Knowledge Graph

- Machine Handlable Knowledge

Outline

1. Knowledge Graph
2. Knowledge Extraction
3. Knowledge Graph Construction
4. Overview and Conclusion

What is a knowledge graph?

What is a knowledge graph?

- Knowledge in graph form!

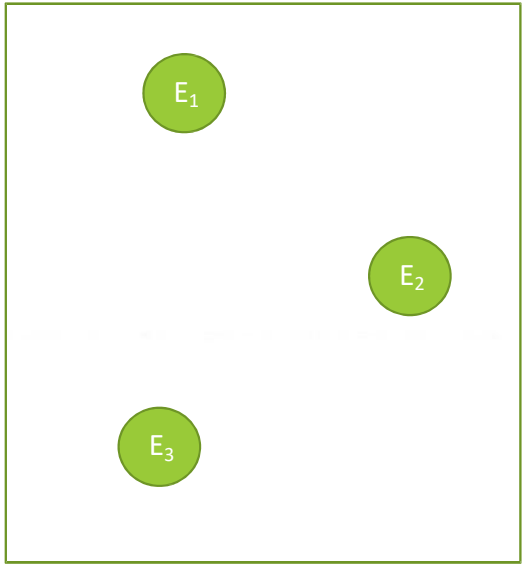
What is a knowledge graph?

- Knowledge in graph form!

- Captures entities, attributes, and relationships

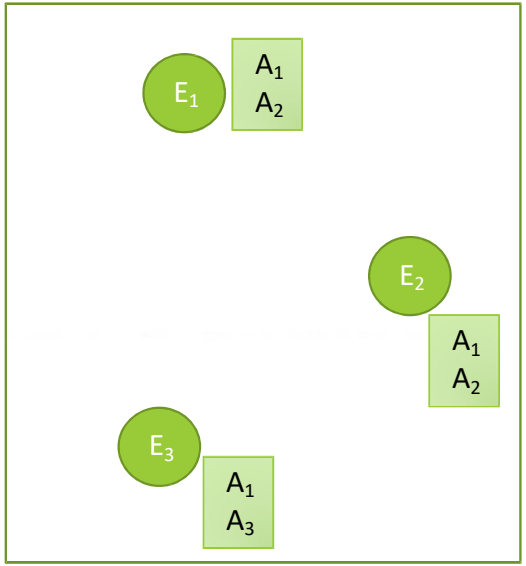
What is a knowledge graph?

- Knowledge in graph form!
- Captures entities, attributes, and relationships
- Nodes are entities



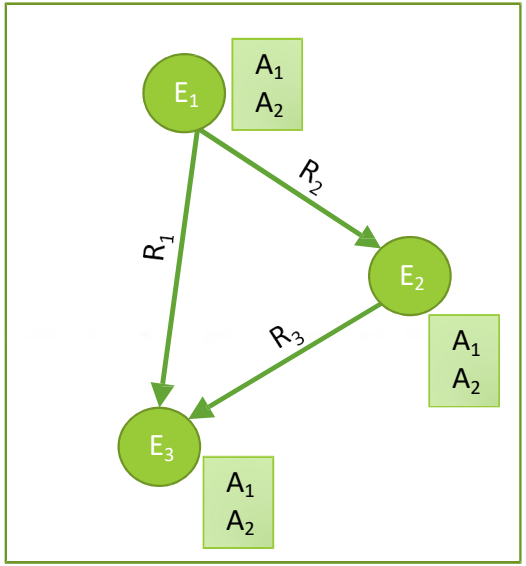
What is a knowledge graph?

