



# Some Recent Work on Knowledge Fusion

Dr. Wang Haofen

Department of Computer Science & Engineering

East China University of Science and Technology (ECUST)

December 27<sup>st</sup>, 2015

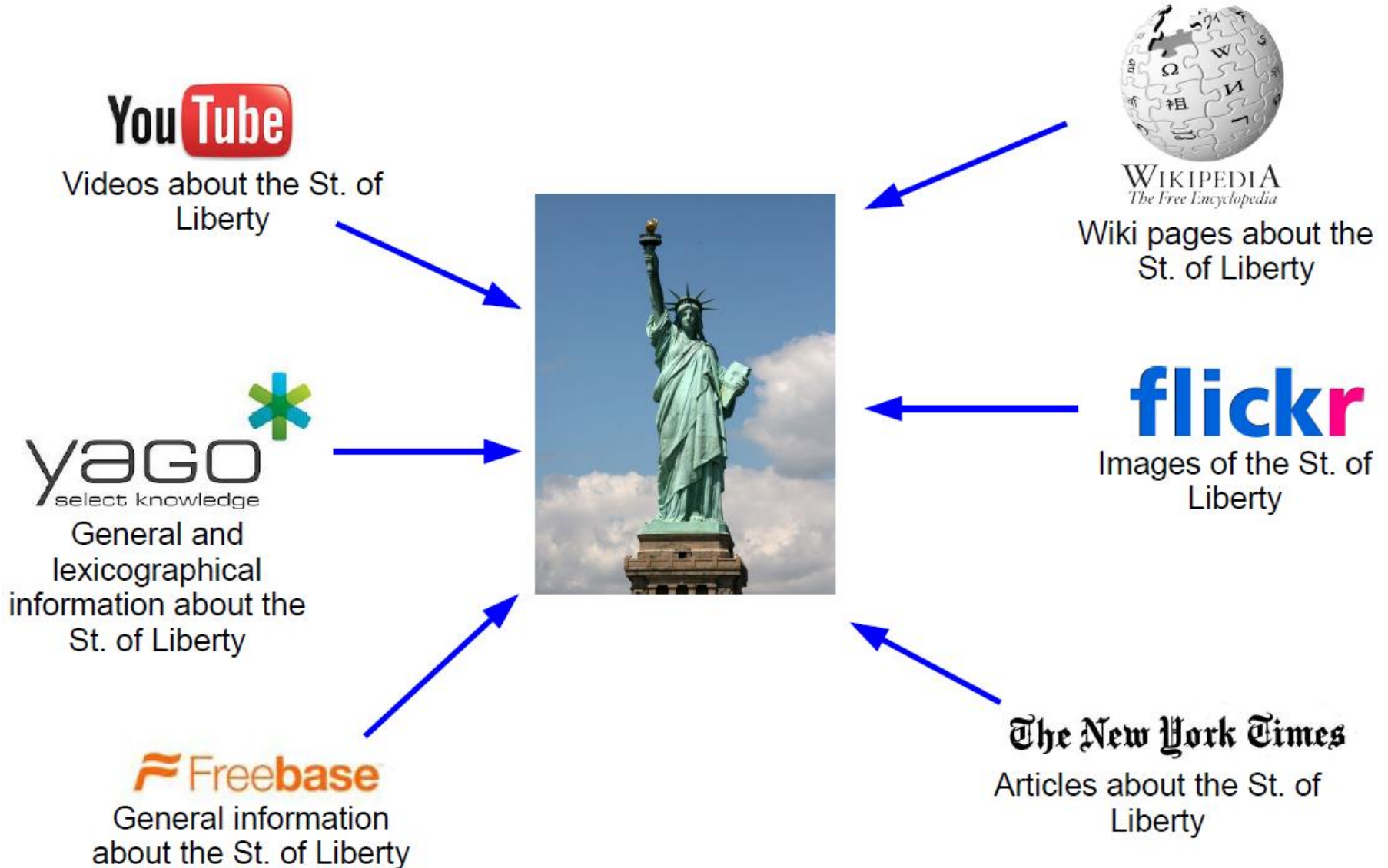




# A Brief Introduction of Knowledge Fusion

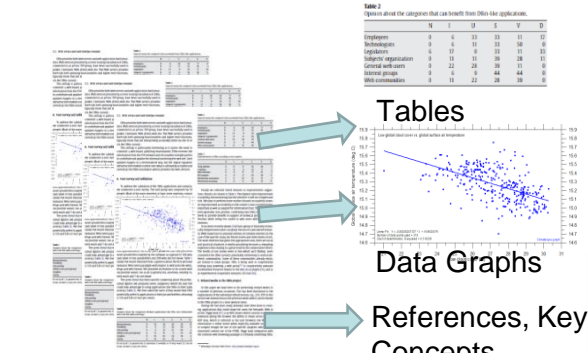


# The same entity in multiple sources



# A common way to publish (vertical) KGs

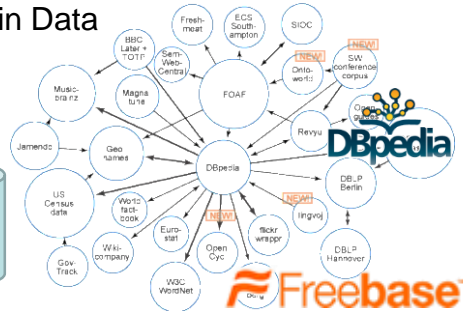
Unstructured/Semi-structured content



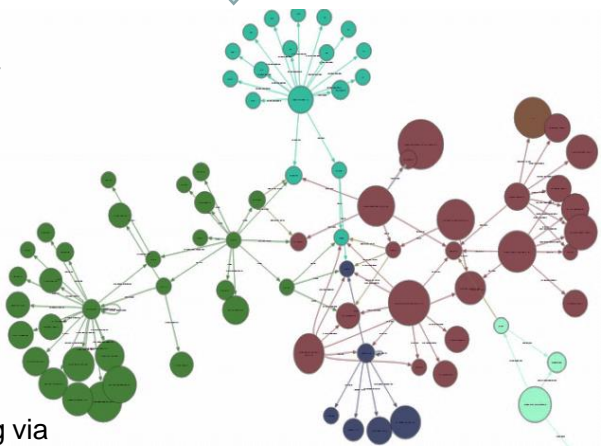
Relational DB



External Domain Data



← Align →



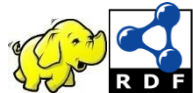
Customer Data



Enrichment and Encoding via Domain Ontology



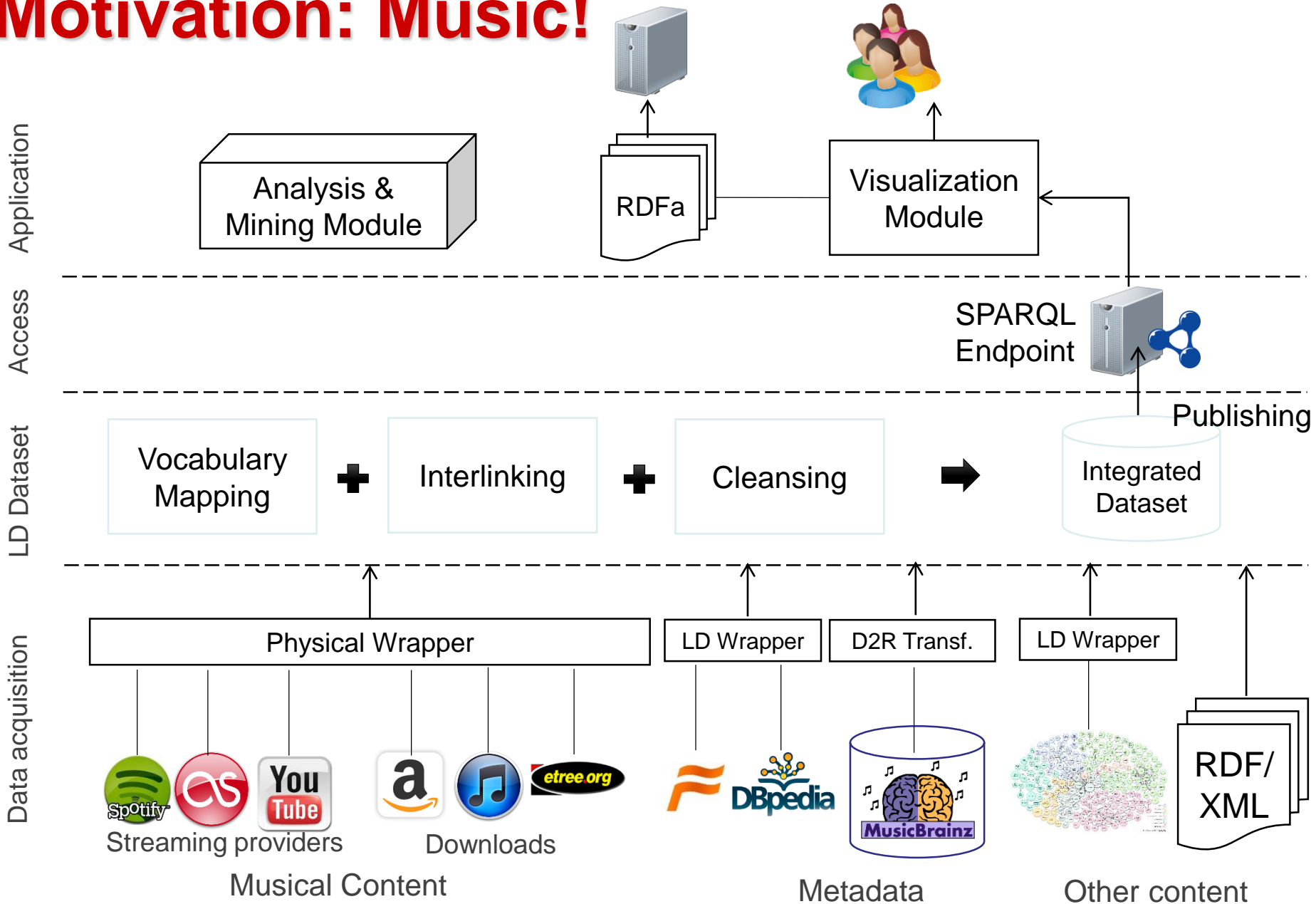
Knowledge Graph



- Search++
- Recommendations
- Vertical applications
- Explorative interfaces



# Motivation: Music!



# Linked RDF Triples on the Web

yago/wordnet: Artist109812338

yago/wordnet: Actor109765278

rdfs:subclassOf

rdfs:subclassOf

yago/wikicategory: ItalianComposer

imdb.com/name/nm0910607/

rdf:type

rdf:type

dbpedia.org/resource/Ennio\_Morricone

imdb.com/title/tt0361748/

dbprop:composedMusicFor

dbpprop:citizenOf

dbpedia.org/resource/Rome

rdf.freebase.com/ns/en.rome

owl:sameAs

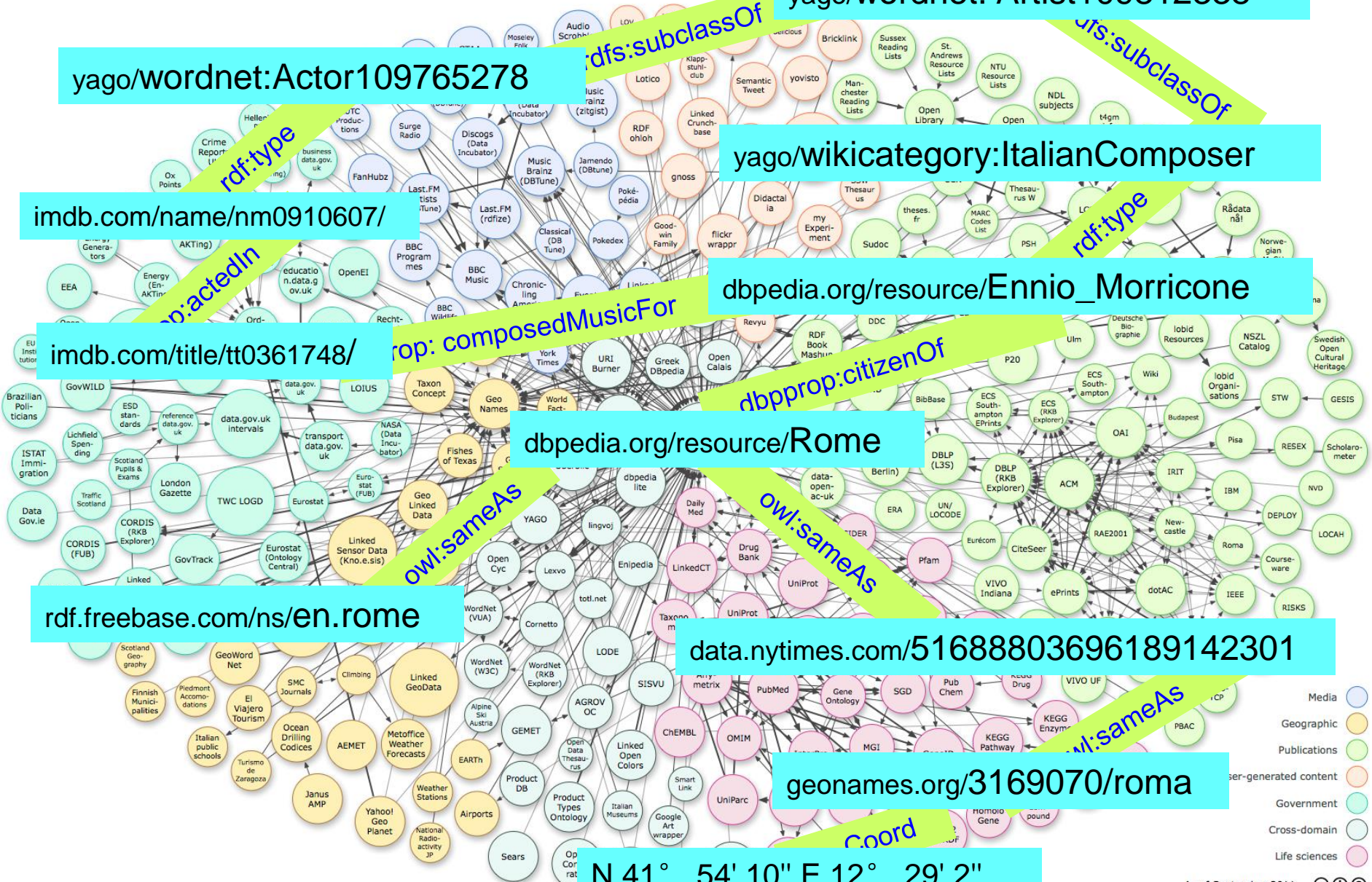
owl:sameAs

data.nytimes.com/51688803696189142301

geonames.org/3169070/roma

owl:sameAs

N 41° 54' 10" E 12° 29' 2"



- Media
- Geographic
- Publications
- Government
- Cross-domain
- Life sciences

# Beyond data-level integration

## Mapping en:Infobox book

Template Mapping <small>(help)</small>	
map to class	Book

Property Mapping <small>(help)</small>	
template property	author
ontology property	author

Property Mapping <small>(help)</small>	
template property	illustrator
ontology property	illustrator

```
}}Infobox book
| author =
| title_orig =
| translator =
| illustrator =
| subject =
| genre =
}}
```

Class <i>Book</i> :	
Properties	
author	
coverArtist	
firstPublicationDate	
illustrator	
isbn	
lastPublicationDate	
...	

## Mapping el:Βιβλίο

Template Mapping <small>(help)</small>	
map to class	Book

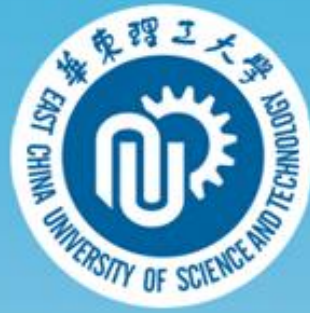
Property Mapping <small>(help)</small>	
template property	συγγραφέας
ontology property	author

Property Mapping <small>(help)</small>	
template property	εικονογράφηση
ontology property	illustrator

```
}}Βιβλίο
| συγγραφέας =
| είδος =
| εκδότης =
| πρώτη_έκδοση =
| ISBN =
| εικονογράφηση =
}}
```





# Instance Matching in Zhishi.me





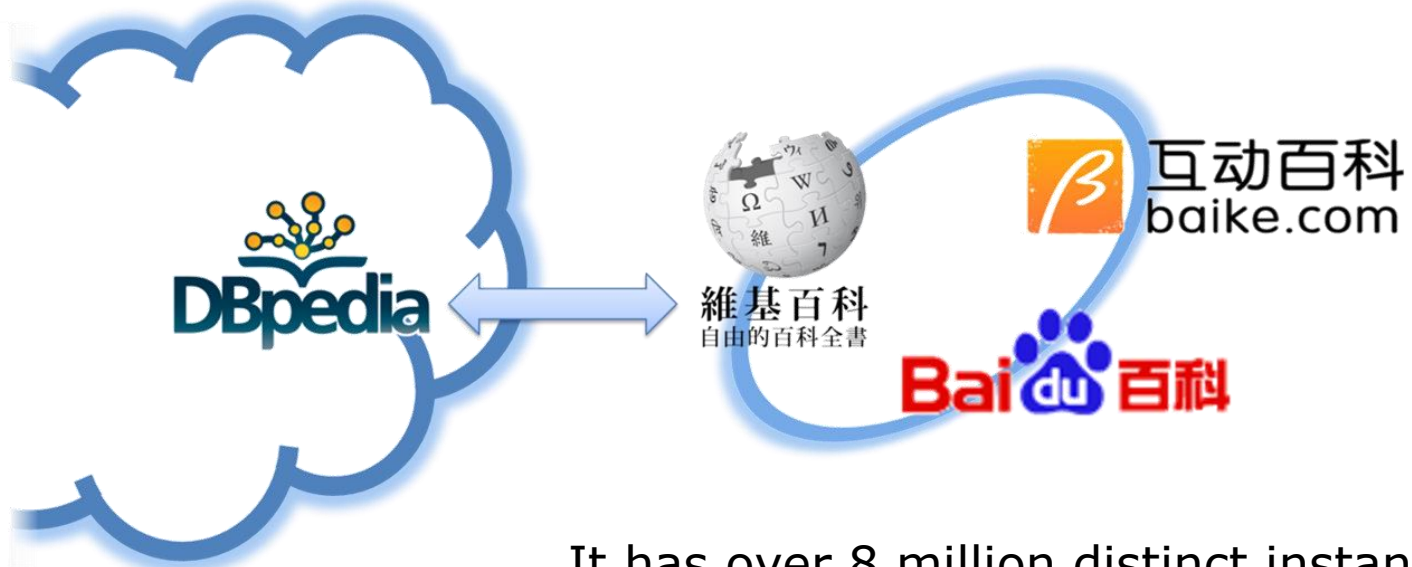
# Zhishi.me

- Zhishi.me (<http://zhishi.me>) is the first effort to publish large scale Chinese semantic data and link them together as a Chinese Linking Open Data (CLOD).

