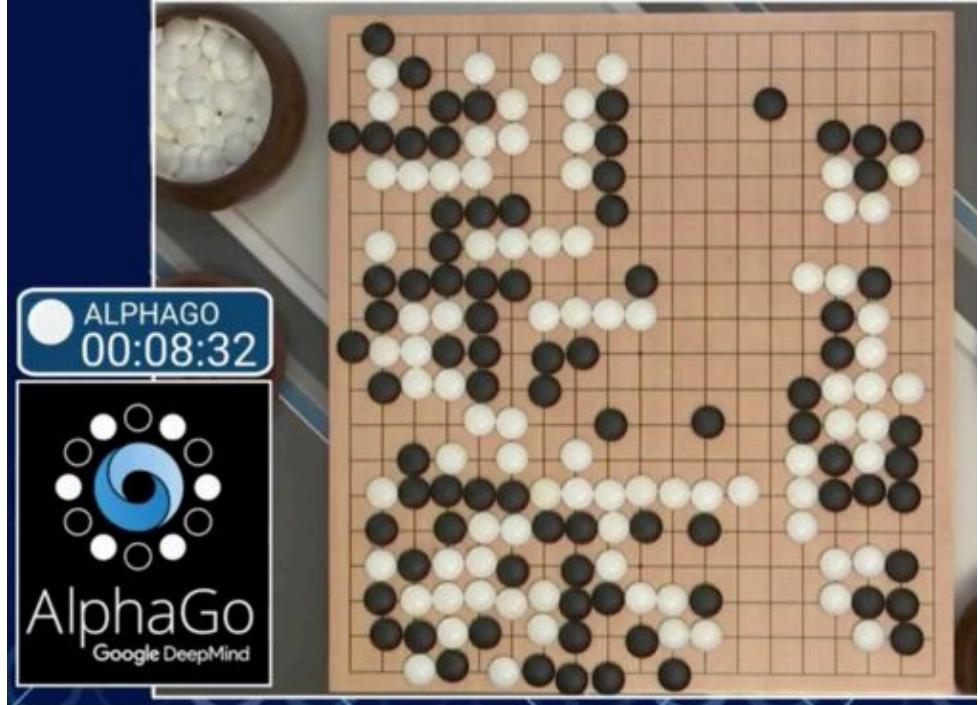


Artificial Intelligence For NLP Lesson-01

开课吧人工智能(与前沿技术学院)

2019.Sept.28

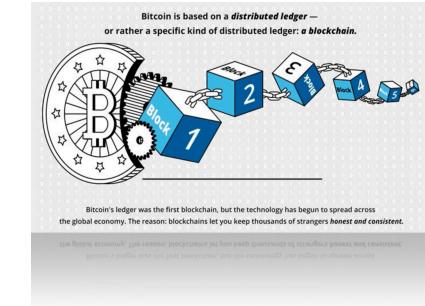
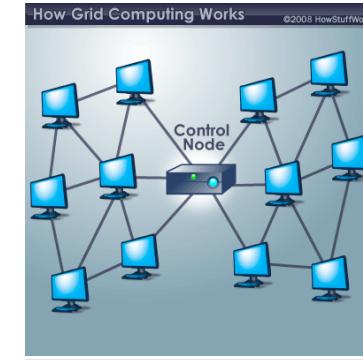
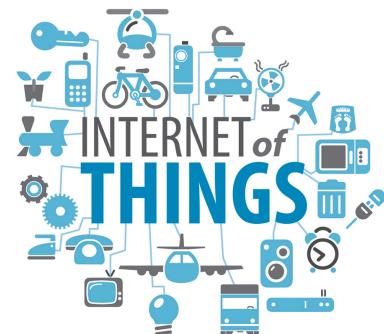
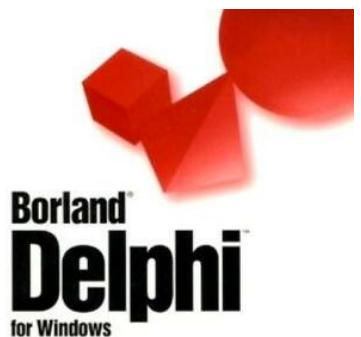


Outline

- i. Course Background
- ii. AI Introduction
- iii. Syntax Tree, Probability Language Model
- iv. 1st Assignment.

