



# **Building a large scale cross lingual knowledge graph**

Juanzi Li

**Knowledge Engineering Group  
Tsinghua University**

2014-12-11

# Outline



- **Knowledge graph basics**
- **The diversity of knowledge graphs**
- **Knowledge graph applications**
- **XIORE: Building a large scale multilingual knowledge graph**
- **Conclusions**



# A Real Experience

Departure	23:00	28 <sup>th</sup> , Sep	Baiyun Airport
Arrival	05:50	29 <sup>th</sup> , Sep	Charles de Gaulle Airport

**Background:** In mid-September, Air France pilots take place over a period of 10 days of mass strikes. Numerous flights are cancelled due to the strikes.

Time Cost: use three days to arrive  
Money Cost: tickets and hotel (about 5000RMB)

Air France strike hits European travel

BY NATALIE HUET, KIRSTI KNOLLE AND TIM HEPHER

Mon Sep 15, 2014 3:40pm EDT

Tweet



Share



Share this



1 OF 2. An Air France Airbus A330 aircraft takes off at Charles-de-Gaulle airport, near Paris, September 14, 2014.

CREDIT: REUTERS/CHRISTIAN HARTMANN

## RELATED VIDEO



Turbulent times for Europe flag carriers

## RELATED NEWS

Air France CEO raises alarm over deadlocked pilot talks

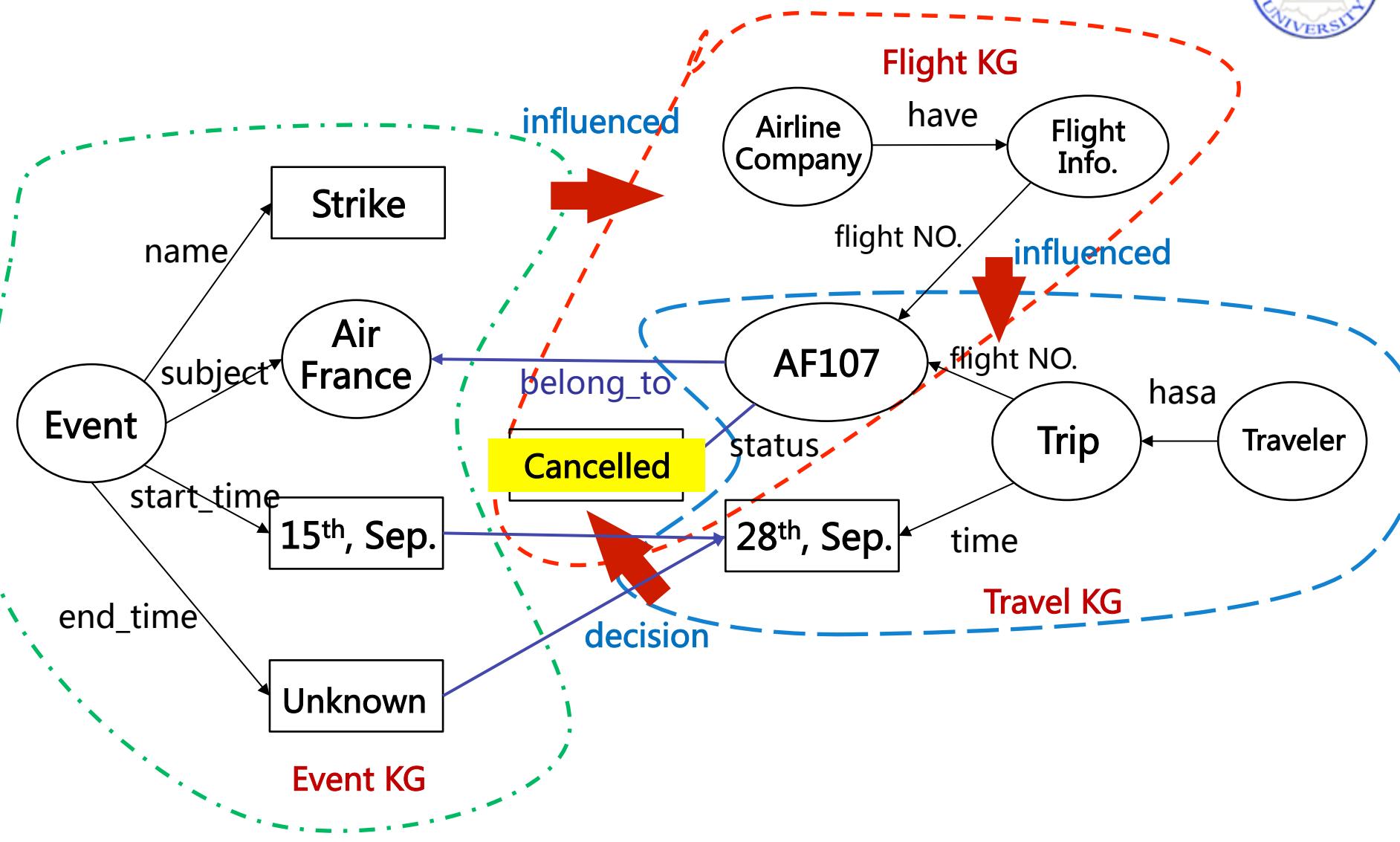
(Reuters) - European travellers face continued disruption as Air France (AIRF.PA) pilots began a week-long strike in a dispute over cost cuts, but a pilots union at the last minute called off a strike planned for Tuesday at Germany's Lufthansa (LHAG.DE).

Air France said it had been forced to cancel six out of 10 flights on Monday and predicted the same level of groundings the next day.

Air France pilots intend to strike until Friday over the carrier's plans to expand the low-cost operations of its Transavia brand by cutting non-fare basis. Their union said they stand ready to

## Air France Strike News

# Semantic Links and Active Information Pushing





# Limitations of Current Web

## ■ Too much information with too little structure and made for **human consumption**

Content search is simplistic, keyword based

- Future requires better search methods: search things, not strings

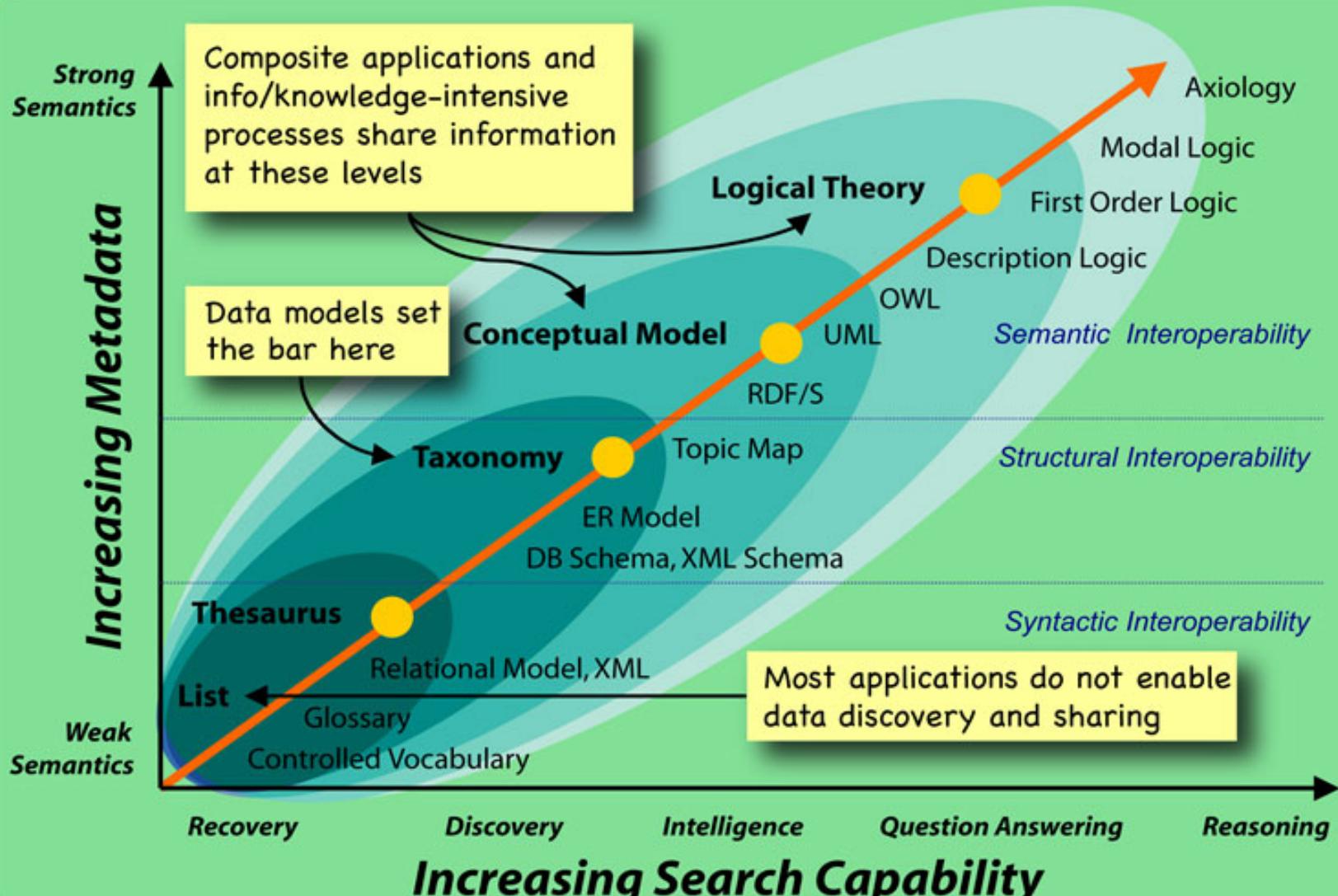
## ■ Web content is heterogeneous both in **content** and in **structure**

- Future requires intelligent information integration: semantic integration

## ■ Humans can derive new (implicit) information from given pieces of information, but on the current Web we can only deal with syntax

- Future requires automated reasoning techniques:
  - $\text{son}(x, y) \& \text{spouse}(y, z) \rightarrow \text{son}(x, z)$

# Content is King or should be!



Source: Dr. Leo Obrst, Mills Davis, Project 10X

# Bring structure to the meaningful content of Web pages

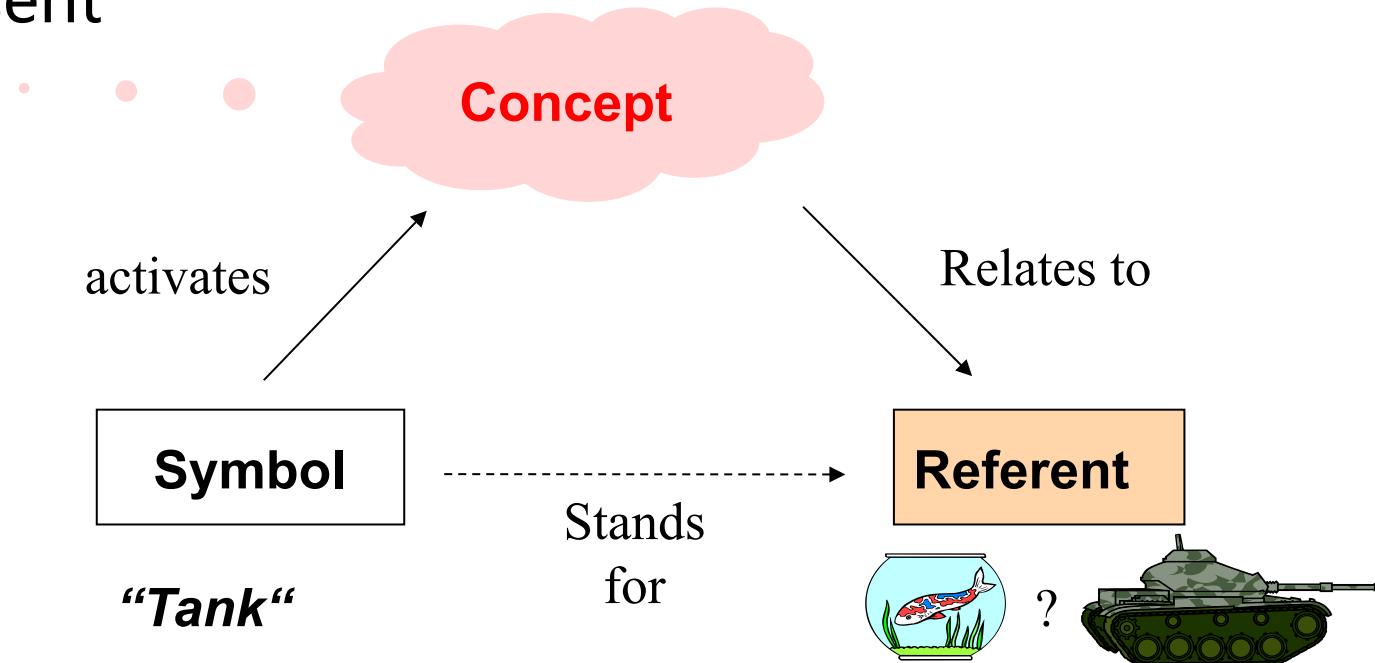


The Semantic Web. Tim Berners-Lee, James Hendler, and Ora Lassila. Scientific American, 2001.

