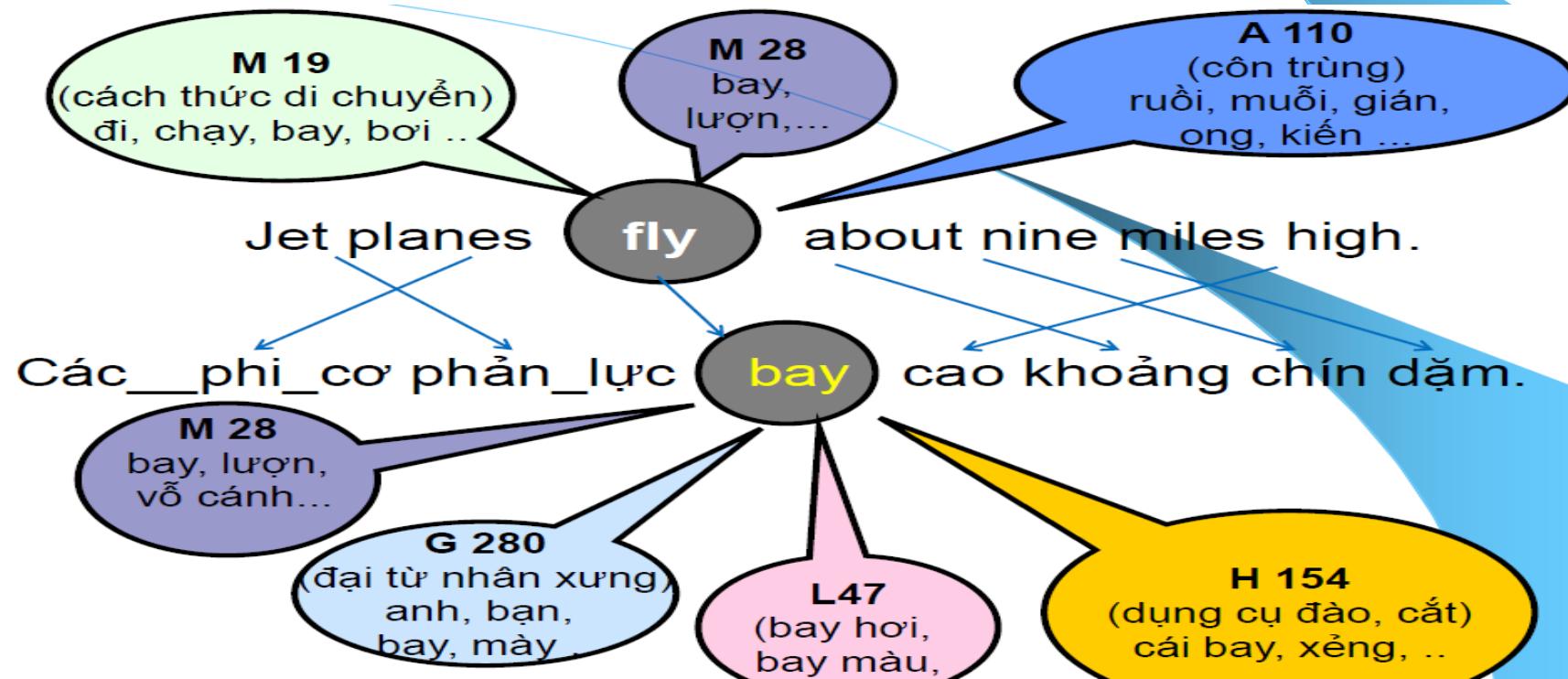
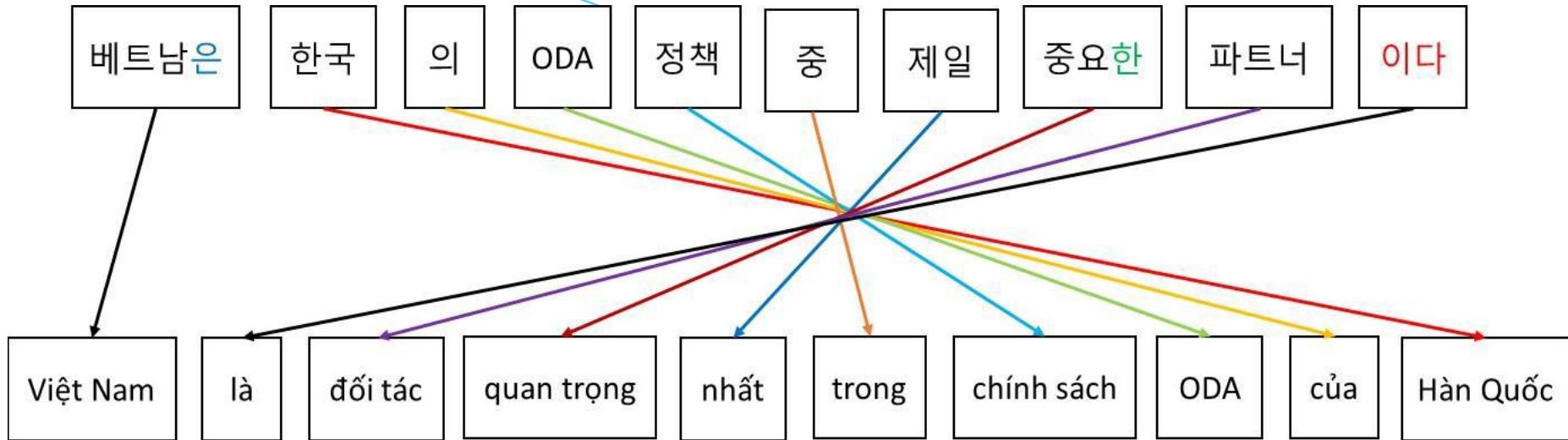
Máy bay trực thăng có thể lên thẳng trên không và đáp thẳng xuống đất. Chúng có thể đứng yên trên không. Máy bay trực thăng không có cánh, một cánh quạt lớn gọi là chong chóng trên đầu chiếc máy bay cung cấp sức nâng.

- * Helicopters can rise straight up into the air and can go straight down.
+ Máy bay trực thăng có thể lên thẳng trên không và đáp thẳng xuống đất.
- * They can stand still in the air.
+ Chúng có thể đứng yên trên không.
- * Helicopters do not have wings.
+ Máy bay trực thăng không có cánh.

Parallel corpus: word alignment



Parallel corpus: semantics tagging



Training Corpus: Q&A ChatBot

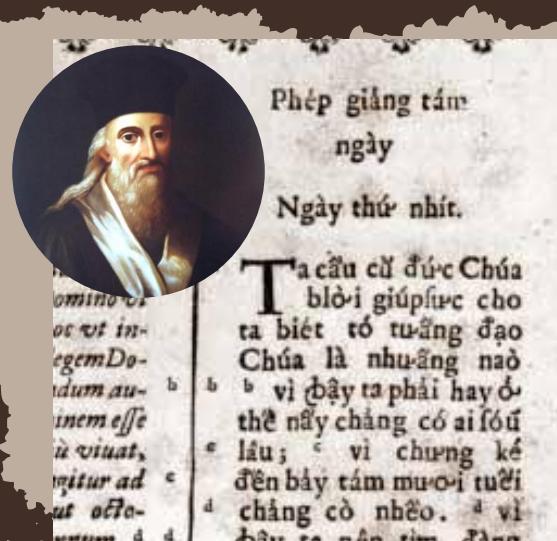
QUESTION	ANSWER
Actiso được trồng nhiều nhất ở đâu?	Actiso được trồng nhiều nhất ở Đà Lạt, Sapa, Tam Đảo.
Đâu là phần của cây Actiso được sử dụng làm thuốc?	Lá, hoa, rễ và thân của cây Actiso đều được dùng làm thuốc.
Thành phần hóa học chứa trong lá Actiso bao gồm những gì?	Lá Actiso chứa các hợp chất như dẫn xuất axit caffeic, flavonoid, lacton sesquiterpene, anthocyan, cyanidin, tannin và tinh dầu dễ bay hơi.
Công dụng của lá Actiso là gì?	Lá Actiso có tác dụng lợi tiểu và được dùng trong điều trị bệnh phù và thấp khớp.
Cách dùng và liều dùng của Actiso như thế nào?	Actiso có thể dùng dưới hình thức thuốc sắc, lá tươi hoặc khô. Liều dùng hàng ngày tùy thuộc vào hình thức và tình trạng bệnh.
Actiso có thể giúp bảo vệ gan như thế nào?	Actiso chứa các chất chống oxy hóa giúp bảo vệ gan và giảm nồng độ độc tố có hại cho gan.
Actiso có tác dụng lợi mật như thế nào?	Actiso giúp tăng bài tiết dịch mật và tăng hoạt động các enzyme giúp tiêu hóa.
Actiso có tác dụng giúp xương khỏe mạnh như thế nào?	Actiso chứa những khoáng chất như mangan, photpho và magiê giúp tăng cường sức khỏe của xương và ngăn ngừa loãng xương.
Actiso có tác dụng giúp trái tim khỏe mạnh như thế nào?	Actiso giúp giảm mức độ cholesterol xấu và tăng mức độ cholesterol tốt, từ đó giúp ổn định huyết áp và giảm các bệnh lý tim mạch.

