Natural Language Processing

第四周 自然语言处理介绍

庞彦

yanpang@gzhu.edu.cn

Spring 2023

Overview

CONTENTS

01 Natural Language Processing

自然语言处理介绍

Spring 2023

Why is NLP important?

Why is NLP important?　　为什么自然语言处理这么重要?

> "Language understanding is the crown jewel in the field of artificial intelligence“
> 语言理解是人工智能领域皇冠上的明珠
>
> Bill Gates

Why is NLP important?

Language

NLP

Why is "natural language" processing?

Natural language is the expression commonly used in daily life, which is what people usually mean by “speaking.“ 讲人话

Natural language: I have a little camel on my back. 我驼背。
(unnatural language: my back is curved) 背部呈弯曲状

Natural language: The baby‘s agent talk to baby’s baby.
(a lot of this kind on Weibo) 微博上这种段子一大把

Core Task

Natural Language Understanding – NLU 自然语言理解
Natural language generation – NLG 自然语言生成

Natural Language Understanding

> 自然语言理解就是希望机器像人一样，具备正常人的语言理解能力

Challenge of NLU

自然语言理解的5个难点：

1.Language diversity　　语言的多样性
2.Language ambiguity　　语言的歧义性
3.Language robustness　　语言的鲁棒性
4.Language knowledge dependence语言的知识依赖
5.Language context　　语言的上下文