# Triangle of Meaning and Ontology



- ❑ How linguistic symbols are related to the objects they represent



[Ogden, Richards, 1923]

*Ontology is the philosophical study of the nature of **being, becoming, existence, or reality**, as well as the **basic categories** of being and their relations.*

--- Wikipedia

# Ontology in Computer Science



- ▣ An **ontology** is a formal, explicit specification of a shared conceptualization – Gruber 1993
  - **Conceptualization:** an abstract model of phenomena in the world by having identified the relevant concepts of those phenomena.
  - **Explicit:** the type of concepts used, and the constraints on their use are explicitly defined.
  - **Formal:** the fact that the ontology should be machine readable.
  - **Shared:** ontology should capture consensual knowledge accepted by the communities



# Ontology Formalization

## □ 5-tuple Representation $O=\{C,R,F,A,I\}$

### ■ *C - concepts*

- **concepts** of the domain or tasks, which are usually organized in **taxonomies**
- **Football Star, Tsinghua University Alumni**

### ■ *R - relations*

- **interactions** between concepts or instances
- **subClassOf , birth\_place**

### ■ *F - functions*

- a special case of relations in which the n-th element of the relationship is unique for the n-1 preceding elements
- **Price-of-a-used-car** is decided by the **car-model, manufacturing data and kilometers**



# Ontology Formalization

## □ 5-tuple Representation

$$O = \{C, R, F, A, I\}$$

### ■ *A - axioms*

- model sentences that are always true
- If A is a child\_of B, and B is a child\_of C, then A is a grand\_child\_of C

### ■ *I - instances*

- to represent specific things
- e.g.: Student called Peter is the instance of Student class

# Ontology Description

## ❑ RDF

- Res

## ❑ RDF I

- Res

- A
- A

**Future**

- Pro

- A

- Stat

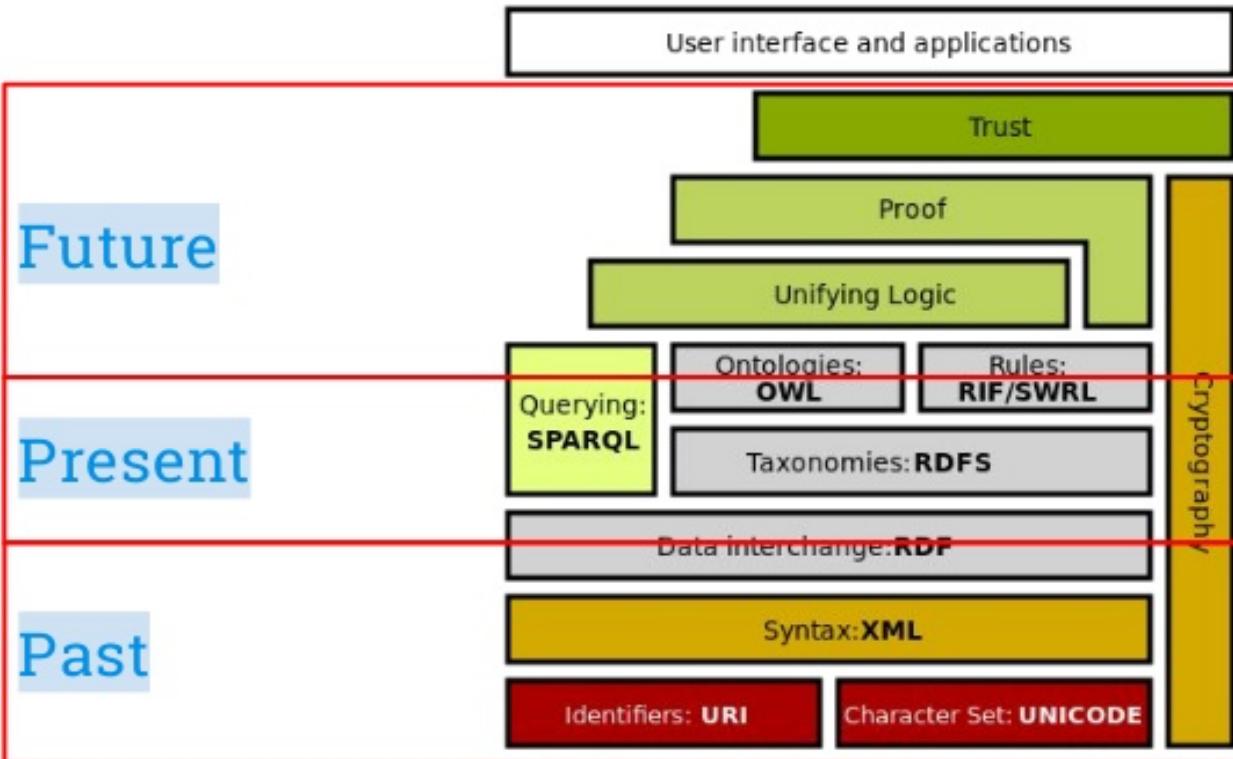
**Present**

- a

- r

## ❑ OWL

**Past**



# Knowledge Graph Definition

## □ Knowledge Graph (KG)

The Knowledge Graph provides **structured and detailed information about the topic** in addition to a list of links to other sites. The goal is that users would be able to use this information to resolve their query without having to navigate to other sites and assemble the information themselves.

--- Wikipedia

## □ Google Knowledge Graph

Graph of entities and their relations.

Over 570 million objects and 18 billion facts.

--- Wikipedia

KG: Knowledge Base? Semantic Network?  
Form of KG: RDF? Graph?

### Leonardo da Vinci

Painter



Leonardo di ser Piero da Vinci was an Italian polymath, painter, sculptor, architect, musician, mathematician, engineer, inventor, anatomist, geologist, cartographer, botanist, and writer. [Wikipedia](#)

**Born:** April 15, 1452, [Vinci, Italy](#)

**Died:** May 2, 1519, [Amboise, France](#)

**Buried:** [Chapel of Saint-Hubert, France](#)

**Periods:** High Renaissance, Pastel Art

**Parents:** Piero Fruosino di Antonio da Vinci, Caterina da Vinci

### Artwork



Mona Lisa  
1517



The Last Supper  
1498



Virgin of the  
Rocks  
1486



Ginevra de'  
Benci  
1478



The Baptism of  
Christ  
1475

[View 25+ more](#)

### People also search for



Michelan...



Raphael



Pablo  
Picasso



Vincent van  
Gogh



Galileo  
Galilei

[View 15+ more](#)

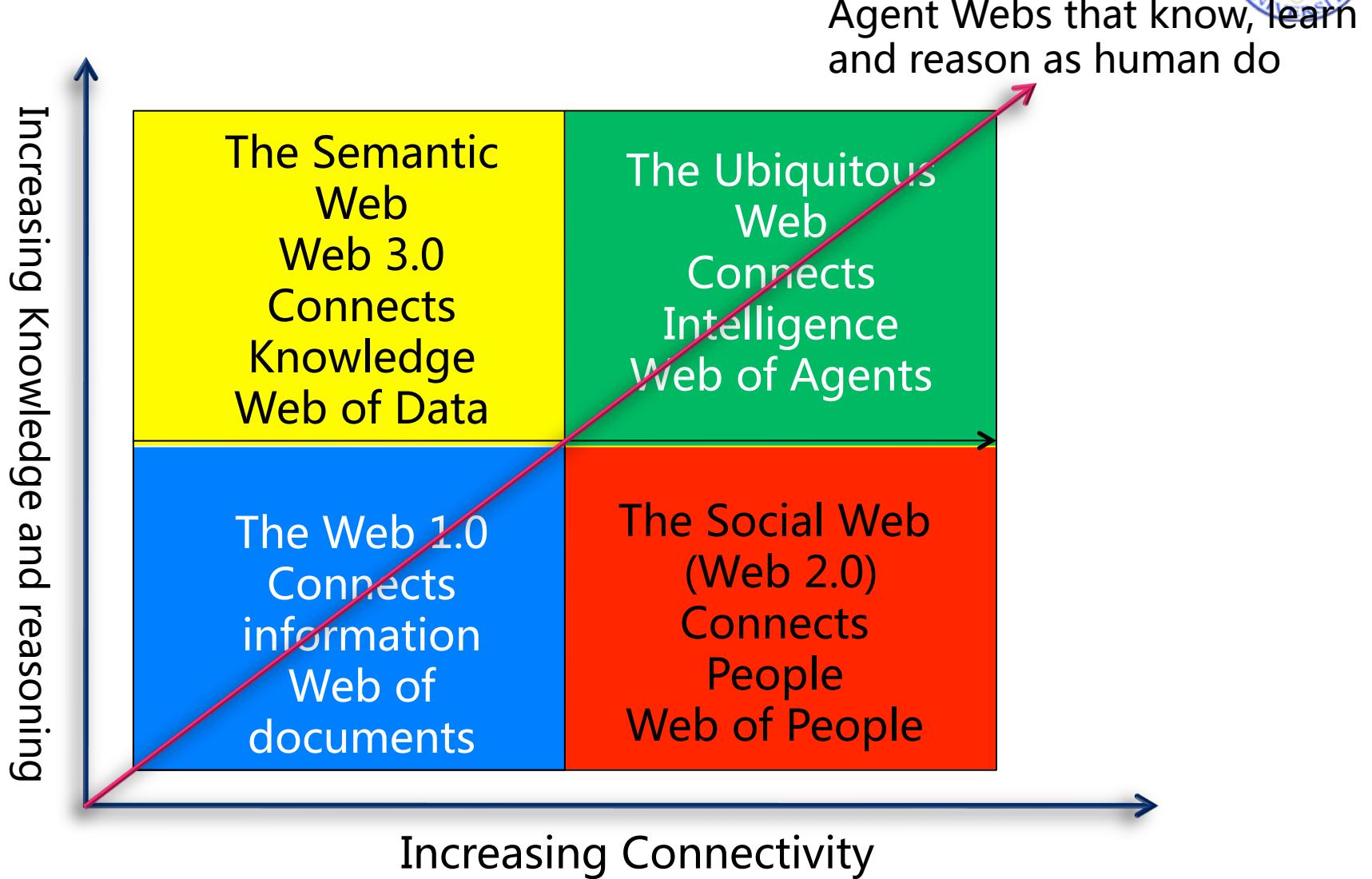




# What's defined in freebase? - Light type system

- **Topic:** one concept or one entity with globally unique ID (Statistics from Freebase: Topics: 46,411,433 Entity Types: 1500 Properties: 35,000 Google Knowledge Graph Entity Types: 1500 Properties: 35,000 Entities: about 570,000,000 zero or more properties, each of that type)
- **Literal:** string, timestamp, or number
- **Type:** property of concepts
- **Property:** a type
- **Schema:** Entity types, zero or more properties, each of that type
- **Domain:** a collection of types which share namespace

# Future of the Web





# Brief Summary

- ❑ Knowledge graph maps objective world knowledge from string description to structural semantic representation.
- ❑ Ontology could be viewed as the conceptual model and basic of knowledge graph.
- ❑ Knowledge graph can describe the concepts in different levels of abstraction and granularities.
- ❑ Knowledge graph is the basic for organizing Internet resources semantically.

While the vision of the Semantic Web has not yet been achieved, the development of the knowledge graph is a good start to make us better understand the information on the Web using world knowledge

# Outline



- **Knowledge graph basics**
- **The diversity of knowledge graphs**
- **Knowledge graph applications**
- **XIORE: Building a large scale multilingual knowledge graph**
- **Conclusions**

# Examples of Knowledge Graph



754 classes  
6M entities  
6000 properties  
500M facts  
live updates



350K classes  
10M entities  
100 relations  
120M facts



15K topics  
40M entities  
4000 properties  
1B facts  
Core of Google KG



A very large multilingual ontology  
50M synsets  
50+ languages  
262M facts

NELL



24M items  
416M edits  
16K users



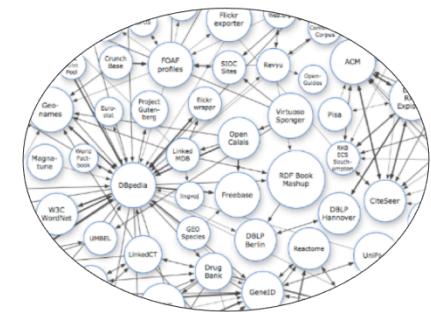
850K concepts  
8M instances  
70K properties

OpenIE  
(Reverb, OLLIE)



Google KG

15K topics  
600M entities  
20B facts



WordNet  
7 euro-languages  
Cross-lingual links



# Domain Independent Knowledge Graphs



Category	Name	Description
Manually Built	ResearchCyc	<a href="http://www.cyc.com/platform/researchcyc">www.cyc.com/platform/researchcyc</a>
	WordNet	<a href="http://wordnet.princeton.edu/">wordnet.princeton.edu/</a>
Wiki-based	DBPedia	<a href="http://wiki.dbpedia.org">wiki.dbpedia.org</a>
	YAGO	<a href="http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/">www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/</a>
	Freebase	<a href="http://developers.google.com/freebase/">developers.google.com/freebase/</a>
	WikiTaxonomy	<a href="http://www.h-its.org/en/research/nlp/wikitaxonomy/">www.h-its.org/en/research/nlp/wikitaxonomy/</a>
	BabelNet	<a href="http://babelnet.org">babelnet.org</a>
	KnowItAll	<a href="http://openie.cs.washington.edu">openie.cs.washington.edu</a>
Open Knowledge Extraction and Machine Learning	NELL	<a href="http://rtw.ml.cmu.edu">rtw.ml.cmu.edu</a>
	Probase	<a href="http://www.microsoft.com/en-us/research/project/probase/">www.microsoft.com/en-us/research/project/probase/</a>
	Baidu ZHIXIN	<a href="http://www.baidu.com">www.baidu.com</a>
Chinese Knowledge Graph	Sogou ZHILIFANG	<a href="http://www.sogou.com">www.sogou.com</a>

# Event Knowledge Graph

## □ Natural Disasters>Earthquake>2010 Haiti Earthquake

### 2010 Haiti earthquake

From Wikipedia, the free encyclopedia

 It has been suggested that [Rome/ Joseph Day](#) be merged into this article. ([Discuss](#)) Proposed since May 2014.  
The 2010 Haiti earthquake was a catastrophic magnitude 7.0 M<sub>w</sub> earthquake, with an epicenter near the town of Léogâne (Ouest Department), approximately 25 kilometres (16 mi) west of Port-au-Prince, Haiti's capital. The earthquake occurred at 16:53 local time (21:53 UTC) on Tuesday, 12 January 2010.<sup>[1]</sup>

By 24 January, at least 52 aftershocks measuring 4.5 or greater had been recorded.<sup>[2]</sup> An estimated three million people were affected by the quake.<sup>[3]</sup> Death toll estimates range from 100,000<sup>[4]</sup> to about 160,000<sup>[5]</sup> to Haitian government figures from 220,000<sup>[1][6]</sup> to 316,000 that have been widely characterized as deliberately inflated by the Haitian government.<sup>[6]</sup> The government of Haiti estimated that 250,000 residences and 30,000 commercial buildings had collapsed or were severely damaged.<sup>[1][6]</sup> There has been a history of national debt, unfair trade policies by other countries, and foreign intervention into national affairs that contributed to the pre-existing poverty and poor housing conditions that exacerbated the death toll.<sup>[1][6]</sup>

The earthquake caused major damage in Port-au-Prince, Jacmel and other settlements in the region. Notable landmark buildings were significantly damaged or destroyed, including the Presidential Palace, the National Assembly building, the Port-au-Prince Cathedral, and the main jail. Among those killed were Archbishop of Port-au-Prince Joseph Serge Miot,<sup>[1][6]</sup> and opposition leader Micha Gaultier.<sup>[1][6]</sup> The headquarters of the United Nations Stabilization Mission in Haiti (MINUSTAH), located in the capital, collapsed, killing many, including the Mission's Chief, Hédi Annabi.<sup>[1][6]</sup>

Many countries responded to appeals for humanitarian aid, pledging funds and dispatching rescue and medical teams, engineers and support personnel. Communication systems, air, land, and sea transport facilities, hospitals, and electrical networks had been damaged by the earthquake, which hampered rescue and aid efforts; confusion over who was in charge, air traffic congestion, and problems with prioritisation of flights further complicated early relief work. Port-au-Prince's morgues were overwhelmed with tens of thousands of bodies. These had to be buried in mass graves.<sup>[2]</sup> As rescues tailed off, supplies, medical care and sanitation became priorities. Delays in aid distribution led to angry appeals from aid workers and survivors, and looting and sporadic violence were observed. On 22 January the United Nations noted that the emergency phase of the relief operation was drawing to a close, and on the following day the Haitian government officially called off the search for survivors.

