



CCAI 2017
中国人工智能大会



自然语言处理的十个发展趋势

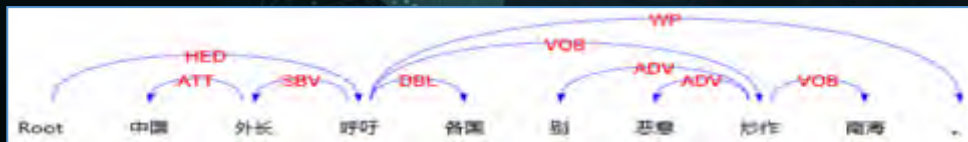
刘挺

哈尔滨工业大学社会计算与信息检索研究中心

2017年7月22日

浅层学习（分层）

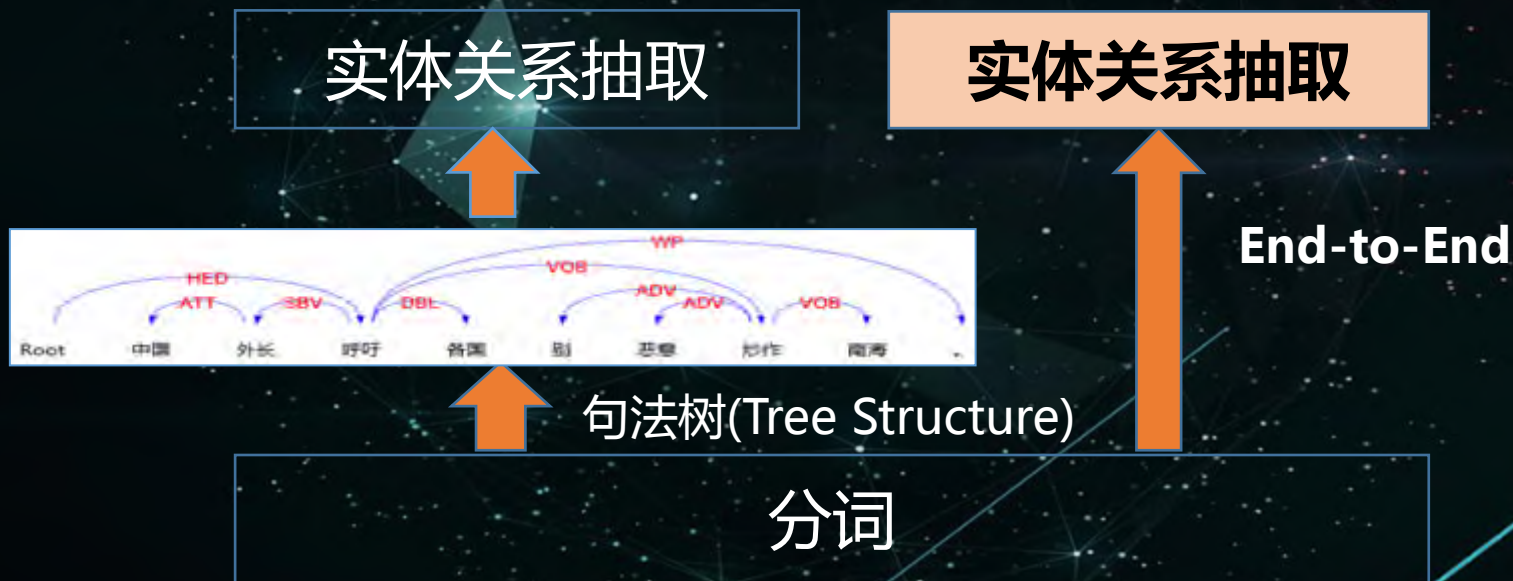
实体关系抽取



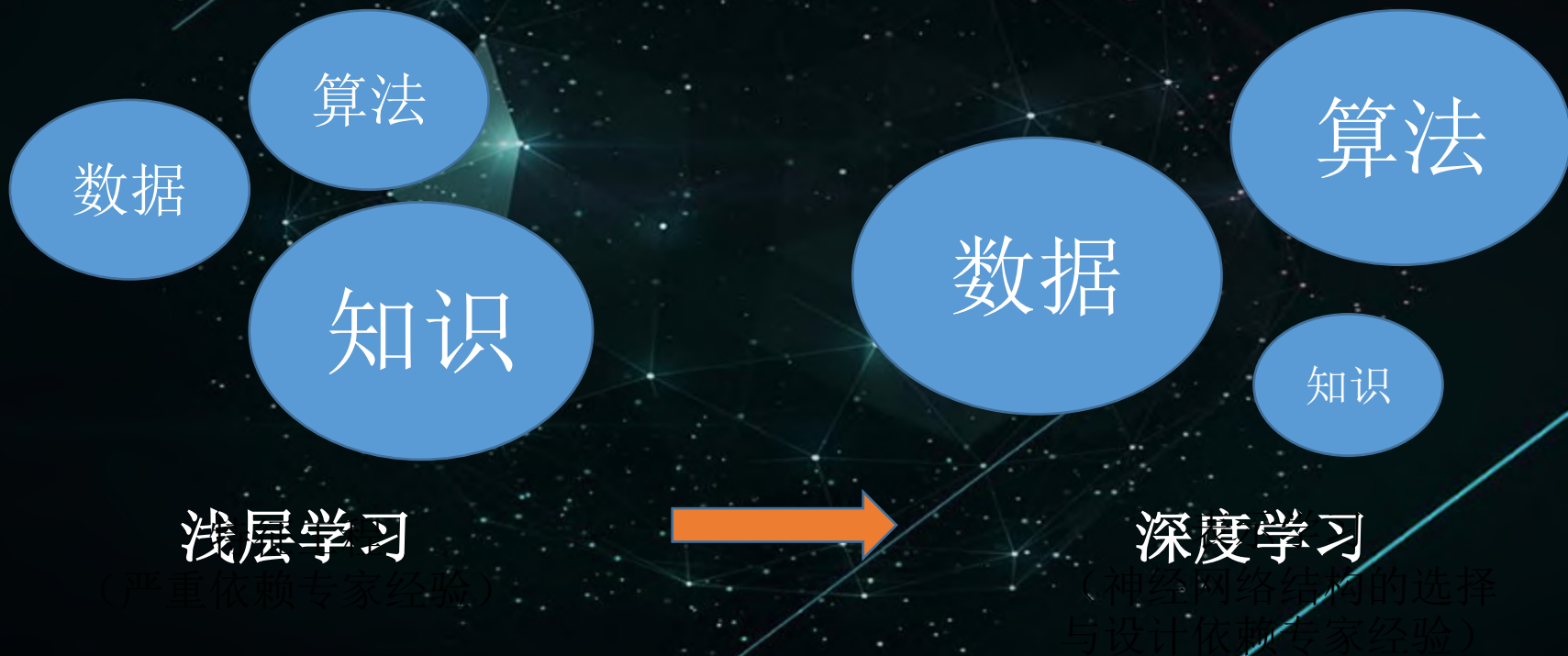
句法树(Tree Structure)

