大数据时代的自然语言处理：前沿与进展

孙茂松
清华大学计算机科学与技术系

第十四届中国少数民族语言文字信息处理学术研讨会
2013年9月14日，兰州

目录

1. 引言

2. 三个重要前沿及其进展

（1）基于深度学习的句子结构预测

（2）开放域信息抽取

（3）知识图谱

3. 题外话

目录

1. 引言

2. 三个重要前沿及其进展

（1）基于深度学习的句子结构预测

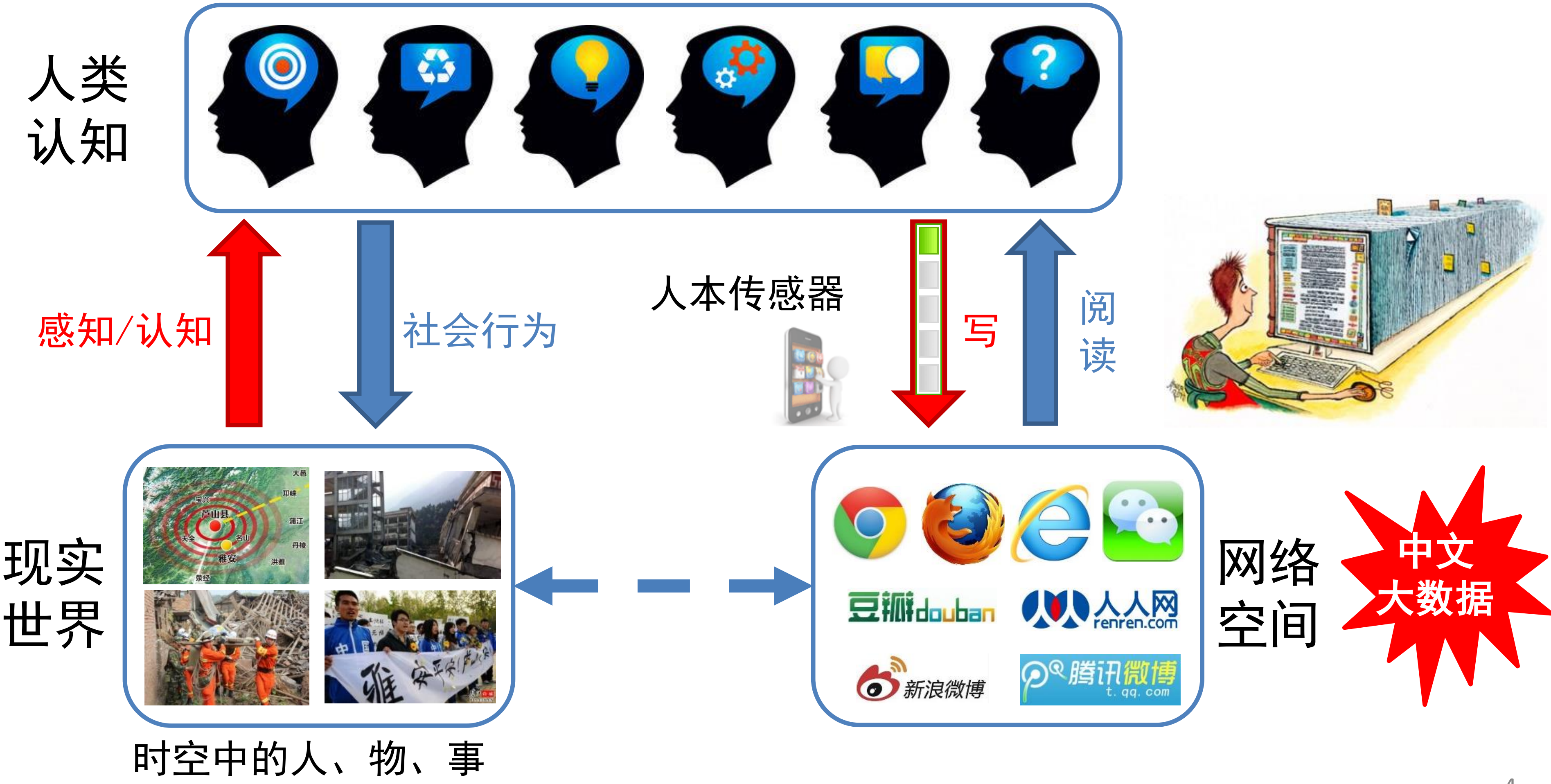
（2）开放域信息抽取

（3）知识图谱

3. 题外话

现实世界、网络空间与人类认知

- 三位一体：有史以来深度和广度最蔚为壮观的虚实结合的空间
  - “人本传感器”：中国网民规模达5.64亿，微博用户3.09亿（2013年1月）
  - “人本传感器”信号：仅新浪微博每日发布超过1亿条微博（2012年12月）



