Knowledge Graphs and Information Retrieval
A Symbiotic Relationship

Sumit Bhatia

IBM Research AI, New Delhi, India

sumitbhatia@in.ibm.com
http://sumitbhatia.net
@sbhatia_

December 6, 2018

Introduction

Have you used Knowledge Graphs?

Have you used Knowledge Graphs?

Most probably, YES!!!

Have you used Knowledge Graphs?

Most probably, YES!!!

What is a Knowledge Graph?

No clear definition...

What is a Knowledge Graph?

No clear definition...

Structured representation of knowledge

What is a Knowledge Graph?

No clear definition...

Structured representation of knowledge

Collection of *facts* about real-world concepts

What is a Knowledge Graph?

No clear definition...

Structured representation of knowledge

Collection of *facts* about real-world concepts

Represented as $< h, r, t >$ triples

What is a Knowledge Graph?

No clear definition...

Structured representation of knowledge

Collection of *facts* about real-world concepts

Represented as $< h, r, t >$ triples

<IIT Delhi, *locatedAt*, New Delhi>

What is a Knowledge Graph?

No clear definition...

Structured representation of knowledge

Collection of *facts* about real-world concepts

Represented as $< h, r, t >$ triples

<IIT Delhi, *locatedAt*, New Delhi>

h,t – entities, represented as nodes in the graph

What is a Knowledge Graph?

No clear definition...

Structured representation of knowledge

Collection of *facts* about real-world concepts

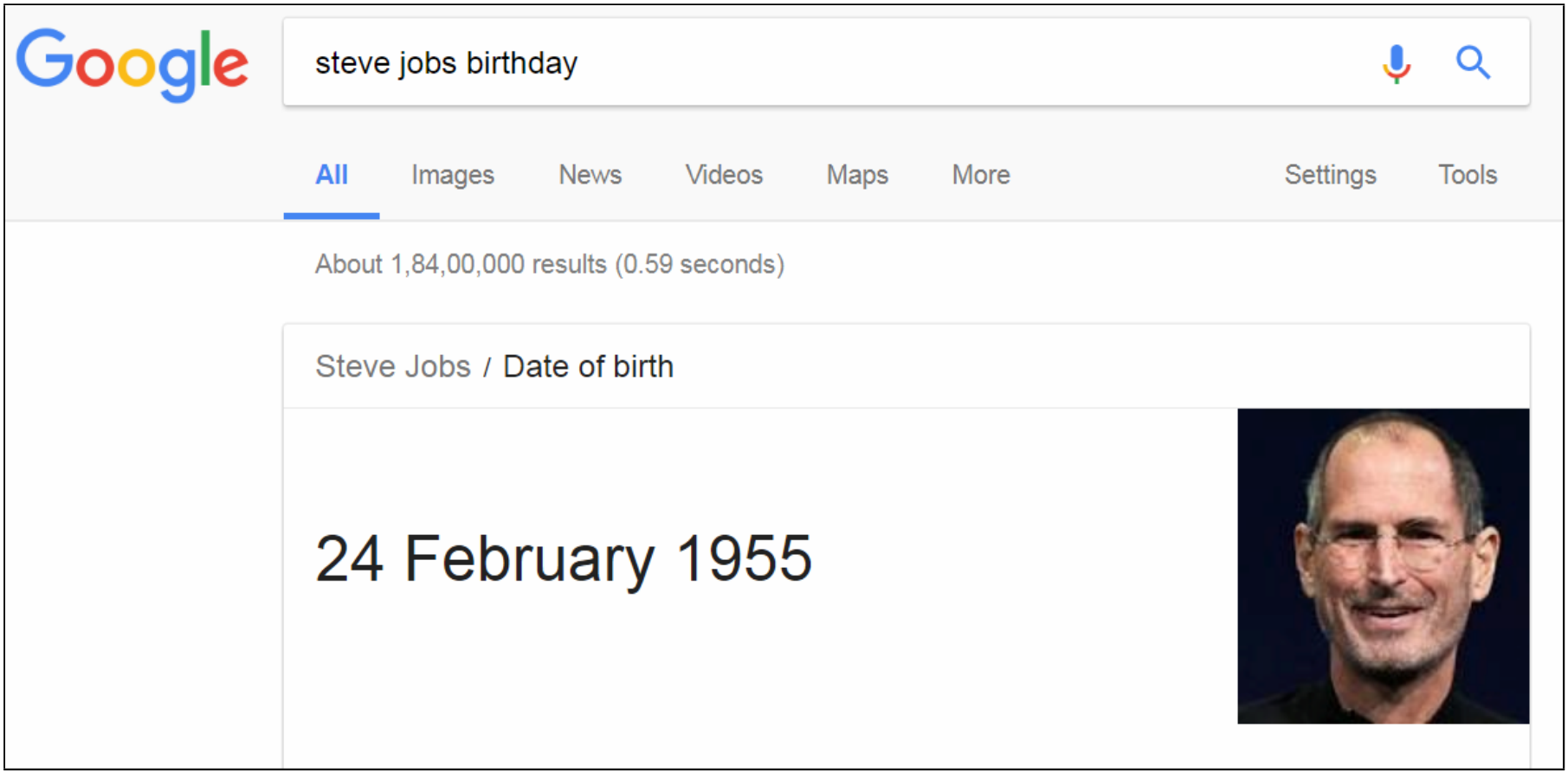
Represented as $< h, r, t >$ triples

<IIT Delhi, *locatedAt*, New Delhi>

h,t – entities, represented as nodes in the graph

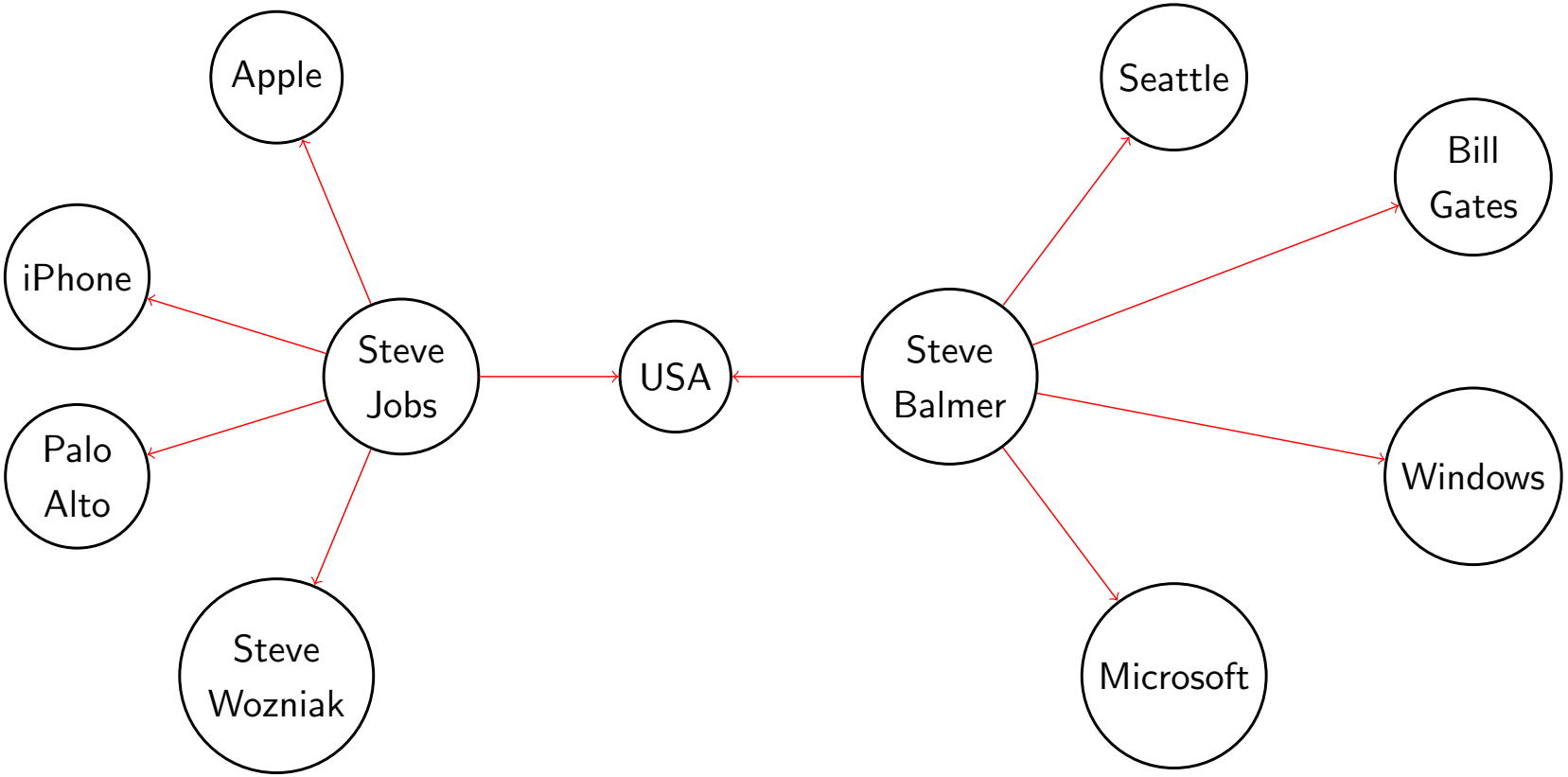
r – relation between entities, represented as an edge in the graph

Knowledge Graph



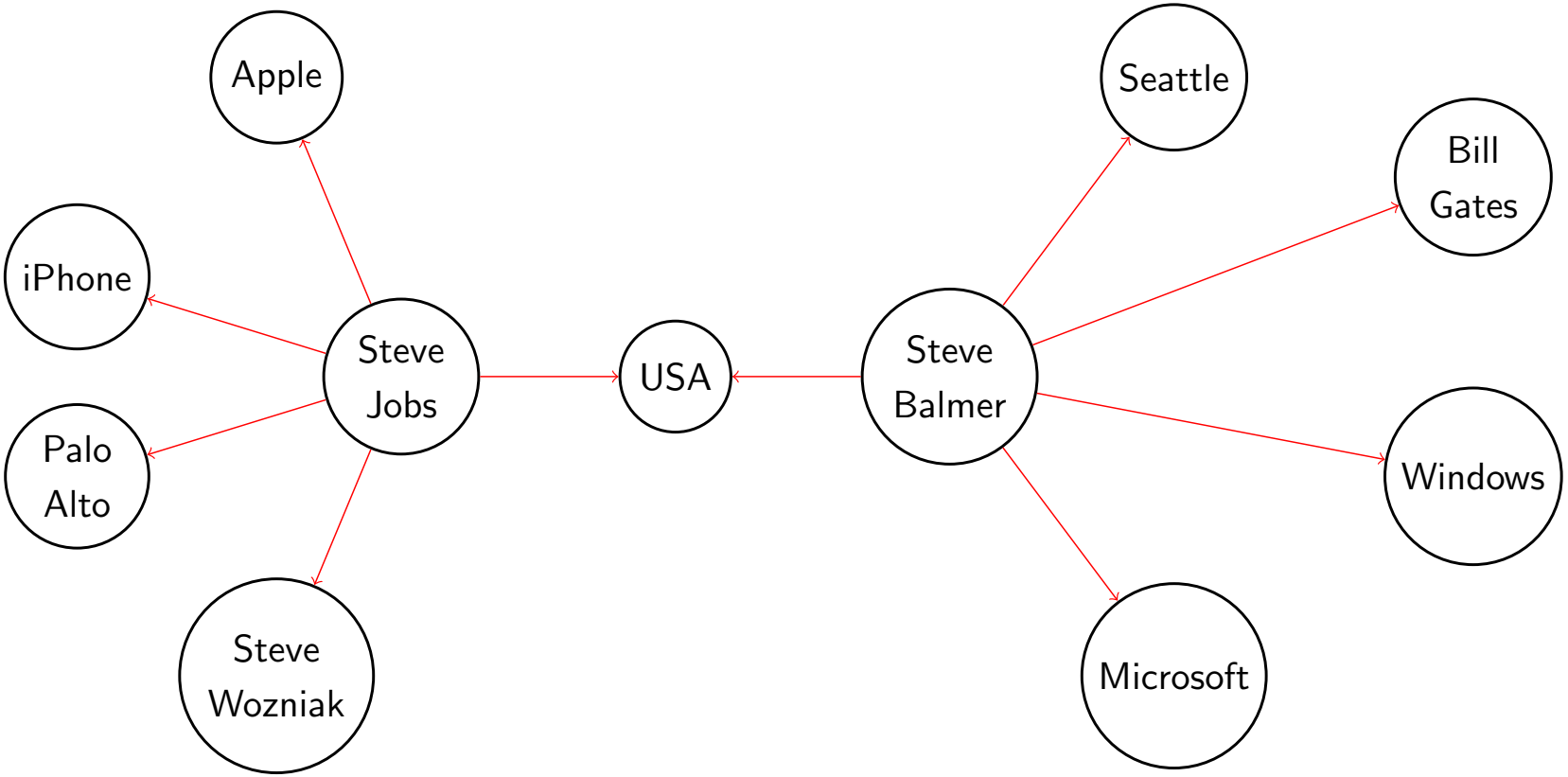
Knowledge Graph

Let us consider this graph:



Knowledge Graph

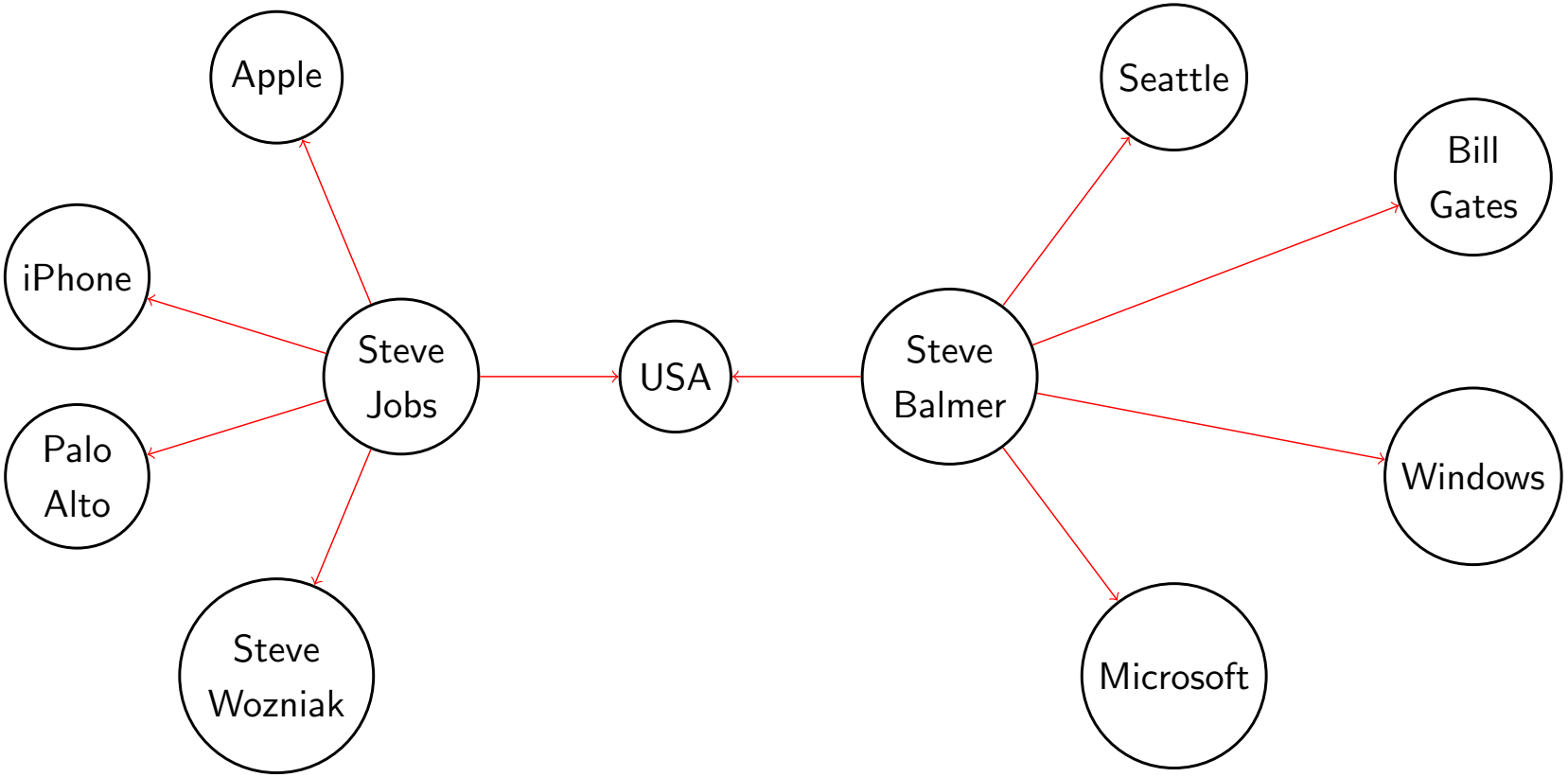
Let us consider this graph:



Open World Assumption:

Knowledge Graph

Let us consider this graph:



Open World Assumption: What is not known to be true is unknown

Knowledge Graph

What makes a Knowledge Graph

Knowledge Graph

What makes a Knowledge Graph

- Not every data in graph format is a Knowledge Graph

Knowledge Graph

What makes a Knowledge Graph

- Not every data in graph format is a Knowledge Graph
- Citation network, web graph, road network, etc. are not knowledge graphs

Knowledge Graph

What makes a Knowledge Graph

- Not every data in graph format is a Knowledge Graph
- Citation network, web graph, road network, etc. are not knowledge graphs
- Recall: A *Knowledge Graph* consists of *entities* and their *relationships*

Knowledge Graph

What makes a Knowledge Graph

- Not every data in graph format is a Knowledge Graph
- Citation network, web graph, road network, etc. are not knowledge graphs
- Recall: A *Knowledge Graph* consists of *entities* and their *relationships*

T-Box (terminology box): Defines the real world concepts and relations between them. Also known as Ontology/schema.

Knowledge Graph

What makes a Knowledge Graph

- Not every data in graph format is a Knowledge Graph
- Citation network, web graph, road network, etc. are not knowledge graphs
- Recall: A *Knowledge Graph* consists of *entities* and their *relationships*

T-Box (terminology box): Defines the real world concepts and relations between them. Also known as Ontology/schema.

A-Box (axiom box): Represents the facts, statements, or axioms

Knowledge Graph

Let us consider an example

Knowledge Graph

Let us consider an example

Our T-BOX defines these classes of entities – PERSON, MALE, FEMALE

Knowledge Graph

Let us consider an example

Our T-BOX defines these classes of entities – PERSON, MALE, FEMALE

Relations could be defined as: MALE *isA* PERSON; FEMALE *isA* PERSON ;PERSON *hasMother* FEMALE

Knowledge Graph

Let us consider an example

Our T-BOX defines these classes of entities – PERSON, MALE, FEMALE

Relations could be defined as: MALE *isA* PERSON; FEMALE *isA* PERSON ;PERSON *hasMother* FEMALE

Consider this A-Box

- Rahul *isA* MALE
- Sonia *isA* FEMALE
- Priyanka *isA* FEMALE

Knowledge Graph

Let us consider an example

Our T-BOX defines these classes of entities – PERSON, MALE, FEMALE

Relations could be defined as: MALE *isA* PERSON; FEMALE *isA* PERSON ;PERSON *hasMother* FEMALE

Consider this A-Box

- Rahul *isA* MALE
- Sonia *isA* FEMALE
- Priyanka *isA* FEMALE
- Rahul *hasMother* Sonia

Knowledge Graph

Let us consider an example

Our T-BOX defines these classes of entities – PERSON, MALE, FEMALE

Relations could be defined as: MALE *isA* PERSON; FEMALE *isA* PERSON ;PERSON *hasMother* FEMALE

Consider this A-Box

- Rahul *isA* MALE
- Sonia *isA* FEMALE
- Priyanka *isA* FEMALE
- Rahul *hasMother* Sonia
- Sonia *hasMother* Rahul

Knowledge Graph

Let us consider an example

Our T-BOX defines these classes of entities – PERSON, MALE, FEMALE

Relations could be defined as: MALE *isA* PERSON; FEMALE *isA* PERSON ;PERSON *hasMother* FEMALE

Consider this A-Box

- Rahul *isA* MALE
- Sonia *isA* FEMALE
- Priyanka *isA* FEMALE
- Rahul *hasMother* Sonia
- Sonia *hasMother* Rahul
- Sonia *hasMother* Priyanka

A Typical Knowledge Graph Entity

Here is the Wikipedia page for Narendra Modi

Narendra Modi

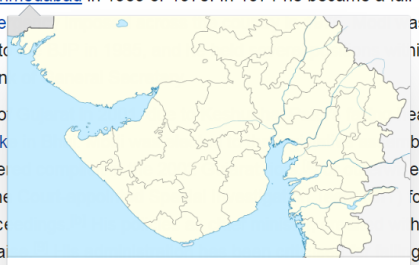
From Wikipedia, the free encyclopedia

Narendra Damodardas Modi (pronounced [ˈnəre:ndrə da:mo:dərˈda:s ˈmo:di:] (listen); born 17 September 1950) is an Indian politician serving as the 14th and current Prime Minister of India since 2014. He was the Chief Minister of Gujarat from 2001 to 2014, and is the Member of Parliament for Varanasi. Modi is a member of the Bharatiya Janata Party (BJP) and the Rashtriya Swayamsevak Sangh (RSS).

Born to a Gujarati family in Vadnagar, Modi helped his father sell tea as a child and later ran his own stall. He was introduced to the RSS at the age of eight, beginning a long association with the organisation. He left home after graduating from school, partly because of an arranged marriage which he rejected. Modi travelled around India for two years and visited a number of religious centres. He returned to Gujarat and moved to Ahmedabad in 1969 or 1970. In 1971 he became a full-time worker for the RSS. During the state of eme as forced to go into hiding. The RSS assigned him t in the party hierarchy until 2001, rising to the ran

Modi was appointed Chief Minister o ealth and poor public image following the earthquak bly soon after. His administration has been conside e criticised for its handling of it, however a Supreme found no evidence to initiate prosecution proc h encouraging economic growth, have received pra g to significantly improve health, poverty,

Modi led the BJP in the 2014 genera Sabha, the first time a single party had achieved ment from Varanasi. Since taking office, Modi's ment in the Indian economy, increased spending are and social s atten centralised



Ahmedabad, also pronounced as, **Amdavad** in Gujarati, is the largest city and former capital of the Indian state of Gujarat. It is the administrative headquarter of the Ahmadabad district and the seat of the Gujarat High Court. Ahmedabad's population of 5,633,927 makes

Narendra Modi

MP

14th Prime Minister of India	
Incumbent	
Assumed office	26 May 2014
President	Pranab Mukherjee Ram Nath Kovind
Preceded by	Manmohan Singh
14th Chief Minister of Gujarat	
In office	7 October 2001 – 22 May 2014

A Typical Knowledge Graph Entity

Here is the Wikidata page for Narendra Modi

Narendra Modi (Q1058)

Prime Minister of India

edit

In more languages Configure

Language	Label	Description	Also known as
English	Narendra Modi	Prime Minister of India	
Hindi	नरेन्द्र मोदी	भारत के प्रधानमंत्री	नरेन्द्र दामोदरदास मोदी नमो नरेन्द्र भाई नरेंद्रभाई मोदी मोदी
Bangla	নরেন্দ্র মোদী	ভারতের সর্বকালের শ্রেষ্ঠ প্রধানমন্ত্রী।	
Telugu	నరేంద్ర మోడి	భారతదేశ ప్రధాన మంత్రి	మోదీ నమో నరేంద్ర భాయ్ నరేంద్ర భాయ్ మోడీ నరేంద్ర దామోదర్ దాస్ మోడీ

All entered languages

Statements

instance of	human 2 references	edit
		+ add value
image		edit

A Typical Knowledge Graph Entity

Here is the Wikidata page for Narendra Modi

educated at	Gujarat University		edit
	academic degree	master's degree	
	end time	1983	
	academic major	political science	
	0 references		
			+ add reference
	University of Delhi		edit
	academic major	political science	
	end time	1978	
	0 references		
			+ add reference
			+ add value
residence	New Delhi		edit
	0 references		
			+ add reference
	7, Lok Kalyan Marg		edit
	0 references		
			+ add reference
			+ add value
member of political party	Bharatiya Janata Party		edit
	0 references		
			+ add reference

A Typical Knowledge Graph Entity

Here is the Wikidata page for Narendra Modi

family name	Modi 0 references	edit add reference add value
date of birth	17 September 1950 1 reference	edit add value
place of birth	Vadnagar located in the administrative territorial entity: Bombay State country: India 0 references	edit add reference add value
spouse	Jashodaben Narendrabhai Modi 1 reference	edit add value
native language	Gujarati 1 reference	edit add value

Some Knowledge Graphs

- Google, Bing, Yahoo, Yandex - all have one
- dbPedia (`https://wiki.dbpedia.org/`) – created out of Wikipedia infoboxes ($\approx 800M$ triples)
- Wikidata (https://www.wikidata.org/) – Comprehensive, structured knowledge base from Wikipedia ($\approx 7B$ triples)
- YAGO [25] – developed and maintained by Max-Planck Institute ($\approx 200M$ triples)
- Many proprietary, custom knowledge graphs in different domains

Knowledge Graphs and Information Retrieval

Knowledge Graphs and Information Retrieval

- KGs provide a structured view of your data

Knowledge Graphs and Information Retrieval

- KGs provide a structured view of your data
- Ontologies need to be there, highly precise and sensitive to errors/noise

Knowledge Graphs and Information Retrieval

- KGs provide a structured view of your data
- Ontologies need to be there, highly precise and sensitive to errors/noise
- Information Retrieval is mostly over unstructured data, probabilistic in nature, comparatively more robust to noise