分词

深度学习（端到端）



趋势二：学习模式从浅层学习到深度学习

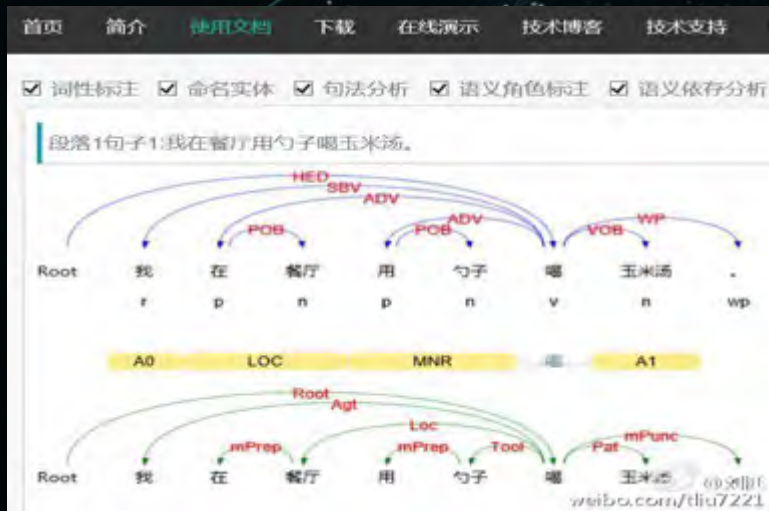


趋势三：NLP平台化从封闭走向开放

首页 简介 使用文档 下载 在线演示 技术博客 技术支持

词性标注 命名实体 句法分析 语义角色标注 语义依存分析

段落1句子1:我在餐厅用勺子喝玉米汤。



weibo.com/tliu7221

哈工大LTP平台
钱伟长一等奖2010
省一等奖2016

Configuration c_i I **booked** a **ticket to** **Google**
Stack Buffer

Action $c_i \rightarrow c_{i+1}$ **SHIFT** **LEFT_ARC** **RIGHT_ARC**

Derivation $c_0, a_0 \rightarrow c_1, a_1 \rightarrow c_2, a_2 \rightarrow c_3, a_3 \rightarrow c_4, a_4$

谷歌SyntaxNet



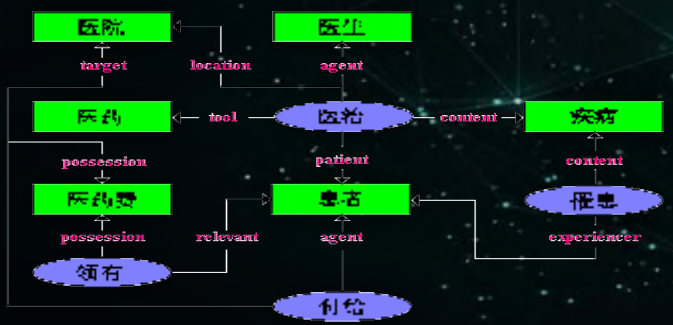
代词 动词 数词 量词 副词 形容词 副词 形容词 助词 名词

腾讯文智平台

趋势四：语言知识从人工构建到自动构建



北大语法词典
(俞士汶教授)
国家二等奖



知网/HowNet
(董振东先生)



大词林
(哈工大SCIR)

知识问答：图灵测试的各种翻版



2017年，搜狗“智能狗”
挑战《一站到底》



2014年，百度“小度”
挑战《芝麻开门》

2011年，IBM Watson
挑战《危险边缘》

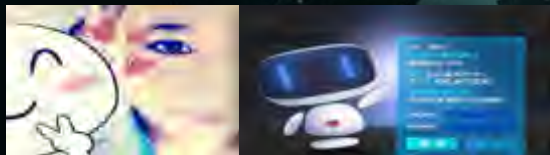


趋势五：对话机器人 从通用到场景化



场景化
任务执行
(2016年至今)

2016年
哈工大SCIR-笨笨



聊天机器人 (2014年至今)

2014年, 微软小冰、百度小度



Siri




语音助手 (2012年-2014年)

2012年, 百度、搜狗、讯飞

2011年, 苹果

趋势六：文本理解与推理从浅层分析向深度理解迈进

你问我答
我叫“童童”，今年四岁
我能用一个问题回答问题，考考我吧



《白羊和黑羊》
一天，一只白羊从南面上了一座木桥，一只黑羊从北面上了独木桥。他们同时来到桥当中，白羊说：“你退回去，让我先过桥！”黑羊说：“你退回去，让我先过桥！”它们谁也不肯让谁，就打了起来。不一会儿，只听到河里“扑通！扑通！”的响声，它们都掉到河里去了。

根据故事内容提问

两只羊掉进哪里去了？

提问

上一个故事 下一个故事

哈工大讯飞
联合实验室 (HFL)
“六龄童阅读理解”

“市议会拒绝给示威者许可证，因为他们害怕暴力。” 请问：谁害怕暴力？

Winograd测试

Original Version	Anonymised Version
<p>Context The BBC producer allegedly struck by Jeremy Clarkson will not press charges against the "Top Gear" host, his lawyer said Friday. Clarkson, who hosted one of the most-watched television shows in the world, was dropped by the BBC Wednesday after an internal investigation by the British broadcaster found he had subjected producer Osin Tymon "to an unprovoked physical and verbal attack."</p>	<p>the <i>em381</i> producer allegedly struck by <i>em212</i> will not press charges against the "<i>em153</i>" host, his lawyer said friday. <i>em212</i>, who hosted one of the most - watched television shows in the world, was dropped by the <i>em381</i> wednesday after an internal investigation by the <i>em180</i> broadcaster found he had subjected producer <i>em193</i> "to an unprovoked physical and verbal attack."</p>
<p>Query Producer X will not press charges against Jeremy Clarkson, his lawyer says.</p>	<p>Producer X will not press charges against <i>em212</i>, his lawyer says.</p>
<p>Answer Osin Tymon</p>	<p><i>em193</i></p>

