



Error Correction in Knowledge Graph

Piyawat Lertvittayakumjorn

14 March 2017

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

Italy

Michelangelo

Contents

Error Correction in Knowledge Graph

- Overview
- Error Classes
- Range Violation Errors
- Problem Formulation
- Discussion
- Definition & Characteristics
- Knowledge Graph Projects
- Strengths & Limitations
- Motivation & Related Concepts
- Standards
- Research Areas

Contents

Error Correction in Knowledge Graph

- Definition & Characteristics
- Knowledge Graph Projects
- Strengths & Limitations
- Motivation & Related Concepts
- Standards
- Research Areas



Japan



ทั้งหมด

ค้นรูป

แผนที่

วิดีโอ

ข่าวสาร

เพิ่มเติม

การตั้งค่า

เครื่องมือ

ผลการค้นหาประมาณ 2,320,000,000 รายการ (0.57 วินาที)

Japan - Wikipedia

<https://en.wikipedia.org/wiki/Japan> ▾ [แปลหน้านี้](#)

Japan is a sovereign island nation in Eastern Asia. Located in the Pacific Ocean, it lies off the eastern coast of the Asia Mainland (east of China, Korea, Russia) ...

[Shinto](#) · [History of Japan](#) · [Japanese people](#) · [Geography of Japan](#)

japan-guide.com - Japan Travel and Living Guide

www.japan-guide.com/ ▾ [แปลหน้านี้](#)

Everything about modern and traditional **Japan** with emphasis on travel and living related information.

[Tokyo](#) · [2017 Cherry Blossom Forecast](#) · [Sign in](#) · [Friends](#)

Japan | World news | The Guardian

<https://www.theguardian.com/world/japan> ▾ [แปลหน้านี้](#)

Inflation returns to **Japan** for the first time in more than a year. Published: 2 Mar ... [Chris Grayling tells](#)

[.Japan](#): some of our trains are superior to yours. [Transport](#) ...



ประเทศญี่ปุ่น

ประเทศ ใน เอเชียตะวันออก

ญี่ปุ่น มีชื่อทางการคือประเทศญี่ปุ่น เป็นประเทศหมู่เกาะในภูมิภาคเอเชียตะวันออก ตั้งอยู่ในมหาสมุทรแปซิฟิก ทางตะวันตกติดกับคาบสมุทรเกาหลี และสาธารณรัฐประชาชนจีน โดยมีทะเลญี่ปุ่นกั้น ส่วนทางทิศเหนือ ติดกับประเทศรัสเซีย มีทะเลโอค็อตสค์ เป็นเส้นแบ่งแดน ... [วิกิพีเดีย](#)

เมืองหลวง: โตเกียว

รหัสโทรศัพท์: +81

พื้นที่: 377,962 ตร.กม.

สกุลเงิน: เยน

ประชากร: 127.3 ล้าน (พ.ศ. 2556) ธนาคารโลก

จักรพรรดิ: สมเด็จพระจักรพรรดิอากิฮิโตะ

จุดท่องเที่ยว

ดูอีกกว่า 15 รายการ



ภูเขาฟูจิ



วัดคิงกะกุ



วัดเซ็นโซ



วัดคิโยะมิซุ



วัดโทได

ความคิดเห็น

นายกญี่ปุ่น

ทั้งหมด คำนรูป วิดีโอ ข่าวสาร แผนที่ เพิ่มเติม การตั้งค่า เครื่องมือ

ผลการค้นหาประมาณ 485,000 รายการ (0.77 วินาที)

ประเทศญี่ปุ่น / นายกรัฐมนตรี

ชินโซ อะเบะ

ตั้งแต่ พ.ศ. 2555



ชินโซ อะเบะ เป็นนายกรัฐมนตรีคนที่ 90 ของประเทศญี่ปุ่นและเป็นคนญี่ปุ่นที่เกิดหลังสงครามโลกครั้งที่ 2 คนแรกที่ได้ดำรงตำแหน่งนายกรัฐมนตรี [วิกิพีเดีย](#)

ข้อมูลเพิ่มเติมเกี่ยวกับ ชินโซ อะเบะ

ความคิดเห็น

Google's Knowledge Graph

What is the knowledge graph ?

- Knowledge graph is a graph of knowledge
- Each node represents an entity
- Each edge represents a relationship



ประเทศญี่ปุ่น

ประเทศ ใน เอเชียตะวันออก

ญี่ปุ่น มีชื่อทางการคือประเทศญี่ปุ่น เป็นประเทศหมู่เกาะในภูมิภาคเอเชียตะวันออก ตั้งอยู่ในมหาสมุทรแปซิฟิก ทางตะวันตกติดกับคาบสมุทรเกาหลี และสาธารณรัฐประชาชนจีน โดยมีทะเลญี่ปุ่นกั้น ส่วนทางทิศเหนือ ติดกับประเทศรัสเซีย มีทะเลโอค็อตสค์ เป็นเส้นแบ่งแดน ... [วิกิพีเดีย](#)

เมืองหลวง: โตเกียว

รหัสโทรศัพท์: +81

พื้นที่: 377,962 ตร.กม.