1/5 Background

Things change, with time goes by.



2/5 AI Introduction

AI is not of magic but mathematical and programming



Our mentors



李明晓 (NLP课程导师)
Catholic University of
Leuven PHD

曾获国家奖学金赴欧洲交流；曾获欧盟伊拉姆斯全额奖学金。现为丰田欧洲研发中心研究员，与鲁汶大学、剑桥大学、德国马普所、苏黎世联邦理工、捷克布拉格理工合作研发无人驾驶项目。研究方向为自然语言处理，深度学习，多模态对话系统



张希敏 (AI英语提升导师)
ETC全球研究院
普林斯顿大学访问学者

ETC官方认证托福培训师，人事部二级口译员。具备多年会议口译、陪同口译经验，已帮助数百位学生考取托福、雅思高分，成功申请海外名校

- And plenty of references: <https://github.com/Computing-Intelligence/References>

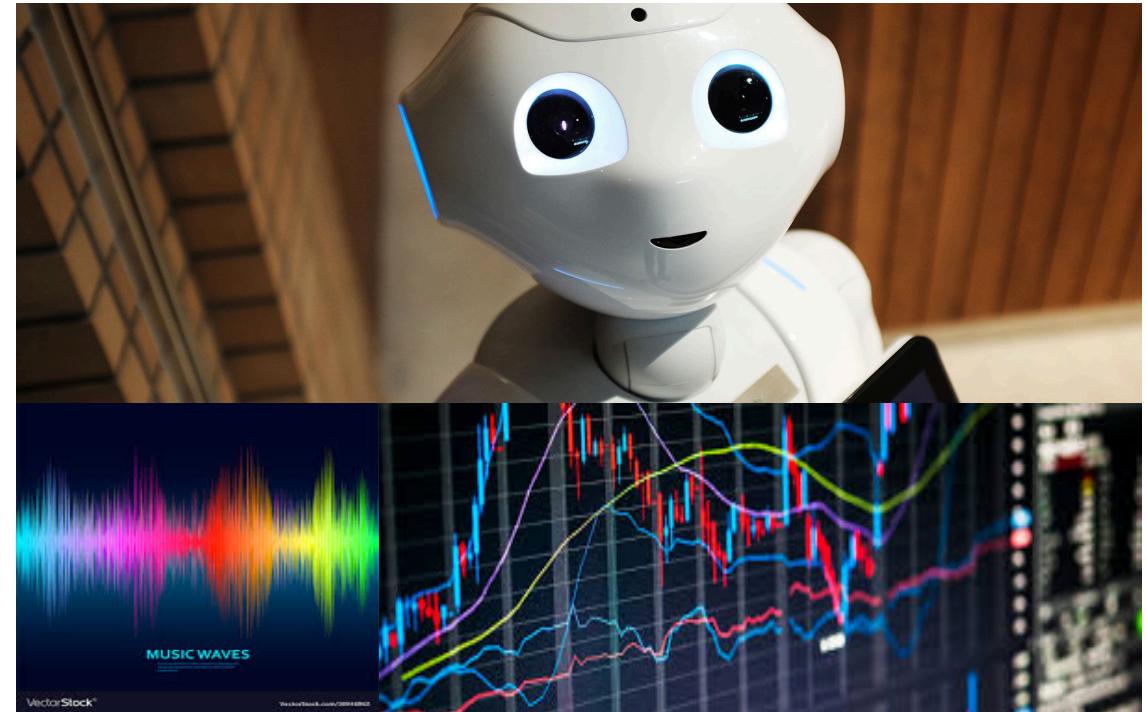
What AI Solved?

- Sources:
 - <https://www.youtube.com/watch?v=vjSohj-lclc>
 - 大众点评 搜索海底捞 选择一家 拨打电话
 - Question: How many *AI methods they using. (split room and talking)*

Question

- Classify the applications or system based dimensions.