Linking Open Data

链接开放数据

ZHISHI.me



It has over 8 million distinct instances and 200 million RDF triples.

您已登录到该词条，更多含义请查看上海市。

上海



中文名称: 上海
外文名称: Municipality of Shanghai
别名: 申
行政区划: 直辖市
所属地区: 华东地区
下辖地区: 黄浦区、徐汇区、静安区等

- 目录
1 历史沿革
2 上海概况
3 行政区划
4 人口概述
5 政府领导
6 繁华都市

历史沿革

春秋时属吴，吴灭后入越，越王无疆大败于楚，后成为楚国重要的封地（故上海别称“申”），前223年秦灭越后设会稽郡（治所在今苏州）并辖有郟县（今金山）、由拳县（今嘉兴）和海盐县等。郟县包括今上海市大部（其余尚未成陆），前207年郟县改名菑县。晋时松江（吴淞江）居民创造一种竹编捕鱼工具“簖”，又因流入海处称“滬”，松江下游一带称“菑涂”，后又改名为“沪”。唐天宝十年（751年），吴郡刺史赵国鼎上奏，奏请划昆山南境、嘉兴东境、海盐北境，设华亭县。宋淳化二年（991年）松江淤浅，船舶无法上海湾。南宋咸亨三年（1267年）华亭县在上海西岸设“上海镇”。元初至元十四年（1277年），华亭县升格为府，翌年更名松江府，辖华亭县（一府一县）。元朝至元十九年（1292年），析华亭县部分，设上海县，均隶于松江府。上海县治在今黄浦区，这是上海建城的开始。

扩展阅读

- 1 上海市人民政府门户网站 http://www.shanghai.gov.cn
2 上海市人民政府新闻办公室 http://www.shibp.gov.cn
3 上海浦东论坛 http://www.pudongba.com/
4 上海外滩地标看视频 http://v.youku.com/v\_showid\_YMjUzNDAsMjM4.html

开放分类: 中国, 直辖市, 上海, 申城, 沪上

“上海”相关词条: 伦敦 纽约 巴黎 东京 芝加哥 法兰克福 香港 洛杉矶 米兰 旧金山 悉尼 多伦多 苏黎世 布鲁塞尔 马德里

Labels

上海, 中国 大陆 第一大城市; 四个中央直辖市之一; 是中国大陆的经济、金融、贸易和航运中心。上海创造和打破了 中国 世界纪录协会 多项世界之最、中国之最。上海位于中国 大陆 海岸线中部的长江口, 拥有中国最大的外贸港口、最大的工业基地。有超过2000万人居住和生活在上海地区, 其中大部分属汉族江浙民系, 通行吴语上海话。上海又是一座新兴的 旅游目的地, 具有深厚的近代城市文化底蕴和众多的历史古迹。如今上海已经发展成为一个 闪耀 全球的国际化大都市, 并致力于在2020年建设成为 国际金融中心 和 航运中心。上海是 2010年世界博览会 举办城市。

中文名称: 上海
外文名称: Municipality of Shanghai
别名: 申
行政区划: 直辖市
所属地区: 华东地区
下辖地区: 黄浦区、徐汇区、静安区等

Internal Links

- 历史沿革
上海概况
行政区划
人口概述
政府领导
繁华都市

历史沿革

春秋时属吴，吴灭后入越，越王无疆大败于楚，后成为楚国重要的封地（故上海别称“申”），前223年秦灭越后设会稽郡（治所在今苏州）并辖有郟县（今金山）、由拳县（今嘉兴）和海盐县等。郟县包括今上海市大部（其余尚未成陆），前207年郟县改名菑县。晋时松江（吴淞江）居民创造一种竹编捕鱼工具“簖”，又因流入海处称“滬”，松江下游一带称“菑涂”，后又改名为“沪”。唐天宝十年（751年），吴郡刺史赵国鼎上奏，奏请划昆山南境、嘉兴东境、海盐北境，设华亭县。宋淳化二年（991年）松江淤浅，船舶无法上海湾。南宋咸亨三年（1267年）华亭县在上海西岸设“上海镇”。元初至元十四年（1277年），华亭县升格为府，翌年更名松江府，辖华亭县（一府一县）。元朝至元十九年（1292年），析华亭县部分，设上海县，均隶于松江府。上海县治在今黄浦区，这是上海建城的开始。

扩展阅读

- 1 上海市人民政府门户网站 http://www.shanghai.gov.cn
2 上海市人民政府新闻办公室 http://www.shibp.gov.cn
3 上海浦东论坛 http://www.pudongba.com/
4 上海外滩地标看视频 http://v.youku.com/v\_showid\_YMjUzNDAsMjM4.html

开放分类: 中国, 直辖市, 上海, 申城, 沪上

“上海”相关词条: 伦敦 纽约 巴黎 东京 芝加哥 法兰克福 香港 洛杉矶 米兰 旧金山 悉尼 多伦多 苏黎世 布鲁塞尔 马德里

上海市

开放分类: 中国城市 中国直辖市 中国省级行政区



“上海”是“上海市”的简称。
上海, 中国 大陆 第一大城市; 四个中央直辖市之一; 是中国大陆的经济、金融、贸易和航运中心。上海创造和打破了 中国 世界纪录协会 多项世界之最、中国之最。上海位于中国 大陆 海岸线中部的长江口, 拥有中国最大的外贸港口、最大的工业基地。有超过2000万人居住和生活在上海地区, 其中大部分属汉族江浙民系, 通行吴语上海话。上海又是一座新兴的 旅游目的地, 具有深厚的近代城市文化底蕴和众多的历史古迹。如今上海已经发展成为一个 闪耀 全球的国际化大都市, 并致力于在2020年建设成为 国际金融中心 和 航运中心。上海是 2010年世界博览会 举办城市。

目录

- 1 概述
2 历史沿革
3 行政区划
4 地理气候

上海市 - 概述

上海位于北纬31度14分, 东经121度29分。地处长江三角洲前缘, 东濒东海, 南临杭州湾, 西接江苏、浙江两省, 北界长江入海口, 正当中国南北海岸线的中部, 交通便利, 腹地广阔, 地理位置优越, 是一个良好的江海港口。

地形

除西南部有少数丘陵山脉外, 上海境内全为坦荡低平的平原, 是长江三角洲冲积平原的一部分。

能源

煤、石油、水力的储藏和生产, 所需能源都要靠其他省市的支援。但是, 上海具有一定数量和较高质量的“二次能源”生产, 产品主要是电力、石油制品、焦煤和煤气(包括液化石油气)。其他可以利用开发的能源还有沼气、风能、潮汐能和太阳能。

Redirects

“上海”是“上海市”的简称。

Infobox Properties

上海市 - 概述
上海位于北纬31度14分, 东经121度29分。地处长江三角洲前缘, 东濒东海, 南临杭州湾, 西接江苏、浙江两省, 北界长江入海口, 正当中国南北海岸线的中部, 交通便利, 腹地广阔, 地理位置优越, 是一个良好的江海港口。

目录

- 1 概述
2 历史沿革
3 行政区划
4 地理气候

上海市 - 概述

上海位于北纬31度14分, 东经121度29分。地处长江三角洲前缘, 东濒东海, 南临杭州湾, 西接江苏、浙江两省, 北界长江入海口, 正当中国南北海岸线的中部, 交通便利, 腹地广阔, 地理位置优越, 是一个良好的江海港口。

地形

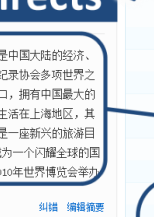
除西南部有少数丘陵山脉外, 上海境内全为坦荡低平的平原, 是长江三角洲冲积平原的一部分。

能源

煤、石油、水力的储藏和生产, 所需能源都要靠其他省市的支援。但是, 上海具有一定数量和较高质量的“二次能源”生产, 产品主要是电力、石油制品、焦煤和煤气(包括液化石油气)。其他可以利用开发的能源还有沼气、风能、潮汐能和太阳能。

上海市

开放分类: 中国城市 中国直辖市 中国省级行政区



“上海”是“上海市”的简称。
上海, 中国 大陆 第一大城市; 四个中央直辖市之一; 是中国大陆的经济、金融、贸易和航运中心。上海创造和打破了 中国 世界纪录协会 多项世界之最、中国之最。上海位于中国 大陆 海岸线中部的长江口, 拥有中国最大的外贸港口、最大的工业基地。有超过2000万人居住和生活在上海地区, 其中大部分属汉族江浙民系, 通行吴语上海话。上海又是一座新兴的 旅游目的地, 具有深厚的近代城市文化底蕴和众多的历史古迹。如今上海已经发展成为一个 闪耀 全球的国际化大都市, 并致力于在2020年建设成为 国际金融中心 和 航运中心。上海是 2010年世界博览会 举办城市。

目录

- 1 概述
2 历史沿革
3 行政区划
4 地理气候

上海市 - 概述

上海位于北纬31度14分, 东经121度29分。地处长江三角洲前缘, 东濒东海, 南临杭州湾, 西接江苏、浙江两省, 北界长江入海口, 正当中国南北海岸线的中部, 交通便利, 腹地广阔, 地理位置优越, 是一个良好的江海港口。

地形

除西南部有少数丘陵山脉外, 上海境内全为坦荡低平的平原, 是长江三角洲冲积平原的一部分。

能源

煤、石油、水力的储藏和生产, 所需能源都要靠其他省市的支援。但是, 上海具有一定数量和较高质量的“二次能源”生产, 产品主要是电力、石油制品、焦煤和煤气(包括液化石油气)。其他可以利用开发的能源还有沼气、风能、潮汐能和太阳能。

上海市

开放分类: 中国城市 中国直辖市 中国省级行政区



“上海”是“上海市”的简称。
上海, 中国 大陆 第一大城市; 四个中央直辖市之一; 是中国大陆的经济、金融、贸易和航运中心。上海创造和打破了 中国 世界纪录协会 多项世界之最、中国之最。上海位于中国 大陆 海岸线中部的长江口, 拥有中国最大的外贸港口、最大的工业基地。有超过2000万人居住和生活在上海地区, 其中大部分属汉族江浙民系, 通行吴语上海话。上海又是一座新兴的 旅游目的地, 具有深厚的近代城市文化底蕴和众多的历史古迹。如今上海已经发展成为一个 闪耀 全球的国际化大都市, 并致力于在2020年建设成为 国际金融中心 和 航运中心。上海是 2010年世界博览会 举办城市。

目录

- 1 概述
2 历史沿革
3 行政区划
4 地理气候

上海市 - 概述

上海位于北纬31度14分, 东经121度29分。地处长江三角洲前缘, 东濒东海, 南临杭州湾, 西接江苏、浙江两省, 北界长江入海口, 正当中国南北海岸线的中部, 交通便利, 腹地广阔, 地理位置优越, 是一个良好的江海港口。

地形

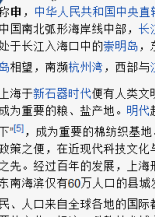
除西南部有少数丘陵山脉外, 上海境内全为坦荡低平的平原, 是长江三角洲冲积平原的一部分。

能源

煤、石油、水力的储藏和生产, 所需能源都要靠其他省市的支援。但是, 上海具有一定数量和较高质量的“二次能源”生产, 产品主要是电力、石油制品、焦煤和煤气(包括液化石油气)。其他可以利用开发的能源还有沼气、风能、潮汐能和太阳能。

上海市

开放分类: 中国城市 中国直辖市 中国省级行政区



“上海”是“上海市”的简称。
上海, 中国 大陆 第一大城市; 四个中央直辖市之一; 是中国大陆的经济、金融、贸易和航运中心。上海创造和打破了 中国 世界纪录协会 多项世界之最、中国之最。上海位于中国 大陆 海岸线中部的长江口, 拥有中国最大的外贸港口、最大的工业基地。有超过2000万人居住和生活在上海地区, 其中大部分属汉族江浙民系, 通行吴语上海话。上海又是一座新兴的 旅游目的地, 具有深厚的近代城市文化底蕴和众多的历史古迹。如今上海已经发展成为一个 闪耀 全球的国际化大都市, 并致力于在2020年建设成为 国际金融中心 和 航运中心。上海是 2010年世界博览会 举办城市。

目录

- 1 概述
2 历史沿革
3 行政区划
4 地理气候

上海市 - 概述

上海位于北纬31度14分, 东经121度29分。地处长江三角洲前缘, 东濒东海, 南临杭州湾, 西接江苏、浙江两省, 北界长江入海口, 正当中国南北海岸线的中部, 交通便利, 腹地广阔, 地理位置优越, 是一个良好的江海港口。

地形

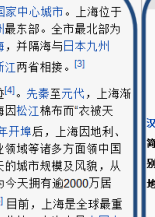
除西南部有少数丘陵山脉外, 上海境内全为坦荡低平的平原, 是长江三角洲冲积平原的一部分。

能源

煤、石油、水力的储藏和生产, 所需能源都要靠其他省市的支援。但是, 上海具有一定数量和较高质量的“二次能源”生产, 产品主要是电力、石油制品、焦煤和煤气(包括液化石油气)。其他可以利用开发的能源还有沼气、风能、潮汐能和太阳能。

上海市

开放分类: 中国城市 中国直辖市 中国省级行政区



“上海”是“上海市”的简称。
上海, 中国 大陆 第一大城市; 四个中央直辖市之一; 是中国大陆的经济、金融、贸易和航运中心。上海创造和打破了 中国 世界纪录协会 多项世界之最、中国之最。上海位于中国 大陆 海岸线中部的长江口, 拥有中国最大的外贸港口、最大的工业基地。有超过2000万人居住和生活在上海地区, 其中大部分属汉族江浙民系, 通行吴语上海话。上海又是一座新兴的 旅游目的地, 具有深厚的近代城市文化底蕴和众多的历史古迹。如今上海已经发展成为一个 闪耀 全球的国际化大都市, 并致力于在2020年建设成为 国际金融中心 和 航运中心。上海是 2010年世界博览会 举办城市。

目录

- 1 概述
2 历史沿革
3 行政区划
4 地理气候

上海市 - 概述

上海位于北纬31度14分, 东经121度29分。地处长江三角洲前缘, 东濒东海, 南临杭州湾, 西接江苏、浙江两省, 北界长江入海口, 正当中国南北海岸线的中部, 交通便利, 腹地广阔, 地理位置优越, 是一个良好的江海港口。

地形

除西南部有少数丘陵山脉外, 上海境内全为坦荡低平的平原, 是长江三角洲冲积平原的一部分。

能源

煤、石油、水力的储藏和生产, 所需能源都要靠其他省市的支援。但是, 上海具有一定数量和较高质量的“二次能源”生产, 产品主要是电力、石油制品、焦煤和煤气(包括液化石油气)。其他可以利用开发的能源还有沼气、风能、潮汐能和太阳能。

Abstracts

维基百科,自由的百科全书



上海市 (ⓘ) 发音 (帮助·信息); 吴语: Zaohe), 简称沪, 别称申, 中华人民共和国中央直辖市、国家中心城市。上海位于中国南北弧形海岸线中部, 长江三角洲东部, 全市最北部为处于长江入海口中的崇明岛, 东向东海, 并隔海与日本九州岛相望, 南濒杭州湾, 西部与江苏、浙江两省相接。

概况

上海为石器时代便有人类文明的遗址。先秦至元代, 上海渐成为重要的银、盐产地, 明代起, 上海因松江棉布而衣被天下, 成为重要的纺织业基地。1843年开埠后, 上海因地利、政策之便, 在近代科技文化及工商业领域等诸多方面居中国之先。经过百年的发展, 上海形成今天的城市规模及风貌, 从东南沿海仅有60万人口的小县城发展成为今天拥有逾2000万居民、人口来自全球各地的国际都会。

外部链接

- 18 外部链接



上海市市花白玉兰

地名出处

吴淞江支流上海浦

国家

中华人民共和国

行政区划类型

直辖市

行政区划代码

31

下级行政区

17市辖区、1县

设市时间

1927年设上海特别市

现任市长

韩正

现任市委书记

俞正声

政府驻地

黄浦区人民大道200号

简称

江苏省、浙江省

城市精神

海纳百川 追求卓越 开明睿智 大气谦和

External Links

- 中国上海网: 上海概览
上海人大公众网
上海政协
上海市人民政府新闻办公室

Categories

10个分类: 上海铁路车站 | 上海 | 中华人民共和国直辖市 | 中国特大城市 | 中国共产党革命圣地 | 太平洋沿海城市 | 长江三角洲城市 | 国家历史文化名城 | 中国沿海开放城市 | 国家森林公园