Knowledge Graphs and Information Retrieval

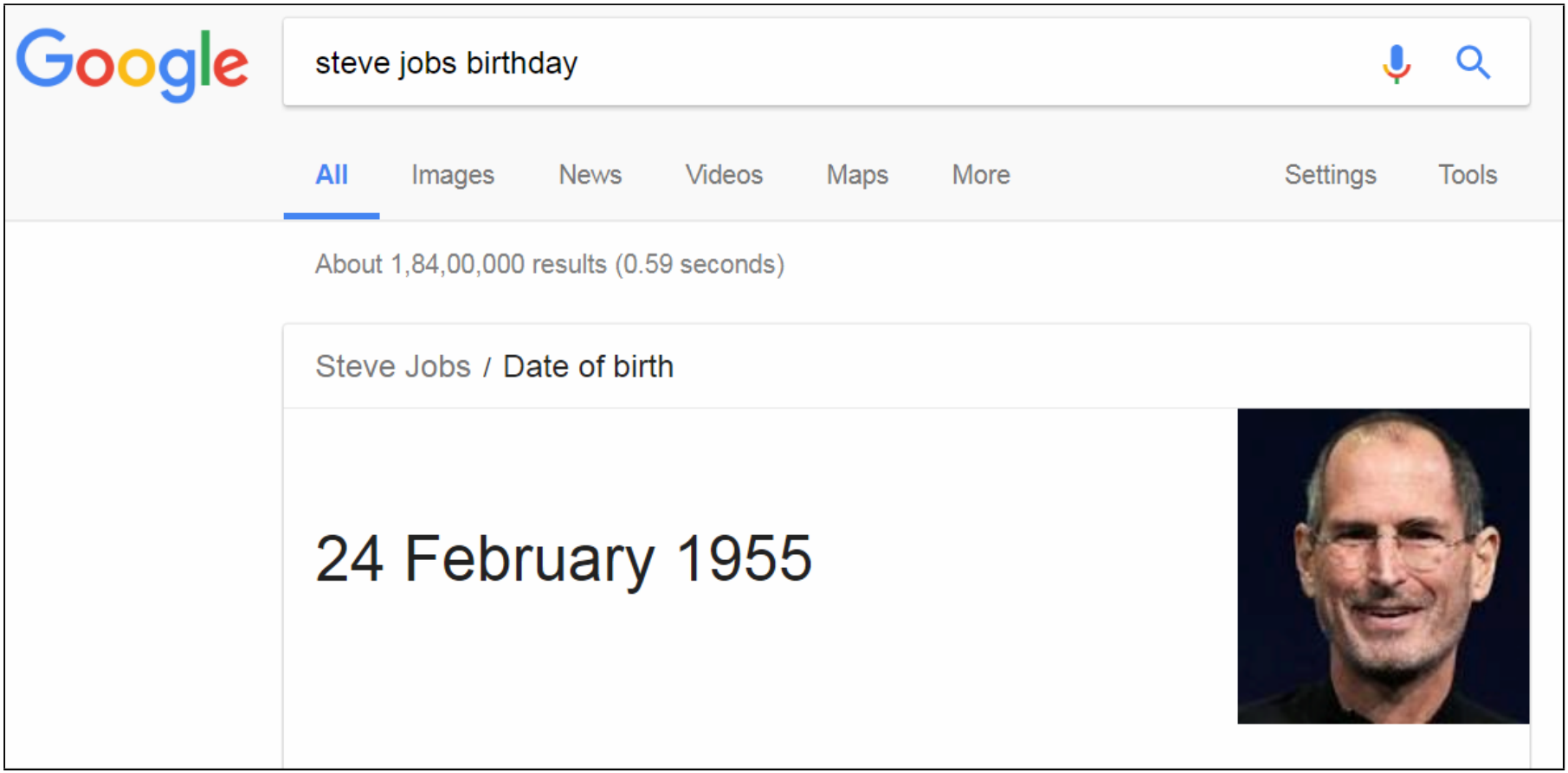
- KGs provide a structured view of your data
- Ontologies need to be there, highly precise and sensitive to errors/noise
- Information Retrieval is mostly over unstructured data, probabilistic in nature, comparatively more robust to noise

Focus of this tutorial: How can these two benefit or complement each other?

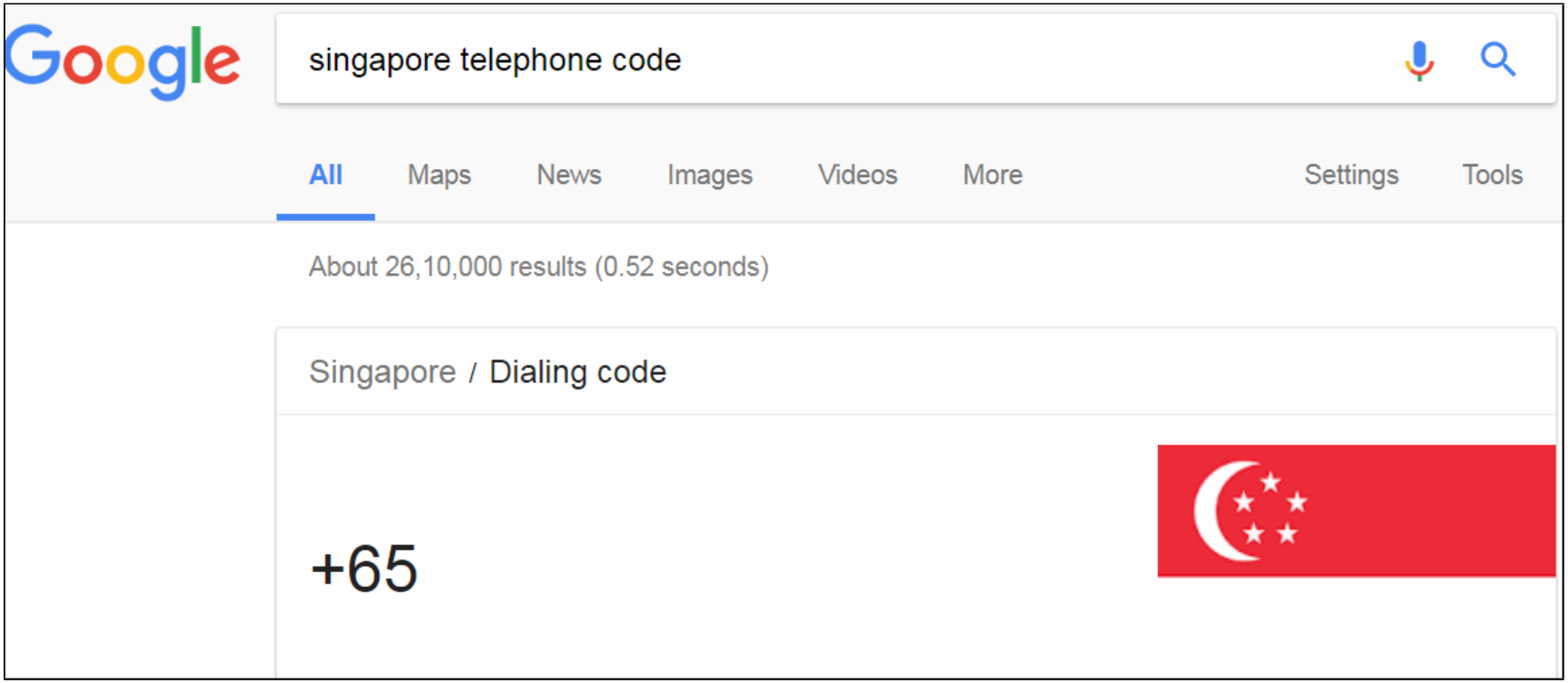
Overview of Tutorial

- Finding Entities of Interest
  - Entity Search and Recommendation
  - Entity Linking and Disambiguation
  - Knowledge Graphs for Query Expansion
- Entity exploration: Knowing more about the entities
  - Relationship Search
  - Path Ranking
- Upcoming Challenges

Finding the Right Entities



Finding the Right Entities



Finding Right Entities

Entities are the *fundamental units* of a Knowledge graph. How to get to the right entities in the graph?

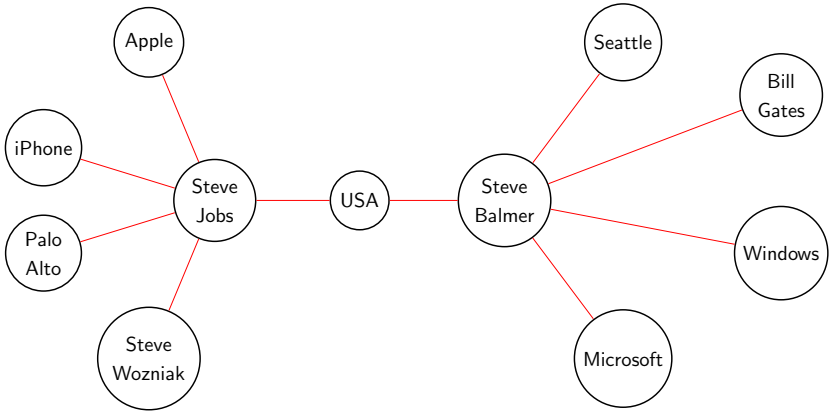
Finding Right Entities

Entities are the *fundamental units* of a Knowledge graph. How to get to the right entities in the graph?
Given a Knowledge Base, $K = \{\mathcal{E}, \mathcal{R}\}$, a document corpus $\mathcal{D}$, and a named entity mention m, map/link the mention m to its corresponding entity $e \in \mathcal{E}$.

Finding Right Entities

Entities are the *fundamental units* of a Knowledge graph. How to get to the right entities in the graph?

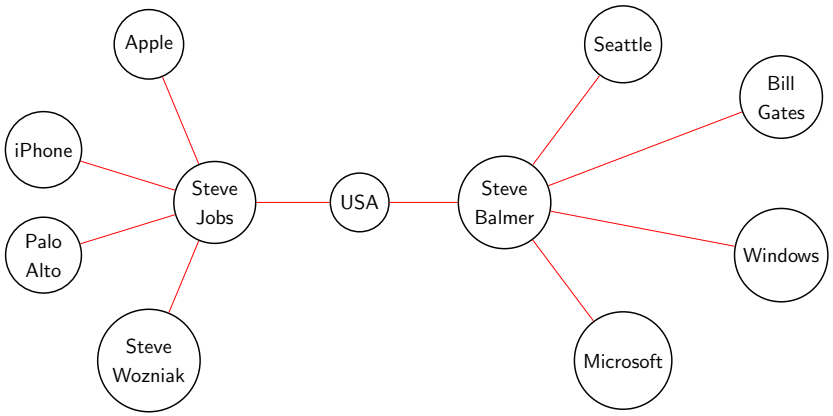
Given a Knowledge Base, $K = \{\mathcal{E}, \mathcal{R}\}$, a document corpus $\mathcal{D}$, and a named entity mention m, map/link the mention m to its corresponding entity $e \in \mathcal{E}$.



Finding Right Entities

Entities are the *fundamental units* of a Knowledge graph. How to get to the right entities in the graph?

Given a Knowledge Base, $K = \{\mathcal{E}, \mathcal{R}\}$, a document corpus $\mathcal{D}$, and a named entity mention m, map/link the mention m to its corresponding entity $e \in \mathcal{E}$.



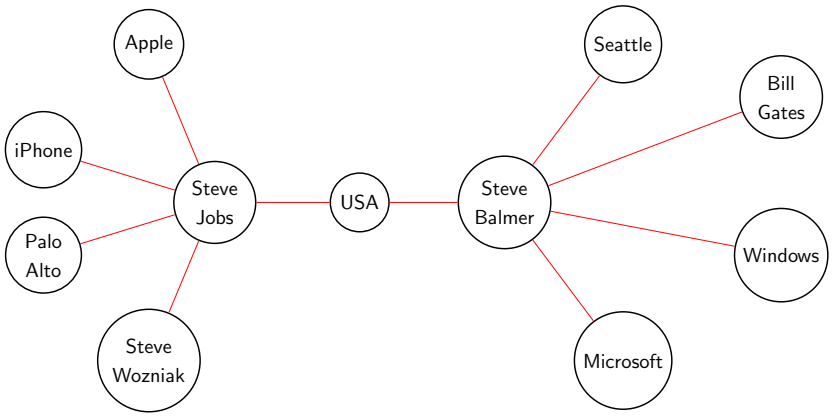
Web Queries:

steve jobs birthday

Finding Right Entities

Entities are the *fundamental units* of a Knowledge graph. How to get to the right entities in the graph?

Given a Knowledge Base, $K = \{\mathcal{E}, \mathcal{R}\}$, a document corpus $\mathcal{D}$, and a named entity mention m, map/link the mention m to its corresponding entity $e \in \mathcal{E}$.



Web Queries:

steve jobs birthday

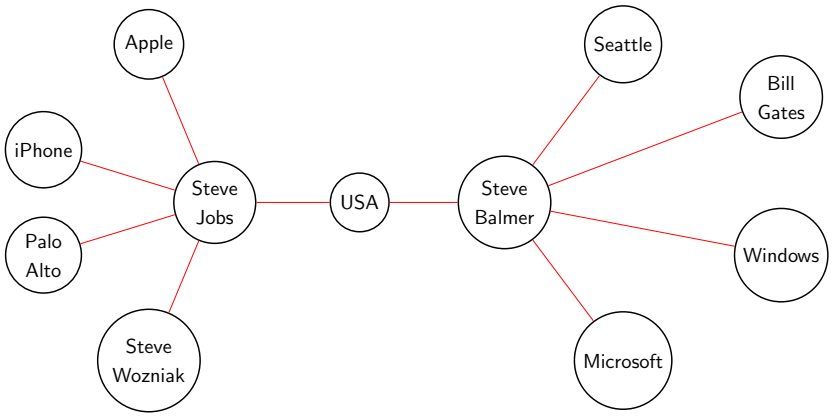
NL Questions:

When did Steve resign from Microsoft?

Finding Right Entities

Entities are the *fundamental units* of a Knowledge graph. How to get to the right entities in the graph?

Given a Knowledge Base, $K = \{\mathcal{E}, \mathcal{R}\}$, a document corpus $\mathcal{D}$, and a named entity mention m, map/link the mention m to its corresponding entity $e \in \mathcal{E}$.



Web Queries:

steve jobs birthday

NL Questions:

When did Steve resign from Microsoft?

NL Text:

....Jobs and Wozniak started Apple Computers from their garage...

Challenges

- Same entity can be represented by multiple surface forms

Challenges

- Same entity can be represented by multiple surface forms
  Barack Obama, Barack H. Obama, President Obama, Senator Obama

Challenges

- Same entity can be represented by multiple surface forms
  Barack Obama, Barack H. Obama, President Obama, Senator Obama
  President of the United States

Challenges

- Same entity can be represented by multiple surface forms
  Barack Obama, Barack H. Obama, President Obama, Senator Obama
  President of the United States
- Same surface form could refer to multiple entities

Challenges

- Same entity can be represented by multiple surface forms
  Barack Obama, Barack H. Obama, President Obama, Senator Obama
  President of the United States
- Same surface form could refer to multiple entities
  Michael Jordan – Basketball player or Berkeley professor

Challenges

- Same entity can be represented by multiple surface forms
  Barack Obama, Barack H. Obama, President Obama, Senator Obama
  President of the United States
- Same surface form could refer to multiple entities
  Michael Jordan – Basketball player or Berkeley professor
  when did *steve* leave apple?

