



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

趋势一：语义表示从符号表示到分布表示

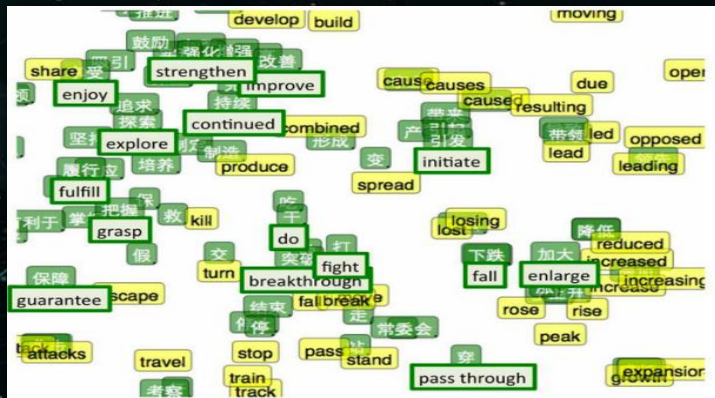
- 符号表示

- 离散、高维、稀疏
- One-Hot表示，词袋表示等



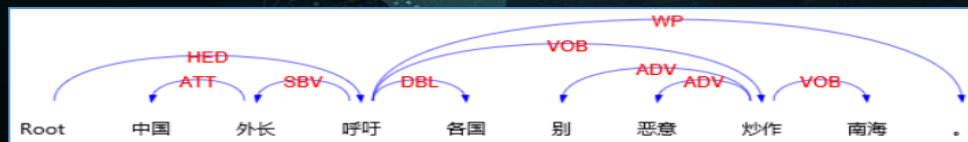
