

Building Knowledge (on Things) for AI

Seung-won Hwang
Data Intelligence Lab
Yonsei University

Text Analytics without Knowledge

Does Anne Hathaway News Drive Berkshire Hathaway's Stock?

ALEXIS C. MADRIGAL | MAR 18, 2011 | TECHNOLOGY

Share Tweet ...

TEXT SIZE
- +

In the first episode of [The Atlantic Interview](#), Chimamanda Ngozi Adichie talks with Jeffrey Goldberg and Ta-Nehisi Coates about race and identity. [Listen and subscribe](#) to the podcast.

Given the awesome correlating powers of today's stock trading computers, the idea may not be as far-fetched as you think.



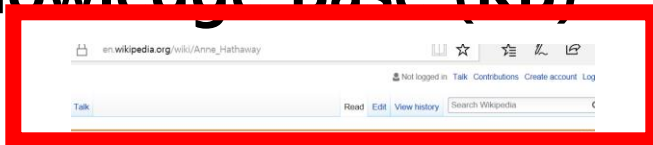
A couple weeks ago, Huffington Post blogger Dan Mirvish noted a funny trend: when Anne Hathaway was in the news, Warren Buffett's Berkshire Hathaway's shares went up. He pointed to [six dates going back to 2008](#) to show the

Is human any better?

Indeed, **Hathaway**'s life has seemed especially charmed recently. Prior to her successful awards season run, the actress married **Adam Shulman** in

Knowledge on Things (KoT)

Knowledge Base (KB)



Does Anne Hathaway News Drive Berkshire Hathaway's Stock



anne Hathaway

Wikipedia, the free encyclopedia

This article is about the actress. For the wife of William Shakespeare, see *Anne Hathaway (wife of Shakespeare)*. For other uses, see *Anne Hathaway (disambiguation)*.

Jacqueline Hathaway (born November 12, 1982) is an American actress and singer. Her breakthrough debut film role was as Mia Thermopolis in the Disney comedy *The Princess Diaries* (2001). Hathaway made a transition to adult roles in box office and ally acclaimed films such as *Brokeback Mountain* (2005), *The Devil Wears Prada* (2006), as Jane Austen in *Becoming Jane* (2007), *Rachel Getting Married* (2008), *Bride Wars* (2009), *Valentine's Day* (2010), as the White Queen in Tim Burton's *Alice in Wonderland* (2010), *Love & Other Drugs* (2010), as DC Comics femme fatale Catwoman in Christopher Nolan's *The Dark Knight Rises* (2012), as Fantine in Tom Hooper's *Les Misérables* (2012), *Interstellar* (2014), *The Intern* (2015), and *Alice Through the Looking Glass* (2016).

Her notable accolades include the Academy Award for Best Supporting Actress, Golden Globe, BAFTA, and SAG Awards for Best Supporting Actress (*Les Misérables*), Critics' Choice Movie Award for Best Actress (*Rachel Getting Married*), Satellite Award for Best Actress – Motion Picture Musical or Comedy (*Love & Other Drugs*), Saturn Award for Best Supporting Actress (*The Dark Knight Rises*), Emmy Award for Outstanding Voice-Over Performance (*Once Upon a Time in Springfield*), as well as an Academy, Golden Globe, Screen Actors Guild and Indie Spirit Awards nomination for Best Actress (*Rachel Getting Married*) and a BIFA nomination for Best Performance by an Actress (*Becoming Jane*).^[1] *People* magazine named Hathaway one of its breakthrough stars of the year^[2] and *Times* listed her as one of the 50 Most Beautiful Women in Film.^[3] Hathaway began dating businessman Adam Shulman in 2008. They married in 2012 and have one son, born in March 2016. Hathaway is a board member of The Lollipop Theatre Network,^[4] and advocates gender equality as a UN Women Goodwill Ambassador.^[5] *Forbes* named Hathaway one of the highest paid actresses of 2015,^[6] and since 2017 she is among the highest-grossing actresses of the 21st century.^[6]

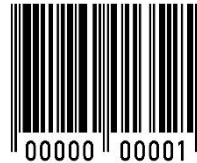
Contents [hide]

Anne Hathaway



Hathaway in 2017

Born	Anne Jacqueline Hathaway November 12, 1982 (age 34) Brooklyn, New York, U.S.
Alma mater	NYU Tisch School of the Arts Vassar College
Occupation	Actress, singer
Years active	1999–present
Spouse(s)	Adam Shulman (m. 2012)
Children	1
Awards	Full list









Tagged text

Topics

Does [Anne Hathaway](#) News Drive [Berkshire Hathaway's Stock](#)

Freebase

Other Modality of KoT

Rank	Name	Net Worth	Age	Source	Country of Citizenship
	#1 Bill Gates	\$75 B	61	Microsoft	United States
	#2 Amancio Ortega	\$67 B	80	Zara	Spain
	#3 Warren Buffett	\$60.8 B	86	Berkshire Hathaway	United States
	#4 Carlos Slim Helu	\$50 B	76	telecom	Mexico
	#5 Jeff Bezos	\$45.2 B	53	Amazon.com	United States
...					
	#40 Paul Allen	\$17.5 B	63	Microsoft, investments	United States

Paul Gardner Allen (born January 21, 1953) is an American business magnate, investor and philanthropist. He is best known as the co-founder of **Microsoft**, alongside **Bill Gates**. As of August 2016, he was estimated to be the 40th richest person in the world, with an estimated wealth of \$18.6 billion.^[2]



Problem Space (Tutorial@KDD15)

- **Growth: knowledge graphs are incomplete!**
 - *Link prediction*: add relations
 - *Ontology matching*: connect graphs
 - *Knowledge extraction*: extract new entities and relations from web/text
- **Validation: knowledge graphs are not always correct!**
 - *Entity resolution*: merge duplicate entities, split wrongly merged ones
 - *Error detection*: remove false assertions
- **Interface: how to make it easier to access knowledge?**
 - *Semantic parsing*: interpret the meaning of queries
 - *Question answering*: compute answers using the knowledge graph
- **Intelligence: can AI emerge from knowledge graphs?**
 - *Automatic reasoning* and planning
 - Generalization and abstraction

KoT Productized (circa 2017)

Paul Gardner Allen (born January 21, 1953) is an American business magnate, investor and philanthropist. He is best known as the co-founder of **Microsoft**, alongside **Bill Gates**. As of August 2016, he was estimated to be the 40th richest person in the world, with an estimated wealth of \$18.6 billion.^[2]