<b>Contents</b> [hide]
1 Background
2 Geology
2.1 Aftershocks
2.2 Tsunami
3 Damage to infrastructure
3.1 Essential services
3.2 General infrastructure
4 Conditions in the aftermath
5 Casualties
6 Early response
7 Rescue and relief efforts
8 Recovery
8.1 Status of the recovery
8.1.1 2011
8.1.2 2012
8.1.3 2013
9 See also
10 References
11 Further reading
12 External links

topics

infobox

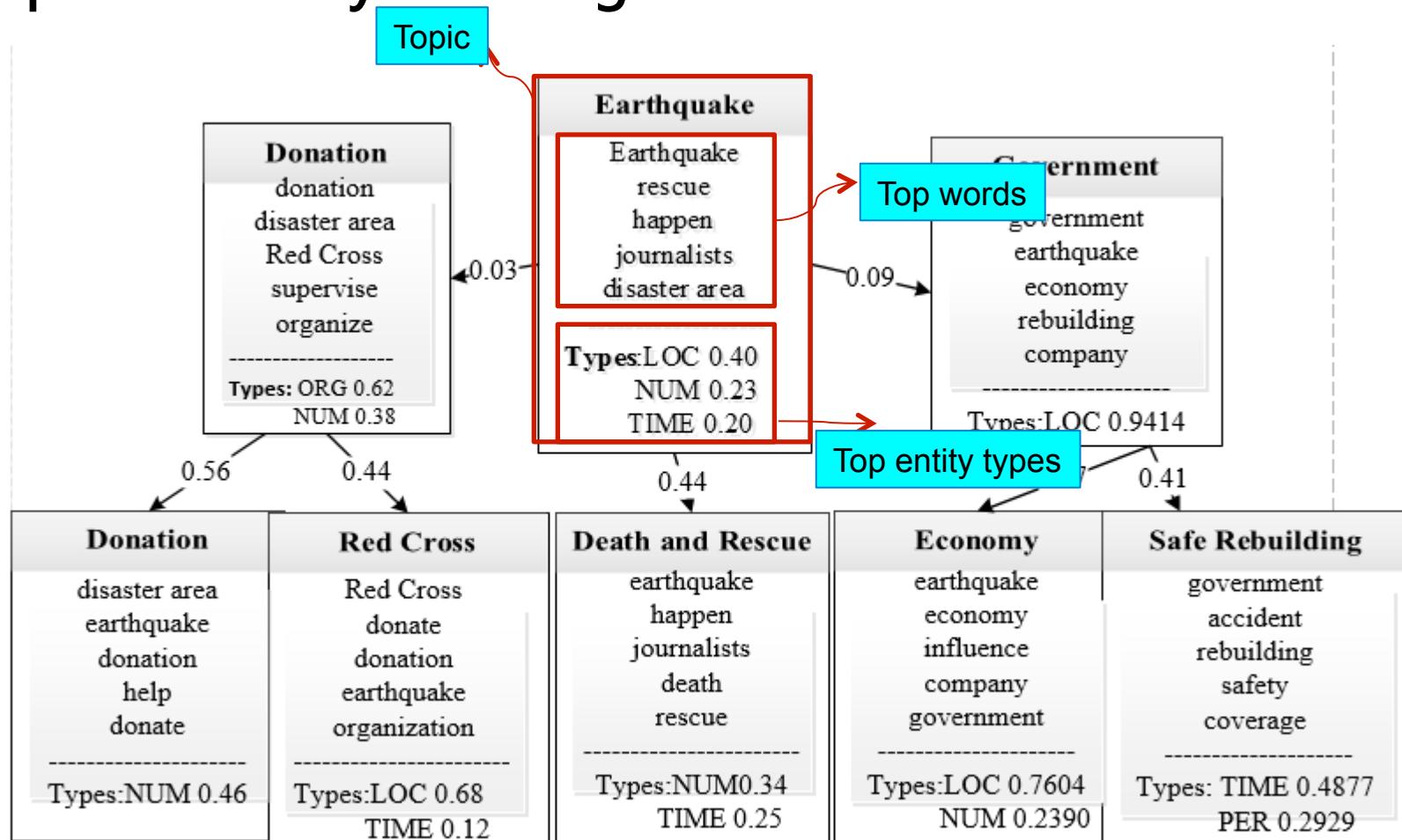
Property

2010 Haiti earthquake	
Date	12 January 2010
Origin time	16:53
Magnitude	7.0 M <sub>w</sub>
Depth	13 km (8.1 mi)
Epicenter	 18°27'25"N 72°31'59"W
Areas affected	Haiti, Dominican Republic
Max. intensity	MM X <sup>[1]</sup>
Peak acceleration	0.5 g <sup>[2]</sup>
Tsunami	Yes (localized) <sup>[3]</sup>
Casualties	100,000 to 316,000 deaths (the higher figure is from a government estimate widely charged with being deliberately inflated; <sup>[4]</sup> a figure of about 160,000 is provided in a 2010 University of Michigan study; <sup>[5]</sup> the 100,000 figure is suggested by the U.S. Geological Survey.) <sup>[6]</sup>

# Event Knowledge Graph



## □ Topic Hierarchy Learning From Similar News Events



# Knowledge Graph for Computation



## □ Computational Knowledge Engine WolframAlpha

**MATHEMATICS**

$2x^4 + 4x^2 - 2x - 3$   
 $x^2 + \frac{1}{x^2} - x -$

Elementary Math · Numbers · Plotting · Algebra · Matrices · Calculus · Geometry · Trigonometry · Discrete Math · Number Theory · Applied Math · Logic · Functions · Definitions · ...

**WORDS & LINGUISTICS**

a1 \_ \_ 1a \_ \_  
calculate  
f a [L<sub>1</sub>] [I<sub>1</sub>] [N<sub>1</sub>] [G<sub>2</sub>] [U<sub>1</sub>] [I<sub>1</sub>]

Spanish: *belleza* (beauty)  
Mandarin: 美 (měi) (beautiful)

Word Properties · Dictionary · Lookup · Word Puzzles · Anagrams · Languages · Document Length · Morse Code · Soundex · Number Names · Character Encodings · ...

**UNITS & MEASURES**

30.093 meters  
 $3.009 \times 10^6$  cm (centimeters)  
98.732 feet  
16.25 nmi (nautical miles)

Conversions · Calculations · Comparisons · Industrial & Construction Analysis · Batteries · Bulk Materials · Paint · Display Formats · Ring Sizes · Shoe Sizes · ...

**STEP-BY-STEP SOLUTIONS** PRO

Move terms with  $x$  to the left hand side.  
Subtract 2 from both sides:  
 $(4x - [2x]) - 6 = (2x - [2x]) + 8$

Next step · Show all steps

**STATISTICS & DATA ANALYSIS**

Descriptive Statistics · Statistical Inference · Regression · Statistical Distributions · Random Variables · Probability · ...

**PEOPLE & HISTORY**

People · Genealogy · Names · Occupations · Political Leaders · Historical Events · Historical Periods · Historical Countries · Historical Numerals · Historical US Money · Inventions · ...

**DATES & TIMES**

Tokyo      Champaign, Illinois  
4:26:37 am JT      2:26:37 pm CDT

Date Computations · Time Zones · Calendars · Holidays · Geological Time · Birthstones · Birth Flowers · Wedding Anniversaries · ...

**CHEMISTRY**

Elements · Compounds · Ions · ...

**CULTURE & MEDIA**

The New York Times  
Paramount Analytics · Natasha Tavite · ...

**plot  $x^3 - 6x^2 + 4x + 12$**

Input interpretation: plot  $x^3 - 6x^2 + 4x + 12$

Plots:

**5" by 7" photo vs 8" by 10" photo**

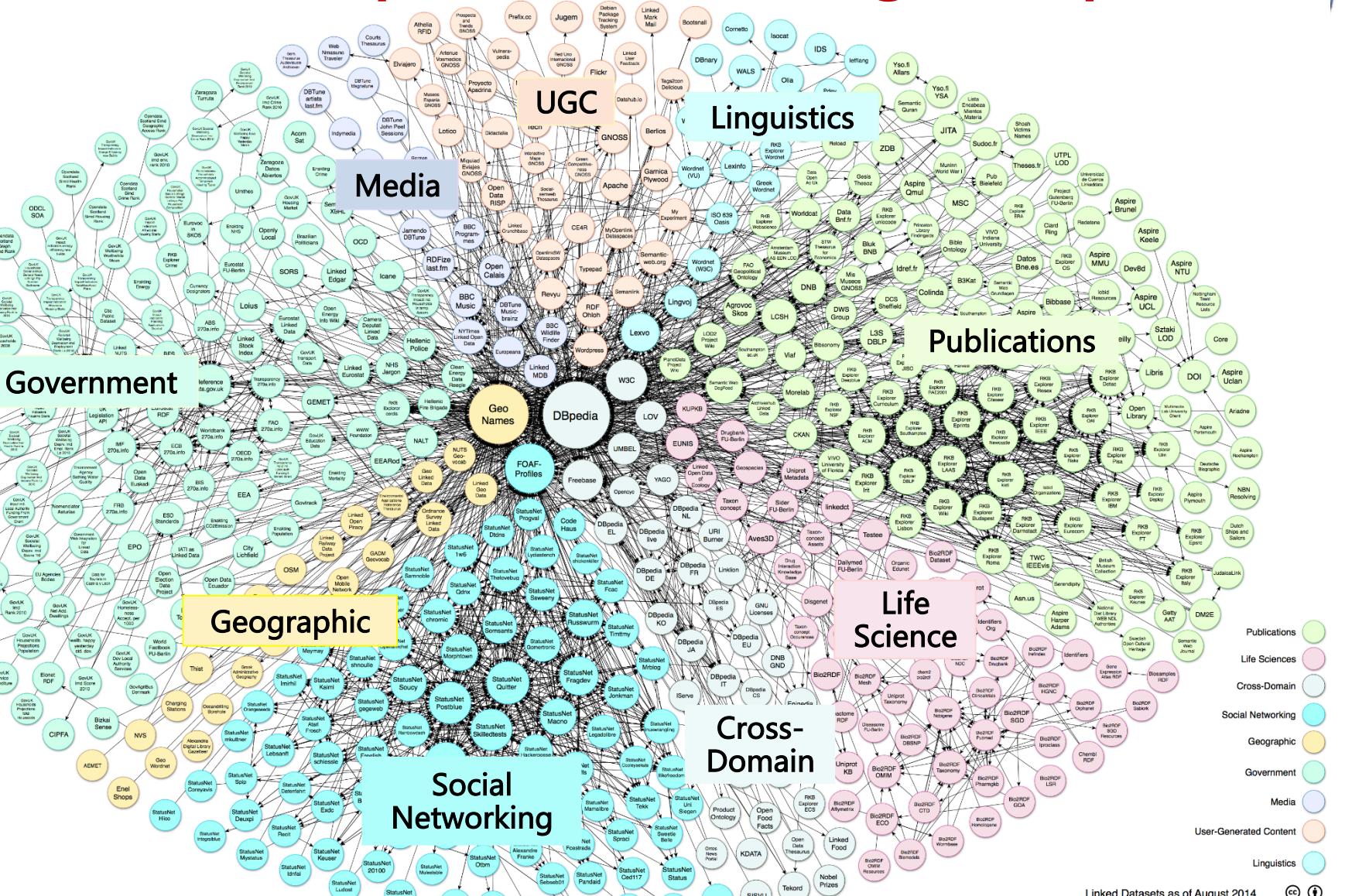
Input interpretation: 5 in × 7 in (display size) | 8 in × 10 in (display size)

Basic properties:

	5 in × 7 in	8 in × 10 in
width	12.7 cm (centimeters)	20.32 cm (centimeters)
height	17.78 cm (centimeters)	25.4 cm (centimeters)
diagonal	21.8 cm (centimeters)	32.5 cm (centimeters)
aspect ratio	5 : 7 (1 : 1.40)	4 : 5 (1 : 1.25)
pixel count at 72 pixels/in	0.1814 megapixels	0.4147 megapixels
pixel count at 300 pixels/in	3.15 megapixels	7.2 megapixels
raw memory at 72 pixels/in	544 kB (kilobytes) (24-bit)	1.24 MB (megabytes) (24-bit)
raw memory at 300 pixels/in	9.45 MB (megabytes) (24-bit)	21.6 MB (megabytes) (24-bit)