สกุลเงิน: เยน

ประชากร: 127.3 ล้าน (พ.ศ. 2556) หนาแน่นโลก

จักรพรรดิ: สมเด็จพระจักรพรรดิอากิฮิโตะ

จุดท่องเที่ยว

ดูอีกกว่า 15 รายการ



ภูเขาฟูจิ



วัดคิงกะกุ



วัดเซ็นโซ



วัดคิโยมิซุ



วัดโทได

ความคิดเห็น

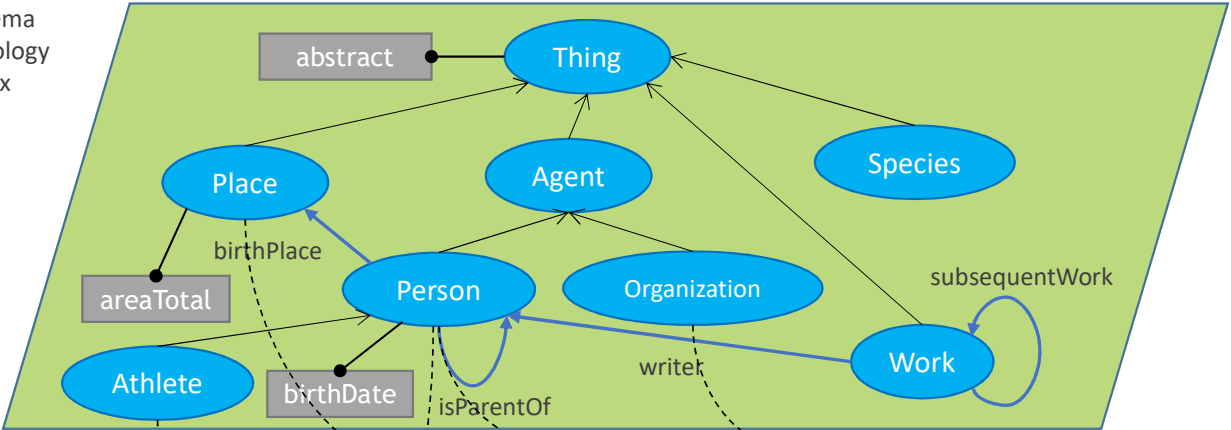
Let's construct the knowledge graph

What is the knowledge graph ?

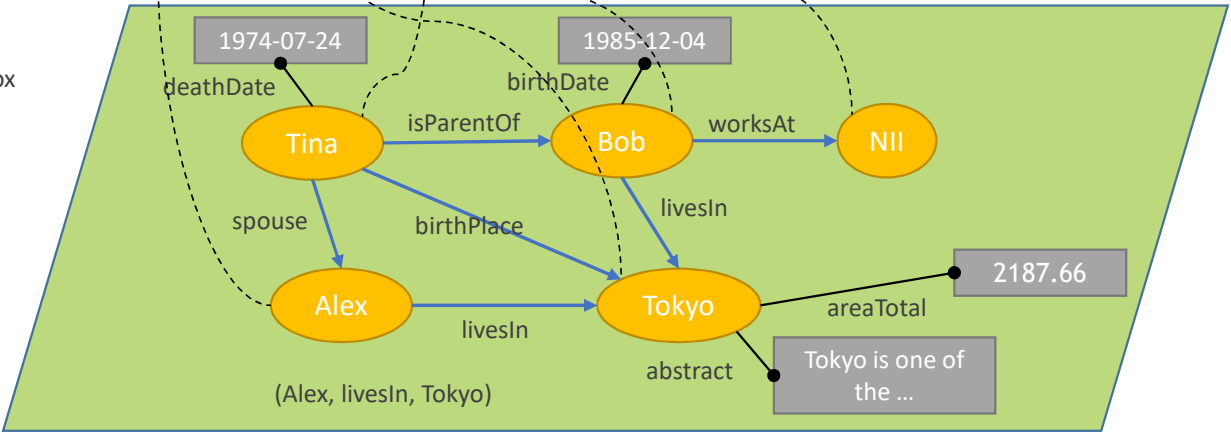
- Knowledge graph (KG) is a graph of knowledge
- Each node represents an entity
- Each edge represents a relationship
- Each triple represents one knowledge
 - <Subject, Predicate, Object>
 - Ex. <Japan, capitalCity, Tokyo>, <Japan, area, 377962 km²>
 - Subject must be an entity
 - Object can be an entity or a literal (value)







Ontology: a specification of a conceptualization.