Challenges

- Same entity can be represented by multiple surface forms
  Barack Obama, Barack H. Obama, President Obama, Senator Obama
  President of the United States
- Same surface form could refer to multiple entities
  Michael Jordan – Basketball player or Berkeley professor
  when did *steve* leave apple?
- Out of KG mentions

Entity Linking

Related problems:

- Record linkage/de-duplication in databases
- Entity Resolution/name matching
- Co-reference resolution, Word Sense disambiguation

Entity Linking Process

Entity Linking Process

Entity Linking Process

Named Entity Recognition

Well studied in NLP [19]

open source software like Stanford NLP toolkit [18]

Entity Linking Process

Named Entity Recognition Well studied in NLP [19] open source software like Stanford NLP toolkit [18]

Use of dictionaries

Entity Linking Process

Named Entity Recognition
Well studied in NLP [19]
open source software like Stanford NLP toolkit [18]

Use of dictionaries

Ranking target entities based on:

- graph based features
- text/document based features

Candidate Entity List Generation

Candidate Entity List Generation

- Much of the variation between different entity linking algorithms could be explained by quality of candidate search components [14]

Candidate Entity List Generation

- Much of the variation between different entity linking algorithms could be explained by quality of candidate search components [14]
- Acronym expansions and coreference resolutions lead to significant performance gains [14]

Candidate Entity List Generation

- Much of the variation between different entity linking algorithms could be explained by quality of candidate search components [14]
- Acronym expansions and coreference resolutions lead to significant performance gains [14]
- The candidate set should be exhaustive enough but not too big to affect efficiency

Candidate Entity List Generation

Dictionary based Methods

An offline dictionary of entity names created out of external sources mapping different possible surface forms of entity names to their corresponding entities in the KG

Candidate Entity List Generation

Dictionary based Methods

An offline dictionary of entity names created out of external sources mapping different possible surface forms of entity names to their corresponding entities in the KG

- Domain specific sources like Gene name dictionary [20]

Candidate Entity List Generation

Dictionary based Methods

An offline dictionary of entity names created out of external sources mapping different possible surface forms of entity names to their corresponding entities in the KG

- Domain specific sources like Gene name dictionary [20]
- Wikipedia/DBPedia
  - Page Titles
  - Disambiguation/Redirect pages
  - Anchor text of Wikipedia in links

Candidate Entity List Generation

Dictionary based Methods

An offline dictionary of entity names created out of external sources mapping different possible surface forms of entity names to their corresponding entities in the KG

- Domain specific sources like Gene name dictionary [20]
- Wikipedia/DBPedia
  - Page Titles
  - Disambiguation/Redirect pages
  - Anchor text of Wikipedia in links
- Anchor text from Web pages to Wikipedia articles

Candidate Entity List Generation

Dictionary based Methods

An offline dictionary of entity names created out of external sources mapping different possible surface forms of entity names to their corresponding entities in the KG

- Domain specific sources like Gene name dictionary [20]
- Wikipedia/DBPedia
  - Page Titles
  - Disambiguation/Redirect pages
  - Anchor text of Wikipedia in links
- Anchor text from Web pages to Wikipedia articles
- Acronym expansions

Candidate Entity List Generation

Surface Form	Entity Canonical Form
Barack Obama	< Barack Obama,Person>
Barack H. Obama	<Barack Obama,Person>
USA	<United States of America, Country>
America	<United States of America,Country>
Big Apple	<New York, City>
NYC	<New York, City>
NY	<New York, City>

Candidate Entity List Generation

Surface Form	Entity Canonical Form
Barack Obama	< Barack Obama,Person>
Barack H. Obama	<Barack Obama,Person>
USA	<United States of America, Country>
America	<United States of America,Country>
Big Apple	<New York, City>
NYC	<New York, City>
NY	<New York, City>
NY	<New York, State>

Candidate Entity List Generation

Surface Form	Entity Canonical Form
Barack Obama	< Barack Obama,Person>
Barack H. Obama	<Barack Obama,Person>
USA	<United States of America, Country>
America	<United States of America,Country>
Big Apple	<New York, City>
NYC	<New York, City>
NY	<New York, City>
NY	<New York, State>

Simple term match – partial or exact

...Obama visited Singapore in 2016...

Matches: Barack Obama, Mount Obama, Michelle Obama,..., etc.

Candidate Entity Ranking

Candidate Entity Ranking

The candidate entity set can be big!

Candidate Entity Ranking

The candidate entity set can be big!

For KORE50 dataset:

- 631 candidates on an average per mention in YAGO [25]
- 2000+ in Watson KG [5]

Candidate Entity Ranking

The candidate entity set can be big!
For KORE50 dataset:

- 631 candidates on an average per mention in YAGO [25]
- 2000+ in Watson KG [5]

Approaches for ranking can be clubbed under two broad categories:

- Text based
- Graph structure based

Candidate Entity Ranking

. . . Obama is in Hawaii this week. . .

{Barack Obama, Michelle Obama, Mt. Obama}

Candidate Entity Ranking

. . . Obama is in Hawaii this week. . .

{Barack Obama, Michelle Obama, Mt. Obama}

- Similarity between entity name and mention
  - Term overlap, edit distance, etc.

Candidate Entity Ranking

. . . Obama is in Hawaii this week. . .

{Barack Obama, Michelle Obama, Mt. Obama}

- Similarity between entity name and mention
  - Term overlap, edit distance, etc.
- Entity Popularity – Wikipedia page views [13, 12]

Candidate Entity Ranking

. . . Obama is in Hawaii this week. . .

{Barack Obama, Michelle Obama, Mt. Obama}

- Similarity between entity name and mention
  - Term overlap, edit distance, etc.
- Entity Popularity – Wikipedia page views [13, 12]
- Wikipedia/web anchor text/ inlinks [22, 15]

Candidate Entity Ranking

. . . Obama is in Hawaii this week. . .

{Barack Obama, Michelle Obama, Mt. Obama}

- Similarity between entity name and mention
  - Term overlap, edit distance, etc.
- Entity Popularity – Wikipedia page views [13, 12]
- Wikipedia/web anchor text/ inlinks [22, 15]

. . . when did Steve leave apple. . .

{Steve Jobs,Steve Wozniak,Steve Ballmer}

Candidate Entity Ranking

. . . Obama is in Hawaii this week. . .

{Barack Obama, Michelle Obama, Mt. Obama}

- Similarity between entity name and mention
  - Term overlap, edit distance, etc.
- Entity Popularity – Wikipedia page views [13, 12]
- Wikipedia/web anchor text/ inlinks [22, 15]

. . . when did Steve leave apple. . .

{Steve Jobs,Steve Wozniak,Steve Ballmer}

Context Matters!

Candidate Entity Ranking

Role of Context

. . . when did Steve leave apple. . .

{Steve Jobs,Steve Wozniak,Steve Ballmer}

Candidate Entity Ranking

Role of Context

. . . when did Steve leave apple. . .

{Steve Jobs,Steve Wozniak,Steve Ballmer}

- Mention context
  - text of the document/paragraph in which the mention appears
  - a window of terms around the mention

Candidate Entity Ranking

Role of Context

. . . when did Steve leave apple. . .

{Steve Jobs,Steve Wozniak,Steve Ballmer}

- Mention context
  - text of the document/paragraph in which the mention appears
  - a window of terms around the mention
- Entity context representations
  - Wikipedia article
  - Text around anchors
  - Domain specific models: abstracts of papers containing gene name in titles

Candidate Entity Ranking

Role of Context

. . . when did Steve leave apple. . .

{Steve Jobs,Steve Wozniak,Steve Ballmer}

- Mention context
  - text of the document/paragraph in which the mention appears
  - a window of terms around the mention
- Entity context representations
  - Wikipedia article
  - Text around anchors
  - Domain specific models: abstracts of papers containing gene name in titles

Compute similarity between mention and entity context representations

Candidate Entity Ranking

Graph Based Features Focus on *strength* between entities, often useful in **collective entity linking**

Candidate Entity Ranking

Graph Based Features Focus on *strength* between entities, often useful in **collective entity linking**

- Simplest graph based measure – Entity Popularity

$$pop(e) = \frac{nbrCount(e)}{\sum_{e' \in \mathcal{E}} nbrCount(e')} \tag{1}$$

In Wikipedia graph, inlinks and outlinks can be used to compute popularity

Candidate Entity Ranking

Graph Based Features Focus on *strength* between entities, often useful in **collective entity linking**

- Simplest graph based measure – Entity Popularity

$$pop(e) = \frac{nbrCount(e)}{\sum_{e' \in \mathcal{E}} nbrCount(e')} \tag{1}$$

  In Wikipedia graph, inlinks and outlinks can be used to compute popularity

Next we review some measures useful for collective entity linking

Candidate Entity Ranking

Linking/Resolving/Disambiguating Multiple Entities simultaneously

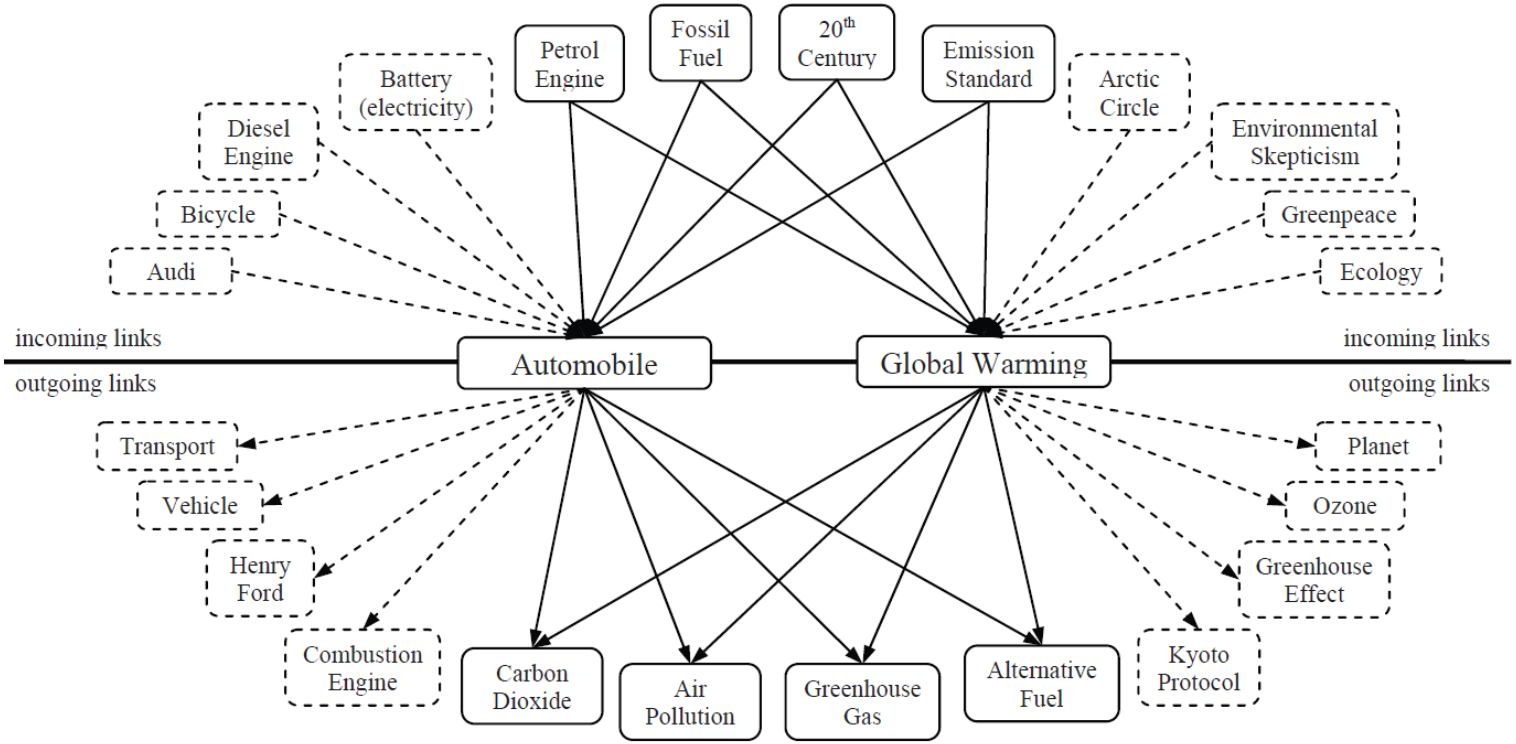


Image Source: [28]

Candidate Entity Ranking

Brad and Angelina were holidaying in Paris.

Candidate Entity Ranking

Brad and Angelina were holidaying in Paris.

- Jaccard Index

$$J(a, b) = \frac{|A \cap B|}{|A \cup B|} \tag{2}$$

Candidate Entity Ranking

Brad and Angelina were holidaying in Paris.

- Jaccard Index

$$J(a, b) = \frac{|A \cap B|}{|A \cup B|} \tag{2}$$

- Milne-Witten Similarity [28]

$$MW(a, b) = \frac{log(max(|A|, |B|)) - log(|A \cap B|)}{log(|\mathcal{N}|) - log(min(|A|, |B|))} \tag{3}$$

where, A and B are the set of neighbors of entities a and b, respectively.

Candidate Entity Ranking

Brad and Angelina were holidaying in Paris.

- Jaccard Index

$$J(a, b) = \frac{|A \cap B|}{|A \cup B|} \tag{2}$$

- Milne-Witten Similarity [28]

$$MW(a, b) = \frac{log(max(|A|, |B|)) - log(|A \cap B|)}{log(|\mathcal{N}|) - log(min(|A|, |B|))} \tag{3}$$

where, A and B are the set of neighbors of entities a and b, respectively.