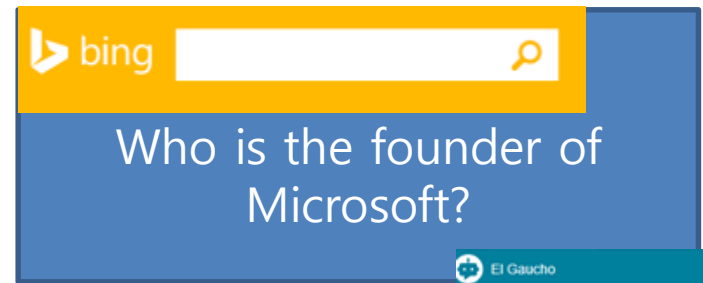
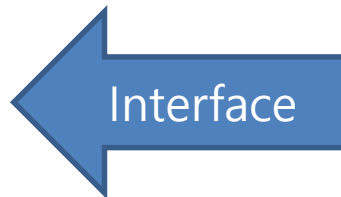
폴 앨런
사업가

폴 가드너 앨런은 빌 게이츠와 더불어 마이크로소프트를 공동 창업한 사업가이자 시애틀 사운더스 FC의 구단주로, 최근 들어 여러 부자 목록에 이름이 등재되었으며, 포브스 지에 따르면 그의 재산은 2006년 현재 세계 6위로, 50억 달러 상당의 마이크로소프트 주식을 포함, 총 227억 달러에 달한다. 위키백과

출생: 1953년 1월 21일 (63세), 미국 워싱턴 주 시애틀
 저서: 아이디어맨, 아이디어맨 원클릭 세트(전2권)
 부모: 페이 G. 알렌, 케네스 S. 알렌
 학력:레이크사이드 스쿨, 워싱턴 주립 대학교
 형제자매: 조디 패튼
 설립 단체: 마이크로소프트

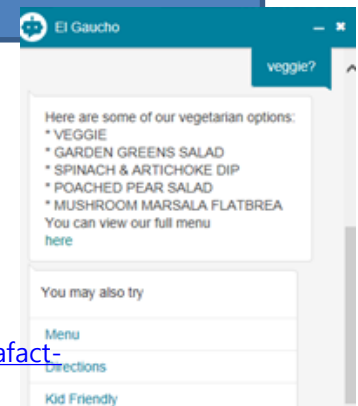
함께 찾은 검색어

10개 이상 항목 더보기



	A	B
1		Founder
2	Microsoft	
3	Google	
4	Facebook	

<https://www.microsoft.com/en-us/research/video/video-abstract-insta-fact-bringing-knowledge-office-apps/>



KoT Directions from 2017 Conferences



Direction 1: Interactive Learning on Things

The image shows a Bing search results page for the query "singapore history". The search bar at the top contains the text "singapore history" and a magnifying glass icon. To the right of the search bar, there are links for "Sign in" and "19502" along with a menu icon. Below the search bar, there are tabs for "Web", "Images", "Videos", "Maps", and "News". The "Web" tab is selected. Below the tabs, there are suggestions: "Also try: singapore history timeline · the culture of singapore · singapore colonial ...". The search results show "14,300,000 RESULTS" and "Any time" filter. The main result is "History of Singapore" with a brief description: "The written **history of Singapore** dates back to the third century. Later, the Kingdom of Singapura rose in importance during the 14th century under the rule of Srivijayan prince Parameswara and Singapore became an important port, until it was destroyed by Acehese raiders in 1613." Below this is a link to the Wikipedia article: "History of Singapore - Wikipedia, the free encyclopedia" with the URL "en.wikipedia.org/wiki/History_of_Singapore". At the bottom of the result box, there is a "Chat" button circled in blue. To the right, an InfoBot chat window is open, displaying the question "what about ethnicity?" and the answer: "Singapore - Ethnic groups: Chinese 74.2%, Malay 13.3%, Indian 9.2%, other 3.3% (2013)". The chat window also shows a map of Singapore and the Singapore flag.

Search results for "singapore history" on Bing. The main result is "History of Singapore" from Wikipedia, which states: "The written **history of Singapore** dates back to the third century. Later, the Kingdom of Singapura rose in importance during the 14th century under the rule of Srivijayan prince Parameswara and Singapore became an important port, until it was destroyed by Acehese raiders in 1613." Below the text is a link to the full Wikipedia article: en.wikipedia.org/wiki/History_of_Singapore.

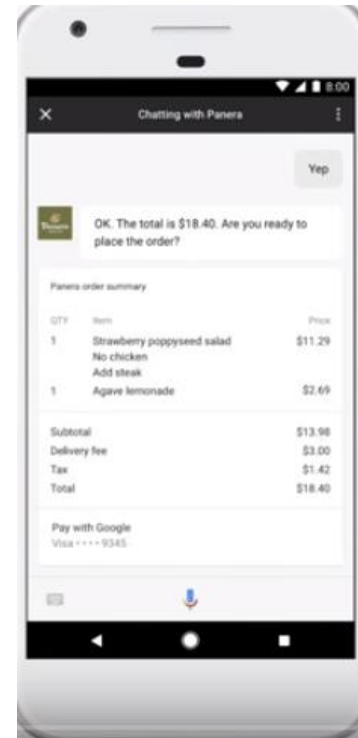
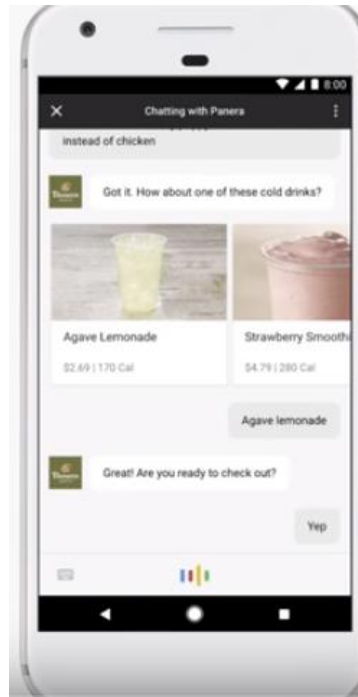
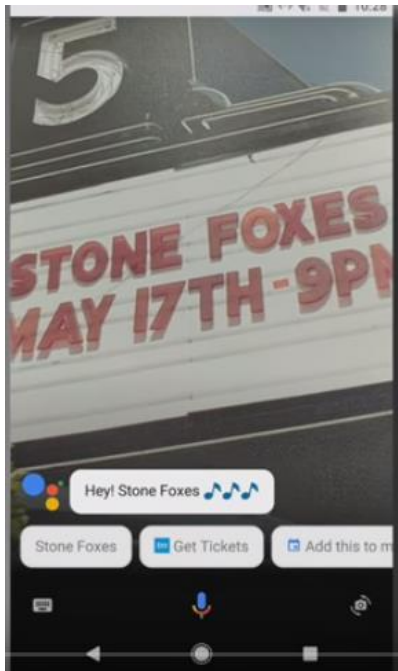
At the bottom of the result box, there is a "Chat" button circled in blue, which is part of the InfoBot interface. The InfoBot chat window is open, showing the question "what about ethnicity?" and the answer: "Singapore - Ethnic groups: Chinese 74.2%, Malay 13.3%, Indian 9.2%, other 3.3% (2013)". The chat window also displays a map of Singapore and the Singapore flag.