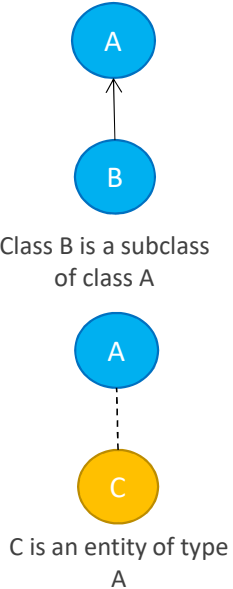
Schema
Ontology
T-box



KG
A-box



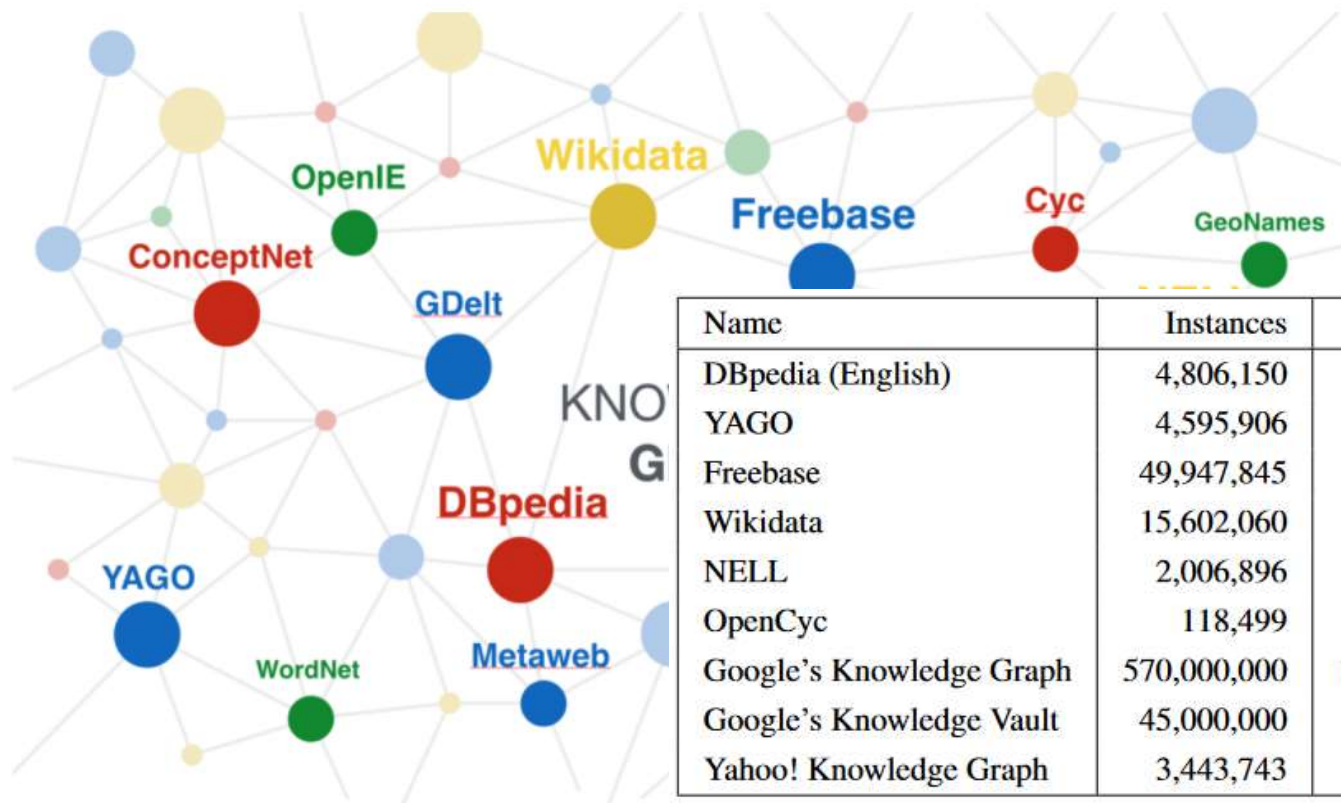
-  - Entity
-  - Instance / Object
-  - Relations
-  - ObjectProperty
-  - DatatypeProperty
-  - Literal



Characteristics of Knowledge Graph

- Directed Graph
- Attributed (Node-labeled, Edge-labeled)
- Dynamic Graph

Knowledge Graph Projects



Knowledge Graph Projects

- [NELL](#): Never-Ending Language Learning (Read the web)
- [Cyc](#) & [OpenCyc](#): The world's common sense knowledge base
- [Freebase](#): A public, editable knowledge graph with schema template
- [GDELT](#): monitors the world's broadcast, print, and web news
- [DBPedia](#): extracted from structured data in Wikipedia
- [WordNet](#) & [ConceptNet](#): lexical database and semantic network
- [YAGO](#): derived from Wikipedia, WordNet, and GeoNames.
- WikiData, Google's KV, Yahoo, Facebook, OpenIE, Microsoft, DeepDive, Prospera, ...

Advantages of Knowledge Graph

- Flexible
- Machine Understandable
- Suitable for
 - Question-Answering
 - Data Exploration


us president's birthdays




ทั้งหมด ค้นรูป วิดีโอ ข่าวสาร แผนที่เพิ่มเติม การตั้งค่า เครื่องมือ

ผลการค้นหาประมาณ 4,250,000 รายการ (0.73 วินาที)

ดอนัลด์ ทรัมป์ / วันเกิด

14 มิถุนายน 2489 (อายุ 70)



 เมลาเนีย ทรัมป์ 26 เมษายน 2513	 ฮิลลารี คลินตัน 26 ตุลาคม 2490	 บารัค โอบามา 4 สิงหาคม 2504
--	--	---

ความคิดเห็น

Limitations of Knowledge Graph (1)

- One Triple = <Subject, Predicate, Object>
= S+V+O
= One Knowledge

Limitations of Knowledge Graph (2)

- But, not all knowledges can be represented with S+V+O
 - S + V Ex. Balloons can fly
 - S + not + V + O Ex. Muslims do not eat pork
 - S + V + O + T Ex. Thailand's PM was Ms.Yingluck during Aug 2011 - May 2014
 - S + V + O₁ + O₂ Ex. Germany paid 64 billions dollars to The Allies of World War I
 - Sequences Ex. Cooking Recipes
 - ...