Google Deepmind: 挖词填空

趋势七：文本情感分析从事实性文本到情感文本



哈工大，微博情绪地图



SMP情感分析论坛

趋势十：NLP+行业 与领域深度结合，为行业创造价值

NLP+教育 (HFL)

辅助判案

案情摘要: 案例3: 盗窃

案情描述: 被告人施某于2015年4月20日11时许, 在铜仁清溪自治县铜仁镇乐群超市地下车库试衣间外衣架上的黑色手持包逃走, 包内有人民币15215元, 棕色卡包, 银行卡等案发后, 被告人施某于2015年4月23日被公安机关抓获归案前款, 赃物追回退还失主

原拟判决: 有期徒刑十个月, 缓刑一年, 罚金15000元
《中华人民共和国刑法》第二百六十四条、第七十二条、第七十三条

关键词: 15215 盗窃 杨仁德

罪名: 盗窃

刑期: 五个月

罚金: 1472150

法律条款: 中华人民共和国刑法第二百六十四条、第五十二条

《中华人民共和国刑法》第二百六十四条
【盗窃】盗窃公私财物, 数额较大的, 或者多次盗窃、入户盗窃、携带凶器盗窃、扒窃的, 处三年以下有期徒刑、拘役或者管制, 并处或者单处罚金数额巨大或者有其他严重情节的, 处三年以上十年以下有期徒刑, 并处罚金; 数额特别巨大或者有其他特别严重情节

作文批改

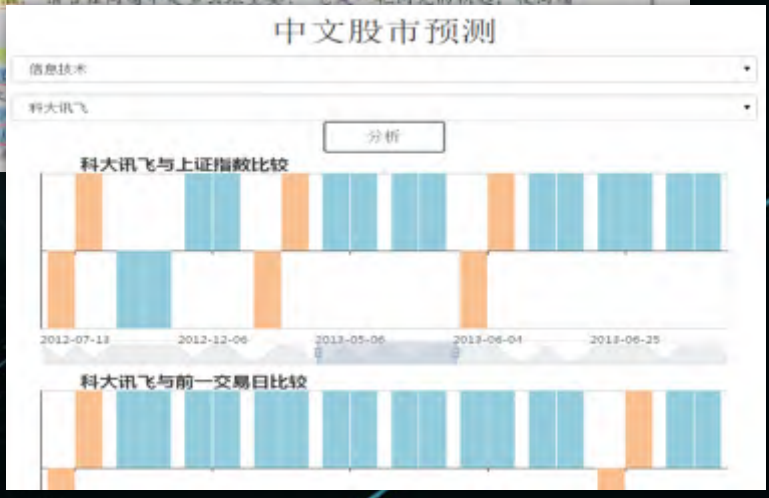
全文语言流畅, 行文自然舒适, 并且内容比较充实, 支撑了作文的整体结构。
在结构上, 段落排比的运用, 既增强了条理性, 也增添了语言气势和表达效果; 并且引用的材料内容详实俱有说服力, 亦易让读者接受, 认同作者观点。

语言是沟通的钥匙
假如沟通是一扇窗, 那么语言就是这扇窗的钥匙。

如果沟通的沟通的另一方面是语言的交流, 那么语言的钥匙就引领着你走向何方, 望满天, 如果沟通的沟通的另一方面是一望无际的沙漠, 那么语言的钥匙就引领着你走向鸟语花香的绿洲; 如果沟通的沟通的另一方面是浩瀚无垠的大海, 那么语言的钥匙就引领着你驶向如金浪翻滚的港湾, 语言在沟通中是多么地重要! 它是一把闪光的钥匙, 使沟通直接到达人的内心深处。

你知道其他的语言事迹吗?
《海天的物》海天的知识, 李白的《蜀道难》一文, 写的老泪纵横之美, 今年, 我也受到友人的赞赏, 所以, 它需要语言为它传达信息。

94分



NLP+司法 (HFL)

NLP+金融
(哈工大)

谢谢！