- 分布表示

- 连续、低维、稠密
- 词、短语、句子、篇章
- 便于计算语言单元之间的语义距离



浅层学习（分层）

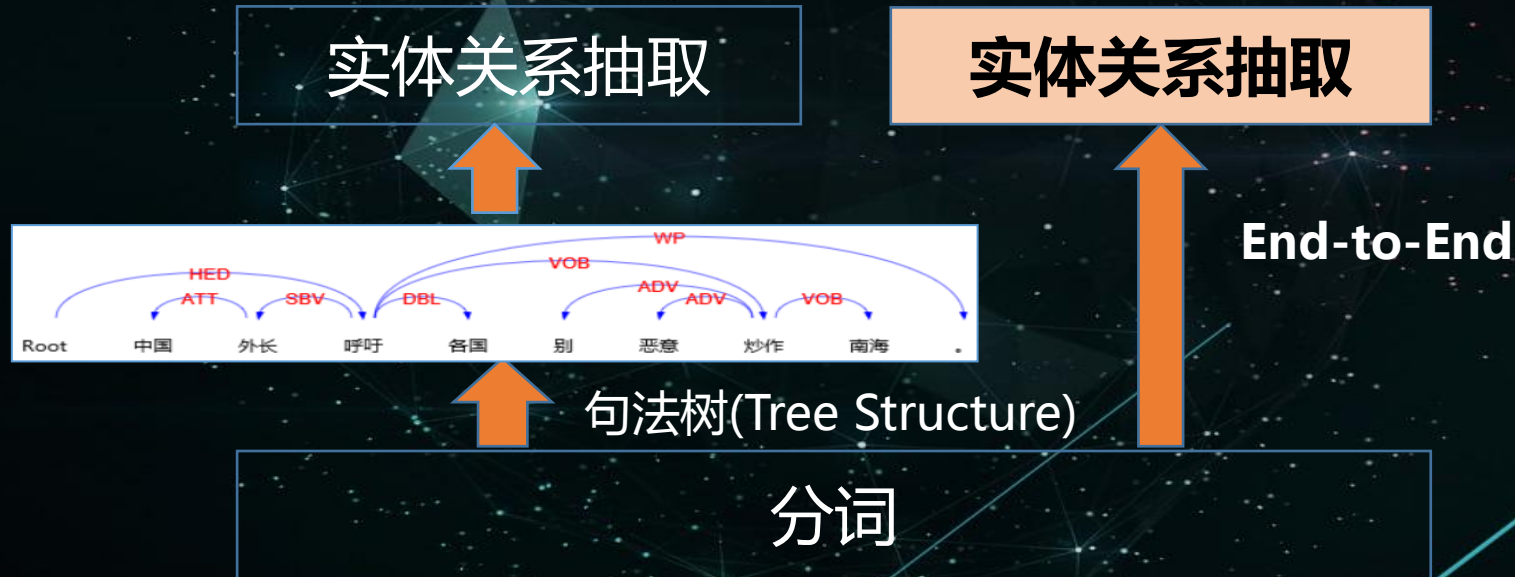
实体关系抽取



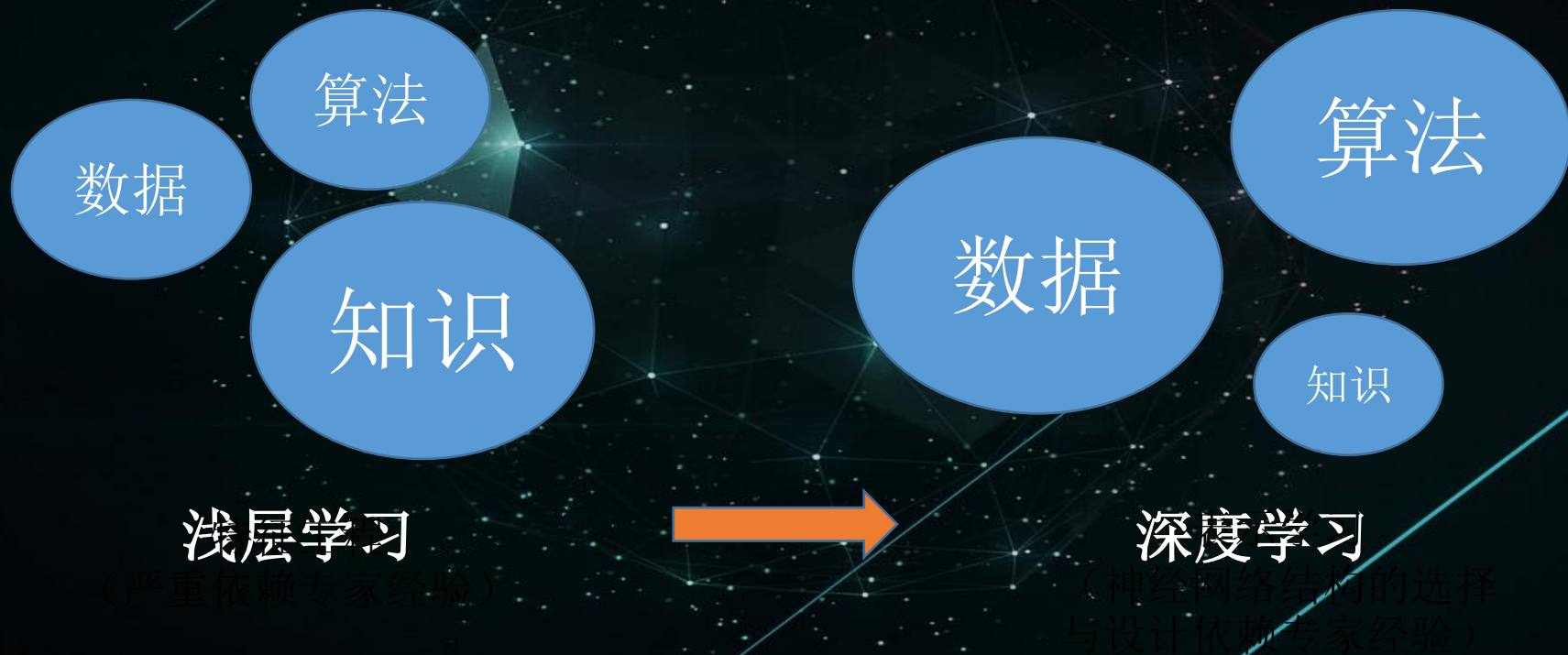
句法树(Tree Structure)