<http://www.wolframalpha.com/>

# Domain Specific Knowledge Graph



Linked Datasets as of August 2015

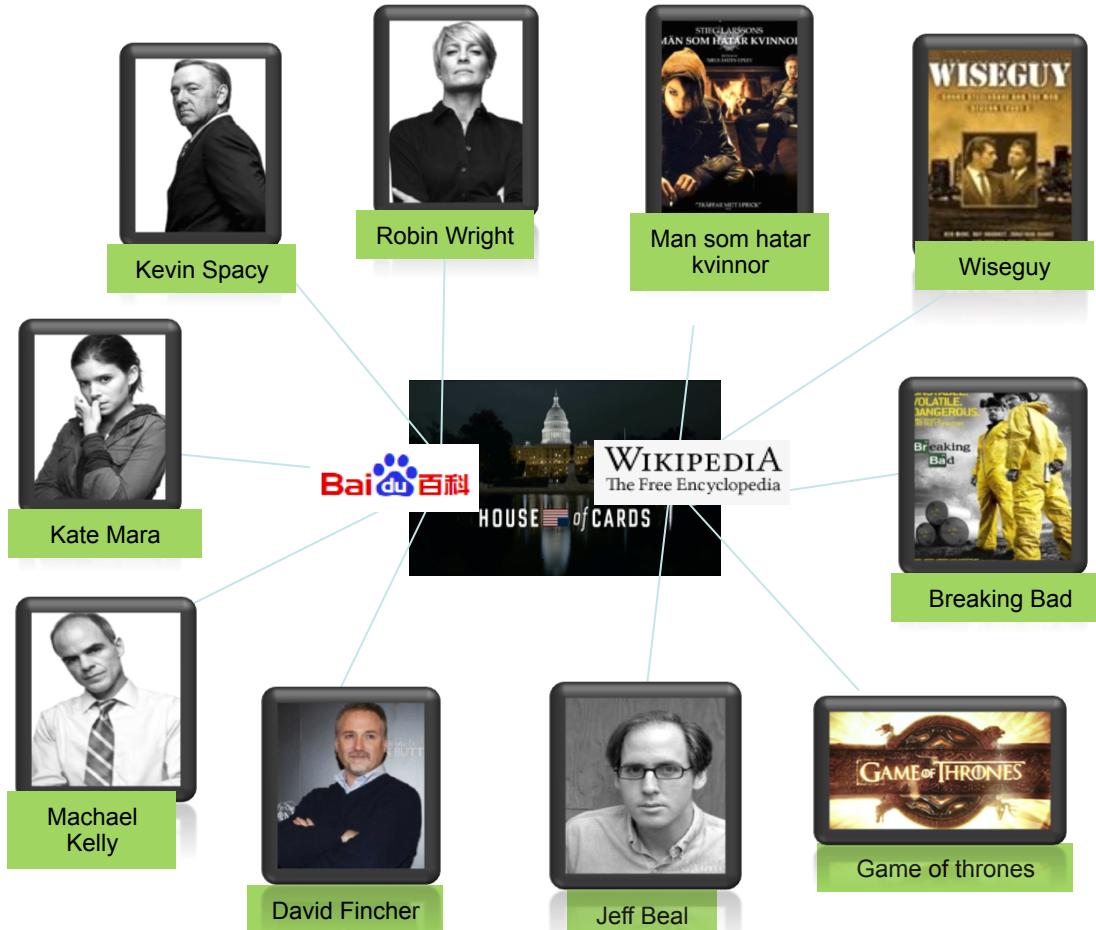


# Movie Domain Knowledge Graph



## Movie Data Integration

### Relation Entities:



### isA Validation:

	Piranha women	isA
	Baidu 百科	Film or TV?
	Freebase	Film
	豆瓣电影	Film

### Bilingual Text Align:

movie:directed\_by  
“**David Fincher**”@en,  
“**大卫·芬奇**”@zh .

movie:genres  
“**Drama**”@en,  
“**剧情**”@zh .

movie:summary  
“**House of Cards** is directed  
by **David Fincher** and acted by  
**Kevin SPacy...**”@en,  
“**《纸牌屋》是由大卫·芬  
奇执导，鲍尔·蒙立威改编，凯文·史  
蒂夫主演...**”@zh .



# Cross-lingual Knowledge Graph



Meaning: actor<sup>1</sup> • ID: bn:00001176n • Type: Concept

**Senses:**

- W English actor<sup>1</sup>, histrion<sup>1</sup>, player<sup>3</sup>, thespian<sup>1</sup>, role player<sup>2</sup>
- W لاعب التئار، مُؤدي، مُمثل، مُمثل هنرجي | joueur, musicien, agent, comédien, acteur, personnage, ηθοποιός, θεατρίνος, שחקן, attore, interprete, istrione, アクター,俳優, 出演者, 名優, 役者, 男優, actor, actor dramático, actuación, artista, comediante, ejecutor teatral, histrionista, histrón, intérprete
- W English Actor, 演員, ممثل, 演員, Acteur, Schauspieler, Ηθοποιός, שחקן, Attore, 俳優, Aktør, Actuación
- W English American actress, Movie actress, Joyu, Synthetic actor, Film actress, Craig Watkinson, Film Actor, Theatre actor, Theatre actress, Stage actress, Actors and Actresses, Stage actor, Television actor, Actors, Actors and actresses, Film actor, Stage actors, Player (dating), American actor, Actresses, Actor & Actress, ممثلة, ممثل, 男演员, 女演员, 戏子, 女演员, 男演员, 演员, Attrice futur, Jeu de l'acteur, Comédien, Comédiens, Attrice, Jeu dramatique, Comédienne, Comedien, Acteurs, Schauspielkunst, Fernsehdarsteller, Schauspielerin, Bühnenschauspieler, Theaterschauspielerin, Theaterschauspieler, Aktrice, Fernsehschauspieler, Schauspieler, شחקن تھاتھوڑا, شوکن کلٹوں, شوکنیت سلولیورا, شوکنیت کلٹوں, شوکنیت سلولیورا, شوکنیت کلٹوں, شوکنیت سلولیورا, Attrice (spettacolo), Attore (spettacolo), Attrice, Attori, 舞台女優, 男優, 實寫俳優, 舞台俳優, 旅役者, 俳優業, バックスター, 役者, 映画俳優, Киноактриса, Киноактер, Киноактриска, Актёр, Киноактер, Актриса, Actuacion, Actriz, Actores, Actor, Actuar, Actor de cine, Actriz
- Q English actor, player, acteur, Schauspielerin, Darsteller, Schauspieler, Darstellerin, attore, 俳優, actriz, actris, actor
- [w] English actor, 演員, ممثل, acteur, actress, Schauspieler, Schauspielerin, Ηθοποιός, Ήθοποιήση, اداگ, شوکن, အမိန္ဒရာ, အမိန္ဒရီ, attore, attrice, 俳優, 役者, 女優, aktør, aktrisa, artist, artistka, actor, actriz
- [[ English actor / actress, 演員, ممثل, 演員, acteur, Schauspieler, Ηθοποιός, شوکن, အမိန္ဒရာ, အမိန္ဒရီ, attore, 俳優, actor, actriz
- Yellow English actress, 美国女演员, 戏剧演员, 电视演员, 演员, 电影演员, , 演员, , 美国演员, , actrice américaine, actrice de cinéma, acteur de théâtre, acteur de télévision, acteurs, acteur, actrices, joueur, acteur américain, us-amerikanische schauspielerin, filmschauspielerin, schauspielerin, theaterschauspieler, fernsehschauspieler, schauspieler, schauspielerinnen, spieler,

演员/Actor

<http://babelnet.org/>

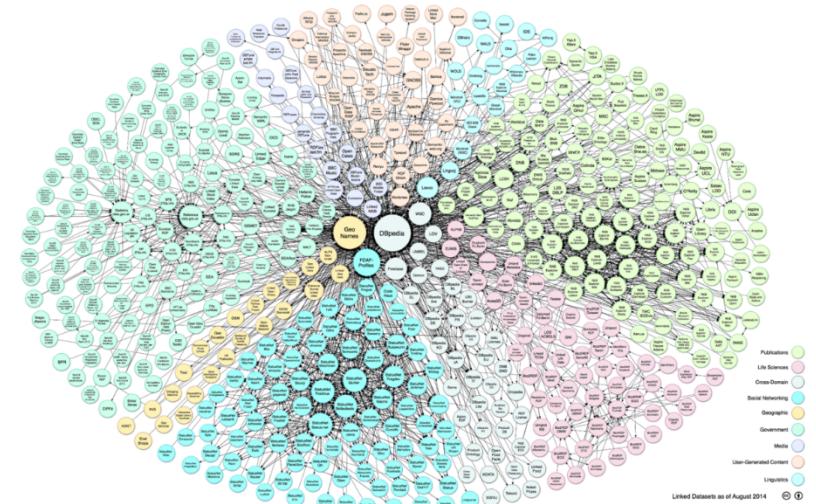


Google KG



# Brief Summary

- ❑ Knowledge graph is the structured representation of processed data.
- ❑ Knowledge graph can represent:
  - Entities and their relations
  - Event knowledge
  - Knowledge for computation
  - Domain independent knowledge
  - Domain specific knowledge
  - Cross-lingual knowledge
  - .....
- Knowledge Graph is the basic of semantic interoperability



# Outline



- Knowledge graph basics
- The diversity of knowledge graphs
- Knowledge graph applications
- XIORE: Building a large scale multilingual knowledge graph
- Conclusions

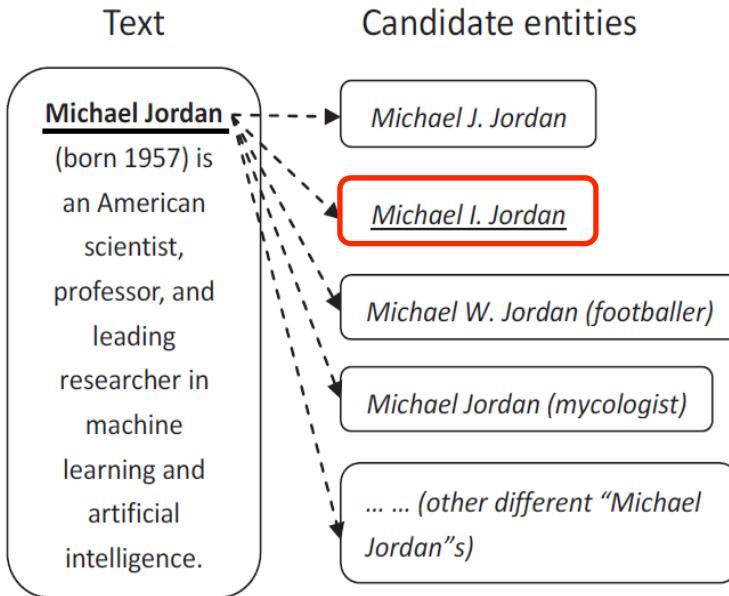
# What does Knowledge graph bring to us?



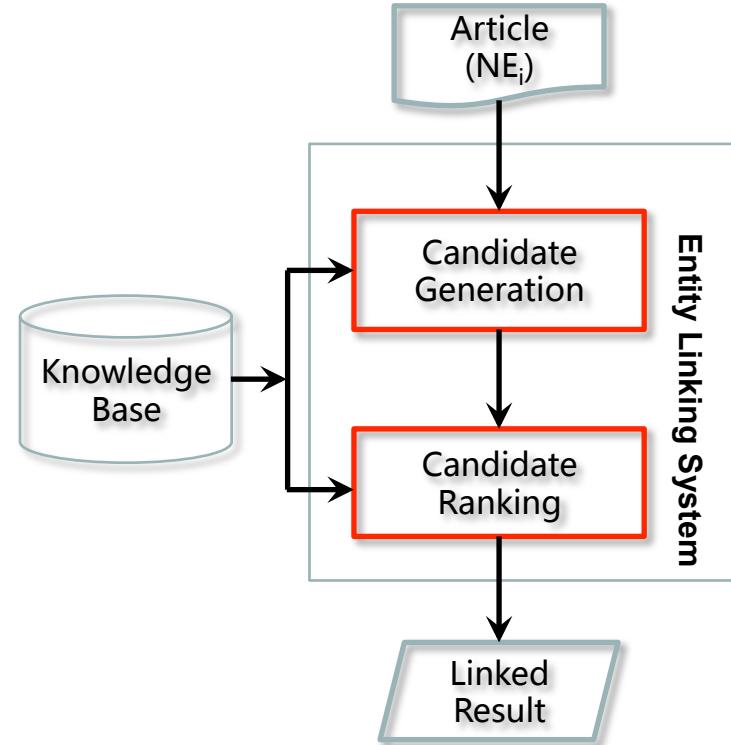
- ❑ Annotated entities and relations-semantic integration
- ❑ Indexing and retrieving entities and relations not just strings-search from strings to entities and relations
- ❑ Question-answering
- ❑ Inferring new facts: learning rules from KG

# Entity Linking

## Problem Definition



## Architecture



Entity linking describes the task of matching each entity mention found in natural language texts (structured, unstructured, long, short) to its corresponding entity in the knowledge base.

# Semantic Integration

## ❑ KG-based Semantic Annotation



### Barack Obama

44th U.S. President

Barack Hussein Obama II is the 44th and current President of the United States, and the first African American to hold the office. [Wikipedia](#)

**Born:** August 4, 1961 (age 53), Honolulu, Hawaii, United States

**Spouse:** Michelle Obama (m. 1992)

**Office:** President of the United States since 2009

**Presidential term:** January 20, 2009 –

**Parents:** Ann Dunham, Barack Obama, Sr.

**Siblings:** Maya Soetoro-Ng, Mark Okoth Obama Nyesandjo, more

### Recent posts on Google+



Barack Obama

4,769,268 followers • Shared publicly

"It's long past time for us to raise the minimum wage." —President Obama Add your name if you agree: [#With1010](http://ofa.bo/b1Do) 10 Oct 2014



### Images



### Texts



Barack Obama

Shared publicly - Oct 11, 2014

#With1010

"It's long past time for us to raise the minimum wage." —President Obama

Add your name if you agree: [#With1010](http://ofa.bo/b1Do)

A higher minimum wage would give a raise to

**28 MILLION AMERICANS**

What would \$10.10 mean for you?

