

NLP is your friend
NLP – Introduction to DS



20181 – ICT - HUST

NLP Introduction

- NLP Introduction
- Word segmentation – Maximum matching
- POS tagging – HMMs

Course Information

- Web page:
<http://is.hust.edu.vn/~hieunk/courses/IT4142E/>
- Lecturer: Nguyen Kiem Hieu
- Email: hieunk@soict.hust.edu.vn
- Office: 1002 B1

NLP is your friend

What is NLP?

Main tasks in NLP

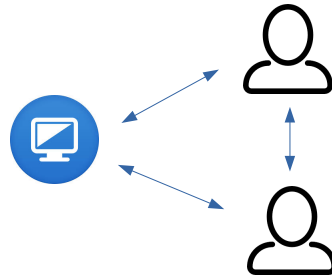
Main challenges in NLP

NLP is your friend
What is NLP?

Natural languages are used for communication between human and human, human and machine

Format: speech, text, image

Categorization: natural languages, programming languages



5

NLP is your friend
What is NLP?

Understanding, **modeling** natural languages

Human – machine **communication** using natural languages

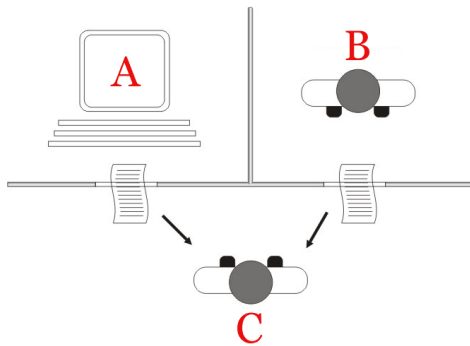
Human – human **communication** using natural languages

Text **mining** for automation and decision making

6

NLP is your friend
What is NLP?

- 1950: Turing test



7

NLP is your friend
What is NLP?

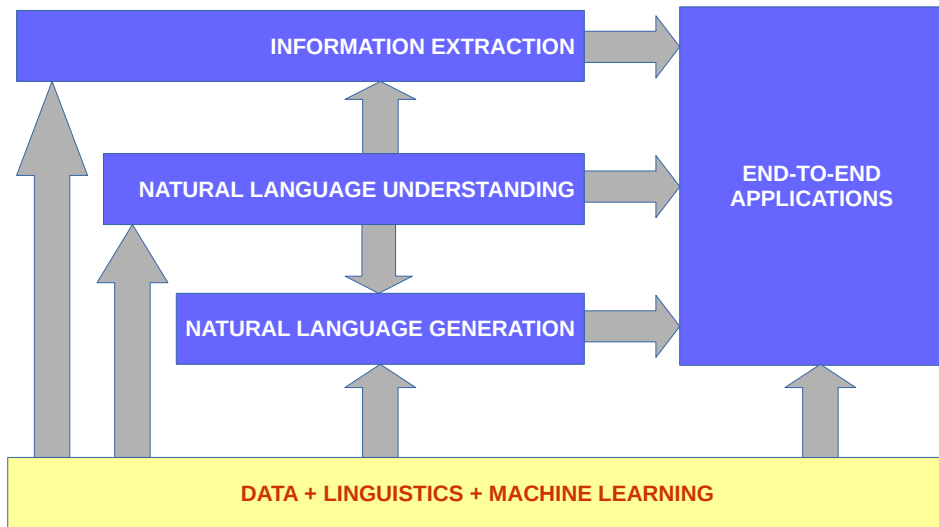
NLP is related to:

Linguistics, psychology, philosophy

Artificial intelligence, machine learning, big data

8

NLP is your friend
Main tasks in NLP



9

NLP is your friend
Main tasks in NLP
Natural language understanding

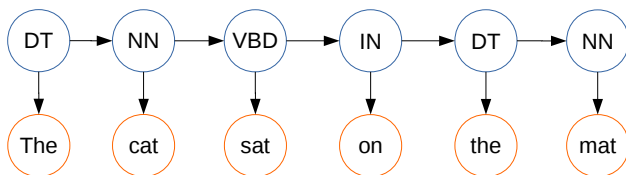
- Word segmentation

“xử lý ngôn ngữ tự nhiên”
(process language natural)

10

NLP is your friend
Main tasks in NLP
Natural language understanding

- Part-of-speech tagging



11

NLP is your friend
Main tasks in NLP
Natural language understanding

- Phrase chunking

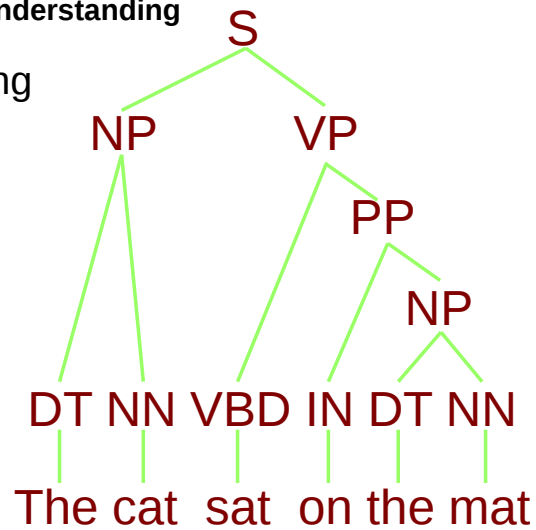
[The cat]_{NP} [sat]_{VP} [on]_{PP} [the mat]_{NP}

