From Graph to Knowledge Graph: Algorithms and Applications

Module 5: Knowledge Graph Inference and Applications

Outline

- Knowledge Graph Inference
  - WHAT - definition
  - WHY – why do we need it and why it is important?
  - HOW
    - Within existing KG – modeling approaches
      - Graph feature model
      - Latent feature models (KG embeddings)
    - Use external sources – QnA system

- Knowledge Graph Applications
  - Entity Recommendation
  - Question and Answering

Knowledge Graph Inference

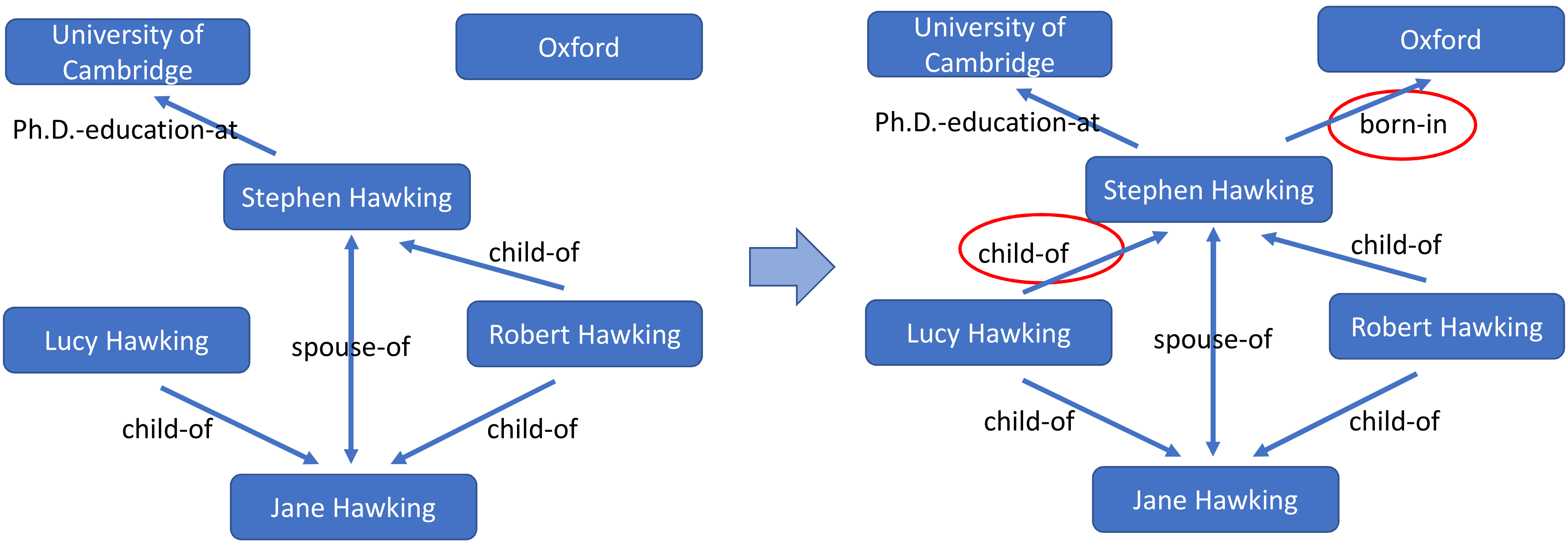
- ***WHAT – definition***

- WHY – why do we need it and why it is important?

- HOW
  - Within existing KG
    - Modeling overview
    - Graph feature model -- Path Ranking Algorithm (PRA)
    - Latent feature models (KG embeddings)
  - Use external sources – QnA system

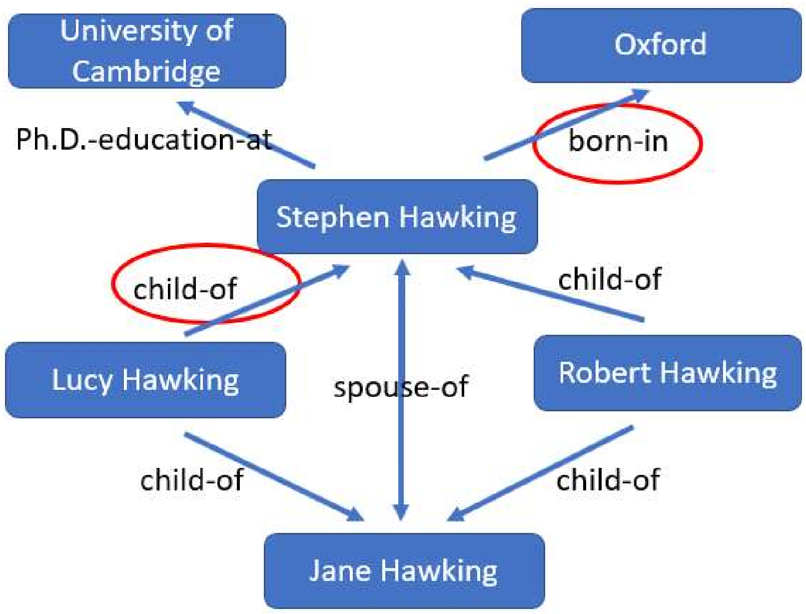
- General assumptions
- Key components in model
- Milestone models
  - RESCAL and tensor factorization
  - Structured Embedding
  - Neural Tensor Networks (NTN)
  - TransE and DISTMULT

Problem Definition



Problem Definition

- Add knowledge from existing graph
- With external source
  - Using external documents to confirm the existence of relationship
- Without external source
  - Reasoning within the graph
  - Link prediction problem



Knowledge Graph Inference

- WHAT – definition
- ***WHY – why do we need it and why it is important?***
- HOW
  - Within existing KG
    - Modeling overview
    - Graph feature model -- Path Ranking Algorithm (PRA)
    - Latent feature models (KG embeddings)
  - Use external sources – QnA system

Why is Knowledge Graph Inference Important?

- Knowledge Base is largely incomplete

- Need systematic and scalable approaches to complete knowledge graph

Freebase

Relation	Percentage unknown	
	All 3M	*Top 100K*
PROFESSION	68%	24%
PLACE OF BIRTH	71%	13%
NATIONALITY	75%	21%
EDUCATION	91%	63%
SPOUSES	92%	68%
PARENTS	94%	77%
CHILDREN	94%	80%
SIBLINGS	96%	83%
ETHNICITY	99%	86%

Incompleteness of Freebase for some relations that apply to entities of type PERSON. Left: all 3M Freebase PERSON entities. Right: only the 100K most frequent PERSON entities.

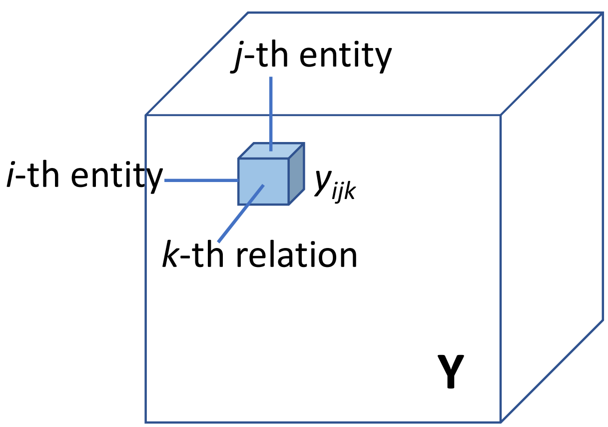
West, et al., WWW'14, Knowledge Base Completion via Search-Based Question Answering

Knowledge Graph Inference

- WHAT – definition
- WHY – why do we need it and why it is important?

- HOW
  - ***Within existing KG***
    - ***Modeling overview***
      - Graph feature model -- Path Ranking Algorithm (PRA)
      - Latent feature models (KG embeddings)
  - Use external sources – QnA system

Statistical Relational Learning



$$Y_{ijk} = \begin{cases} 1, & \text{if the triple } (e_i, r_k, e_j) \text{ exists;} \\ 0, & \text{otherwise.} \end{cases}$$

Element-wise

$$Y \in \{0, 1\}^{Ne \times Ne \times Nr}$$

adjacency tensor
(adjacency matrix)

$$P(Y)$$

Estimate the joint-distribution

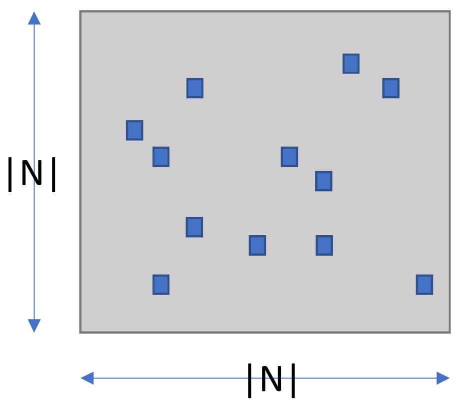
$$P(yijk)$$

Predict *unobserved* triples

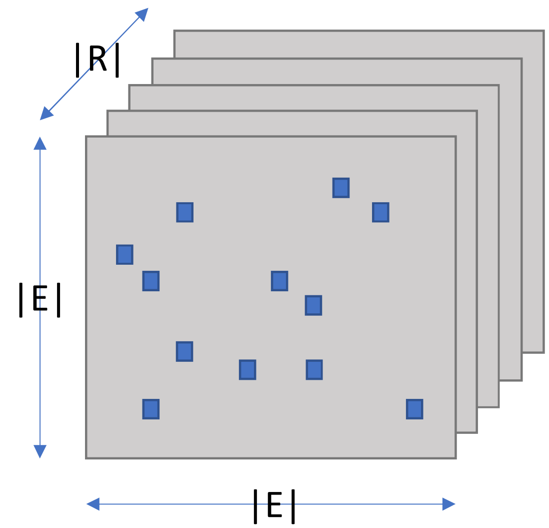
Nickel, et al., *Proceedings of the IEEE* 2016, A review of relational machine learning for knowledge graphs.

Graph vs Knowledge Graph Relations

- Single relation vs Multiple relations



|N|: Number of Nodes in Graph



|E|: Number of Entities in Knowledge Graph
|R|: Number of Relations in Knowledge Graph

Knowledge Graph Inference

- WHAT – definition
- WHY – why do we need it and why it is important?

- HOW
  - ***Within existing KG***
    - Modeling overview
    - ***Graph feature model -- Path Ranking Algorithm (PRA)***
    - Latent feature models (KG embeddings)
  - Use external sources – QnA system

Modeling technique

- ***Graph Feature*** Models
  - Intuition: Similar entities are likely to be related
  - "Similarity":
    - Local – Common Neighbors
    - Global – Random Walk
    - *Quasi-local* – ***Random Walk with Bounded Length***

Path Ranking Algorithm

- ***Latent Feature*** Models (***Embedding***)

Graph Feature Model - PRA

- Path Ranking Algorithm (PRA)
  - A relation path $P=(R_1, \ldots, R_n)$ is a sequence of relations
  - A PRA model scores a source-target node pair by a linear function of their path features

    $$score(s,t) = \sum_{p \in \mathrm{P}} f_p(s,t)\,\theta_p$$

    where P is the set of all relation paths with length <= L,
    $f_p(s,t) = Prob(s \rightarrow t; p)$

Lao, et al., EMNLP'11, Random Walk Inference and Learning in A Large Scale Knowledge Base

Graph Feature Model - PRA

- Training
  - For a relation R and a set of node pairs $\{(s_i, t_i)\}$,
  - Construct a training dataset $D = \{(x_i, y_i)\}$, where
  - x_i is a vector of all the path features for (s_i, t_i), and
  - y_i indicates whether $R(s_i, t_i)$ is true or not
  - Train a logistic function to predict $P(Y|X, \theta)$
    - θ is the parameter vector
  - θ is estimated using L1, L2-regularized logistic regression

Lao, et al., EMNLP'11, Random Walk Inference and Learning in A Large Scale Knowledge Base

Knowledge Graph Inference

- WHAT – definition
- WHY – why do we need it and why it is important?

- HOW
  - ***Within existing KG***
    - Modeling overview
    - Graph feature model -- Path Ranking Algorithm (PRA)

    - ***Latent feature models (KG embeddings)***

  - Use external sources – QnA system

Modeling technique

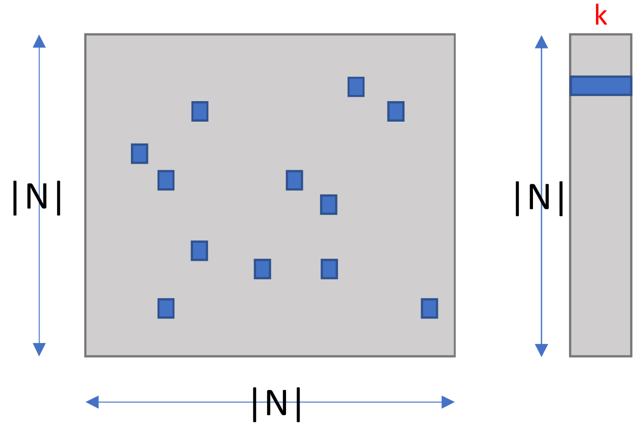
- ***Graph Feature*** Models

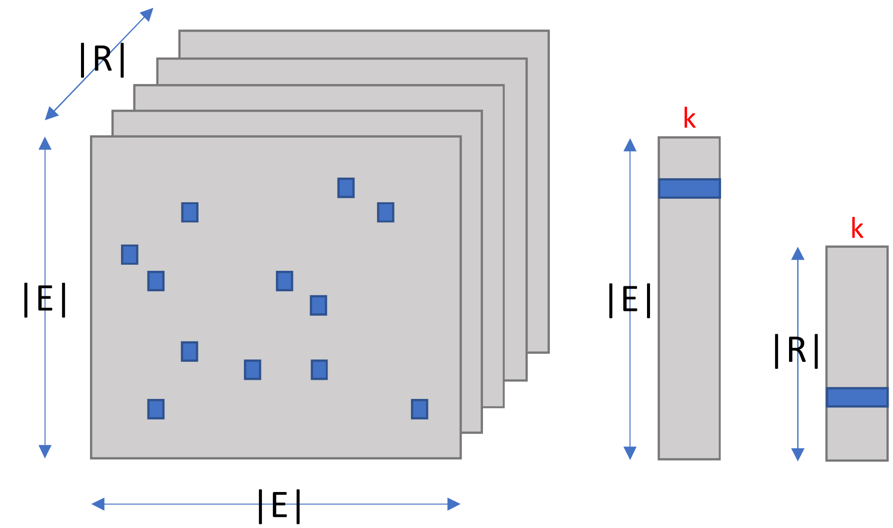
- ***Latent Feature*** Models (***Embedding***)
  - General assumption
  - Key components in model
    - *Entity* representation
    - *Relationship* (edge) representation
    - *Entity-Relation* interaction
    - Embedding *scoring function*
  - Milestone models
    - RESCAL and tensor factorization
    - Structured Embedding
    - Neural Tensor Networks (NTN)
    - TransE and DISTMULT

Graph vs Knowledge Graph

-- Relations and Embeddings

- Goal: learn the representation from the Graph/Knowledge Graph





|N|: Number of Nodes in Graph
k: Dimensionality

|E|: Number of Entities in Knowledge Graph
|R|: Number of Relations in Knowledge Graph
k: Dimensionality

Embedding Approaches - General Assumption

- Entities are vectors (low-dimensional space)
- Relation types are operators (on entity vectors) in forms of matrix or vector
- Embeddings are trained to define a similarity score on triples such that "the score of correct triple > the score of corrupted triple", for example:

Score(Seahawks, HomeField, Seattle) > Score(Seahawks, HomeField, LA)