人类认知

感知/认知

社会行为

人本传感器

写

阅读

现实世界

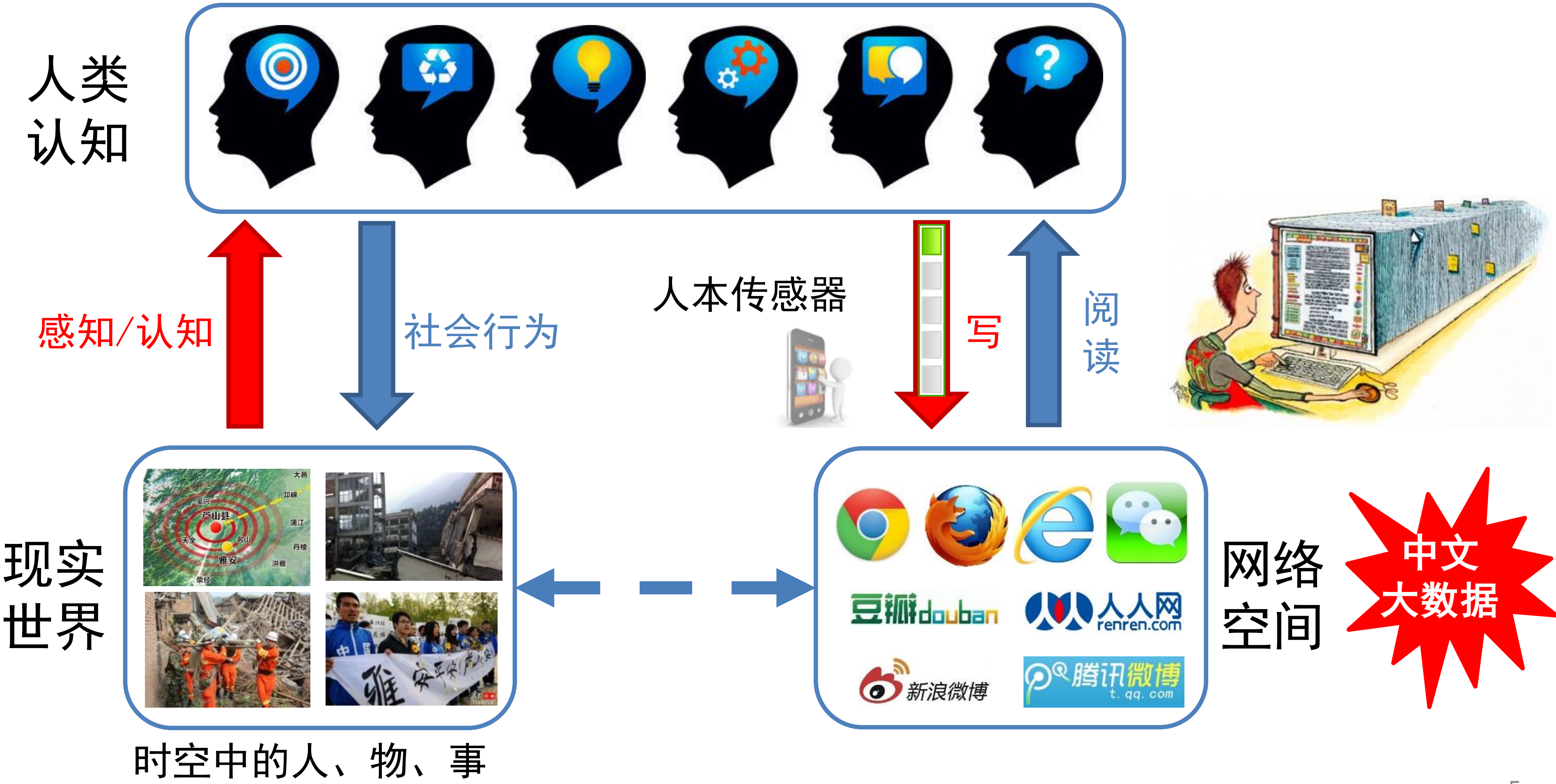
网络空间

中文大数据

时空中的人、物、事

机器阅读理解互联网

- 人类“管中窥豹式”阅读难以形成对虚实空间完整准确的认识
- 机器阅读理解网络空间的中文信息是实现网络洞察力的关键



人类认知

感知/认知

社会行为

人本传感器

写

阅读

现实世界

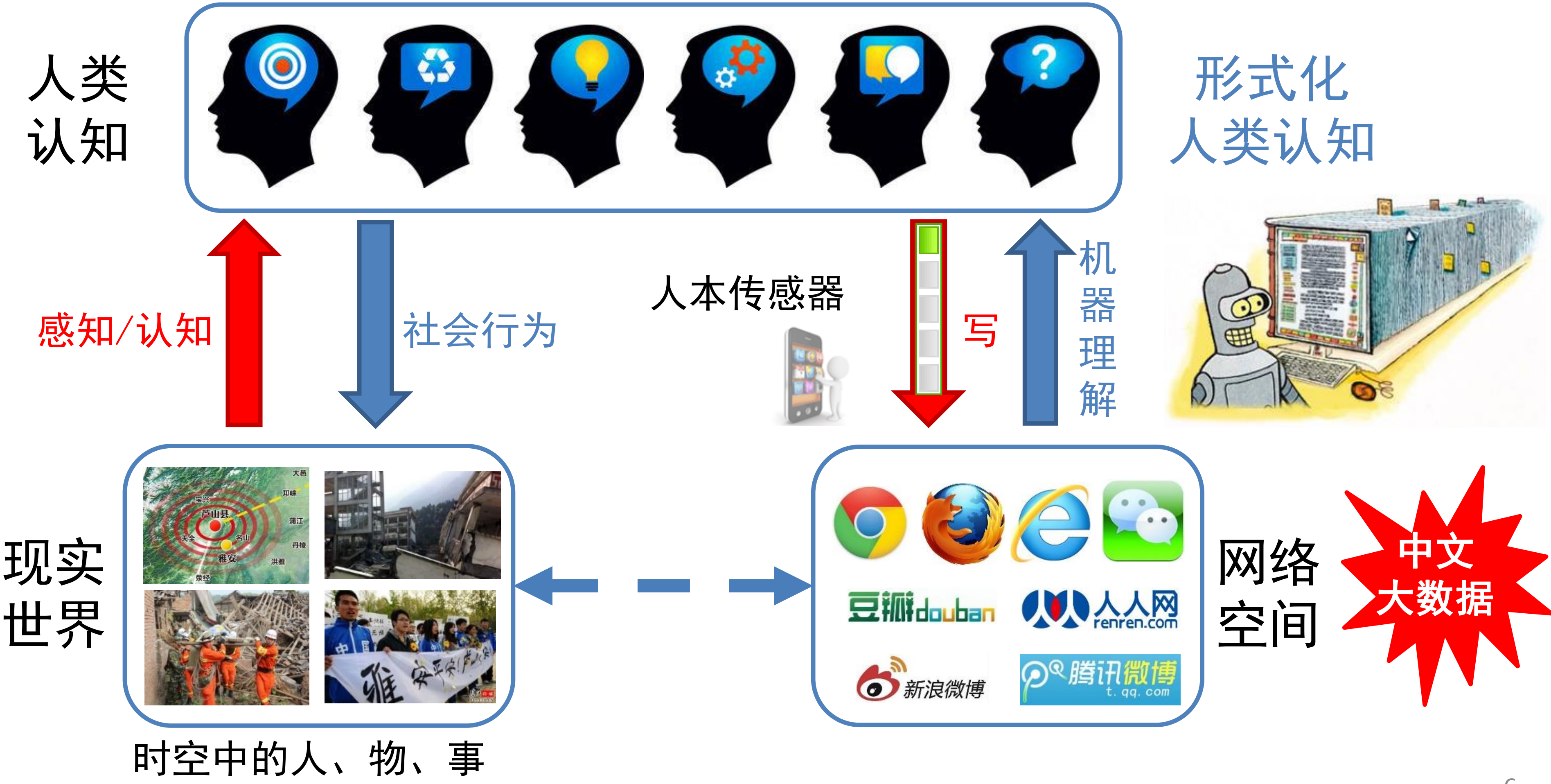
时空中的人、物、事

网络空间

中文大数据

机器阅读理解互联网

- 人类“管中窥豹式”阅读难以形成对虚实空间完整准确的认识
- 机器阅读理解网络空间的中文信息是实现网络洞察力的关键



人类认知

形式化人类认知

感知/认知

社会行为

人本传感器

写

机器理解

现实世界

网络空间

中文大数据

时空中的人、物、事

目录

1. 引言

2. 三个重要前沿及其进展

（1）基于深度学习的句子结构预测

（2）开放域信息抽取

（3）知识图谱

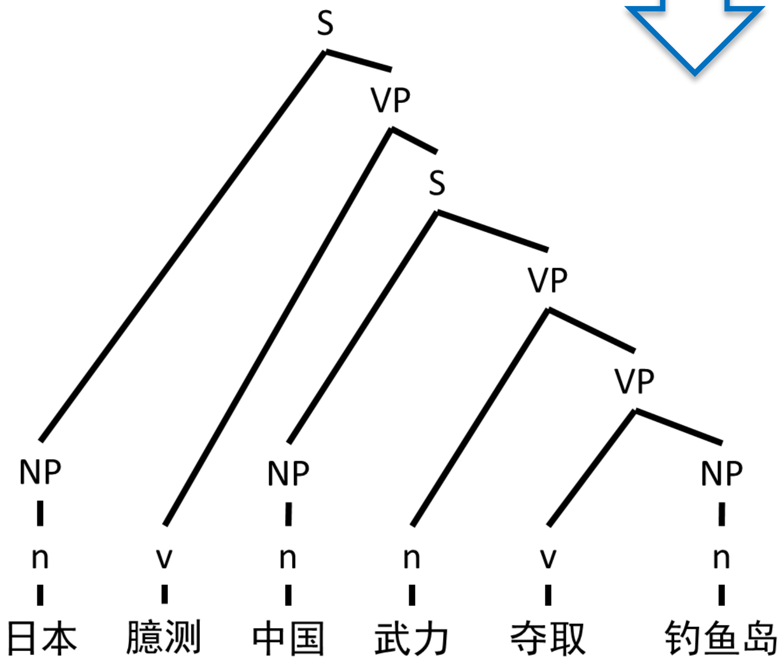
3. 清华最近NLP相关工作

自然语言处理的根本任务

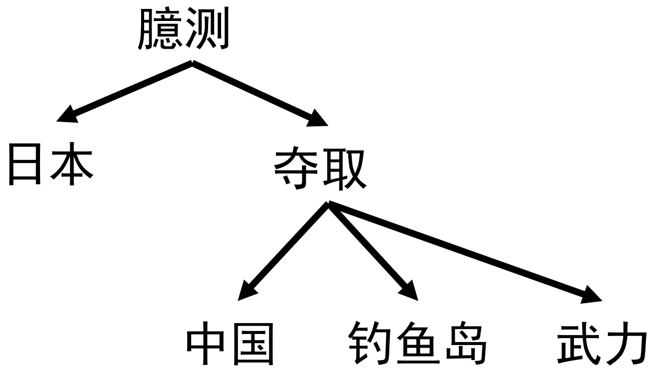
输入： 日本臆测中国武力夺取钓鱼岛

结构预测

输出：



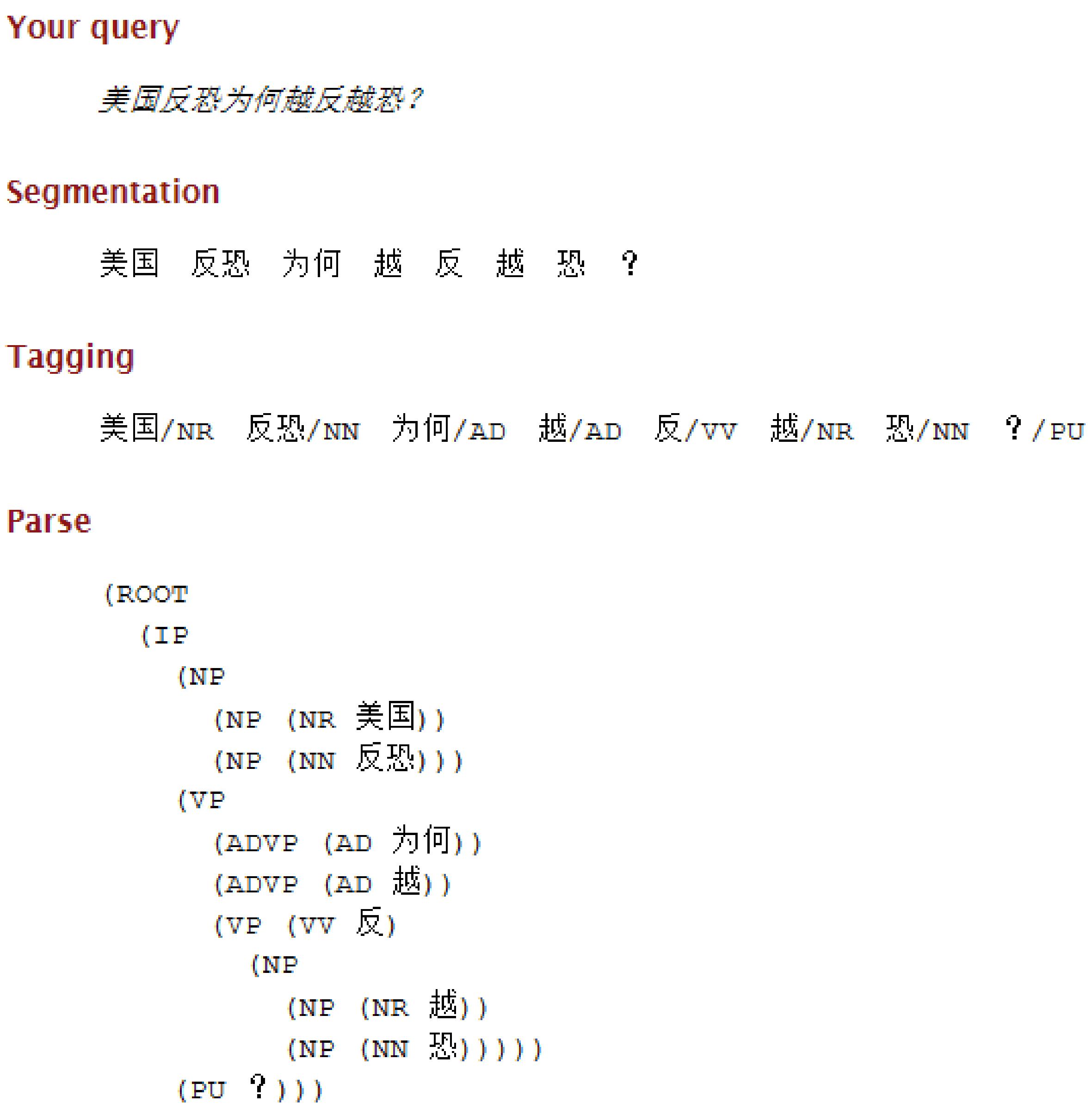
句法结构



语义结构

语言计算的本质是结构预测

从例句说起



Your query

美国反恐为何越反越恐？

Segmentation

美国 反恐 为何 越 反 越 恐 ？

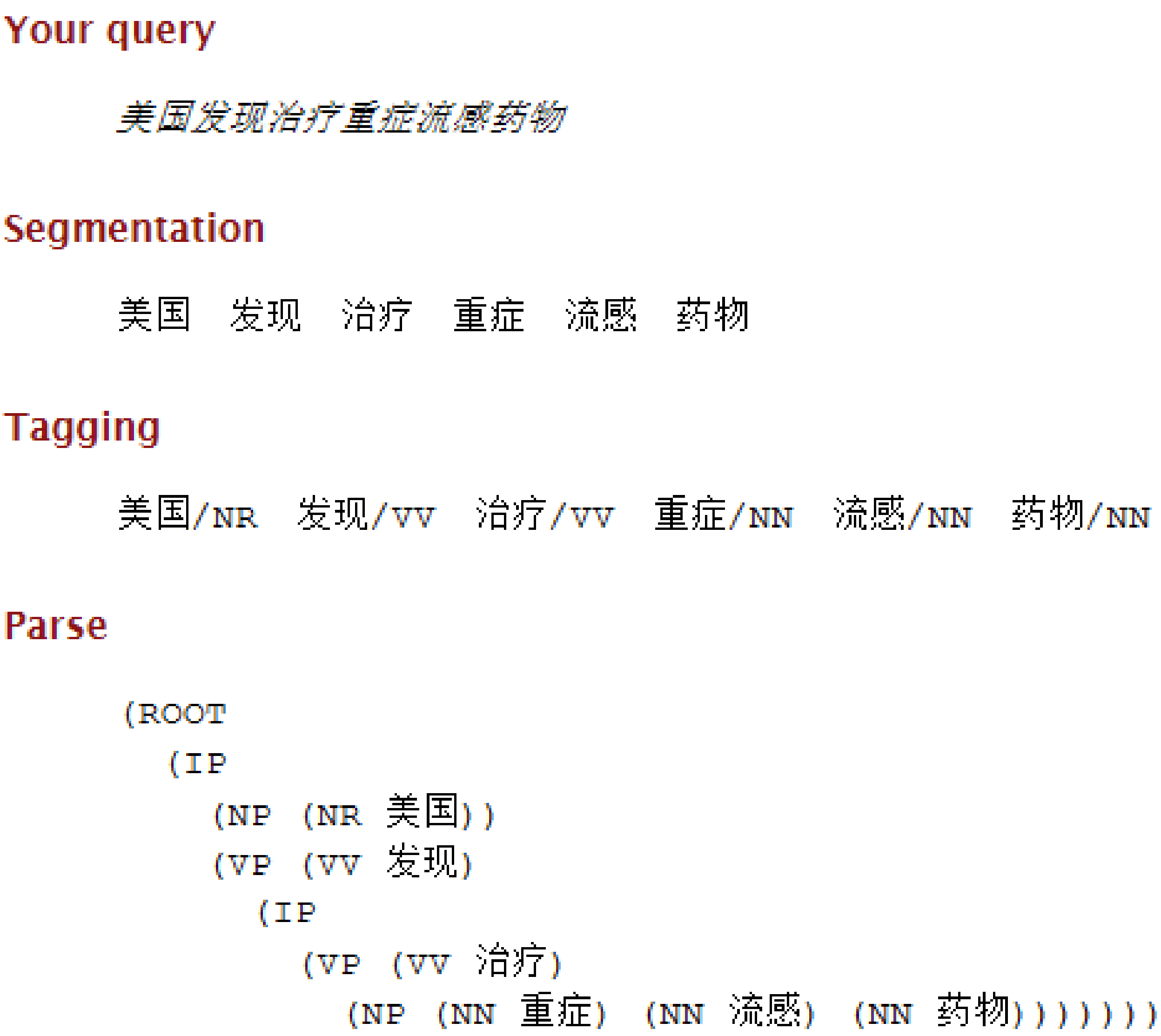
Tagging

美国/NR 反恐/NN 为何/AD 越/AD 反/VV 越/NR 恐/NN ？/PU

Parse

```
(ROOT
  (IP
    (NP
      (NP (NR 美国))
      (NP (NN 反恐)))
    (VP
      (ADVP (AD 为何))
      (ADVP (AD 越))
      (VP (VV 反)
        (NP
          (NP (NR 越))
          (NP (NN 恐)))))
    (PU ？)))
```

从例句说起



Your query

美国发现治疗重症流感药物

Segmentation

美国　发现　治疗　重症　流感　药物

Tagging

美国/NR　发现/VV　治疗/VV　重症/NN　流感/NN　药物/NN

Parse

```
(ROOT
  (IP
    (NP (NR 美国))
    (VP (VV 发现)
      (IP
        (VP (VV 治疗)
          (NP (NN 重症) (NN 流感) (NN 药物)))))))
```

汉语是世界上最难被计算机理解的语言之一

- 汉语具有显著的特点

特点	例子
复杂名词短语	中国北京红十字芦山抢险救援队“五一”节期间工作掠影
形式标记和形态变化	机器翻译，翻译人员，翻译小说
流水句 （成分省略）	她弯下腰来飞快地割着麦子，一把一把沉甸甸的，今年收成真是不错，心情不由得欢快起来。