- Adamic Adar [1]

$$AA(a, b) = \sum_{n \in A \cup B} log(\frac{1}{degree(n)}) \tag{4}$$

Candidate Entity Ranking

- These features can be used in supervised or unsupervised settings
- Choice of features depend on data/domain at hand. Many features are specific for Wikipedia, that may not be applicable to other textual data.
- Trade off between accuracy and efficiency while designing your systems

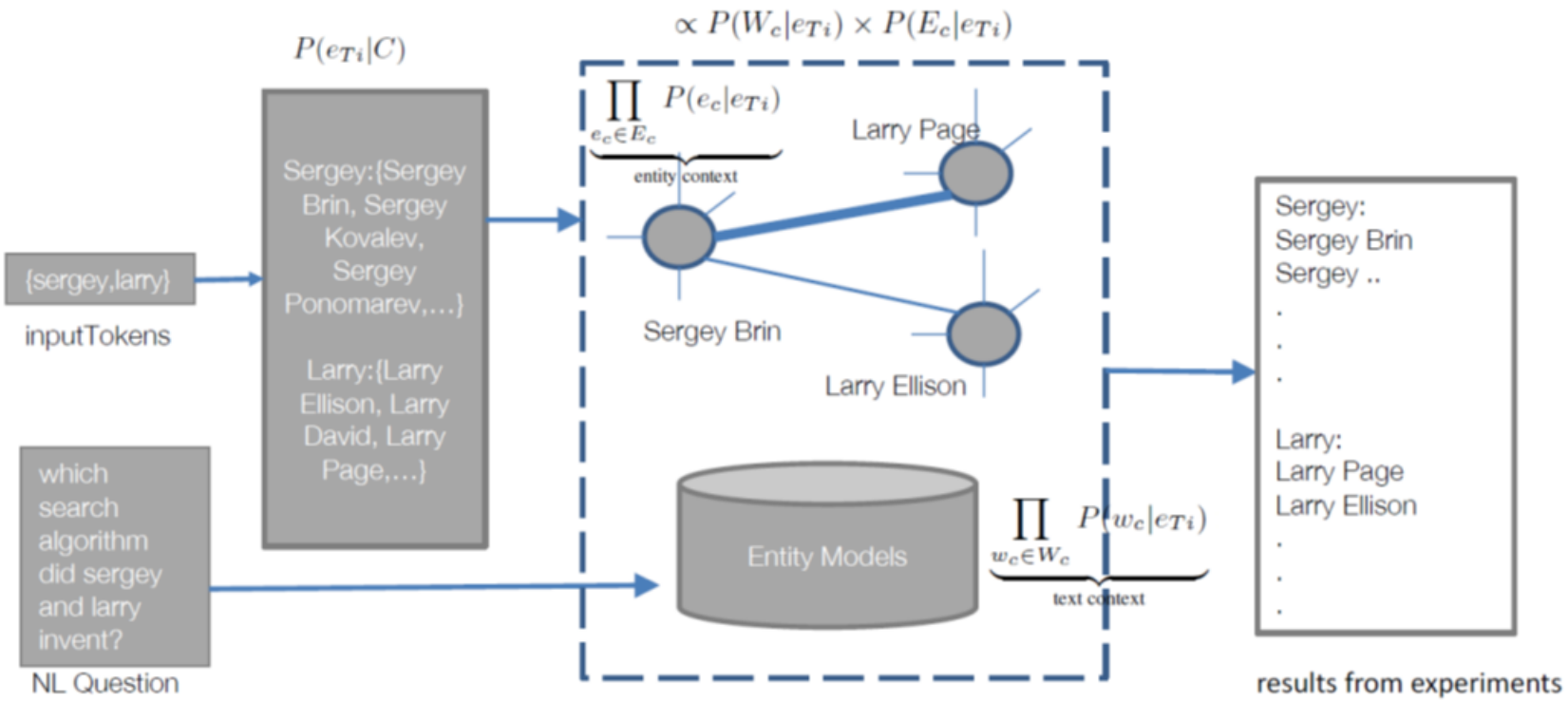
Entity Linking as implemented in Watson KG[1]

Which search algorithm did Sergey and Larry invent?

[1] S. Bhatia and A. Jain. "Context Sensitive Entity Linking of Search Queries in Enterprise Knowledge Graphs". In: *International Semantic Web Conference*. Springer. 2016, pp. 50–54.

Entity Linking as implemented in Watson KG[1]

Which search algorithm did Sergey and Larry invent?



[1] S. Bhatia and A. Jain. "Context Sensitive Entity Linking of Search Queries in Enterprise Knowledge Graphs". In: *International Semantic Web Conference*. Springer. 2016, pp. 50–54.

Knowledge Graphs for Query Expansion

Even before Knowledge Graphs became popular, query expansion using structured knowledge was shown to be useful [8].
Advancements in entity linking make it more appealing

Knowledge Graphs for Query Expansion

Even before Knowledge Graphs became popular, query expansion using structured knowledge was shown to be useful [8]. Advancements in entity linking make it more appealing

- All documents in collection are annotated with entity mentions

Knowledge Graphs for Query Expansion

Even before Knowledge Graphs became popular, query expansion using structured knowledge was shown to be useful [8]. Advancements in entity linking make it more appealing

- All documents in collection are annotated with entity mentions
- Perform entity linking for queries

Knowledge Graphs for Query Expansion

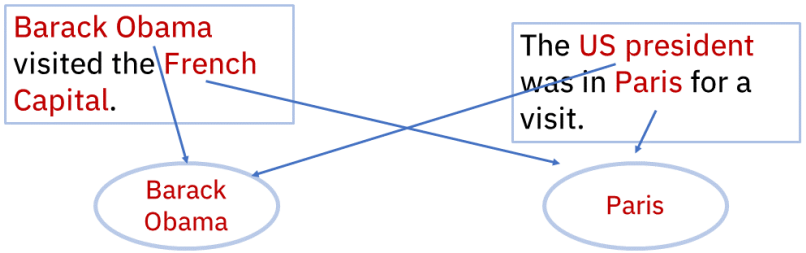
Even before Knowledge Graphs became popular, query expansion using structured knowledge was shown to be useful [8]. Advancements in entity linking make it more appealing

- All documents in collection are annotated with entity mentions
- Perform entity linking for queries
- Produce a ranking based on matched terms and knowledge concepts

Knowledge Graphs for Query Expansion

Even before Knowledge Graphs became popular, query expansion using structured knowledge was shown to be useful [8]. Advancements in entity linking make it more appealing

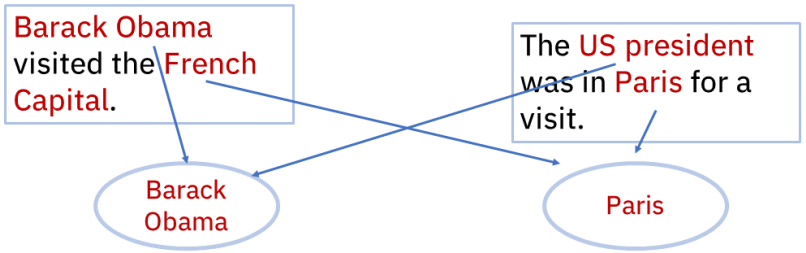
- All documents in collection are annotated with entity mentions
- Perform entity linking for queries
- Produce a ranking based on matched terms and knowledge concepts
- Consider the query *obama in paris*



Knowledge Graphs for Query Expansion

Even before Knowledge Graphs became popular, query expansion using structured knowledge was shown to be useful [8]. Advancements in entity linking make it more appealing

- All documents in collection are annotated with entity mentions
- Perform entity linking for queries
- Produce a ranking based on matched terms and knowledge concepts
- Consider the query *obama in paris*



Additional features and entity properties from KG can also be utilized [29, 29]

Entity Exploration

Entity Exploration

We found the entity of interest.

Knowing more about the entity

- Finding entities related to entity of interest
- Properties of entities
- Going beyond immediate neighborhood of the entity

Entity Retrieval

- Entity Box in web queries
- Lots of useful information about the query entity
- $\approx$ 40% of all web queries are entity queries [21]
- Many QA queries can be answered by the underlying Knowledge Base

Entity Retrieval

Related Entity Finding track at TREC [3]

Entity Retrieval

Related Entity Finding track at TREC [3]

Input: Entity Name and Search Intent

Entity Retrieval

Related Entity Finding track at TREC [3]

Input: Entity Name and Search Intent

Output: Ranked list of entity documents – entities embedded in documents

Entity Retrieval

Related Entity Finding track at TREC [3]

Input: Entity Name and Search Intent

Output: Ranked list of entity documents – entities embedded in documents

Example:

Query: Blackberry

Intent:Carriers that carry Blackberry phones

Example Answers:Verizon, AT&T, etc.

Entity Retrieval

Components of Related Entity Ranking [9][2]

[2] M. Bron, K. Balog, and M. De Rijke. "Ranking related entities: components and analyses". In: *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM. 2010, pp. 1079–1088.

Entity Retrieval

Components of Related Entity Ranking [9][2]

For a given input entity e_s, type T of target entity, and a relation description R, we wish to rank the target entities as follows:

$$P(e|e_s, T, R) \propto \underbrace{P(R|e_s, e)}_{\text{Context Modeling}} \times \underbrace{P(e|e_s)}_{\text{Co-occurrence}} \times \underbrace{P(T|e)}_{\text{Type Filtering}} \quad (5)$$

[2] M. Bron, K. Balog, and M. De Rijke. "Ranking related entities: components and analyses". In: *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM. 2010, pp. 1079–1088.

Entity Retrieval

Components of Related Entity Ranking [9][2]

For a given input entity e_s, type T of target entity, and a relation description R, we wish to rank the target entities as follows:

$$P(e|e_s, T, R) \propto \underbrace{P(R|e_s, e)}_{\text{Context Modeling}} \times \underbrace{P(e|e_s)}_{\text{Co-occurrence}} \times \underbrace{P(T|e)}_{\text{Type Filtering}} \tag{5}$$

[2] M. Bron, K. Balog, and M. De Rijke. "Ranking related entities: components and analyses". In: *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM. 2010, pp. 1079–1088.

Entity Retrieval

Co-occurrence

$$P(e|e_s) = \frac{cooc(e, e_s)}{\sum_{e' \in E} cooc(e', e_s)}$$

Entity Retrieval

Co-occurrence

$$P(e|e_s) = \frac{cooc(e, e_s)}{\sum_{e' \in E} cooc(e', e_s)}$$

Type Filtering

- Wikipedia categories
- Named entity recognizer tools

Entity Retrieval

Co-occurrence

$$P(e|e_s) = \frac{cooc(e, e_s)}{\sum_{e' \in E} cooc(e', e_s)}$$

Type Filtering

- Wikipedia categories
- Named entity recognizer tools

Context Modeling

Co-occurrence language model Θ_{ee_s} approximated by documents in which e, E_s co-occur

$$P(R|e, e_s) = \prod_{t \in R} P(t|\Theta_{ee_s})$$

Entity Recommendation for Web Queries I

Entity recommendations for web search queries[7][3]

[3] R. Blanco et al. "Entity recommendations in web search". In: *International Semantic Web Conference*. Springer. 2013, pp. 33–48.

Entity Recommendation for Web Queries I

Entity recommendations for web search queries[7][3]

- Co-occurrence features
  - query logs, user sessions
  - flickr and twitter tags

[3] R. Blanco et al. "Entity recommendations in web search". In: *International Semantic Web Conference*. Springer. 2013, pp. 33–48.

Entity Recommendation for Web Queries I

Entity recommendations for web search queries[7][3]

- Co-occurrence features
  - query logs, user sessions
  - flickr and twitter tags
- frequency

[3] R. Blanco et al. "Entity recommendations in web search". In: *International Semantic Web Conference*. Springer. 2013, pp. 33–48.

Entity Recommendation for Web Queries I

Entity recommendations for web search queries[7][3]

- Co-occurrence features
  - query logs, user sessions
  - flickr and twitter tags
- frequency
- Graph theoretic features
  - Page rank on entity graph
  - Common neighbors between two entities

[3] R. Blanco et al. "Entity recommendations in web search". In: *International Semantic Web Conference*. Springer. 2013, pp. 33–48.

Entity Recommendation for Web Queries II

Learning to rank using text and graph based features[23][4]

[4] M. Schuhmacher, L. Dietz, and S. Paolo Ponzetto. "Ranking entities for web queries through text and knowledge". In: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM. 2015, pp. 1461–1470.

Entity Recommendation for Web Queries II

Learning to rank using text and graph based features[23][4]

- Given a web query, retrieve relevant documents,

[4] M. Schuhmacher, L. Dietz, and S. Paolo Ponzetto. "Ranking entities for web queries through text and knowledge". In: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM. 2015, pp. 1461–1470.

Entity Recommendation for Web Queries II

Learning to rank using text and graph based features[23][4]

- Given a web query, retrieve relevant documents,
- Identify entities present in them using entity linking methods

[4] M. Schuhmacher, L. Dietz, and S. Paolo Ponzetto. "Ranking entities for web queries through text and knowledge". In: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM. 2015, pp. 1461–1470.

Entity Recommendation for Web Queries II

Learning to rank using text and graph based features[23][4]

- Given a web query, retrieve relevant documents,
- Identify entities present in them using entity linking methods
- Rank these entities using graph theoretic and text based features

[4] M. Schuhmacher, L. Dietz, and S. Paolo Ponzetto. "Ranking entities for web queries through text and knowledge". In: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM. 2015, pp. 1461–1470.

Entity Recommendation for Web Queries II

Learning to rank using text and graph based features[23][4]

- Given a web query, retrieve relevant documents,
- Identify entities present in them using entity linking methods
- Rank these entities using graph theoretic and text based features
- Reformulates entity retrieval/recommendation as ad hoc document retrieval

[4] M. Schuhmacher, L. Dietz, and S. Paolo Ponzetto. "Ranking entities for web queries through text and knowledge". In: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM. 2015, pp. 1461–1470.