Embedding Approaches - General Assumption

- Training embedding methods
  - Training by ranking triples from the knowledge graph and generated negative samples
  - Optimization through stochastic gradient descent
  - For each triple from the training set such as (Seahawks, HomeField, Seattle):
    - Generate negative samples by randomly replacing one entity or the relation type
      - (Seahawks, HomeField, LA)
      - (Seahawks, Coach, Seattle)
      - (Patriots, HomeField, Seattle)
    - Check the similarity score
      Score(Seahawks, HomeField, Seattle) > Score(Seahawks, HomeField, LA) + 1
    - If not satisfied, parameters of the considered triples are updated

Latent Feature Models (Embedding)

- ***Entity* representation**
  - Low dimensional vector: e_{i}
  - Initialization
    - Random
    - Average word vector with pre-trained vectors (V_{word}), e.g.

$$e_{homo\ sapiens} = 0.5 \times (V_{homo} + V_{sapiens})$$

- *Relationship* (edge) representation
- *Entity-Relation* interaction
- Embedding *scoring function*

Latent Feature Models (Embedding)

- ***Entity*** representation
- ***Relationship* (edge) representation**
  - Each relation type as *matrix*:
    - W_k : bilinear weight matrix
    - A_k : linear feature map
  - Each relation type as *vector*: r_k
- ***Entity-Relation*** interaction
- Embedding ***scoring function***

Latent Feature Models (Embedding)

- *Entity* representation
- *Relationship* (edge) representation
- ***Entity-Relation* interaction**
    - Linear : $A_k \, e_i$
    - Bilinear: $e_i^T W_k \, e_j$
- Embedding *scoring function*

Latent Feature Models (Embedding)

- *Entity* representation
- *Relationship* (edge) representation
- *Entity-Relation* interaction
- **Embedding *scoring function***
    - Margin-based ranking loss

Maximize the margin btw existing & non-existing triples

$$\sum_{(s,r,o)\in T} \sum_{(s',r,o')\in T'(s,r,o)} max(0, 1 + f(s',r,o') - f(s,r,o))$$

- Negative sampling loss

Negative log-likelihood of the correct triples & sampled corrupted triples

$$-\sum_{(s,r,o)\in T} \left(\log\sigma\big(f(s,r,o)\big) + \sum_{(s',r,o')\in T'(s,r,o)} \log\sigma\big(-f(s',r,o')\big) \right)$$

Milestones for KG embeddings

Time

DistMult

Yang, et al. ***ICLR***, Embedding entities and relations for learning and inference in knowledge bases

2015

TransE

Bordes, et al. ***NIPS***, Translating embeddings for modeling multi-relational data

Neural Tensor Networks (NTN)

Socher, et al. ***NIPS***, Reasoning with neural tensor networks for knowledge base completion

2013

Structured Embedding (SE)

Bordes, et al. ***AAAI***, Learning structured embedding of knowledge base

RESCAL

Nickel, et al. ***ICML***, A three-way model for collective learning on multi-relational data

2011

RESCAL

2011

RESCAL

Nickel, et al. *ICML,* A three-way model for collective learning on multi-relational data

$$X_k \approx A R_k A^T$$

- The optimization problem

$$\min_{A, R_k} loss(A, R_k) + reg(A, R_k)$$

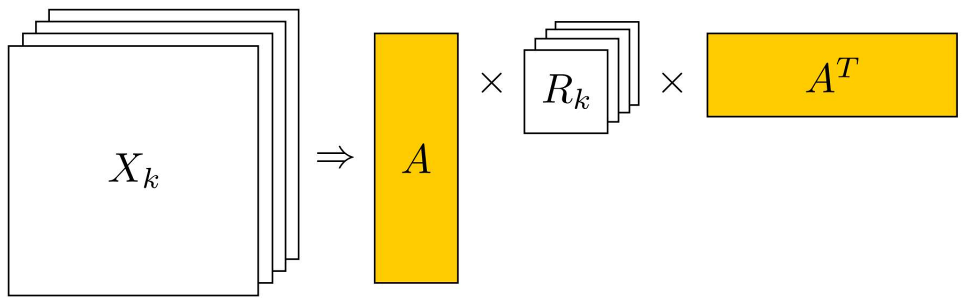
- Loss function

$$loss(A, R_k) = \frac{1}{2} \sum_k \|\mathcal{X}_k - A R_k A^T\|_F^2$$

- Regularization term

$$reg(A, R_k) = \frac{1}{2}\lambda \left(\|A\|_F^2 + \sum_k \|R_k\|_F^2 \right)$$

- Learned using alternating-least squares algorithm



Entity representation : Vector
Relation representation : Matrix
Entity-relation interaction : Bilinear
Scoring function:

RESCAL

2011 **RESCAL**

Nickel, et al. ***ICML,*** A three-way model for collective learning on multi-relational data

Tensor Factorization

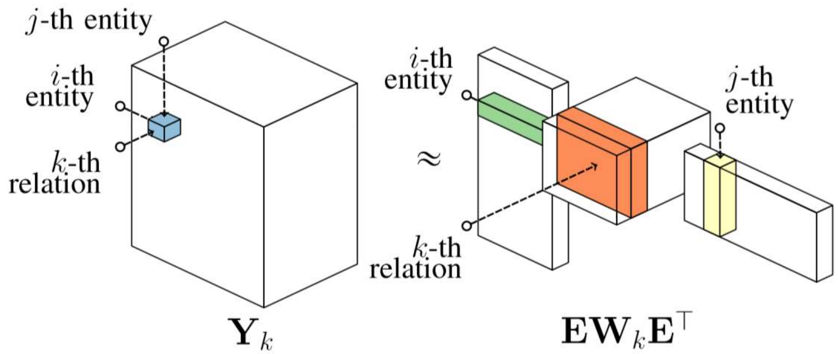


Fig. 4. RESCAL as a tensor factorization of the adjacency tensor $\mathbf{Y}$.

Nickel, et al., *Proceeding of the IEEE*, 2016.
A review of relational machine learning for knowledge graphs

- W, V, U are learned by alternating least squares
- Does not take advantage of the symmetry of the tensor

Structured Embedding

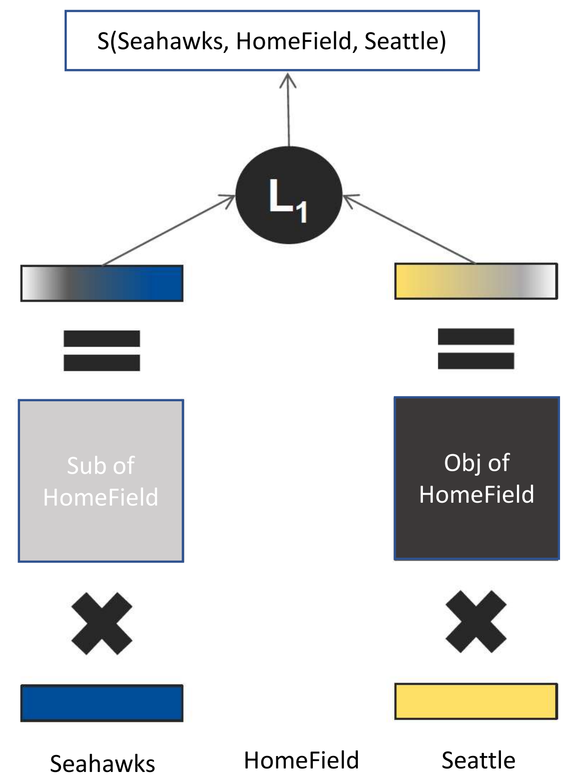
Structured Embedding (SE)

Bordes, et al. *AAAI*, Learning structured embedding of knowledge base

2011

Entity representation : Vector
Relation representation : Matrix
Entity-relation interaction : Linear
Scoring function: margin-based ranking loss

- Each entity is represented by 1 vector
- Each relation is represented by 2 matrices
- Score function: L1 distance between projected embeddings



Neural Tensor Networks

2013

Neural Tensor Networks (NTN)

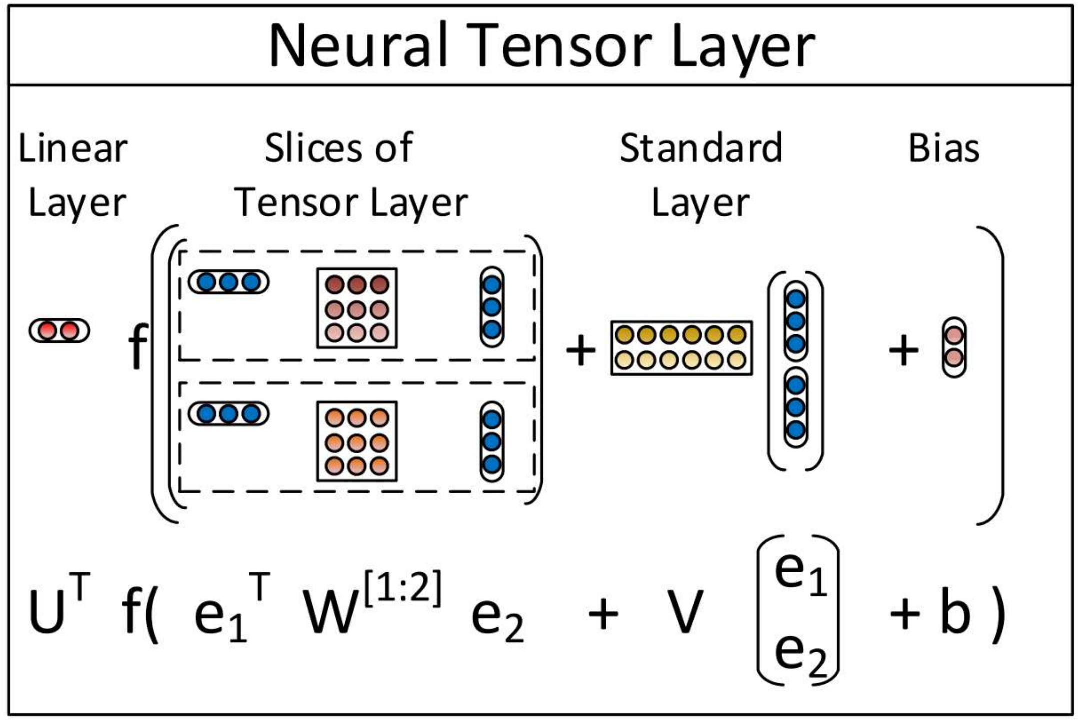
Socher, et al. ***NIPS***, Reasoning with neural tensor networks *for knowledge base completion*

Entity representation : Vector
(initialized with pre-trained word vector)
Relation representation : Tensor (Matrix & vector)
Entity-relation interaction : Bilinear + Linear
Scoring function: margin-based ranking loss

- An entity is represented by 1 vector
- A relationship is represented by a tensor:
  - k matrices and k vectors (in the example k == 2)
- Non-linearity (tanh) is also applied
- Score function is defined as:

$$u_R^T f\left(e_1^T W_R^{[1:k]} e_2 + V_R \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} + b_R\right)$$

- Very powerful model with high capacity for each relation



TransE

TransE

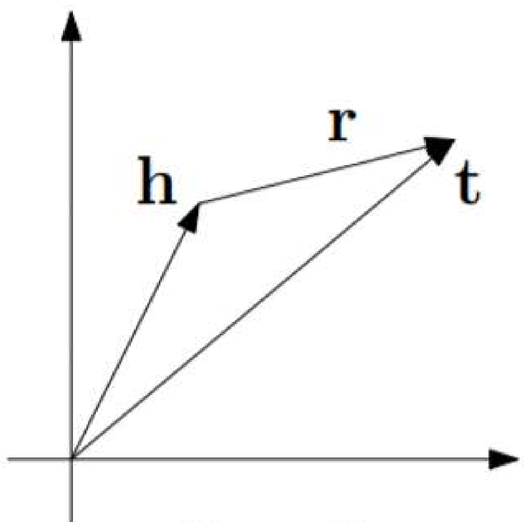
Bordes, et al. ***NIPS***, *Translating* embeddings for modeling multi-relational data

2013

Entity representation : Vector
Relation representation : Vector
Entity-relation interaction : Linear
Scoring function: margin-based ranking loss

- Modeling relations as translations
- Intuition, we assume:
  Seahawks + HomeField ≈ Seattle
- The score function is defined as:

S(Seahawks, HomeField, Seattle) = $-||e_{Seahawks} + e_{HomeField} - e_{Seattle}||_2$



TransE

DISTMULT

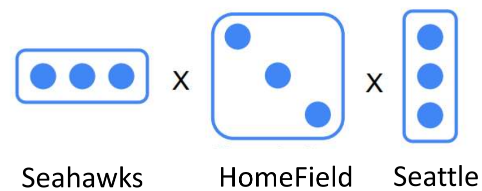
DISTMULT

Yang, et al. ***ICLR***, *Embedding entities and relations for learning and inference in knowledge bases*

2015

Entity representation : Vector
Relation representation : Vector
Entity-relation interaction : Bilinear
Scoring function: margin-based ranking loss

- An entity is represented by 1 vector
- A relationship is represented by 1 vector
- A simplification of the Bilinear method
- Very efficient and effective model
- Normally generate the best result



Scalability [# of params]

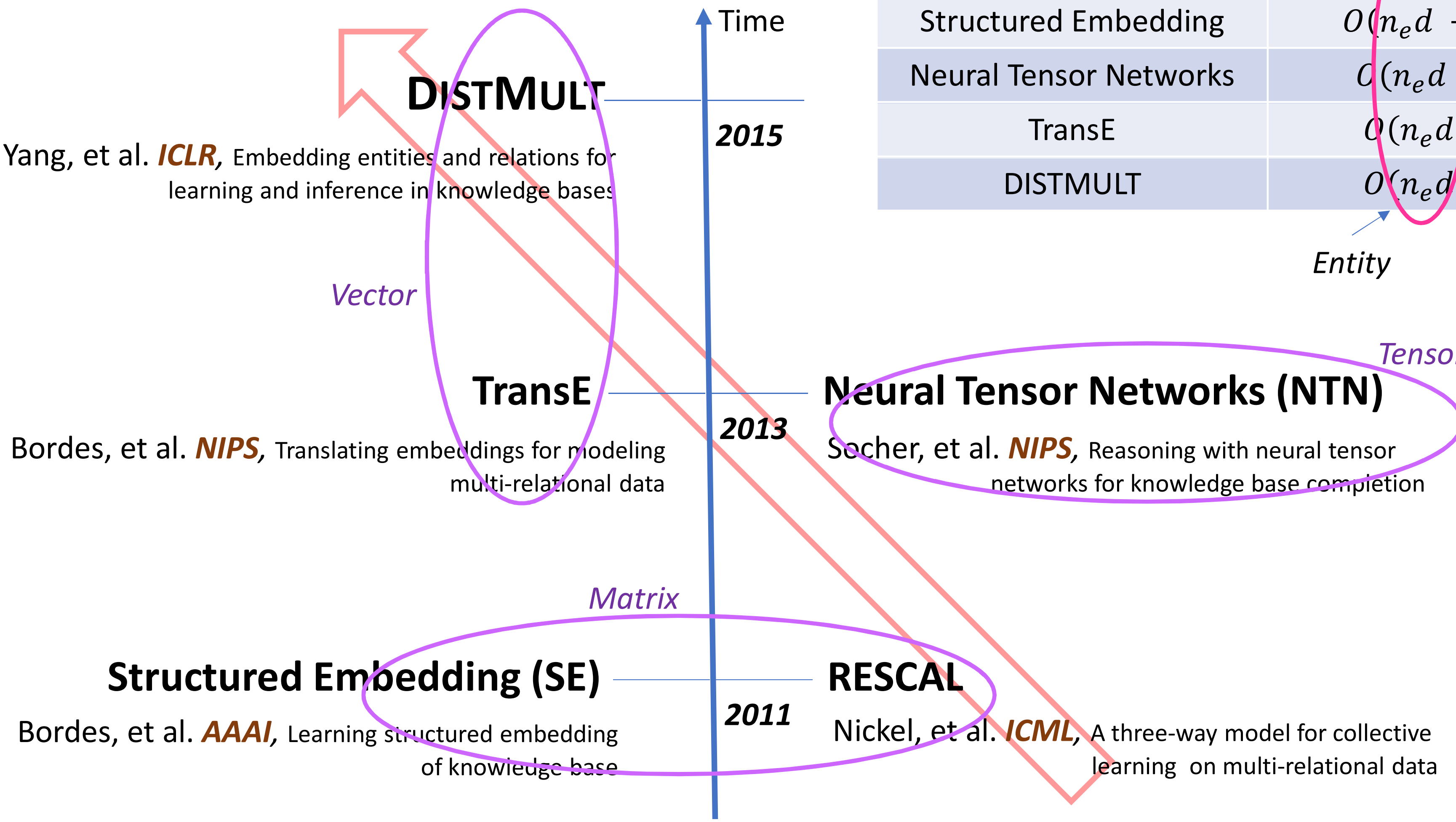
Method	# of Parameters
RESCAL	$O(n_e d + n_r d^2)$
Structured Embedding	$O(n_e d + 2 n_r d^2)$
Neural Tensor Networks	$O(n_e d + n_r d^3)$
TransE	$O(n_e d + n_r d)$
DISTMULT	$O(n_e d + n_r d)$

Matrix

Tensor

Vector

Entity *Relation*



Summary of KG embeddings

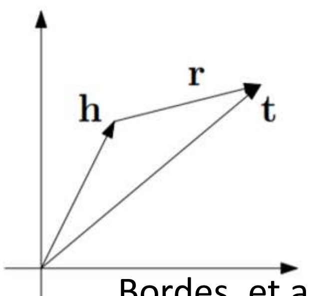
Who is the ***best*** ?