需要性能高、覆盖能力强的汉语句子结构预测模型

- 目前语言计算主流模型可分为两类，但均存在很大局限性
- 互联网中文理解亟需建立能处理大规模开放域文本深层结构的语言计算模型

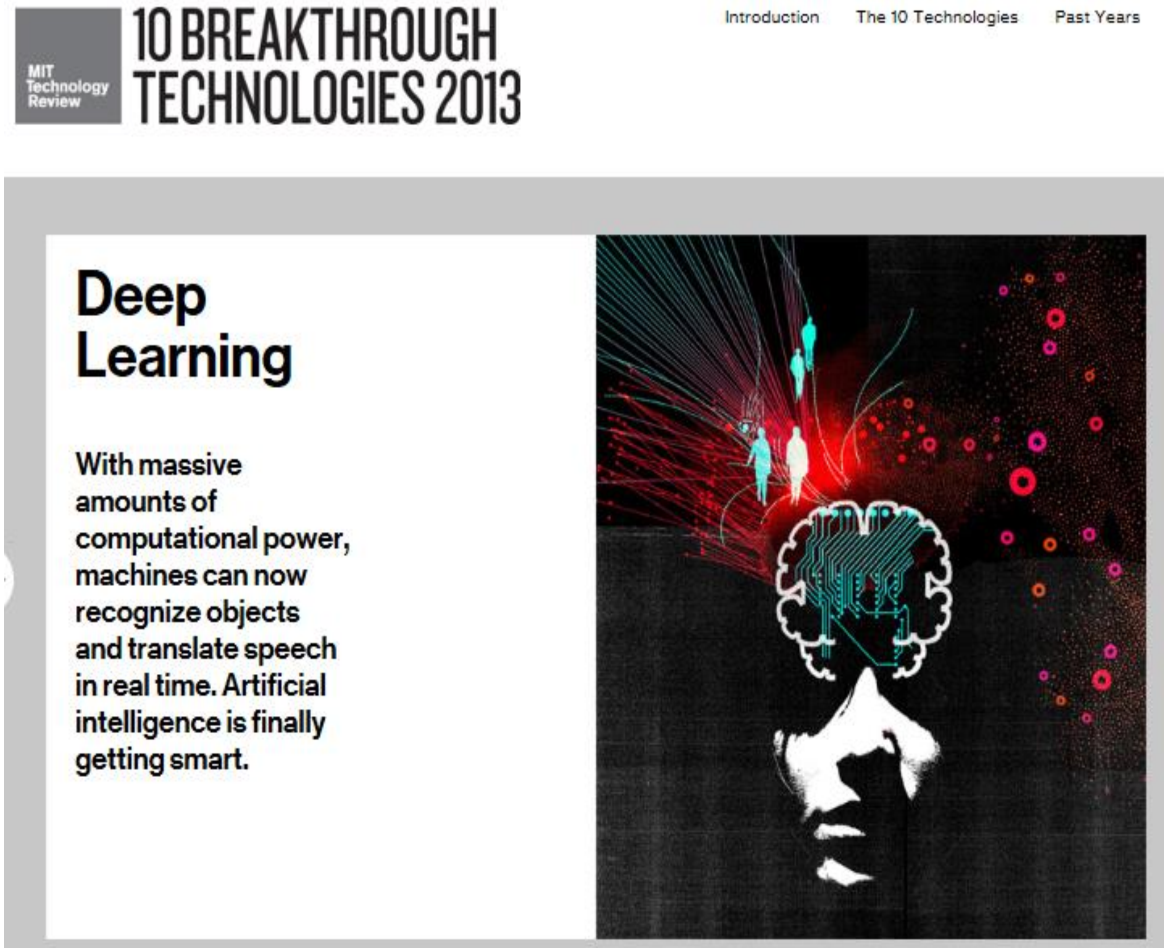
语言计算模型	语言结构	模型训练所需语料库	可用训练数据规模	对互联网的覆盖能力
马尔科夫模型	表层	无标注	极大	强
条件随机场模型	深层	有标注	较小	弱
?	深层	无标注& 有标注	极大& 较小	强

可能的策略：深度学习

- 深度学习：通过学习出模型的“深层结构”对数据中存在的复杂关系进行建模（本质上是一种数学模型）

Geoffrey Hinton
深度信念网络DBN（2006）
英国皇家学会院士

Judea Pearl
概率图模型（ 2011年获图灵奖）
美国工程院院士



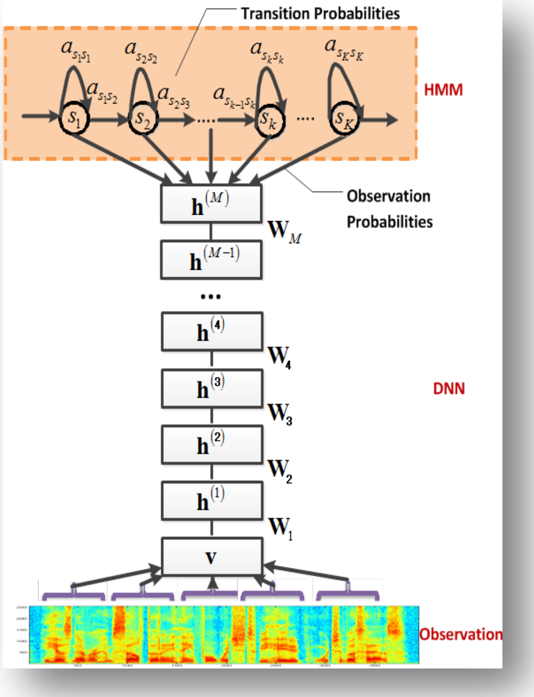
10 BREAKTHROUGH TECHNOLOGIES 2013

Deep Learning

With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart.

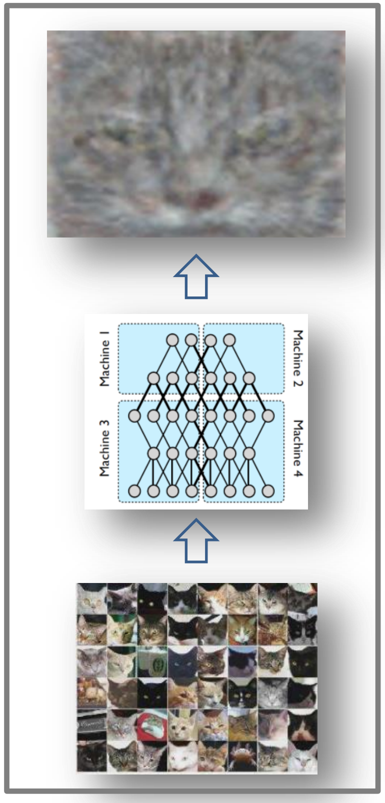
深度学习的显著进展

- 优良的计算性质：可望突破“表层结构”的限制，适合小规模有标注样本和极大规模无标注样本的融合学习
- 深度学习在英文语音识别和图像识别中取得突破



微软语音识别

错误率减少30%以上



16000多个处理器、10亿个内部连接组成的“虚拟大脑”，从1000万帧YouTube的无标签图片中自主“学会”了猫的概念。

谷歌虚拟大脑
（Google Brain）

一个关于深度学习的基本事实

- 强烈反差：

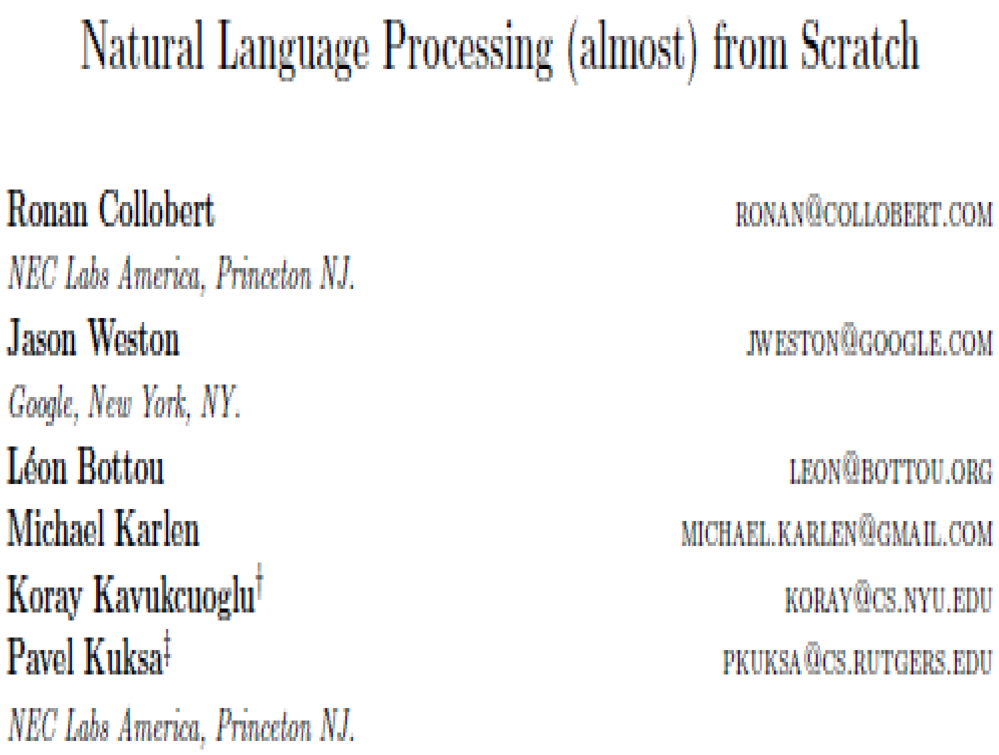
  针对语言理解的深度学习尚未取得成功

  - 语音图像：基于视觉或音频的“底层认知特征”
  - 语言理解：基于词法、句法和语义等“高层认知特征”