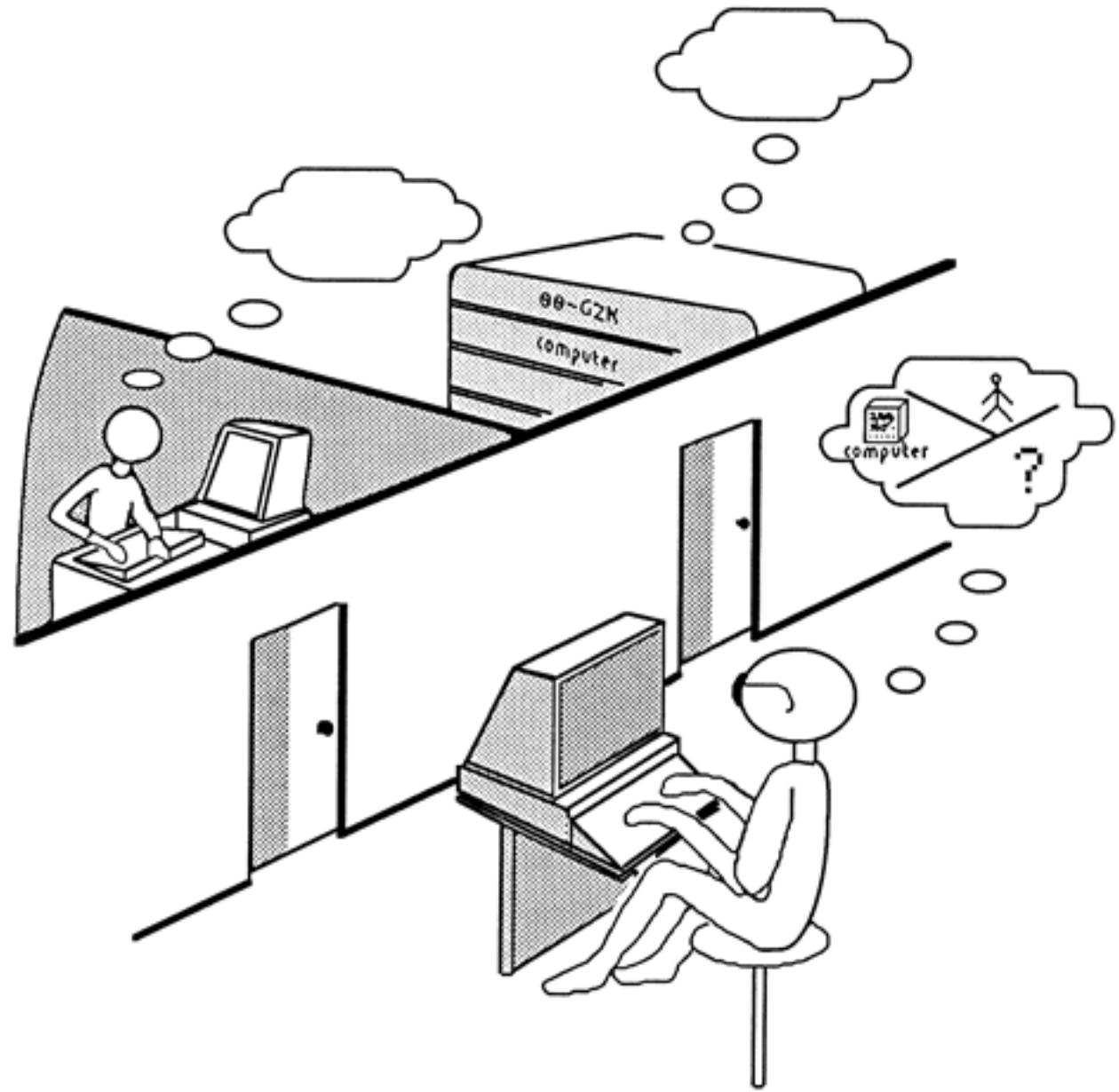
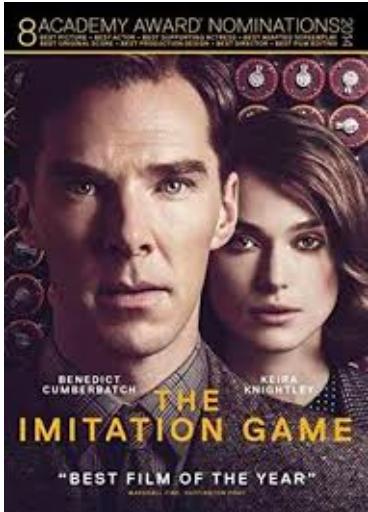
- Auto Composition
- Voice Recognition
- Stock Prediction
- Service Robot
- Anti Money Laundering