Entity Exploration

Till now, we have focused on finding entities

Entity Exploration

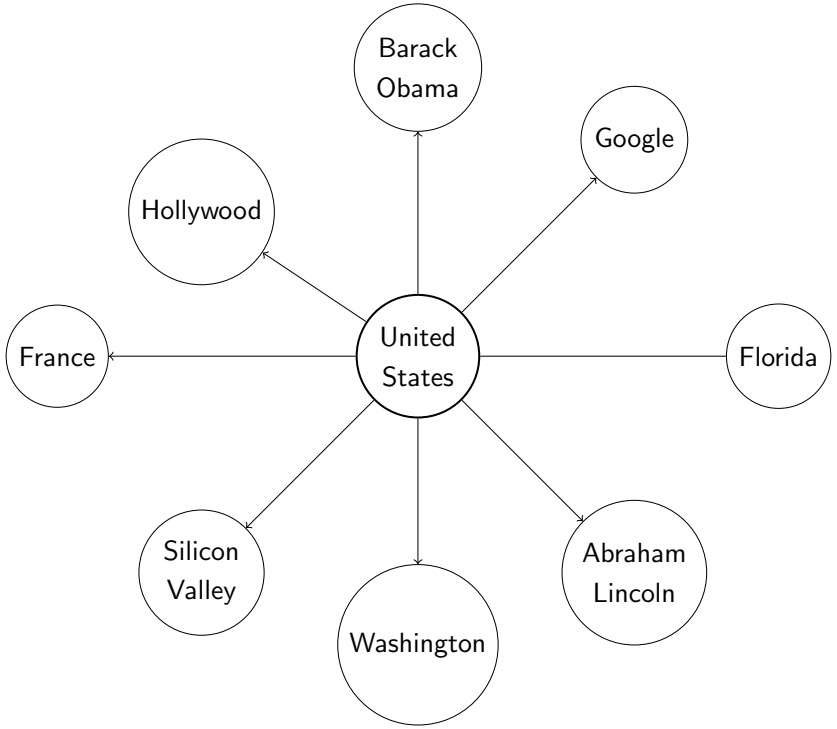
Till now, we have focused on finding entities

Let us focus our attention now on finding *about* entities

Entity Exploration

Till now, we have focused on finding entities

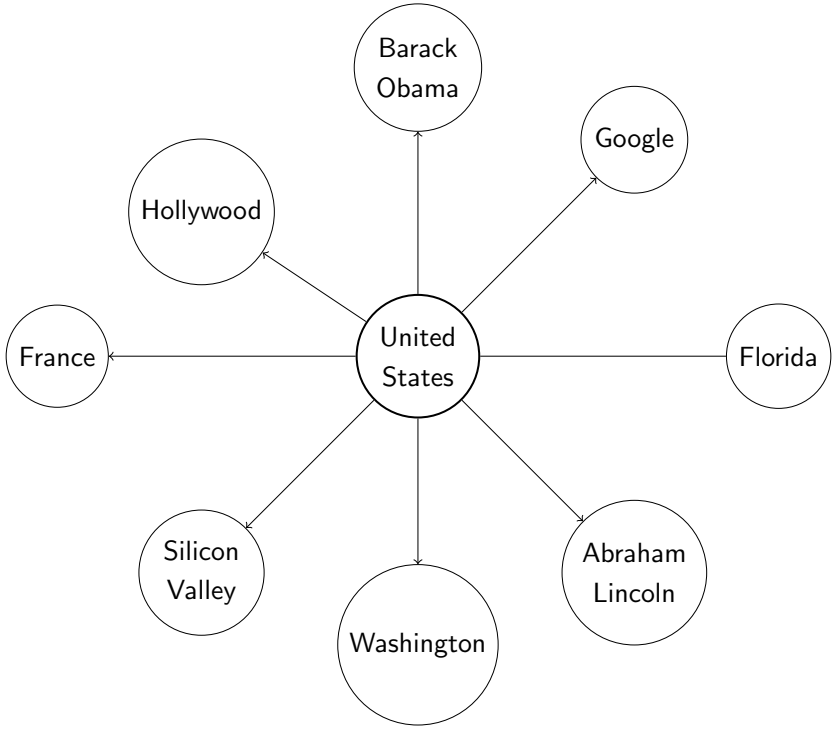
Let us focus our attention now on finding *about* entities



Entity Exploration

Till now, we have focused on finding entities

Let us focus our attention now on finding *about* entities



Relationships of similar types can be clustered and then explored based on user requirements [30]

Entity Exploration – Fact Ranking

What are the most important facts about an entity?[5]

[5] S. Bhatia et al. "Separating Wheat from the Chaff–A Relationship Ranking Algorithm". In: *International Semantic Web Conference*. Springer. 2016, pp. 79–83.

Entity Exploration – Fact Ranking

What are the most important facts about an entity?[5] Given a source entity e_s, we wish to compute the probability $P(r, e_t|e_s)$

$$P(r, e_t|e_s) \propto \underbrace{P(e_t)}_{\text{entity prior}} \times \underbrace{P(e_s|e_t)}_{\text{entity affinity}} \times \underbrace{P(r|e_s, e_t)}_{\text{relationship strength}} \tag{6}$$

[5] S. Bhatia et al. "Separating Wheat from the Chaff–A Relationship Ranking Algorithm". In: *International Semantic Web Conference*. Springer. 2016, pp. 79–83.

Entity Exploration – Fact Ranking

What are the most important facts about an entity?[5] Given a source entity e_s, we wish to compute the probability $P(r, e_t|e_s)$

$$P(r, e_t|e_s) \propto \underbrace{P(e_t)}_{\text{entity prior}} \times \underbrace{P(e_s|e_t)}_{\text{entity affinity}} \times \underbrace{P(r|e_s, e_t)}_{\text{relationship strength}} \tag{6}$$

Entity Prior:

$$P(e_t) \propto relCount(e_t) \tag{7}$$

[5] S. Bhatia et al. "Separating Wheat from the Chaff–A Relationship Ranking Algorithm". In: *International Semantic Web Conference*. Springer. 2016, pp. 79–83.

Entity Exploration – Fact Ranking

What are the most important facts about an entity?[5] Given a source entity e_s, we wish to compute the probability $P(r, e_t|e_s)$

$$P(r, e_t|e_s) \propto \underbrace{P(e_t)}_{\text{entity prior}} \times \underbrace{P(e_s|e_t)}_{\text{entity affinity}} \times \underbrace{P(r|e_s, e_t)}_{\text{relationship strength}} \tag{6}$$

Entity Prior:

$$P(e_t) \propto relCount(e_t) \tag{7}$$

Entity Affinity

$$P(e|e_t) = \frac{\sum_{r_i \in R(e_s, e_t)} w(r_i) \times r_i}{\sum_{r_i \in R(e_t)} w(r_i) \times r_i} \tag{8}$$

[5] S. Bhatia et al. "Separating Wheat from the Chaff–A Relationship Ranking Algorithm". In: *International Semantic Web Conference*. Springer. 2016, pp. 79–83.

Entity Exploration – Fact Ranking

What are the most important facts about an entity?[5] Given a source entity e_s, we wish to compute the probability $P(r, e_t|e_s)$

$$P(r, e_t|e_s) \propto \underbrace{P(e_t)}_{\text{entity prior}} \times \underbrace{P(e_s|e_t)}_{\text{entity affinity}} \times \underbrace{P(r|e_s, e_t)}_{\text{relationship strength}} \tag{6}$$

Entity Prior:

$$P(e_t) \propto relCount(e_t) \tag{7}$$

Entity Affinity

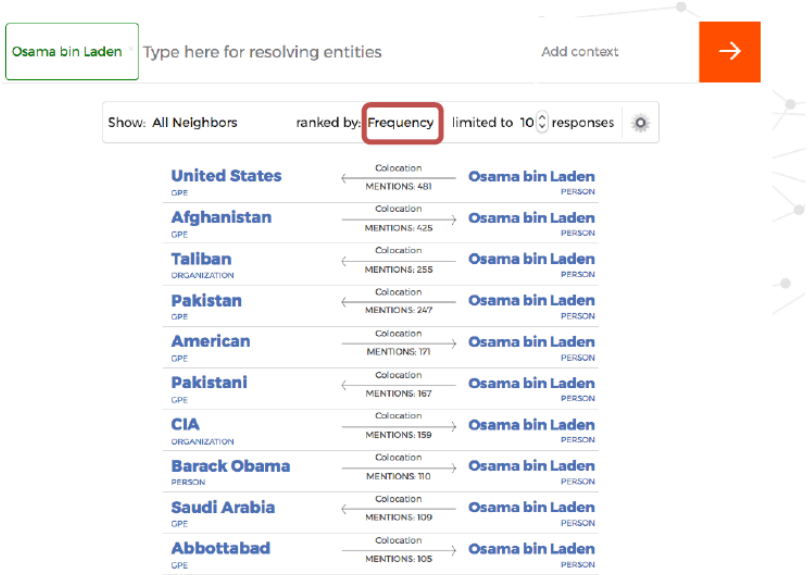
$$P(e|e_t) = \frac{\sum_{r_i \in R(e_s, e_t)} w(r_i) \times r_i}{\sum_{r_i \in R(e_t)} w(r_i) \times r_i} \tag{8}$$

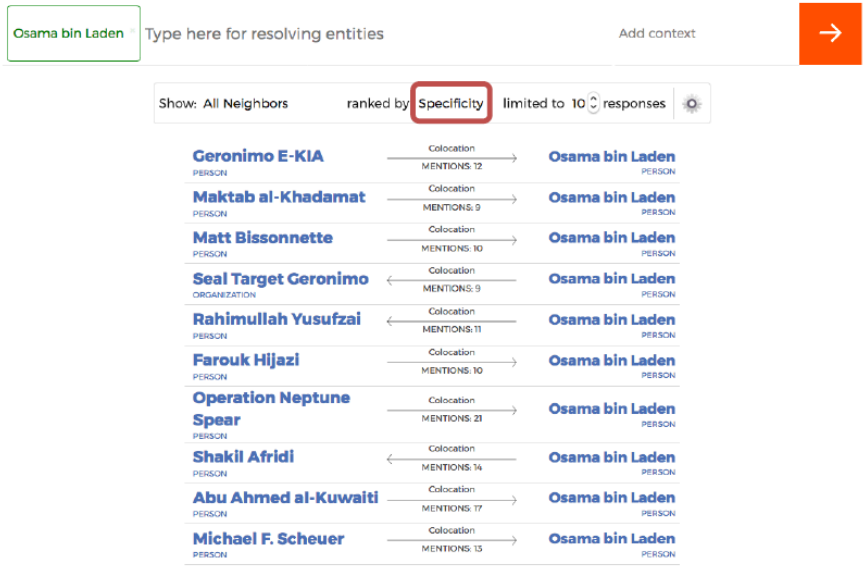
Relationship Strength

$$P(r|e_s, e_t) = \frac{mentionCount(r, e_s, e_t)}{\sum_{r \in R(e_s, e_t)} mentionCount(r, e_s, e_t)} \tag{9}$$

[5] S. Bhatia et al. "Separating Wheat from the Chaff–A Relationship Ranking Algorithm". In: *International Semantic Web Conference*. Springer. 2016, pp. 79–83.

Entity Exploration - Fact Ranking





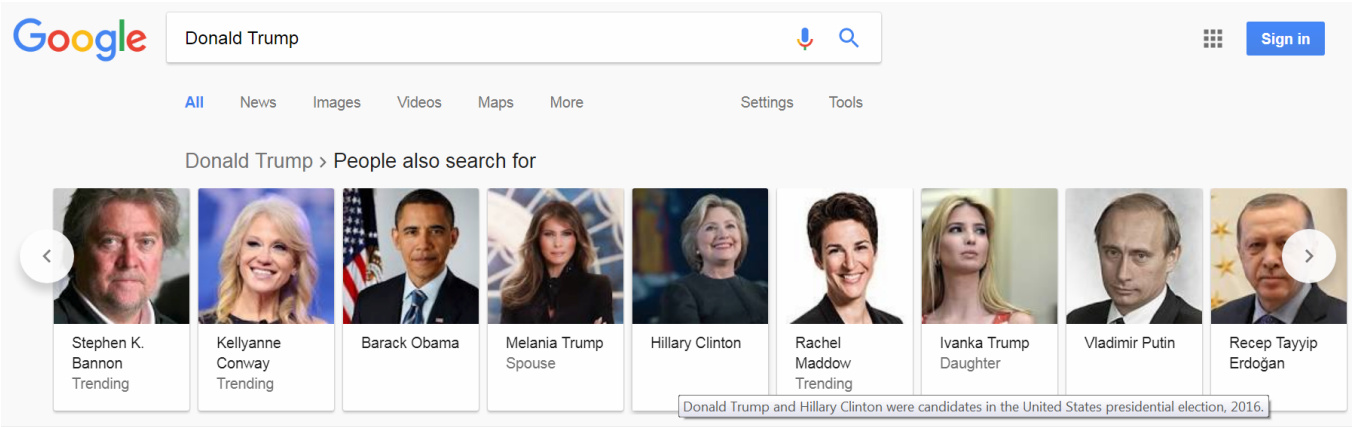
Entity Exploration – Moving Beyond the Neighborhood

Till now, we have limited our attention to relations of the entity and it's immediate neighborhood.

Entity Exploration – Moving Beyond the Neighborhood

Till now, we have limited our attention to relations of the entity and it's immediate neighborhood.

What lies after that?