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

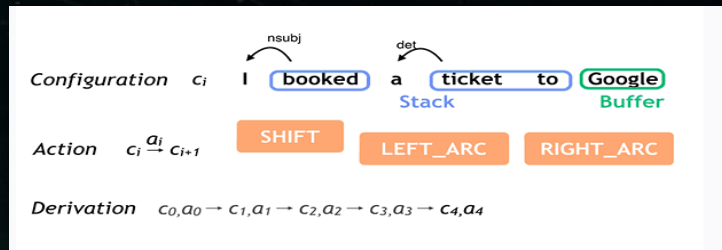
首页 简介 使用文档 下载 在线演示 技术博客 技术支持

☒ 词性标注 ☒ 命名实体 ☒ 句法分析 ☒ 语义角色标注 ☒ 语义依存分析

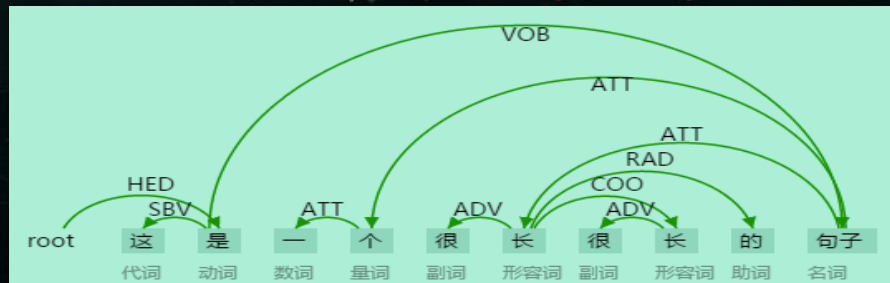
段落1句子1:我在餐厅用勺子喝玉米汤。

weibo.com/tliu7221

哈工大LTP平台
钱伟长一等奖2010
省一等奖2016

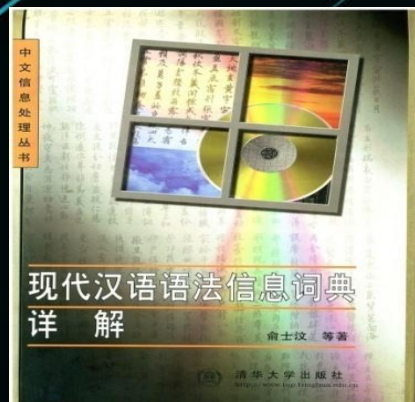


谷歌SyntaxNet



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



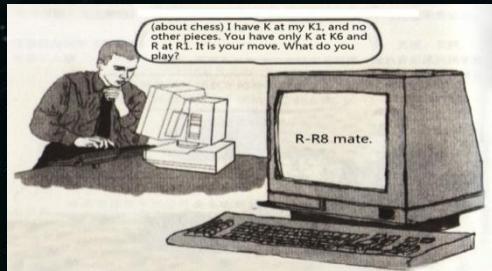
2017年，搜狗“智能狗”
挑战《一站到底》



2014年，百度“小度”
挑战《芝麻开门》



2011年，IBM Watson
挑战《危险边缘》

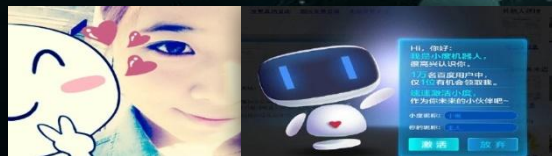


趋势五：对话机器人 从通用到场景化



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




2011年, 苹果



语音助手 (2012年-2014年)

2012年, 百度、搜狗、讯飞

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
我能用一个词回答问题，考考我吧

《白羊和黑羊》

一天，一只白羊从南面上独木桥，一只黑羊从北面上独木桥。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
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 Oisin Tymon “to an unprovoked physical and verbal attack.” ...	the <i>ent381</i> producer allegedly struck by <i>ent212</i> will not press charges against the “ <i>ent153</i> ” host, his lawyer said friday. <i>ent212</i> , who hosted one of the most - watched television shows in the world, was dropped by the <i>ent381</i> wednesday after an internal investigation by the <i>ent180</i> broadcaster found he had subjected producer <i>ent193</i> “to an unprovoked physical and verbal attack.” ...
Query Producer X will not press charges against Jeremy Clarkson, his lawyer says.	Producer X will not press charges against <i>ent212</i> , his lawyer says.
Answer Oisin Tymon	<i>ent193</i>

