



深度学习与智能问答

刘 康 （中科院自动化研究所）

冯岩松 （北京大学）

2016年10月14日

搜索急需一场变革

- IDC统计，全球数据量以每年58%的速度增长，2020年将达到40ZB，是2011年数据总量的22倍



IDC统计2014年数据量是多少

网页 新闻 贴吧 知道 音乐 图片 视频 地图 文库 更多»

百度为您找到相关结果约359,000个

域名商2014年度报告:中国数据域名总量跌至22万 - IDC评述网
2015年1月26日 - IDC评述网(idcps.com)01月26日报道:据域名统计机构WebHosting.info数据,截至2014年12月29日,中国数据域名总量已降至221,397个,年净减80,640个,与...
www.idcps.com/news/201... - 百度快照 - 70%好评

IDC | 199IT互联网数据中心 | 中文互联网数据研究资讯中心-199IT
2014年Q4全球PC出货量之差 Gartner与IDC数据相差290万台 2014年第四季度PC出货量呈现稳定趋势,由于统计的方式不同,两家大型统计公司在第四季度PC出货量数字上略有分...
www.199it.com/archives... - 百度快照 - 91%好评

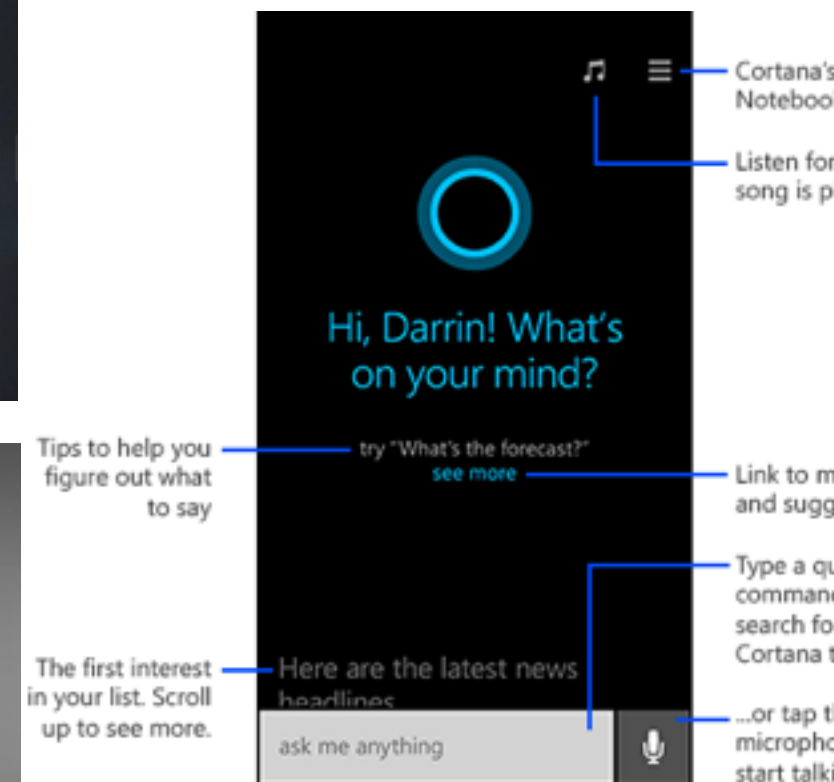
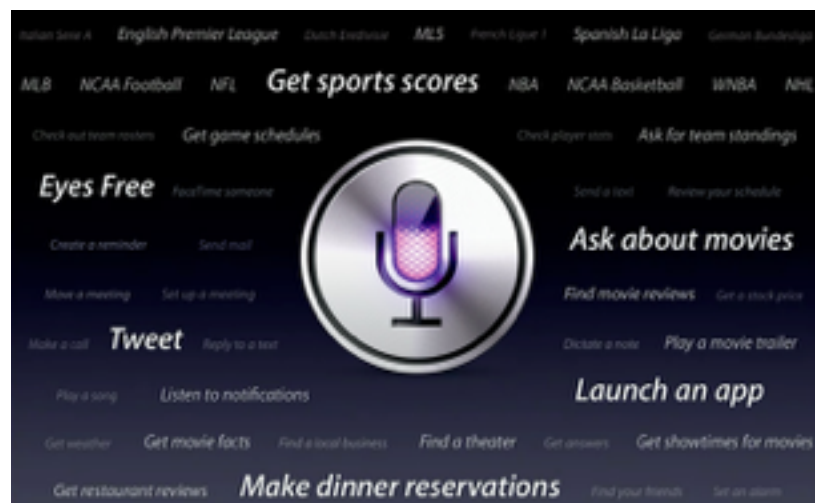
2014年Q4全球PC出货量之差 Gartner与IDC数据相差290万台
2015年1月14日 - 两家大型统计公司在第四季度PC出货量数字上略有分歧:根据Gartner统计2014年第4季度PC出货量达8370万台,同比增长1%;根据IDC统计,2014年第四季度出货量...
www.199it.com/archives... - 百度快照 - 91%好评