- 深度学习在中文计算方面尚未见公开报道成果

- 语言深度学习存在重要的理论创新空间
  - 高层认知特征的表示及其学习
  - 适合于语言计算的大规模人工神经网络模型

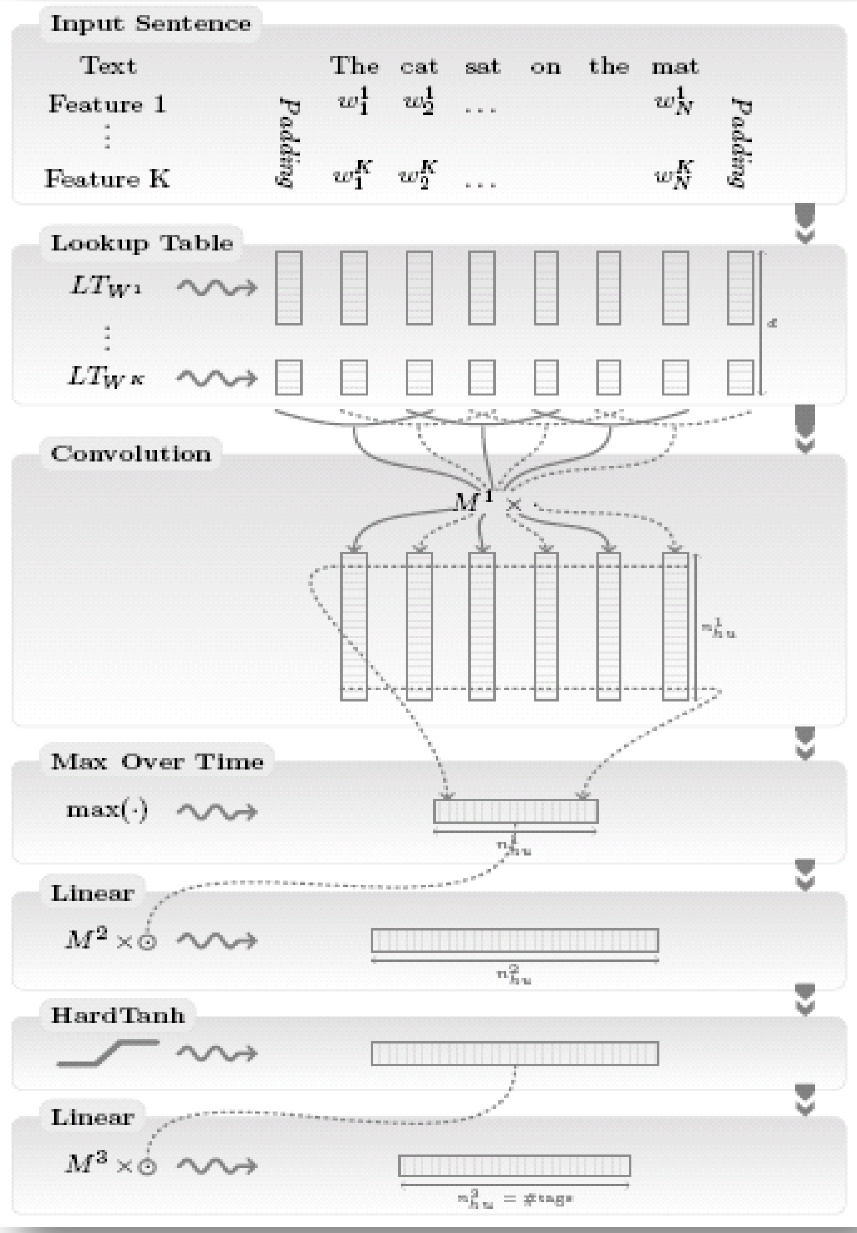
代表性工作之一：从几乎一无所有开始的自然语言处理



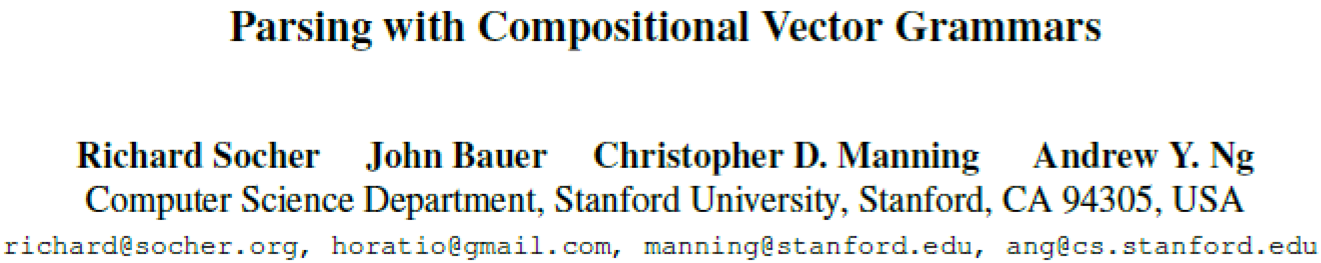
Natural Language Processing (almost) from Scratch

Ronan Collobert RONAN@COLLOBERT.COM
NEC Labs America, Princeton NJ.
Jason Weston JWESTON@GOOGLE.COM
Google, New York, NY.
Léon Bottou LEON@BOTTOU.ORG
Michael Karlen MICHAEL.KARLEN@GMAIL.COM
Koray Kavukcuoglu[†] KORAY@CS.NYU.EDU
Pavel Kuksa[‡] PKUKSA@CS.RUTGERS.EDU
NEC Labs America, Princeton NJ.

在多项自然语言处理任务中
与经典主流方法结果具有可比性



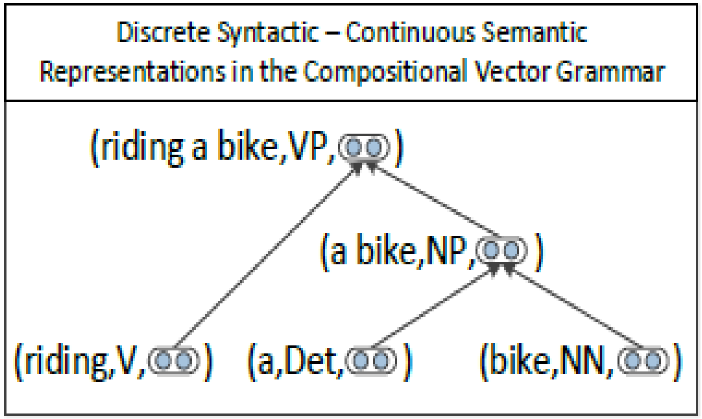
代表性工作之二：基于组合向量语法的分析



Parsing with Compositional Vector Grammars

Richard Socher John Bauer Christopher D. Manning Andrew Y. Ng
Computer Science Department, Stanford University, Stanford, CA 94305, USA
richard@socher.org, horatio@gmail.com, manning@stanford.edu, ang@cs.stanford.edu

- Small sets of discrete categories such as NP and VP does not capture the full syntactic nor semantic richness of linguistic phrases Lexicalizing phrases or splitting categories only partly address the problem at the cost of huge feature spaces and sparseness.
- Compositional Vector Grammar (CVG), which combines PCFGs with a recursive neural network that learns syntactico-semantic, compositional vector representations.
- The CVG improves the PCFG of the Stanford Parser by 3.8% to obtain an F1 score of 90.4%. It is fast to train, about 20% faster than the current Stanford factored parser.



Discrete Syntactic – Continuous Semantic Representations in the Compositional Vector Grammar

Figure 1: Example of a CVG tree with (category,vector) representations at each node. The vectors for nonterminals are computed via a new type of recursive neural network which is conditioned on syntactic categories from a PCFG.

目录

1. 引言

2. 三个重要前沿及其进展

（1）基于深度学习的句子结构预测

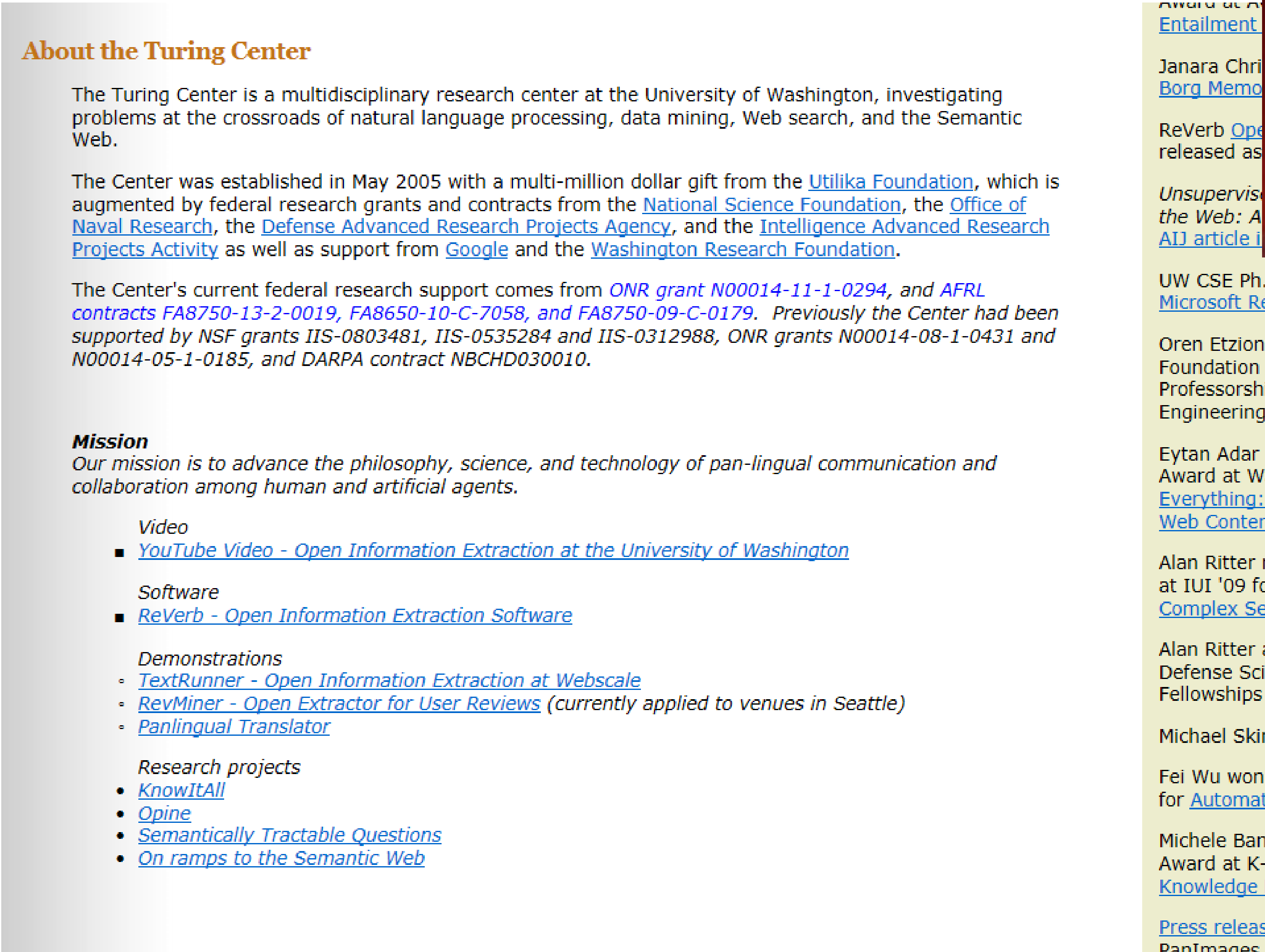
（2）开放域信息抽取

（3）知识图谱

3. 题外话

华盛顿大学图灵中心：ReVerb

http://turing.cs.washington.edu/index.htm



About the Turing Center

The Turing Center is a multidisciplinary research center at the University of Washington, investigating problems at the crossroads of natural language processing, data mining, Web search, and the Semantic Web.

The Center was established in May 2005 with a multi-million dollar gift from the Utilika Foundation, which is augmented by federal research grants and contracts from the National Science Foundation, the Office of Naval Research, the Defense Advanced Research Projects Agency, and the Intelligence Advanced Research Projects Activity as well as support from Google and the Washington Research Foundation.

The Center's current federal research support comes from *ONR grant N00014-11-1-0294, and AFRL contracts FA8750-13-2-0019, FA8650-10-C-7058, and FA8750-09-C-0179. Previously the Center had been supported by NSF grants IIS-0803481, IIS-0535284 and IIS-0312988, ONR grants N00014-08-1-0431 and N00014-05-1-0185, and DARPA contract NBCHD030010.*

Mission
Our mission is to advance the philosophy, science, and technology of pan-lingual communication and collaboration among human and artificial agents.

Video
- *YouTube Video - Open Information Extraction at the University of Washington*

Software
- *ReVerb - Open Information Extraction Software*

Demonstrations
- *TextRunner - Open Information Extraction at Webscale*
- *RevMiner - Open Extractor for User Reviews (currently applied to venues in Seattle)*
- *Panlingual Translator*

Research projects
- *KnowItAll*
- *Opine*
- *Semantically Tractable Questions*
- *On ramps to the Semantic Web*

Award at A
Entailment

Janara Chri
Borg Memo

ReVerb Ope
released as

Unsupervis
the Web: A
AIJ article i

UW CSE Ph.D. alum Doug Downey is 2010 Microsoft Research Faculty Fellow

Oren Etzioni awarded a Washington Research Foundation Endowed Entrepreneurship Professorship in Computer Science & Engineering, summer 2009

Eytan Adar received Best Student Paper Award at WSDM '09 for The Web Changes Everything: Understanding the Dynamics of Web Content

Alan Ritter received Best Student Paper Award at IUI '09 for Learning to Generalize for Complex Selection Tasks

Alan Ritter and Tom Lin awarded National Defense Science and Engineering Graduate Fellowships