And, can you figure out some more?

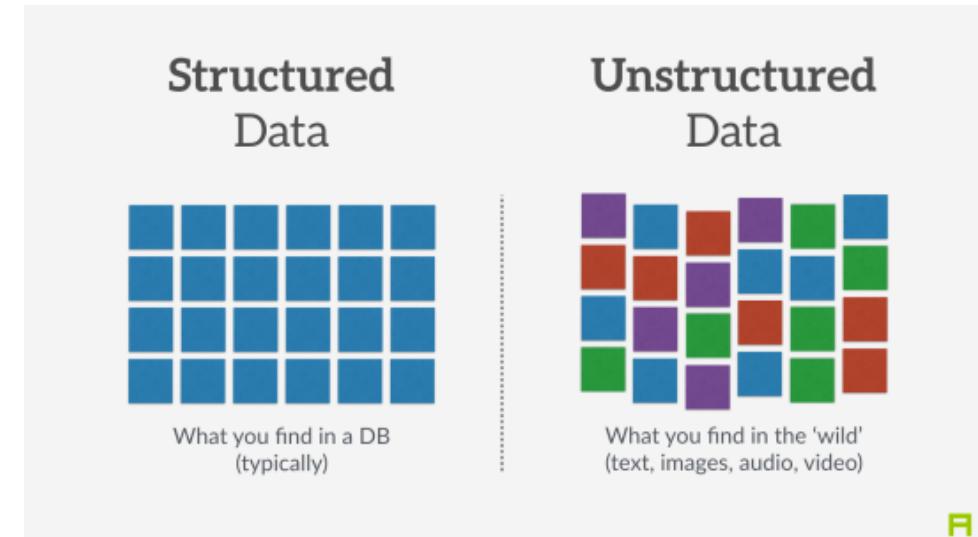
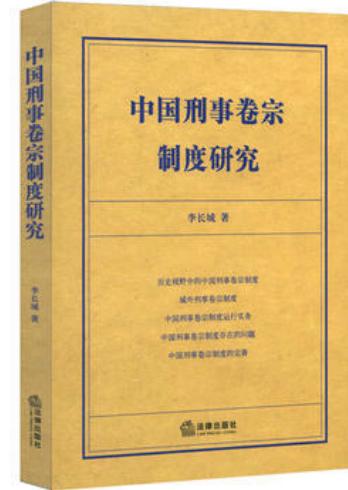
Turing test

- 1. Machine Intelligence
- 2. Imitation Game



Why Natural Language is so hard?

- 1. Text is Logic
- 2. Diversity
- 3. Unstructured
- 4. ...