Collect Ancient Texts

Before 17th century :



17th century :



After 17th century :

Dầu lòng hai á Tố-nga, 1)
Túy-kiều là chị, em là Túy-vân.
Mai cốt-cách, tuyết tinh-thần, (2)
một người một vẻ, mươi phân vẹn mươi.
Vân xem trang-trọng tốt vời,
khuôn lung đầy-dặn, nét ngotrời nở-nang
Hoa cười ngọc thốt đoan-trang, (3)
mây thua nước tóc, tuyết nhường màu da.
Kiều càng sắc-sảo mặn-mà,
so bè tài sắc lại là phần hơn.
Gương thu thủy, vút xuân sơn, 4)
hoa ghen thua thẳm, liêu hòn kẽm xanh;
Một đôi nghiêng nước nghiêng thành,
sắc dành đổi một, tài dành hòa hai:

Now :

15.Đầu lòng hai á tố nga,
Thúy Kiều là chị, em là Thúy Vân.
Mai cốt cách, tuyết tinh thần,
Mỗi người một vẻ, mươi phân vẹn mươi.
Vân xem trang trọng khác vời,
Khuôn trang đầy đặn, nét ngài nở nang.
Hoa cười ngọc thốt đoan trang,
Mây thua nước tóc, tuyết nhường màu da.
Kiều càng sắc sảo, mặn mà,
So bè tài, sắc, lại là phần hơn.

Ancient Corpus

sino	chinese	Viet_raw
<p>Kinh thảo Đa Sĩ.</p> <p>Tiên thi, Đa Sĩ thị hữu công lao hoành hành quốc trung, hiếp gian cự nữ, bức dâm nhân thê, cố Duy Sản sứ Chân thảo chí.</p> <p>Thời hữu tinh vẫn ư doanh trung.</p> <p>Duy Sản quân tiến chí Chi Linh [huyện danh] dữ Cào giáp chiến ư Nam Giản.</p> <p>Tì tướng danh Hạnh tử ư trận.</p> <p>Duy Sản kiến tặc khiêu chiến dục kích chi, chư tướng gián bất thính.</p> <p>Tặc hưu khiêu chiến, Duy Sản phản nộ phân đạo tiến chiến, thân tien sĩ tốt.</p> <p>Duy Sản cập Nguyễn Thượng giai vị Trần Cảo sở hoạch, chí Vạn Kiếp hành quán sát chí.</p> <p>Cào trực đáo Bồ Đề.</p> <p>Đế mệnh Thiết Sơn bá Trần Chân tiến thảo, đại phá chí, thích sát thậm đà.</p> <p>Cào toại thoán vu Lạng Nguyên, bắt cảm phục xuất, di Nguyệt Giang vi giới.</p> <p>Chân dữ Cào quân trương tri.</p> <p>Cào truyền kỉ tử thăng tiêm hiệu Tuyên Hoà.</p> <p>Hậu Cào trước.</p>	<p>京討多士. 先是多士恃有功勞衡衡國中脅奸 虜女逼淫人妻故惟僥使真討之. 時有星隕於營中. 惟僥軍進至至靈 [縣名] 與曠夾戰 於南澗. 裨將名幸死於陣. 惟僥見賊挑戰欲擊之諸將諫不聽. 賊又挑戰惟僥忿怒分道進戰身先士卒. 惟僥及阮尚皆為陣曠所獲至萬劫行館殺支. 曠直到菩提. 帝命鐵山伯陳真進討大破之刺殺甚多. 曠遂竄于諒原不敢復出以月江為界. 真與曠軍相持. 曠傳其子昇借號宣和. 後曠削.</p>	<p>Kinh_sư đánh bọn Đa_Sĩ .</p> <p>Trước đó , Đa_Sĩ cậy có công_lao hoành_hành trong Kinh , cưỡng_hiếp con_gái chưa chồng , gian_dâm vợ của người khác , cho_nên_Duy_Sản sai Chân về đánh .</p> <p>Bấy_giờ , có sao_sa xuống trung_doanh .</p> <p>Quân_Duy_Sản tiến đến Chi_Linh , giáp_chiến với quân Cào ở xã Nam_Giản .</p> <p>Viên_tỷ_tướng tên là Hạnh chết tại_trận .</p> <p>Duy_Sản thấy giặc khiêu_chiến , có ý_muốn đánh , các tướng can không được .</p> <p>Giặc_lại khiêu_chiến .</p> <p>Duy_Sản tức giận chia đường tiến đánh , tự mình xông lên trước quân_linh .</p> <p>Duy_Sản và Nguyễn_Thượng đều bị Trần_Cảo bắt được , đem về hành quán ở Vạn_Kiếp giết chết .</p> <p>Cào tiến thắng đến Bồ Đề , vua sai Thiết_Son_bá Trần_Chân tiến đánh , phá tan được , chém_giết rất nhiều .</p> <p>Cào phái chạy trốn về Lạng_Nguyên không dám ra đánh nữa , lấy sông Nguyệt làm ranh_giới .</p> <p>Chân cầm_cự với quân Cào .</p> <p>Cào truyền ngôi cho con là Cung , tiếm xưng niên_hiệu là Tuyên_hoà .</p> <p>Sau Cào cạo</p>
<p>giai giải tán.</p> <p>Hạ, từ nguyệt, thời cốc đại phong nhãm, nãi phản khiển bồi trúc các xú đê lô dì phòng thuỷ hoạn.</p> <p>Thị nguyệt, Yên Định huyện, Đan Nê Thượng xã, Đồng Cò son băng.</p> <p>Mệnh quan vãng cáo té chí.</p> <p>Nhị thập tứ nhật, Dần thì, hữu tinh hiện vu Tây Nam phuông, hình như hồng quyên.</p> <p>Nhị thập bát nhật, Dậu thì, hữu tinh trực đằng Tây phuông tầu, hình như bạch thắt.</p> <p>Nhuận tú nguyệt, mệnh Thái phó Thanh quân công Trịnh Tráng, Thái bảo Vạn quận công Trịnh Xuân đằng đốc lính tượng mã sỹ tốt vãng phạt nguy Hào quận đồ đằng vu Yên Dũng địa.</p>	<p>皆解散. 夏四月時穀大豐稔乃分遣培築各處堤路以防水患. 是月安定縣丹泥上社銅礮山崩. 命官往告祭之. 二十四日寅時有彗星見于西南方形如紅絹. 二十八日酉時有星直騰西方走形如帛疋. 閏四月命太傅清郡公鄭椿太保萬郡公鄭椿等督領象馬士卒往伐偽豪郡徒黨于安勇地.</p>	<p>tan_vỡ .</p> <p>Mùa hạ , tháng 4 , bấy_giờ lúa rất tốt , bèn chia sai bồi_dập đê_điều các xú đê phòng nạn lụt .</p> <p>Tháng áy , núi Đồng_Cò ở xã Đan_Nê_Thượng , huyện Yên_Def bị lở .</p> <p>Sai_quan đến cáo_té .</p> <p>Ngày 2 , giờ Dần , có sao_Chổi mọc ở phuông tây_nam , hình_như tôm hùm đỏ .</p> <p>Ngày 28 , giờ Dậu , có ngôi_sao bay thẳng về phuông tây , hình_như tôm hùm Tháng 4 nhuận , sai bọn Thái phó Thanh quận_công Trịnh_Tráng và Thái_bảo_Vạn quận_công Trịnh_Xuân đốc lính voi ngựa , quân_linh đi đánh bè_đằng Hào quận công nguy ở vùng Yên_Dũng .</p>
		

Data production

THE NEW NEW WORLD

How Cheap Labor Drives China's A.I. Ambitions



Workers at the headquarters of Ruijin Technology Company in Jiaxian, in central China's Henan Province. They identify objects in images to help artificial intelligence make sense of the world. Yan Cong for The New York Times

Data is the new oil, it has been said for years now. If data is the new oil, then China is already the largest producer with its factories packed with laborers working hard to annotate images and data for machine learning (*Analytics India Magazine*).

Vietnamese Computational Linguistics

Rank	Language Name	Primary Country	Population
1	CHINESE, MANDARIN	China	885,000,000
2	SPANISH	Spain	332,000,000
3	ENGLISH	United Kingdom	322,000,000
4	BENGALI	Bangladesh	189,000,000
5	HINDI	India	182,000,000
6	PORTUGUESE	Portugal	170,000,000
7	RUSSIAN	Russia	170,000,000
8	JAPANESE	Japan	125,000,000
9	GERMAN, STANDARD	Germany	98,000,000
10	CHINESE, WU (Ngô)	China	77,175,000
11	JAVANESE	Indonesia, Java, Bali	75,500,800
12	KOREAN	Korea, South	75,000,000
13	FRENCH	France	72,000,000
14	VIETNAMESE	Vietnam	67,662,000
15	TELUGU	India	66,350,000
16	CHINESE, YUE (Việt)	China	66,000,000

Abilities of Vietnamese language perception?

Việt Nam giành huy chương Vàng môn kiếm chém đồng đội

VietNam+ 06/12/2019 38 liên quan

Đây là tấm huy chương Vàng thứ 4 của đoàn thể thao Việt Nam trong ngày thi đấu chính thức thứ 6 của SEA Games 30.