Michael Skinner wins TopCoder Competition

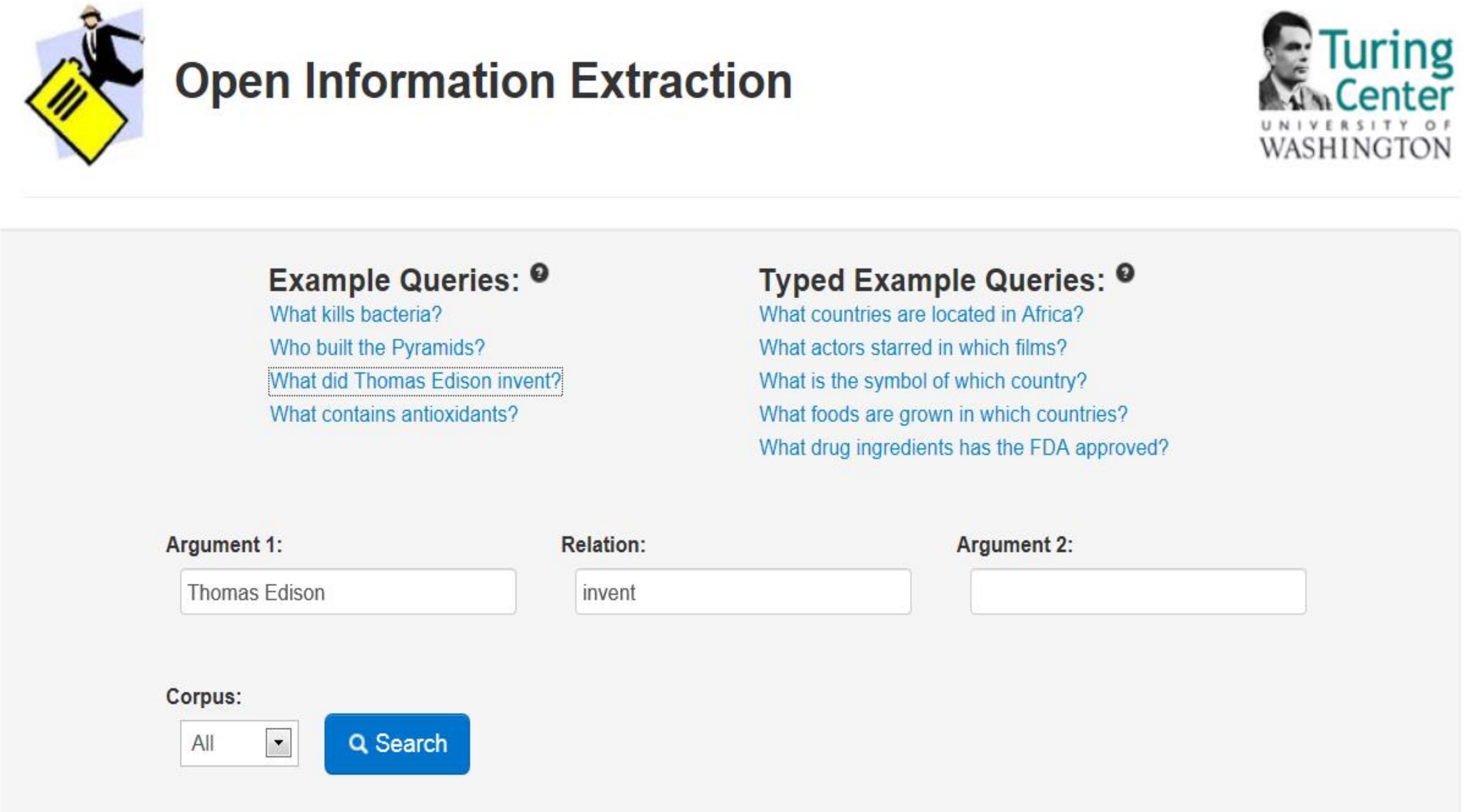
Fei Wu won the best paper prize at CIKM '07 for Automatically Semantifying Wikipedia

Michele Banko received Best Student Paper Award at K-CAP '07 for Strategies for Lifelong Knowledge Extraction from the Web

Press release announcing the launch of PanImages

华盛顿大学图灵中心：ReVerb

- ReVerb, an open-source extractor, which extracted over 1,000,000,000 assertions from the Web.



Open Information Extraction

Example Queries:

- What kills bacteria?
- Who built the Pyramids?
- What did Thomas Edison invent?
- What contains antioxidants?

Typed Example Queries:

- What countries are located in Africa?
- What actors starred in which films?
- What is the symbol of which country?
- What foods are grown in which countries?
- What drug ingredients has the FDA approved?

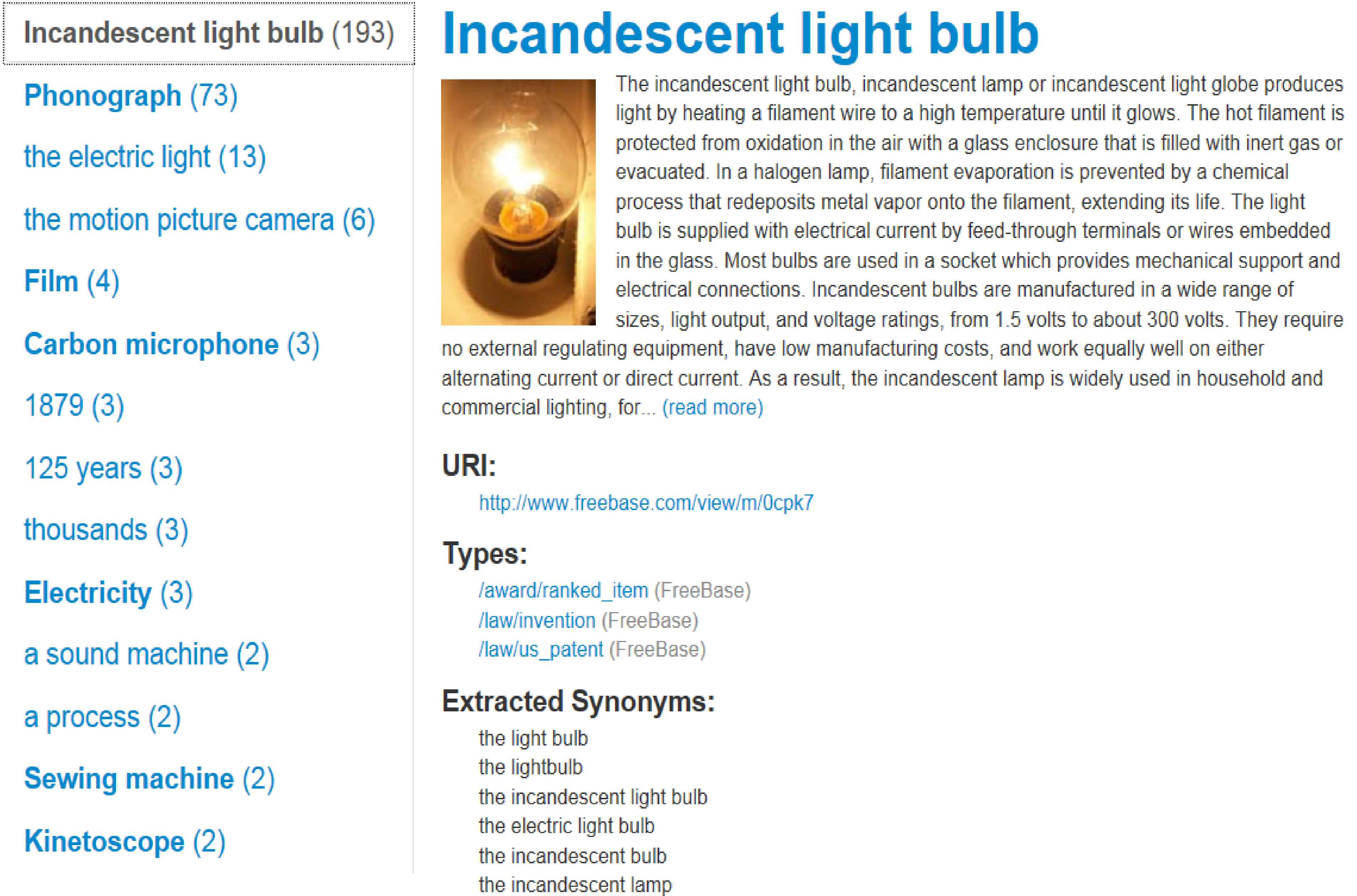
Argument 1: Thomas Edison

Relation: invent

Argument 2:

Corpus: All

华盛顿大学图灵中心：ReVerb



Incandescent light bulb (193)

Phonograph (73)

the electric light (13)

the motion picture camera (6)

Film (4)

Carbon microphone (3)

1879 (3)

125 years (3)

thousands (3)

Electricity (3)

a sound machine (2)

a process (2)

Sewing machine (2)

Kinetoscope (2)

Incandescent light bulb

The incandescent light bulb, incandescent lamp or incandescent light globe produces light by heating a filament wire to a high temperature until it glows. The hot filament is protected from oxidation in the air with a glass enclosure that is filled with inert gas or evacuated. In a halogen lamp, filament evaporation is prevented by a chemical process that redeposits metal vapor onto the filament, extending its life. The light bulb is supplied with electrical current by feed-through terminals or wires embedded in the glass. Most bulbs are used in a socket which provides mechanical support and electrical connections. Incandescent bulbs are manufactured in a wide range of sizes, light output, and voltage ratings, from 1.5 volts to about 300 volts. They require no external regulating equipment, have low manufacturing costs, and work equally well on either alternating current or direct current. As a result, the incandescent lamp is widely used in household and commercial lighting, for... (read more)

URI:

http://www.freebase.com/view/m/0cpk7

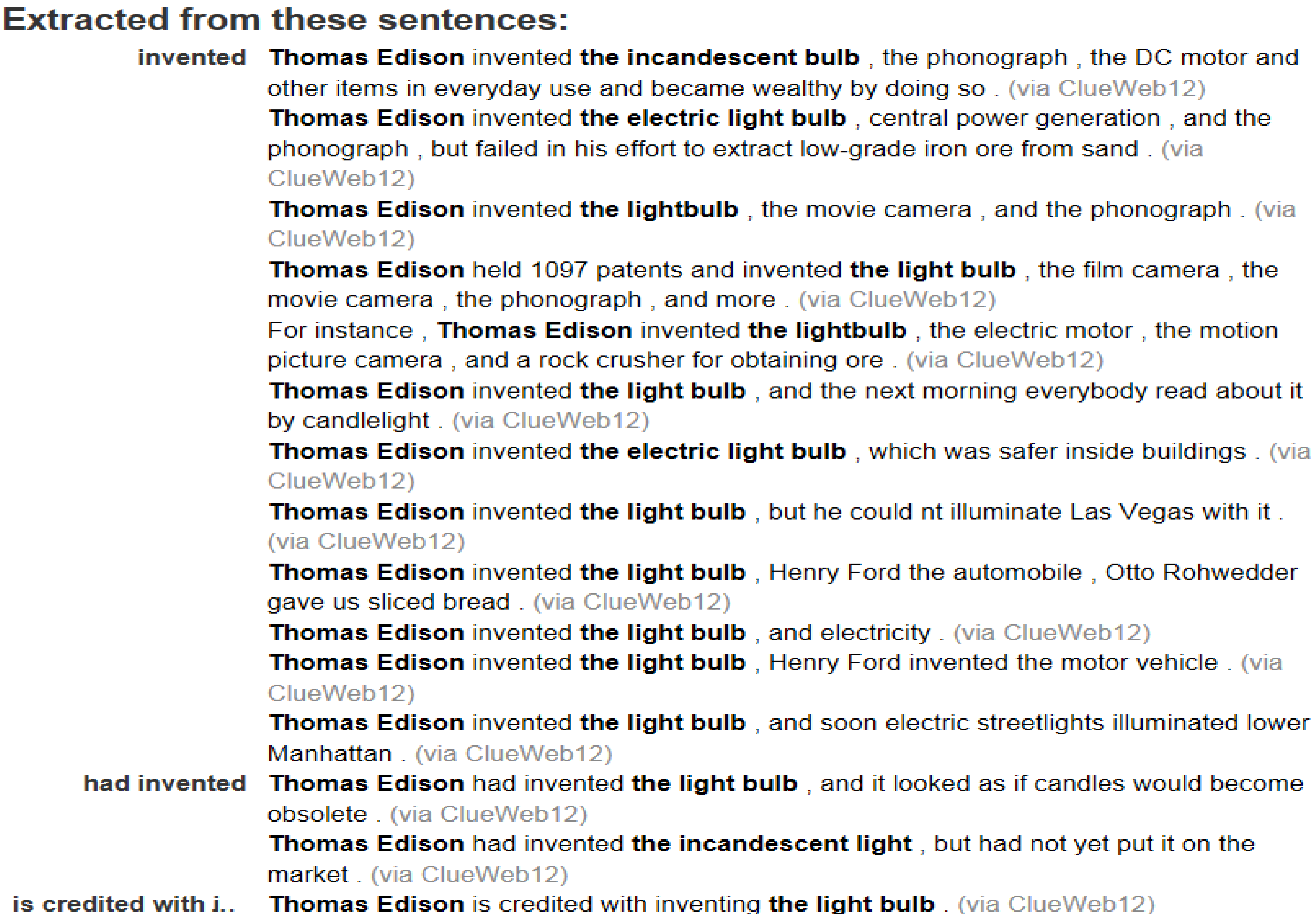
Types:

/award/ranked_item (FreeBase)
/law/invention (FreeBase)
/law/us_patent (FreeBase)

Extracted Synonyms:

the light bulb
the lightbulb
the incandescent light bulb
the electric light bulb
the incandescent bulb
the incandescent lamp