Time

DISTMULT

2015

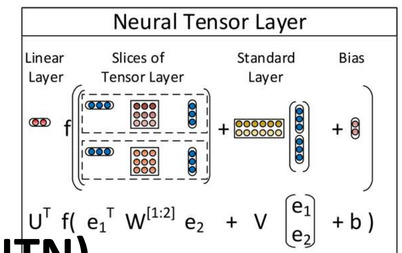
Yang, et al. ***ICLR***, Embedding entities and relations for learning and inference in knowledge bases



TransE

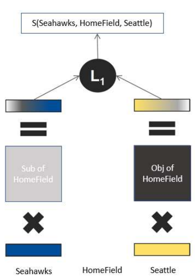
2013

Bordes, et al. ***NIPS***, Translating embeddings for modeling multi-relational data



Neural Tensor Networks (NTN)

Socher, et al. ***NIPS***, Reasoning with neural tensor networks for knowledge base completion

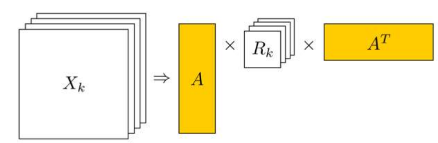


Structured Embedding (SE)

2011

Bordes, et al. ***AAAI***, Learning structured embedding of knowledge base

Tensor Factorization



RESCAL

Nickel, et al. ***ICML***, A three-way model for collective learning on multi-relational data

Knowledge Graph Inference

- WHAT – definition
- WHY – why do we need it and why it is important?

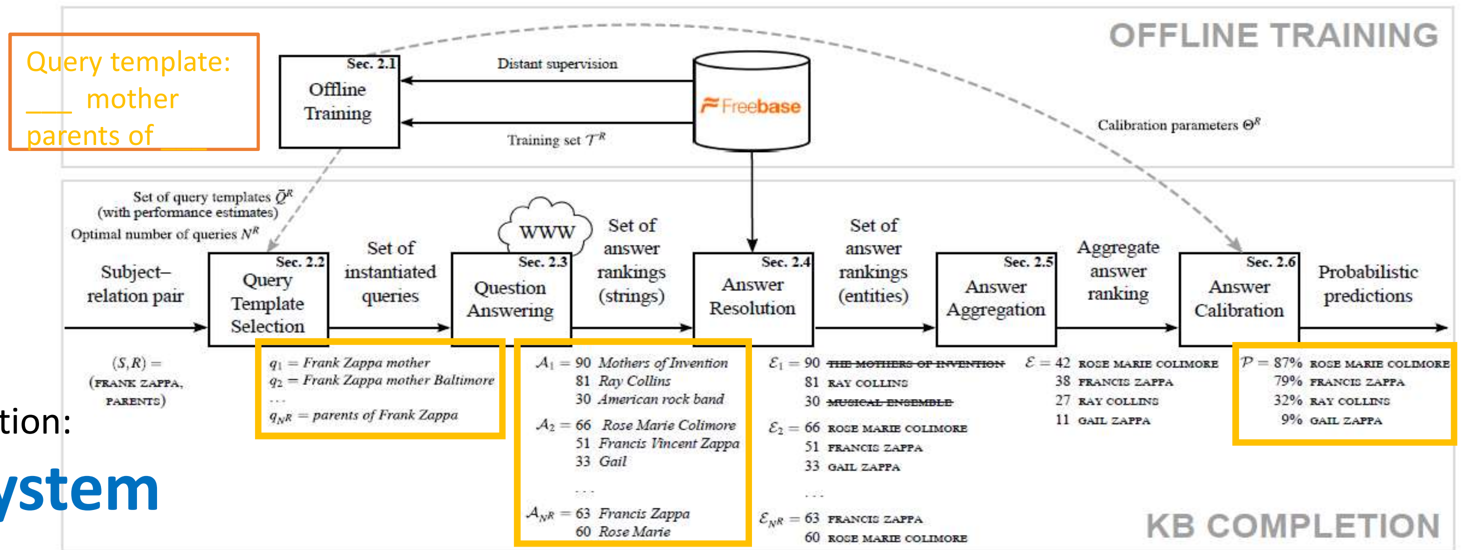
- HOW
  - Within existing KG
    - Modeling overview
    - Graph feature model -- Path Ranking Algorithm (PRA)
    - Latent feature models (KG embeddings)

  - ***Use external sources – QnA system***

Knowledge Graph Inference

– Using Q&A Method

- Input: subject-relation pairs (FRANK ZAPPA, PAERENTS)
- Output: answer entity (ROSE MARIE COLMORE, …)



Leverage ***external*** Information:

An in-house QA system

West, et al., WWW'14, Knowledge Base Completion via Search-Based Question Answering

Knowledge Graph Inference - Summary

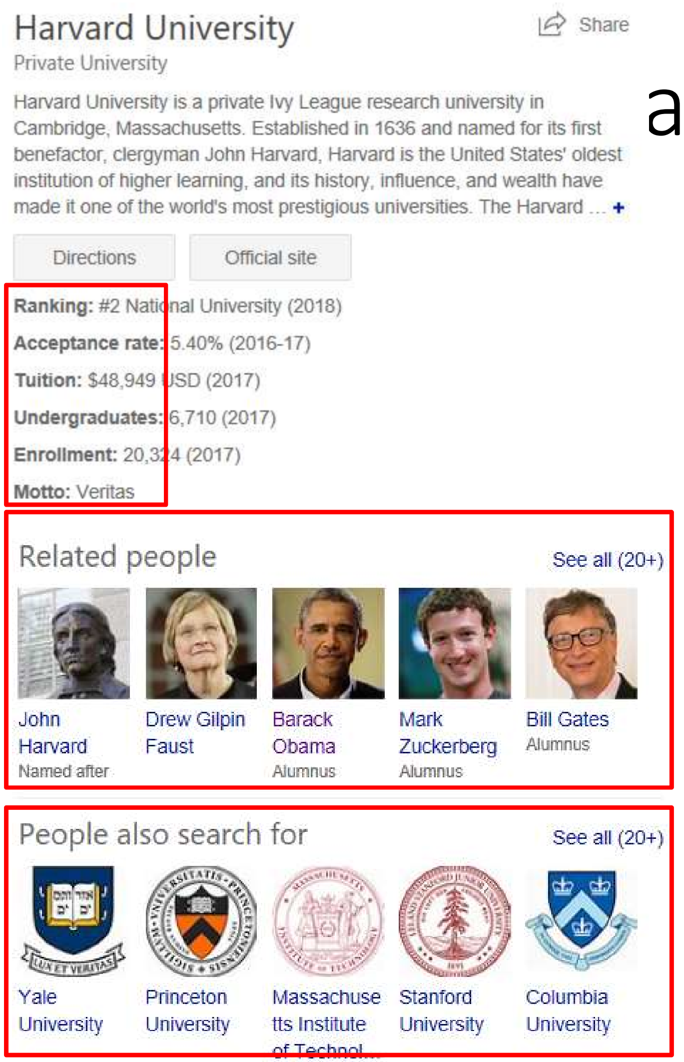
- WHAT – "inference" from existing graph
  - With graph itself – reasoning / link prediction
  - With ***external*** source
- WHY – knowledge graph is largely ***incomplete***
- HOW
  - Within existing KG
    - Modeling overview
    - Graph feature model -- Path Ranking Algorithm (PRA)
    - Latent feature models (KG embeddings)
  - Use external sources – QnA system

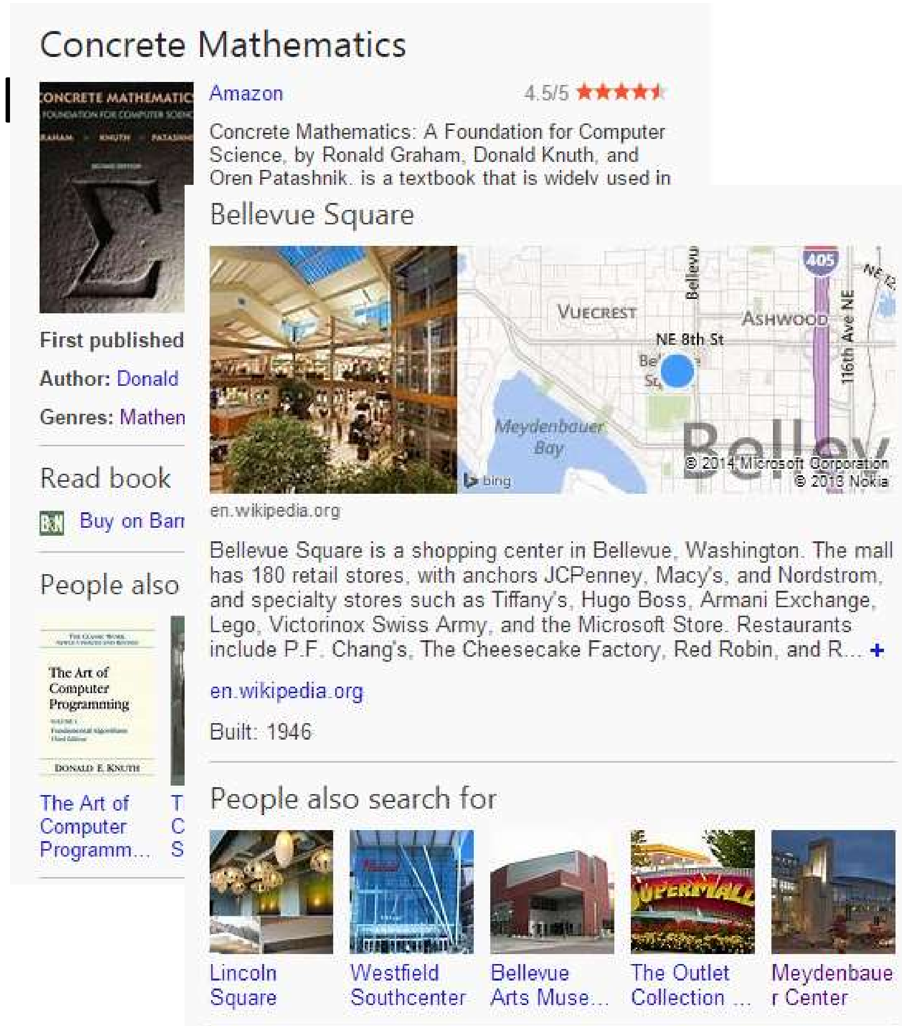
- General assumptions
- Key components in model
- Milestone models
  - RESCAL and tensor factorization
  - Structured Embedding
  - Neural Tensor Networks (NTN)
  - TransE and DISTMULT

Knowledge Graph Applications

- ***Entity Recommendation***
  - Co-occurrence based
  - Similarity based
    - Textual
    - Embedding
- Question and Answering
  - KG Based
  - Web Based with KB enrichment

En atio


Harvard University
Share
Private University
Harvard University is a private Ivy League research university in Cambridge, Massachusetts. Established in 1636 and named for its first benefactor, clergyman John Harvard, Harvard is the United States' oldest institution of higher learning, and its history, influence, and wealth have made it one of the world's most prestigious universities. The Harvard ... +
Directions
Official site
Ranking: #2 National University (2018)
Acceptance rate: 5.40% (2016-17)
Tuition: $48,949 USD (2017)
Undergraduates: 6,710 (2017)
Enrollment: 20,324 (2017)
Motto: Veritas
Related people
See all (20+)
John Harvard
Named after
Drew Gilpin Faust
Barack Obama
Alumnus
Mark Zuckerberg
Alumnus
Bill Gates
Alumnus
People also search for
See all (20+)
Yale University
Princeton University
Massachusetts Institute of Technol...
Stanford University
Columbia University


Concrete Mathematics
Amazon
4.5/5
Concrete Mathematics: A Foundation for Computer Science, by Ronald Graham, Donald Knuth, and Oren Patashnik, is a textbook that is widely used in
First published
Author: Donald
Genres: Mathem
Read book
Buy on Barn
People also
The Art of Computer Programm...
Bellevue Square
Vuecrest
Ashwood
NE 8th St
116th Ave NE
405
Meydenbauer Bay
bing
© 2014 Microsoft Corporation
© 2013 Nokia
en.wikipedia.org
Bellevue Square is a shopping center in Bellevue, Washington. The mall has 180 retail stores, with anchors JCPenney, Macy's, and Nordstrom, and specialty stores such as Tiffany's, Hugo Boss, Armani Exchange, Lego, Victorinox Swiss Army, and the Microsoft Store. Restaurants include P.F. Chang's, The Cheesecake Factory, Red Robin, and R... +
en.wikipedia.org
Built: 1946
People also search for
Lincoln Square
Westfield Southcenter
Bellevue Arts Muse...
The Outlet Collection ...
Meydenbauer Center

$P(entity|entity)$

- $P(Florence|Italy) = \frac{Freq(Florence, Italy)}{Freq(Italy)}$ Co-occurrence

- $P(Florence|Italy) = \frac{Sim(Florence,Italy)}{\sum Sim(*, Italy)}$ Cosine Similarity

Deshpande, et al., ACM TOIS 2004, Item-Based Top-N Recommendation Algorithms

Knowledge Graph Applications

- ***Entity Recommendation***
    - ***Co-occurrence based***
    - Similarity based
        - Textual
        - Embedding
- Question and Answering
    - KG Based
    - Web Based with KB enrichment