# Equivalent Instances



## ● Baidu Baike

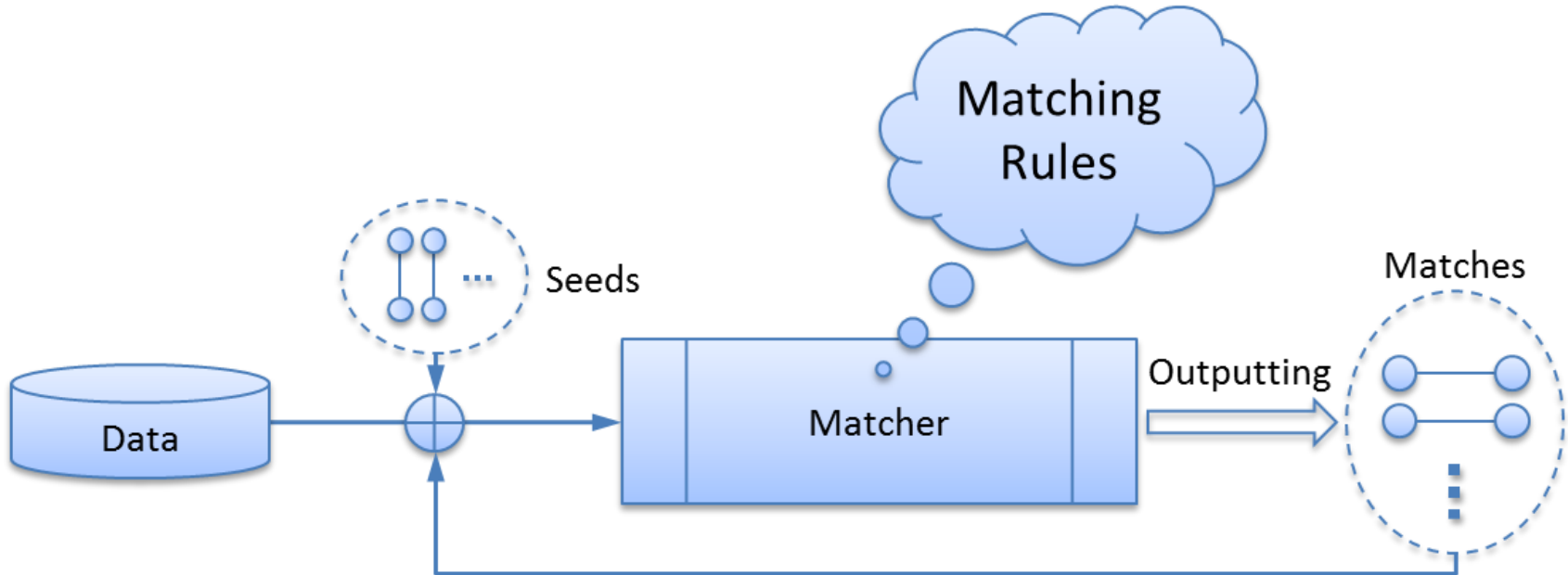
baidu:大熊猫	
baidu:标签	“大熊猫”
baidu:拉丁学名	“ <i>Ailuropoda melanoleuca</i> ”
baidu:纲	baidu:哺乳纲
...	

## ● Hudong Baike

hudong:大熊猫	
hudong:中文学名	“大熊猫”
hudong:二名法	“ <i>Ailuropoda melanoleuca</i> ”
hudong:纲	hudong:哺乳纲
...	



# Our Solution

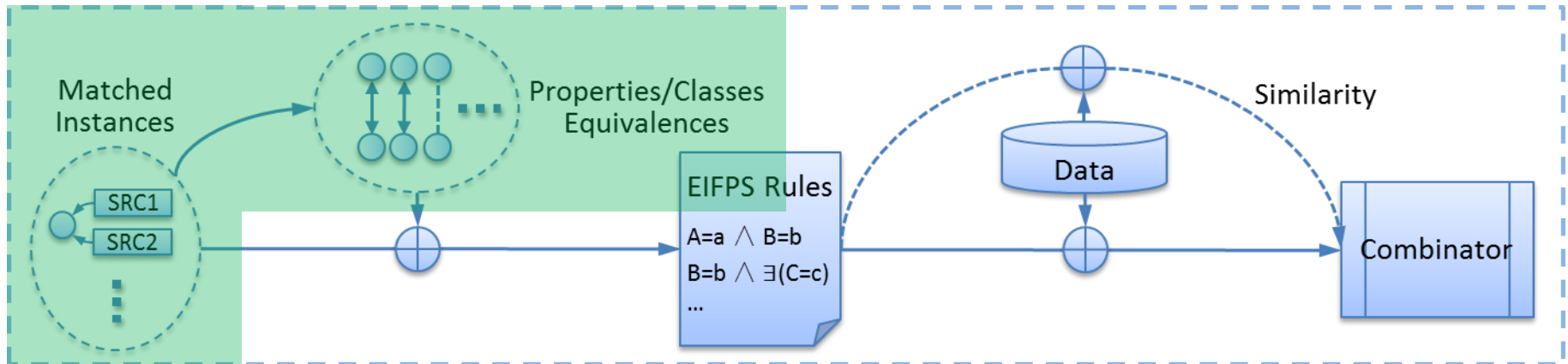


- Automatically discovering and refining dataset-specific matching rules in iterations
  - Deriving these rules by finding the most discriminative data characteristics for a given data source pair.



# Workflow

## – Mining Properties Equivalences

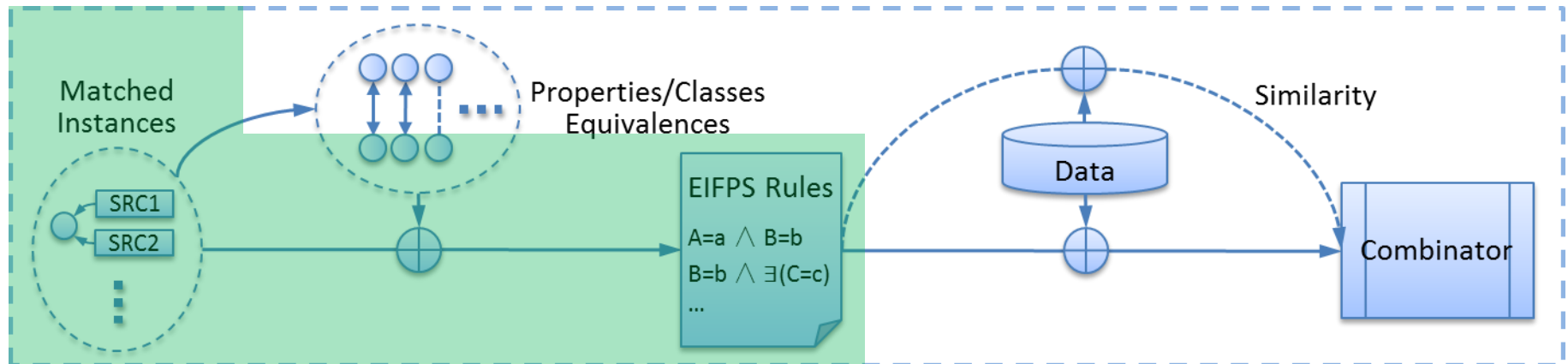


- For each pair of existing matched instances, their property-value pairs are merged.

Values	Property_1	Property_2
“大熊猫”	baidu:标签	hudong:中文学名
“Ailuropoda melanoleuca”	baidu:拉丁学名	hudong:二名法
“白鳍豚”	baidu:标签	hudong:中文学名
“桂花”	baidu:标签	hudong:中文学名
...	...	...

# Workflow

## – Mining Matching Rules

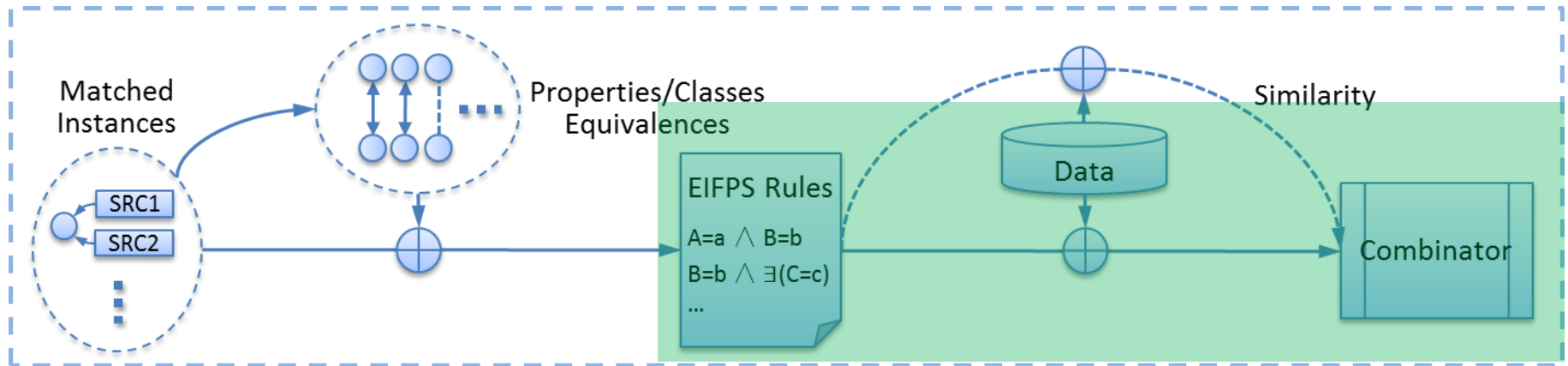


### ● Matching rule (frequent set mining):

- baidu:x and hudong:x are matched, iff.
- $\text{valueOf}(\text{baidu}:\text{标签}) = \text{valueOf}(\text{hudong}:\text{中文学名})$
- and
- $\text{valueOf}(\text{baidu}:\text{拉丁学名}) = \text{valueOf}(\text{hudong}:\text{二名法})$
- and
- $\text{valueOf}(\text{baidu}:\text{纲}) = \text{valueOf}(\text{hudong}:\text{纲})$

# Workflow

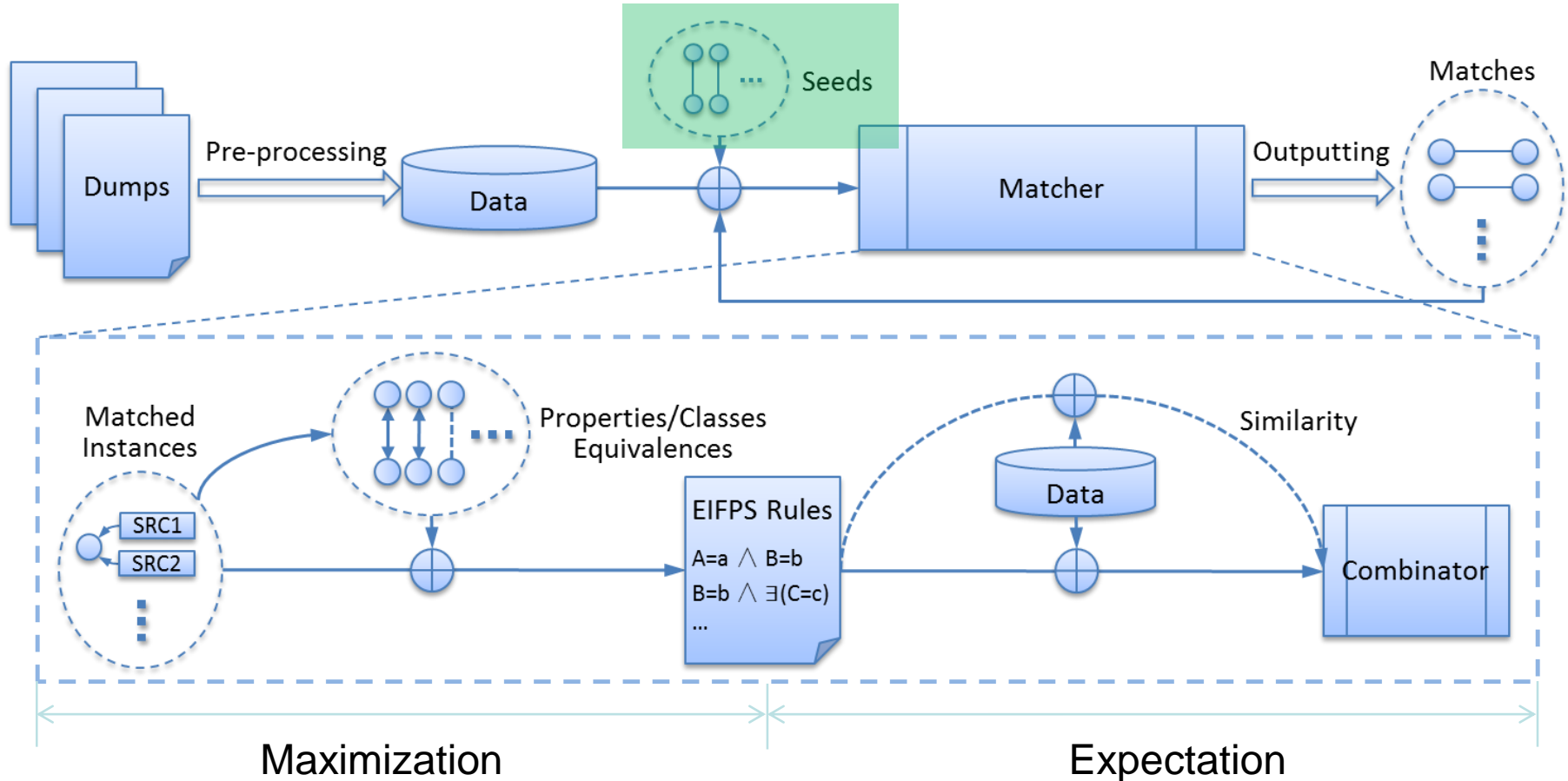
## – Generating Matches



- Applying the obtained rule(s) on the unlabeled data to generate matches' candidates.
- The combiner is used to combine confidence values of a match's candidate.

# Workflow

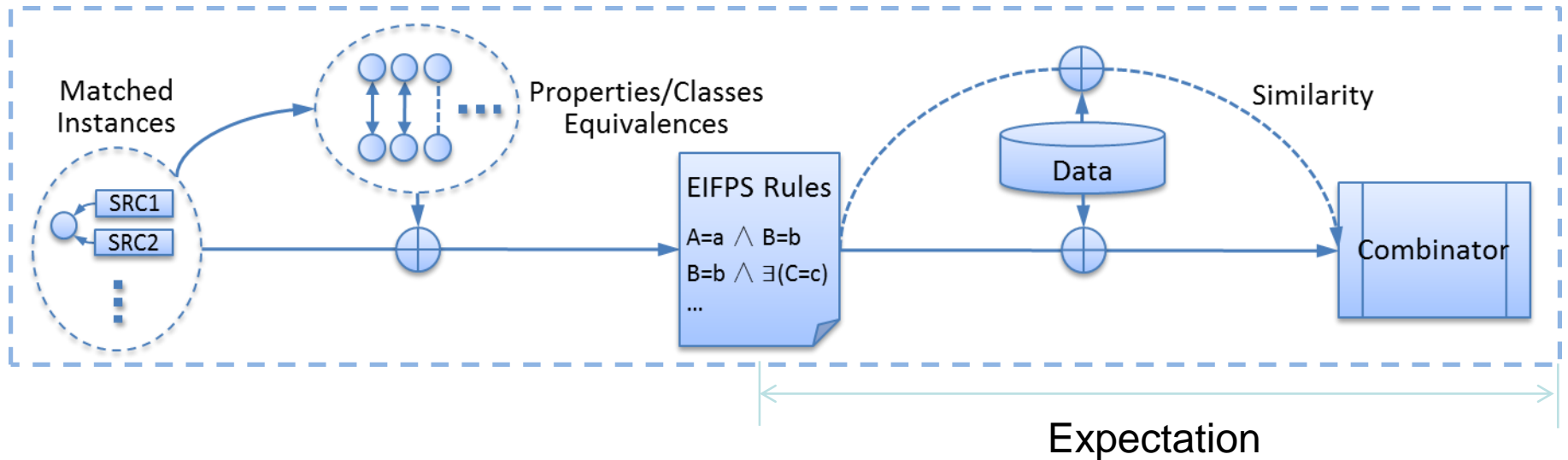
## – the Wrapper Algorithm



- The wrapper is an implementation of Expectation-Maximization iterations.