Direction 2: Actuation on Things

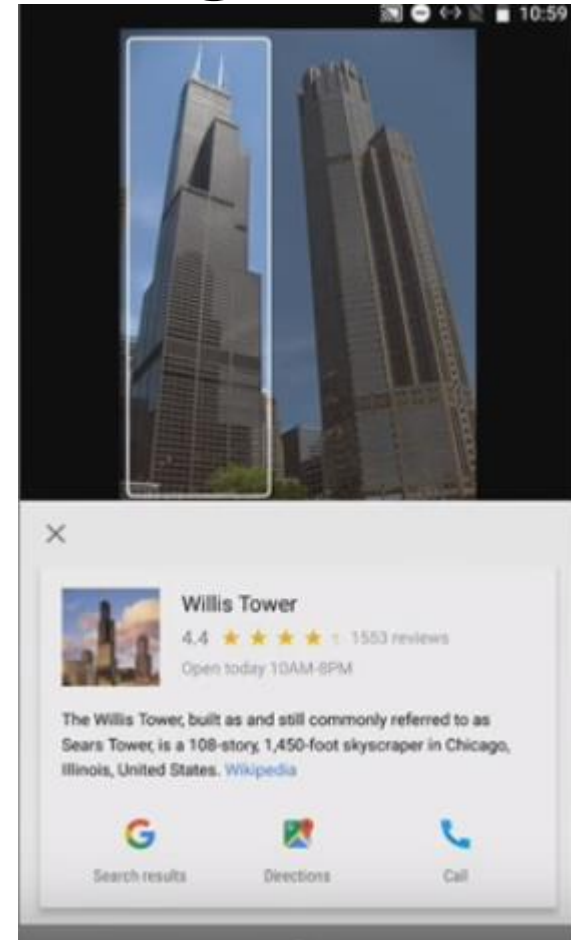
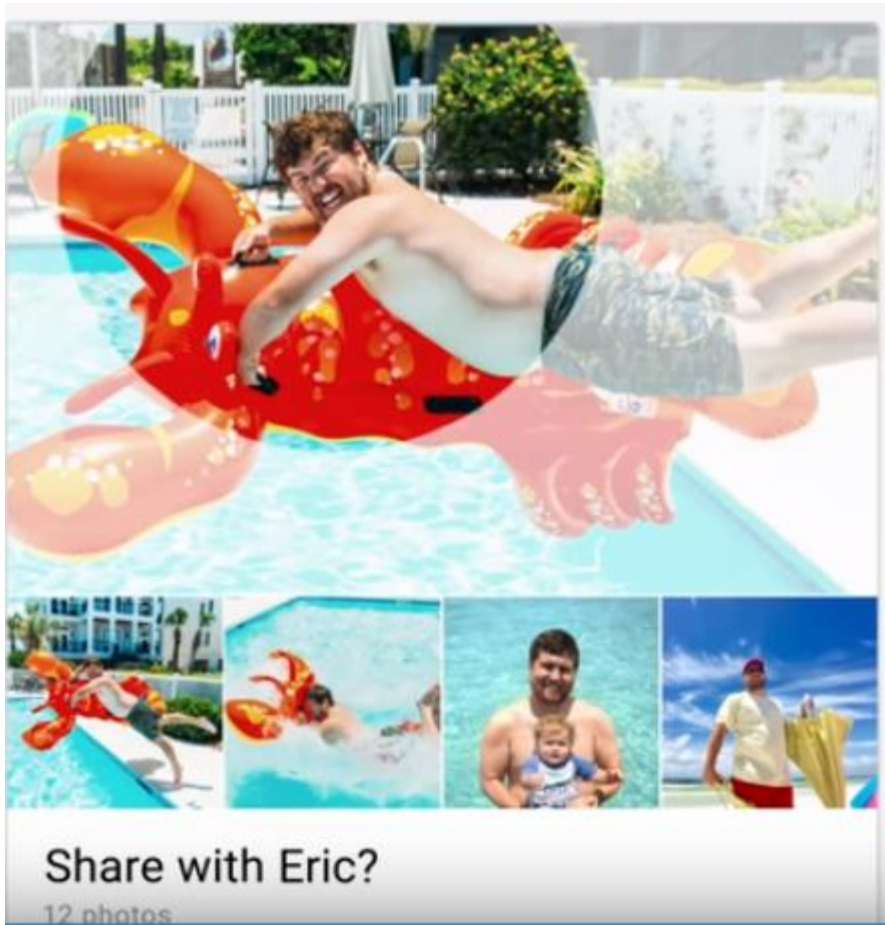
- Google Lens: camera (perception) into search → action bot