$P(entity|entity)$ – Co-occurrence

- Sources
  - Search user behavior
    - Within Queries
    - Across Queries
    - User Url Clicks
  - Wikipedia
    - Wikipedia Pages
    - Wikipedia Categories/Templates
    - Wikipedia Revision Histories
  - Web documents

The Wisdom of Crowds

Image credit: https://commons.wikimedia.org/

$P(entity|entity)$ – Co-occurrence

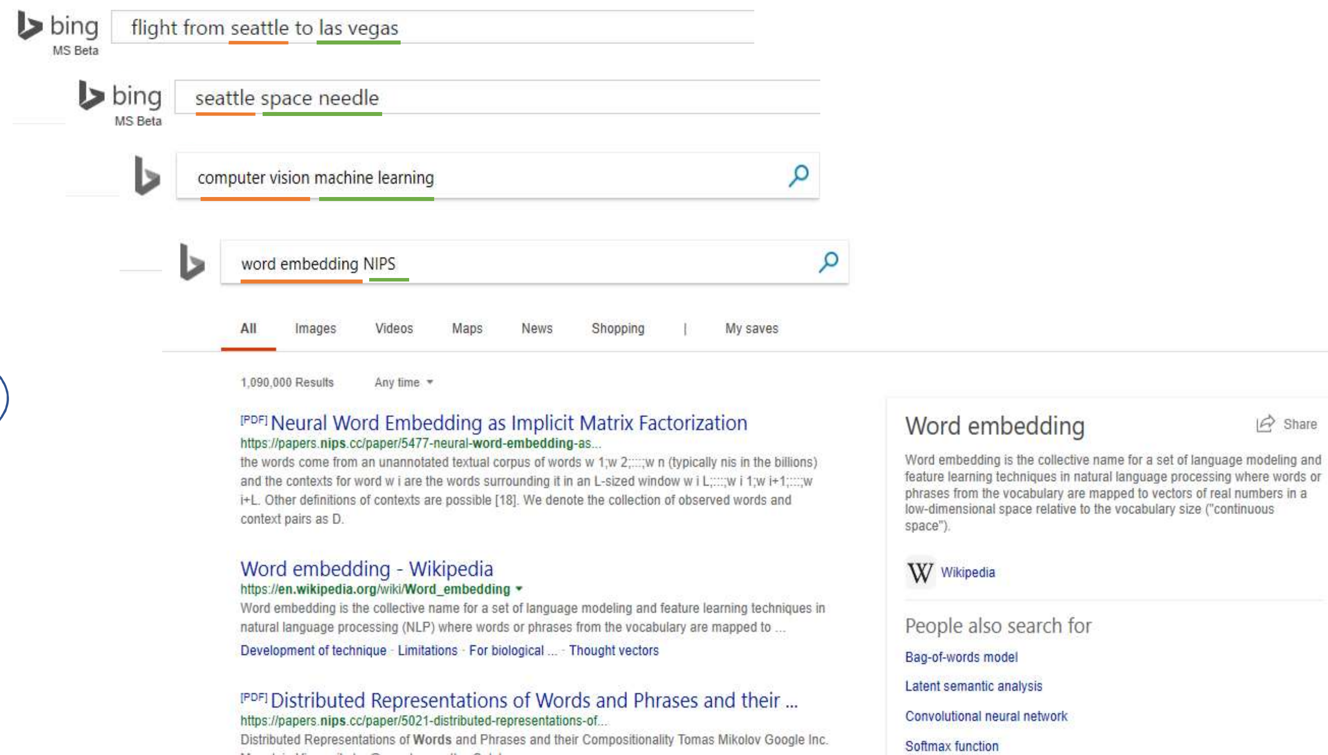
- Sources
  - ***Search user behavior***
    - Within Queries
    - Across Queries
    - User Url Clicks
  - Wikipedia
  - Web documents

Image credit: https://commons.wikimedia.org/

$P(entity|entity)$ – Co-occurrence

- Within Queries

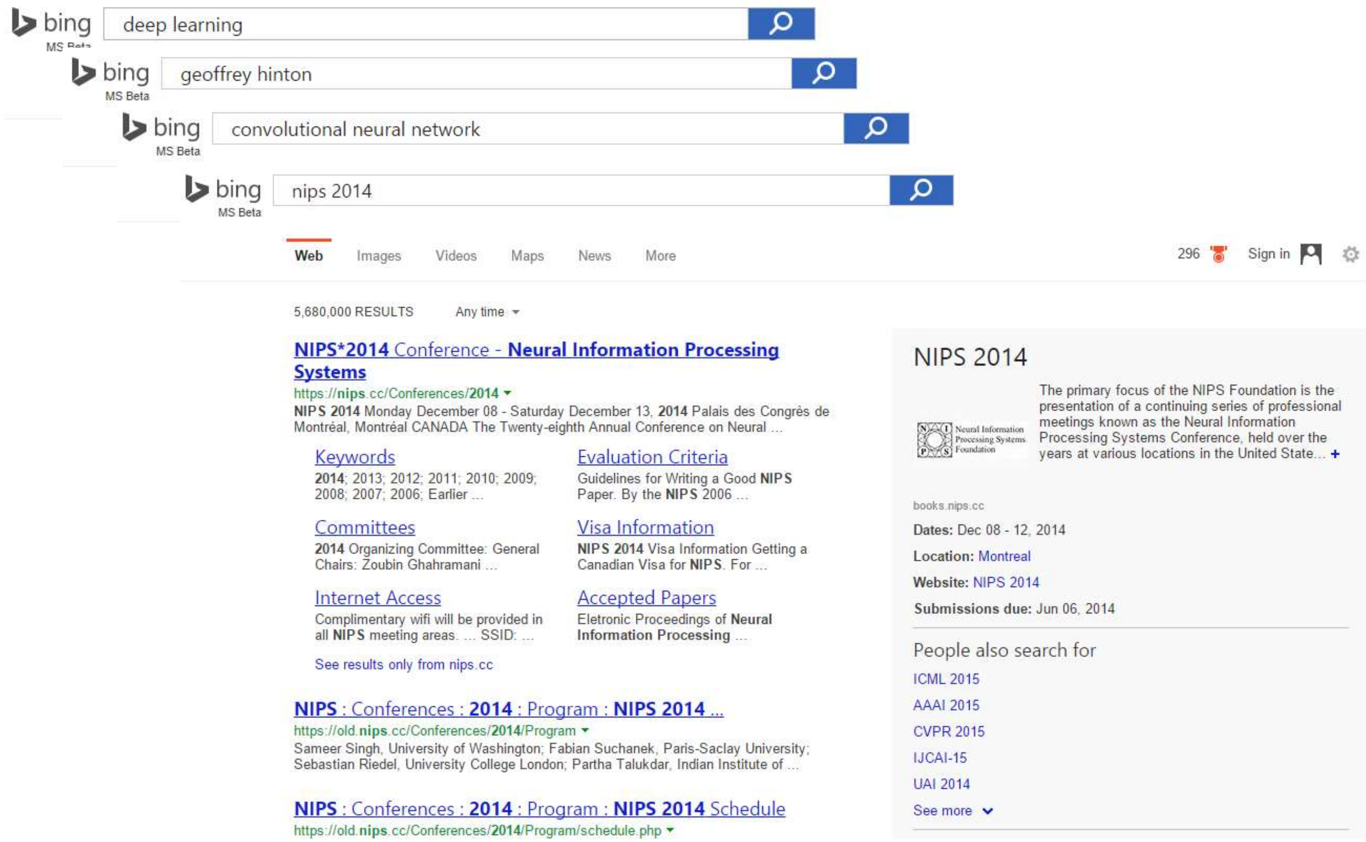
How to extract those entities?



Blanco, et al., ISWC'13. Entity Recommendations in Web Search

$P(entity|entity)$ – Co-occurrence

- Across Queries



Blanco, et al., ISWC''13. Entity Recommendations in Web Search

$P(entity|entity)$ – Co-occurrence

- User Url Clicks

$P(entity|entity)$ – Co-occurrence

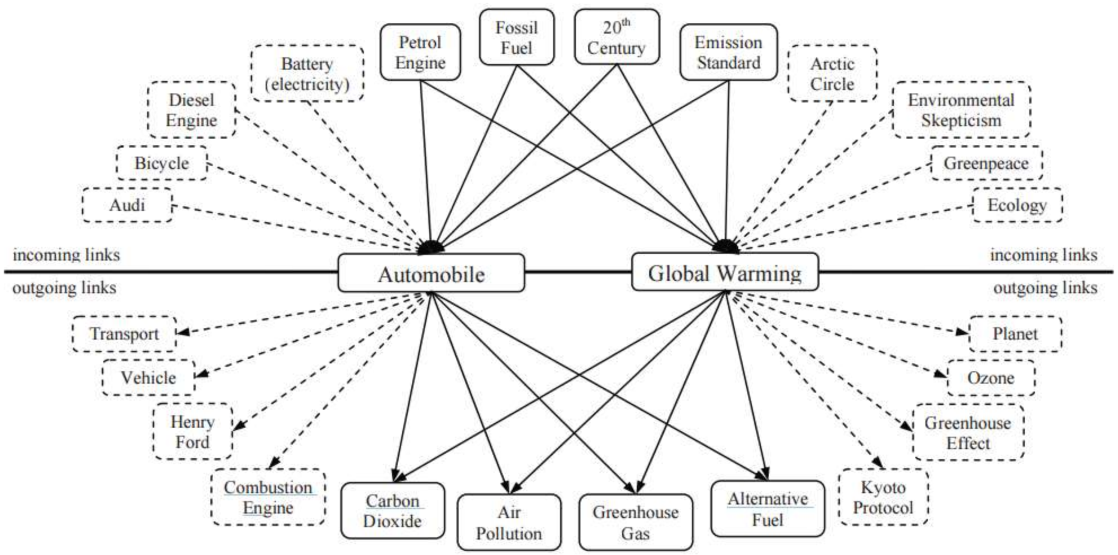
- Sources
  - Search user behavior
  - ***Wikipedia***
    - Wikipedia Pages
    - Wikipedia Categories/Templates
    - Wikipedia Revision Histories
  - Web documents

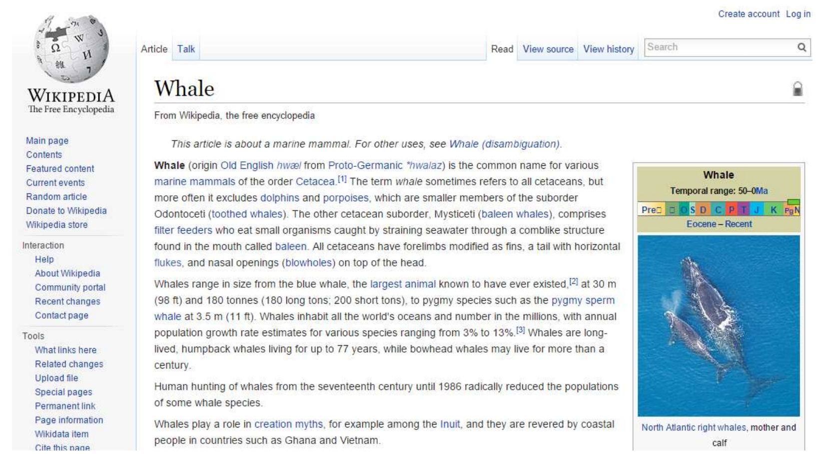
The Wisdom of Crowds

Image credit: https://commons.wikimedia.org/

$P(entity|entity)$ – Co-occurrence

- Wikipedia Pages

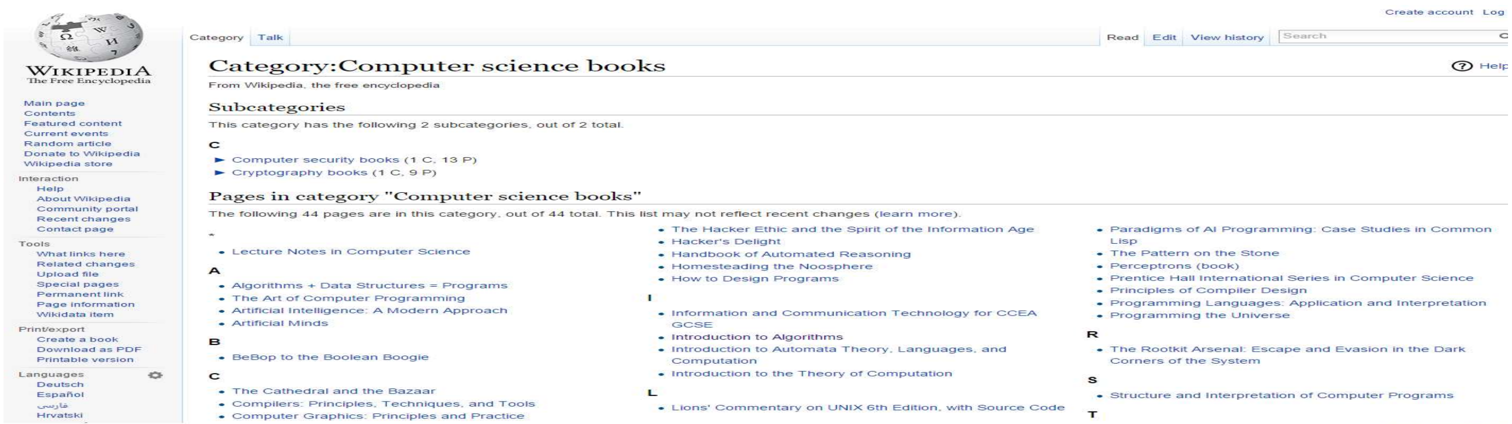


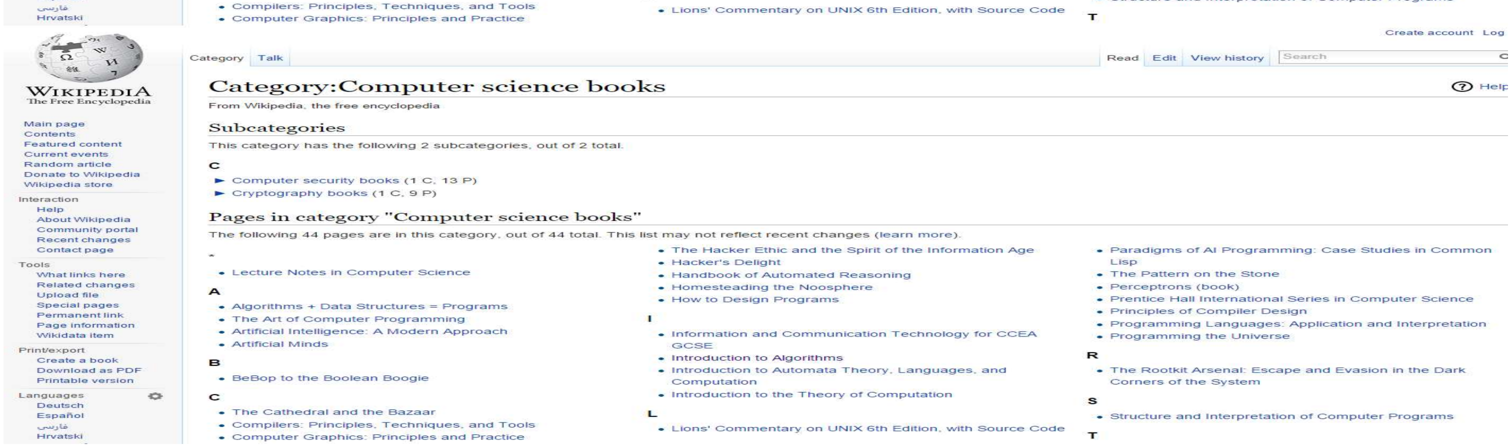


Milne, et al., AAAI'08. An Effective, Low-Cost Measure of Semantic Relatedness Obtained from Wikipedia Links