Related Terms and Motivation

- Semantic Web, Web of Data, Linked Data

I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web - the content, links, and transactions between people and computers. A "Semantic Web", which makes this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The "intelligent agents" people have touted for ages will finally materialize.

Weaving the Web, Tim Berners-Lee (1999)

Semantic Web

- Web metadata

```
<meta name="keywords" content="computing, computer studies, computer" />
<meta name="description" content="Cheap widgets for sale" />
<meta name="author" content="John Doe" />
```

- Semantic HTML

- `<i>` → ``
- ``
- Semantic tags such as `section`, `article`, `footer`, `progress`, `nav`, `aside`, `mark`, and `time`.

Linked Data

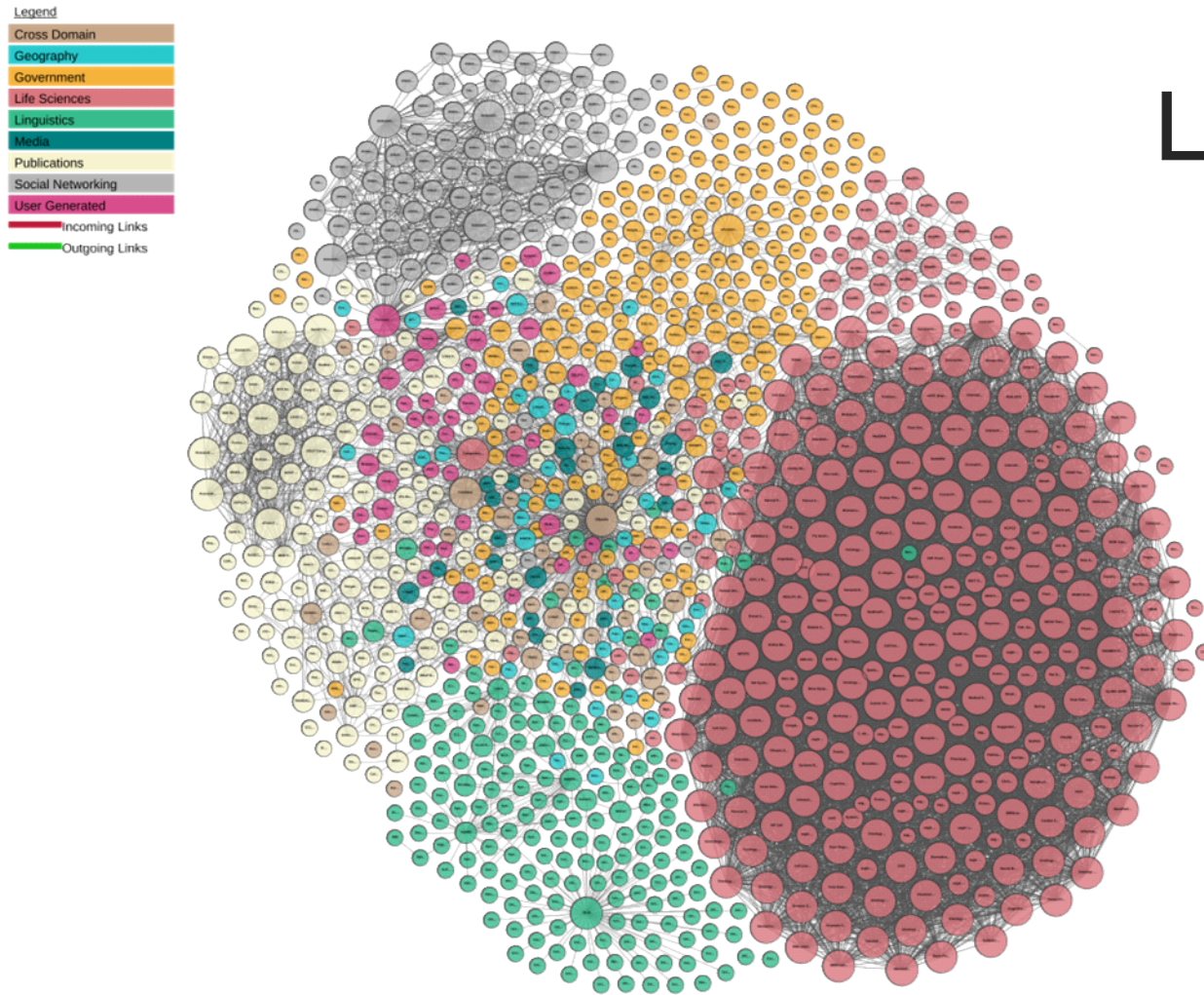
- Publishing and connecting structured data on the web