Thích

Bình luận

Chia sẻ

Á hậu, MC bình thản khai nhận
nhiều lần bán dâm ngàn USD cho
chách ở trụ sở công an

Viết Dũng - Theo Trí Thức Trẻ, 06/09/2018 17:34



**BÙI TIẾN DŨNG THỔ LỘ
VIỆC VỢ CÓ BẦU VỚI CỰU
HLV TRƯỞNG ĐỘI TUYỂN
VIỆT NAM TẠI SÂN HÀNG
ĐẤY**

[ttvn.vn](#) | 07/07/2019 12:00 AM

Bùi Tiến Dũng và Viettel đã có chiến thắng tối thiểu 1-0 trước CLB TPHCM ở vòng 14 V.League 2019 diễn ra vào tối 7/7. Sau trận đấu, anh cũng có cuộc gặp mặt ngắn với





"BỎ QUA"

- Qua cũng có nói là, nếu như em thấy qua có gì không phải thì em hãy bỏ qua! Vậy mà cuối cùng cô ấy bỏ qua...

-Thì ông nói vậy mà ?

- Hồng phai, ý của qua là nếu như qua không phải thì hãy bỏ qua, chứ đừng có bỏ qua! Vậy mà cô ấy không chịu bỏ qua, cô ấy bỏ qua...

NĂM CON HỔ - NHÂM DẦN 2022

Thật ra, thì NĂM CON HỔ và NĂM CON HỒ có hai nghĩa hoàn toàn khác nhau.

Một bên là NĂM CON HỒ và bên kia là NĂM CON HỔ.

Nếu hiểu theo nghĩa NĂM CON HỒ thì nó sẽ là NĂM CON HỒ, còn nếu ta hiểu theo nghĩa của NĂM CON HỒ thì nó phải được hiểu là NĂM CON HỔ.

Vietnamese-native speaker: our strong point.

=> Our strong point: Vietnamese-native speakers

Inquiry about Machine Translation for Vietnamese Inbox x   

임행선 <hs00.lim@samsung.com> 8/9/13 Star Up Down

to me

Dear Professor Dinh Dien,

This is the Software Center at Samsung Electronics in Korea. Our lab is currently researching the development of Korean <-> Vietnamese machine translation. While searching for Vietnamese universities and companies which have expertise in MT, we came across your name. We wonder whether you have conducted research on MT for Vietnamese language, and whether you have an ongoing research or project. If you share with us how things are with you, it will very helpful to us.

We also need the info on MT companies which work on Vietnamese. If you know any company or institution which supports Vietnamese with its own MT engine, please let us know.

Thank you in advance.

Best regards,
Haengsun Eunice Lim

Haengsun Eunice Lim
Mobile. +82-10-2320-5040 / Tel. +82-31-279-2395
E-mail. hs00.lim@samsung.com

Web Platform Lab/ Software Center
Samsung Electronics Co., LTD in Suwon, Korea

Inquiry about Machine Translation for Vietnamese



Inbox x



임행선 <hs00.lim@samsung.com>

to me ▼

8/9/13



Dear Professor Dinh Dien,

This is the Software Center at Samsung Electronics in Korea. Our lab is currently researching the development of Korean <-> Vietnamese machine translation.

While searching for Vietnamese universities and companies which have expertise in MT, we came across your name.

We wonder whether you have conducted research on MT for Vietnamese language, and whether you have an ongoing research or project. If you share with us how things are with you, it will be very helpful to us.

We also need the info on MT companies which work on Vietnamese. If you know any company or institution which supports Vietnamese with its own MT engine, please let us know.

Thank you in advance.

Best regards,

Haengsun Eunice Lim

Haengsun Eunice Lim

Mobile. +82-10-2320-5040 / Tel. +82-31-279-2395

E-mail. hs00.lim@samsung.com

Web Platform Lab/ Software Center

Samsung Electronics Co., LTD in Suwon, Korea

Sent: Tuesday, January 19, 2016 1:58 PM

To: ddien@fit.hcmus.edu.vn

Subject: Acquiring Vietnamese treebank

Dear Prof. Dinh Dien,

HyunJeong Choe <hyunjeongc@google.com>

1/21/16



to me

Thank you so much your prompt reply!

If your treebank contain 300k, then we would like to acquire the entire set.

We are Natural language understanding team under Google research team and focusing on several NLP projects. We'd like to use your treebank to train our Vietnamese segmenter, PoS tagger and NER tagger. These analyzer will be used several Google projects such as conversational search.

The Licensee may use the data internally only. The Licensee may not:

1. Distribute the data;
2. Publish any research in which the data was used without providing a citation acknowledging that the data was developed by the Computation Linguistics Center of HCMUS.

Best,

-HJ

Date: 15-Oct-2015

From: Kohei Saito <AdvancedLinguistics@gmail.com>

Subject: Vietnamese; Computational Linguistics; Morphology; Phonology; Semantics; Syntax: Analytic Linguistic Project Manager, Google, Inc., Singapore

University or Organization: **Google, Inc.**

Department: Natural Language Understanding

Job Location: Singapore, Singapore

Job Title: Analytic Linguistic Project Manager [Vietnamese]

Job Rank: Analytic Linguistic Project Manager; Manager

Specialty Areas: Computational Linguistics; Morphology; Phonology; Semantics; Syntax

Required Language(s): **Vietnamese (vie)**

Description:

The role of the Analytic Linguistic Project Manager is to consult with Natural Language Understanding Researchers on creating guidelines and setting standards for a variety of NLP projects as well as to manage the work of a team of junior linguists to achieve high quality data output.

This includes:

- Training, managing and overseeing the work of a team of junior linguists
- Creating guidelines for semantic, syntactic and morphological projects
- Evaluating and analyzing data quality
- Consulting with researchers and engineers on the development of linguistic databases

Job requirements:

- **Native-level speaker of Vietnamese** and fluent in English
- **Master's degree or higher in Linguistics or Computational Linguistics**, specializing in semantics, syntax, morphology or lexicography
- Ability to quickly grasp technical concepts; should have an interest in natural language processing
- Excellent oral and written communication skills
- Good organizational skills
- Previous project management and people management experience preferred
- Some programming language or previous experience working in a Linux environment a plus

Hivan Fagnano hivan.fagnano

Số hóa > Công nghệ

Thứ tư, 30/9/2020, 13:00 (GMT+7)

to clc 

Greetings,

Our company is looking for
well known multinational co
We're building up a team of
studies.

The project, which would last 3-5 months, involves performing quality control tasks of audio-recorded files \ voice overs in the linguists' native language, so phonetic transcription, pronunciation transcription and proofreading skills are required.

Please note that the right candidates must be native speakers of Vietnamese.

By visiting your site, I've noticed the the Computational Linguistic Center focuses on spelling checker, grammar checker, Text Translation, Contrastive Linguistics, etc. and I thought students from your course might be considered good candidates, as we need native Vietnamese speakers with the above mentioned skills.

Would it be possible talking with teachers from the Center?

Apple tuyển người nói tiếng Việt làm Siri

Apple đăng tuyển nhân sự thành thạo tiếng Việt trên trang tuyển dụng của mình, nhiều khả năng sẽ phát triển trợ lý ảo Siri cho thị trường Việt Nam.

Trên trang tuyển dụng, Apple mới bổ sung vị trí chuyên viên Ngôn ngữ Việt Nam cho mảng trí tuệ nhân tạo và học máy. Người được tuyển dụng sẽ làm việc trong đội ngũ phát triển Siri, tại văn phòng ở khu Ang Mo Kio (Singapore). Siri là trợ lý ảo của Apple và là một trong những ứng dụng thực tế nhất về AI mà Apple đang phát triển.

MY PRODUCTS

Dictionary

An entry in the Chinese-Vietnamese Dictionary:

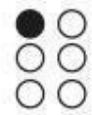
```
<WORD>
  <HEAD>油然</HEAD>
  <PHONETIC>yóurán</PHONETIC>
  <BODY>
    <TXT_V>Tự nhiên</TXT_V>
    <EXAMPLE>
      <TXT_C>敬慕之心，油然而生</TXT_C>
      <TXT_V>Lòng ngưỡng mộ, tự nhiên mà có</TXT_V>
    </EXAMPLE>
  </BODY>
  <BODY>
    <TXT_V> hơi nước bốc lên</TXT_V>
    <EXAMPLE>
      <TXT_C>油然作云</TXT_C>
      <TXT_V>Hơi nước bốc lên thành mây</TXT_V>
    </EXAMPLE>
  </BODY>
</WORD>
```

An image of the Oxford Advanced Learner's Dictionary, showing its blue and white cover with the title and a small image of the British Parliament building.

An image of the Kim Từ Điển electronic dictionary device. It is a black, handheld unit with a small screen and a keyboard. The screen displays the text "Từ điển DỊCH CÂU dẫn đầu CÔNG NGHỆ". Below the device, it says "ANH - VIỆT - PHÁP - HOA - NHẬT - HÀN - ĐỨC - NGA".

An image of the Kim Từ Điển GD7200M electronic dictionary, shown from a top-down perspective. The screen displays various icons and text related to the dictionary's functions. The model number "GD7200M" is visible at the bottom right of the device.