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

趋势八：社会媒体处理从传统媒体到社交媒体

票房预测



归来

上映日期：2014-05-16

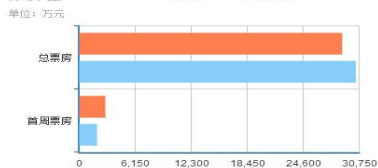
导演：张艺谋

主演：陈道明 / 巩俐 / 张慧雯 / 郭涛

类型：剧情 / 历史 / 爱情

上世纪70年代初，与家人音讯全无隔绝多年的劳改犯陆焉识（陈道明 饰）在一次农场转迁途中逃跑回家。这对隐瞒芭蕾舞梦想的女儿丹凤（张慧雯 饰）带来了巨大压力，她阻止母亲冯婉瑜（巩俐 饰）与父亲的相见。因此夫妻二人近在咫尺却又相隔天涯。文革结束后，陆焉识终于平反回家，但他却发现女儿早已放弃了芭蕾舞梦想成了一名工厂女工，而深爱的妻子冯婉瑜也已经不认识自己。深厚的感情，生活的变故，迫使陆焉识做出了对他来说最荒唐却又最合理的人生选择……本片是张艺谋导演加盟乐视影业后的第一部作品，被编剧邹静之誉为良心之作。©豆瓣

票房值 单位：万元



票房预测



黄金时代

上映日期：2014-10-01

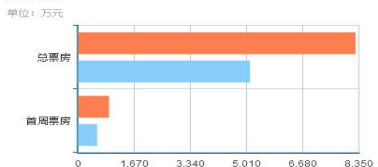
导演：许鞍华

主演：汤唯 / 冯绍峰 / 王传之 / 朱亚文

类型：剧情 / 传记 / 爱情

上世纪二十至四十年代的中国，那是一个民气十足，胸襟天空的时代，一群年轻人经历了一段放任自流的时代，自由地追求梦想与爱情，有人在流离中讨得求容，有人在抗争中企盼家国未来……萧红，一个特立独行的女子，一路流亡，从北方到南方，从哈尔滨到香港，一边躲避战乱，一边经历着令人唏嘘又痛仰心酸的爱情与人生。对生的至强对死的挣扎在她笔下穿流低语，她的人生亦是如此。

票房值 单位：万元



哈工大，电影票房预测

全国社会媒体处理大会
SMP 2016



Input: 无边落木萧萧下
Generated: 一夜寒灯耿耿明
Original: 不尽长江滚滚来

Input: 星垂平野阔
Generated: 月落远林疏
Original: 月涌大江流

Input: 秋雁

Generated poem:
白蘋江上惊秋雁
红蓼洲边起暮鸦
遥指翠微亭下路
行人不见武陵花

作古体诗（清华孙茂松）

BEFORE

[illegible]

橄榄球赛自动报道

(美联社+Automated Insights公司)

作文题：“挫折”（哈工大SCIR）

逆境磨练意志。懦弱的人遇到困难选择了退缩；人生总是坎坎坷坷，荆棘密布，难免会受到困难和挫折的困扰。人生的道路中，它便是战胜一切挫折和困难的不竭动力。重要的是在挫折中能坚持到底，永不言弃，直至击败挫折。只有能面对困难和挫折而毫无惧色的人，才能到达成功的顶峰。又或许我们会选择迎难而上，不屈向困难低头而奋起跨越。面对挫折，要乐观、自信。

遇到挫折就垂头丧气，一蹶不振：失败是不气馁，坦然面对，再接再厉敢再向困难挑战是；面对失败与挫折，我们沮丧，不知所措：稍遇困难、挫折就会一蹶不振，准备放弃。敢于直面人生挫折，勇于接受困难的挑战，即使眼看山穷水尽，仍会峰回路转，柳暗花明。并非自满而不知进取的消极态度。高大的姚明摔倒在地，被对手牵起来，他总是以笑相对，当教练指导谈话，他总是微笑聆听，这个优秀的中国小伙子，正是以其谦和、进取的人生态度潜移默化地影响着中国人乃至世界人民。原打算要去北京而后又去了上海的那个乡下人，抱着积极进取，顽强拼搏的生活态度，树立了乐观向上的世界观，所以他认为去上海就是选择了发财致富的好路子。

勾践卧薪尝胆，忍辱负重，逆境中毫不退缩，挫折中未曾止步。

AFTER

Atlanta Falcons Tony Gonzalez Wins the StatSheet NFL Offensive Player of the Week: 11-13-2012

Atlanta Falcons tight end Tony Gonzalez had a spectacular game to become the StatSheet NFL Offensive Player of the Week. Gonzalez received in 11 passes for 122 yards and two touchdowns in the 31-27 loss to the Saints. To this point in the season, he has 617 yards receiving and six touchdowns.