域名商年度报告:2014年DNSPod域名总量达174万 - IDC评述网
2015年1月15日 - IDC评述网(idcps.com)01月15日报道:据域名统计机构WebHosting.info数据,截至2014年12月29日,DNSPod域名总量已增至174万,年净增长220,393个域名,而...
www.idcps.com/news/201... - 百度快照 - 70%好评

2014年全球软件开发者数量统计_中国行业研究网
2014年1月6日 - 数据分析公司IDC发布最新报告称:2014年全球的软件开发者数量达到1850万,其中1100万是专业的软件开发人员,另外750万是开发爱好者,占全体程序员的4成...
www.chinain.com/news/... - 百度快照 - 94%好评

交互方式的转变需要信息服务模式的转变

- 移动互联网以及可穿戴设备的飞速发展需要高效、准确的自然语言形式的信息服务方式



问答系统是下一代搜索引擎的基本形态



Prof. Oren Etzioni

Turing Center
University of Washington

以直接而准确的方式
回答用户自然语言提
问的自动问答系统将
构成下一代搜索引擎
的基本形态

— 《Nature》2011.8

IBM Watson

- 沃森(Watson): 2011年, IBM研发的超级计算机“沃森”在美国知识竞赛节目《危险边缘Jeopardy!》中上演“人机问答大战”, 战胜人类选手Ken和Brad