华盛顿大学图灵中心：ReVerb



Extracted from these sentences:

invented	**Thomas Edison** invented **the incandescent bulb** , the phonograph , the DC motor and other items in everyday use and became wealthy by doing so . (via ClueWeb12)
	Thomas Edison invented **the electric light bulb** , central power generation , and the phonograph , but failed in his effort to extract low-grade iron ore from sand . (via ClueWeb12)
	Thomas Edison invented **the lightbulb** , the movie camera , and the phonograph . (via ClueWeb12)
	Thomas Edison held 1097 patents and invented **the light bulb** , the film camera , the movie camera , the phonograph , and more . (via ClueWeb12)
	For instance , **Thomas Edison** invented **the lightbulb** , the electric motor , the motion picture camera , and a rock crusher for obtaining ore . (via ClueWeb12)
	Thomas Edison invented **the light bulb** , and the next morning everybody read about it by candlelight . (via ClueWeb12)
	Thomas Edison invented **the electric light bulb** , which was safer inside buildings . (via ClueWeb12)
	Thomas Edison invented **the light bulb** , but he could nt illuminate Las Vegas with it . (via ClueWeb12)
	Thomas Edison invented **the light bulb** , Henry Ford the automobile , Otto Rohwedder gave us sliced bread . (via ClueWeb12)
	Thomas Edison invented **the light bulb** , and electricity . (via ClueWeb12)
	Thomas Edison invented **the light bulb** , Henry Ford invented the motor vehicle . (via ClueWeb12)
	Thomas Edison invented **the light bulb** , and soon electric streetlights illuminated lower Manhattan . (via ClueWeb12)
had invented	**Thomas Edison** had invented **the light bulb** , and it looked as if candles would become obsolete . (via ClueWeb12)
	Thomas Edison had invented **the incandescent light** , but had not yet put it on the market . (via ClueWeb12)
is credited with i..	**Thomas Edison** is credited with inventing **the light bulb** . (via ClueWeb12)

华盛顿大学图灵中心：逻辑推理

- Markov Logic Networks

http://homes.cs.washington.edu/~pedrod/kbmn.pdf

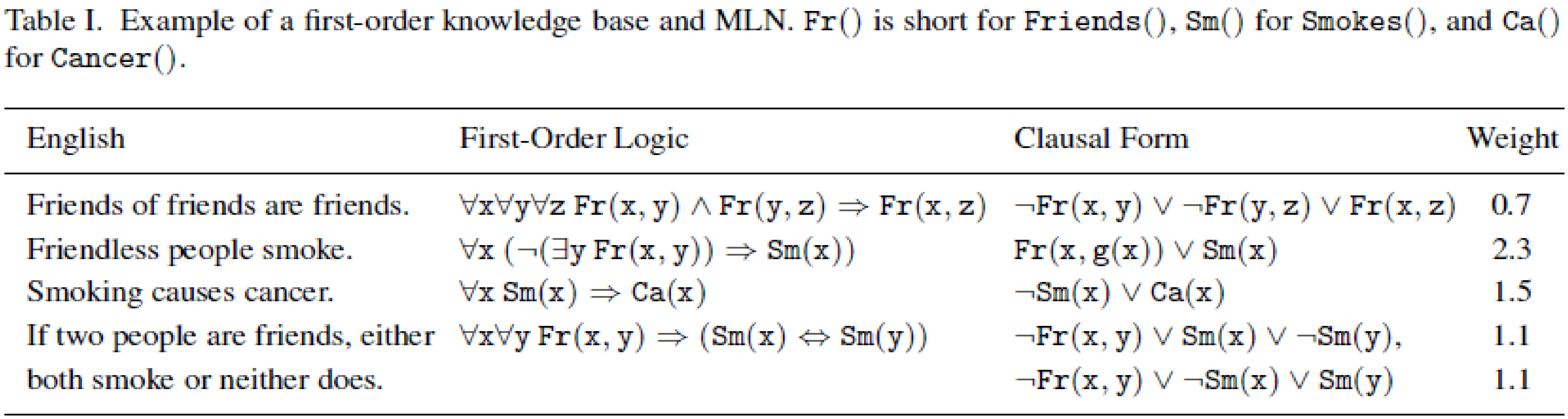


Table I. Example of a first-order knowledge base and MLN. Fr() is short for Friends(), Sm() for Smokes(), and Ca() for Cancer().

English	First-Order Logic	Clausal Form	Weight
Friends of friends are friends.	$\forall x \forall y \forall z\ \mathtt{Fr}(x,y) \wedge \mathtt{Fr}(y,z) \Rightarrow \mathtt{Fr}(x,z)$	$\neg\mathtt{Fr}(x,y) \vee \neg\mathtt{Fr}(y,z) \vee \mathtt{Fr}(x,z)$	0.7
Friendless people smoke.	$\forall x\ (\neg(\exists y\ \mathtt{Fr}(x,y)) \Rightarrow \mathtt{Sm}(x))$	$\mathtt{Fr}(x,g(x)) \vee \mathtt{Sm}(x)$	2.3
Smoking causes cancer.	$\forall x\ \mathtt{Sm}(x) \Rightarrow \mathtt{Ca}(x)$	$\neg\mathtt{Sm}(x) \vee \mathtt{Ca}(x)$	1.5
If two people are friends, either	$\forall x \forall y\ \mathtt{Fr}(x,y) \Rightarrow (\mathtt{Sm}(x) \Leftrightarrow \mathtt{Sm}(y))$	$\neg\mathtt{Fr}(x,y) \vee \mathtt{Sm}(x) \vee \neg\mathtt{Sm}(y),$	1.1
both smoke or neither does.		$\neg\mathtt{Fr}(x,y) \vee \neg\mathtt{Sm}(x) \vee \mathtt{Sm}(y)$	1.1

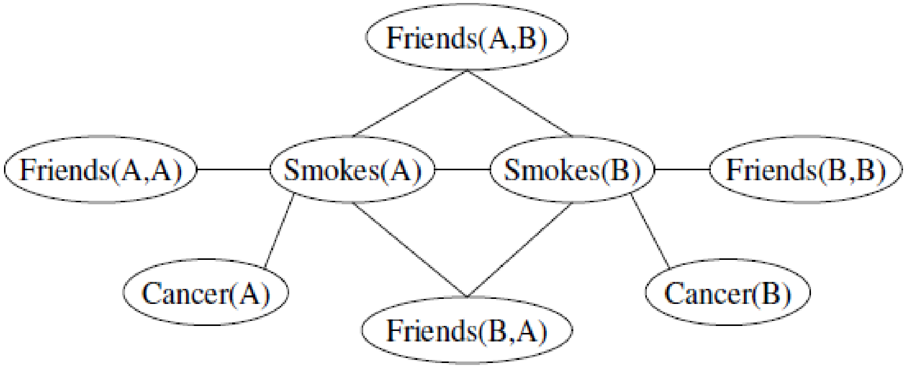
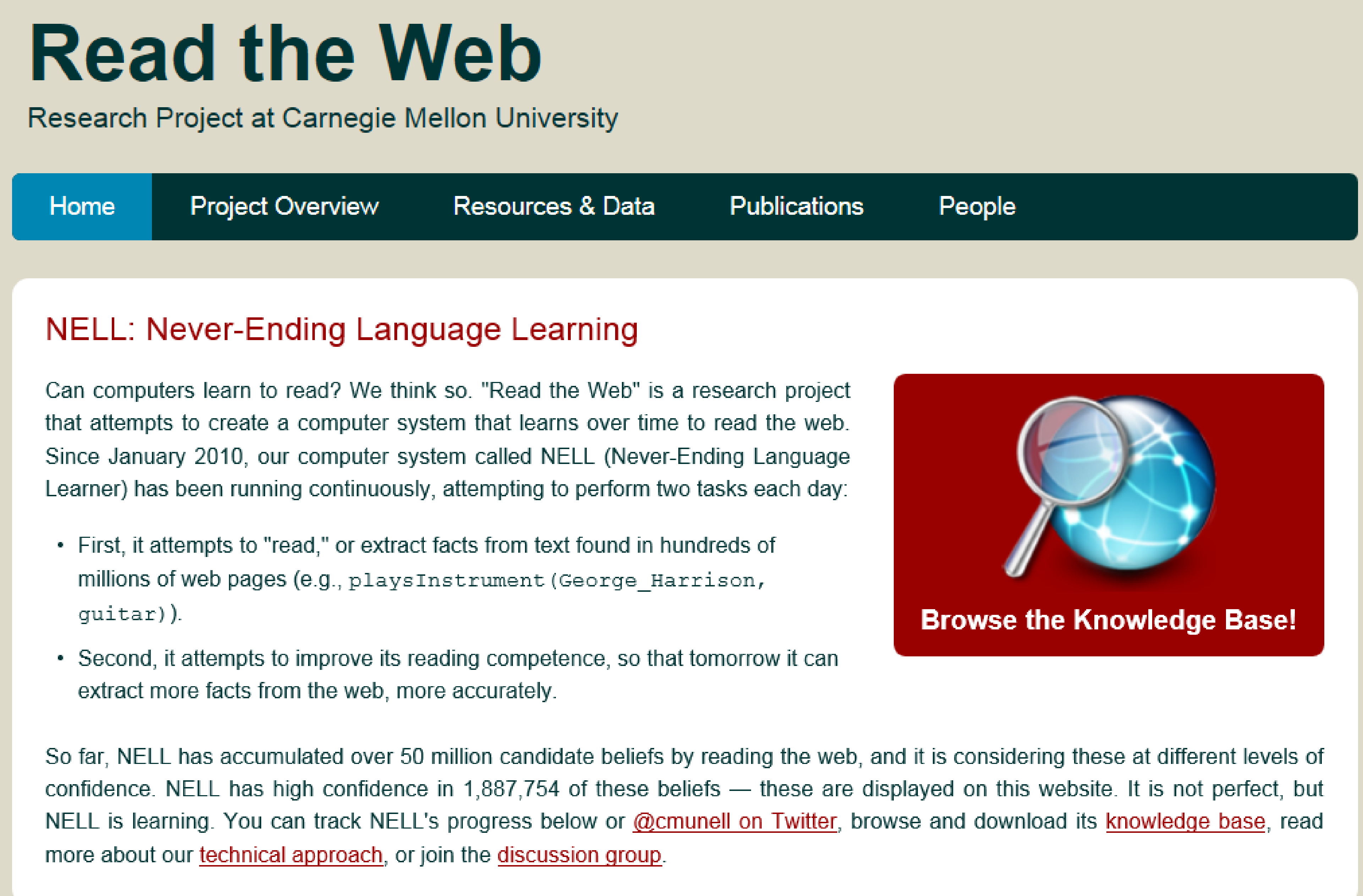


Figure 1. Ground Markov network obtained by applying the last two formulas in Table I to the constants Anna(A) and Bob(B).

10-803: Markov Logic Networks Machine Learning Department, Carnegie Mellon University