Direction 2: Actuation on Things



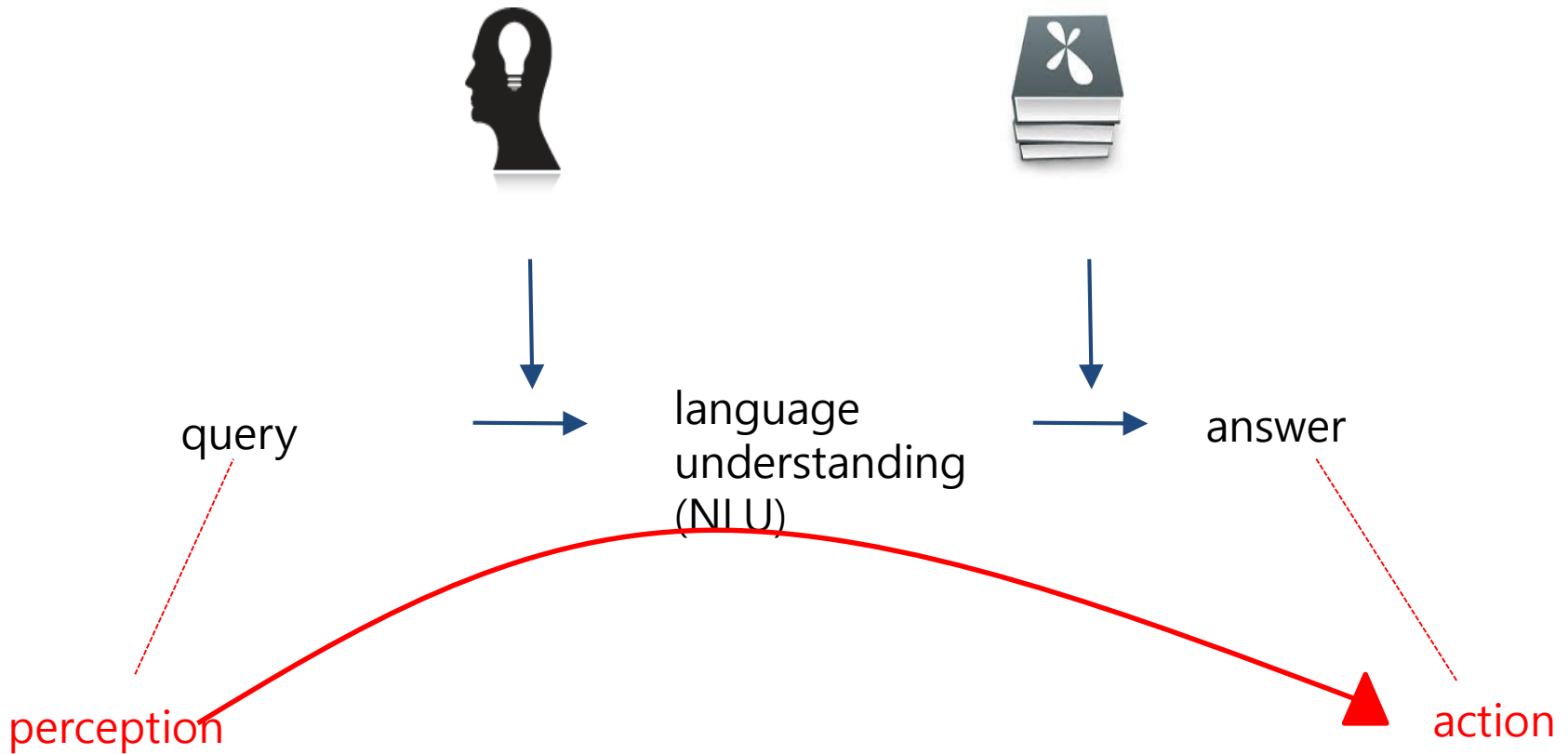
Direction 2: Actuation on Things



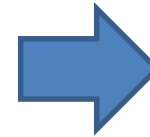
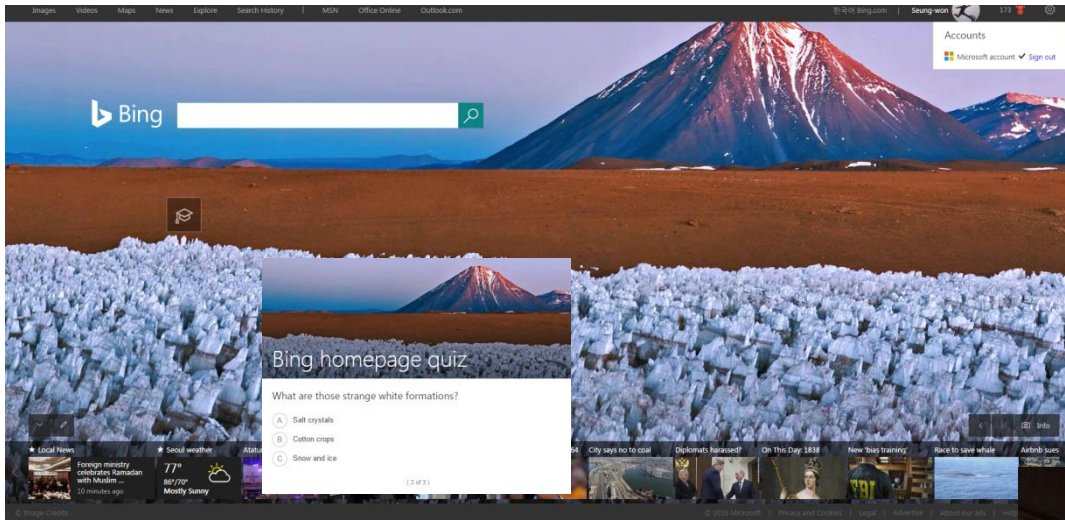
Direction 2: Actuation on Things



DILAB focus



DILAB Project Showcase (1/2)



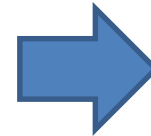
COLING2016

What's the aspect about this thing people are interested?