辅助医疗



金融辅助决策



企业服务

问答系统历史

Template-based QA
Expert System

IR-based QA

KB-based QA
Community QA

1960

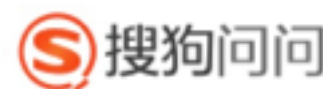
BaseBall
LUNAR
MACSYMA
SHRDLE

1990

MASQUE
TREC



2000

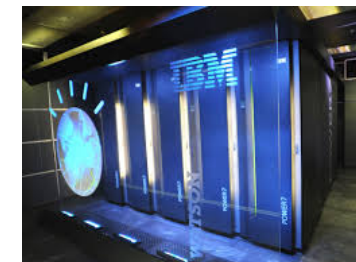


知乎

分享你的知识、经验和见解

搜索问题、话题或人

2010



WolframAlpha
computational...
knowledge engine

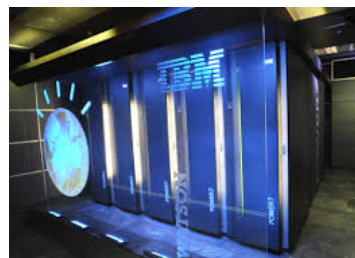


ProBase

问答系统分类

IR-based QA

基于关键词匹配 + 信息抽取，仍然是基于浅层语义分析



Community QA

依赖于网民贡献，问答过程仍然依赖于关键词检索技术



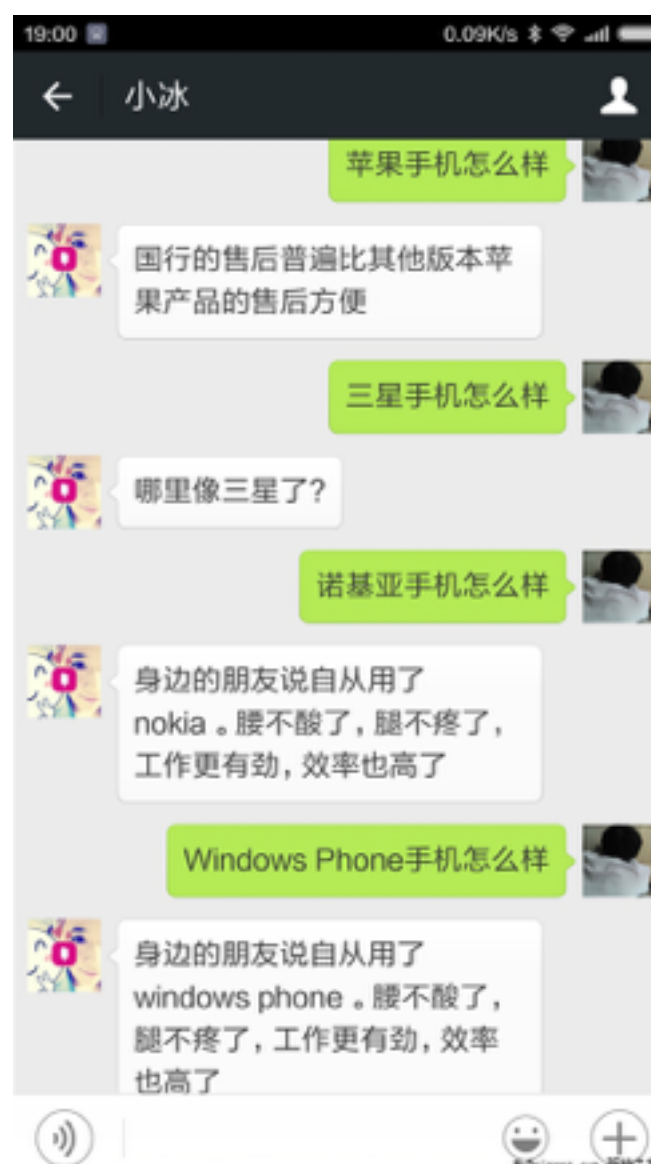
KB-based QA

Knowledge Base



根据问答形式分类

- 一问一答
- 交互式问答
- 阅读理解









Mary journeyed to the den.
Mary went back to the kitchen.
John journeyed to the bedroom.
Mary discarded the milk.
Q: Where was the milk before the den?
A. Hallway

Brian is a lion.
Julius is a lion.
Julius is white.
Bernhard is green.
Q: What color is Brian?
A. White

Sam walks into the kitchen.
Sam picks up an apple.
Sam walks into the bedroom.
Sam drops the apple.
Q: Where is the apple?
A. Bedroom

KB-based QA应用

+NedShare


Web Images Maps Videos More ▾ Search tools

About 155,000 results (0.48 seconds)

Cookies help us deliver our services. By using our services, you agree to our use of cookies.

[OK](#) [Learn more](#)

Ned Poulter

[www.nedpoulter.com/ ▾](http://www.nedpoulter.com/)

by Ned Poulter - in 1,122 Google+ circles
Geek, Blogger & Digital Marketer currently working as the SEO Manager for Miinto. Check out my blogs and contact me here.

Ned Poulter (NedPoulter) on Twitter

[https://twitter.com/NedPoulter ▾](https://twitter.com/NedPoulter)

The latest from Ned Poulter (@NedPoulter). Geek, Film Buff, Blogger, Social Butterfly, Aspiring Indiana Jones. Lover of SEO & Online Marketing. Copenhagen ...

Ned Poulter - Danmark | LinkedIn


[dk.linkedin.com/in/nedpoulter ▾](http://dk.linkedin.com/in/nedpoulter)

Copenhagen Area, Denmark - Head of Digital at Miinto Group
Vis Ned Poulter's (Danmark) faglige profil på LinkedIn. LinkedIn er verdens største erhvervsnetværk, der hjælper fagfolk som Ned Poulter med at finde interne ...