AI Paradigm

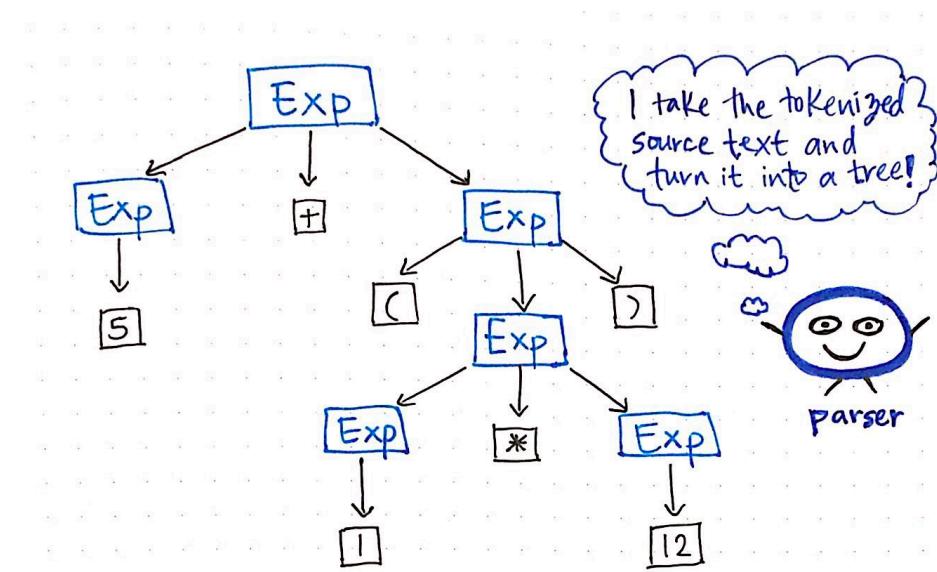
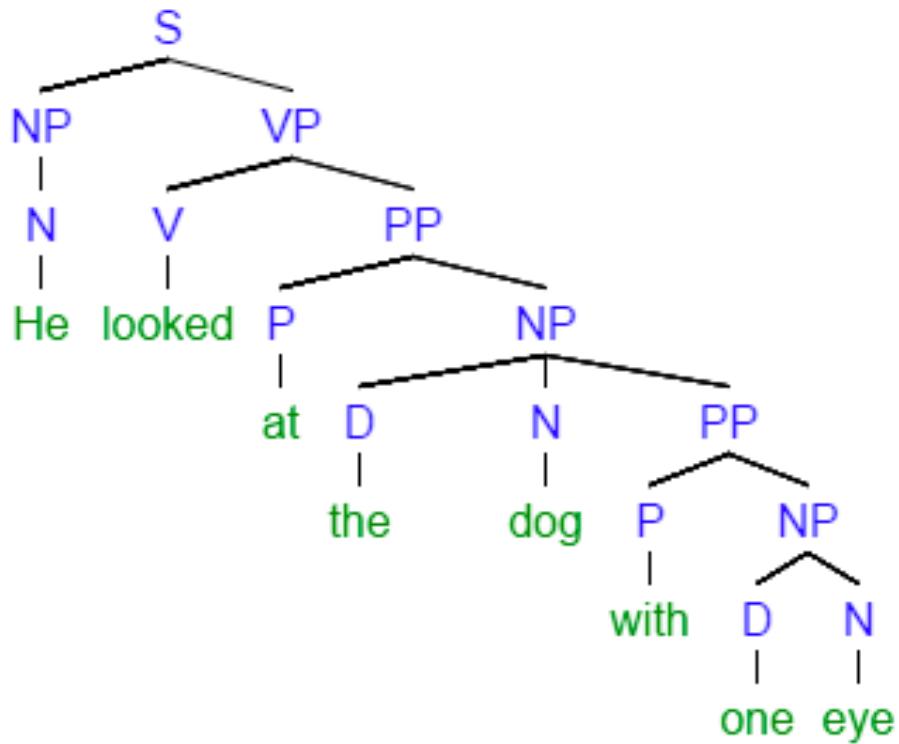
- 1. Rule Based
- 2. Probability Based
- 3. Problem Solving: Search based
- 4. Mathematical or Analytic Based
- 5. Machine Learning (deep learning) Based

1. Rule Based



```
simple_grammar = "''''''  
  
sentence => noun phrase verb_phrase noun_phrase => Article  
Adj* noun  
  
Adj* => null | Adj Adj*  
  
verb_phrase => verb noun_phrase  
Article => 一个 | 这个  
noun => 女人 | 篮球 | 桌子 | 小猫  
verb => 看着 | 坐在 | 听着 | 看见  
Adj => 蓝色的 | 好看的 | 小小的  
.....
```