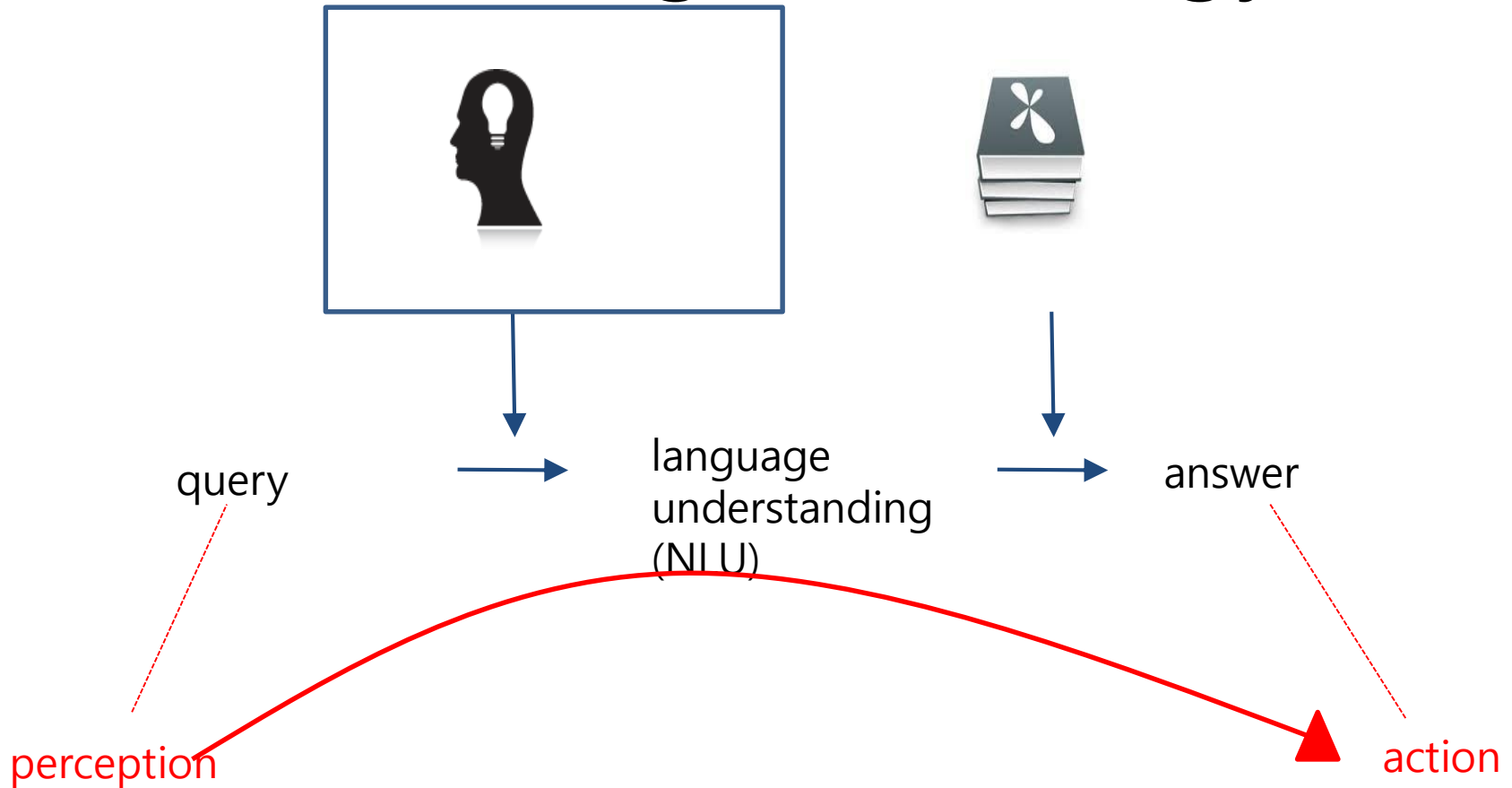
DILAB Project Showcase (2/2)