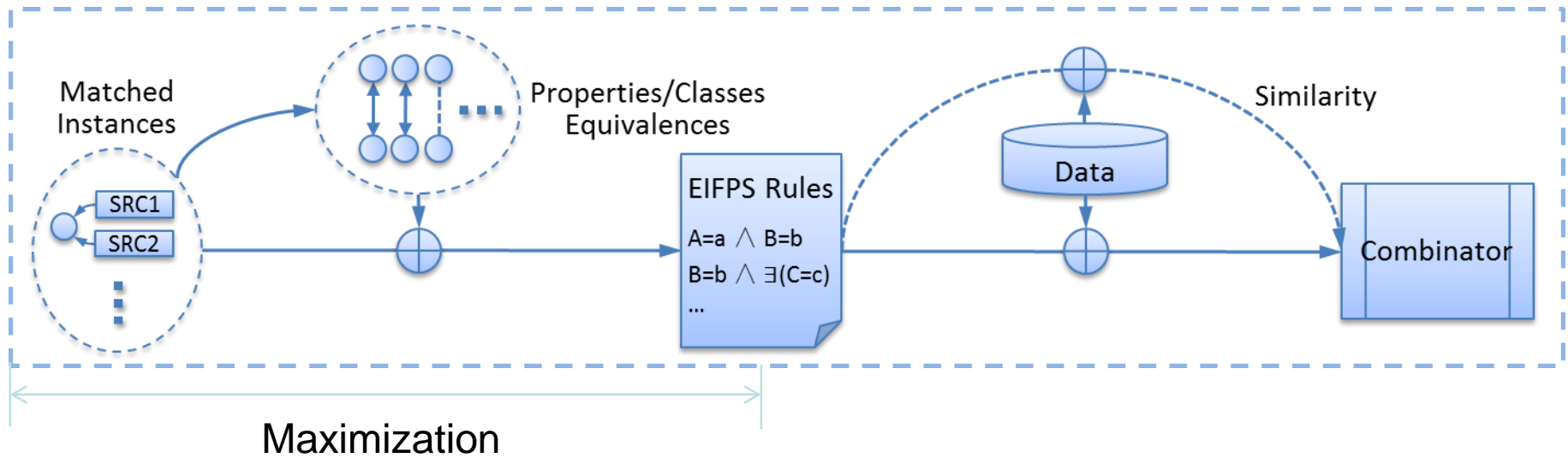


# Workflow – E-step



- The E-step estimates the **missing data (matches)** using the observed data and the current estimate for the **parameters (matching rules)**.

# Workflow – M-step



- The M-step computes **parameters** maximizing the **likelihood function** as the data estimated in E-step are used in line of the actual missing data.

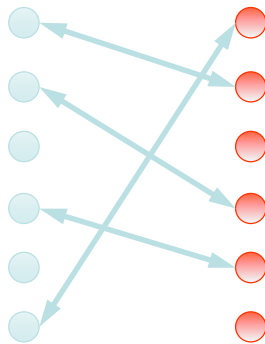
- $M$ : matches
- $\Theta$ : parameters

$$L(\theta; M) = \Pr(M|\theta).$$

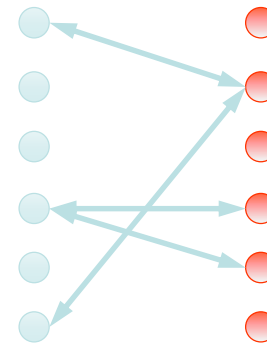
# The Likelihood Function

$$L(\theta; M) \approx \frac{|\text{ConnectedComponent}(M)|}{|\text{Edge}(M)|}$$

- Assuming that no equivalent instances exist in a single data source, we can infer that an instance is equivalent to at most one another from the other data source.
- Incorrect matches in  $M$  may result in a node connecting to more than one other node, which is contrary to the assumption.

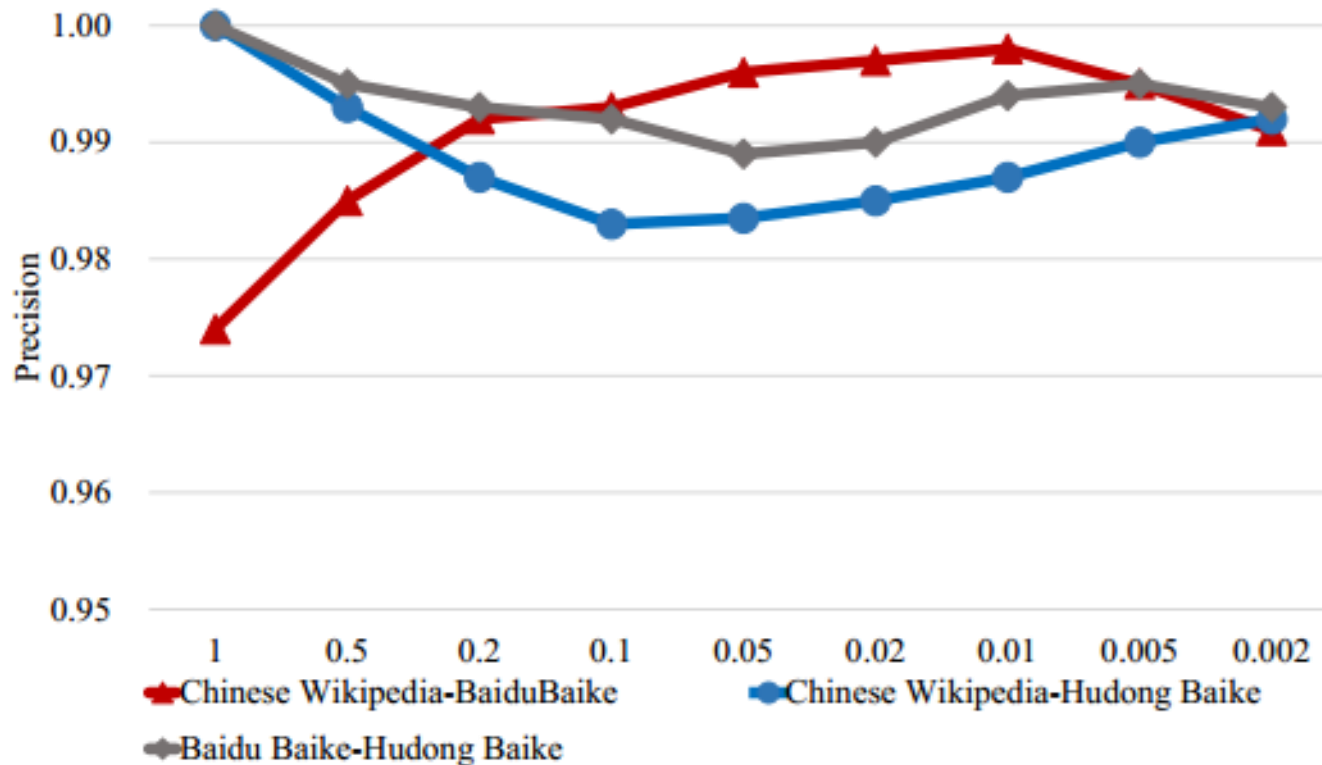


$$P=4/4=1$$



$$P=2/4=0.5$$

# Evaluation – Precisions

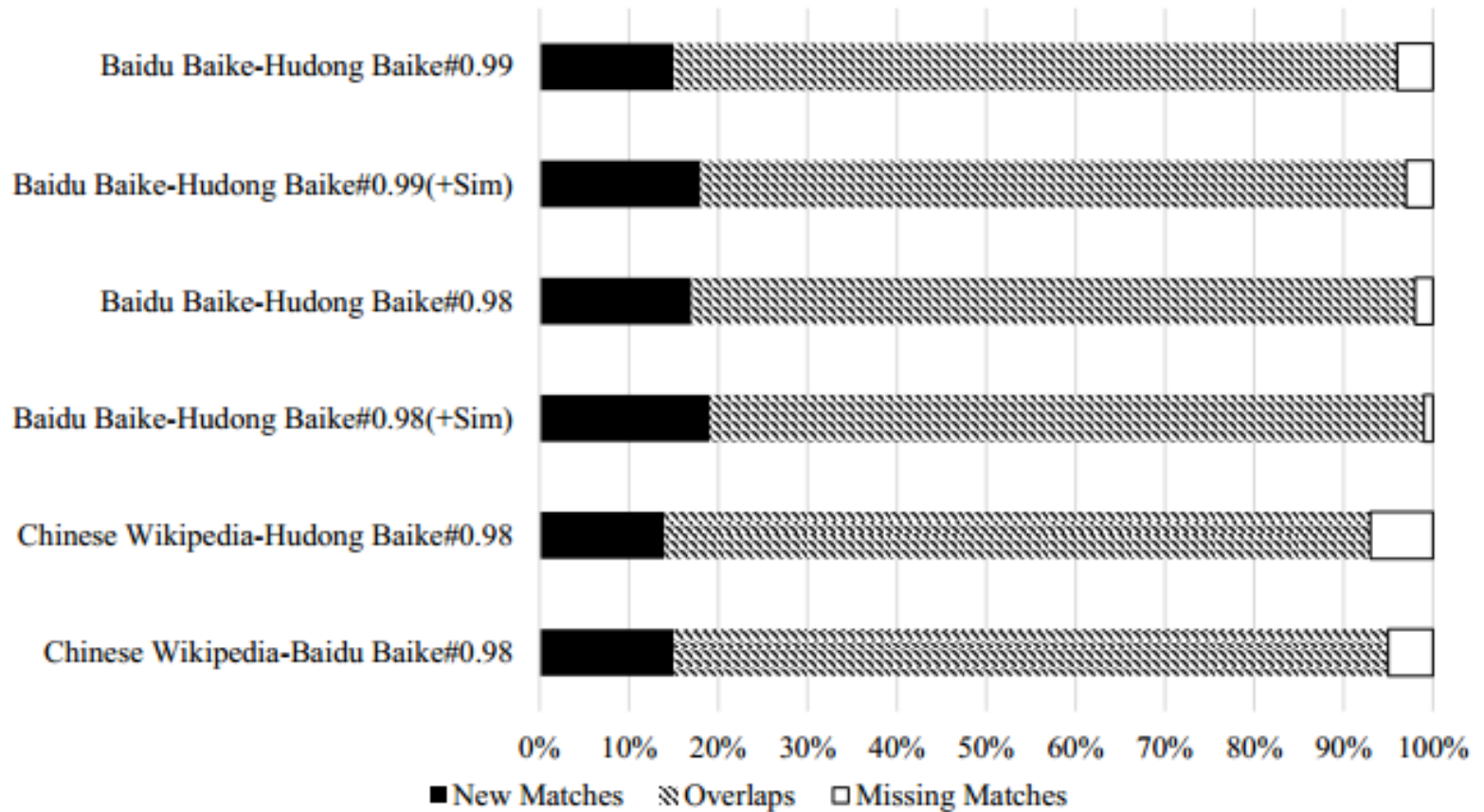


- Sampling a certain number of output matches.
- The X-axis indicates the proportions of selected seeds in complete reference matches.



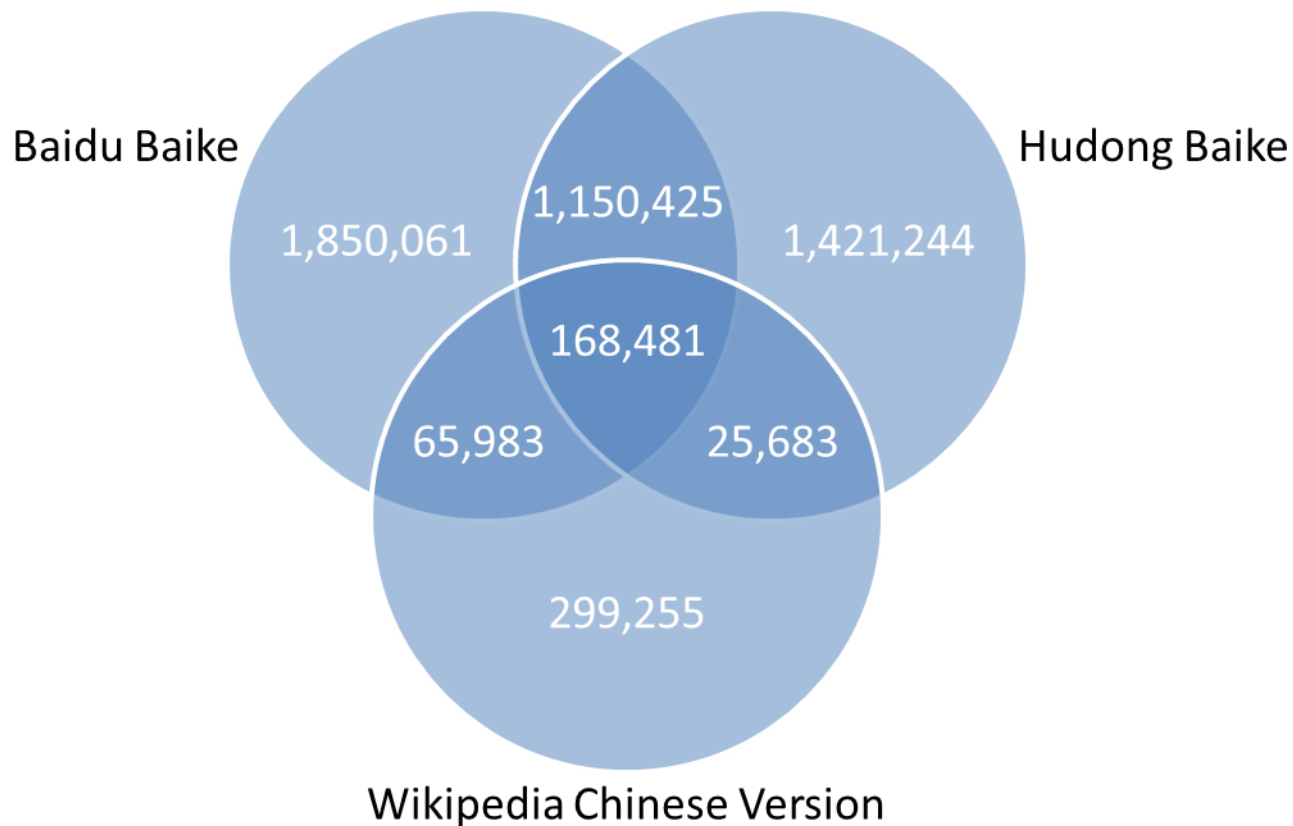
# Evaluation

## – Newly Found Matches



- The match space constituted by reference matches and newly found matches

# Results



- Accuracy: ~97% (manual labeling)

# Web Access to Equivalent Instances

**owl:sameAs**

- hudong:北京 (this)
- baidu:北京
- zhwiki:北京

MERGE PAGE

resources

<owl:sameAs> check box

**owl:sameAs**

- zhwiki:北京市 (this)
- baidu:北京 (this)
- zhwiki:北京
- hudong:北京 (this)
- dbpedia:Beijing
- hudong:北京市
- baidu:北京市

MERGE PAGE

**zhwiki:北京市**  
**hudong:北京**  
**baidu:北京**

Index

- zhishi:abstract
- infobox
- dcterms:subject
- zhishi:thumbnail
- zhishi:relatedPage
- ...

**zhishi:abstract**

北京有着三千余年的建城史和八百五十余年的建都史，最初见于记载的名字为“蓟”。民国时期，称北平。新中国成立后，是中华人民共和国的首都，简称“京”，现为中国四个中央直辖市之一，全国第二大城市及政治、交通和文化中心。北京位于华北平原北端，东南局部地区与天津相连，其余为河北省所环绕。它荟萃了元、明、清以来的中华文化，拥有众多名胜古迹和人文景观，是世界上拥有世界文化遗产最多的城市，每年有超过1亿4700万的旅客。

**infobox**

市花  月季、菊花 → a merged statement

政府驻地  东城区正义路2号

下辖地区  海淀区  西城区  顺义区  大兴区  东城、西城、海淀、丰台、朝阳等

时区  UTC+8 (东八区)

行政区类别  直辖市



友好城市  纽约、东京、巴黎

**dcterms:subject**

中华人民共和国直辖市  京津冀城市群  中国城市  首都  燕京  中国

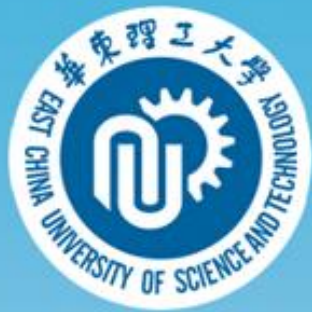
**zhishi:thumbnail**

**zhishi:relatedPage**

圆明园 [SHOW MORE \(182\)](#)

all other statements

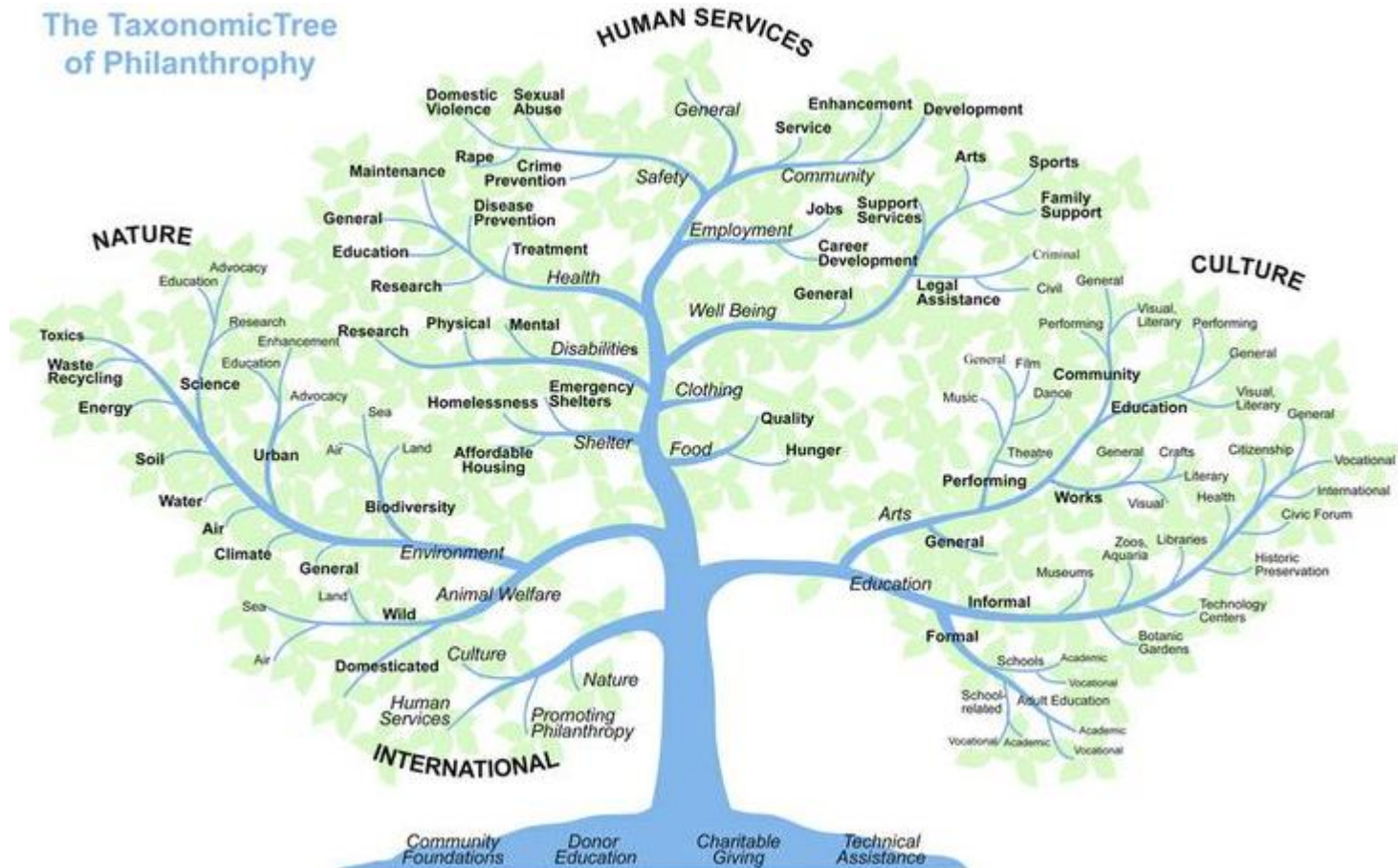


# Cross-lingual Taxonomy Alignment



# Motivation

- Taxonomy refers to a hierarchy of categories which entities are classified to.