Avatar

Director: James Cameron (born August 16, 1954)
Science fiction [Trailer](#)

```
1 <div vocab="http://schema.org/" typeof="Movie">
2   <h1 property="name">Avatar</h1>
3   <div property="director" typeof="Person">
4     Director: <span property="name">James Cameron</span>
5     (born <time property="birthDate" datetime="1954-08-16">August 16, 1954</time>)
6   </div>
7   <span property="genre">Science fiction</span>
8   <a href="../movies/avatar-theatrical-trailer.html" property="trailer">Trailer</a>
9 </div>
```

Linked Open Data



"Linking Open Data cloud diagram 2017, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cyganiak. <http://lod-cloud.net/>"

Linked Data Principle

- Use URIs (Uniform Resources Identifiers) as names for things
 - Use HTTP URIs so that people can look up those names
 - When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
 - Include links to other URIs to allow discovering more things
-
- Ex. <http://data.linkedmdb.org/resource/film/77>

Standards

- RDF (Resource Description Framework)
 - The data model of semantic web
 - “Graph” : Nodes, Edges, Triples <Subject, Predicate, Object>
 - Triples are organized into named graphs, forming quads (4).
- URI (Universal Resource Identifier)
 - Universal identities for things (consistent across databases)
 - In RDF, resources and edges are URIs.
 - Might be written in shorthand
 - Ex. foaf:name = <http://xmlns.com/foaf/0.1/name>

Standards

- RDFS (RDF Schema)
 - Define common vocabularies for RDF based on the Object-oriented concept
 - `rdf:type`, `rdf:Class`, `rdfs:Property`, `rdfs:Domain`, `rdfs:Range`, `rdfs:subClassOf`, `rdfs:subPropertyOf`
 - `rdfs:label`, `rdfs:comment`, `rdfs:seeAlso`

Standards

- OWL (Web Ontology Language)
 - The ontology (schema) language of the Semantic Web
 - Efficient automated reasoning, fast, and flexible
 - Larger than RDFS
 - Datatype: owl:real, xsd:decimal, xsd:negativeInteger, ...
 - Class, DatatypeProperty, ObjectProperty, ...
 - ClassExpressions: Intersection, Union, Cardinality, Existential, ...
 - Axioms: subclass, equivalent, disjoint, irreflexive, functional, ...
 - ...

Standards

- SPARQL (SPARQL Protocol and RDF Query Language)
 - Enables querying the Semantic Web
 - Consists of two parts: query language and protocol

```
SELECT ?s1 ?s2
WHERE {
?s1 dbo:parent ?s2.
?s2 rdf:type dbo:Scientist.
} LIMIT 3 OFFSET 20
```

SPARQL results:

s1	s2
:Amy_Chua	:Leon_O._Chua
:Robert_Tappan_Morris	:Robert_Morris_(cryptographer)
:Gwen_Raverat	:George_Darwin