#With1010

+1473

112



30



**Obama says Univision doing "disservi..."**

Washington Times - 5 hours ago

Speaking at a town-hall meeting in Nashville, President Obama said it's "theoretically" possible that the next administration could undo the ...

**Obama defends immigration action, challenges Congress**

USA TODAY - 8 hours ago

**Obama Urges Undocumented Immigrants To Apply For Relief**

Huffington Post - 7 hours ago

**Obama reassures Latinos on immigration action**

CNN - 4 hours ago

**Obama: Americans Won't Support Reversing Immigration Order**

NBCNews.com - 6 hours ago



USA T... CNN Yahoo ... Breitba... The Te... Internat...

**Explore in depth** (505 more articles)



**President Obama Filled in for Stephen Colbert on Th...**

Slate Magazine (blog) - 5 hours ago

Last night, he scored just about the biggest late-night guest you can get—President Barack Obama—but Obama wasn't content to stay on the ...

NEWS.co...

**President Obama took the reins for a hilarious segment on ...**

NEWS.com.au - 5 hours ago

Tongue in cheek, Barack **Obama** displaces comedian Stephen ...

Sydney Morning Herald - 7 hours ago

**President Obama takes command of 'The Colbert Report'**

Los Angeles Times - 7 hours ago

**President Obama Takes Over the Colbert Report**

The White House (blog) - 13 hours ago



YouTube

YouTube

Sydney...

Los An...

The W...

The Gu...

**Explore in depth** (637 more articles)



## Barack Obama

44th U.S. President

Barack Hussein Obama II is the 44th and current President of the United States, and the first African American to hold the office. [Wikipedia](#)

**Born:** August 4, 1961 (age 53), Honolulu, Hawaii, United States

**Full name:** Barack Hussein Obama II

**Spouse:** Michelle Obama (m. 1992)

**Office:** President of the United States since 2009

**Parents:** Ann Dunham, Barack Obama, Sr.

**Siblings:** Maya Soetoro-Ng, Mark Okoth Obama Ndesandjo, more

### Profiles



Facebook



Twitter



Google+



Myspace

### Recent posts on Google+



# Semantic Search

SafeSearch on ▾

Web Images Videos News Maps More ▾ Search tools

About 368,000,000 results (0.27 seconds)

## [House of Cards \(U.S. TV series\) - Wikipedia, the free ...](#)

[en.wikipedia.org/wiki/House\\_of\\_Cards\\_\(U.S.\\_TV\\_series\)](http://en.wikipedia.org/wiki/House_of_Cards_(U.S._TV_series)) ▾

House of Cards is an American political drama television series, developed and produced by Beau Willimon. It is an adaptation of BBC's mini-series of the same ...  
Season 2 - Season 1 - Episodes - House of Cards (UK TV series)

## [House of Cards \(TV Series 2013– \) - IMDb](#)

[www.imdb.com/title/tt1856010/](http://www.imdb.com/title/tt1856010/) ▾

★★★★★ Rating: 9.1/10 - 161,856 votes

House of Cards -- Francis finds himself shut out and Claire makes a difficult sacrifice ·  
House of Cards -- Francis puts China in the cross-hairs. Claire confronts ...

## In the news

### ['House of Cards' Season 3 Premiere Date: Kevin Spacey Returns in Political Drama Scheduled To Air on February 27 ...](#)

BREATHecast - 9 hours ago

Not long ago, Netflix announced that they would be renewing their popular political series, 'House of Cards' for a third season. The news was ...

### ['House of Cards' films season-three finale in NM](#)

Albuquerque Journal - 1 day ago

### ['House Of Cards' Season 3 Premiere Date Revealed, But Where's The Trailer?](#)

iDigitalTimes.com - 18 hours ago

[More news for House of Cards](#)

## [House of Cards | Facebook](#)

<https://www.facebook.com/HouseofCards> ▾

House of Cards. 1408738 likes · 254552 talking about this. Netflix Original Series  
HOUSE OF CARDS. Season 3 premieres February 27, 2015 exclusively on...

## House of Cards

American Television Series

★★★★★ 9.1/10 - IMDb

★★★★★ 8.1/10 - TV.com



U.S. Rep. Francis Underwood of South Carolina starts out as a ruthless politician seeking revenge in this Netflix original production. Promised the post of Secretary of State in exchange for his support, his efforts help to ensure the election of Garrett Walker to the presidency. But Walker changes ... [More](#)

Network: Netflix

Program creator: Beau Willimon

Awards: Golden Globe Award for Best Performance by an Actress In A Television Series - Drama, more

Writers: Beau Willimon, Michael Dobbs, Sarah Treem, Andrew Davies, Rick Cleveland, Kate Barlow, Keith Huff, Sam Forman

Nominations: Primetime Emmy Award for Outstanding Drama Series, more

## Recent posts on Google+



House of Cards

148,135 followers • Shared publicly



A special message from the White House  
1 Dec 2014



# QA Systems

What killed sir edward heath

Web Images News Videos Maps More ▾ Search tools

About 43,300,000 results (0.44 seconds)

## Pneumonia

Edward Heath, Cause of death

### Edward Heath - Wikipedia, the free encyclopedia

[en.wikipedia.org/wiki/Edward\\_Heath](https://en.wikipedia.org/wiki/Edward_Heath) ▾

Jump to [Illness and death](#) - [edit]. Heath's monument in Salisbury Cathedral. In August 2003, at the age of 87, Heath suffered a pulmonary embolism while ...

[Early life](#) - [Second World War](#) - [Post war](#) - [Member of Parliament](#)

### Former PM Sir Edward Heath dies - BBC News

[news.bbc.co.uk/2/hi/4691051.stm](https://news.bbc.co.uk/2/hi/4691051.stm) ▾

Jul 18, 2005 - Former Conservative prime minister Sir Edward Heath has died at the age of 89. His successor Lady Thatcher said he was a "political giant" and "in every sense the first modern Conservative leader".

 WolframAlpha computational knowledge engine

how long is yangzi river



≡ Exam

Input interpretation:

Yangtze length

Share |

Result

6300 km (kilometers)

Unit conversions:

3915 miles

$6.3 \times 10^6$  meters

# Domain Data Analysis based on KG



## □ movie data analysis in Netflix

- Main audience : middle-aged men
- Popular Genres : Political thriller
- Popular director : David Fincher
- Popular Actor : Kevin Spacey
- View Preference : >1 episodes
- Mining relations based on knowledge graph : mining the contributions to the TV by integrating the 3 factors: Kevin Spac, David Fincher and “BBC production”

## □ More precision and feasibility than traditional methods

### House of Cards

American Television Series

★★★★★ 9.1/10 - IMDb

★★★★★ 76% - Metacritic



U.S. Rep. Francis Underwood of South Carolina starts out as a ruthless politician seeking revenge in this Netflix original production. Promised the post of Secretary of State in exchange for his support, his efforts help to ensure the election of Garrett Walker to the presidency. But Walker changes ... [More](#)

Network: Netflix

Program creator: Beau Willimon

Awards: Golden Globe Award for Best Performance by an Actress In A Television Series - Drama, more

Writers: Beau Willimon, Michael Dobbs, Sarah Treem, Andrew Davies, Rick Cleveland, Kate Barnow, Keith Huff, Sam Forman

Nominations: Primetime Emmy Award for Outstanding Drama Series, more

### Recent posts on Google+



House of Cards

148,130 followers • Shared publicly

Follow



A special message from the White House  
1 Dec 2014

### Cast



Kevin Spacey  
Francis Underwood



Robin Wright  
Claire Underwood



Kate Mara  
Zoe Barnes



Corey Stoll  
Rep. Peter Russo



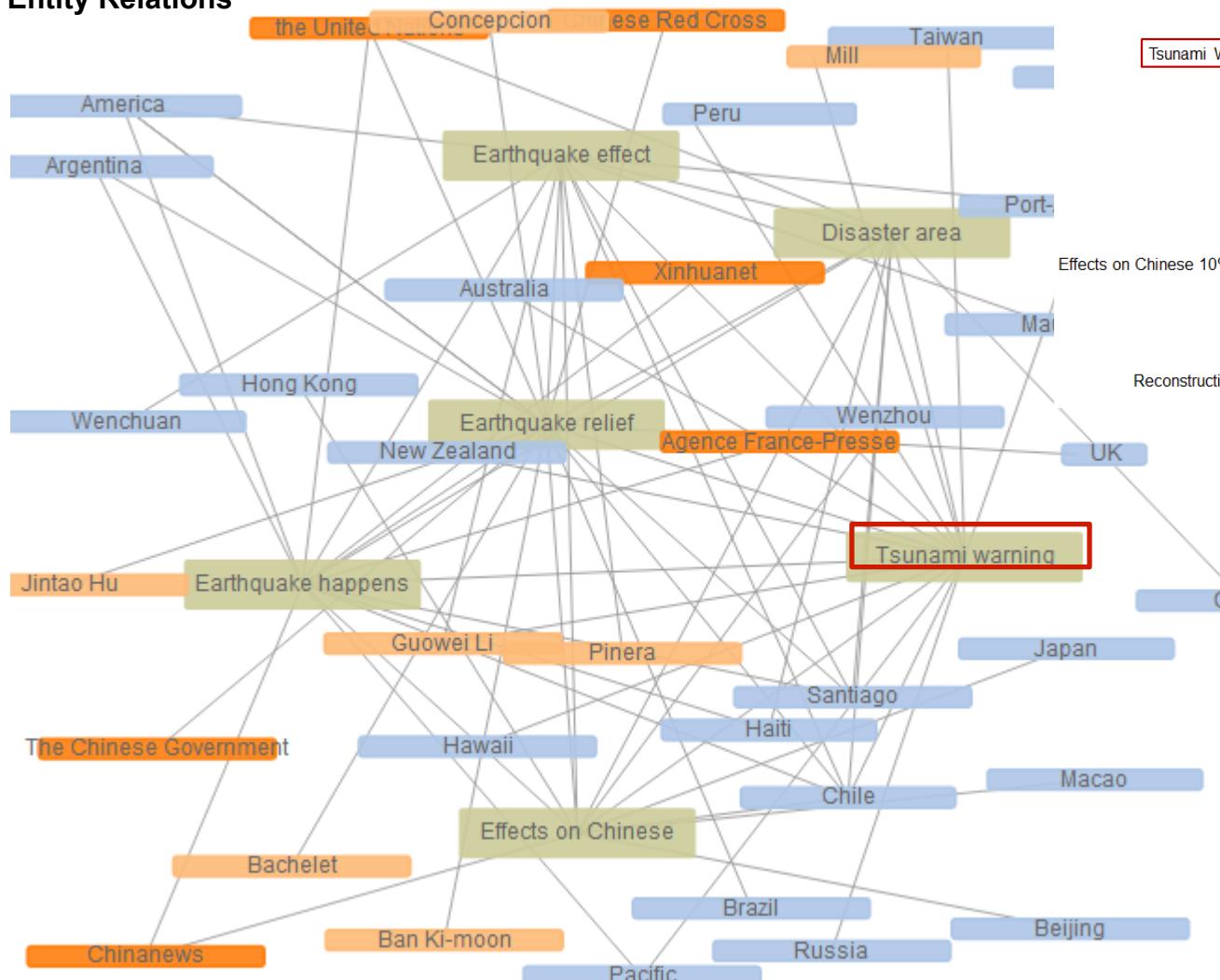
View 15+ more  
Kristen Connolly  
Christina Gallagher

# Knowledge-based News Data Analysis

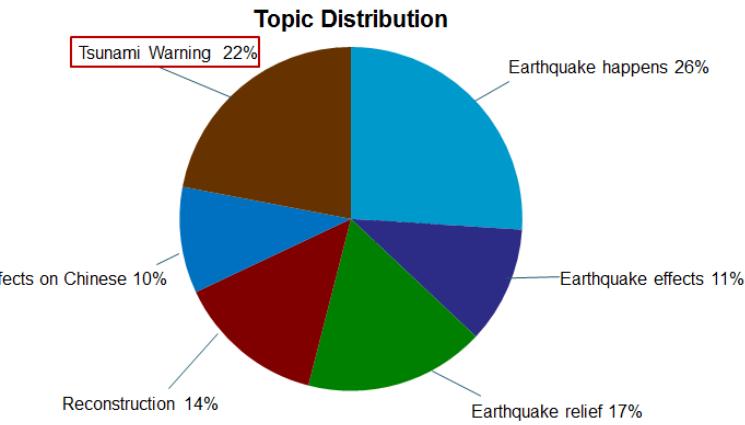


## □ 2010 Chile Earthquake Event

Entity Relations



Topic Distribution





# What else can Knowledge graph do?

- ❑ Knowledge graph adds meaning to the web observatory
- ❑ Gives analysis and interpretation about real world entities and relationships
- ❑ Deep reading: comprehensiveness
- ❑ Knowledge graph can empower intelligent applications

# Outline



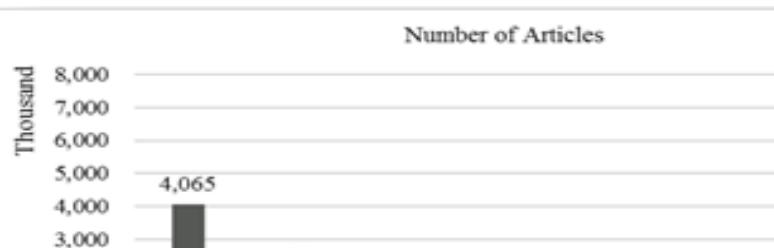
- Knowledge graph basics
- The diversity of knowledge graphs
- Knowledge graph applications
- XIORE: Building a large scale multilingual knowledge graph
- Conclusions