12

NLP is your friend
Main tasks in NLP

Natural language understanding

- Constituent parsing

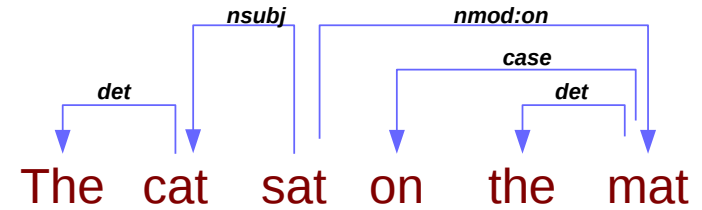


13

NLP is your friend
Main tasks in NLP

Natural language understanding

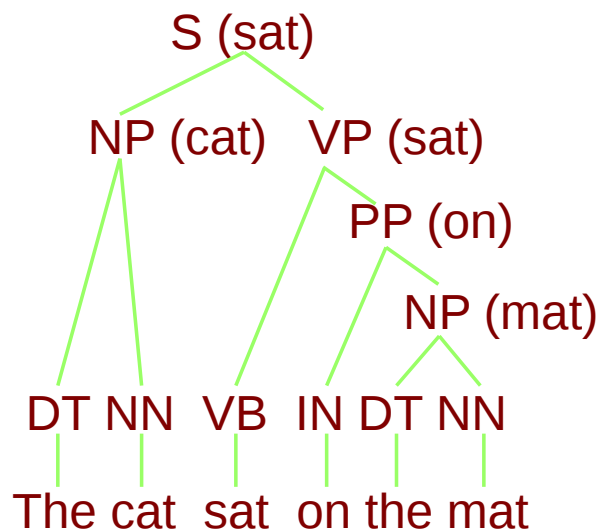
- Dependency parsing



14

NLP is your friend
Main tasks in NLP

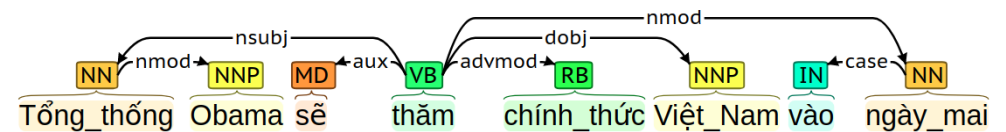
Natural language understanding



15

NLP is your friend
Main tasks in NLP

Natural language understanding



16

NLP is your friend

Main tasks in NLP

Natural language understanding

- Co-reference resolution

doc#1

Alexis Sanchez stepped up his preparations for Manchester United's FA Cup clash against Yeovil when he checked in for training on Thursday. The Chile star is in line to make his debut for the Red Devils in the fourth round tie at Huish Park following his move to Old Trafford. Sanchez met his new team-mates for the first time on Wednesday and could go straight into the matchday squad to face the Glovers. Jose Mourinho has so far given no indication on the strength of team that he will take to Somerset, but Sanchez will be keen to make his debut.

doc#2

Alexis Sanchez checks in for training as he prepares for his Manchester United debut

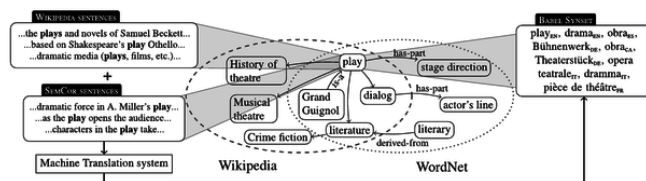
17

NLP is your friend

Main tasks in NLP

Natural language understanding

- Multilingual concept net



19

NLP is your friend

Main tasks in NLP

Natural language understanding

- Semantic parsing

“Mouse love Rice”



“The history of computer mouse”



18

NLP is your friend

Main tasks in NLP

Natural language generation

- Language models

“You are uniformly charming!” cried he, with a smile of associating and now and then I bowed and they perceived à chaise and four to wish for.

Random sentence generated from a Jane Austen trigram model

From Dan Jurafsky 2018

20

NLP is your friend
Main tasks in NLP
Natural language generation

"Arts"	"Budgets"	"Children"	"Education"
NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

The William Randolph Hearst Foundation will give \$1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. "Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services," Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center's share will be \$200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive \$400,000 each. The Juilliard School, where music and the performing arts are taught, will get \$250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual \$100,000 donation, too.