- Language-driven highlighting



Show me Sarah's highlights

DILAB focus: Enabling Technology



isA Extraction



- **Hearst pattern**

NP such as NP, NP, ..., and|or NP

such NP as NP,* or|and NP

NP, NP*, or other NP

NP, NP*, and other NP

NP, including NP,* or | and NP

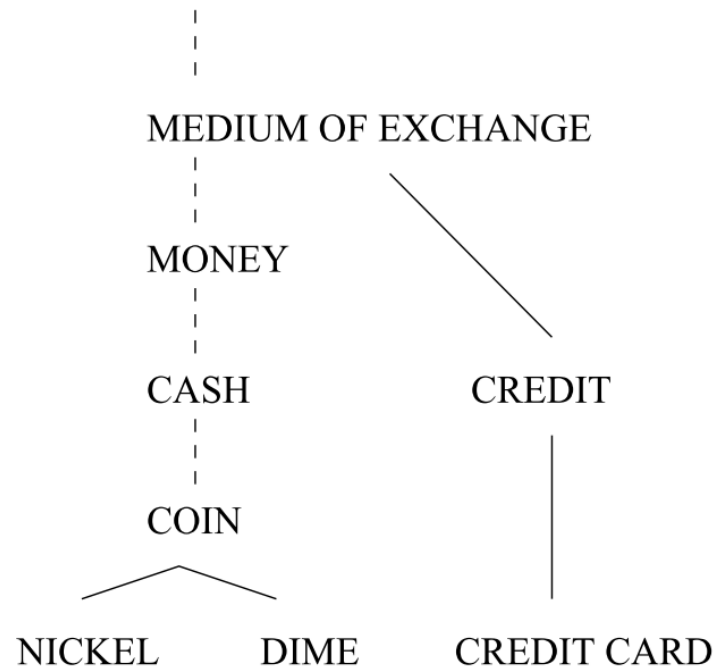
NP, especially NP,* or|and NP

- **... is a ... pattern**

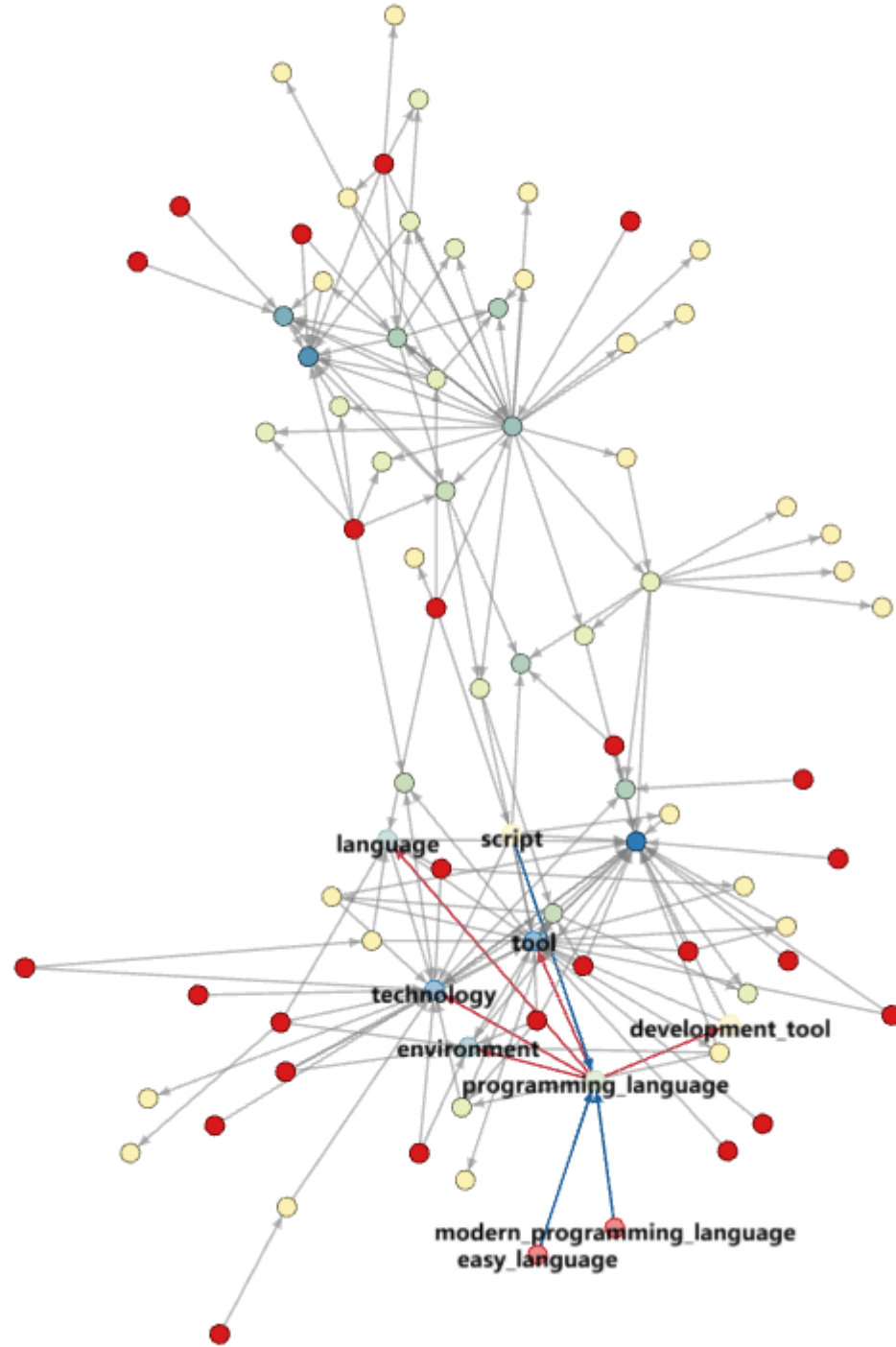
NP is a/an/the NP

- *domestic animals* such as *cats* and *dogs* ...
- animals other than *cats* such as *dogs* ...
- *China* is a *developing country*.

A traditional taxonomy

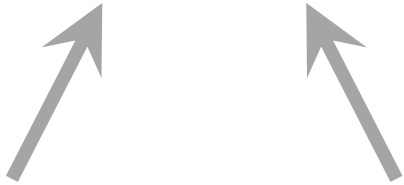


“python”



Typicality

bird



$$P(e|c) = \frac{n(c, e) + \alpha}{\sum_{e_i \in c} n(c, e_i) + \alpha N}$$

$$P(c|e) = \frac{n(c, e) + \alpha}{\sum_{e \in c_i} n(c_i, e) + \alpha N}$$

“robin” is a more *typical* bird than a “penguin”



$$p(\text{robin}|\text{bird}) > p(\text{penguin}|\text{bird})$$

One hot vs. Distributed Representation



Similarity

- microsoft, ibm → 0.933
- google, apple → 0.378 ??

$$sim(t_1, t_2) = \max_{x,y} cosine(c_x(t_1), c_y(t_2))$$

Contextual Concept Learning: Bayesian

$$P(c_k|E) = \frac{P(E|c_k)P(c_k)}{P(E)} \propto P(c_k) \prod_{i=1}^M P(e_i|c_k).$$

Google

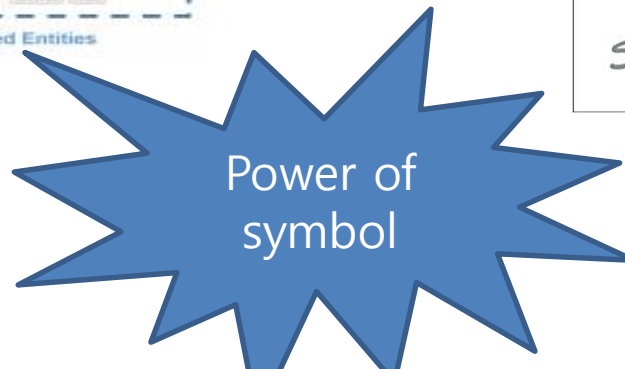
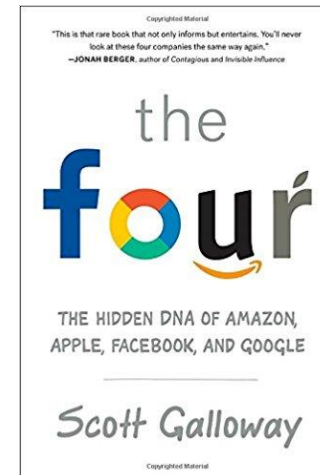
Apple



