
自然语言处理 (Natural Language Processing, NLP)

汇报人： 杨夕

2019.11.23

目录

一 自然语言是什么

二 NLP 是什么

三 NLP 有哪些应用

四 目前公司所用的技术有哪些

五 NLP 所用编程语言有哪些

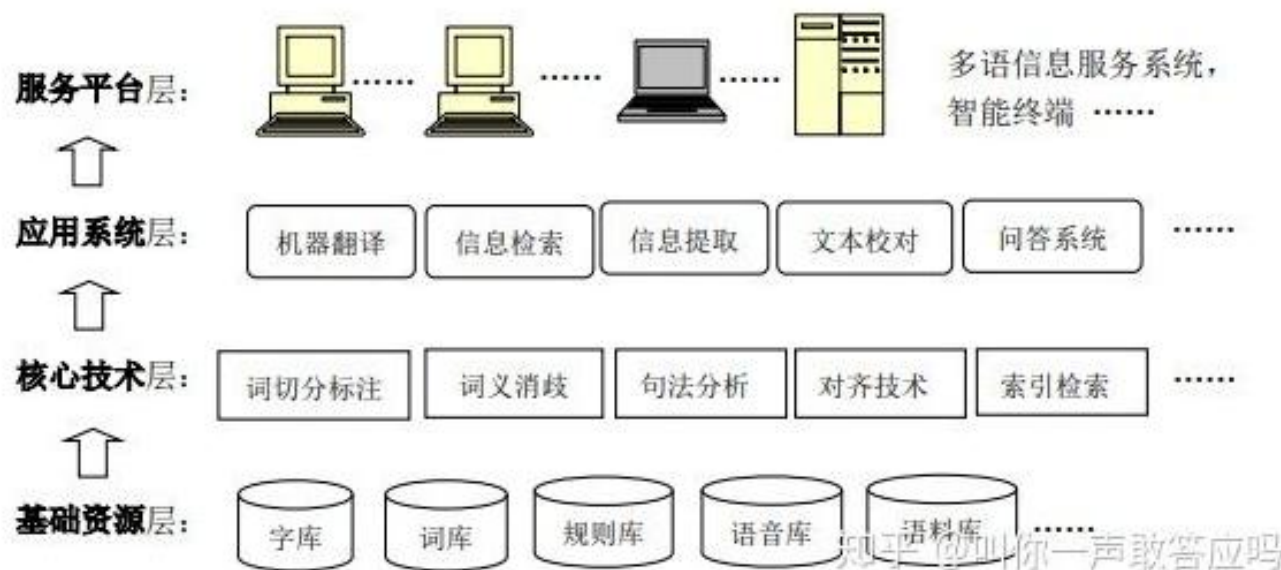
六 NLP 所用技术有哪些

语言出现的目的就是为了人类之间的通信。

交流双方只需了解对方的编解码规则，既可实现不同语种之间的交流，但人机之间交互的情况却复杂得多。究其原因，除了计算机性能因素之外，更重要的原因在于自然语言的复杂性。

用计算机来分析和生成自然语言（文本、语音），目的是让人类可以用自然语言形式跟计算机系统进行人机，交互，从而更便捷、有效地进行信息管理。

中文自动分词 (Chinese word segmentation)
命名实体识别 (Named Entity Recognition)
文本分类 (Text categorization)
关系抽取 (Relationship extraction)
问答系统 (Question answering)
语音识别 (Speech recognition)
机器翻译 (Machine translation)
问答系统 (Question and answer system)
对话系统 (Dialogue system)



中文自动分词（Chinese word segmentation）

命名实体识别（Named Entity Recognition）

文本分类（Text categorization）

关系抽取（Relationship extraction）

人类语言|是|一套|用来|传达|含义|的|系统|。

命名实体识别（Named Entity Recognition，简称NER），又称作“专名识别”，是指识别文本中具有特定意义的实体，主要包括人名、地名、机构名、专有名词等。简单的讲，就是识别自然文本中的实体指称的边界和类别。

04 目前公司所用的技术有哪些? —— 命名实体识别

Tokens	IO	BIO	BIOES	...
5	I-TIME	B-TIME	B-TIME	
号	I-TIME	I-TIME	E-TIME	
特	I-PER	B-PER	B-PER	
朗	I-PER	I-PER	I-PER	
普	I-PER	I-PER	E-PER	
访	O	O	O	
华	I-LOC	B-LOC	S-LOC	
。	O	O	O	

Tokens	词性
董卿	人名
现身	动词
国家级博物馆	地名
看	动词
展	动词

推出:

(董卿, 现身, 国家级博物馆)

Python (主流)

Java

C++

为什么需要词向量？

因为计算机时无法直接识别文字的，所以需要转化为计算机能够识别的语言。

词向量是什么？

词向量就是将单词映射到词空间的一个点。

词向量有哪些？、

one-hot 向量，word2vec...

5号 特朗普 访华

6号 李伟 加班

	词典[5号 特朗普 访华 6号 李伟 加班]						label
5号 特朗普 访华	1	1	1	0	0	0	1
6号 李伟 加班	0	0	0	1	1	1	0

机器学习方法：朴素贝叶斯（NB）、支持向量机（SVN）

深度学习方法：神经网络

这件商品很赞！ 褒义

这件商品太垃圾了！ 贬义

这件商品质量很好！ 褒义

敬请各位大佬批评指正