Entity Exploration – Moving Beyond the Neighborhood

Discovering and Explaining Higher Order Relations Between Entities

Entity Exploration – Moving Beyond the Neighborhood

Discovering and Explaining Higher Order Relations Between Entities

Entity Exploration – Moving Beyond the Neighborhood

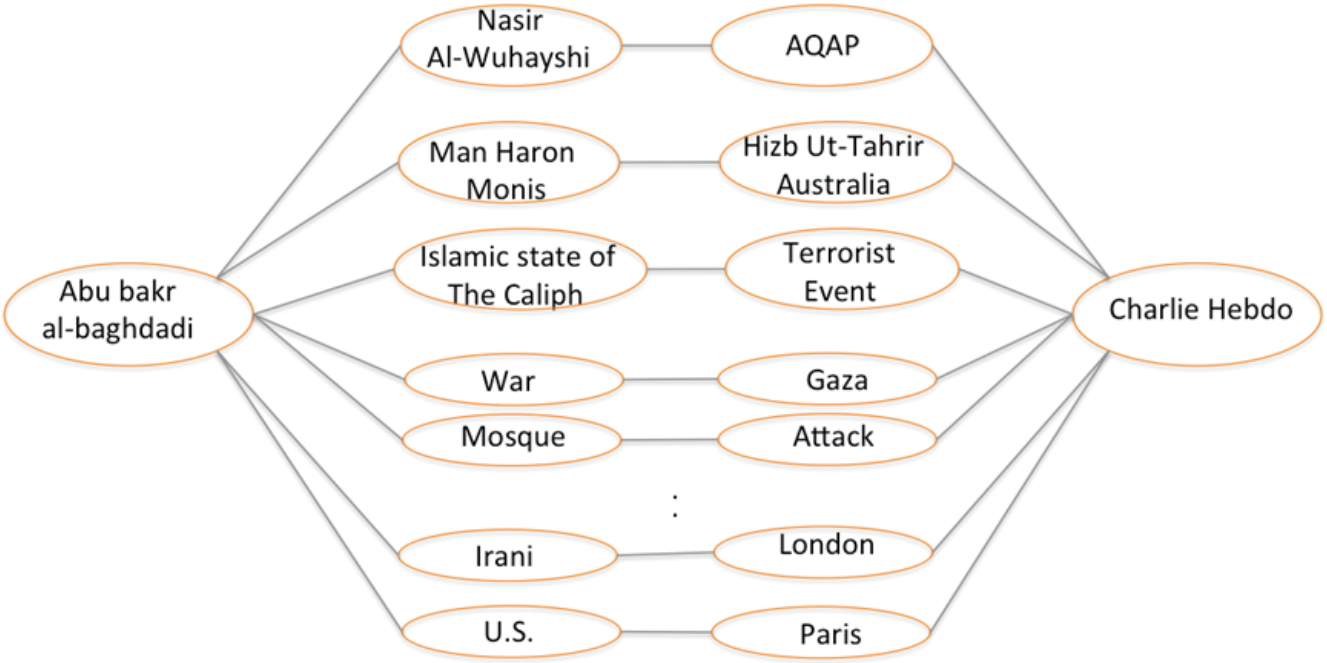
Discovering and Explaining Higher Order Relations Between Entities

Entity Exploration – Moving Beyond the Neighborhood

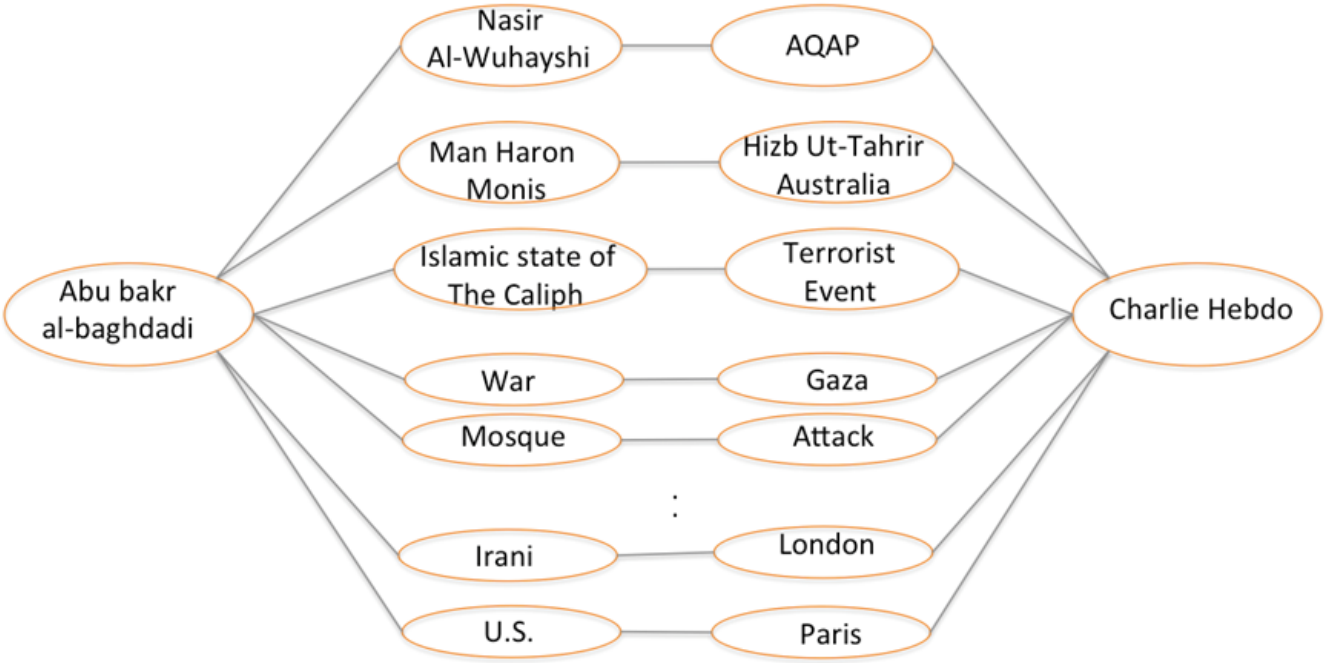
Discovering and Explaining Higher Order Relations Between Entities

Can we tell how are they connected?

Path Ranking



Path Ranking



- Thousands of such paths
- Too generic – obvious relations

Path Ranking

Three components for ranking possible paths [2]

Path Ranking

Three components for ranking possible paths [2]

Specificity: Popular entities given lower scores

$$spec(p) = \sum_{e \in p} spec(e); \text{where: } spec(e) = log(1 + 1/docCount(e)) \quad (10)$$

Reduces generic paths, but boosts noise entities

Path Ranking

Three components for ranking possible paths [2]

Specificity: Popular entities given lower scores

$$spec(p) = \sum_{e \in p} spec(e); \text{where: } spec(e) = log(1 + 1/docCount(e)) \quad (10)$$

Reduces generic paths, but boosts noise entities

Connectivity: A strongly connected path consists of strong edges.

$$score(e_a, e_b) = \vec{d_{e\,a}} \cdot \vec{d_{e\,b}} \quad (11)$$

Path Ranking

Three components for ranking possible paths [2]

Specificity: Popular entities given lower scores

$$spec(p) = \sum_{e \in p} spec(e); \text{where: } spec(e) = log(1 + 1/docCount(e)) \quad (10)$$

Reduces generic paths, but boosts noise entities

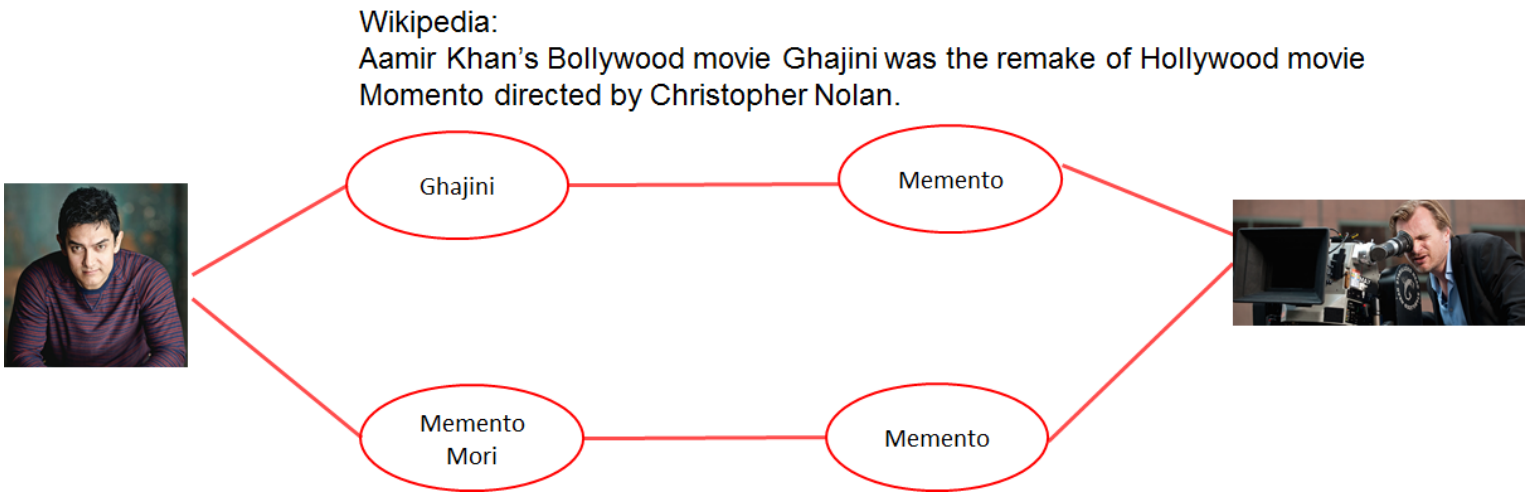
Connectivity: A strongly connected path consists of strong edges.

$$score(e_a, e_b) = \vec{d_{e\,a}} \cdot \vec{d_{e\,b}} \quad (11)$$

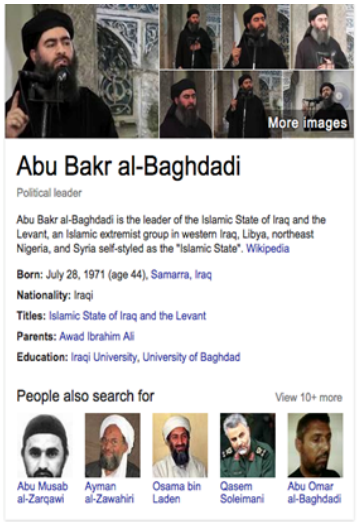
Cohesiveness:

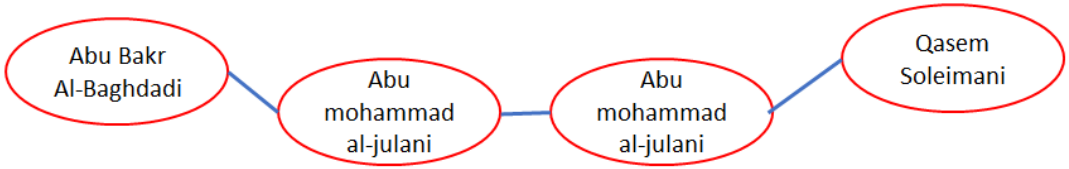
$$score(p) = \sum_{i=2}^{n-1} score(e_i) = \sum_{i=2}^{n-1} \vec{d_{e\,i-1}} \cdot \vec{d_{e_{i+1}}} \quad (12)$$

Path Ranking



Path Ranking





Wikipedia:
Shortly after the Syrian uprising began against the Syrian administration headed by Syrian president Bashar al-Assad, al-Julani moved into Syrian territory and, fully supported by al-Baghdadi....
Bashar was supported by major general Qasem Soleimani ....

Application Example from Life Sciences

Predicting Drug-Drug Interactions(DDI)[6]

[6] A. Fokoue et al. "Predicting drug-drug interactions through large-scale similarity-based link prediction". In: *International Semantic Web Conference*. Springer. 2016, pp. 774–789.

Application Example from Life Sciences

Predicting Drug-Drug Interactions(DDI)[6]

- DDI are a major cause of preventable adverse drug reactions

[6] A. Fokoue et al. "Predicting drug-drug interactions through large-scale similarity-based link prediction". In: *International Semantic Web Conference*. Springer. 2016, pp. 774–789.

Application Example from Life Sciences

Predicting Drug-Drug Interactions(DDI)[6]

- DDI are a major cause of preventable adverse drug reactions
- Clinical studies can not accurately determine all possible DDIs

[6] A. Fokoue et al. "Predicting drug-drug interactions through large-scale similarity-based link prediction". In: *International Semantic Web Conference*. Springer. 2016, pp. 774–789.

Application Example from Life Sciences

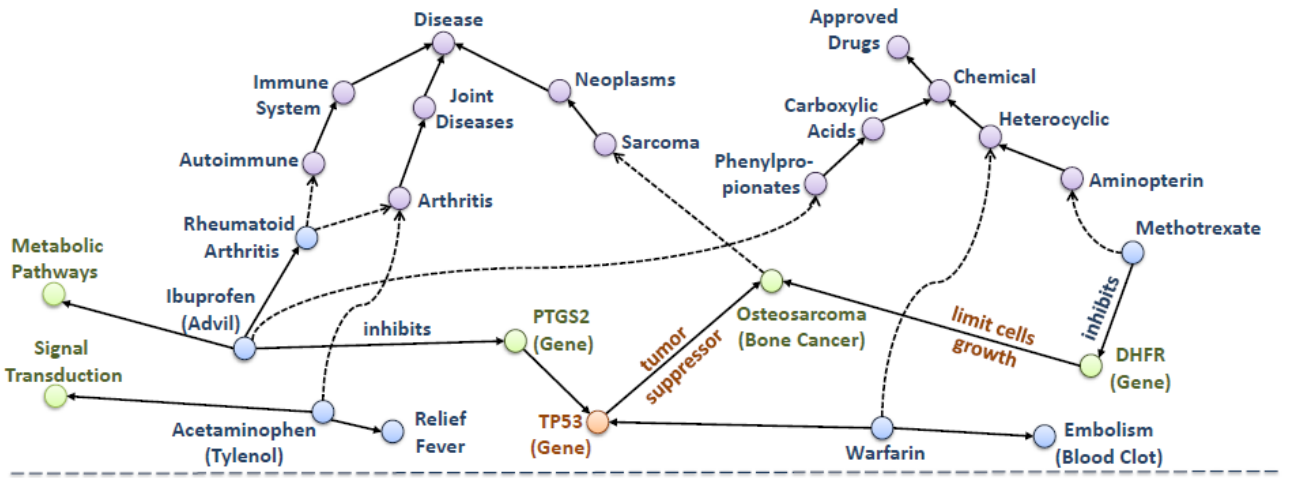
Predicting Drug-Drug Interactions(DDI)[6]

- DDI are a major cause of preventable adverse drug reactions
- Clinical studies can not accurately determine all possible DDIs
- Can we utilize knowledge about drugs to predict possible DDIs?

[6] A. Fokoue et al. "Predicting drug-drug interactions through large-scale similarity-based link prediction". In: *International Semantic Web Conference*. Springer. 2016, pp. 774–789.

Application Example from Life Sciences

Create a KG out of existing information about drugs and their interactions with genes, enzymes, molecules, etc.