This is the first time this season that Gonzalez has won the award. Tight end Jimmy Graham of the Saints is the runner-up (7 rec, 146 rec yds, 2 TD). Honorable mention goes to tight end Greg Olsen from the Panthers (9 rec, 102 rec yds, 2 TD).

StatSheet NFL Defensive Player of the Week: Darius Butler, Indianapolis Colts

A NFL season high for interceptions in a game was matched by Indianapolis Colts cornerback Darius Butler, clearing the path for his StatSheet Defensive Player of the Week win. Butler recorded four solo tackles and two interceptions during the 27-10 win over the Jaguars. For the year, he has eight tackles and two interceptions.

This is the first time Butler has earned the award this season. The runner-up is linebacker Derrick Johnson of the

不怕困难勇往直前。因为大海的雄奇伟力造就了水手们敢于冒险，敢于拼搏，勇往直前的进取性格，磨炼水手们不怕吃苦，不怕牺牲的坚强意志。的杜甫，他不畏艰险，勇于攀登；因为男儿不畏艰险

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

案情举例: 案例3: 盗窃

案情描述: 被告人施某某于2015年4月20日11时许，在梧仁满族自治县梧仁镇乐购超市地下潮衣库试衣间外衣架上的黑色手拎包盗走。包内有人民币15215元，粉色卡包，银行卡等案发后，被告人施某某于2015年4月23日被公安机关抓获归案赃款、赃物追回返还失主

原始判决: 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、七十二条、第七十三条

机器判案 **类案检索**

关键词: 15215 归案 梧仁镇

罪名: 盗窃

刑期: 九个月

罚金: 14731元

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
【盗窃罪】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

作文评语

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。
在修辞上，段落排比的使用，促进结构条理化，也增强了语言气势和表达效果；并且引用的材料内容简练但有说服力，容易让读者接受，认同作者观点。

94分

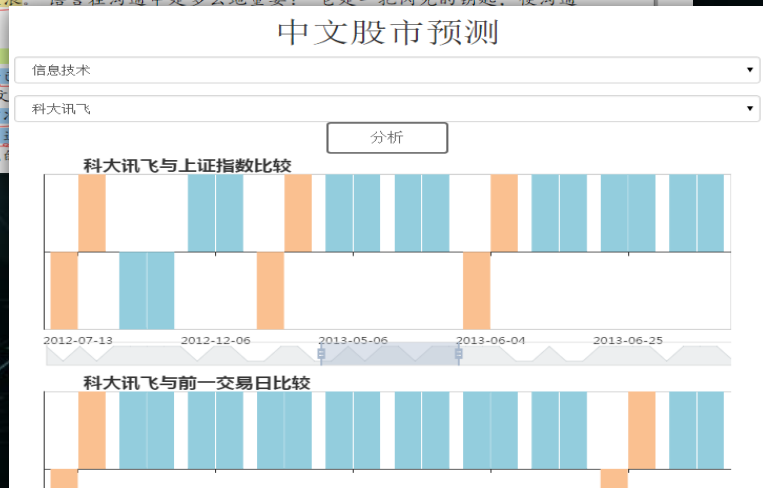
作文原文

语言是沟通的钥匙

假如沟通是一扇门，那么语言就是这扇门的钥匙。

如果沟通的门通向的另一方是漆黑的深夜，那么语言的钥匙便引领着你走向皓月当空，繁星满天；如果沟通的门通向的另一方是一望无际的沙漠，那么语言的钥匙便引领着你走向鸟语花香的绿洲；如果沟通的门通向的另一方是浩瀚无边的大海，那么语言的钥匙便引领着你走向如宗愈般乘风破浪。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心坎上。

恰如其分的语言表达，君”与王勃的“海内存知己，天涯若比邻”。李白《蜀道难》一文有六龙回日之高标，下有冲波逆折之回崖。中感受到友人的关怀，沟通需要语言为它传达彼此



NLP+司法（HFL）

NLP+金融 (哈工大)

谢谢！