KÝ HIỆU CHỮ BRAILLE VIỆT NGỮ



A



Ă



Â



B



C



Đ



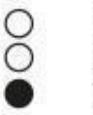
D



E



Ê



F



G



H



I



J



K



L



M



N



O



Ô



Õ



P



Q



R



S



T



U



Ú



Ý



Ý



X



Y



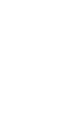
Z



sắc



huyền



hỏi



ngã



năng



hai chấm :



phẩy ,



chấm phẩy ;



chấm câu .



chấm thang !



chấm hỏi ?



Báo viết hoa



(



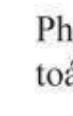
)



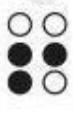
mở "



đóng "



Phép
tính:



+



-



X



:



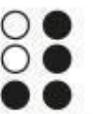
=



<



>



Báo số



0



1



2



3



4



5



6



7

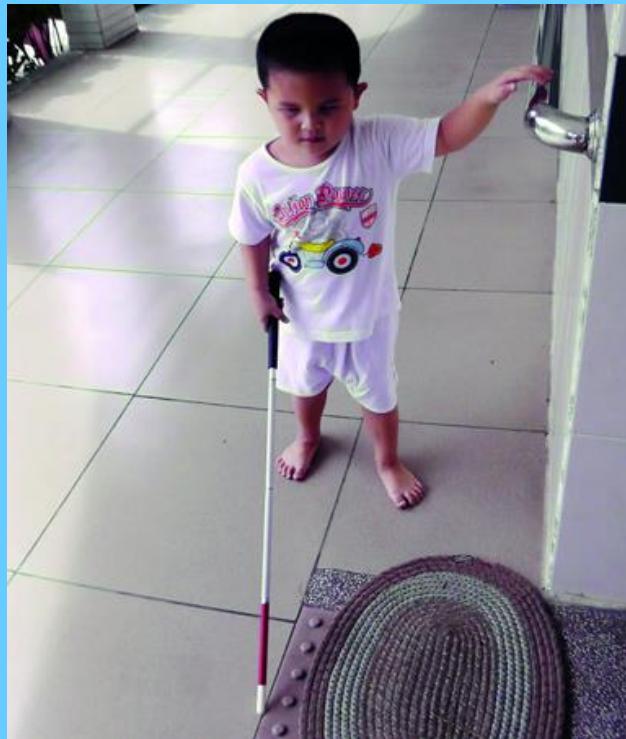


8



9

Trung tâm dữ liệu đa ngữ
Kim Từ Điện (KMDC) chúng
tôi chuyên sản xuất các
Phần mềm có hỗ trợ tiếng
Việt cho người khiếm thị.





Language Technology

making good progress

mostly solved

Spam detection

Let's go to Agra!
Buy V1AGRA ...



Part-of-speech (POS) tagging

ADJ	ADJ	NOUN	VERB	ADV
Colorless	green	ideas	sleep	furiously.

Named entity recognition (NER)

PERSON	ORG	LOC
Einstein	met with	UN officials in Princeton

Sentiment analysis

Best roast chicken in San Francisco!
The waiter ignored us for 20 minutes.



Coreference resolution

Carter told Mubarak he shouldn't run again.

Word sense disambiguation (WSD)

I need new batteries for my **mouse**.



Parsing

I can see Alcatraz from the window!

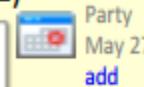
Machine translation (MT)

第13届上海国际电影节开幕...
The 13th Shanghai International Film Festival...



Information extraction (IE)

You're invited to our dinner party, Friday May 27 at 8:30



still really hard

Question answering (QA)

Q. How effective is ibuprofen in reducing fever in patients with acute febrile illness?

Paraphrase

XYZ acquired ABC yesterday
ABC has been taken over by XYZ

Summarization

The Dow Jones is up
The S&P500 jumped
Housing prices rose



Economy is good

Dialog

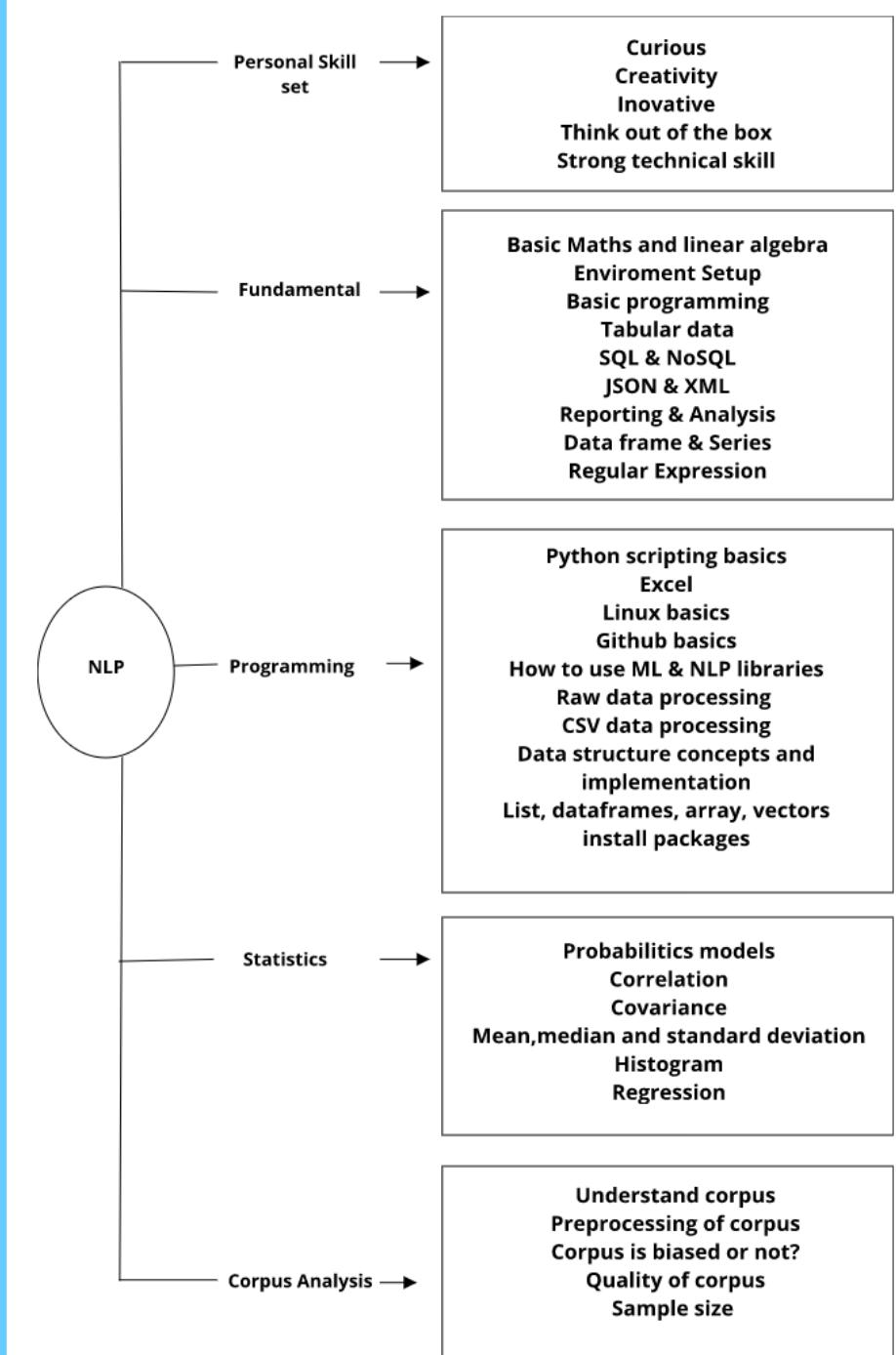
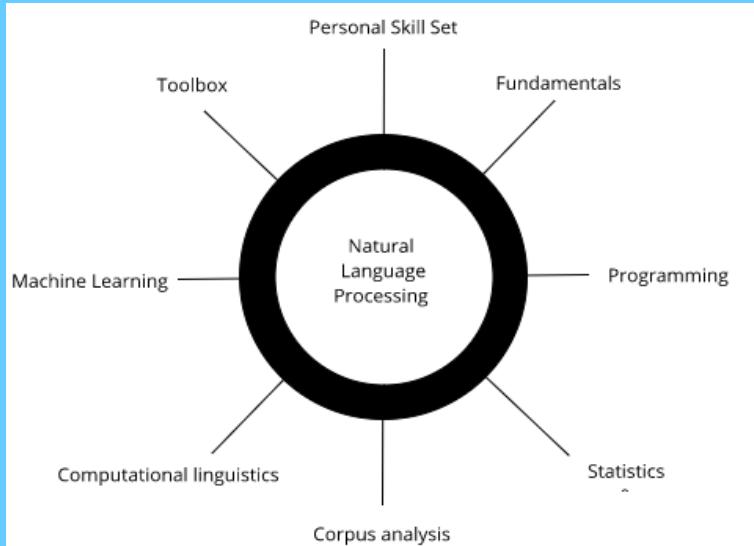
Where is Citizen Kane playing in SF?