DrugBank: Bioinformatics & Cheminformatics Resource

Drug Name	Drug Targets (Genes)	Symptomatic Treatment
Ibuprofen	PTGS2	Rheumatoid Arthritis
Acetaminophen	PTGS2	Relief Fever
Methotrexate	DHFR	Antineoplastic Anti-metabolite
Warfarin	TP53	Embolism (Blood Clot)

CTD: Comparative Toxicogenomics Database

Gene	Interaction
PTGS2	TP53 (Gene)

Gene	Disease
TP53	Osteosarcoma

Chemical	Pathways	Linked Data Source
Ibuprofen	Metabolic Pathways	KEGG
Acetaminophen	Signal Transduction	Reactome
Methotrexate	Immune System	Reactome

Uniprot: Universal Protein Resource

Gene	Function
TP53	Tumor Suppressor
DHFR	Limits Cell Growth

Application Example from Life Sciences

- Given a pair of drugs, extract features based on physiological effect, side effect, targets, drug targets, chemical structure, etc.
- Perform supervised classification using logistic regression
- Retrospective Analysis: Known DDIs til January 2011 as training.
- Could predict $\approx 68\%$ of DDIs discovered after January 2011 till December 2014.

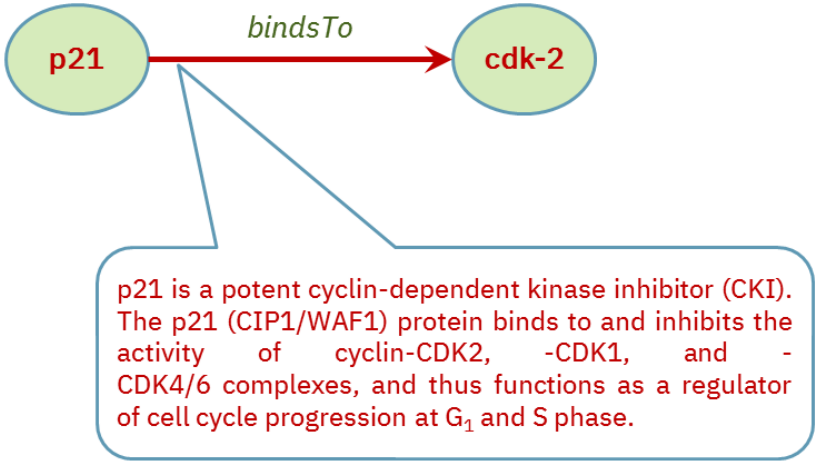
Upcoming Areas and Future Directions

Future Research Directions

Explaining relations present in a graph [26, 16, 4]

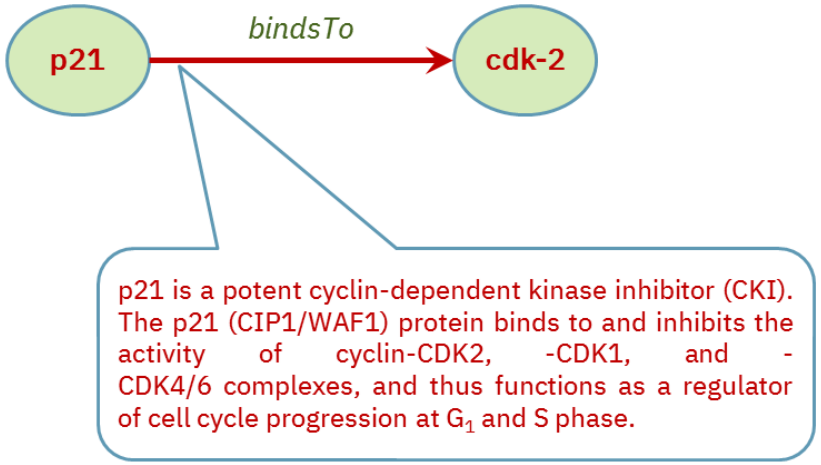
Future Research Directions

Explaining relations present in a graph [26, 16, 4]



Future Research Directions

Explaining relations present in a graph [26, 16, 4]



Graph and text joint modeling [27, 31]

Future Research Directions

- Reasoning over Knowledge Graphs
  - KG Completion [10, 24, 17]
  - Complex QA Systems

Future Research Directions

- Reasoning over Knowledge Graphs
  - KG Completion [10, 24, 17]
  - Complex QA Systems

Question: Which type of scientist would study the relationship between simple machines and energy?
(A) chemist; (B) biologist; (C) physicist; (D) geologist

Future Research Directions

- Reasoning over Knowledge Graphs
  - KG Completion [10, 24, 17]
  - Complex QA Systems

Question: Which type of scientist would study the relationship between simple machines and energy?
(A) chemist; (B) biologist; (C) physicist; (D) geologist

Question: Which is a physical property of an apple?
(A) what color it is; (B) how pretty it is; (C) how much it costs; (D) when it was picked

Future Research Directions

Complex QA Systems Allen Institute for AI has a complex QA challenge running

Future Research Directions

Complex QA Systems Allen Institute for AI has a complex QA challenge running

- Easy Set: Best system has an accuracy of 68.90%; IR-Solver has 62.55%
  - Challenge Set: Best system has an accuracy of 42.32%; IR-solver has 20.26%

Future Research Directions

Complex QA Systems Allen Institute for AI has a complex QA challenge running

- Easy Set: Best system has an accuracy of 68.90%; IR-Solver has 62.55%
  - Challenge Set: Best system has an accuracy of 42.32%; IR-solver has 20.26%

IR-Solver is pretty basic – tf-idf scoring

Future Research Directions

Complex QA Systems Allen Institute for AI has a complex QA challenge running

- Easy Set: Best system has an accuracy of 68.90%; IR-Solver has 62.55%
  - Challenge Set: Best system has an accuracy of 42.32%; IR-solver has 20.26%

IR-Solver is pretty basic – tf-idf scoring

I know we can certainly do much better

Future Research Directions

Complex QA Systems Allen Institute for AI has a complex QA challenge running

- Easy Set: Best system has an accuracy of 68.90%; IR-Solver has 62.55%
  - Challenge Set: Best system has an accuracy of 42.32%; IR-solver has 20.26%

IR-Solver is pretty basic – tf-idf scoring

I know we can certainly do much better (because I have done with a student)

Future Research Directions

Complex QA Systems Allen Institute for AI has a complex QA challenge running

- Easy Set: Best system has an accuracy of 68.90%; IR-Solver has 62.55%
  - Challenge Set: Best system has an accuracy of 42.32%; IR-solver has 20.26%

IR-Solver is pretty basic – tf-idf scoring

I know we can certainly do much better (because I have done with a student)

Currently, we have around 70% on Easy Set and 38% on challenge set using language modeling, query expansion, and relevance feedback.

Future Research Directions

Complex QA Systems Allen Institute for AI has a complex QA challenge running

- Easy Set: Best system has an accuracy of 68.90%; IR-Solver has 62.55%
  - Challenge Set: Best system has an accuracy of 42.32%; IR-solver has 20.26%

IR-Solver is pretty basic – tf-idf scoring

I know we can certainly do much better (because I have done with a student)

Currently, we have around 70% on Easy Set and 38% on challenge set using language modeling, query expansion, and relevance feedback.

Goal is to understand limitations of IR and where advanced reading

Conclusions

- KG can provide structure to your unstructured data!

Conclusions

- KG can provide structure to your unstructured data!
- They thus complement traditional IR paradigm of unstructured data

Conclusions

- KG can provide structure to your unstructured data!
- They thus complement traditional IR paradigm of unstructured data
- Many time-tested IR principles can in turn benefit Knowledge Graph exploration

Conclusions

- KG can provide structure to your unstructured data!
- They thus complement traditional IR paradigm of unstructured data
- Many time-tested IR principles can in turn benefit Knowledge Graph exploration
- We wanted to provide an overview of tools/techniques that have worked well in the past, and challenges you may face

Conclusions

- KG can provide structure to your unstructured data!
- They thus complement traditional IR paradigm of unstructured data
- Many time-tested IR principles can in turn benefit Knowledge Graph exploration
- We wanted to provide an overview of tools/techniques that have worked well in the past, and challenges you may face
- Be careful in selecting the KG appropriate for your domain and requirements.

Conclusions

- KG can provide structure to your unstructured data!
- They thus complement traditional IR paradigm of unstructured data
- Many time-tested IR principles can in turn benefit Knowledge Graph exploration
- We wanted to provide an overview of tools/techniques that have worked well in the past, and challenges you may face
- Be careful in selecting the KG appropriate for your domain and requirements.
- Keep in mind the scale and efficiency issues

Thanks!!!
Suggestions and Questions Welcome!

Slides available at `http://sumitbhatia.net/source/knowledge-graph-tutorial.html`

References i

[1] L. A. Adamic and E. Adar. "Friends and neighbors on the web". In: *Social networks* 25.3 (2003), pp. 211–230.

[2] N. Aggarwal, S. Bhatia, and V. Misra. "Connecting the Dots: Explaining Relationships Between Unconnected Entities in a Knowledge Graph". In: *International Semantic Web Conference*. Springer. 2016, pp. 35–39.

[3] K. Balog et al. "Overview of the TREC 2009 entity track". In: *In Proceedings of the Eighteenth Text REtrieval Conference*. 2009.

[4] S. Bhatia, P. Dwivedi, and A. Kaur. "Thats Interesting, Tell Me More! Finding Descriptive Support Passages for Knowledge Graph Relationships". In: *International Semantic Web Conference*. Springer. 2018, pp. 250–267.

[5] S. Bhatia and A. Jain. "Context Sensitive Entity Linking of Search Queries in Enterprise Knowledge Graphs". In: *International Semantic Web Conference*. Springer. 2016, pp. 50–54.

[6] S. Bhatia et al. "Separating Wheat from the Chaff–A Relationship Ranking Algorithm". In: *International Semantic Web Conference*. Springer. 2016, pp. 79–83.

References ii

[7] R. Blanco et al. "Entity recommendations in web search". In: *International Semantic Web Conference*. Springer. 2013, pp. 33–48.

[8] R. C. Bodner and F. Song. "Knowledge-based approaches to query expansion in information retrieval". In: *Conference of the Canadian Society for Computational Studies of Intelligence*. Springer. 1996, pp. 146–158.

[9] M. Bron, K. Balog, and M. De Rijke. "Ranking related entities: components and analyses". In: *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM. 2010, pp. 1079–1088.

[10] R. Das et al. "Chains of reasoning over entities, relations, and text using recurrent neural networks". In: *arXiv preprint arXiv:1607.01426* (2016).

[11] A. Fokoue et al. "Predicting drug-drug interactions through large-scale similarity-based link prediction". In: *International Semantic Web Conference*. Springer. 2016, pp. 774–789.

[12] A. Gattani et al. "Entity extraction, linking, classification, and tagging for social media: a wikipedia-based approach". In: *Proceedings of the VLDB Endowment* 6.11 (2013), pp. 1126–1137.

References iii

[13] S. Guo, M.-W. Chang, and E. Kiciman. "To Link or Not to Link? A Study on End-to-End Tweet Entity Linking.". In: *HLT-NAACL*. 2013, pp. 1020–1030.

[14] B. Hachey et al. "Evaluating Entity Linking with Wikipedia". In: *Artif. Intell.* 194 (Jan. 2013), pp. 130–150.

[15] J. Hoffart et al. "Robust disambiguation of named entities in text". In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. 2011, pp. 782–792.

[16] J. Huang et al. "Generating Recommendation Evidence Using Translation Model.". In: *IJCAI*. 2016, pp. 2810–2816.

[17] Y. Lin et al. "Learning Entity and Relation Embeddings for Knowledge Graph Completion.". In: *AAAI*. 2015, pp. 2181–2187.

[18] C. D. Manning et al. "The Stanford CoreNLP Natural Language Processing Toolkit". In: *Association for Computational Linguistics (ACL) System Demonstrations*. 2014, pp. 55–60.

[19] D. Nadeau and S. Sekine. "A survey of named entity recognition and classification". In: *Lingvisticae Investigationes* 30.1 (2007), pp. 3–26.

References iv

[20] M. e. a. Nagarajan. "Predicting Future Scientific Discoveries Based on a Networked Analysis of the Past Literature". In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '15. Sydney, NSW, Australia: ACM, 2015, pp. 2019–2028.

[21] J. Pound, P. Mika, and H. Zaragoza. "Ad-hoc Object Retrieval in the Web of Data". In: *Proceedings of the 19th International Conference on World Wide Web*. WWW '10. Raleigh, North Carolina, USA: ACM, 2010, pp. 771–780.

[22] L. Ratinov et al. "Local and global algorithms for disambiguation to wikipedia". In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics. 2011, pp. 1375–1384.

[23] M. Schuhmacher, L. Dietz, and S. Paolo Ponzetto. "Ranking entities for web queries through text and knowledge". In: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM. 2015, pp. 1461–1470.

[24] R. Socher et al. "Reasoning with neural tensor networks for knowledge base completion". In: *Advances in neural information processing systems*. 2013, pp. 926–934.

References v

[25] F. M. Suchanek, G. Kasneci, and G. Weikum. "Yago: a core of semantic knowledge". In: *Proceedings of the 16th international conference on World Wide Web*. ACM. 2007, pp. 697–706.

[26] N. Voskarides et al. "Learning to explain entity relationships in knowledge graphs". In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and The 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL-IJCNLP 2015)*. 2015, p. 11.

[27] Z. Wang et al. "Knowledge Graph and Text Jointly Embedding.". In: *EMNLP*. Vol. 14. 2014, pp. 1591–1601.

[28] I. H. Witten and D. N. Milne. "An effective, low-cost measure of semantic relatedness obtained from Wikipedia links". In: (2008).

[29] C. Xiong and J. Callan. "Query expansion with Freebase". In: *Proceedings of the 2015 international conference on the theory of information retrieval*. ACM. 2015, pp. 111–120.

References vi

[30] Y. Zhang, G. Cheng, and Y. Qu. "Towards exploratory relationship search: A clustering-based approach". In: *Joint International Semantic Technology Conference*. Springer. 2013, pp. 277–293.

[31] H. Zhong et al. "Aligning Knowledge and Text Embeddings by Entity Descriptions.". In: *EMNLP*. 2015, pp. 267–272.