# Motivations

## □ Limitations of existing knowledge graphs

- Imbalance of knowledge in different languages
  - Statistics from Wikipedia
- Low integration of multilingual knowledge
  - Only less than 5% English articles have inter language links in Wikipedia to Chinese articles
- Lack of research on the knowledge complement of different languages
  - Making use of the knowledge in other languages



A balanced, integrated multilingual knowledge graph is essential for intelligent multilingual information processing and services

# Chinese Wiki Knowledge Bases

## ■ Wikipedia

- 4,951,692 articles in English
- 837,596 articles in Chinese



## ■ Hudong Baike

- 13,301,945 articles
- 9,774,730 contribution users



## ■ Baidu Baike

- 12,341,484 articles
- 5,558,289 contribution users



# Motivations

## ■ Multilingual knowledge graph: Multilingual Entities & Relationships

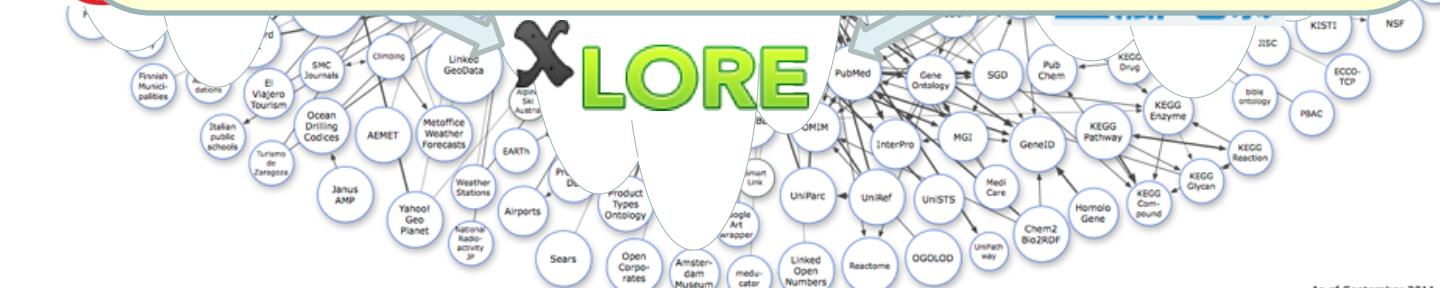


- Enhance the linked data internationalization and the globalization of knowledge sharing among different languages on the Web
- Facilitate the cross-lingual language processing such as cross-lingual information retrieval, machine translation and Q&A etc



# Our objectives

- Cross-lingual knowledge linking
  - Cross-lingual knowledge extraction
  - Cross-lingual Taxonomy Derivation
  - Cross lingual knowledge graph-XLore



As of September 2011 



# CROSS LINGUAL KNOWLEDGE LINKING

Zhichun Wang, Juanzi Li, Zhigang Wang, Jie Tang

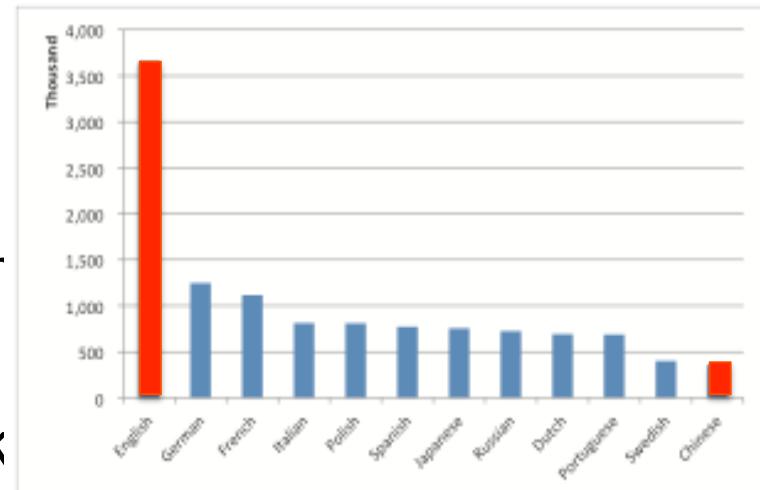
WWW'2012, IJCAI'2013

# Motivation

- Articles in different languages are highly unbalanced within Wikipedia

- 4,096,240 English articles
- 602,972 Chinese articles
- 218 thousand Cross lingual lin (October 2011) between Ch-Er

- Find missing crosslingual link
  - [Sorg 2008] German-English
  - [Oh 2008] Japanese-English



It is infeasible to create cross-lingual links with sufficient coverage by **only using Wikipedia data**.



# Knowledge linking

## □ Wiki knowledge bases

- Baidu Baike, 12,341,484 Chinese articles
  - <http://baike.baidu.com/>
- Hudong, 13,301,945 Chinese articles
  - <http://www.hudong.com/>



- The problem:
  - Given two knowledge bases  $K_1$  and  $K_2$
  - For each  $a_i \in K_1$ , find  $a_j \in K_2$  such that
$$a_i \Leftrightarrow a_j$$
  - Crosslingual knowledge linking
    - $K_1$  and  $K_2$  are in different languages

# Knowledge Linking

**Titles**

**Links**

**Categories**

**Authors**

**Cross lingual links**

The diagram illustrates the concept of Knowledge Linking across different platforms. It shows how entities from one platform can be linked to their counterparts on another. The entities are represented by colored boxes: red for titles, blue for links, light blue for categories, yellow for authors, and green for cross-lingual links.

**Wikipedia (Red Box):** Shows the title "Anaerobic exercise" and its definition. A red box labeled "Titles" highlights the title itself. A green arrow points from the "anaerobic metabolism" section to the Baidu page, indicating a cross-lingual link.

**Baidu (Blue Box):** Shows the title "无氧运动" (Anerobic Exercise) and its definition. A blue box labeled "Titles" highlights the title itself. A green arrow points from the "Fox and Haskell formula" chart to the Wikipedia page, indicating a cross-lingual link.

**Wikidata (Light Blue Box):** Shows the category "Exercise physiology". A light blue box labeled "Categories" highlights the category itself. A green arrow points from the "Fox and Haskell formula" chart to the Wikidata page, indicating a cross-lingual link.

**Authors (Yellow Box):** Shows the list of contributors for the Wikipedia page. A yellow box labeled "Authors" highlights the contributor list. A green arrow points from the "Fox and Haskell formula" chart to the Authors section, indicating a cross-lingual link.

**Cross Lingual Links (Green Box):** Shows the "无氧运动" entry on Baidu. A green box labeled "Cross lingual links" highlights the entry. A green arrow points from the "Fox and Haskell formula" chart to the Baidu page, indicating a cross-lingual link.

# Linkage Factor Graph

## Constraint factor

Linking cardinalities constraint

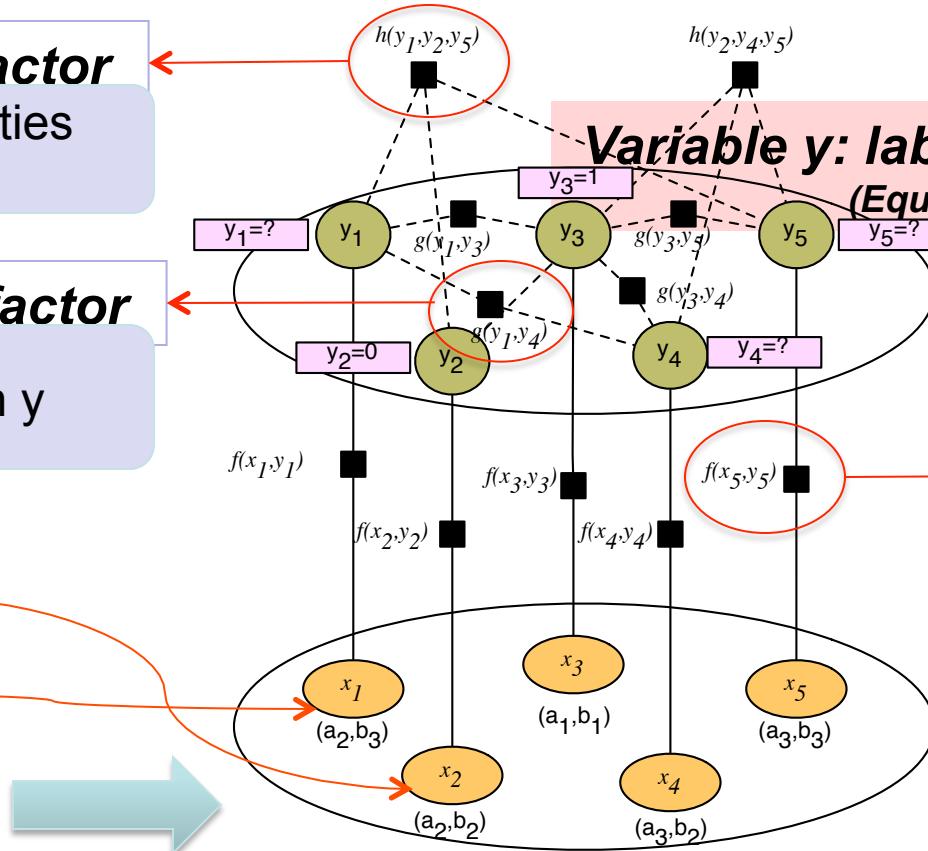
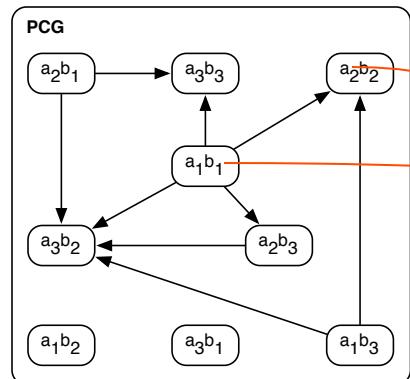
## Correlation factor

relation between  $y$

**Variable  $y$ : label of article pair  
(Equivalent or Inequivalent)**

## Attribute factor

posterior probability of  $y$  given  $x$

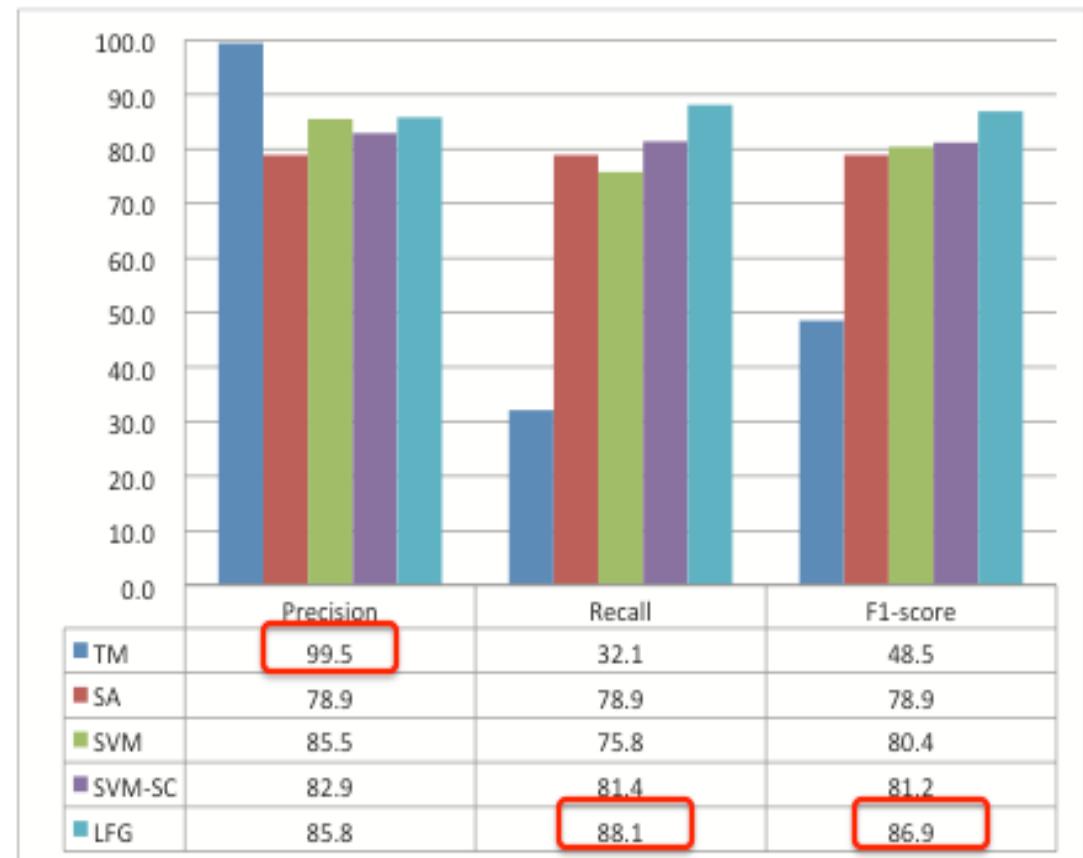


**Variable  $x$ : attribute vector of article pair**

# Experiments

## □ Results ( 3-fold cross validation )

- TM gets high precision but low recall ;
- SVM gets better precision while SVM-SC gets better recall, SVM-SC outperforms SVM by 0.8% in terms of F1-score.
- LFG achieves the best recall and F1-score among all these methods.