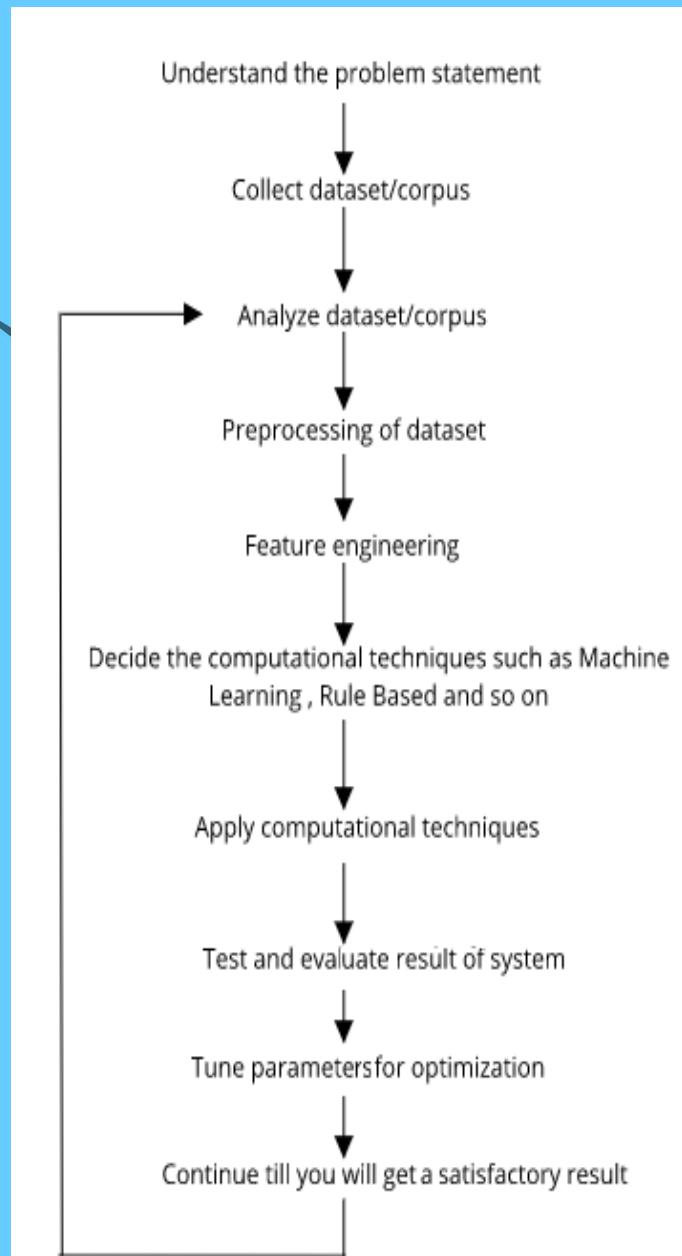
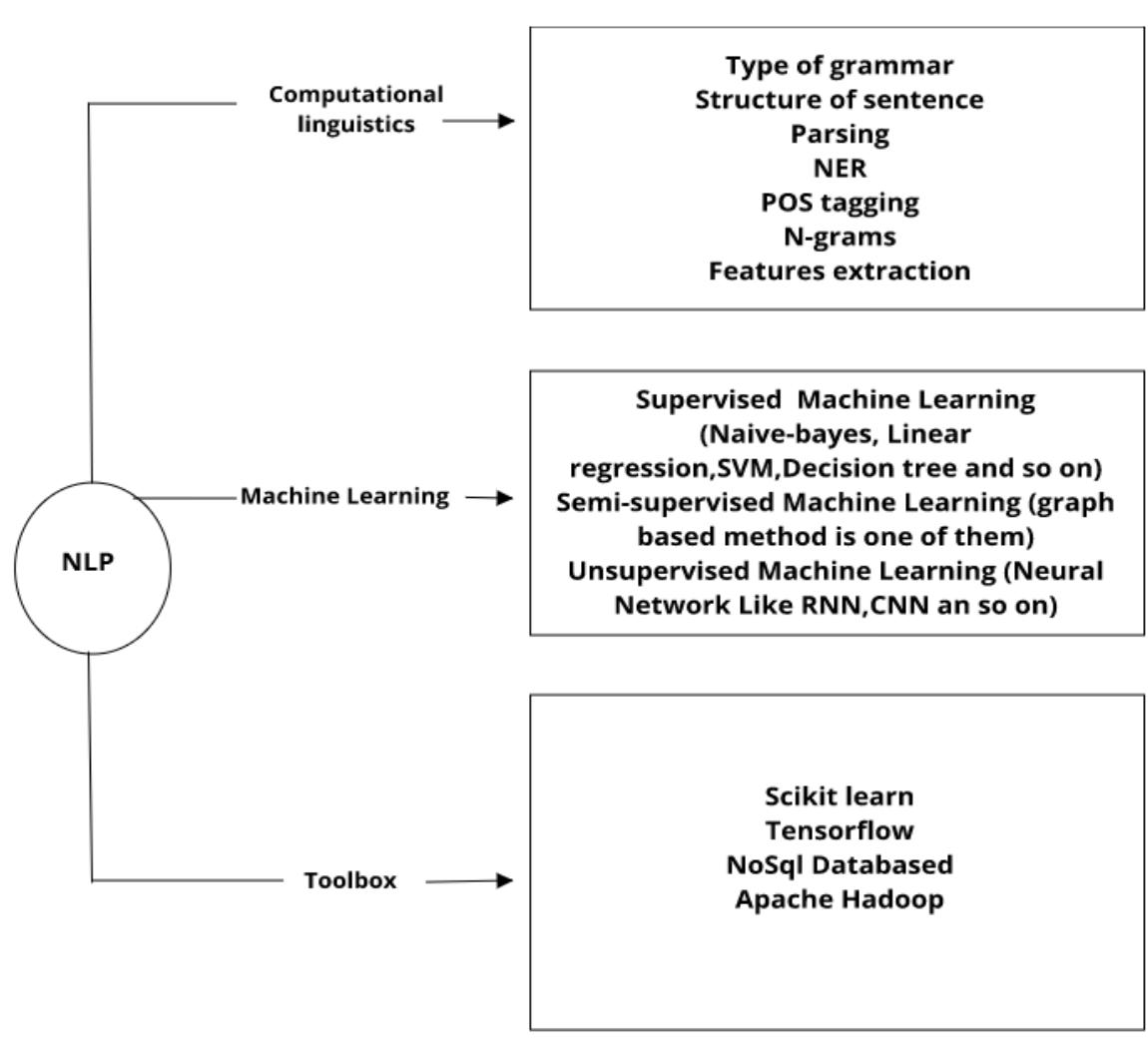
Castro Theatre at 7:30. Do you want a ticket?



What do you need for NLP?



What will you learn in this NLP course?



Course content

Part I:

Introduction to Natural Languages

Review Formal Languages

Rule-based NLP: parsing

Corpus-based NLP: corpus linguistics

PART II:

AI-ML, DL, LLM, Gen-AI,

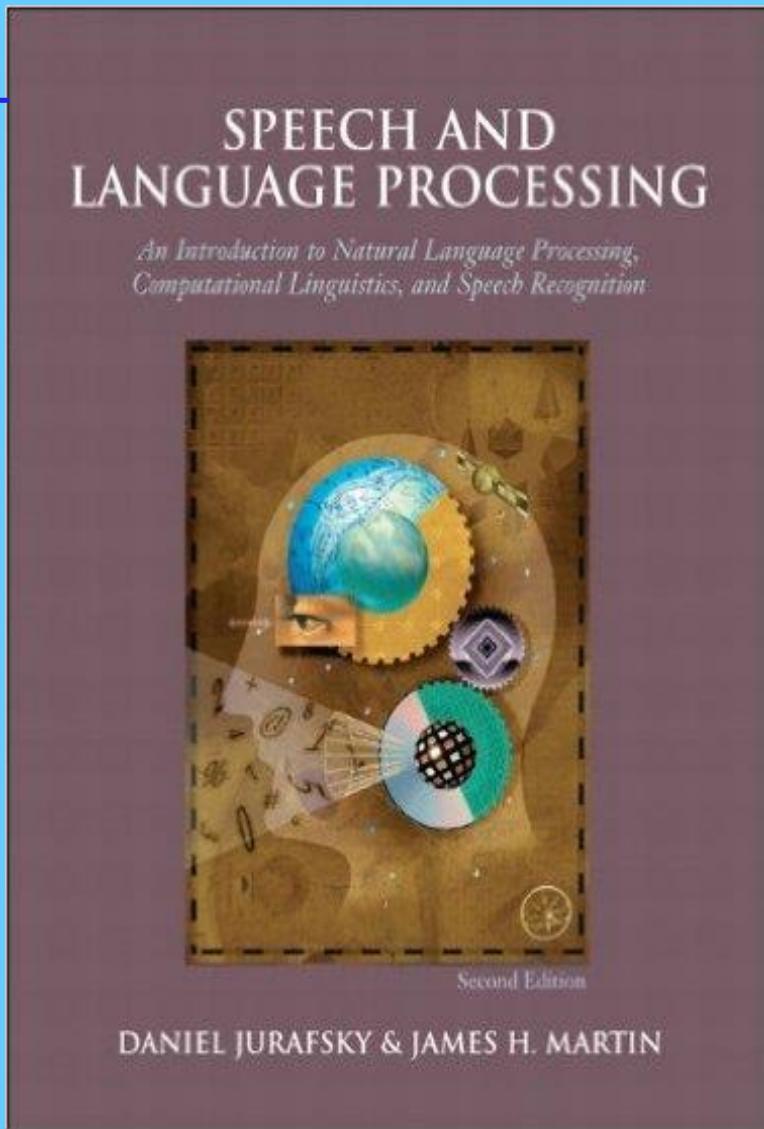
NLP project (seminar):

Enterprise: customized AI-Chatbot, PE techniques.