HTTP/1.1 200 OK

Date: Mon, 21 May 2012 23:43:38 GMT

Content-Type: application/sparql-results+xml; charset=UTF-8

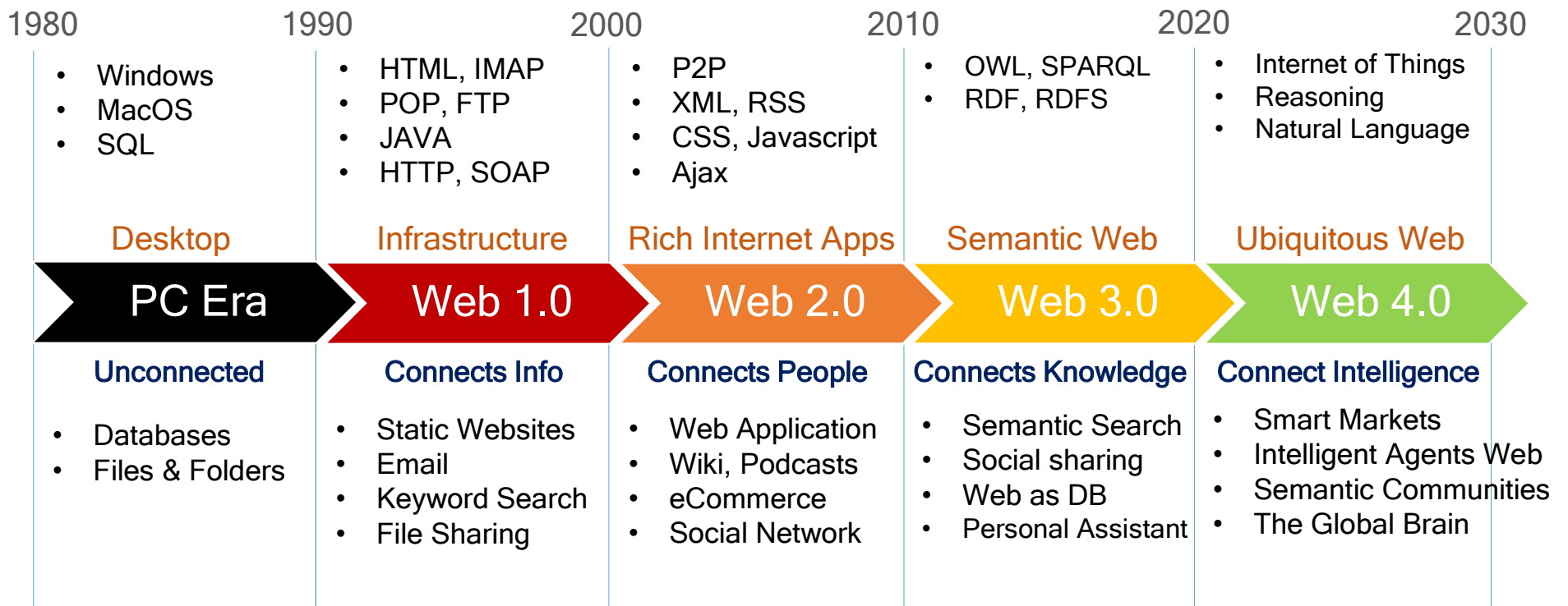
Connection: keep-alive

Server: Virtuoso/06.04.3132 (Linux) x86_64-generic-linux-glibc25-64 VDB

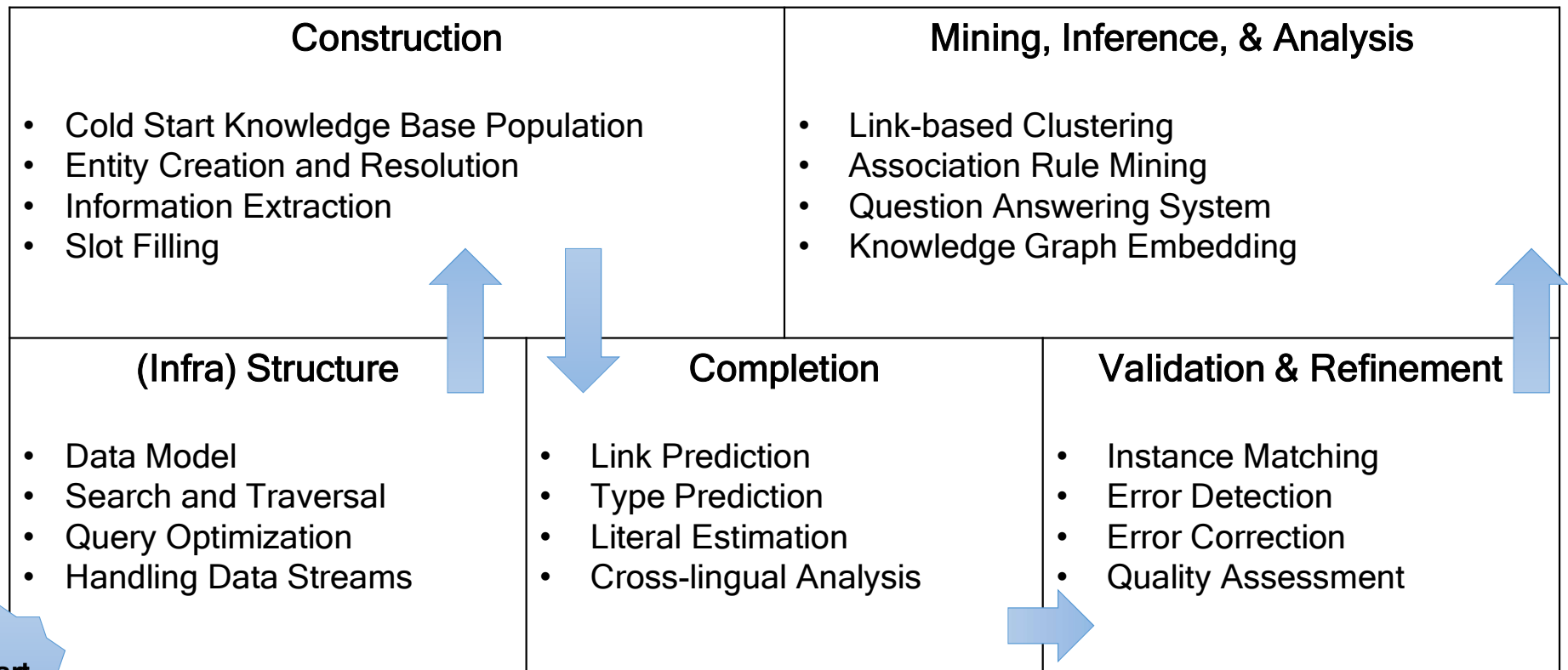
Content-Length: 96743

Accept-Ranges: bytes

Web Technologies Timeline



Research Areas of Knowledge Graph



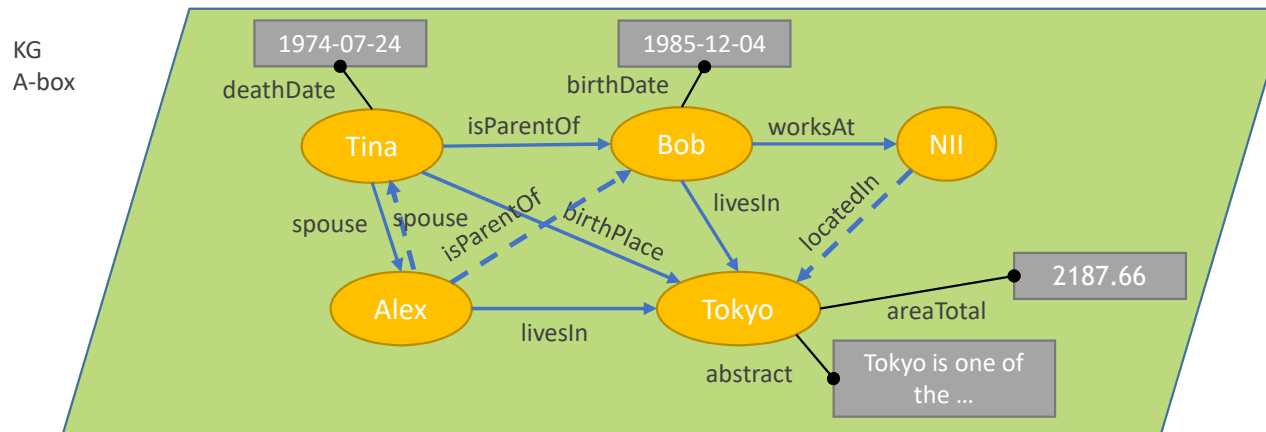
Start

Contents

Error Correction in Knowledge Graph

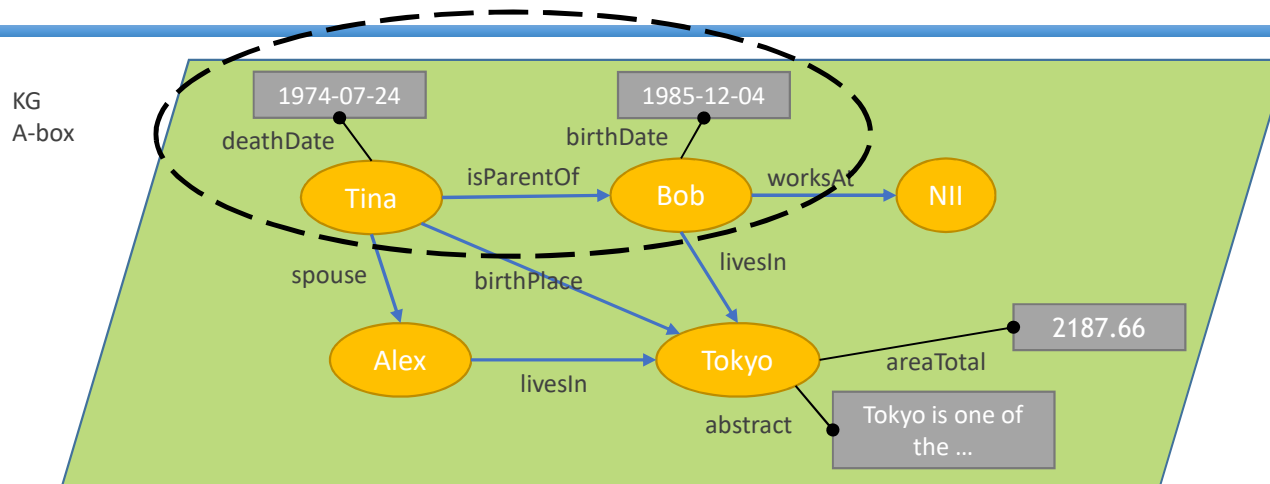
- Overview
- Error Classes
- Range Violation Errors
- Problem Formulation
- Discussion

KG is not complete



- Knowledge Graph Completion
- Rules Learning : Given current knowledge graph, learn ...
 - $(A, \text{spouse}, B) \text{ :- } (B, \text{spouse}, A)$ 100%
 - $(A, \text{isParentOf}, B) \text{ :- } (A, \text{spouse}, C), (C, \text{isParentOf}, B)$ 85%
 - $(A, \text{locatedIn}, B) \text{ :- } (C, \text{worksAt}, A), (C, \text{livesIn}, B)$ 70%

KG is not correct



- Error Detection in Knowledge Graph
 - Numerical Outlier Detections
 - Detecting Inconsistencies in Knowledge Graph
 - etc.
- Error Correction in Knowledge Graph

Real Example

- http://dbpedia.org/page/Donald_Trump

About: Donald Trump

An Entity of Type : person, from Named Graph : <http://dbpedia.org>, within Data Space : dbpedia.org

Donald John Trump (born June 14, 1946) is an American businessman, politician, and television personality. He is a candidate for the Republican nomination for President of the United States in the 2016 election. Trump is the chairman and president of The Trump Organization and the founder of Trump Entertainment Resorts, a gaming and hotel enterprise. His extensive self promotion, outspoken manner, career, personal life and wealth have made him a celebrity.