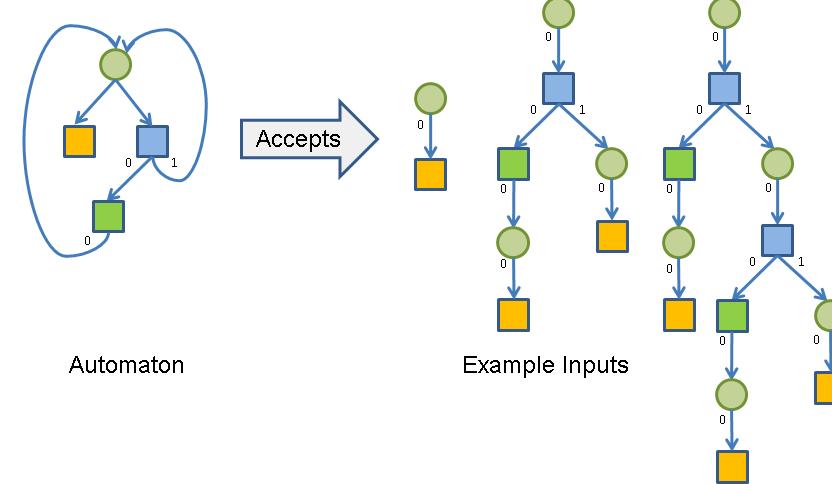
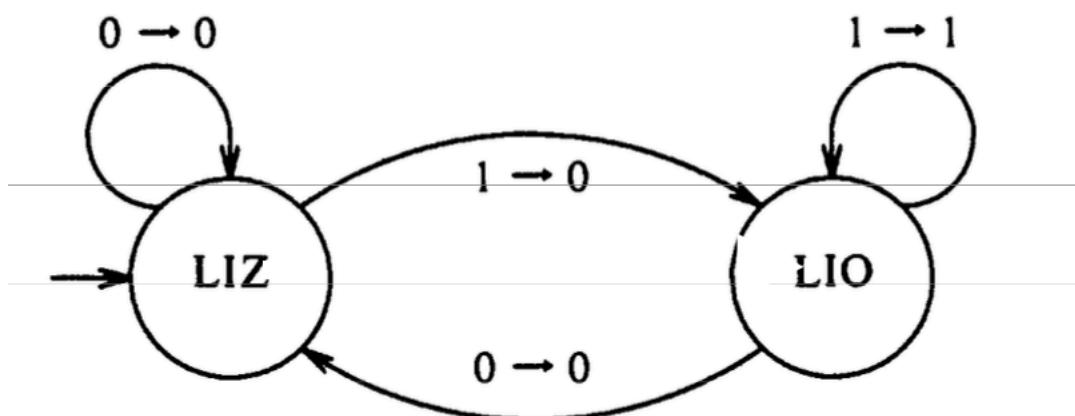
1. Rule Based: Syntax Tree



*First comes the **Lexical analysis** phase, followed by the **syntax analysis** phase, which will generate a **parse tree**.