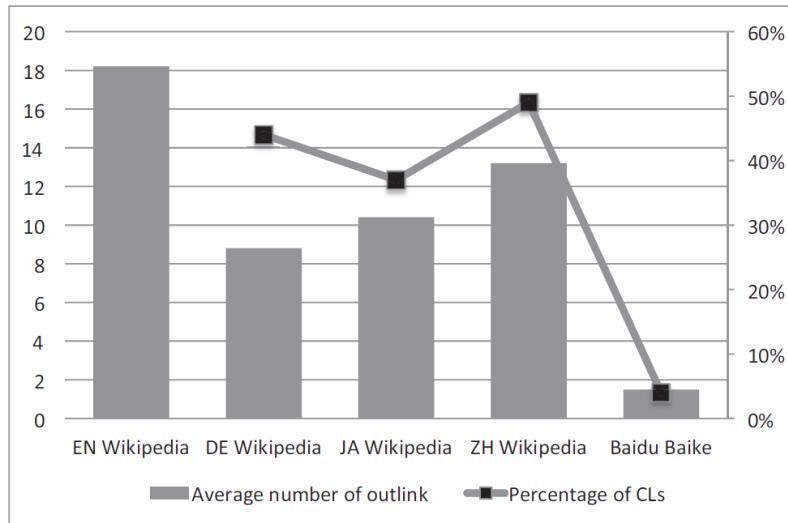


# Boosting Knowledge Linking via Link Annotation



## □ Background

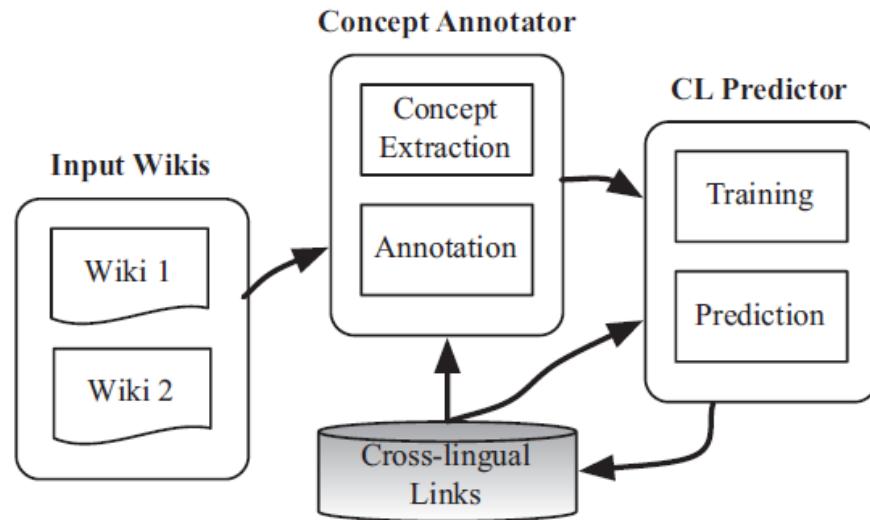
- LFG method relies on the number of links
- The discovered cross-lingual links are limited in one iteration



# Boosting Knowledge Linking via Link Annotation

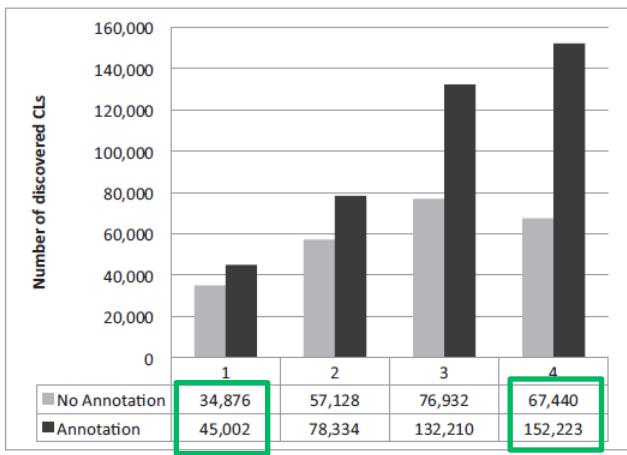


## ■ Framework



## ■ Experiment Result

#Seed CLs	Model	Before Annotation			After Annotation		
		Precision	Recall	F1-measure	Precision	Recall	F1-measure
0.05 Mil. CLs	SVM	92.1	35.0	50.7	78.5	37.2	50.5
	RM	93.3	36.0	<b>52.0</b>	92.4	38.6	<b>54.5</b>
0.10 Mil. CLs	SVM	79.7	35.0	48.6	86.9	50.4	63.8
	RM	84.6	37.4	<b>51.9</b>	96.6	49.3	<b>65.3</b>
0.15 Mil. CLs	SVM	80.9	35.9	49.7	88.1	57.3	69.5
	RM	93.5	38.2	<b>54.2</b>	93.7	56.2	<b>70.2</b>
0.20 Mil. CLs	SVM	84.7	37.3	51.8	88.8	68.1	77.1
	RM	94.5	37.9	<b>54.1</b>	95.9	67.2	<b>79.0</b>





# CROSS LINGUAL KNOWLEDGE EXTRACTION

Zhigang Wang, Zhixing Li, Juanzi LI, Jie Tang

ACL'2013, ISWC'2013

# Motivation

## □ Status of infoboxes

- Missing

- Lots of wiki articles do not have an infobox

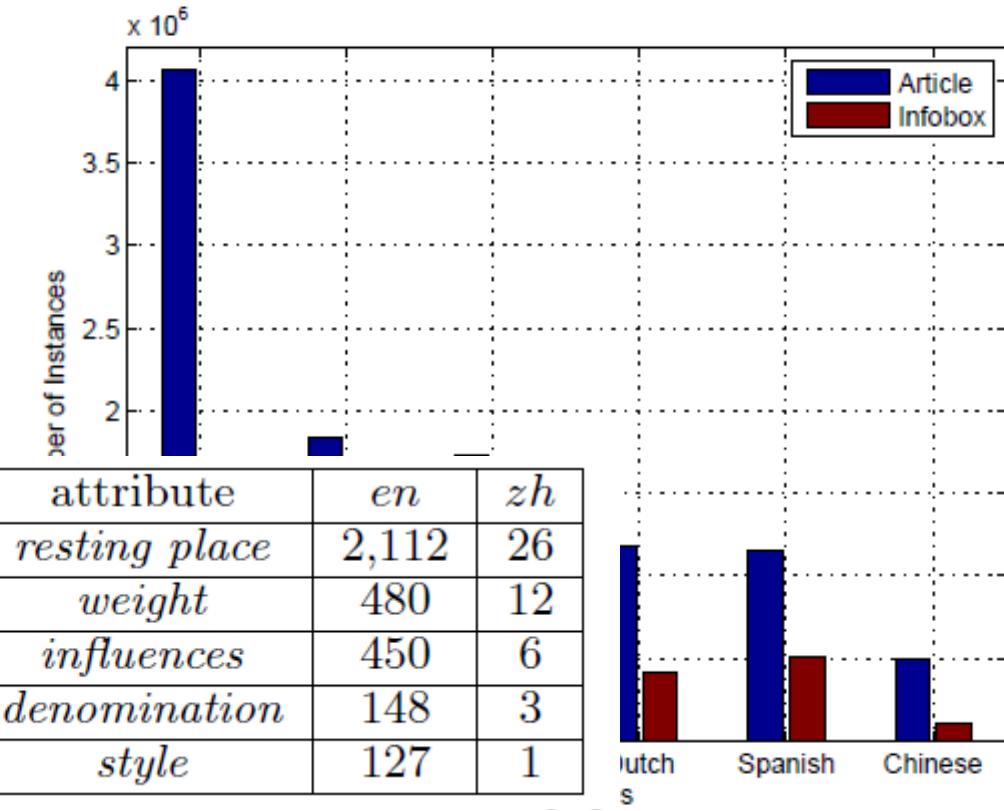
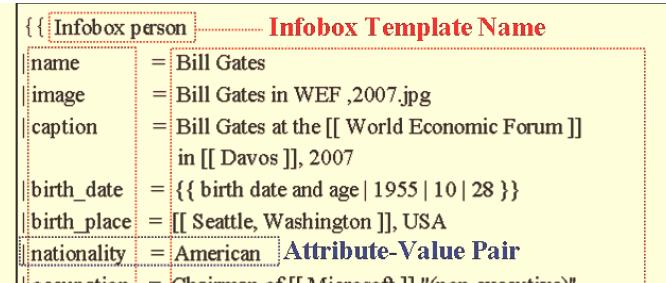
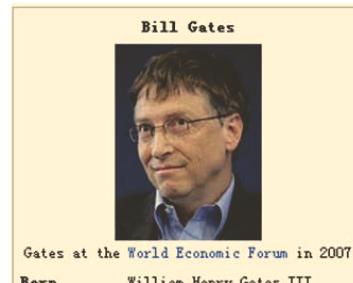
- Incomplete

- Lots of infoboxes have only a few attribute-value Pairs

## □ Wikipedia Multi-linguality

- High skewed between different versions

attribute	<i>en</i>	<i>zh</i>
<i>name</i>	820,99	1,486
<i>birth date</i>	77,850	1,481
<i>occupation</i>	66,768	1,279
<i>nationality</i>	20,048	730
<i>awards</i>	2,310	38



# Related Work

## ■ Extract directly from the article texts.

- DBPedia, Fei Wu's work
- Sparseness of training data, Fei Wu's work

## ■ Translation-based method.

- Limited

Table 1: Functionality of Translation Based Method

- Equivalent value |

Has infobox	Has cl	cl has infobox	#Article	Translation Method
Y	Y	Y	43,158	Y
Y	Y	N	11,360	N
Y	N	-	52,323	N
N	Y	Y	44,445	Y
N	Y	N	117,444	N
N	N	-	160,047	N

ding attribute-

*Has infobox*: this Chinese article has an infobox

*Has cl*: this Chinese article has an cross-lingual link

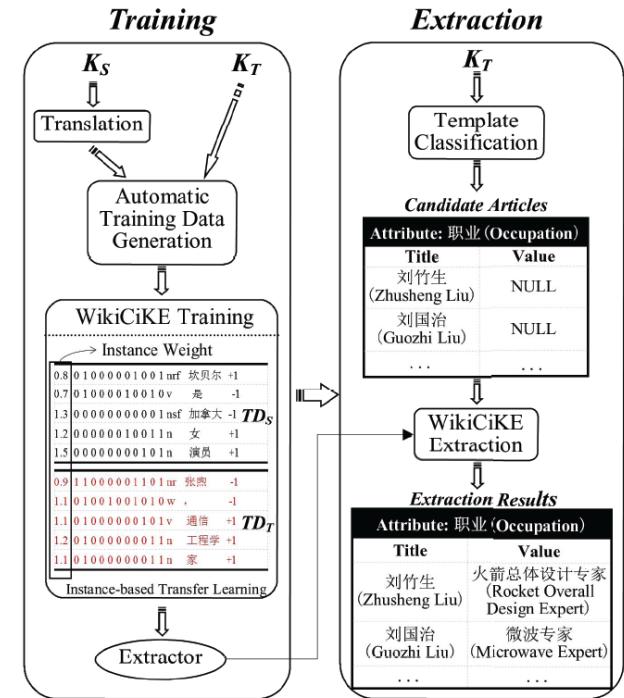
*cl has infobox*: Cross-lingual linked English article has an infobox

*Translation Method*: Y - works; N - fails

# Cross-lingual Knowledge Extraction

## ■ Approach

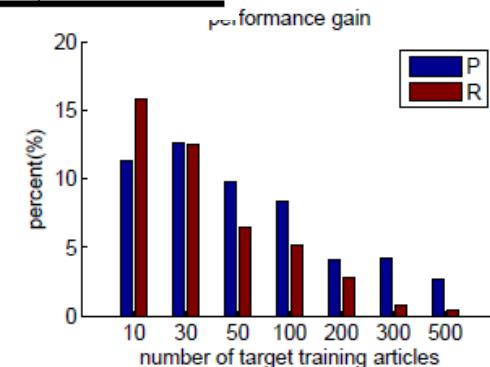
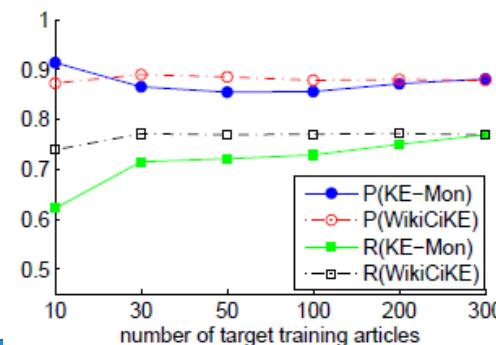
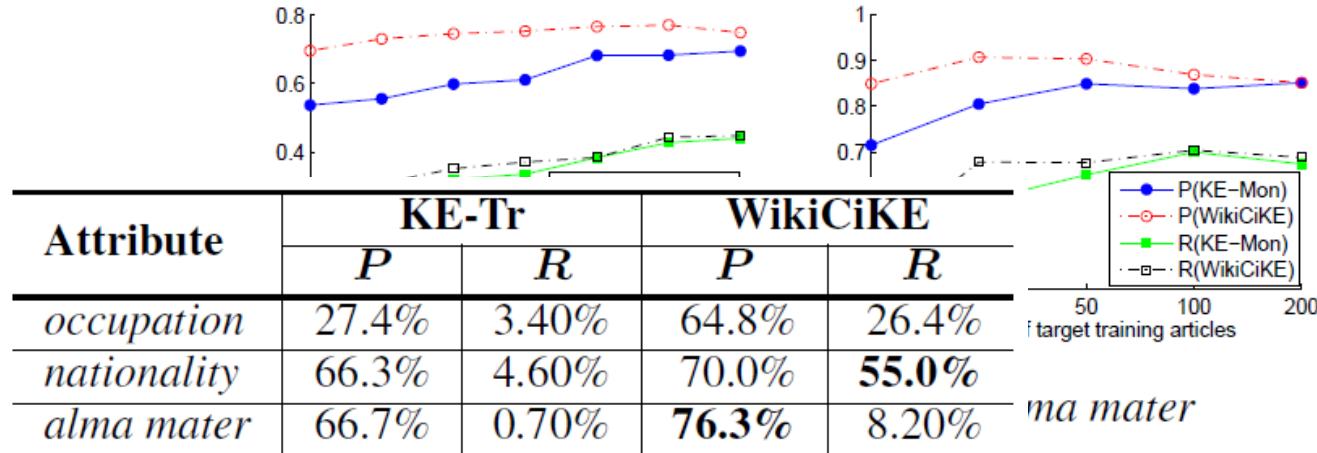
- Train an extractor for each attribute
- Automatically generate the training data
  - Chinese – Target
  - English – Source/Auxiliary
- Transfer learning-based extractor training
  - SVM
  - TrAdaBoost
- Extract the missing value from the texts



# Cross-lingual Knowledge Extraction

## ■ Experiments

- Comparison with mono-lingual extraction method
- Comparison with translation-based method