# Motivation

- Taxonomies are prevalent on the Web.

The image shows two examples of product catalogues. On the left is the eBay 'All Categories' page, featuring a navigation bar with letters A through M and a list of categories under 'A', including Antiques, Art, and various sub-categories like Antiquities, Architectural & Garden, and Asian Antiques. On the right is the JD.com website, displaying a grid of categories in Chinese, such as 电子书刊 (E-books), 音像 (Audio/Video), 英文原版 (English Originals), 文艺 (Arts and Literature), 少儿 (Children's), 人文社科 (Humanities and Social Sciences), 经管励志 (Business and Inspiration), 生活 (Lifestyle), 科技 (Technology), and 教育 (Education).

**ebay**

All Categories

A · B · C · D · E · F · G · H · I · J · K · L · M

**A** **Antiques**

Antiquities  
Architectural & Garden  
Asian Antiques  
Books & Manuscripts  
Decorative Arts  
Ethnographic  
Furniture  
Home & Hearth

**Art**

Direct from the Artist  
Art from Dealers & Resellers  
Wholesale Lots

Linens & Textiles  
Maps, Atlases & ( )  
Maritime  
Mercantile, Trade  
Musical Instrumer  
Periods & Styles  
Primitives  
Restoration & Ca

**JD.COM**

**电子书刊** 电子书 网络原创 数字杂志 多媒体图书

**音像** 音乐 影视 教育音像

**英文原版** 少儿 商务投资 英语学习与考试 小说 传记 励志

**文艺** 小说 文学 青春文学 传记 艺术

**少儿** 少儿 0-2岁 3-6岁 7-10岁 11-14岁

**人文社科** 历史 哲学 国学 政治/军事 法律 宗教 心理学 文化 社会科学

**经管励志** 经济 金融与投资 管理 励志与成功

**生活** 生活 健身与保健 家庭与育儿 旅游 动漫/幽默

**科技** 科技 工程 建筑 医学 科学与自然 计算机与互联网 体育/运动

**教育** 教材教辅 教育与考试 外语学习 新闻出版 语言文字

## Product Catalogue

# Motivation

- Taxonomies are prevalent on the Web.

Top: World: Chinese Simplified (29,440)

- [地区](#) (13,027)



- [休闲](#) (572)
- [体育](#) (255)
- [健康](#) (766)
- [儿童与青少年@](#) (248)
- [参考](#) (2,254)
- [商业](#) (5,764)
- [家庭](#) (120)
- [新闻](#) (444)
- [游戏](#) (591)
- [社会](#) (1,608)
- [科学](#) (946)
- [艺术](#) (970)
- [计算机](#) (1,672)
- [购物](#) (451)

相关类别:

- [Regional: Asia: China](#) (1,756)
- [World: Chinese Traditional](#) (10,134)

## Web Site Directory

A

- Accounts
- Advertising
- Answers
- Autos

B

- Beauty

C

- Celebrity

D

- Dating
- Developer Network
- Downloads

YAHOO!

F

- Fantasy Sports
- Finance
- Flickr
- Food

G

- Games
- Groups

H

- Health
- Help
- Homes

L

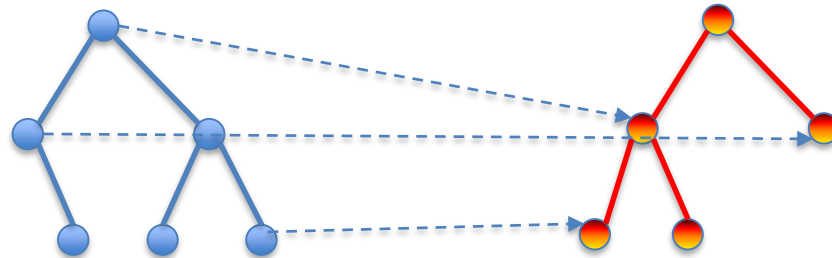
- Local

# Motivation

- As the advent of more and more multilingual resources, the Web has become ***a global information space***.
- To facilitate schema knowledge sharing across languages, we need to deal with the problem of **cross-lingual taxonomy alignment**.

one Chinese taxonomy

one English taxonomy



# Problem and Challenge

- **Problem 1:** Previous work strongly relies on domain-specific information (e.g. book instances in book domain or calculation items in finance domain).
- **Problem 2:** Previous work strongly relies on string similarities to capture different kinds of features.

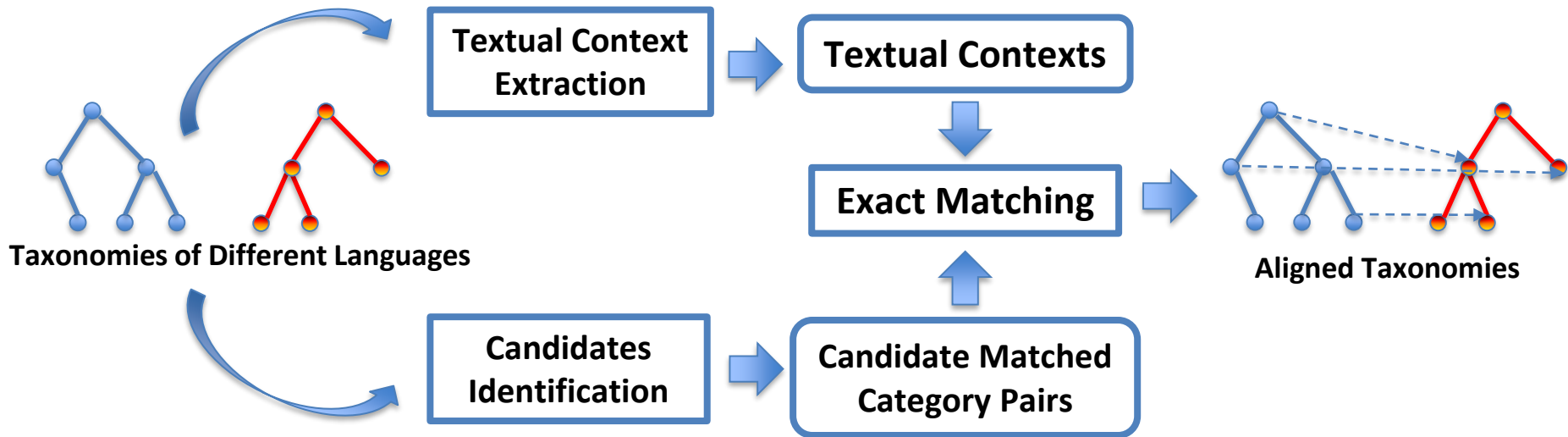
“户外/运动服”  “Outdoor/Sportswear”  “Athletic Apparel”

**Google Translate** **Dissimilar in String Similarity**

(Prytkova, Weikum, and Spaniol 2015) & (Spohr, Hollink, and Cimiano 2011)

# Approach

- Our proposed approach consists of three components:
  - Candidates identification
  - Textual Context Extraction
  - Exact Matching





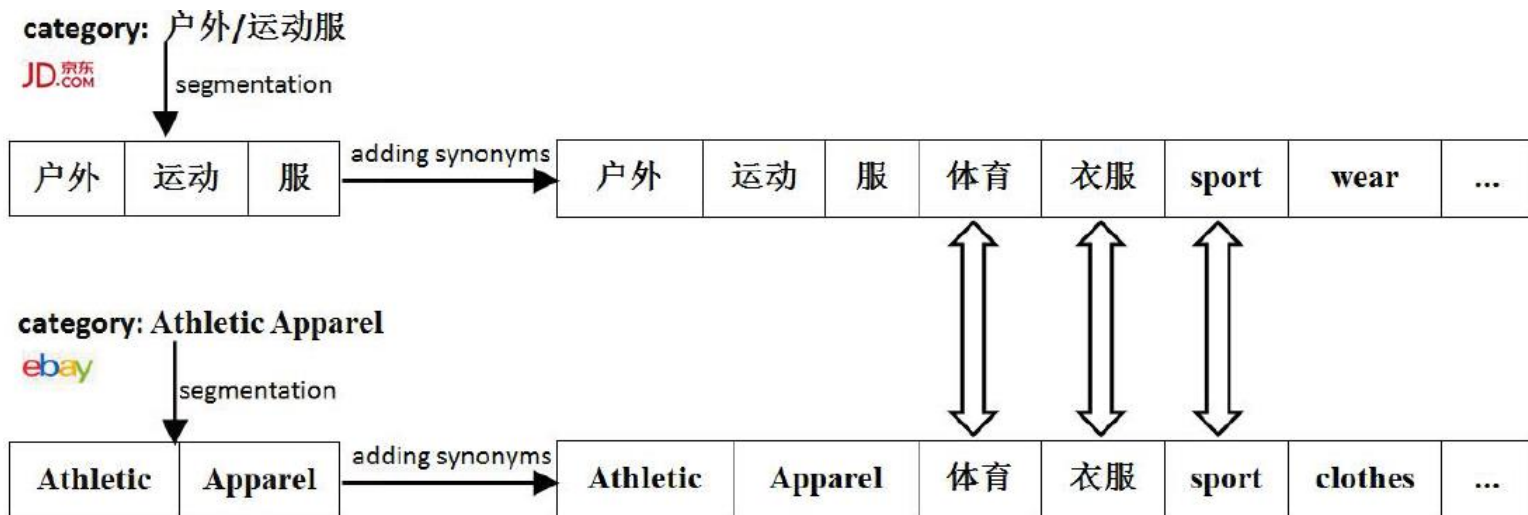
# Approach – Candidates Identification

- Given two taxonomies, we first aim to find **all the possible matched categories** in the target taxonomy for each category in the source taxonomy.
- To ensure the recall of the alignment results, we utilize a Web-scale multilingual synonym thesaurus, i.e. **BabelNet** to capture **cross-lingual string similarities** between categories **in word level**.

$$CSS(c^s, c^t) = \begin{cases} 1, & \text{if } W_{c^s} \cap W_{c^t} \neq \emptyset \\ 0, & \text{otherwise} \end{cases}$$

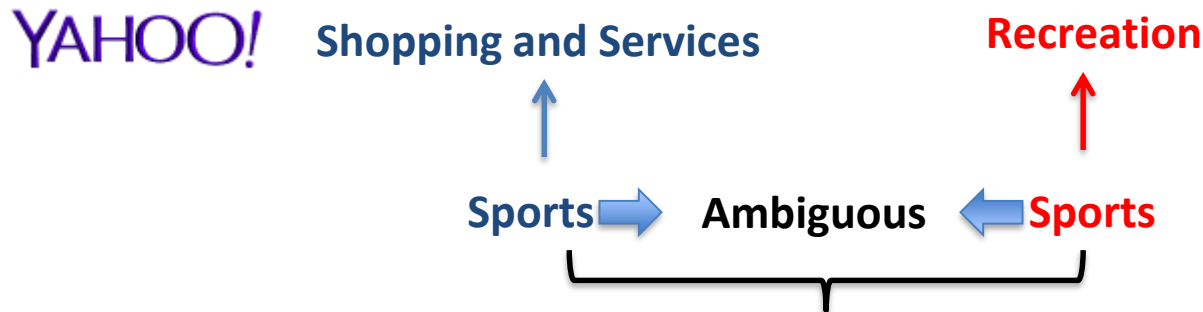
# Approach – Candidates Identification

- The key idea is “*two categories of different languages may be relevant if they share the same or synonymous words*”.



# Approach – Textual Contexts Extraction

- Before performing exact matching, we need to extract textual contexts for each category.
- Here, we choose to acquire textual information by querying the Web with the search engine **Google**.



*can not simply submit the category label to Google*



*jointly submit category “Sports” and its parents “Shopping and Services” or “Recreation” to **avoid ambiguity***

# Approach – Exact Matching

- The snippets (i.e. textual contexts) extracted from the Web are:
  - short documents
  - various in wording styles (a tweet or a piece of news)

## Sports & Recreation Center

[wp.wpi.edu/reccenter/](http://wp.wpi.edu/reccenter/) ▾ 翻译此页

Hello Everyone! Happy Winter! The Sports & Recreation Center is happy to announce WPI Family & Friends winter break time from Monday, January 4-Tuesday, ...

## Sports and Recreation — University of Leicester

[www2.le.ac.uk > Offices](http://www2.le.ac.uk/offices) ▾ 翻译此页

We have two fabulous Sports Centres both with a pool, sauna, steam room, spa pool and extensive other indoor and outdoor facilities. To find out more about ...

## Recreational Sports & Fitness | American University

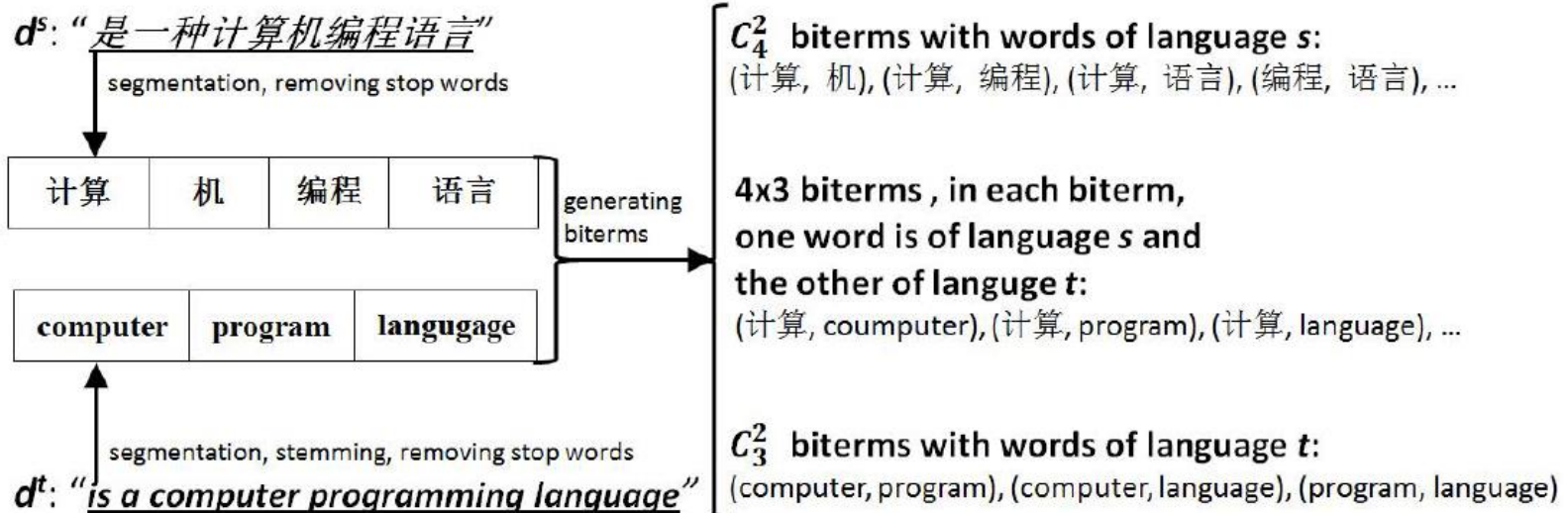
[www.american.edu/recfit/](http://www.american.edu/recfit/) ▾ 翻译此页

Recreational Sports and Fitness News: Please note the following facility closures/modified hours in December: Finals/Holiday Hours begin on Monday, ...

- The Bag-of-Words model or BiLDA (Vulić et al. 2015) may not work well in this scenario.