Automata

- Input: 011010111
- Output: 001000011



- LIZ: Last Input Zero, LIO: Last Input One

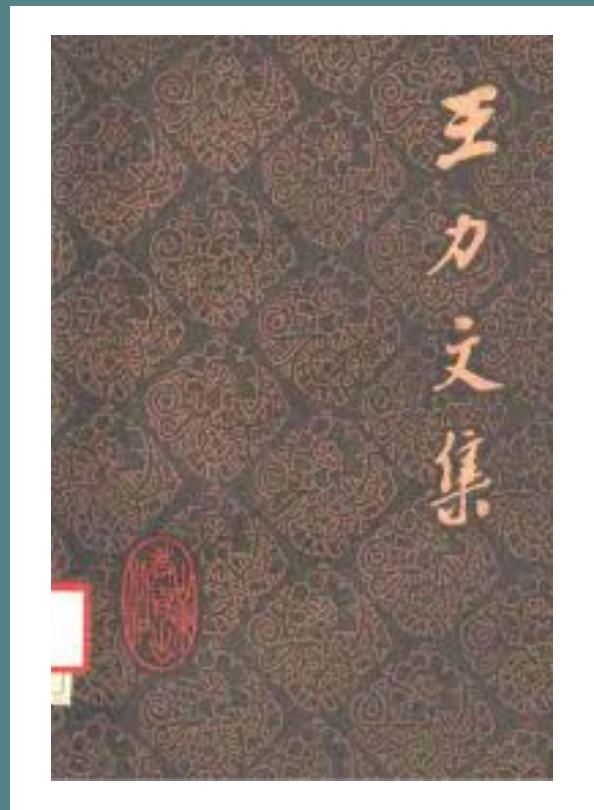
2. Probability Based

- A1. 前天早上吃晚饭的时候
- A2. 前天早上吃早饭的时候

- B1. 正是一个好看的小猫
- B2. 真实一个好看的小猫

- C1. 我无言以对，简直
- C2. 我简直无言以对

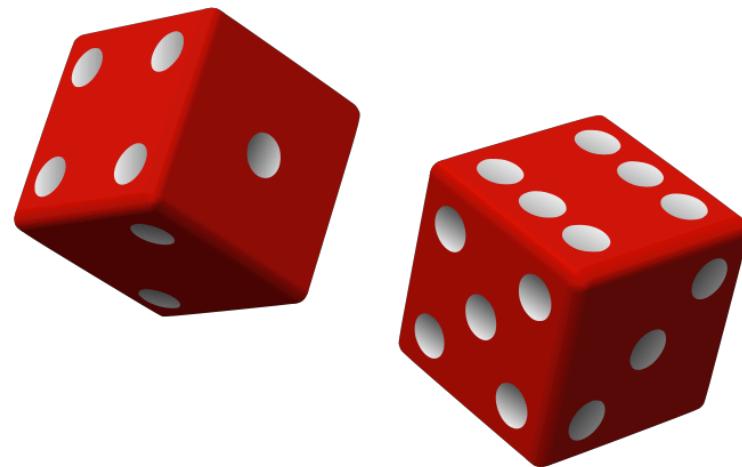
名称	定义	语法特点	类别	举例
(一) 名词	表示人或事物名称的词	①前面可以加数量词(一副对联) ②前面不能加不、很之类的副词(不行联、很对联)。 ③后面不能加状态助词“了”(楹联了)	①具体名词 ②抽象名词	人、牛、山、水、对联 友谊、立场、观点、思想
(附) 方位词	中表示方向位置的词	常用在名词或名调性短语的后面		东、西、南、北、前、后、中间、下边
(二) 代词	具有替代或指示作用的词分	①能够替代或指示替代或各类实词。 ②一般不带修饰成分。	①人称代词 ②指示代词 ③疑问代词	我、你、他、我们、你们 这、那、这里、那边 谁、什么、哪、多少
		①前面可以加副词(刚走、很想)。	①不及物动词	醒、病、游行、觉悟



2. Probability Based

- "Every time I fire a linguist, the performance of the speech recognizer goes up"

----- Frederick Jelinek (18 November 1932 - 14 September 2010)



2. Probability Based: Language Model

- Language Model:
The probability of
a sentence.

N-gram Models

1-gram (Unigram)

$$P(w_i) = \frac{C(w_i)}{\sum_{\forall k} C(w_k)} = \frac{C(w_i)}{N}$$

of tokens

2-gram (Bigram)

$$P(w_{i+1}|w_i) = \frac{C(w_i, w_{i+1})}{\sum_{\forall k} C(w_i, w_k)} = \frac{C(w_i, w_{i+1})}{C(w_i)}$$

Assignment-01

- 1. Self Review
- 2. Build Sentence Generation System Using Syntax Tree and Language Model
- 3. (Optional) Chat Bot Using Pattern
- 4. (Optional) Reading Turing's Machine Intelligence Paper:
See in our github