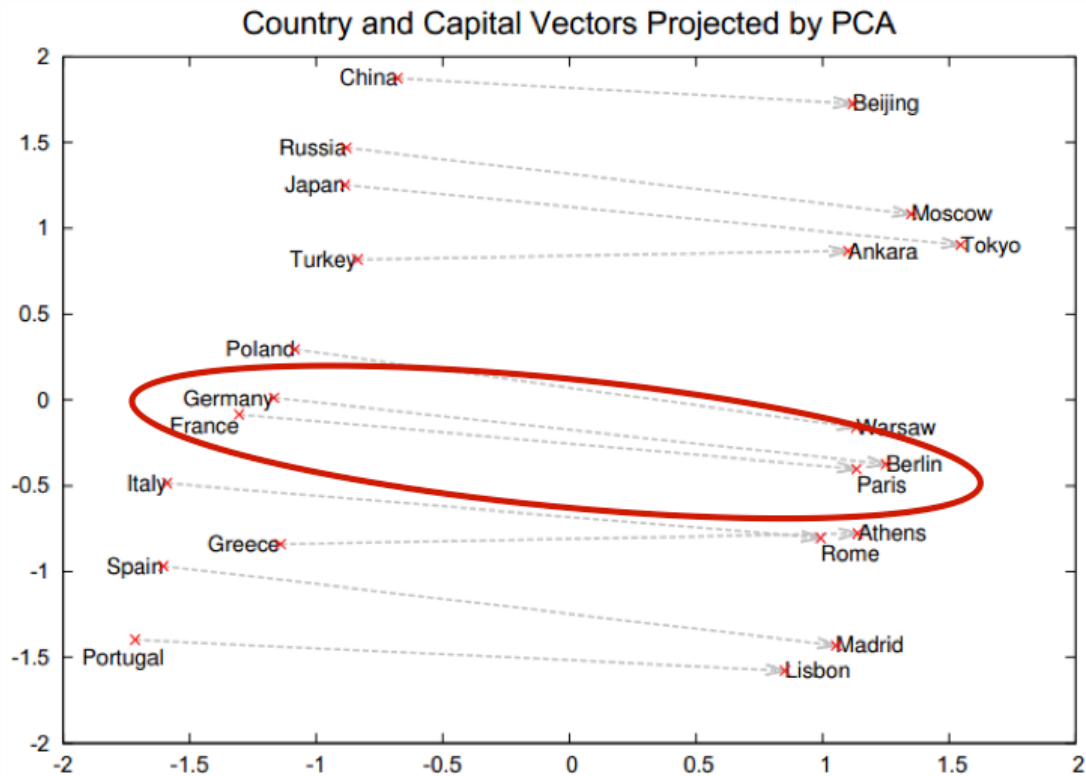
IT company

Conceptualization for Explainability [IJCAI2017]

Google Apple Amazon FB



Complements Neural Embedding (Word2Vec)



Search Query Parsing for Action

What are
Things?
Actions?

- watch harry potter



watch

- harry potter watch



buy

Conceptualizing FrameNet for Verb Understanding [AAAI2016]

Frame: Apply_heat

FE1

FE2

FE3

FE4

She was **FRYING** eggs and bacon and mushrooms on a camp stove in Woolley's billet.



Concept	P(c FE)	Instance	P(w FE)
heat source	0.19	Stove	0.00019
Large metal	0.04	Radiator*	0.00015
Kitchen appliance	0.02	Oven	0.00015
		Grill*	0.00014
		Heater*	0.00013
		Fireplace*	0.00013
		Lamp*	0.00013
		Hair dryer*	0.00012
		Candle*	0.00012

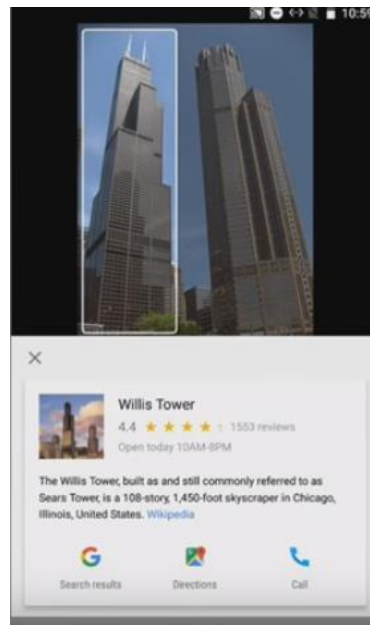


2017 Sampler

- Knowledge base
 - Growth: Multimodal KB [TKDE17]
 - Validate: Extracting/correcting knowledge from text corpus [AAAI17]
 - Interface: KBQA [VLDB17]
 - Reasoning: Adjective understanding [PAKDD2017]

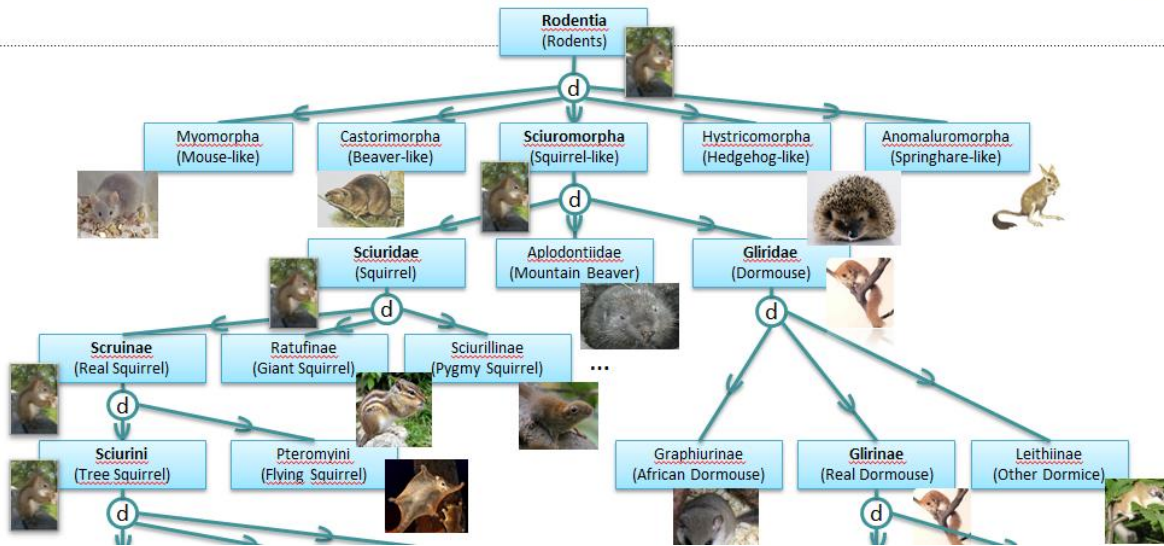
Growth:
TKDE
2017

Multimodal Entity Linking

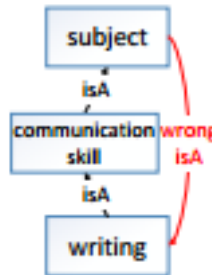
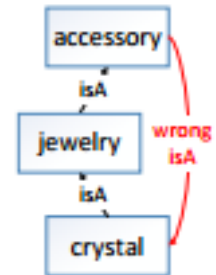
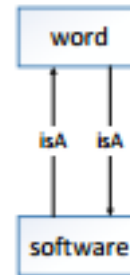


Eiffeltoring 에펠탑 torre Eiffel আইফেল টাওয়ার eiffel torre kull
Eifel Айфеловата кула Eiffelov toranj Эйфелева вежа 艾菲尔塔
eiffelova věž エッフェル塔 eiffel tower Eiffel Dorrea πύργος
του Άιφελ tour eiffel Eyfel qülləsi Эйфелева башня برج ایفل