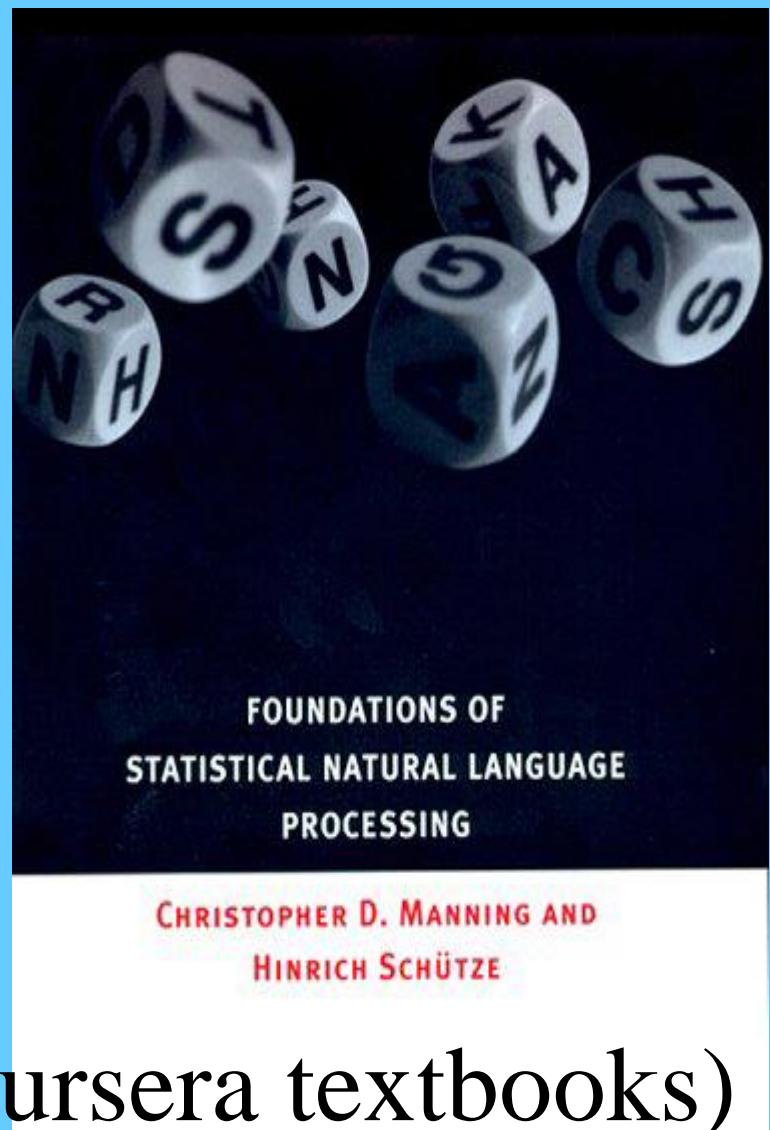
VNese-NLP applications: Sino-Nôm OCR, translation, mining, ...

Textbooks (theory)

01



02



(Coursera textbooks)

Textbooks (practice)

03 **Python**

Natural Language Processing

Explore NLP with machine learning
and deep learning techniques

Jalaj Thanaki

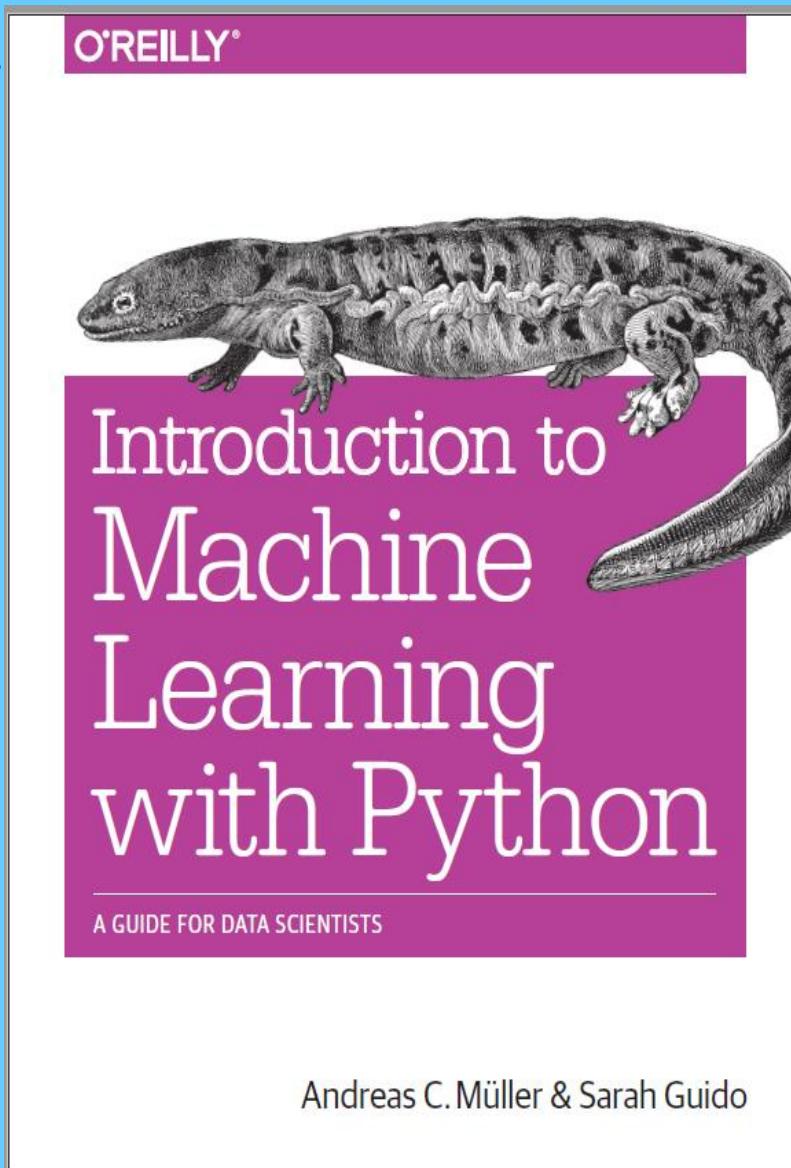
Packt

BIRMINGHAM - MUMBAI

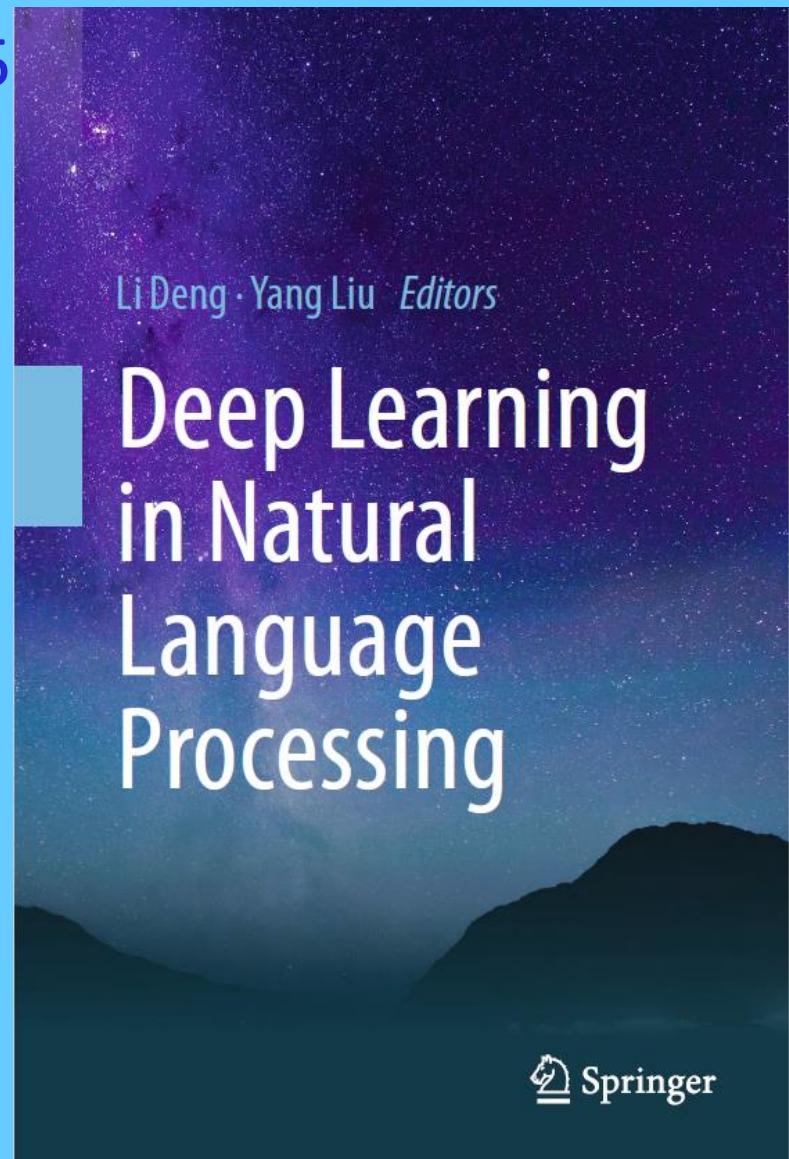
- >  Introduction
- >  Corpus & Dataset
- >  Structure of Sentences
- >  Preprocessing
- >  Feature Engineering & NLP Algorithms
- >  Advanced Feature Engineering & NLP Algorithms
- >  Rule-based System for NLP
- >  Machine Learning for NLP Problems
- >  Deep Learning for NLU & NLG Problems
- >  Advanced Tools
- >  Improve your NLP Skills
- >  Installation Guide