# Approach – Exact Matching

- We propose a new bilingual topic model, called *Bilingual Biterm Topic Model (BiBTM)* to model textual contexts.
- Different from *BTM* (Yan et al. 2013; Cheng et al. 2014), a biterm used in *BiBTM* denotes an unordered word-pair co-occurring in a pair of bilingual documents ( $d^s, d^t$ ).

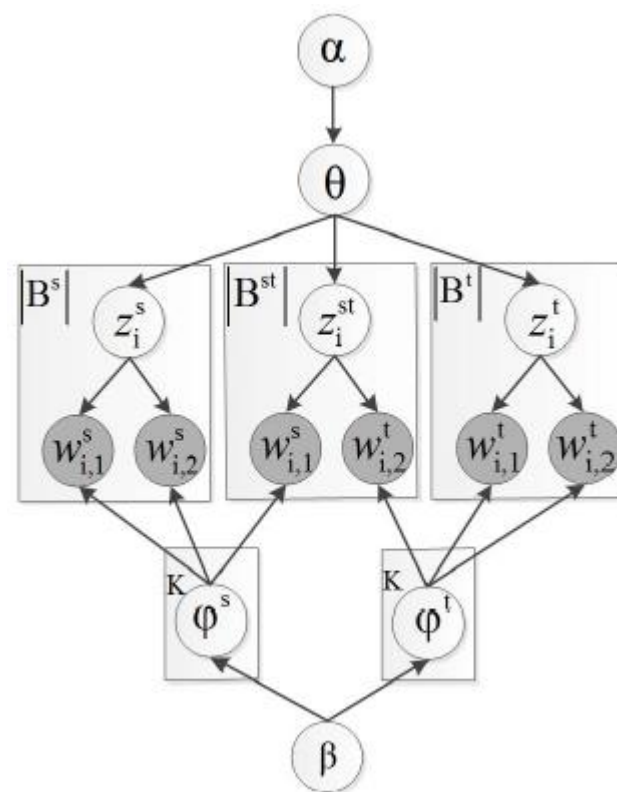




# Approach – Exact Matching

- The probability of generating the whole bilingual paired document corpus  $\mathbf{B}$  given hyperparameters  $\alpha$  and  $\beta$  can be expressed as:

$$\begin{aligned}
 P(\mathbf{B}|\alpha, \beta) &= \prod_{i=1}^{|\mathbf{B}^s|} \int \int \sum_{k=1}^K \theta_k \varphi_{k,w_{i,1}^s} \varphi_{k,w_{i,2}^s} d\theta d\varphi^s \\
 &\times \prod_{i=1}^{|\mathbf{B}^{st}|} \int \int \int \sum_{k=1}^K \theta_k \varphi_{k,w_{i,1}^s} \varphi_{k,w_{i,2}^t} d\theta d\varphi^s d\varphi^t \\
 &\times \prod_{i=1}^{|\mathbf{B}^t|} \int \int \sum_{k=1}^K \theta_k \varphi_{k,w_{i,1}^t} \varphi_{k,w_{i,2}^t} d\theta d\varphi^t
 \end{aligned}$$



# Approach – Exact Matching

- After a sufficient number of iterations, we can estimate the parameters:

$$\theta_k = \frac{\alpha + n_k}{K\alpha + |\mathbf{B}|}$$

$$\varphi_{k,w^s}^s = \frac{\beta + n_{w^s|k}}{W^s\beta + n_{\cdot^s|k}}$$

$$\varphi_{k,w^t}^t = \frac{\beta + n_{w^t|k}}{W^t\beta + n_{\cdot^t|k}}$$

- We then utilize the following formula (Yan et al. 2013) to infer the topic distribution of the context for category  $c$  and it contains  $N_c$  biterms  $\{b_j\}_{j=1}^{N_c}$ .

$$P(z|c) = \sum_{j=1}^{N_c} P(z = k|b_j)P(b_j|c)$$

# Approach – Exact Matching

- $P(b_j|c)$  is estimated by empirical distribution:

$$P(b_j|c) = \frac{n(b_j)}{\sum_{j=1}^{N_c} n(b_j)}$$

- $P(z = k|b_j)$  can be computed via Bayes' formula based on the parameters learned in *BiBTM*:

$$P(z = k|b_j) = \begin{cases} \frac{\theta_k \cdot \varphi_{k,w_{j,1}}^s \cdot \varphi_{k,w_{j,2}}^s}{\sum_{k'=1}^K \theta_{k'} \cdot \varphi_{k',w_{j,1}}^s \cdot \varphi_{k',w_{j,2}}^s}, & \text{if } b_j \in \mathbf{B}^s \\ \frac{\theta_k \cdot \varphi_{k,w_{j,1}}^s \cdot \varphi_{k,w_{j,2}}^t}{\sum_{k'=1}^K \theta_{k'} \cdot \varphi_{k',w_{j,1}}^s \cdot \varphi_{k',w_{j,2}}^t}, & \text{if } b_j \in \mathbf{B}^{st} \\ \frac{\theta_k \cdot \varphi_{k,w_{j,1}}^t \cdot \varphi_{k,w_{j,2}}^t}{\sum_{k'=1}^K \theta_{k'} \cdot \varphi_{k',w_{j,1}}^t \cdot \varphi_{k',w_{j,2}}^t}, & \text{if } b_j \in \mathbf{B}^t \end{cases}$$

- The final relevance score is computed as the cosine similarity between topic vectors.

# Experiments

- We evaluated our proposed approach on two different kinds of **real world datasets**, which are publicly available at <https://github.com/jxls080511/080424>.
- Task description:
  - **Cross-lingual Product Catalogue Alignment:**  
JD.com→eBay.com
  - **Cross-lingual Web Site Directory Alignment:**  
Chinese Dmoz.org→Yahoo Directory

# Experiments

Overall results: MRR values

Approach	JD.com→ eBay.com	Chinese Dmoz.org→ Yahoo Directory
RSVM	0.195	0.261
BiBTM	0.199	0.246
CSS+RSVM	0.210	0.301
CSS+BOW	0.423	0.489
CSS+BiLDA	0.553	0.679
CSS+BiBTM	<b>0.597</b>	<b>0.719</b>

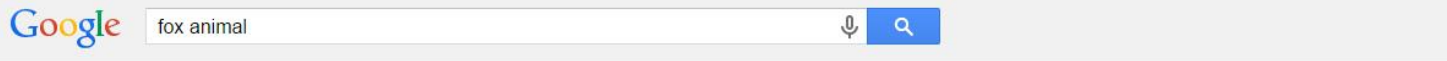




# Effective Online Knowledge Graph Fusion



# Knowledge Graph



Web Images Videos Shopping News More Search tools

About 140,000,000 results (0.36 seconds)

## Fox - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/**Fox** ▼ Wikipedia ▼  
This article is about the animal. For the American broadcast television network, see Fox Broadcasting Company. For other uses, see Fox (disambiguation).  
Canidae - Red fox - Fox Broadcasting Company - Vulpes

## Fox (Vulpes vulpes) - Animals - A-Z Animals - Animal Facts ...

a-z-animals.com/animals/fox/ ▼  
The fox is a scavenger carnivorous dog, generally found in urban city areas in the northern Hemisphere. The fox is a nocturnal mammal, meaning that the fox only ...

## Red Fox - National Geographic

animals.nationalgeographic.com/animals/.../... ▼ National Geographic Society ▼  
The red fox's resourcefulness has earned it a legendary reputation for ... Find out what National Geographic Society is doing to save animals all over the world, ...

## What to Do About Foxes : The Humane Society of the United ...


www.humanesociety.org/animal... ▼ The Humane Society of the United States ▼  
Sep 18, 2014 - There's no need to worry if you see a fox outside during the day—he is ... fox? Immediately take any pet who is bitten by any wild animal to your ...

## Fox Animal Hospital - Veterinarian In Evanston, IL USA ...

foxanimalhospital.com/ ▼  
Fox Animal Hospital is a full service veterinary medical facility, located in Evanston, Illinois. The professional and courteous staff at Fox Animal Hospital seeks to ...

## Foxes: Facts & Pictures - LiveScience

www.livescience.com/27168-foxes.html ▼

A Knowledge Card for "Fox" with a red border. It features a grid of six images of foxes in various settings. Below the images is the title "Fox" and the category "Animal". The card lists several facts: "Foxes are small-to-medium-sized, omnivorous mammals belonging to several genera of the Canidae family. Wikipedia", "Scientific name: Vulpini", "Speed: 3.7 – 8.1 mph (Red fox, Trotting)", "Rank: Tribe", "Daily sleep: 9.8 h (Red fox, In captivity)", "Mass: 4.9 – 31 lbs (Red fox, Adult), 1.5 – 3.5 lbs (Fennec fox, Adult), 7.7 lbs (Arctic fox, Adult, Male)", "Gestation period: 49 – 58 d (Red fox), 50 – 52 d (Fennec fox, In Wild), 52 d (Arctic fox)", and "Height: 9.8 – 12 in. (Arctic fox, Adult, At Shoulder), 8 in. (Fennec fox, Adult), 14 – 20 in. (Red fox, Adult, At Shoulder)".

**Fox**  
Animal

Foxes are small-to-medium-sized, omnivorous mammals belonging to several genera of the Canidae family. Wikipedia

**Scientific name:** Vulpini

**Speed:** 3.7 – 8.1 mph (Red fox, Trotting)

**Rank:** Tribe

**Daily sleep:** 9.8 h (Red fox, In captivity)

**Mass:** 4.9 – 31 lbs (Red fox, Adult), 1.5 – 3.5 lbs (Fennec fox, Adult), 7.7 lbs (Arctic fox, Adult, Male)

**Gestation period:** 49 – 58 d (Red fox), 50 – 52 d (Fennec fox, In Wild), 52 d (Arctic fox)

**Height:** 9.8 – 12 in. (Arctic fox, Adult, At Shoulder), 8 in. (Fennec fox, Adult), 14 – 20 in. (Red fox, Adult, At Shoulder)

Knowledge Card

# Knowledge Card



fox animal



Web Images Videos Maps News Explore

2 Sign in

14,800,000 RESULTS Any time

## Fox - Wikipedia, the free encyclopedia

[en.wikipedia.org/wiki/Fox](https://en.wikipedia.org/wiki/Fox)

In the wild, the typical lifespan of a fox is one to three years, although individuals may live up to ten years. Unlike many canids, foxes are not always pack animals.

## NatGeo: Red Fox - Animals - Animal Pictures - Wild Animal ...

[animals.nationalgeographic.com/animals/mammals/red-fox](https://animals.nationalgeographic.com/animals/mammals/red-fox)

Learn all you wanted to know about red foxes with pictures, videos, photos, facts, and news from National Geographic.

Facts · Video · Kids · Photos · Nutria · Red Panda

## Videos of fox animal

[bing.com/videos](https://bing.com/videos)



Fox Gene Study Could Make ANY

YouTube

Red Fox

YouTube

Fox plays with terrier - funny

YouTube

Fox

YouTube

See more videos of fox animal

## Fox (*Vulpes vulpes*) - Animals - A-Z Animals - Animal ...

[a-z-animals.com/animals/fox](https://a-z-animals.com/animals/fox)

The fox is a scavenger carnivorous dog, generally found in urban city areas in the northern Hemisphere. The fox is a nocturnal mammal, meaning that the fox only goes ...

## Fox Animal Information, Information about Fox Animal

[www.animalport.com/animal-information/fox-animal-information.html](https://www.animalport.com/animal-information/fox-animal-information.html)

Fox animal information will help us know a few facts about the life of a fox. It also briefs us about its life cycle.

## What to Do About Foxes : The Humane Society of the ...

[www.humanesociety.org/animals/foxes/tips/...problems\\_with\\_foxes.html](https://www.humanesociety.org/animals/foxes/tips/...problems_with_foxes.html)

September 18, 2014. What to Do About Foxes. Foxes out during the day are no cause for alarm—but if you need to send a fox family on their way or are worried about ...



## Fox

Foxes are small-to-medium-sized, omnivorous mammals belonging to several genera of the Canidae family. Foxes are slightly smaller than a medium-size domestic dog, with a flattened skull, upright triangular ears, a pointed, slightly upturned snout, and a long bushy tail.

[en.wikipedia.org](https://en.wikipedia.org)

Scientific name: Canidae

Body length: 0.8 - 1.3 ft (Fennec fox) · 1.5 - 3 ft on average (Red fox, Without tail) · 2.7 ft on average (Arctic fox, Female) · 2.8 ft on average (+)

Daily sleep: 9.8 hours on average (Red fox, In captivity)

Gestation period: 49 - 58 days (Red fox) · 50 - 52 days (Fennec fox, In wild) · 52 days (Arctic fox)

Height: 0.7 ft (Fennec fox) · 0.8 - 1 ft (Arctic fox) · 1.1 - 1.6 ft on average (Red fox)

Lifespan: 5 years (Red fox, In wild)

People also search for

See all (15+)



Gray wolf

Coyote

Dog

Lion

Rabbit

# Difference Between Knowledge Card and Infobox in Wikipedia

- values are usually much shorter
- provide missing data in Wikipedia
- update more often than Wikipedia

## FOX

Broadcasting television network · fox.com

The Fox Broadcasting Company, is an American commercial broadcast television network that is owned by the Fox Entertainment Group division of 21st Century Fox. Wikipedia

**Founded:** October 9, 1986, Los Angeles, CA

**CEO:** Peter Rice

**Parent organization:** [News Corporation](#)

**Founders:** Rupert Murdoch, Barry Diller



Fox Broadcasting Company

# FOX

<b>Type</b>	Broadcast television network
<b>Country</b>	United States
<b>Availability</b>	National International
<b>Founded</b>	October 1985; 30 years ago by <a href="#">Rupert Murdoch</a> and <a href="#">Barry Diller</a>
<b>Slogan</b>	<i>We Are FOX</i>
<b>Headquarters</b>	<a href="#">Los Angeles, California</a>
<b>Owner</b>	Fox Television Group (Fox Entertainment Group)
<b>Parent</b>	21st Century Fox
<b>Key people</b>	<a href="#">Dana Walden</a> <a href="#">Gary Newman</a> (co-chairpersons/co-CEOs, Entertainment)
<b>Launch date</b>	October 9, 1986 (on-air operations) April 5, 1987 (primetime launch)



# A Closer Look At Knowledge Card



Card Label

Fox

Foxes are small-to-medium-sized, omnivorous mammals belonging to several genera of the Canidae family. [Wikipedia](#)

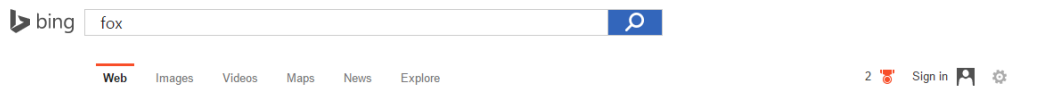
Properties

**Scientific name:** *Vulpini*  
**Speed:** 3.7 – 8.1 mph (Red fox, Trotting)  
**Rank:** Tribe  
**Daily sleep:** 9.8 h (Red fox, In captivity)  
**Mass:** 4.9 – 51 lbs (Red fox, Adult), 1.5 – 3.5 lbs (Fennec fox, Adult), 7.7 lbs (Arctic fox, Adult, Male)  
**Gestation period:** 49 – 58 d (Red fox), 50 – 52 d (Fennec fox, In Wild), 52 d (Arctic fox)  
**Height:** 9.8 – 12 in (Arctic fox, Adult, At Shoulder), 8 in. (Fennec fox, Adult), 14 – 20 in. (Red fox, Adult, At Shoulder)

Values (String, Numeric and Object)

Feedback

# A Closer Look At Knowledge Card



68,600,000 RESULTS Any time ▾

**FOX - Official Site**  
[www.fox.com](http://www.fox.com)

Full Episodes, Clips and the latest information about all of your favorite FOX shows.

#### Full Episodes

Watch full episodes of your favorite FOX shows.

#### Shows

Get the latest buzz and watch full episodes and clips from ...

#### Schedule

Check out TV listings and the full schedule of FOX ...

#### Empire

FOX Networks, Business News; Fox Movie Channel; ...

#### Bones

Check out the latest buzz on Bones Skip to main ...

[See results only from fox.com](#)

#### Glee

Official site features an overview of the show, ...

#### Fox Now

Watch Full Episodes of every FOX show on all your ...

#### Gotham

FOX Networks, Business News; Fox Movie Channel; ...

#### Family Guy

Check out the latest buzz on Family Guy Skip to ... the ...

#### New Girl

Megan Fox is heading to New Girl in Season 5 to ...

Log in

Watch

**Fox News Channel - Official Site**

[www.foxnews.com](http://www.foxnews.com) Official site

Breaking News, Latest News and Current News from FOXNews.com. Breaking news and video. Latest Current News: U.S., World, Entertainment, Health, Business, ...

See results for

#### FOX

TV Network  
The Fox Broadcasting Company, is an American commercial broadcast television network that ...



#### Fox News Channel

Fox News Channel, also known as Fox News, is an American basic cable and satellite news television channel that is owned by the Fox Entertainment Group.



#### Fox

Foxes are small-to-medium-sized, omnivorous mammals belonging to several genera of the Canidae family. Foxes are slightly smaller than a medium-size domestic dog, ...



#### Fox Sports

United States  
Fox Sports is the programming division of the Fox Broadcasting Company, owned by 21st Century Fox, t...



#### Fox Sports Networks

Fox Sports Networks, formerly known as Fox Sports Net, is the collective name for a group of regional sports networks in the United States that are primarily owned ...