$P(entity|entity)$ – Co-occurrence

- Wikipedia Categories & Templates

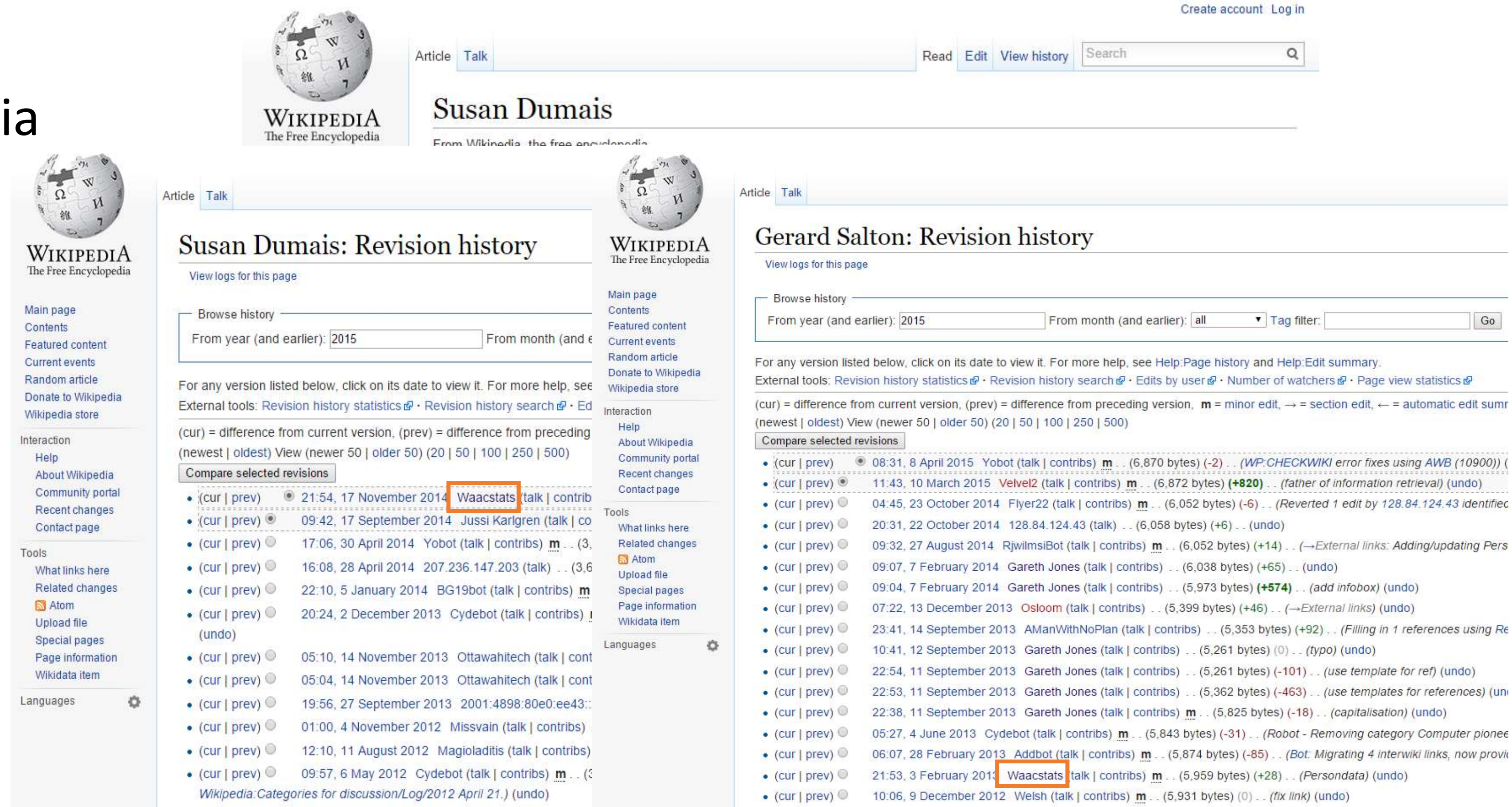




Strube, et al., AAAI'06. WikiRelate! Computing Semantic Relatedness Using Wikipedia

$P(entity|entity)$ – Co-occurrence

- Wikipedia Revision History



$P(entity|entity)$ – Co-occurrence

- Sources
  - Search user behavior
  - Wikipedia
  - ***Web documents***

Image credit: https://commons.wikimedia.org/

$P(entity|entity)$ – Co-occurrence

- Web documents

Research Publications - Data Mining Research Group at CS, UIUC

Added by Han, Jiawei, last edited by Han, Jiawei on Apr 16, 2015 (view change)

2015

1. Ahmed El-Kishky, Yanglei Song, Chi Wang, Clare R. Voss, Jiawei Han, "Scalable Topical Phrase Mining from Text Corpora", PVLDB 8(3): 305 - 316, 2015. Int. Conf. on Very Large Data Bases (VLDB'15), Kohala Coast, Hawaii, Sept. 2015)
2. Qi Li, Yaliang Li, Jing Gao, Lu Su, Bo Zhao, Murat Demirbas, Wei Fan, and Jiawei Han, "A Confidence-Aware Approach for Truth Discovery on Long-Tail Da (Also, in Proc. 2015 Int. Conf. on Very Large Data Bases (VLDB'15), Kohala Coast, Hawaii, Sept. 2015)
3. Chenguang Wang, Yangqiu Song, Dan Roth, Chi Wang, Jiawei Han, Heng Ji, and Ming Zhang, "Constrained Information-Theoretic Tripartite Graph Clusteri Semantically Similar Relations", in Proc. 2015 Int. Joint Conf. on Artificial Intelligence (IJCAI'15), Buenos Aires, Agentina, July 2015.
4. Jialu Liu, Jingbo Shang, Chi Wang, Xiang Ren, Jiawei Han, "Mining Quality Phrases from Massive Text Corpora", in Proc. of 2015 ACM SIGMOD Int. Conf. Data (SIGMOD'15), Melbourne, Australia, May 2015.
5. Fangbo Tao, Bo Zhao, Ariel Fuxman, Yang Li, Jiawei Han, "Leveraging Pattern Semantics for Constructing Entity Taxonomies in Enterprises", in Proc. of 20 World-Wide Web (WWW'15), Florence, Italy, May 2015.
6. Huan Gui, Ya Xu, Anmol Bhasin, Jiawei Han, "Network A/B Testing: From Sampling to Estimation", in Proc. of 2015 Int. Conf. on World-Wide Web (WWW'15 May 2015.
7. Jialu Liu, Chi Wang, Jing Gao, Quanquan Gu, Charu Aggarwal, Lance Kaplan, and Jiawei Han, " GIN: A Clustering Model for Capturing Dual Heterogeneity Data", in Proc. of 2015 SIAM Int. Conf. on Data Mining (SDM'15), Vancouver, Canada, Apr. 2015.
8. Mengting Wan, Yunbo Ouyang, Lance Kaplan, Jiawei Han, " Graph Regularized Meta-path Based Transductive Regression in Heterogeneous Information N 2015 SIAM Int. Conf. on Data Mining (SDM'15), Vancouver, Canada, Apr. 2015.
9. Jingjing Wang, Hongbo Deng, and Jiawei Han, "Topic Periodicity Discovery from Text Data", in Proc. SPIE Vol. 9499, Next-Generation Analyst III, Baltimor 2015.
10. Wei Feng, Chao Zhang, Wei Zhang, Jiawei Han, Jianyong Wang, Charu Aggarwal, Jianbin Huang, " STREAMCUBE: Hierarchical Spatio-temporal Hashtag Exploration over the Twitter Stream", in Proc of 2015 IEEE Int. Conf on Data Engineering (ICDE'15), Seoul, Korea, Apr. 2015.

WWW 2015

Dates: May 20 - 22, 2015

Location: Florence

Website: WWW 2015

Abstracts due: Nov 03, 2014

Submissions due: Nov 10, 2014

Notification date: Jan 17, 2015

Final version due: Mar 08, 2015

People also search for

CIKM 2015

AAAI 2015

SIGIR 2015

KDD 2015

VLDB 2015

IJCAI-15

ICDE 2015

Entity Linking

- How to extract entities from Queries and Documents?
  - Through Entity Linking!

Entity linking

In natural language processing, entity linking, named entity disambiguation, named entity recognition and disambiguation or named entity normalization is the task of determining the identity of entities mentioned in text. It is distinct from named entity recognition in that it identifies not the occurrence of names, but their reference.

en.wikipedia.org

Academic conferences: AAAI 2016 · ACL 2015 · CIKM 2015 · WWW 2015 · SIGIR 2015 · COLING 2014 · IJCAI-15 · EMNLP 2015 +

People also search for

Information extraction

Question answering

Natural language processing

Explicit semantic analysis

Automatic summarization

See more

Knowledge Graph Applications

- ***Entity Recommendation***
  - Co-occurrence based
  - ***Similarity based***
    - ***Textual***
    - Embedding
- Question and Answering
  - KG Based
  - Web Based with KB enrichment

$P(entity|entity)$ – Similarity

- TF*IDF scores based on Wikipedia Corpus

tf–idf

Share

tf–idf, short for term frequency–inverse document frequency, is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus. It is often used as a weighting factor in information retrieval and text mining.

Wikipedia

People also search for

Latent Dirichlet allocation

Vector space model

Okapi BM25

Text mining

F1 score

See more

Data from: Wikipedia

Suggest an edit

Florence

From Wikipedia, the free encyclopedia

"Firenze" and "Florentine" redirect here. For other uses, see Florence (disambiguation), Florentin (disambiguation) (disambiguation).

Florence (/ˈflɒrəns/; Italian: *Firenze* [fiˈrɛntse] (listen), alternative obsolete form: *Fiorenza*; Latin: *Florentia*) is the capital city of the Italian region of Tuscany and of the province of Florence. It is the most populous city in Tuscany, with approximately 380,000 inhabitants, expanding to over 1,520,000 in the metropolitan area.[2]

Florence is famous for its history: a centre of medieval European trade and finance and one of the wealthiest cities of the time,[3] it is considered the birthplace of the Renaissance, and has been called "the Athens of the Middle Ages".[4] A turbulent political history includes periods of rule by the powerful Medici family, and numerous religious and republican revolutions.[5] From 1865 to 1871 the city was the capital of the recently established Kingdom of Italy.

The Historic Centre of Florence attracts millions of tourists each year, and Euromonitor International ranked the city as the world's 89th most visited in 2012, with 1.8 million visitors.[6] It was declared a World Heritage Site by UNESCO in 1982. The city is noted for its culture, Renaissance art and architecture and monuments.[7] The city also contains numerous museums and art galleries, such as the Uffizi Gallery and the Palazzo Pitti, and still exerts an influence in the fields of art, culture and politics.[8] Due to Florence's artistic and architectural heritage, it has been ranked by *Forbes* as one of the most beautiful cities in the world.[9]

Rome

From Wikipedia, the free encyclopedia

This article is about the city in Italy. For the civilization of classical antiquity, see Ancient Rome. For other uses

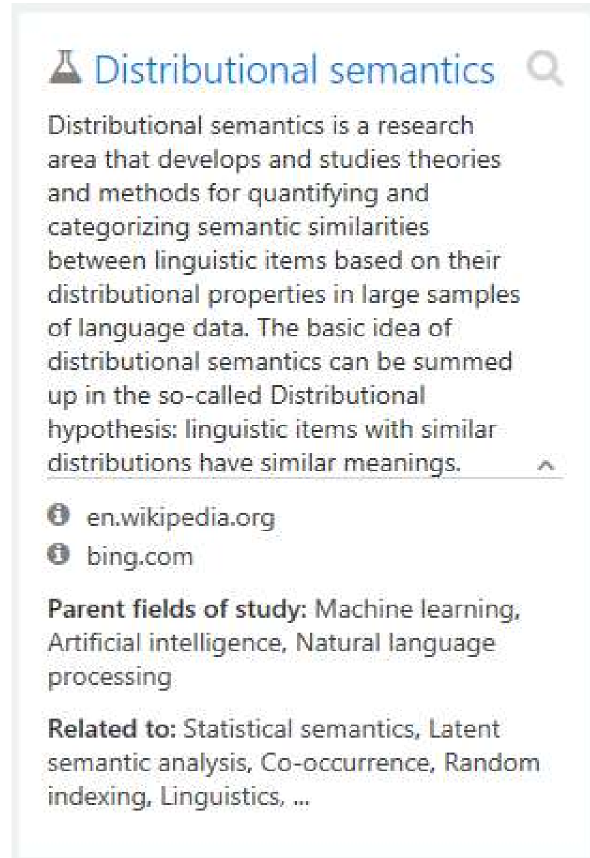
Rome (/ˈroʊm/, Italian: *Roma* [ˈroːma] (listen), Latin: *Rōma*) is a city and special *comune* (named "Roma Capitale") in Italy. Rome is the capital of Italy and region of Lazio. With 2.9 million residents in 1,285 km^2 (496.1 sq mi), it is also the country's largest and most populated *comune* and fourth-most populous city in the European Union by population within city limits. The Metropolitan City of Rome has a population of 4.3 million residents.[2] The city is located in the central-western portion of the Italian Peninsula, within Lazio (Latium), along the shores of Tiber river. Vatican City is an independent country within the city boundaries of Rome, the only existing example of a country within a city: for this reason Rome has been often defined as capital of two states.[3][4]

Rome's history spans more than two and a half thousand years. While Roman mythology dates the founding of Rome at only around 753 BC, the site has been inhabited for much longer, making it one of the oldest continuously occupied cities in Europe.[5] The city's early population originated from a mix of Latins, Etruscans and Sabines. Eventually, the city successively became the capital of the Roman Kingdom, the Roman Republic and the Roman Empire, and is regarded as one of the birthplaces of Western civilization. It is referred to as "Roma Aeterna" (The Eternal City) [6] and "Caput Mundi" (Capital of the World), two central notions in ancient Roman culture.