Ned Poulter - author on State of Digital

[www.stateofdigital.com/author/ned-poulter/ ▾](http://www.stateofdigital.com/author/ned-poulter/)

by Ned Poulter - in 1,122 Google+ circles
Ned Poulter is an author on State of Digital and SEO Manager at Miinto. Find his articles and bio here.




Ned Poulter


1,126 followers on Google+
Copenhagen based SEO, Blogging & Digital Marketing enthusiast who thrives on everything geeky. I love traveling and never stop learning....


Contact info


From Ned Poulter's profile

Email 


From Google Contacts - Visible only to you

Email 

Phone 

Address 

Recent posts



Fantastic Visualisation of The Marketing Technology Landscape in 2014 [Infographic]. Well Worth Saving + Sharing 9 Jan 2014

Personal email

Alternative emails

Mobile phone number

Home address

KB-based QA应用

 who is the winner of nba finals in 2016  

[全部](#) [新闻](#) [图片](#) [视频](#) [地图](#) [更多 ▾](#) [搜索工具](#)

找到约 21,400,000 条结果 (用时 0.62 秒)

NBA总决赛 (2016 年)

克利夫兰骑士



 球员名单和概述


反馈

[2016 NBA Playoffs | Schedules, Brackets, Highlights and Video ...](#)



www.nba.com/playoffs/ ▾ [翻译此页](#)





Find playoff schedules, brackets of your favorite teams and video highlights of all the **2016 NBA Playoffs** games on NBA.com.



KB-based QA 应用

 **WolframAlpha** computational knowledge engine

Enter what you want to calculate or know about:

how big is China  

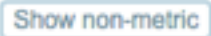
 Examples  Random

Assuming "how big" is international data | Use as [referring to socioeconomic data](#) or [referring to species](#) or [referring to administrative divisions](#) instead

Assuming total area | Use [population](#) instead

Input interpretation:

China total area

Result: 

$9.597 \times 10^6 \text{ km}^2$ (square kilometers) (world rank: 4th)

Unit conversions:

$9.597 \times 10^{12} \text{ m}^2$ (square meters)

3.705 million mi^2 (square miles)

$1.033 \times 10^{14} \text{ ft}^2$ (square feet)

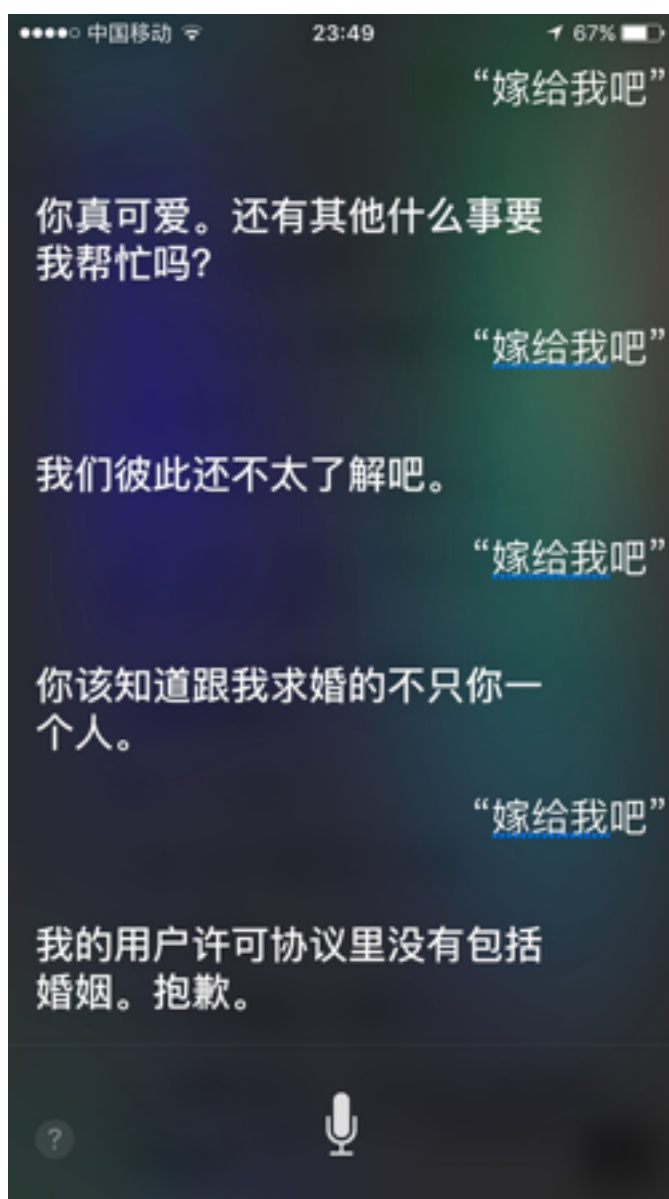
Comparisons as area:

$\approx 0.96 \times$ total area of Canada ($9.98467 \times 10^6 \text{ km}^2$)

$\approx 0.996 \times$ total area of the United States ($9.63142 \times 10^6 \text{ km}^2$)

\approx largest extent of the Roman Empire ($\approx 9 \text{ Mm}^2$)

个人助理



答题机器人

- “多达一”机器人：日本富士通联合日本国立信息学研究所的高考机器人项目（2011）
- 微软创始人Paul Allen与华盛顿大学计划制造能够通过高中生物考试的系统
- 863：基于大数据的类人智能关键技术与系统



问答



- 传统问答方法 (Symbolic Representation)

- 基于关键词检索
- 基于文本蕴含推理
- 基于逻辑表达式

- 基于深度学习的问答方法 (Distributed Representation)

- LSTM
- Attention Model
- Memory Network

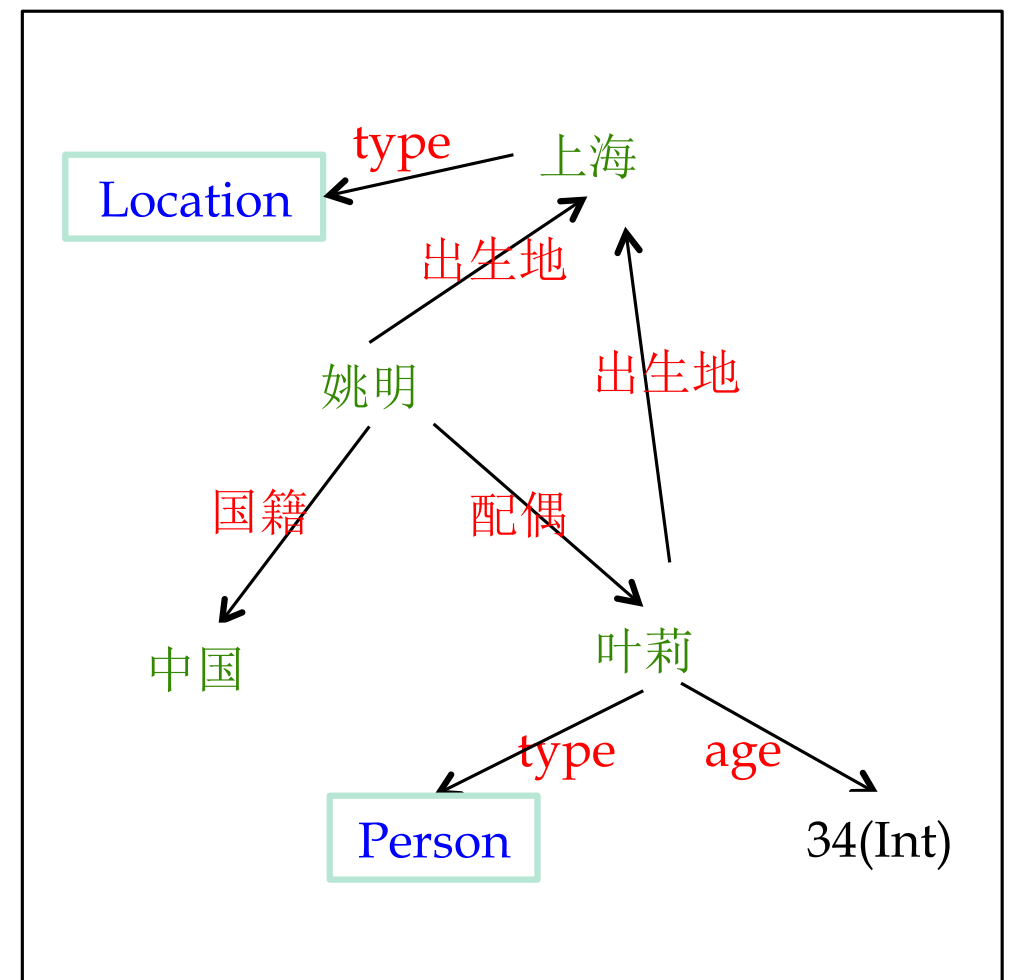
基于符号表示(传统)的知识库问答

姚明的老婆的国籍是?

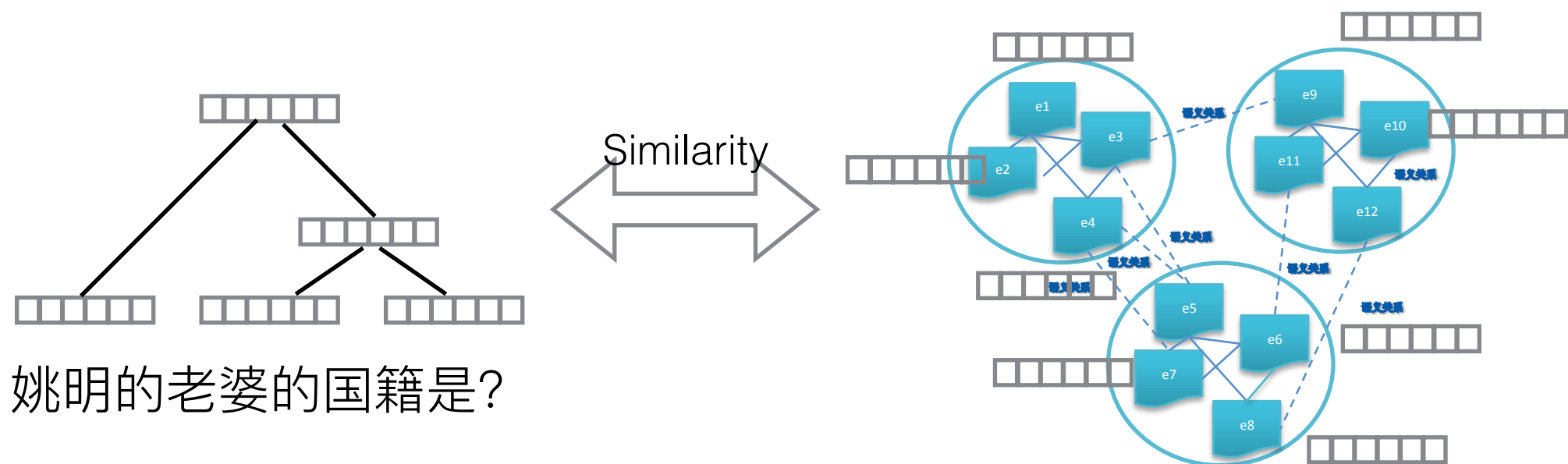
语义解析

查询

```
SELECT DISTINCT ?x
WHERE {
  ?y 国籍 ?x.
  res:姚明 配偶 ?y.
}
```



基于分布式表示(DL)的知识库问答



This Talk

- Part 1, 15:00-15:50: 传统基于符号表示的知识库问答（冯岩松）
- 15:50-16:00: 休息
- Part 2, 16:00-16:30: 基于深度学习的知识库问答（刘康）
- Part 3, 16:30-17:00: 基于深度学习的对话系统（冯岩松）
- Part 4, 17:00-17:30: 基于深度学习的篇章阅读理解（刘康）