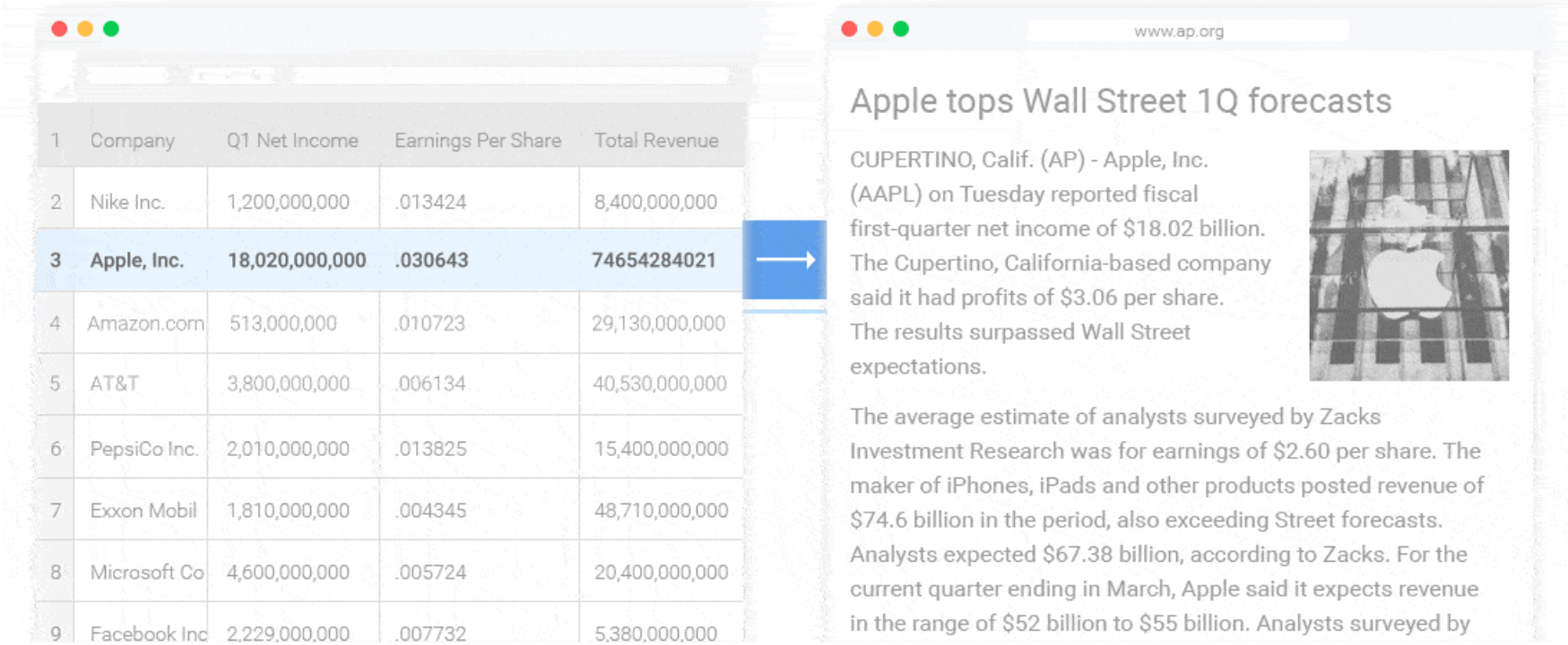
Challenge of NLU

Application of NLU

Natural Language Generation

Natural Language Generation

自然语言生成 – NLG 有2种方式:
1.text – to – text： 文本到语言的生成
2.data – to – text ： 数据到语言的生成



Step of NLG


NLG 的6个步骤
1
2
3
4
5
6
内容确定
Content Determination
文本结构
Text Structuring
句子聚合
Sentence Aggregation
语法化
Lexicalisation
参考表达式生成
REG
语言实现
Linguistic Realisation