Handmade vs. Noisy KB



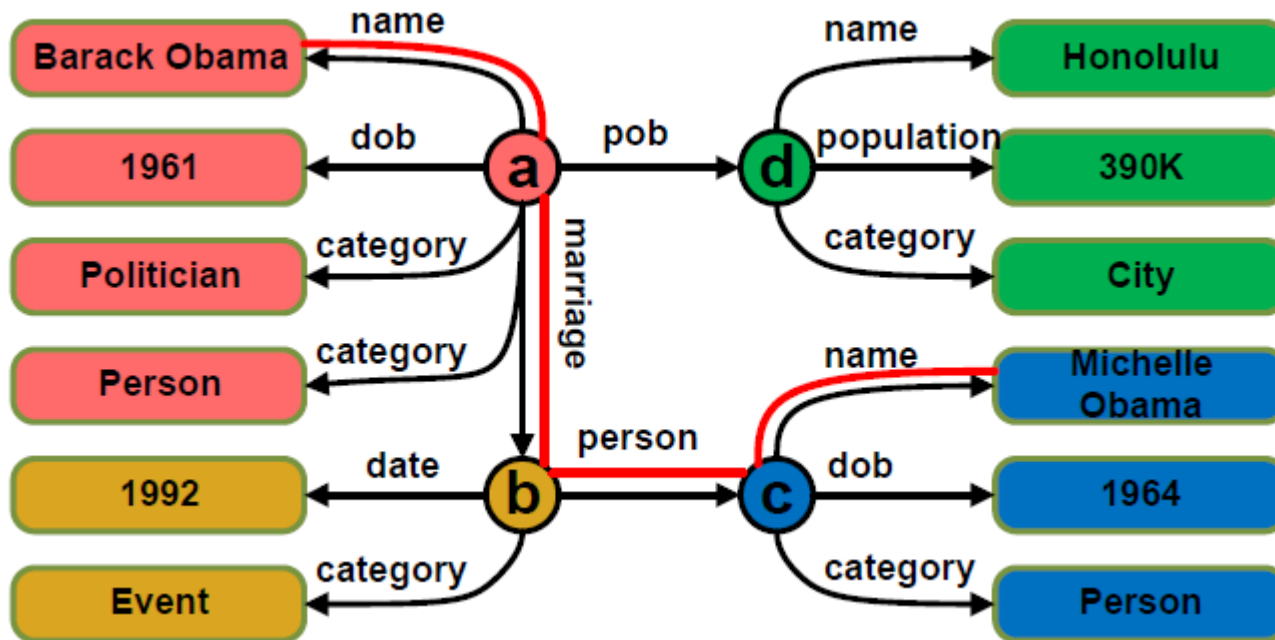
Org chart with perfect hierarchy
($h=1$)



Cycle ($h=0$)

QA on Things

- When was Barack Obama's wife born?



QA Corpus Mining (Offline)

- Questions represented as KB-template **t**
 - How many people are in \$city?
 - What is the population of \$city?
- Template mapped to predicate **p**
 - City-> population

Template	Predicate	F
What are symptoms of \$disease?	symptom → name	2079
Where is \$location?	contained_by → name	1761
What be capital of \$location?	administrative_capital → name	1356
What to do in \$location?	tourist_attraction → name	923
What is there to do in \$group?	tourist_attraction → name	746
What is sign of \$disease?	symptom → name	616
Where is \$location located?	contained_by → name	467
What movie is \$person in?	film → name	463
What movie have \$person been in?	film → name	361
What causes \$disease?	risk_factor → name	331

QA pairs as
supervision

-(q,e,v) triples
observed

-(p,t) hidden

-EM optimization

Online

$$P(v|q_0) = \sum_{e,p,t} P(q_0)P(v|e,p)P(p|t)P(t|e,q_0)P(e|q_0)$$

- $P(e|q)$: NER
- $P(t|e,q)$: what is the likely concept of e ?
- $P(v|e,p)$: fact table lookup (or conflict resolution)

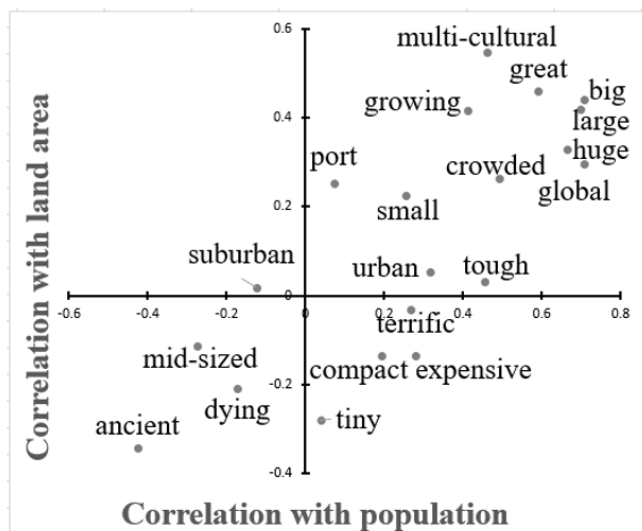
Gradable Adjective Understanding

- “Is **New York** bigger than **San Francisco**?”
- Symbolic
Ex) “big city such as Seoul”, “google is big company”,
“Elephants are larger than dogs”.
- Neural
Ex) skip-gram, GloVe
- Problem : Symbolic → Observation sparsity
Neural → Missing sense of similarity

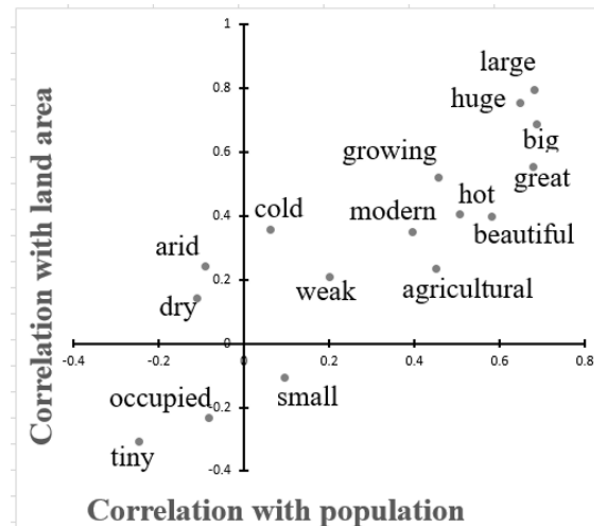
Gradable Adjective Understanding

Correlation between data and graded score

- Large city $\rightarrow \rho$ with population \uparrow
- Large country $\rightarrow \rho$ with land area \uparrow



(a) City



(b) Country

Correlations between adjectives and population/land area

Conclusion

- Internet enables to collect human intelligence.
- Human intelligence feeds AI applications on (learning about / actuating) Things
- For details, visit:
 - <http://dilab.yonsei.ac.kr/~swhwang>