# CROSS-LINGUAL KNOWLEDGE VALIDATION BASED TAXONOMY DERIVATION FROM HETEROGENEOUS ONLINE WIKIS

Zhigang Wang Juanzi Li Shuangjie Li, Jie Tang

AAAI'2014

# Cross-lingual Knowledge Validation Based Taxonomy Derivation

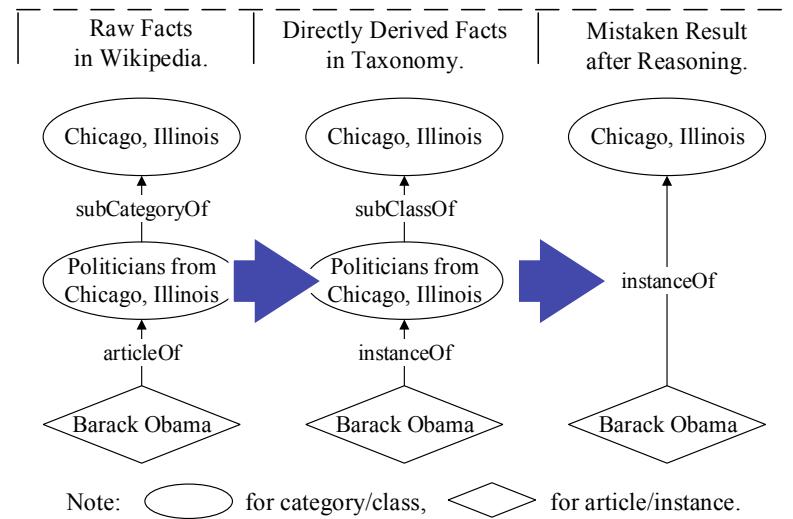


## □ Background

- Wikipedia's category system is important
- The user-generated subsumption relations are not exactly isA relations

## □ Existing methods

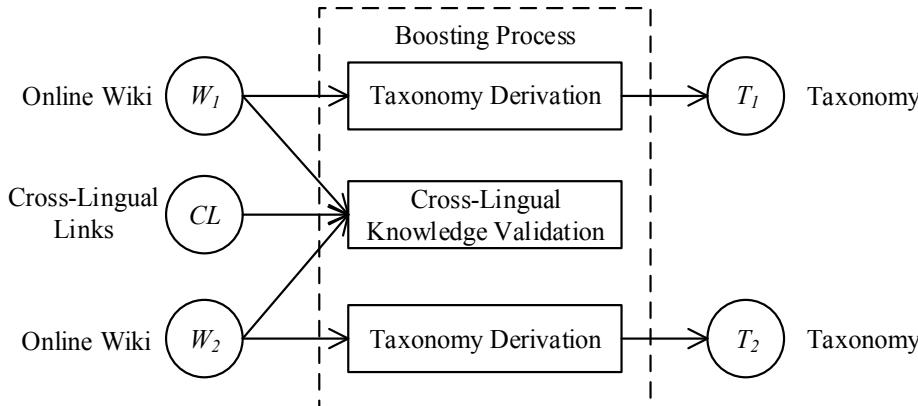
- The heuristic-based methods
  - Language dependent features, e.g. singular/plural forms of head words
- The corpus-based methods
  - Rely on a high-quality large-scale corpus
  - The corpus are usually domain dependent



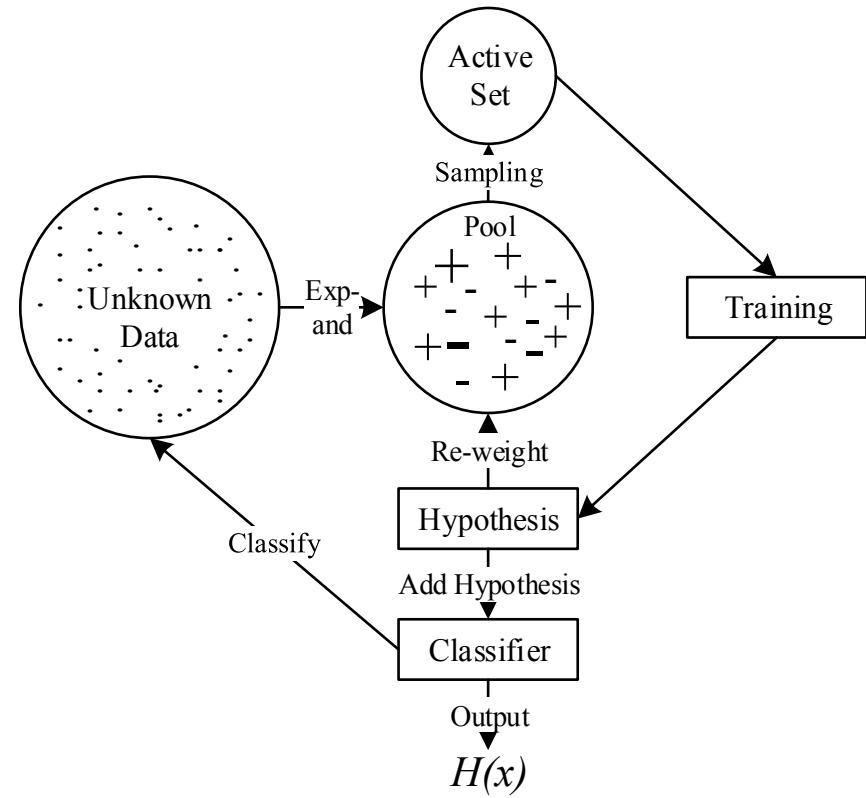
# Cross-lingual Knowledge Validation Based Taxonomy Derivation



## ❑ Framework



## ❑ Dynamic Adaptive Boosting Model



## ❑ Weak Classifier

- Linguistic Heuristic Features
  - English/Chinese Features based on rdfs:label
  - Common Features based on rdfs:comment
- Structural Features
  - Normalized Google Distance
  - Concept, Instance, Property

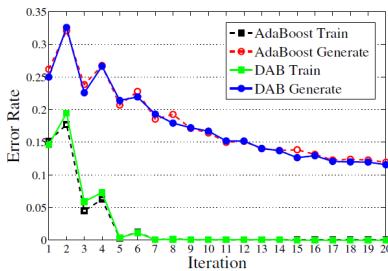
# Cross-lingual Knowledge Validation Based Taxonomy Derivation



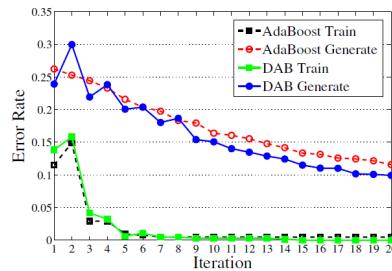
## □ Experimental Results

Table 2: Performance of Cross-lingual Taxonomy Derivation with Different Methods (%)

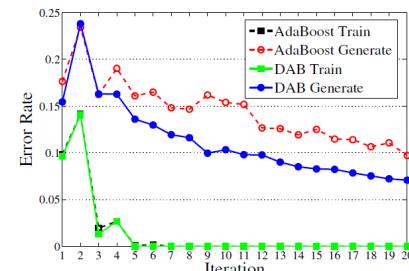
Methods	English SubClassOf			Chinese SubClassOf			English InstanceOf			Chinese InstanceOf		
	P	R	F1									
HL	87.1	81.3	84.1	91.4	91.4	91.4	94.3	89.4	91.8	42.4	51.9	46.7
DT	88.7	86.9	87.8	90.9	92.0	91.4	91.9	95.6	93.7	46.8	58.1	51.8
AdaBoost	90.8	90.9	90.9	91.4	92.3	91.8	94.3	94.1	94.2	51.4	63.9	57.0
DAB	90.7	91.8	<b>91.2</b>	91.1	95.2	<b>93.1</b>	94.1	97.7	<b>95.9</b>	77.8	75.0	<b>76.4</b>



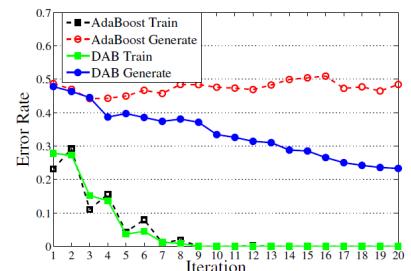
(a) English subClassOf



(b) Chinese subClassOf



(c) English instanceOf



(d) Chinese instanceOf

Boosting Contribution Analysis.

**AAAI2014:** Cross-lingual Knowledge Validation Based Taxonomy Derivation from Heterogeneous Online Wikis.

Zhigang Wang, Juanzi Li, Shuangjie Li, Mingyang Li, Jie Tang, Kuo Zhang, and Kun Zhang.



# SYSTEM AND APPLICATIONS

Zhigang Wang, Mingyang Li, Peng Zhang, Zhixing Li, Juanzi LI



# XLPRE: English-Chinese Knowledge Base

## ■ XLORE

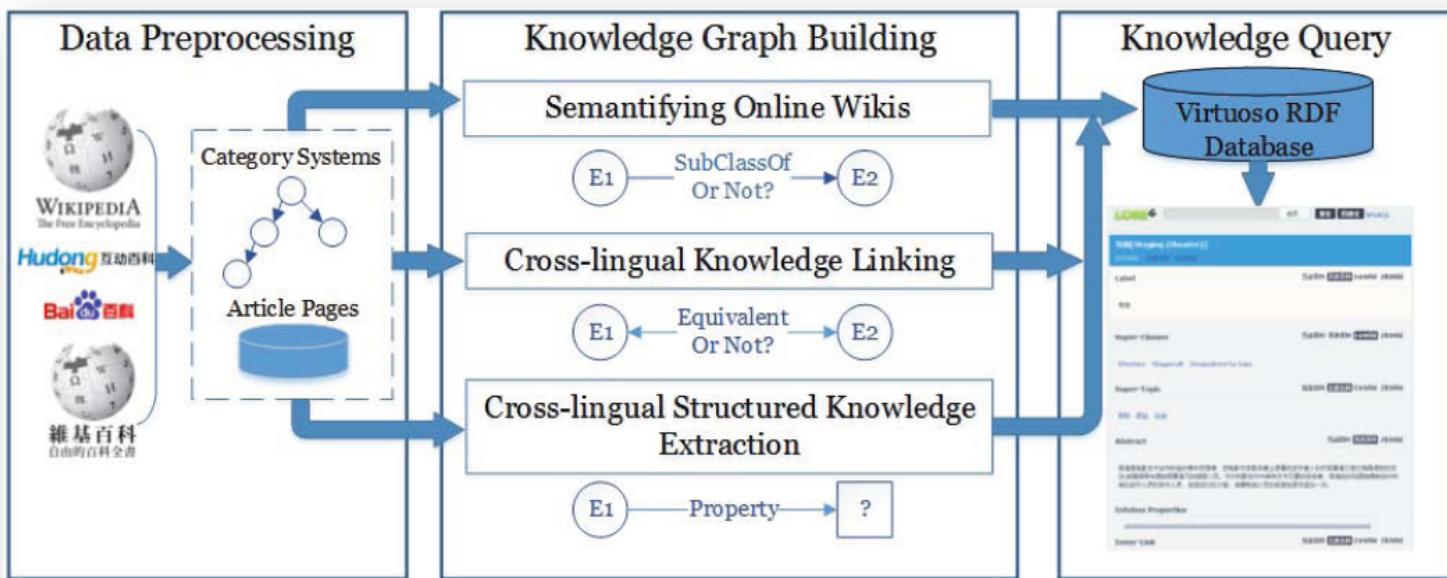
- Large scale bi-lingual knowledge base
- Based on four online wiki encyclopedias
  - Baidu Baike, Hudong Baike and Chinese Wikipedia
  - English Wikipedia

## ■ Knowledge Base

- Machine readable knowledge encyclopedia
- Schema-level: concept(class), property
- Instance-level: instance(individual, entity, ...)

# XLore: English-Chinese Knowledge Base

## ■ Framework





# XLore: English-Chinese Knowledge Base

Properties Instances Publications

**XLORE** All Search SPARQL

**About**

XLore is to extract structured information from heterogenous cross-lingual online wikis and to share the extracted knowledge on the Web. To the best of our knowledge, XLore is the first large-scale knowledge graph with balanced quantity of Chinese-English knowledge. Currently, XLore contains **856,146** classes, **71,596** properties and **7,854,301** instances. It gives a new way for building a large-scale knowledge graph with balanced quantity of knowledge across any two different languages.

**856146 Classes**    **71596 Properties**    **7854301 Instances**

Class Taxonomy

Classes		Properties		Instances	
English	639,020	96.26%	15,380	27.24%	3,879,121
Chinese	88,615	13.35%	51,618	91.44%	7,409,519
Cross-lingual	63,895	9.63%	10,519	18.69%	432,598
Total	663,740		56,449		10,856,042

文化[Culture] Visualization

Label

Hudong Baike Baidu Baike EnWiki ZhWiki

文化

Sub Classes

Baidu Baike Hudong Baike EnWiki ZhWiki

饮食文化 语言 建筑 宗教 次文化 宗教文化 铁器时代 各国文化 电视剧

Super Topics

Baidu Baike Hudong Baike EnWiki ZhWiki

<http://xlore.org/>



# Summary

- **Multilingual ontology building**
  - multilingual concept and property linking
- **Cross lingual knowledge Embedding and computing using deep learning**
- **Event knowledge representation and learning from wiki knowledge bases**
- **More Web based multilingual knowledge applications are expected**



# Conclusions

- ❑ Knowledge graph maps the Internet from the string description to the semantic representation of objective world knowledge.
- ❑ Knowledge graph is the keystone for semantically linking big data.
- ❑ Knowledge is the infrastructure for us to understand the information on the Internet.
- ❑ Can deep learning make knowledge computable?
  - China-Beijing=French-Paris
- ❑ Xlore : a large scale multilingual knowledge graph

A little semantic, a long way to go.

We are on the way ...



**Thanks for your attention!**  
**Q&A**

Juanzi Li  
Tsinghua University  
[lijuanzi@tsinghua.edu.cn](mailto:lijuanzi@tsinghua.edu.cn)