$P(entity|entity)$ – Similarity

- Challenges
  - Textual Similarity suffers the vocabulary mismatch problem
    - “USA” and “United States of America” are semantically equivalent, yet share no terms in common

- Solution
  - Project entities into latent space that can semantically represent the entities



Knowledge Graph Applications

- ***Entity Recommendation***
    - Co-occurrence based
    - ***Similarity based***
        - Textual
        - ***Embedding***
- Question and Answering
    - KG Based
    - Web Based with KB enrichment

$P(entity|entity)$ – Word Embedding

- Word Embedding

$$\mathbf{s}(t) = f\left(\mathbf{U}\mathbf{w}(t) + \mathbf{W}\mathbf{s}(t{-}1)\right)$$

$$\mathbf{y}(t) = g\left(\mathbf{V}\mathbf{s}(t)\right)$$

$$f(z) = \frac{1}{1+e^{-z}}, \quad g(z_m) = \frac{e^{z_m}}{\sum_k e^{z_k}}$$

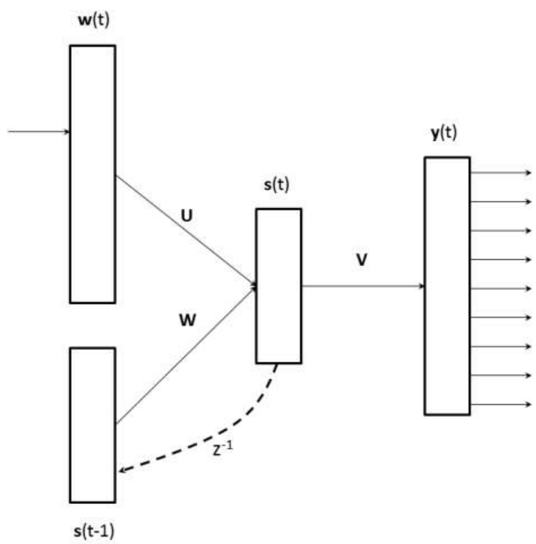
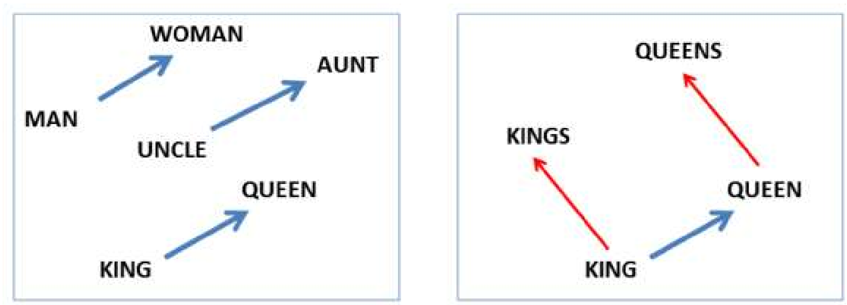


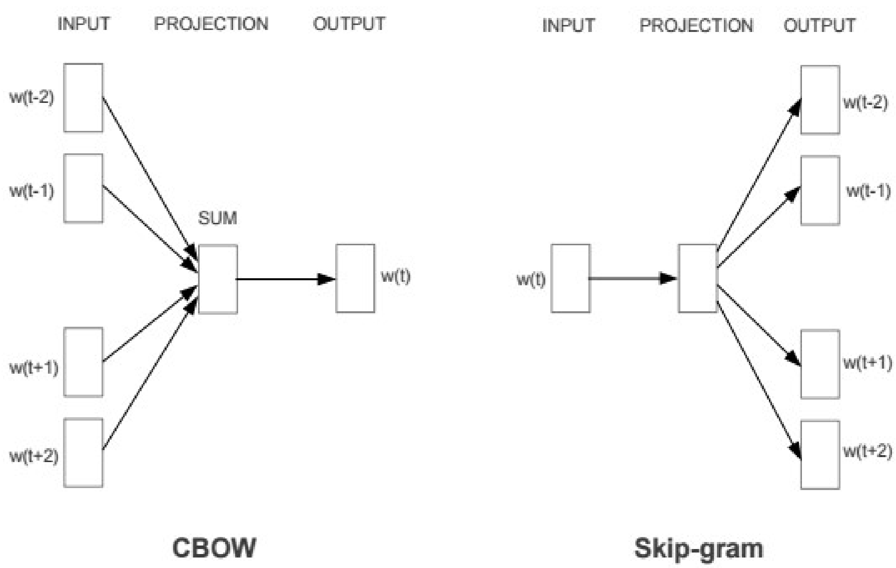
Figure 1: Recurrent Neural Network Language Model.

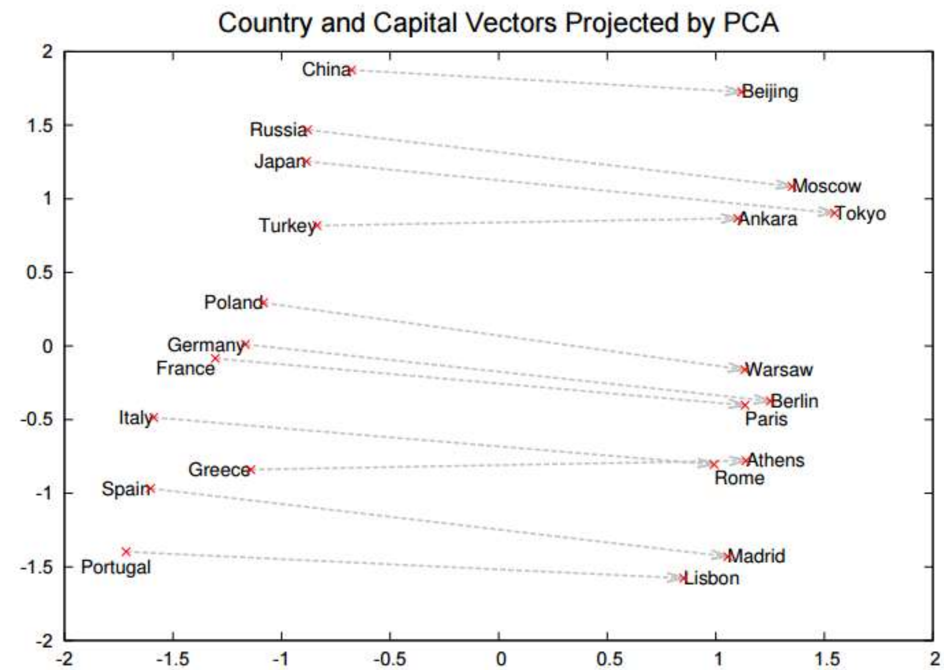


Mikolov, et al., ACL'13. Linguistic Regularities in Continuous Space Word Representations

$P(entity|entity)$ – Word Embedding

- Word Embedding





Mikolov, et al., ICLR'13. Efficient Estimation of Word Representations in Vector Space

$P(entity|entity)$ – Word Embedding

Relationship	Example 1	Example 2	Example 3
France - Paris	Italy: Rome	Japan: Tokyo	Florida: Tallahassee
big - bigger	small: larger	cold: colder	quick: quicker
Miami - Florida	Baltimore: Maryland	Dallas: Texas	Kona: Hawaii
Einstein - scientist	Messi: midfielder	Mozart: violinist	Picasso: painter
Sarkozy - France	Berlusconi: Italy	Merkel: Germany	Koizumi: Japan
copper - Cu	zinc: Zn	gold: Au	uranium: plutonium
Berlusconi - Silvio	Sarkozy: Nicolas	Putin: Medvedev	Obama: Barack
Microsoft - Windows	Google: Android	IBM: Linux	Apple: iPhone
Microsoft - Ballmer	Google: Yahoo	IBM: McNealy	Apple: Jobs
Japan - sushi	Germany: bratwurst	France: tapas	USA: pizza

$$y = x_b - x_a + x_c$$

$$w^* = argmax_w \frac{x_w y}{\|x_w\| \|y\|}$$

Mikolov, et al., NIPS'13. Distributed Representations of Words and Phrases and their Compositionality

$P(entity|entity)$ – Word Embedding

- How to apply Word Embedding in Entities?
  - Perform entity linking on documents
  - Treat each entity as a single word
  - Learning the representation using Word2Vec or Glove

Knowledge Graph Applications

- ***Entity Recommendation - Summary***
    - Co-occurrence based
    - Similarity based
        - Textual
        - Embedding

- Question and Answering
    - KG Based
    - Web Based with KB enrichment

$Entity\ Recommendation - \ P(entity|entity)$

- $P(Florence|Italy) = \frac{Freq(Florence, Italy)}{Freq(Italy)}$

Co-occurrence

Sources

- Search user behavior
- Wikipedia
- Web documents

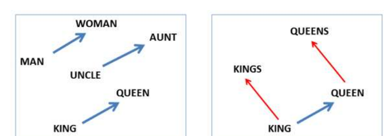
The Wisdom of Crowds

- $P(Florence|Italy) = \frac{Sim(Florence, Italy)}{\sum Sim(*, Italy)}$

Cosine Similarity

Textual (TF-IDF)

Embedding



$$P(entity|entity)$$

- Co-occurrence and Textual Similarity methods work well
- Textual Similarity method is very topic- or genre-related
- Word Embedding might not always work (depend on training data)

Co-occurrence

Whale	Susan Dumais	Tom Cruise
Dolphin	C. J. van Rijsbergen	Nicole Kidman
Pinniped	W. Bruce Croft	Brad Pitt
Shark	Eric Horvitz	Steven Spielberg
Killer whale	George Furnas	Tom Hanks
Humpback whale	Thomas Landauer	John Travolta

Textual Similarity

Whale	Susan Dumais	Tom Cruise
Beluga whale	C. J. van Rijsbergen	Leonardo DiCaprio
Toothed whale	W. Bruce Croft	Nicole Kidman
Killer whale	Harry Shum	Clint Eastwood
Pygmy killer whale	Gerard Salton	Mark Rathbun
Humpback whale	Jaime Teevan	L. Ron Hubbard

Word Embedding

Whale	Susan Dumais	Tom Cruise
Penguin	Gary William Flake	Katie Holmes
Humpback whale	Andrei Broder	Suri Cruise
Killer whale	Bill Buxton	Nicole Kidman
Turtle	Harry Bruce	Cameron Diaz
Rat	Richard Rashid	Connor Antony

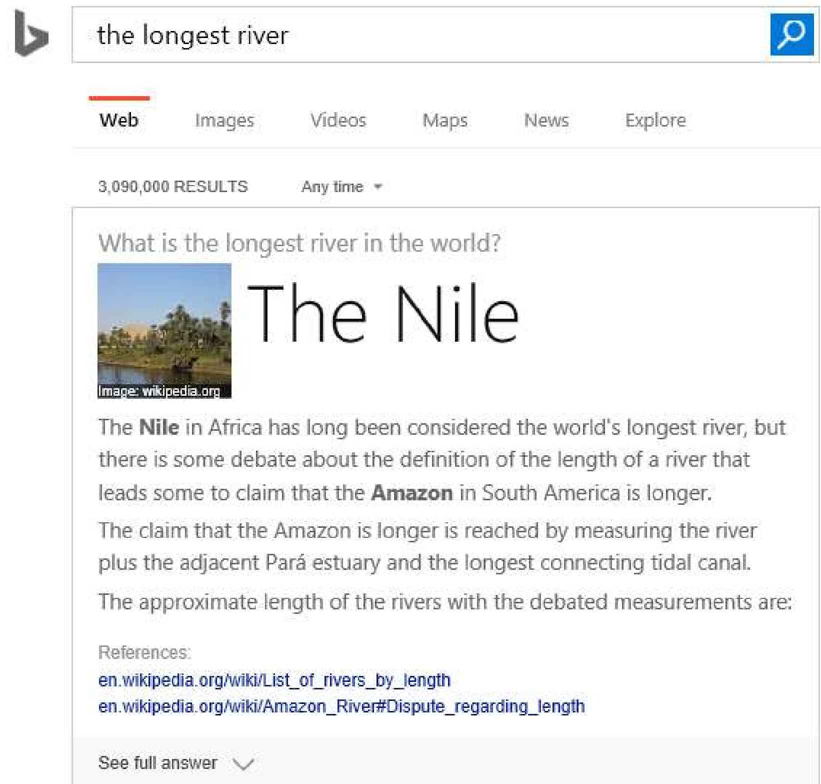
Knowledge Graph Applications

- Entity Recommendation
  - Co-occurrence
  - Similarity
    - Textual
    - Embedding

- ***Question and Answering***
  - ***What is it?***
  - KG Based
  - Web Based with KB enrichment

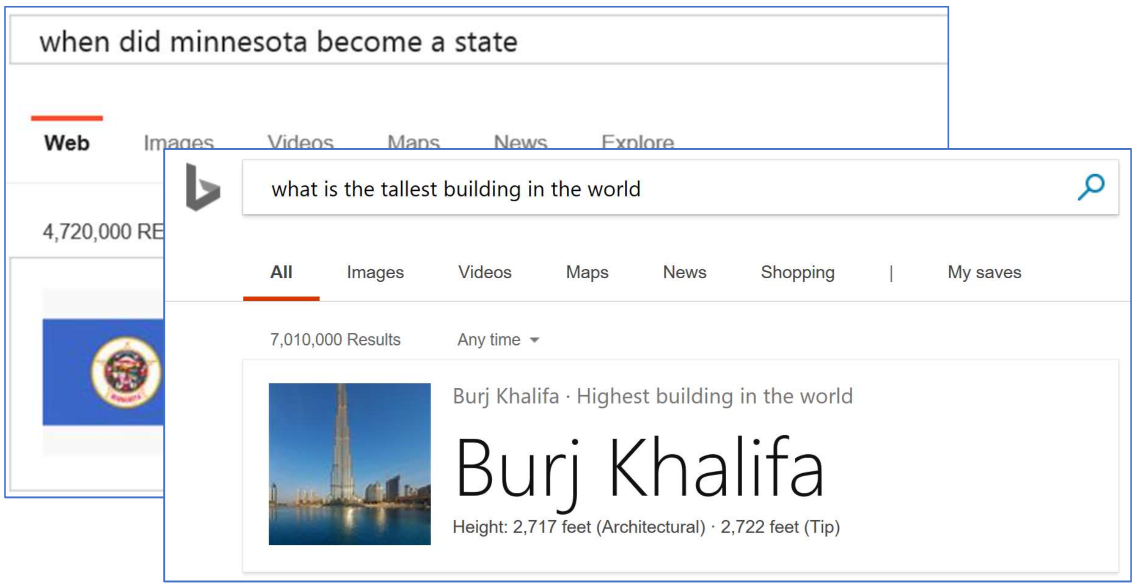
Classical NLP Problems

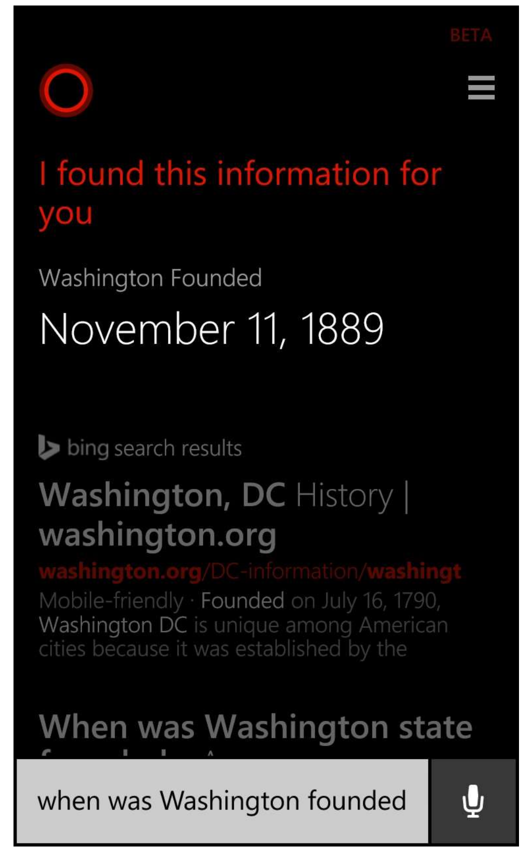
- Question Answering



Classical NLP Problems

- Question Answering





Knowledge Graph Applications

- Entity Recommendation
- ***Question and Answering***
  - What is it?
  - ***KG Based***
  - Web Based with KB enrichment

Question and Answering (KG Based)

- General problem setting
  - Information Source: A “Knowledge Graph”
  - Input: A natural language question (instead of a formal “query”)
  - Output: Answer

KB-QA Datasets

Sample questions from WebQuestions

- What character did Natalie Portman play in Star Wars? ⇒ ***Padme Amidala***
- What currency do you use in Costa Rica? ⇒ ***Costa Rican colon***
- What did Obama study in school? ⇒ ***political science***
- What do Michelle Obama do for a living? ⇒ ***writer, lawyer***
- What killed Sammy Davis Jr? ⇒ ***throat cancer***