Application of NLG

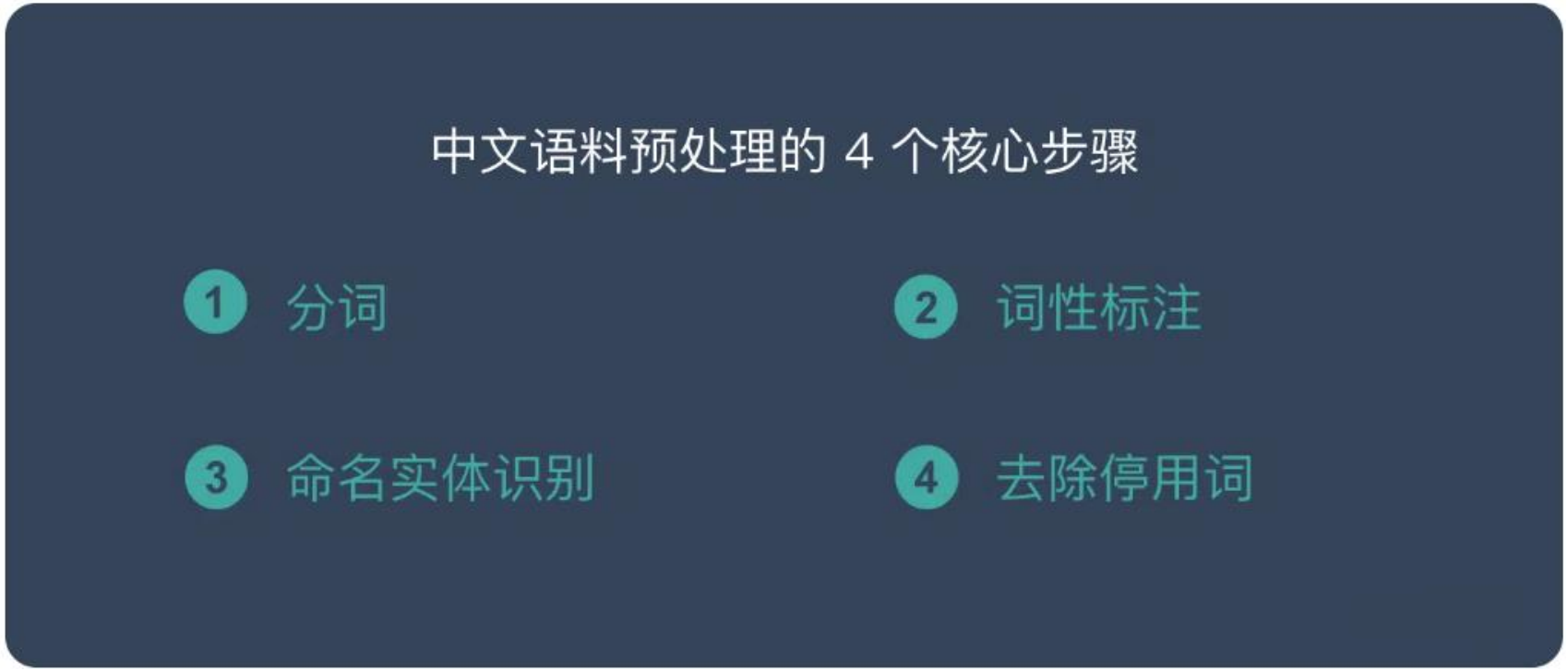
Summary

Traditional NLP

Deep Learning NLP

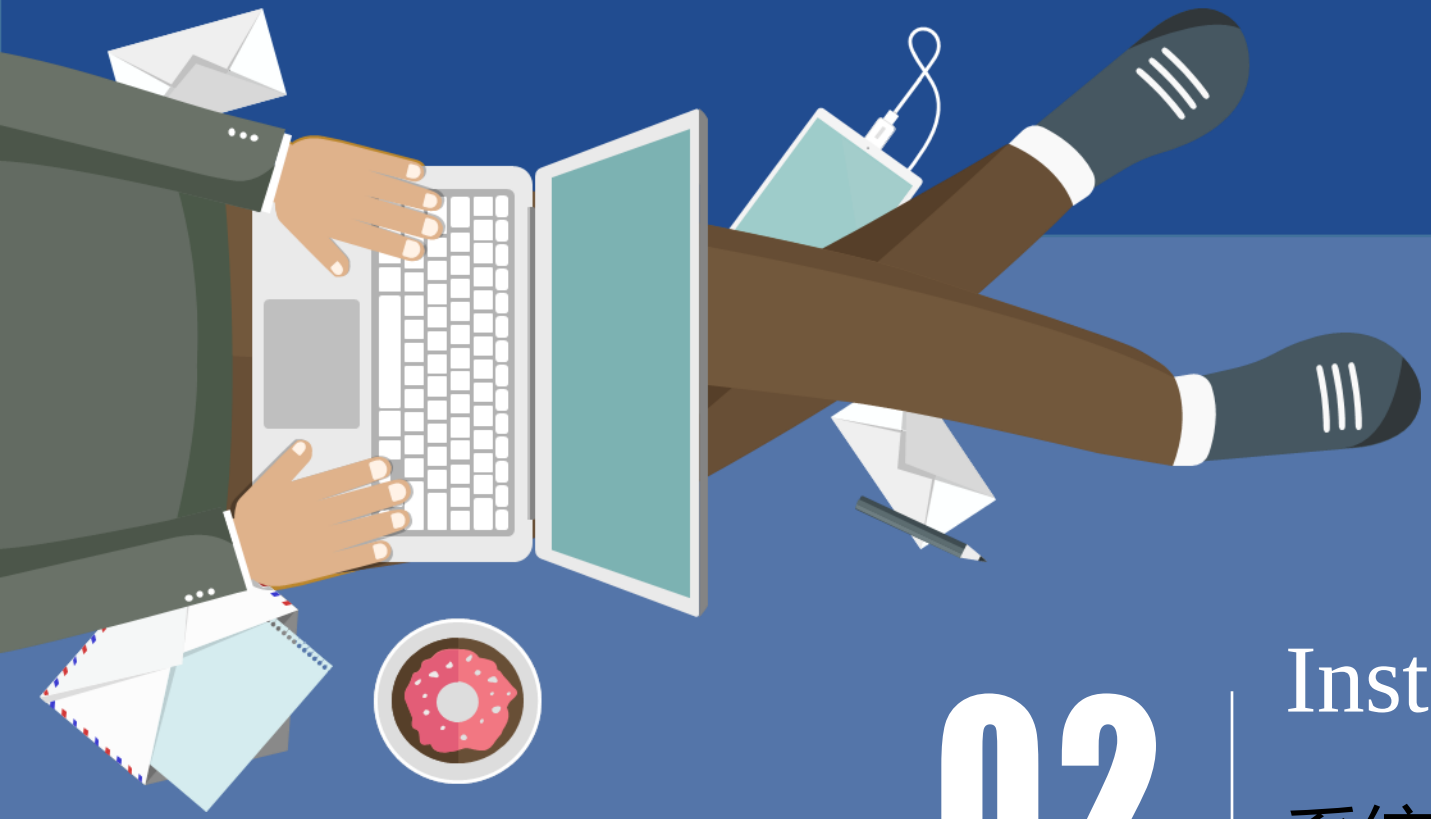
English NLP corpus preprocessing

Chinese NLP corpus preprocessing



Chinese vs English

Chinese vs English



02 Install packages

系统搭建

Spring 2023

Ubuntu

- Install Ubuntu 20.04 **alongside** Windows https://ubuntu.com/;
  安装双系统

Or

- Install Ubuntu 20.04 inside Windows (started from the APP store)
  Windows系统里面直接安装Ubuntu