Reference books (ML techniques)

04



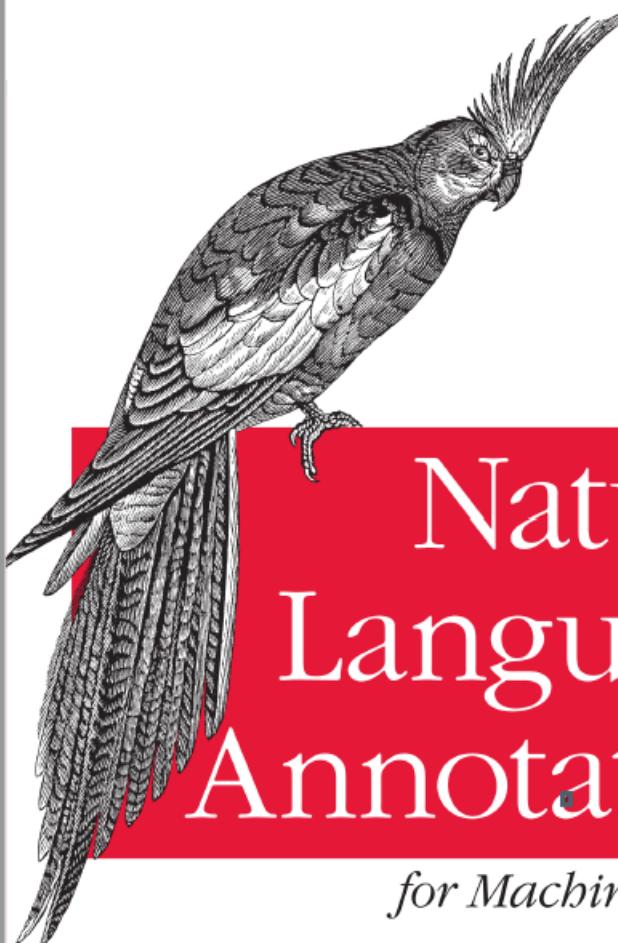
05



Reference books (training corpus)

06

A Guide to Corpus-Building for Applications



- > Preface
- > Chapter 1. The Basics
- > Chapter 2. Defining Your Goal and Dataset
- > Chapter 3. Corpus Analytics
- > Chapter 4. Building Your Model and Specification
- > Chapter 5. Applying and Adopting Annotation Standards
- > Chapter 6. Annotation and Adjudication
- > Chapter 7. Training: Machine Learning
- > Chapter 8. Testing and Evaluation
- > Chapter 9. Revising and Reporting
- > Chapter 10. Annotation: TimeML
- > Chapter 11. Automatic Annotation: Generating TimeML
- > Chapter 12. Afterword: The Future of Annotation
- > Appendix A. List of Available Corpora and Specifications

Reference books (linguistics)



MORGAN & CLAYPOOL PUBLISHERS

07 Linguistic Fundamentals for Natural Language Processing

*100 Essentials from
Morphology and Syntax*

Emily M. Bender

**SYNTHESIS LECTURES ON
HUMAN LANGUAGE TECHNOLOGIES**

Graeme Hirst, *Series Editor*

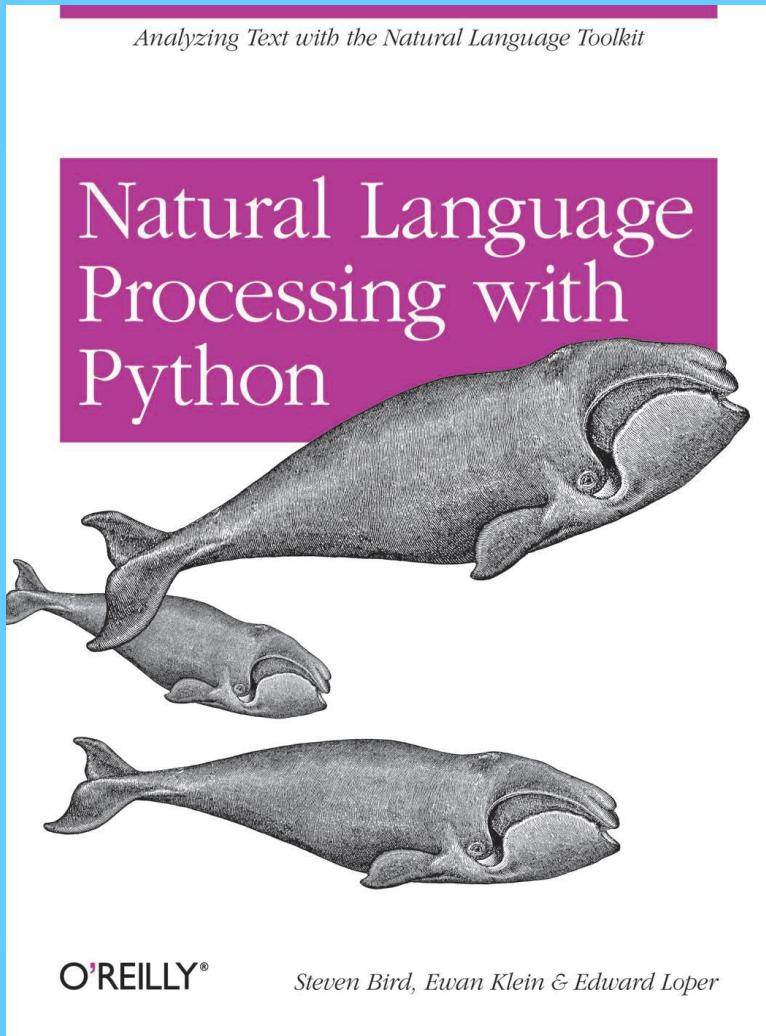
- > [Introduction/motivation](#)
- > [Morphology: Introduction](#)
- > [Morphophonology](#)
- > [Morphosyntax](#)
- > [Syntax: Introduction](#)
- > [Parts of speech](#)
- > [Heads, arguments and adjuncts](#)
- > [Argument types and grammatical functions](#)
- > [Mismatches between syntactic position and semantic roles](#)
- > [Resources](#)

Reference books (Vietnamese)

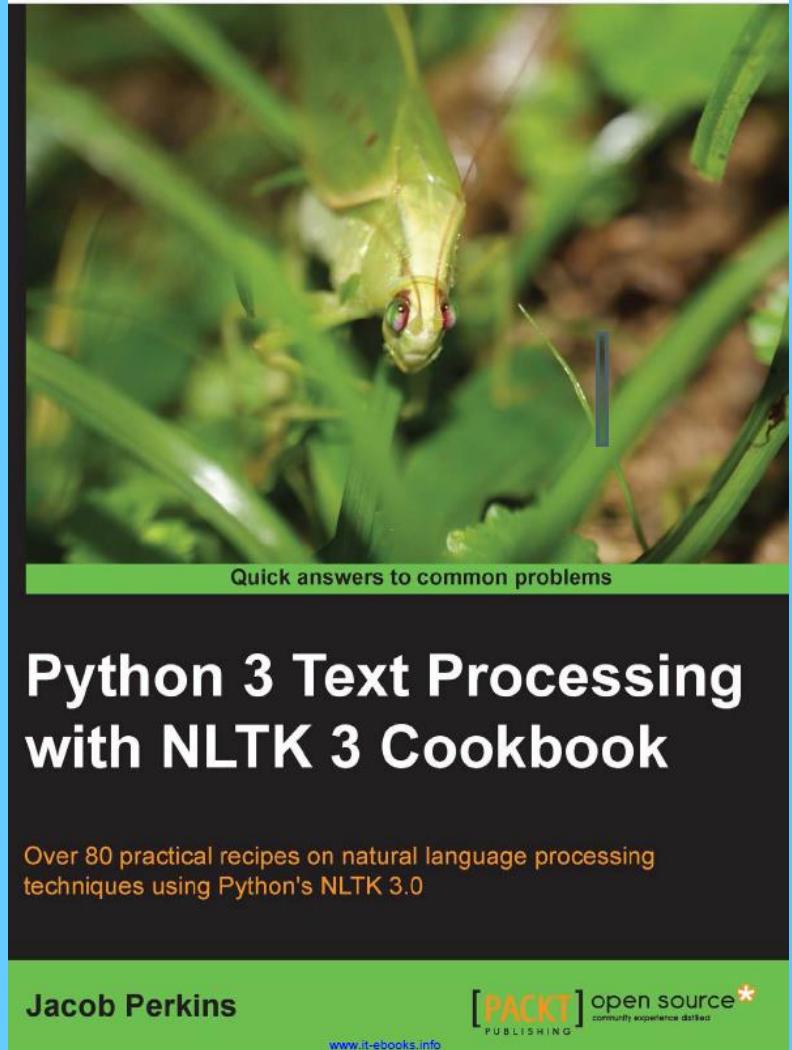


Reference books (programming)