- ***WebQuestions Dataset*** [Berant+ 13]
  - 5,810 questions crawled from Google Suggest API and answered using Amazon MTurk
  - 3,778 training, 2,032 testing
  - A question may have multiple answers → using Avg. F1 (~accuracy)
- ***Free917*** [Cai & Yates, ACL-13]
  - 917 English questions labeled with lambda expressions with predicates & constants defined in Freebase
- ***Simple Questions*** [Bordes et al., arXiv:1506.02075]
  - 108,442 questions paired with Freebase triples
  - Multi-argument relations (CVT) don't seem to be included
- ***WebQuestionsSP*** [Yih et al., ACL-16]
  - Full semantic parses of WebQuestions in SPARQL, along with updated answers and additional entity/relation information

Key Steps and Challenges

KB-based QA main steps:

- Question understanding
  - Question analysis
  - Phrase mapping
  - Disambiguation
- Query construction

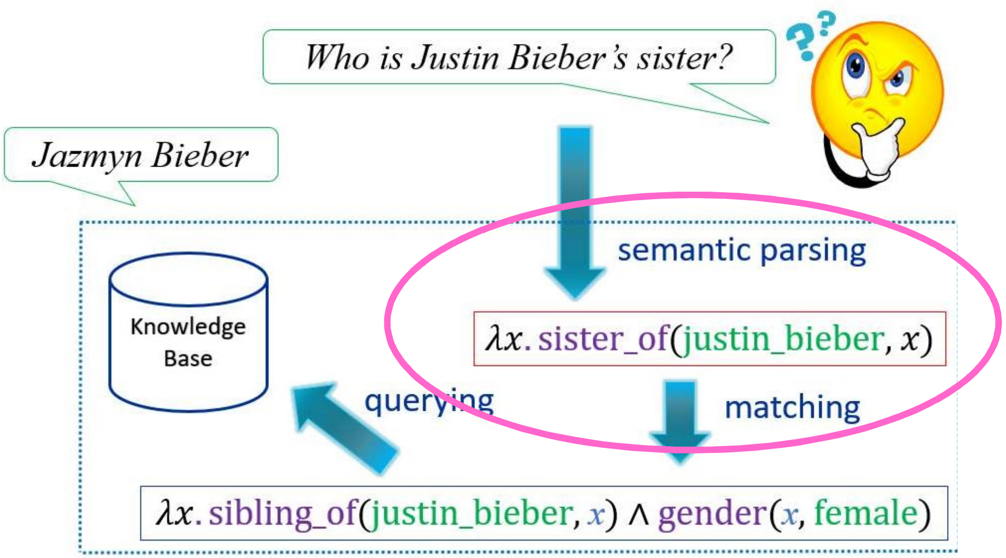
Challenges:

- **Language mismatch**
  - Lots of ways to ask the same question
    - “Who played the role of Meg on Family Guy?”
    - “What is the name of the actress for Meg on Family Guy?”
    - “In the TV show Family Guy, who is the voice for Meg?”
  - Need to map questions to the predicates defined in KB
    - tv.tv_program.regular_cast – tv.regular_tv_appearance.actor
- **Large search space**
  - Some Freebase entities have >160,000 immediate neighbors
- **Compositionality**
  - “What movies are directed by the person who won the most Academy and Golden Globe awards combined?”

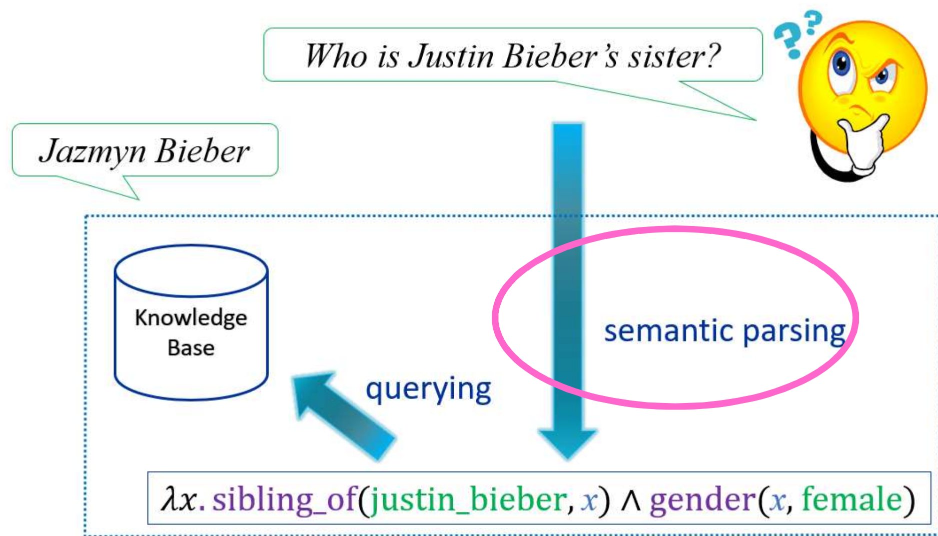
Approaches

- Semantic Parsing
  - Generic semantic parsing and then ontology matching
  - KB-specific semantic parsing
- Embedding

Generic Semantic Parsing (e.g., [Kwiatkowski+ 13])



KB-Specific Semantic Parsing (e.g., [Berant+ 13])



SEMPRE – Bridging
[Berant et al., EMNLP-2013]

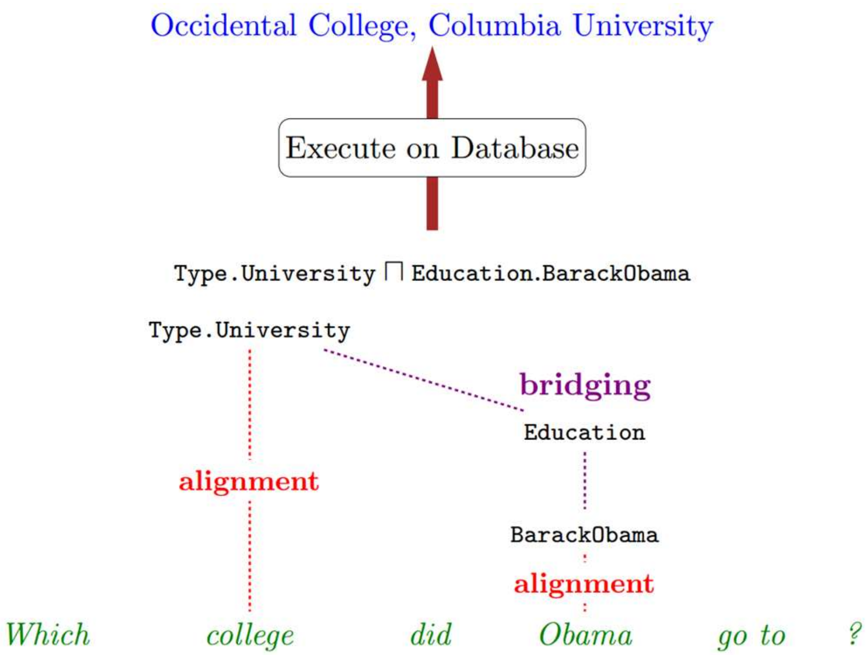
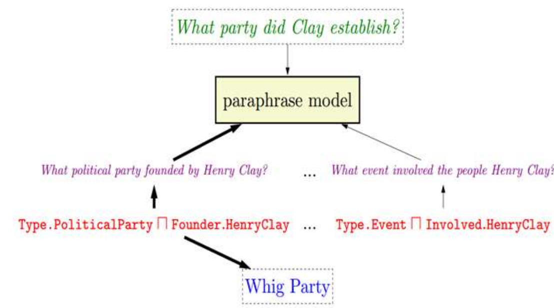
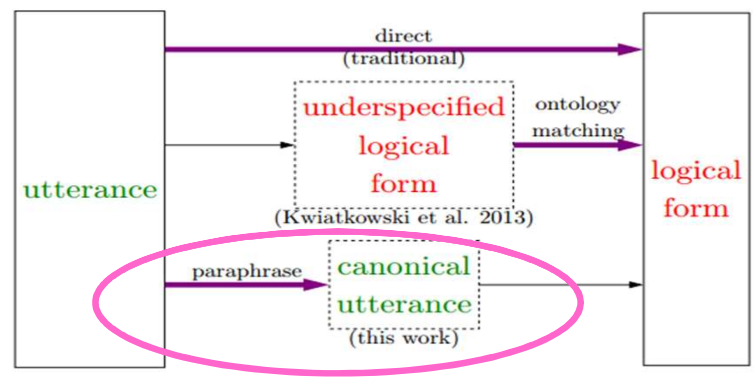


Fig.1 of [Berant et al., 2013]

<u>Bridging</u>: Hypothesizing predicates to be connected when the type constraints are satisfied

SEMPRE – Paraphrasing
[Berant & Liang, ACL-14]





<u>Paraphrasing</u>: to handle the mismatch btw language and the KB

https://nlp.stanford.edu/software/sempre/

Embeddings [Bordes et al., EMNLP-2014]

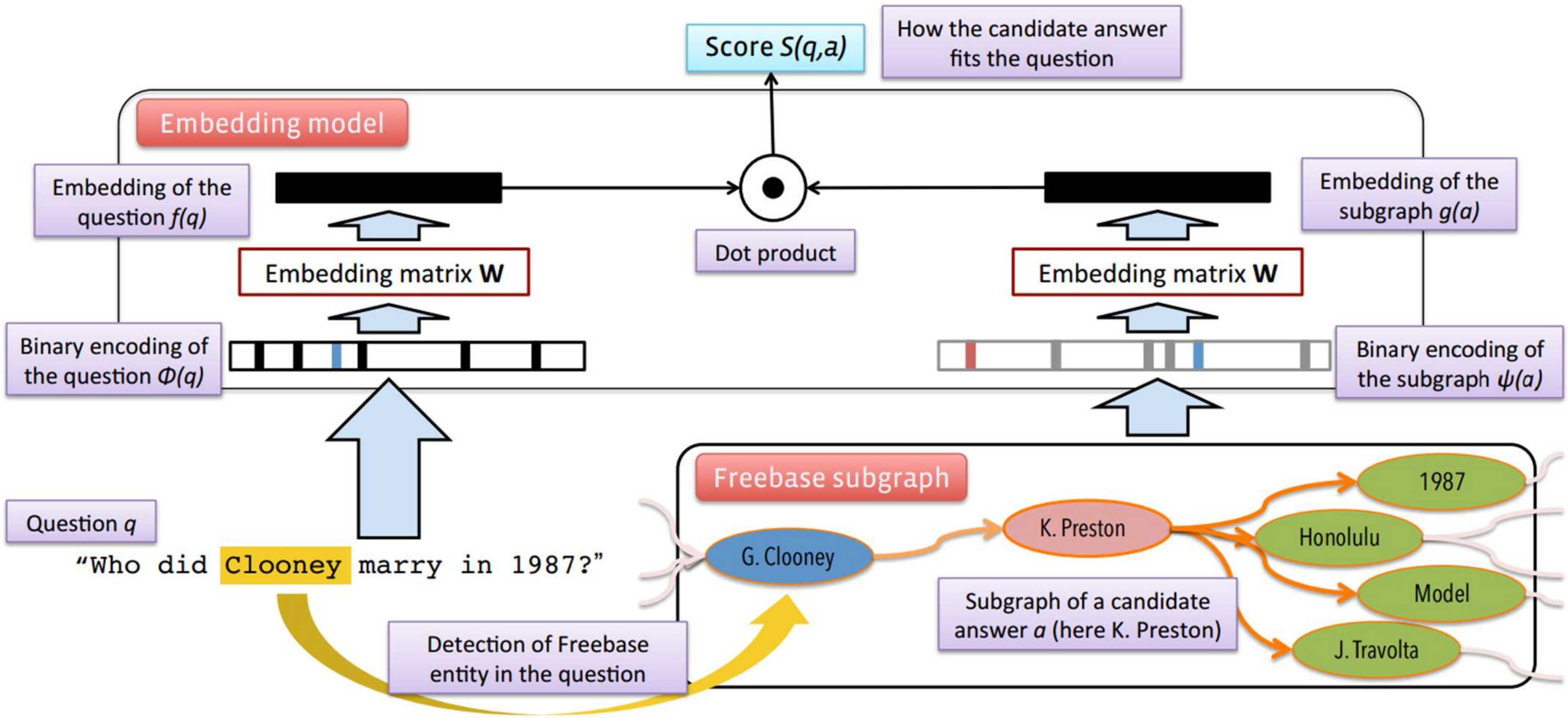


Fig.1 of [Bordes et al., 2014]

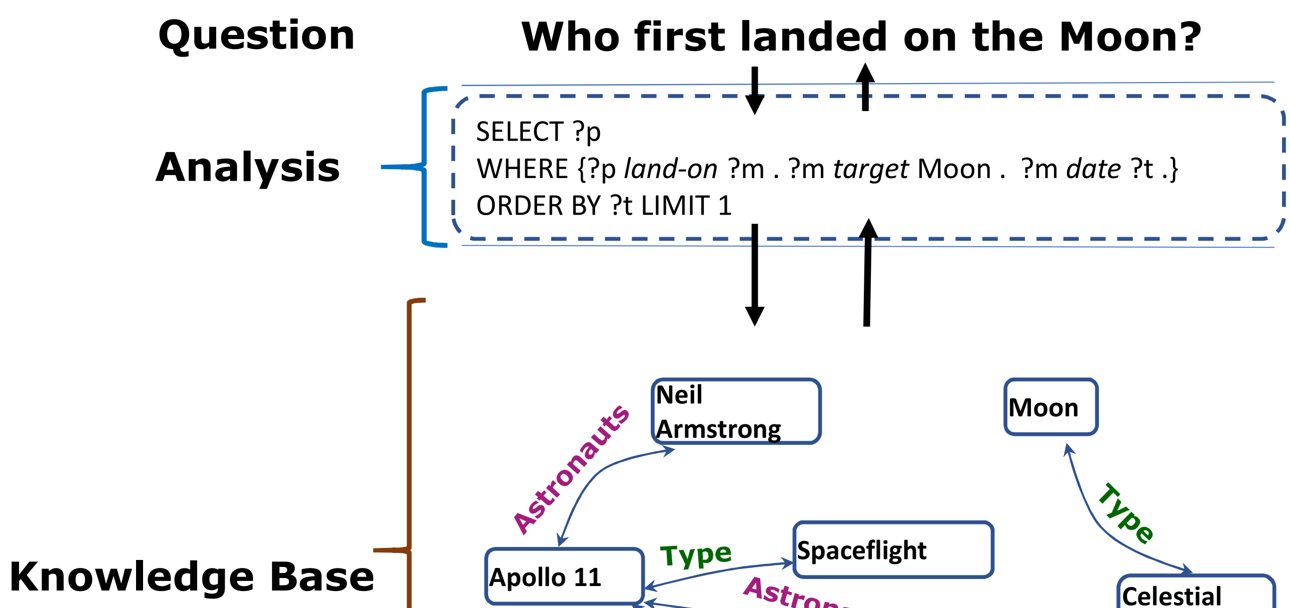
Summary

- Recent work on question answering with KB
  - Task: Answering WebQuestions using Freebase
  - Most approaches aim for semantic parsing of questions
- Challenges
  - How to leverage multiple resources to handle language mismatch?
  - How to handle compositionality correctly and efficiently?
- Very active research problem
  - Many new methods being proposed (e.g., [Berant & Liang, TACL-15], [Reddy et al., TACL-16], [Xu et al., ACL-16])

Knowledge Graph Applications

- Entity Recommendation
- ***Question and Answering***
  - What is it?
  - KG Based
  - ***Web Based with KB enrichment***

Issues with QA with KG



Issues:
- Semantic parsing is difficult due to ontology mismatch
- Knowledge base is incomplete (missing entities/relations/attribute values)