More images

## Fox

Animal

Foxes are small-to-medium-sized, omnivorous mammals belonging to several genera of the Canidae family. [Wikipedia](#)

**Scientific name:** Vulpini

**Speed:** 3.7 – 8.1 mph (Red fox, Trotting)

**Rank:** Tribe

**Daily sleep:** 9.8 h (Red fox, In captivity)

**Mass:** 4.9 – 31 lbs (Red fox, Adult), 1.5 – 3.5 lbs (Fennec fox, Adult), 7.7 lbs (Arctic fox, Adult, Male)

**Gestation period:** 49 – 58 d (Red fox), 50 – 52 d (Fennec fox, In Wild), 52 d (Arctic fox)

**Height:** 9.8 – 12 in. (Arctic fox, Adult, At Shoulder), 8 in. (Fennec fox, Adult), 14 – 20 in. (Red fox, Adult, At Shoulder)

[Feedback](#)



# Meta Search Engine

**dögpilē**   [Preferences](#)

- web
- images
- video
- news
- shopping
- white pages

Ads related to: Fox

Watch **FOX** Online - Access Now. Stream "**FOX**" Online Here.  
[tvshowinsider.com](#)  
Access Now. Stream "**FOX**" Online Here. Fast & Safe Install. Watch Now.  
[Watch LIVE TV Here - Stream Any Network](#)

Robot Chicken- Adult Swim | [AdultSwim.com](#)  
[www.AdultSwim.com/RobotChicken](#)  
Tune In Sunday @ 12am For The Robot Chicken DC Comics Special on [as]  
[as] on Twitter - [as] on YouTube

**Fox** | [when.com](#)  
[when.com/Fox](#)  
Explore **Fox** Discover More on [When.com!](#)

Web Results

**FOX** - Official Site  
[www.fox.com](#)  
Full Episodes, Clips and the latest information about all of your favorite **FOX** shows.

**Fox News Channel** - Official Site  
[www.foxnews.com](#)  
Breaking News, Latest News and Current News from FOXNews.com. Breaking news and video. Latest Current News: U.S., World, Entertainment, Health, Business, Technology ...

**Fox Sports** - Official Site  
[www.foxsports.com](#)  
Find live scores, player & team news, videos, rumors, stats, standings, schedules & fantasy games on **FOX Sports**.

**Fox Racing** - Official Site  
[www.foxhead.com](#)  
**Fox Head Inc**, aka **Fox Racing**, is the most recognized and best-selling brand of mx apparel in the world today - **Fox** has been a motocross icon since 1974.

**FOX** — Redefining Ride Dynamics  
[www.ridefox.com](#)  
**FOX** redefines ride dynamics for ATVs, mountain bikes, motocross, off-road vehicles, snowmobiles, trucks, and UTVs.

**NFL on FOX** - Official Site  
[www.foxsports.com/nfl](#)  
Find live NFL scores, pro football player & team news, NFL videos, rumors, stats, standings, team schedules & fantasy football games on **FOX Sports**.

Are you looking for?

- [Fox Football](#)
- [Fox Sports](#)
- [Fox News Live](#)
- [Fox NFL Sunday](#)
- [Fox News](#)
- [Fox News Channel](#)
- [Jorja Fox](#)
- [MSNBC](#)



# Why fusing knowledge cards is important?

- get a more comprehensive summary of a given entity
- get complementary information from different resources
- get up-to-date information from the fused cards

# Challenges in Knowledge Card Fusion



## Fox

Animal

Foxes are small-to-medium-sized, omnivorous mammals belonging to several genera of the Canidae family. [Wikipedia](#)

**Scientific name:** Vulpini

**Speed:** 3.7 – 8.1 mph (Red fox, Trotting)

**Rank:** Tribe

**Daily sleep:** 9.8 h (Red fox, In captivity)

**Mass:** 4.9 – 31 lbs (Red fox, Adult), 1.5 – 3.5 lbs (Fennec fox, Adult), 7.7 lbs (Arctic fox, Adult, Male)

**Gestation period:** 49 – 68 d (Red fox), 50 – 52 d (Fennec fox, In wild), 52 d (Arctic fox)

**Height:** 9.8 – 12 in. (Arctic fox, Adult, At Shoulder), 8 in. (Fennec fox, Adult), 14 – 20 in. (Red fox, Adult, At Shoulder)

**Conflicts**

**Complementary**

**Missing**

## Fox

Foxes are small-to-medium-sized, omnivorous mammals belonging to several genera of the Canidae family. Foxes are slightly smaller than a medium-size domestic dog, with a flattened skull, upright triangular ears, a pointed, slightly upturned snout, and a long bushy tail. [en.wikipedia.org](#)

**Scientific name:** Canidae

**Body length:** 0.8 - 1.3 ft (Fennec fox) · 1.5 - 3 ft on average (Red fox, Without tail) · 2.7 ft on average (Arctic fox, Female) · 2.8 ft on average (+)

**Daily sleep:** 9.8 hours on average (Red fox, In captivity)

**Gestation period:** 49 - 58 days (Red fox) · 50 - 52 days (Fennec fox, In wild) · 52 days (Arctic fox)

**Biological classification:** Tribe

**Consists of:** Red fox · Fennec fox · Vulpes · Arctic fox · Urocyon · Alopex

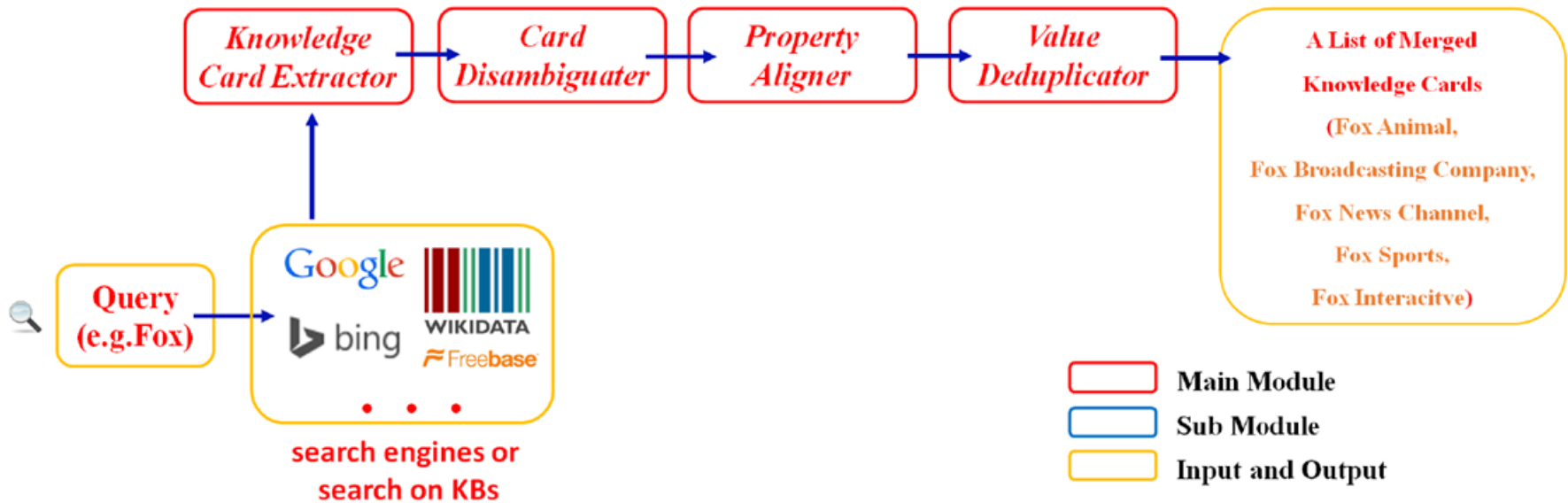
# Challenges

- Ambiguous cards from a same query
- Same value but different expressions
- One to many mappings

# The Relationship Between Ontology Alignment and Knowledge Graph Fusion

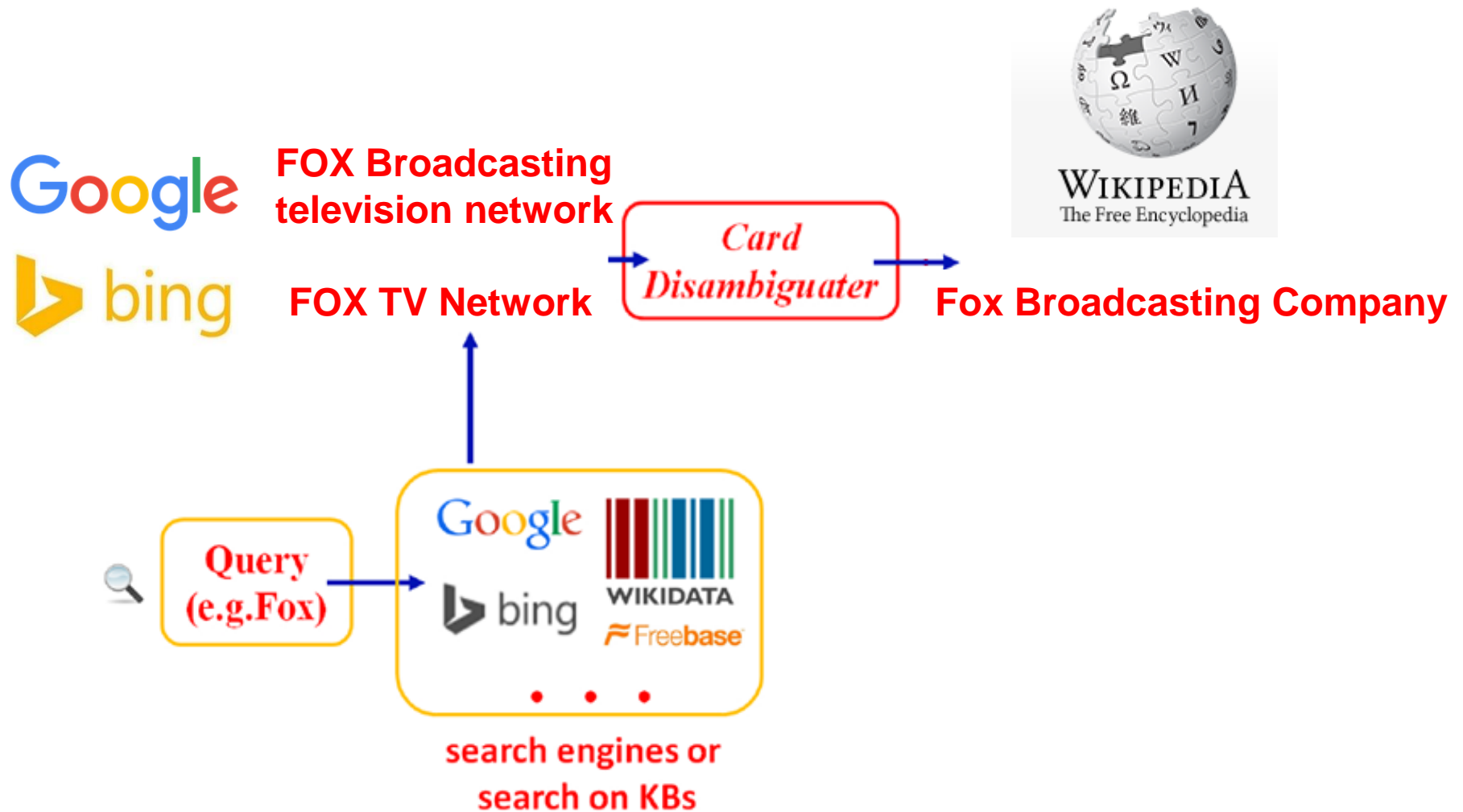
- New branch of Ontology Alignment
- Knowledge cards are fused online only when a query is submitted
- Lack of schema-level information
- Each card only contains a limited number of information
- Instances may use different expressions

# Workflow





# Workflow - Card Disambiguation



# Card Disambiguation

## – commonness computing

$$commonness(m, e) = \frac{|L_{m,e}|}{\sum_{e'} |L_{m,e'}|}$$

Mention: Fox

Candidate Entity:

- (1) Fox Broadcasting Company
- (2) Fox
- (3) NFL on Fox
- .....

0.7629  
0.049  
0.021

**FOX**

Broadcasting television network · [fox.com](http://fox.com)

The Fox Broadcasting Company, is an American commercial broadcast television network that is owned by the Fox Entertainment Group division of 21st Century Fox. [Wikipedia](#)

**Founded:** October 9, 1986, Los Angeles, CA

**CEO:** [Peter Rice](#)

**Parent organization:** [News Corporation](#)

**Founders:** [Rupert Murdoch](#), [Barry Diller](#)



# Card Disambiguation

## – relatedness computing

$$\text{relatedness}(e, v) = 1 - \frac{\log(1 + \max(|L_e|, |L_{e_v}|)) - \log(1 + |L_e \cap L_{e_v}|)}{\log(|WP|) - \log(1 + \min(|L_e|, |L_{e_v}|))}$$

Card: Fox

Object Values:

(1) Los Angeles

0.9142

(2) Peter Rice

0.8142

(3) Barry Diller

0.9774

.....

FOX

Broadcasting television network ·  
[fox.com](http://fox.com)

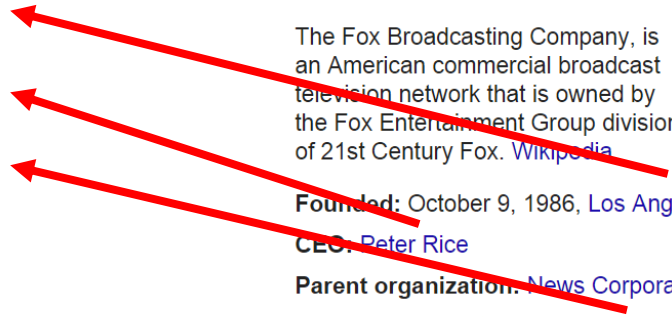
The Fox Broadcasting Company, is an American commercial broadcast television network that is owned by the Fox Entertainment Group division of 21st Century Fox. [Wikipedia](#)

**Founded:** October 9, 1986, Los Angeles, CA

**CEO:** [Peter Rice](#)

**Parent organization:** [News Corporation](#)

**Founders:** [Rupert Murdoch](#), [Barry Diller](#)



# Card Disambiguation - Combining

Card: Fox

Object Values:

- (1) Los Angeles 0.9142
- (2) Peter Rice 0.8142
- (3) Barry Diller 0.9774

**Fox Broadcasting Company**

From Wikipedia, the free encyclopedia

*"FOX" and "Fox Network" redirect here. For "Fox" television channels outside the United States, see Fox (disambig). For the film studio that the network was named after, see 20th Century Fox. For other uses, see Fox (disambiguation).*

**This article has multiple issues. Please help improve it or discuss these issues on the talk page.**

- This article **needs additional citations for verification.** (*August 2013*)
- This article **possibly contains original research.** (*August 2013*)

The **Fox Broadcasting Company**<sup>[a]</sup> (commonly referred to as **Fox**, stylized as **FOX**,<sup>[b]</sup><sup>[c]</sup>) is an American commercial broadcast television network that is owned by the Fox Entertainment Group division of 21st Century Fox. The network is headquartered on Pico Boulevard in Los Angeles, with additional major offices and production facilities at the Fox Television Center in Los Angeles and in Manhattan. It is the third largest major television network in the world, based on total revenues, assets and international coverage.

Launched on October 9, 1986 as a competitor to the three longer-established U.S. commercial television networks, ABC, NBC and CBS, Fox went on to become the most successful attempt at a fourth television network. It was the highest-rated broadcast network in the 18–49 demographic from 2004 to 2012, and earned the position as the most-watched American television network in total viewership during the 2007–08 season.<sup>[d]</sup><sup>[e]</sup>

Fox and its affiliated companies operate many entertainment channels in international markets, although these do not necessarily air the same programming as the U.S. network. Most viewers in Canada have access to at least one U.S.-based Fox affiliate, either over-the-air or through a pay television provider; although Fox's National Football League telecasts and most of its prime time programming are subject to simultaneous substitution regulations for cable and satellite providers imposed by the Canadian Radio-television and Telecommunications Commission (CRTC) to protect rights held by domestically based networks.

The network is named after sister company 20th Century Fox, and indirectly for producer William Fox, who founded one of the movie studio's predecessors, Fox Film. Fox is a member of the North American Broadcasters Association and the National Association of Broadcasters.