http://homes.cs.washington.edu/~pedrod/803/

卡内基梅隆大学:永不停止的语言学习



Read the Web

Research Project at Carnegie Mellon University

Home | Project Overview | Resources & Data | Publications | People

NELL: Never-Ending Language Learning

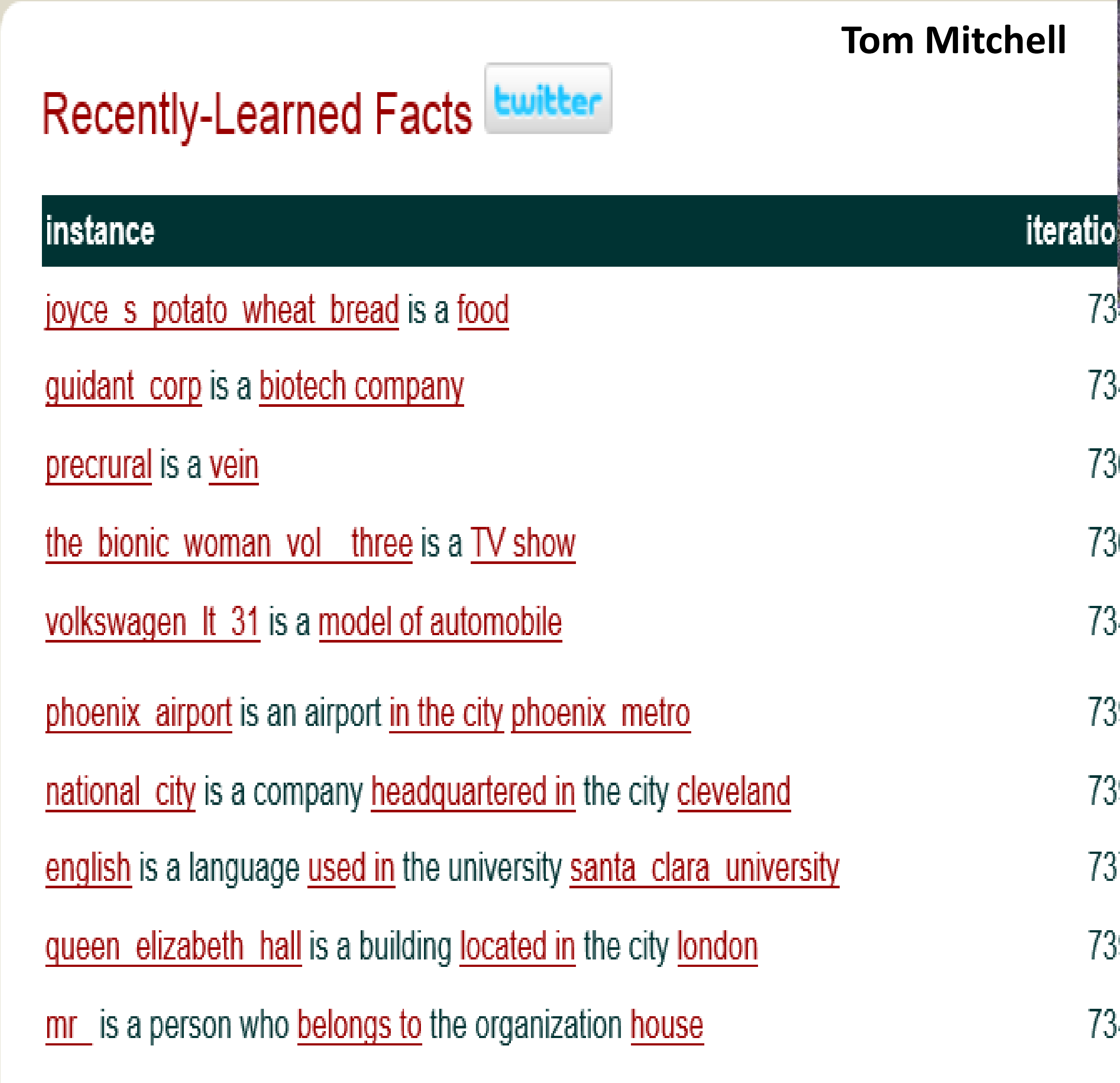
Can computers learn to read? We think so. "Read the Web" is a research project that attempts to create a computer system that learns over time to read the web. Since January 2010, our computer system called NELL (Never-Ending Language Learner) has been running continuously, attempting to perform two tasks each day:

- First, it attempts to "read," or extract facts from text found in hundreds of millions of web pages (e.g., `playsInstrument(George_Harrison, guitar)`).
- Second, it attempts to improve its reading competence, so that tomorrow it can extract more facts from the web, more accurately.

Browse the Knowledge Base!

So far, NELL has accumulated over 50 million candidate beliefs by reading the web, and it is considering these at different levels of confidence. NELL has high confidence in 1,887,754 of these beliefs — these are displayed on this website. It is not perfect, but NELL is learning. You can track NELL's progress below or @cmunell on Twitter, browse and download its knowledge base, read more about our technical approach, or join the discussion group.

卡内基梅隆大学：永不停止的语言学习



Tom Mitchell

Recently-Learned Facts twitter

instance	iteratio		
joyce_s_potato_wheat_bread is a food	734	24-may-2013	92.5
guidant_corp is a biotech company	734	24-may-2013	99.5
precrural is a vein	736	02-jun-2013	97.7
the_bionic_woman_vol__three is a TV show	736	02-jun-2013	94.4
volkswagen_lt_31 is a model of automobile	734	24-may-2013	95.4
phoenix_airport is an airport in the city phoenix_metro	739	09-jun-2013	96.9
national_city is a company headquartered in the city cleveland	739	09-jun-2013	93.8
english is a language used in the university santa_clara_university	737	04-jun-2013	93.8
queen_elizabeth_hall is a building located in the city london	739	09-jun-2013	96.9
mr_ is a person who belongs to the organization house	734	24-may-2013	96.9

目录

1. 引言

2. 三个重要前沿及其进展

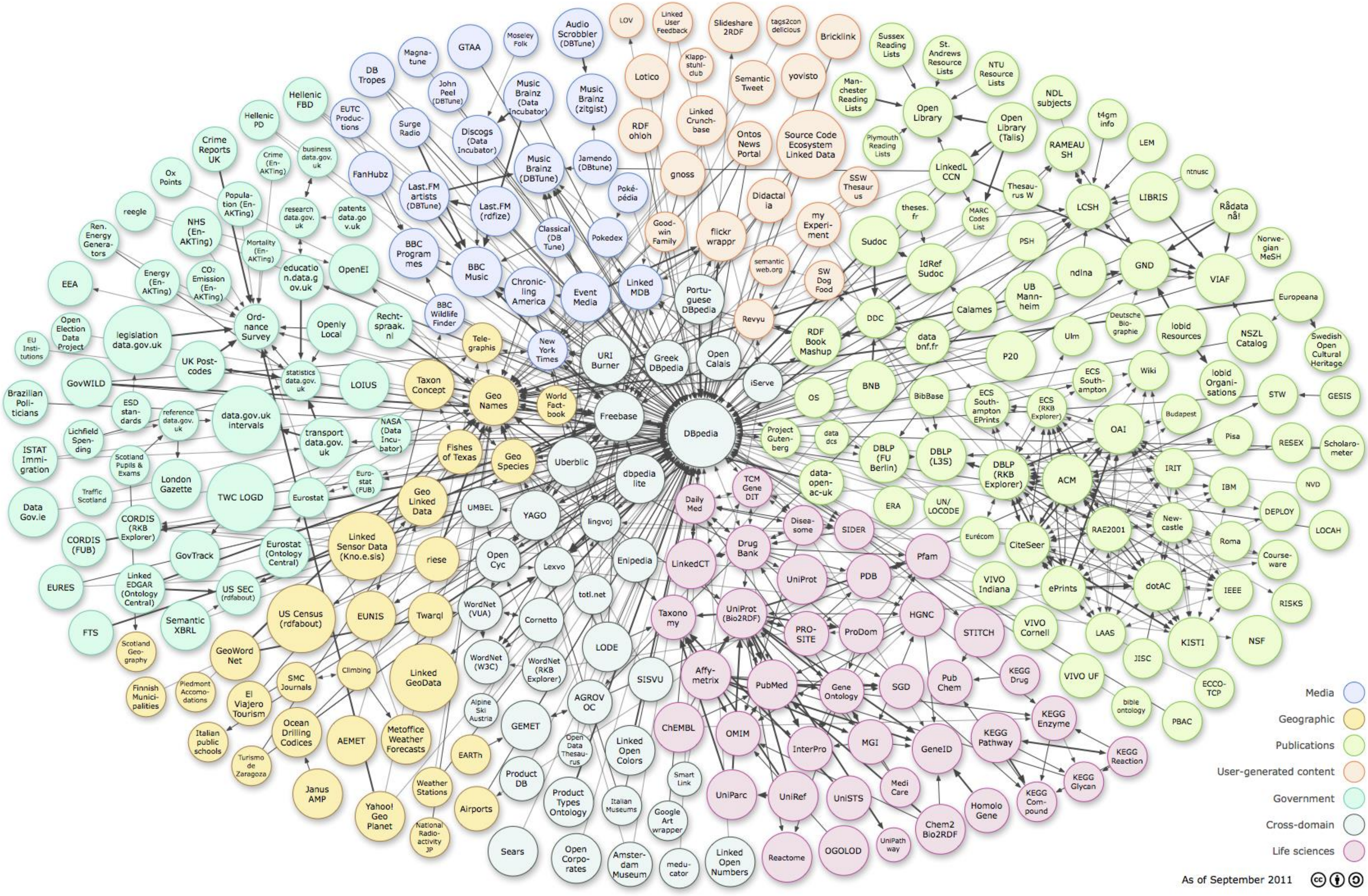
（1）基于深度学习的句子结构预测

（2）开放域信息抽取

（3）知识图谱

3. 题外话

链接数据（Linked Data）



链接数据（Linked Data）

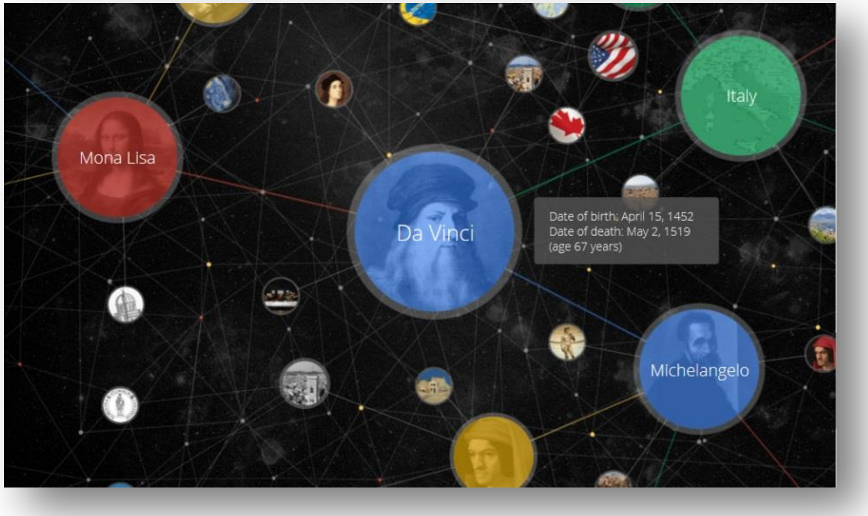
- 现状统计（2011年9月）

Domain	***No of Triples***	***% of Cloud***	***No of Links***	***% of Links***
Media	698.000.000	10,4%	1.238.000	0,8%
Publications	212.000.000	3,2%	4.922.000	3,3%
Life Sciences	2.429.000.000	36,1%	133.199.000	89,4%
Geographic Data	3.097.000.000	46,0%	4.038.000	2,7%
User Generate Content	76.000.000	1,1%	1.559.000	1,0%
Cross-Domain	214.000.000	3,2%	3.992.000	2,7%
Total	***6.726.000.000***		***148.948.000***	

6.7billion facts

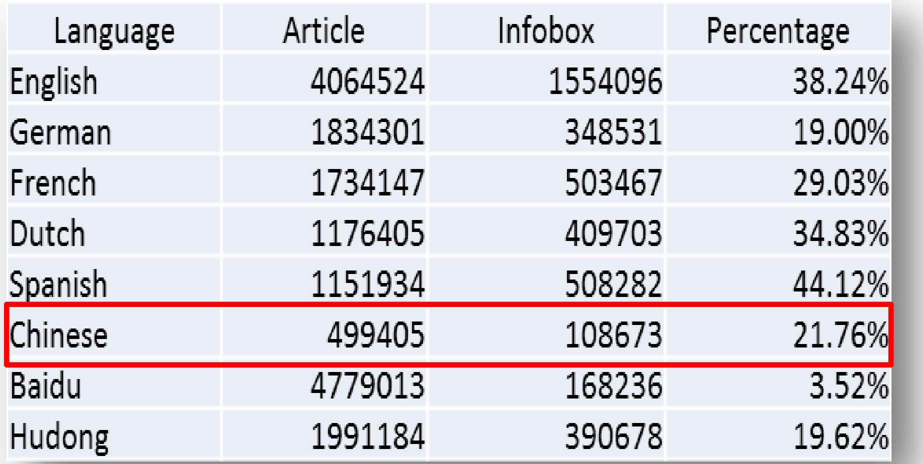
超大规模知识图谱

- 互联网中文理解需要大规模、高覆盖率的知识资源
- 目前的知识资源难以满足中文理解的需求，以Google知识图谱（5亿个实体，35亿个事实）为例：
  - 主要描述实体以及实体之间关系，对于复杂事件的描述甚少
  - 英文知识图谱关于中国的内容很少
  - 中文知识图谱正在构建中，主要挑战之一是infobox信息匮乏
- 百度知识图谱与搜狗知立方也面临类似的问题



Date of birth: April 15, 1452
Date of death: May 2, 1519
(age 67 years)

Google知识图谱



Language	Article	Infobox	Percentage
English	4064524	1554096	38.24%
German	1834301	348531	19.00%
French	1734147	503467	29.03%
Dutch	1176405	409703	34.83%
Spanish	1151934	508282	44.12%
Chinese	499405	108673	21.76%
Baidu	4779013	168236	3.52%
Hudong	1991184	390678	19.62%

维基百科仅有21%的中文文章有infobox

项羽 [编辑]

维基百科，自由的百科全书

项羽（前232年－前202年），名**籍**，字**羽**，以字行。古代中国将领，出生于楚国下相（今江苏省宿迁市宿城区）人，7岁后随叔父项梁迁吴中（今江苏苏州市）。秦末时被楚怀王封为"**鲁公**"，在前207年的决定性战役钜鹿之战中统率楚军大破秦军，秦亡后自封"**西楚霸王**"，统治黄河及长江下游的梁楚九郡，后在楚汉战争中的垓下之战为刘邦所败，突围至长江北岸乌江自刎。