08



09



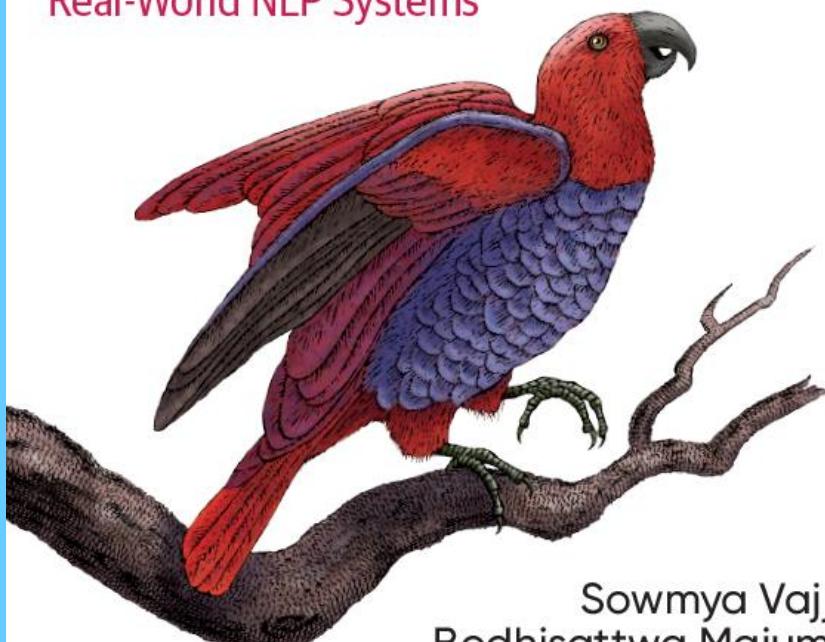
Reference books (applications)

10

O'REILLY®

Practical Natural Language Processing

A Comprehensive Guide to Building
Real-World NLP Systems



Sowmya Vajjala,
Bodhisattwa Majumder,
Anuj Gupta & Harshit Surana

- > Preface
- ▽ Part I. Foundations
 - > Chapter 1. NLP: A Primer
 - > Chapter 2. NLP Pipeline
 - > Chapter 3. Text Representation
- ▽ Part II. Essentials
 - > Chapter 4. Text Classification
 - > Chapter 5. Information Extraction
 - > Chapter 6. Chatbots
 - > Chapter 7. Topics in Brief
- ▽ Part III. Applied
 - > Chapter 8. Social Media
 - > Chapter 9. E-Commerce and Retail
 - > Chapter 10. Healthcare, Finance, and Law
- ▽ Part IV. Bringing It All Together
 - > Chapter 11. The End-to-End NLP Process

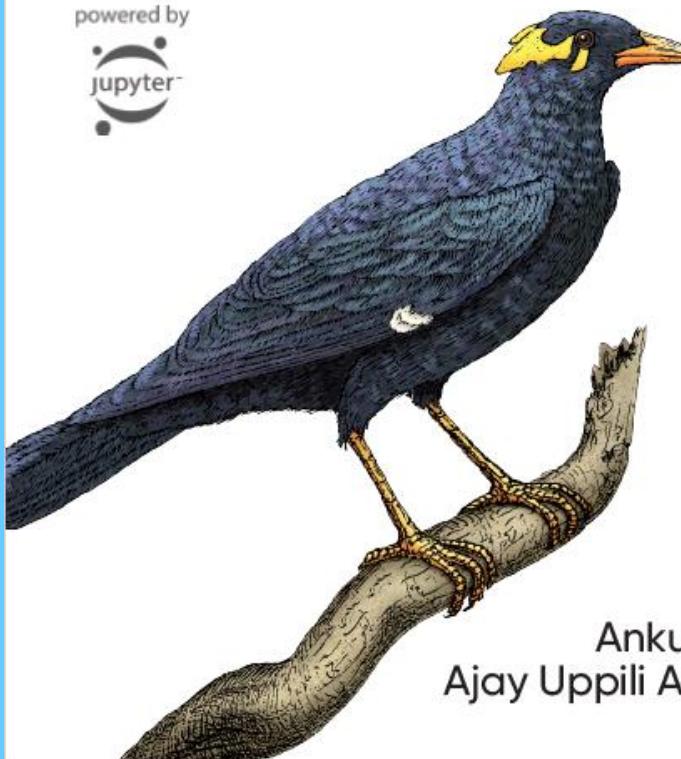
Reference books (applications)

11 O'REILLY®

Applied Natural Language Processing in the Enterprise

Teaching Machines to Read, Write & Understand

powered by



Ankur A. Patel &
Ajay Uppili Arasanipalai

- ▼ Part I. Scratching the Surface
 - > Chapter 1. Introduction to NLP
 - > Chapter 2. Transformers and Transfer Learning
 - > Chapter 3. NLP Tasks and Applications
- ▼ Part II. The Cogs in the Machine
 - > Chapter 4. Tokenization
 - > Chapter 5. Embeddings: How Machines "Understand" Words
 - > Chapter 6. Recurrent Neural Networks and Other Sequence Models
 - > Chapter 7. Transformers
 - > Chapter 8. BERTology: Putting It All Together
- ▼ Part III. Outside the Wall
 - > Chapter 9. Tools of the Trade
 - > Chapter 10. Visualization
 - > Chapter 11. Productionization
 - > Chapter 12. Conclusion
- > Appendix A. Scaling

Reference course (Coursera)

The image shows a screenshot of a video player interface. At the top left is the Coursera logo. Next to it is the text "1. Natural Language Processing, Dan Jurafsky (Stanford University)". Below this is a word cloud graphic with words like "sentences", "shows", "containing", "probability", "grammar", "model", "question", "answers", "discourse", "based", "just", "three", "text", "information", "STATS", "problem", "rule". To the right of the word cloud is the title "Introduction to NLP" in large, bold, black font. Below the title is a video frame showing a man with glasses and a beard speaking. In the bottom left corner of the video frame is a button labeled "MORE VIDEOS".

What is Natural
Language Processing?

→ Main reference slides: Stanford Uni.

Reference websites (papers, dataset)

Association for Computational Linguistics (ACL)

ACL Anthology <https://aclanthology.org>

- A Digital Archive of Top Research Papers in CL, NLP.

<EACL> European Chapter of the ACL

<NAACL> North American chapter of the ACL

<EMNLP> Empirical Methods in Natural Language Processing

CoNLL: Conference on Computational Natural Language Learning

COLING: International Conference on Computational Linguistics

❖ Vietnamese NLP:

<https://vlsp.org.vn/> (Vietnamese Language Speech Processing)

<https://www.clc.hcmus.edu.vn/> (Computational Linguistics

Center, Uni. of Science, HCMC-VNU).



Welcome to the ACL Anthology!

The ACL Anthology currently hosts 79296 papers on the study of computational linguistics and natural language processing.

Subscribe to the mailing list
to receive announcements
and updates to the
Anthology.

...with abstracts (17.97 MB)

ACL Events

Non-ACI Events

Computational Linguistics Journal (2022)

Contents

- Computational Linguistics, Volume 48, Issue 1 - March 2022 [10 papers](#)
- Computational Linguistics, Volume 48, Issue 2 - June 2022 [8 papers](#)
- Computational Linguistics, Volume 48, Issue 3 - September 2022 [7 papers](#)

Show all abstracts ▾

↑up

[bib \(full\)](#) Computational Linguistics, Volume 48, Issue 1 - March 2022

[pdf](#) [bib](#)

Obituary: Martin Kay

Ronald M. Kaplan | Hans Uszkoreit

[pdf](#) [bib](#) [abs](#)

To Augment or Not to Augment? A Comparative Study on Text Augmentation Techniques for Low-Resource NLP

Gözde Güл Şahin

[pdf](#) [bib](#) [abs](#)

Novelty Detection: A Perspective from Natural Language Processing

Tirthankar Ghosal | Tanik Saikh | Tameesh Biswas | Asif Ekbal | Pushpak Bhattacharyya

[pdf](#) [bib](#) [abs](#)

Improved N-Best Extraction with an Evaluation on Language Data

Johanna Björklund | Frank Drewes | Anna Jonsson

[pdf](#) [bib](#) [abs](#)

Linguistic Parameters of Spontaneous Speech for Identifying Mild Cognitive Impairment and Alzheimer Disease

Veronika Vincze | Martina Katalin Szabó | Ildikó Hoffmann | László Tóth | Magdolna Pákáski | János Kálman | Gábor Gosztolya

[pdf](#) [bib](#) [abs](#)

Deep Learning for Text Style Transfer: A Survey

Di Jin | Zhijing Jin | Zhiting Hu | Olga Vechtomova | Rada Mihalcea