**Contents** [hide]

1 History

1.1 Origin

1.2 1980s: Establishment of the network

1.2.1 Foundations

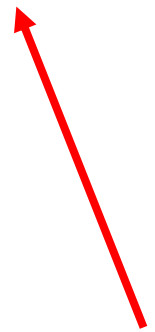
1.2.2 Beginning of the network

**Fox Broadcasting Company**

<b>Type</b>	Broadcast television network
<b>Country</b>	United States
<b>Availability</b>	National, International
<b>Founded</b>	October 9, 1986; 30 years ago
	By Rupert Murdoch and Barry Diller
<b> slogan</b>	<i>Fx Are FOX</i>
<b>Headquarters</b>	Los Angeles, California
<b>Owner</b>	Fox Television Group (Fox Entertainment Group)
<b>Parent</b>	21st Century Fox

$$relatedness(e, V_o) = 1 - \prod_{v \in V_o} (1 - relatedness(e, v))$$

$$\hat{e} = \arg \max_e (commonness(m, e) \times relatedness(e, V_o))$$



**FOX**

Broadcasting television network · fox.com

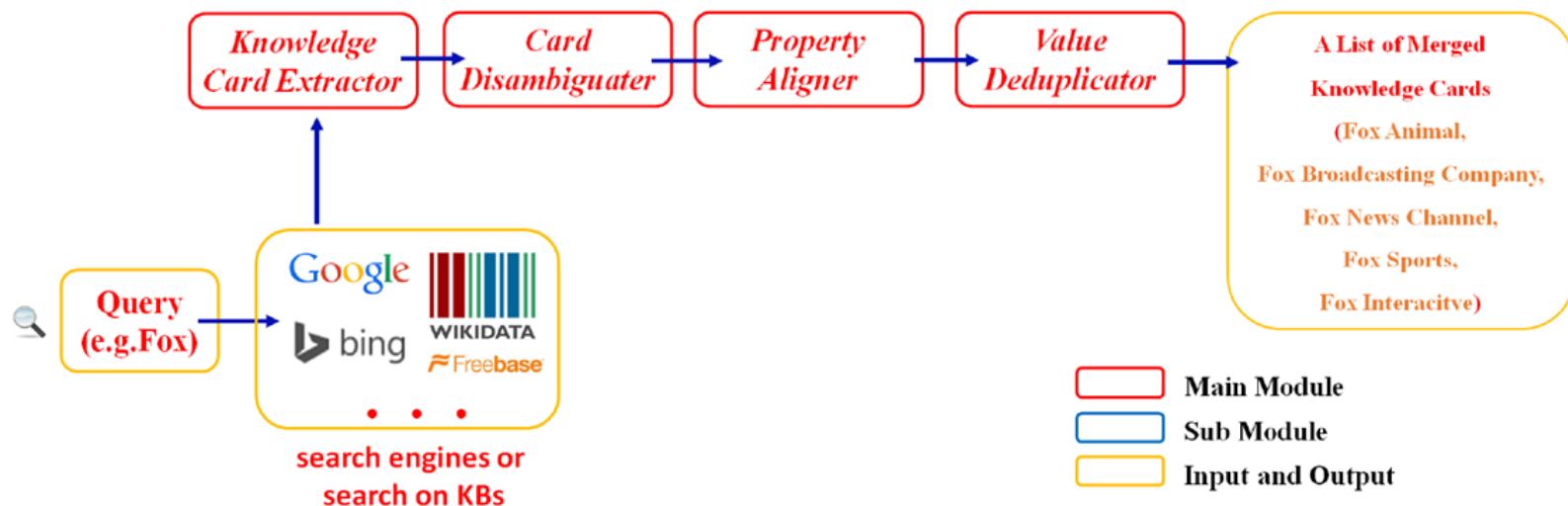


The Fox Broadcasting Company, is an American commercial broadcast television network that is owned by the Fox Entertainment Group division of 21st Century Fox. [Wikipedia](#)

- Founded:** October 9, 1986, Los Angeles, CA
- CEO:** Peter Rice
- Parent organization:** News Corporation
- Founders:** Rupert Murdoch, Barry Diller

# Workflow - Property Alignment

- a learning-based method with a post-processing to predict whether two properties can be aligned.



# Property Alignment

Google

Founded: October 9,  
1986, Los Angeles,  
CA

bing

Place Founded: Los  
Angeles, CA

Property  
Aligner

Founded (Place Founded)

Card  
Disambiguator



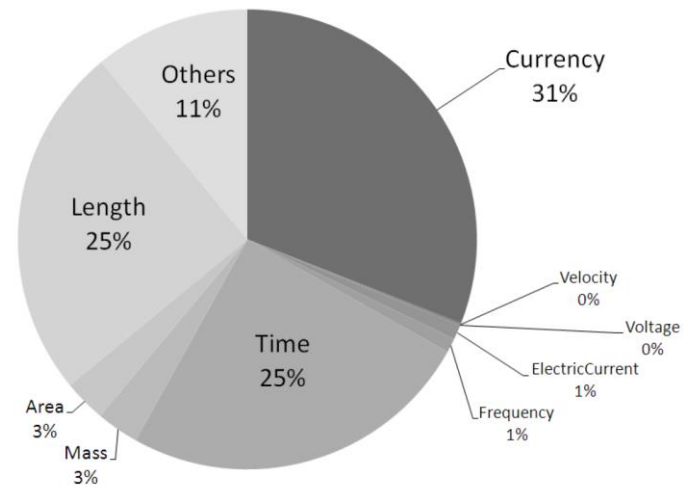
Fox Broadcasting Company



# Property Alignment

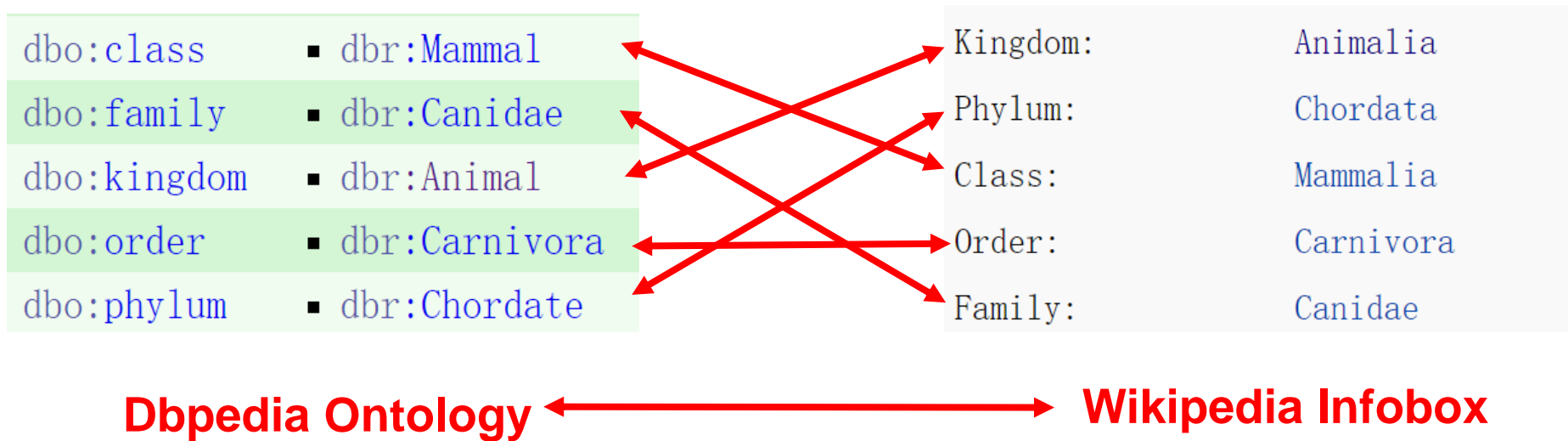
## ● Data Normalization

- object values (link completion)
- string/numeric values



# Property Alignment

## ● Training Data



# Property Alignment

## ● Feature Extraction

- Property Similarity (PS) Similarity ----- Property-based
  - Value Overlap Ratio (VOR)
  - Value Match Ratio (VMR)
  - Value Similarity Variance (VSV)
- } Value-based Similarity

# Post - Processing

- **Property mutual exclusion filtering**
  - A property is disjoint with any other property in the same card, and cannot be aligned with any of them.
- **Object value range validation**
  - If two properties can be aligned, their ranges should be compatible.

# Workflow - Value Deduplication

- groups equivalent values of aligned properties into value clusters.



**Founded (Place Founded):**  
October 9, 1986, Los  
Angeles, CA; Oct  
1985 · Los Angeles, CA

**Founded: October 9, 1986**  
**(Oct 1985), Los Angeles, CA**

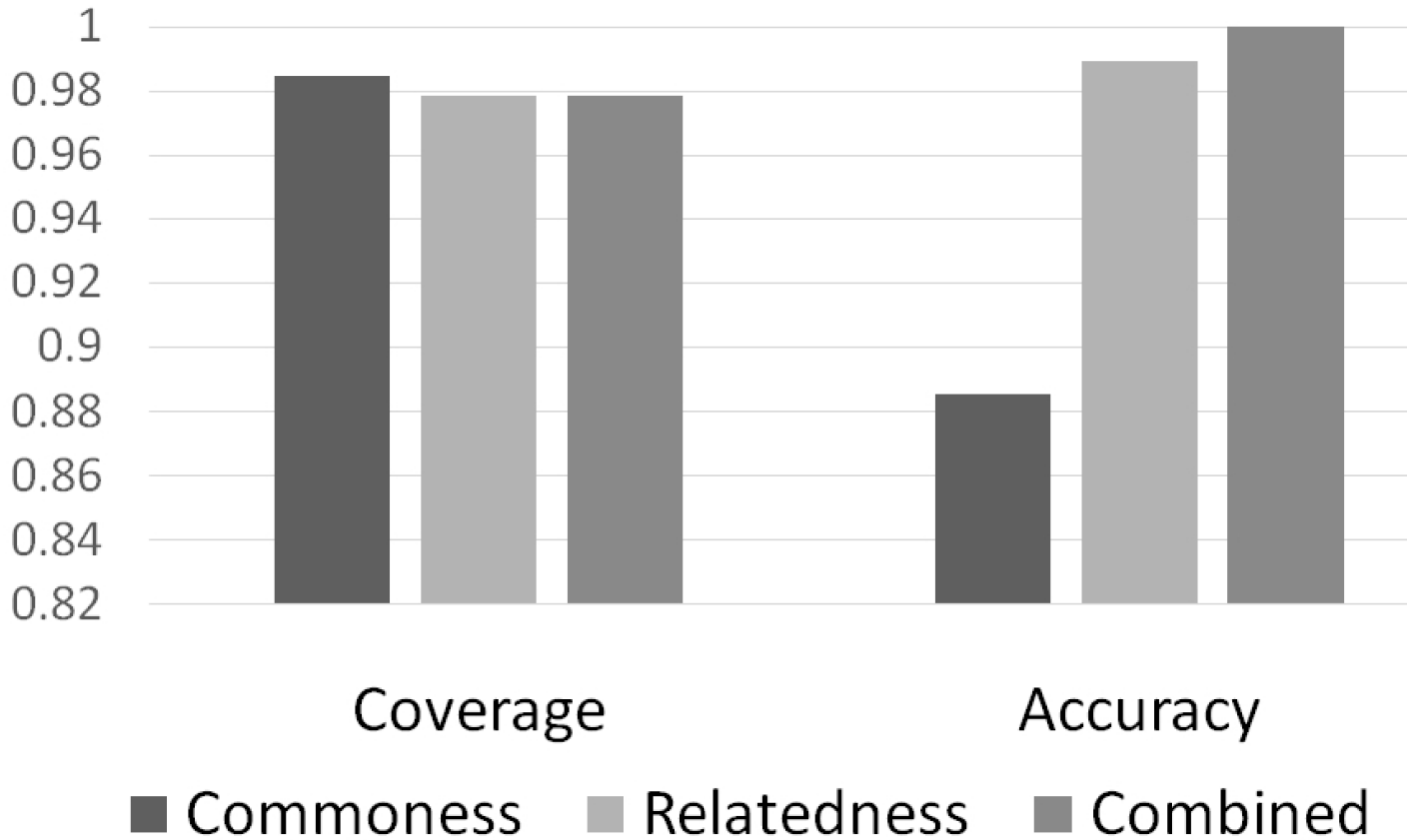
# Experiment - Setup

- select 154 queries which includes both ambiguous and disambiguous queries.
- 464 knowledge cards have been collected
- we randomly selected 3,487 attribute value pairs and manually label whether two properties in each pair can be aligned. As a result, 480 pairs are positive and 3,007 are negative.

<http://kcf.hiekn.com/download/experiment.tar.gz>



# Experiment - Card Disambiguation



# Experiment - Property Alignment

- All features have some positive impacts to **boost the performance** of property alignment.
- The model learned by **Random Forest** achieves the best F1-Measure, which is chosen as our learning model with 6000 training data only.
- The model especially with pre-processing can actually predict aligned property pairs with a relatively high precision (**more than 0.8**) and almost perfect recall.
- After post-processing, the precision **increases significantly** at the expense of a slight drop of recall.

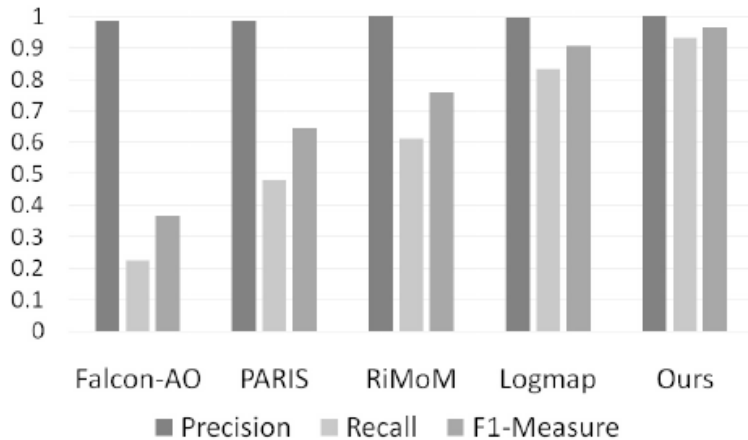
# Ontology Alignment Tools

- Selected Tools:

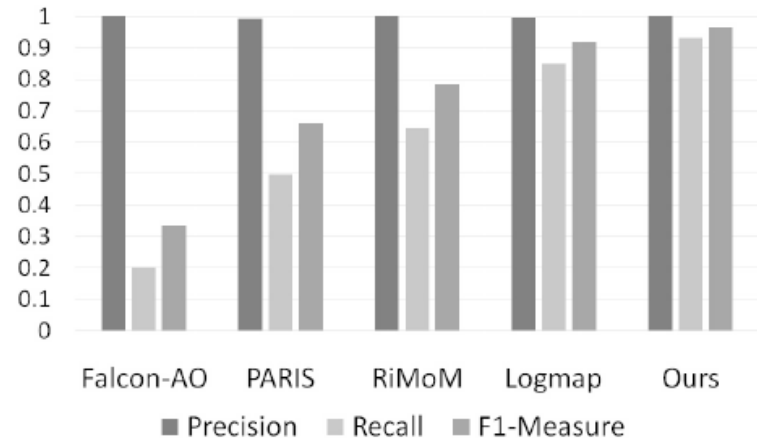
RiMoM , Logmap, Falcon-AO, and PARIS

- The former two are among the top-3 tools of the OAEI campaign<sup>8</sup> in recent three years.
- The latter two also support both alignment tasks and have been widely used in practical applications.

# Experiment - Instance Alignment



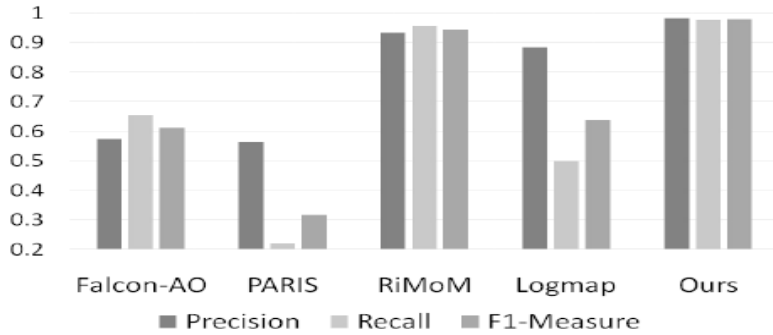
(a) Original ontology



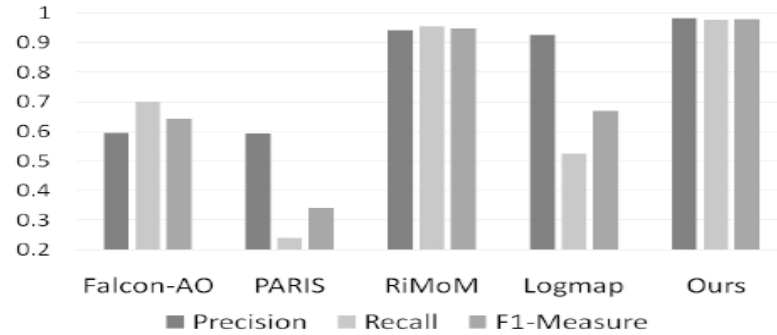
(b) + value and unit normalization

- OA tools **fail to complete missing links** of object values, which can actually reduce the ambiguities, and thus have positive impacts on instance alignments.
- **Unit conversion** is still out of its ability scope.

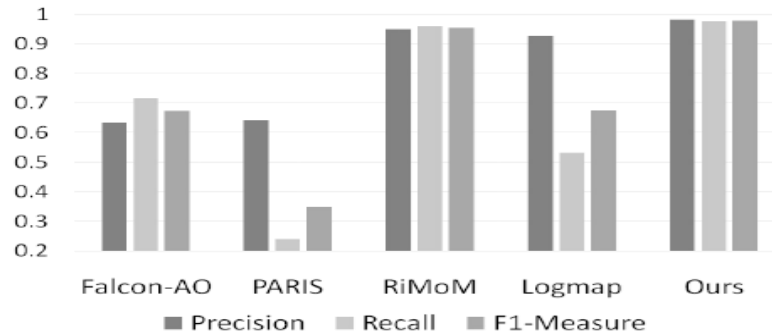
# Experiment - Property Alignment



(a) Original ontology



(b) + subject link completion



(c) + value and unit normalization

- **Dealing with one-to-many property alignments** is the gap between OA tools with ours.

**E.g.** Born: July 18, 1918, Mvezo, South Africa; Die: December 5, 2013, Houghton

Estate, Johannesburg, South Africa“ **VS.** Lived: Jul 18, 1918 - Dec 05, 2013 (age 95)

# References

- An effective rule miner for instance matching in a web of data, CIKM 2012
- Cross-lingual Taxonomy Alignment with Bilingual Biterm Topic Model, AAAI 2016
- Effective Online Knowledge Graph Fusion, ISWC 2015



**Thank you!**  
**Any questions?**

Dr. Haofen Wang

East China University of Science and Technology (ECUST)