项羽的勇武可称天下无敌[1]（古人对其有"羽之神勇，千古无二"的评价[2]），被称为中国数千年来最为勇猛的将领。[3][4]"霸王"一词，专指项羽。

项羽

项羽画像，上官周《晚笑堂竹庄画传》

西楚霸王	
姓	项
名	籍
字	羽
出生	前232年 楚国下相
逝世	前202年 楚国乌江
亲属	显示▼

目录 [隐藏]

关羽（？ －220年）[1]，字**云长**，本字**长生**，司隶河东解人（今山西省运城市），约生于东汉桓帝延熹年间[2]，东汉末年三国时期刘备的重要将领。

关羽最为特殊之处是其倍受中华文化历代推崇，由于其忠义和勇武的形象，多被民众尊称为**关公**、**关老爷**，又多次被后代帝王褒封，直至“武帝”，故也被称为**关圣帝君**、**关圣帝**、**关帝君**、**关帝**等。道教奉为五文昌之一，又尊为“**文衡圣帝**”，或“**协天大帝**”、“**翊汉天尊**”。中国佛教界奉其为护法神之一，称为“**伽蓝菩萨**”。民间由于《三国演义》等传统作品的影响，普遍认为关羽与刘备、张飞是结义兄弟，关羽排行第二，故又俗称其为**关二爷**、**关二哥**，直至现当代的某些社会群体与场合中，仍然经常出现祭拜关羽的情况。

目录 [隐藏]

生平 [编辑]

关羽

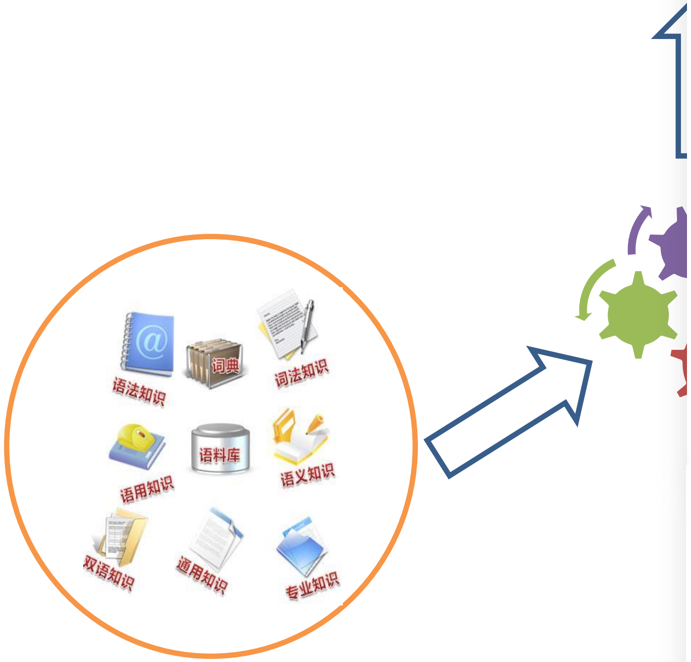
前将军

国家	汉
时代	三国
主君	刘备
姓	关
名	羽
姓名	关羽
字	长生→云长
封爵	汉寿亭侯
封号	亭侯
尊号	关公
本贯	解县
出生	不明 东汉河东解县（今山西省运城市）
逝世	献帝建安二十四年十二月（219年） 东汉（三国）荆州临沮（今湖北省襄阳南漳县）
谥号	壮缪侯
墓葬	关羽墓

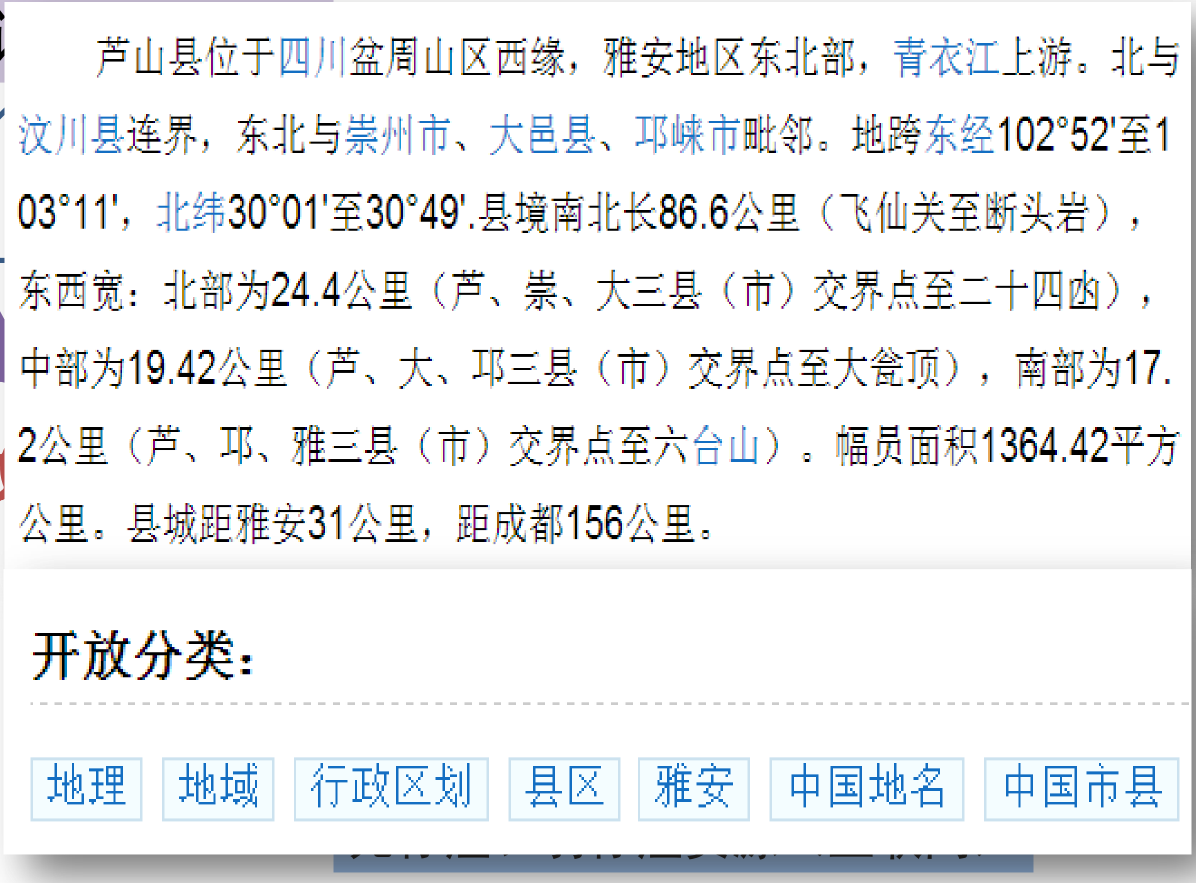
超大规模知识图谱的建构思路

- 精标注资源与海量无标注、弱标注资源的融合

适合互联网中文理解的知i



精标注资源（专家）



芦山县位于四川盆周山区西缘，雅安地区东北部，青衣江上游。北与汶川县连界，东北与崇州市、大邑县、邛崃市毗邻。地跨东经102°52'至103°11'，北纬30°01'至30°49'.县境南北长86.6公里（飞仙关至断头岩），东西宽：北部为24.4公里（芦、崇、大三县（市）交界点至二十四凼），中部为19.42公里（芦、大、邛三县（市）交界点至大瓮顶），南部为17.2公里（芦、邛、雅三县（市）交界点至六台山）。幅员面积1364.42平方公里。县城距雅安31公里，距成都156公里。

开放分类：

地理 地域 行政区划 县区 雅安 中国地名 中国市县

目录

1. 引言

2. 三个重要前沿及其进展

（1）基于深度学习的句子结构预测

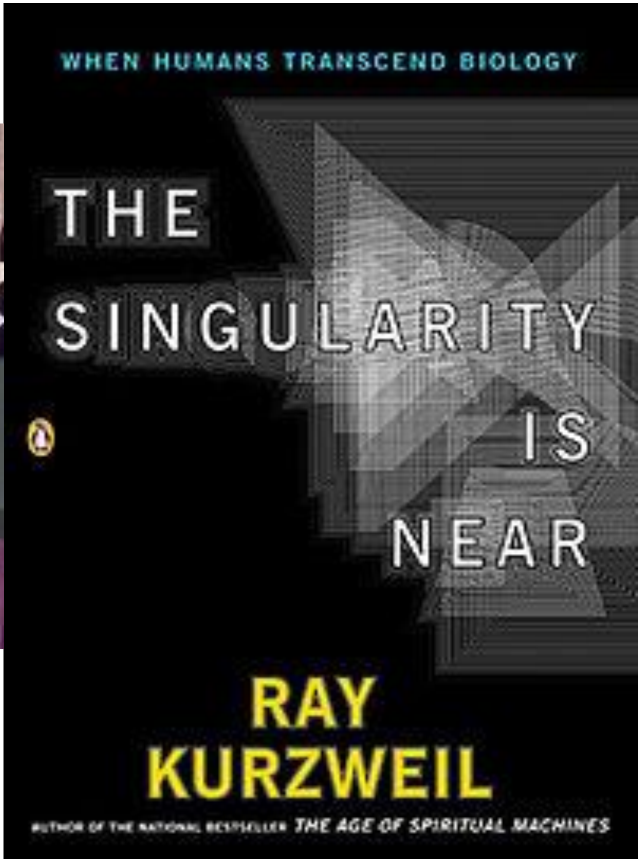
（2）开放域信息抽取

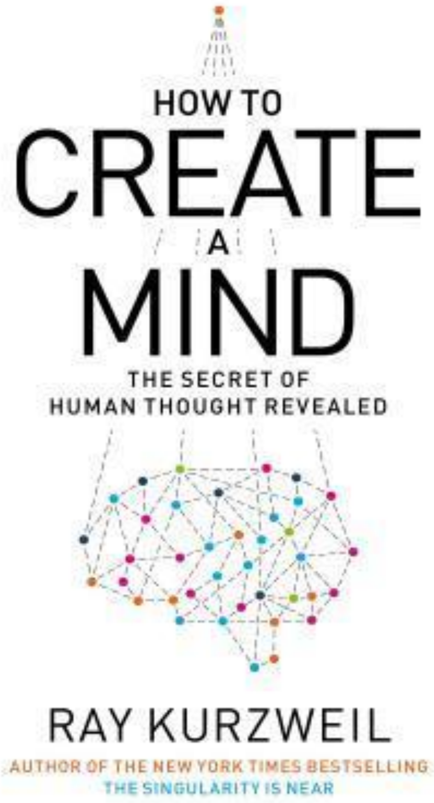
（3）知识图谱

3. 题外话

不太远的愿景：奇点临近?

Kurzweil “I set the date for the Singularity — representing a profound and disruptive transformation in human capability — as 2045”


WHEN HUMANS TRANSCEND BIOLOGY
THE SINGULARITY IS NEAR
RAY KURZWEIL
AUTHOR OF THE NATIONAL BESTSELLER THE AGE OF SPIRITUAL MACHINES


HOW TO CREATE A MIND
THE SECRET OF HUMAN THOUGHT REVEALED
RAY KURZWEIL
AUTHOR OF THE NEW YORK TIMES BESTSELLING
THE SINGULARITY IS NEAR

一个难得的学术交流机会

“第十二届全国计算语言学会议(CCL 2013)及第一届基于自然标注大数据的自然语言处理国际学术研讨会(NLP-NABD 2013)

“知识图谱”研讨会

深度广度兼具的邀请报告

http://210.29.169.226/CNCCL2013/home.html

Maosong Sun Min Zhang
Dekang Lin Haifeng Wang (Eds.)

Chinese Computational Linguistics *and* Natural Language Processing Based on Naturally Annotated Big Data

12th China National Conference, CCL 2013 *and*
First International Symposium, NLP-NABD 2013
Suzhou, China, October 2013, Proceedings

LNAI 8202

Sun • Zhang • Lin • Wang (Eds.)

LNAI 8202

Chinese Computational Linguistics *and* Natural Language Processing Based on Naturally Annotated Big Data

CCL/
NLP-NABD
2013

中国中文信息学会
CIPS
Chinese Information Processing Society of China

Springer

谢谢！

欢迎访问：

清华大学自然语言处理与社会人文计算实验室网站：

http://nlp.csai.tsinghua.edu.cn/site2/

孙茂松微博：

http://weibo.com/u/1970879995