Knowledge Base is largely incomplete

Freebase

Relation	Percentage unknown	
	All 3M	*Top 100K*
PROFESSION	68%	24%
PLACE OF BIRTH	71%	13%
NATIONALITY	75%	21%
EDUCATION	91%	63%
SPOUSES	92%	68%
PARENTS	94%	77%
CHILDREN	94%	80%
SIBLINGS	96%	83%
ETHNICITY	99%	86%

West, et al., WWW'14. Knowledge Base Completion via Search-Based Question Answering

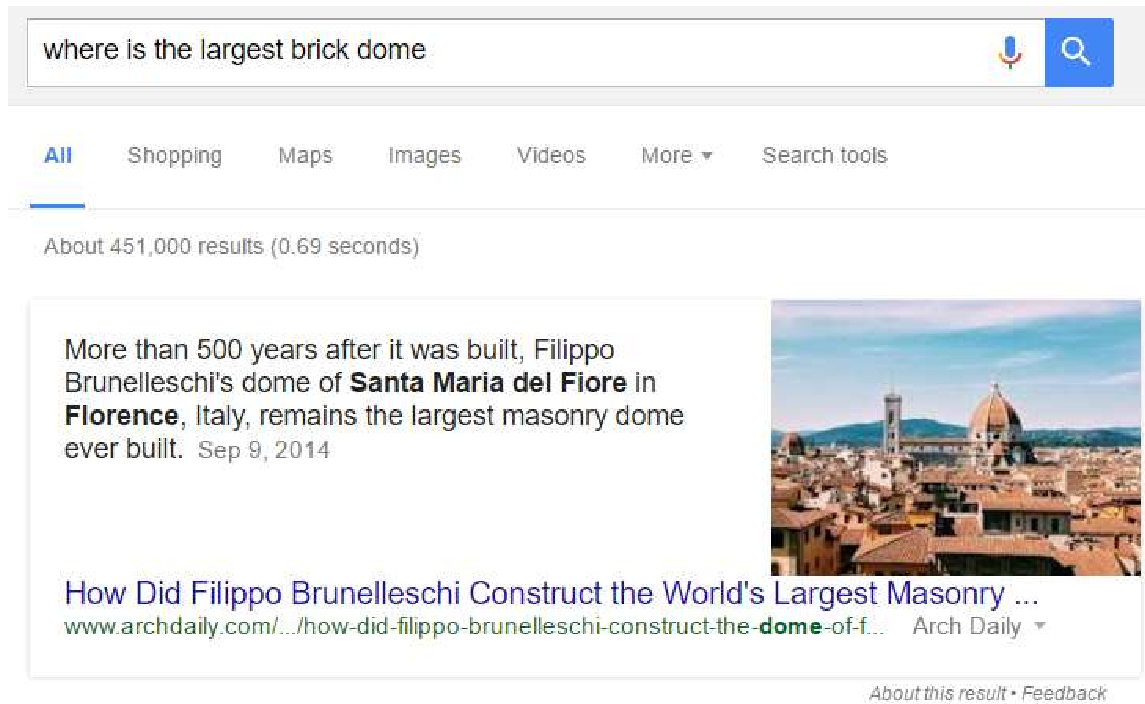
Knowledge Base is largely incomplete

Q: Where is the largest brick dome?

<u>Answer</u>

Florence Cathedral

The Cattedrale di Santa Maria del Fiore is the main church of Florence, Italy. Il Duomo di Firenze, as it is ordinarily called, was begun in 1296 in the Gothic style ... It remains the largest brick dome ever constructed.
en.wikipedia.org



Knowledge Bases

Issues:

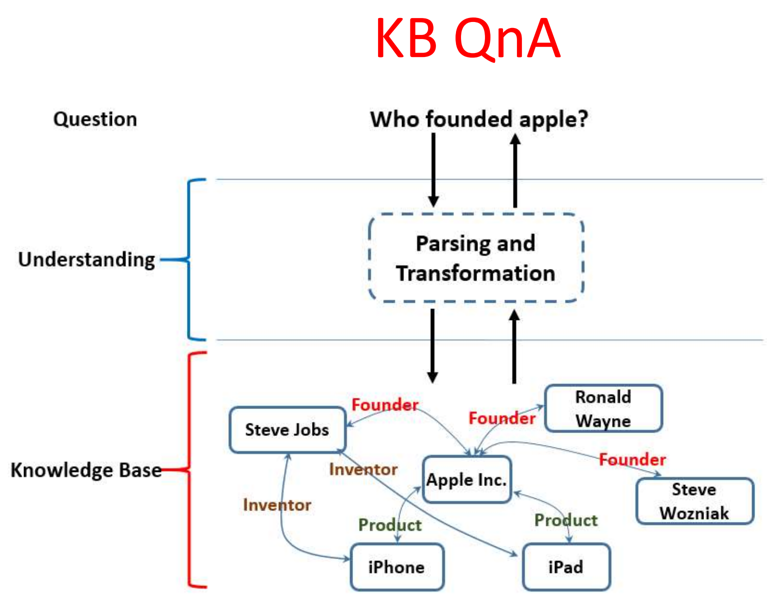
- Semantic parsing is difficult due to ontology mismatch
- Knowledge base is incomplete (missing entities/relations/attribute values)

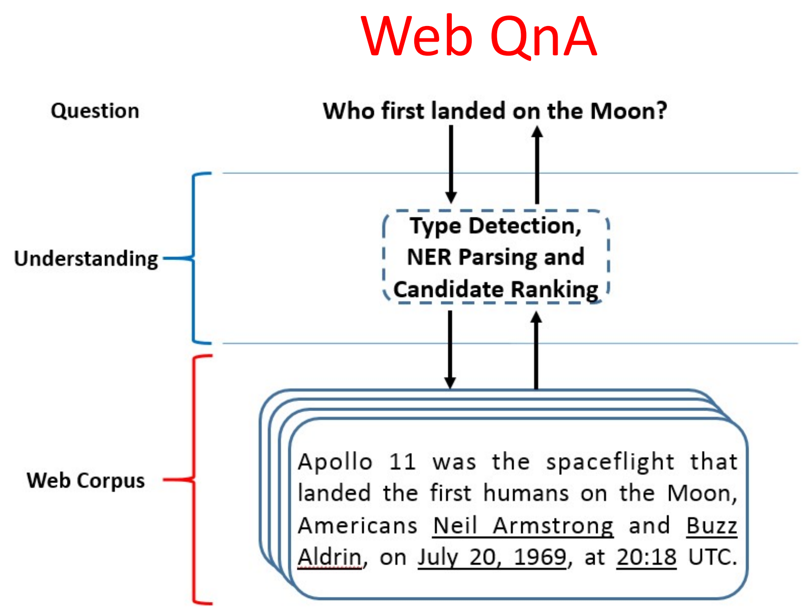
Web

Advantages:

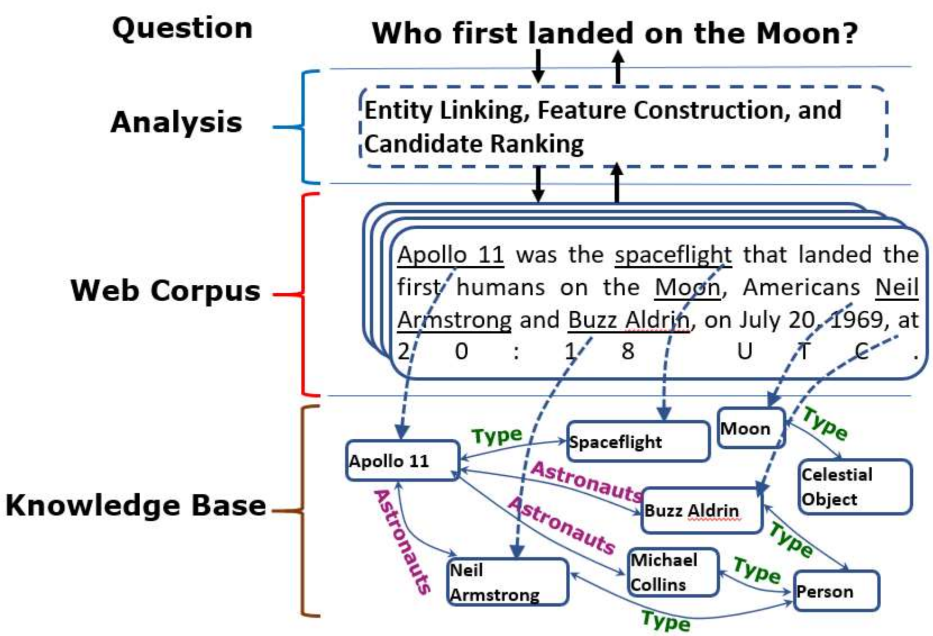
- Contains abundant information
- Redundancy on the Web could help confirm the answers

Factoid Answer based on Web Documents





Question Answering via Semantic Enrichment



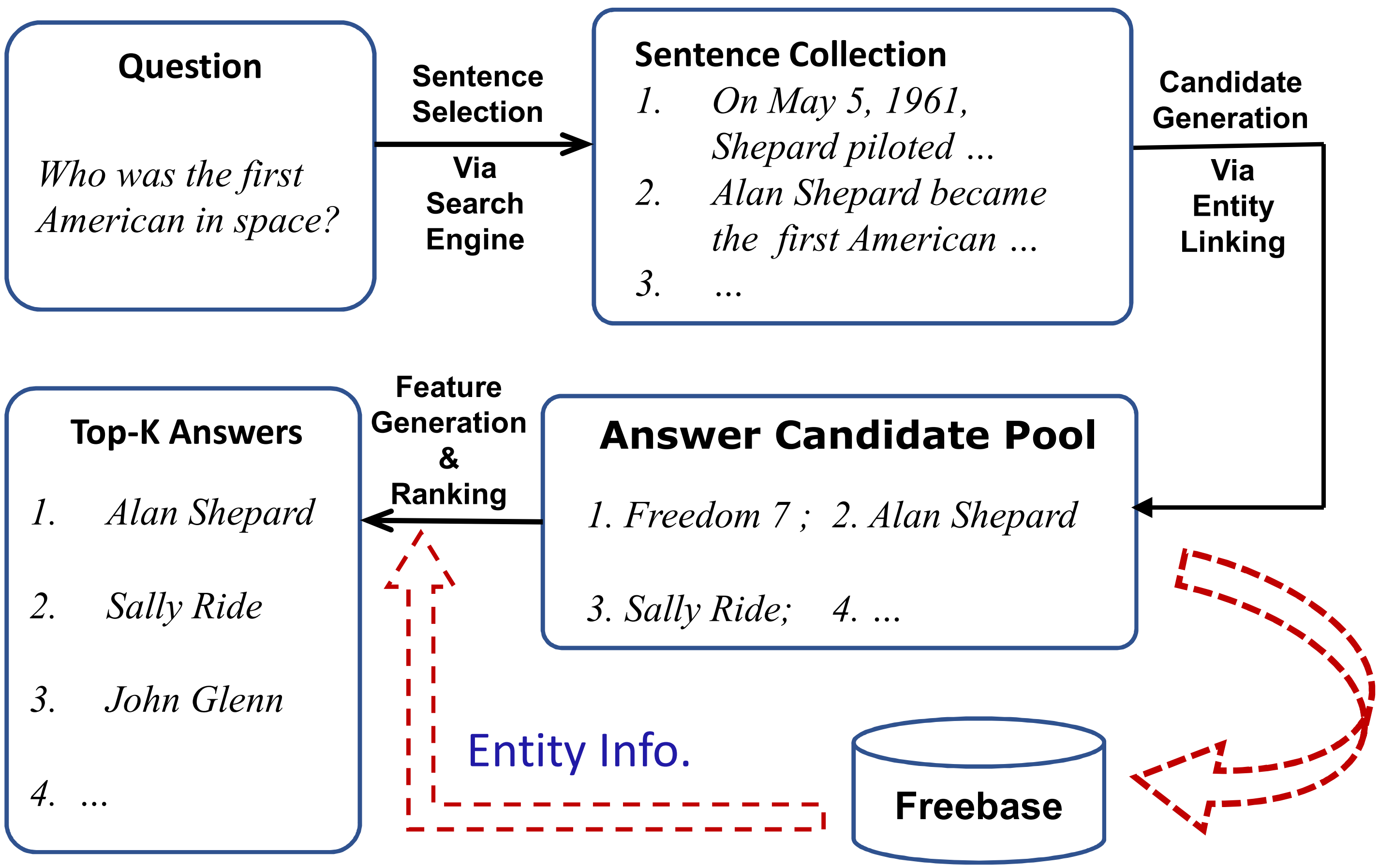
Advantages:

- Generate better answer candidates
  - Entities in Freebase
  - Mentions of the same entity merged to one candidate
- Able to leverage entity information in Freebase
  - Semantic text relevance features for ranking
  - More fine-grained answer type checking

5% ~ 20% improvement in MRR

Open Domain Question and Answering via Semantic Enrichment [Huan Sun, et al., WWW 2015]

System Framework



Open Domain Question and Answering via Semantic Enrichment [Huan Sun, et al., WWW 2015]

Experiments – Data and Evaluation

- TREC Datasets (well-formed questions)
  - Training: 1,700 (entity) questions (TREC 8-11)
  - Testing: 202 (entity) questions (TREC 12)

Example questions:
1. What are pennies made of?
2. What is the tallest building in Japan?
3. Who sang "Tennessee Waltz"?

- Bing Queries (queries with question intent)
  - Training: 4,725 queries; Testing: 1,164 queries

Example queries:
1. the highest flying bird
2. indiana jones named after
3. designer of the golden gate bridge

- QuASE (**Qu**estion **A**nswering via **S**emantic **E**nrichment)
  - Includes other basic features (e.g., candidate freq.)
  - Ranker learner: MART (Multiple Additive Regression Trees)
- Baselines
  - AskMSR+ [Tsai+ '15] – Web-based QA system
  - SEMPRE [Berant+ '14] – Semantic parsing QA using Freebase
- Evaluation Metrics
  - MRR: Mean Reciprocal Rank
    - Determined by the top-ranked correct answer

5% ~ 20% improvement in MRR

Summary – Knowledge Graph Applications

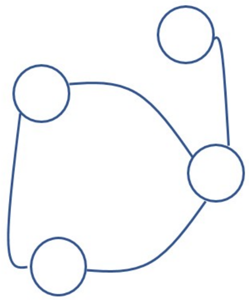
- Entity Recommendation
  - Co-occurrence based
  - Similarity based
    - Textual
    - Embedding
- Question and Answering
  - KG Based
  - Web Based with KB enrichment

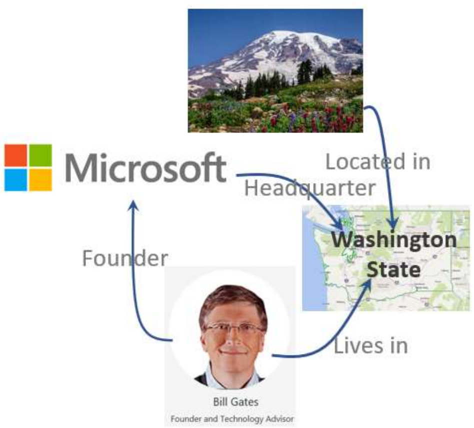
Summary – Knowledge Graph

- History and representation
- KG construction
- KG inference
- KG applications

Summary – Graph & Knowledge Graph

- Graph
  - Graph fundamental - properties
  - Graph applications
  - Graph embedding





Knowledge in the ***graph*** form

- Knowledge Graph
  - KG fundamental - definition / representation
  - KG construction & inference (embedding)
  - KG applications