<code>dbp:birthDate</code>	<ul style="list-style-type: none">▪ 1946-06-14 (xsd:date)
<code>dbp:birthName</code>	<ul style="list-style-type: none">▪ Donald John Trump (en)
<code>dbp:birthPlace</code>	<ul style="list-style-type: none">▪ <code>dbr:United_States</code>▪ <code>dbr:New_York_City</code>▪ <code>dbr:Queens</code>
<code>dbp:children</code>	<ul style="list-style-type: none">▪ <code>dbr:Ivanka_Trump</code>▪ <code>dbr:Eric_Trump</code>▪ <code>dbr:Donald_Trump,_Jr.</code>▪ <code>dbr:Donald_Trump</code>

Donald Trump is a child of himself !!

Real Example

- http://dbpedia.org/page/Marla_Maples

About: [Marla Maples](#)

An Entity of Type : person, from Named Graph : <http://dbpedia.org>, within Data Space : [dbpedia.org](#)

Marla Ann Maples (born October 27, 1963) is an American actress and television personality, best known for her six-year marriage to celebrity businessman and 2016 presidential candidate Donald Trump.

Property	Value
dbo:abstract	<ul style="list-style-type: none">▪ Marla Ann Maples (born October 27, 1963) is an American actress and television personality, best known for her six-year marriage to celebrity businessman and 2016 presidential candidate Donald Trump. (en)
dbo:activeYearsStartYear	<ul style="list-style-type: none">▪ 1986-01-01 (xsd:date)
dbo:birthDate	<ul style="list-style-type: none">▪ 1963-10-27 (xsd:date)
dbo:birthName	<ul style="list-style-type: none">▪ Marla Ann Maples (en)
dbo:birthPlace	<ul style="list-style-type: none">▪ dbr:Cohutta_Georgia▪ dbr:Georgia_(U.S._state)
dbo:birthYear	<ul style="list-style-type: none">▪ 1963-01-01 (xsd:date)
dbo:child	<ul style="list-style-type: none">▪ dbr:Donald_Trump
dbo:occupation	<ul style="list-style-type: none">▪ dbr:Celebrity▪ dbr:Marla_Maples__1▪ dbr:Actor
dbo:spouse	<ul style="list-style-type: none">▪ dbr:Donald_Trump

Donald Trump is her husband and her child !!

Error Classification

#	Error Class	Type Consistent	Fact (Correct)	Appropriate
-	No Error	/	/	/
S	Semantic Inappropriate	/	/	X
?	Not sure (due to lack of definition)	/	/	?
L	Link Missing	/	/	X
U	Unit Missing (for measurement)	/	/	X
D	Duplicate Values	/	/	X
C	Collective Information	/	/	X
E	Link to wrong entity	/	X	X
V	Incorrect Value/External Link	/	X	X
I	Invalid Link	/	X	X
F	Fragment	/	X	X
T	Type-conflicting relation	X	~	X

Range violation errors

- The object of a triple does not have the type required by the range of the triple's property
- Examples
 - < Chandan_Dasgupta, **academicAdvisor**, **Harvard_University** >
 - **academicAdvisor** requires an object with type Person
 - **Harvard University** does not contain the type Person, but University
 - < Bonkers (album), **compiler**, **Neophyte (band)** >
 - **compiler** requires an object with type Person
 - **Neophyte** does not contain the type Person, but band
 - < Latin America, **language**, **Mayan languages** >
 - **language** requires an object with type Language
 - **Mayan languages** does not contain the type Language, but it is actually a language

Object Incorrect

Range Incorrect

Object Type
Incorrect / Missing

Examples

- Non-Person objects being a parent of another object

```
SELECT ?type COUNT(DISTINCT(?o)) as ?cnt
WHERE {
  ?s dbo:spouse ?o.
  ?o a ?type.
FILTER
  regex(?type,"http://dbpedia.org/ontology")
FILTER NOT EXISTS {?o a dbo:Person.}
} GROUP BY ?type
```

type	cnt
dbpedia:ontology/Company	13
dbpedia:ontology/GovernmentAgency	1
dbpedia:ontology/Group	2
dbpedia:ontology/PeriodicalLiterature	2
dbpedia:ontology/MilitaryUnit	4
dbpedia:ontology/WrittenWork	2
dbpedia:ontology/Legislature	1
dbpedia:ontology/GivenName	7
dbpedia:ontology/Organisation	27
dbpedia:ontology/Newspaper	2
dbpedia:ontology/Work	2
dbpedia:ontology/PoliticalParty	1
dbpedia:ontology/EthnicGroup	14
dbpedia:ontology/Agent	27
dbpedia:ontology/Bank	1
dbpedia:ontology/Band	2
dbpedia:ontology/Surname	1
dbpedia:ontology/Name	8

The Framework

- Property p and its range
- All triples with property p

```
graph TD; A[Property p and its range  
All triples with property p] --> B[Refine property's range]; B --> C[Infer objects' type]; C --> D[Find correct objects];
```

Refine property's range

- Appropriate range
- Triples violating the range

Infer objects' type

- Triples with incorrect objects

Find correct objects