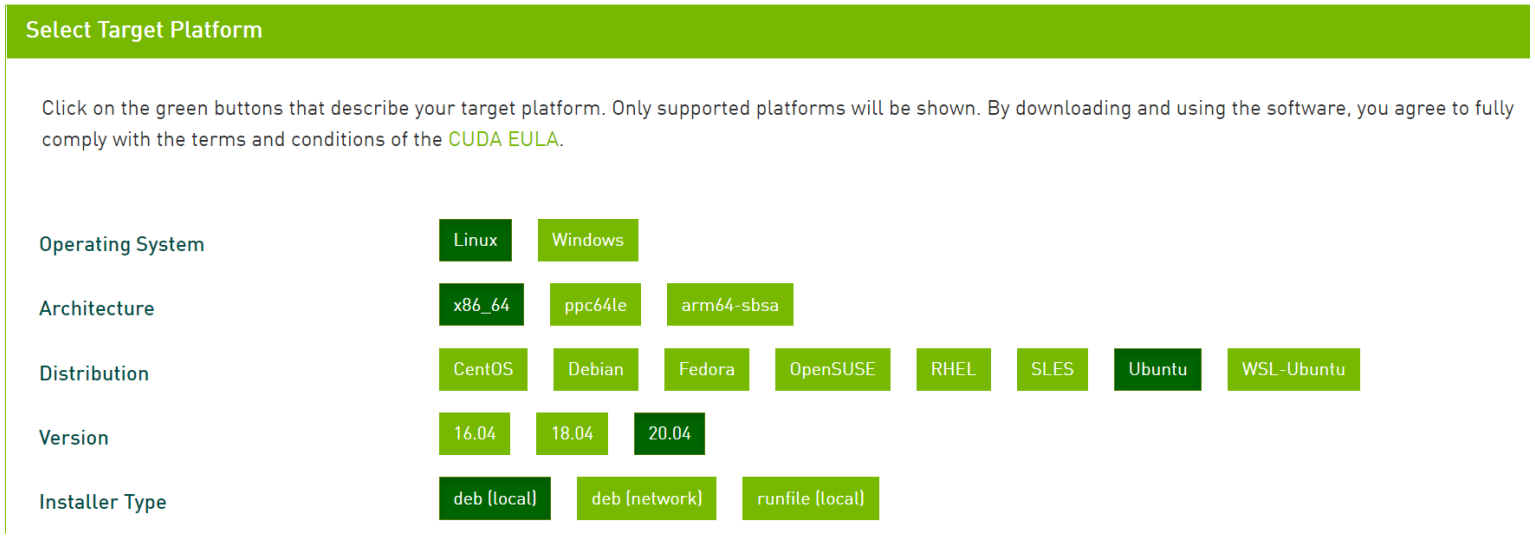
Terminal

- Type Ctrl-Alt-T to open your terminal. 打开终端

Only for GPU

- For GPU, install the **graphic driver, CUDA 11.3 and the corresponding cudnn** under your Ubuntu System. 安装显卡驱动、CUDA、cudnn库

- Ref. https://www.linkedin.com/pulse/install-keras-based-tensorflow-gpu-caffe-gpu-ubuntu-1804-yan-pang/



Conda

- Install Anaconda or Mini Conda under your Ubuntu System
  安装Anaconda或者Mini Conda

- Ref. https://www.anaconda.com/

Create a new environment

- Create a new environment under the installed Anaconda
  在conda下面创建环境

- Ref. https://conda.io/projects/conda/en/latest/user-guide/tasks/manage-environments.html

Python

- Install Python 3.9 or 3.10 under created environment.

PyTorch

- Use commands, pip and conda, to install Pytorch (version depends on your Graphic Card), and other tools under the same environment.
  利用pip或conda命令安装PyTorch（注意选择好版本）

- Ref. https://pytorch.org/get-started/locally/

VS Code

- Install VS Code under the Ubuntu System. You will use this IDE to build all projects in the future. 安装VS Code编译器（建议）

- Ref. https://code.visualstudio.com/download

Q&A