21

NLP is your friend
Main tasks in NLP
Information extraction

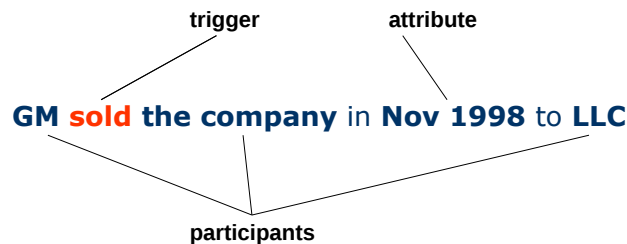
- Named entity recognition

[GM]_{ORG} sold the company in [Nov 1998]_{TIME} to [LLC]_{ORG}

22

NLP is your friend
Main tasks in NLP
Information extraction

- Event extraction



23

NLP is your friend
Main tasks in NLP
Information extraction

- Entity linking



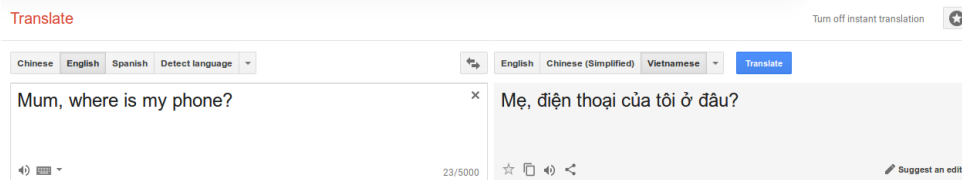
Alexis Sanchez stepped up his preparations for Manchester United's FA Cup clash against Yeovil when he checked in for training on Thursday. The Chile star is in line to make his debut for the Red Devils in the fourth round tie at Huish Park following his move to Old Trafford. Sanchez met his new team-mates for the first time on Wednesday and could go straight into the matchday squad to face the Glovers. Jose Mourinho has so far given no indication on the strength of team that he will take to Somerset, but Sanchez will be keen to make his debut.



24

NLP is your friend
Main tasks in NLP
End-to-end applications

- Machine translation



25

NLP is your friend
Main tasks in NLP
End-to-end applications

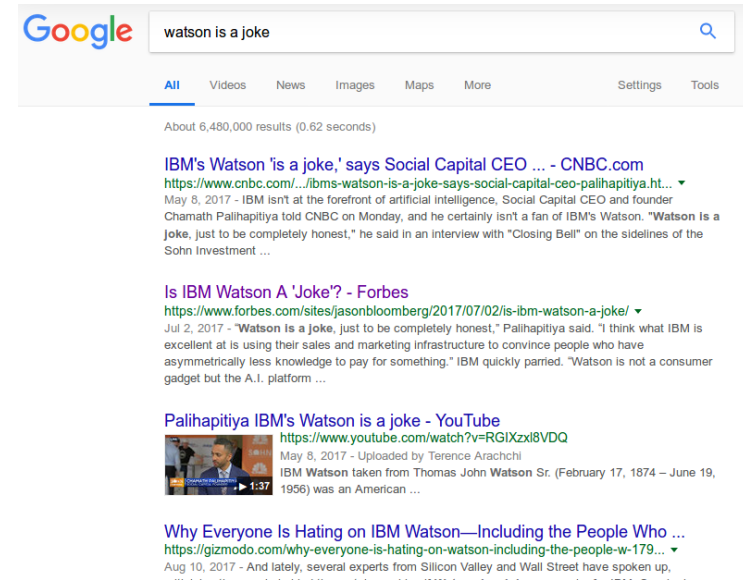
- Chatbot



Life is not easy, even for a bot.

27

NLP is your friend
Main tasks in NLP
End-to-end applications



26

NLP is your friend
Main challenges in NLP

- Languages are ambiguous

28

NLP is your friend

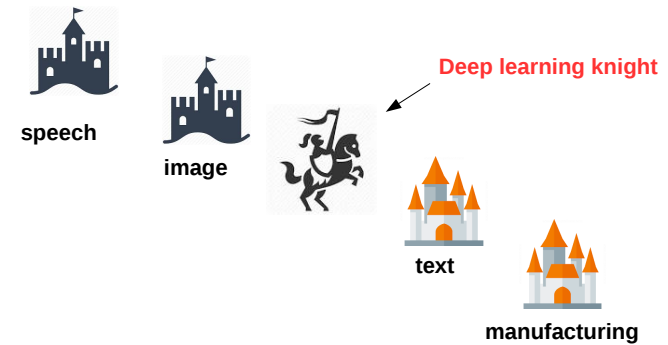
Main challenges in NLP

- Languages are ambiguous
- Languages are diverse and dynamic

NLP là Bạn

Các thách thức chính trong NLP

- Languages are ambiguous
- Languages are diverse and dynamic
- How to apply machine learning to NLP tasks?



29

30

Q&A

hieunk@soict.hust.edu.vn

31