- Knowledge in graph form!
- Captures entities, attributes, and relationships
- Nodes are entities
- Nodes are labeled with attributes (e.g., types)



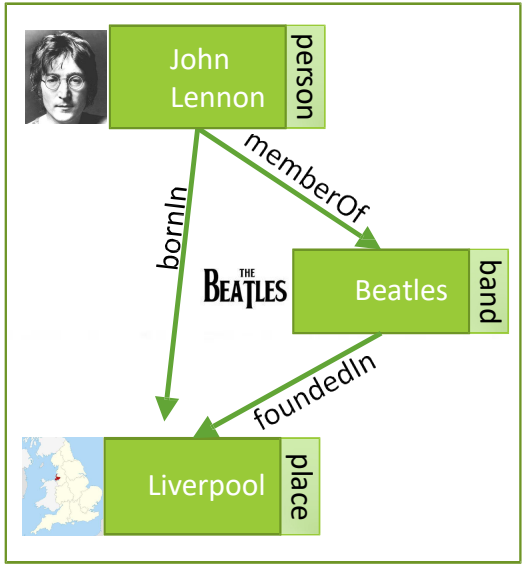
What is a knowledge graph?

- Knowledge in graph form!
- Captures entities, attributes, and relationships
- Nodes are entities
- Nodes are labeled with attributes (e.g., types)
- Typed edges between two nodes capture a relationship between entities



Example knowledge graph

- Knowledge in graph form!
- Captures entities, attributes, and relationships
- Nodes are entities
- Nodes are labeled with attributes (e.g., types)
- Typed edges between two nodes capture a relationship between entities

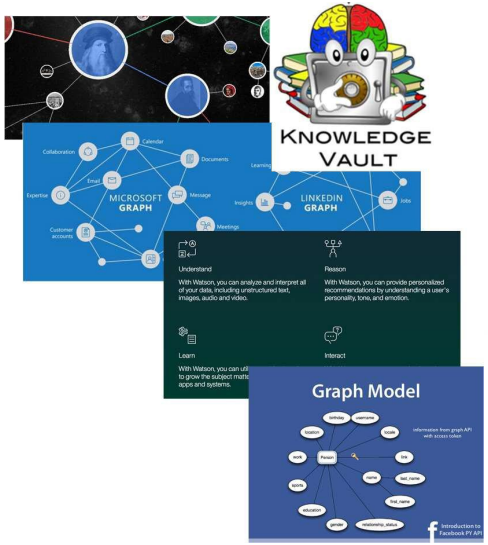


Why knowledge graphs?

- Humans:
  - Combat information overload
  - Explore via intuitive structure
  - Tool for supporting knowledge-driven tasks
- AIs:
  - Key ingredient for many AI tasks
  - Bridge from data to human semantics
  - Use decades of work on graph analysis

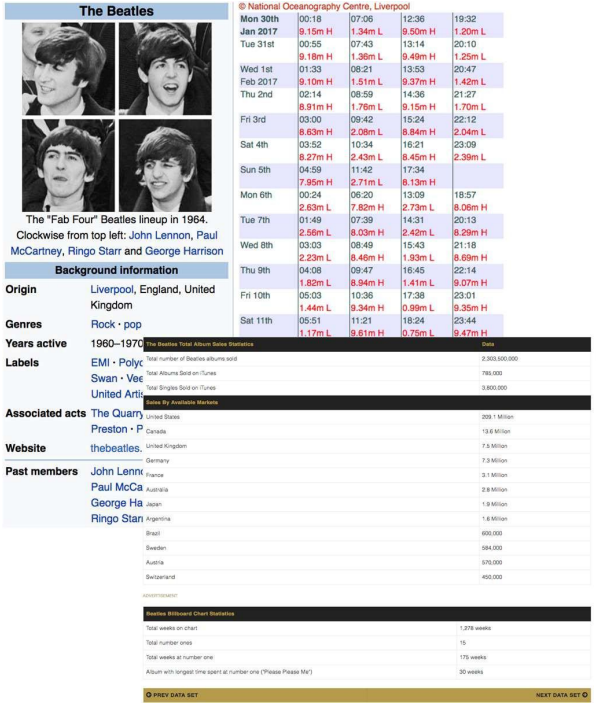
Knowledge Graphs & Industry

- Google Knowledge Graph
  - Google Knowledge Vault
- Amazon Product Graph
- Facebook Graph API
- IBM Watson
- Microsoft Satori
  - Project Hanover/Literome
- LinkedIn Knowledge Graph
- Yandex Object Answer



Where do knowledge graphs comefrom?

- Structured Text
  - Wikipedia Infoboxes, tables, databases, social nets



The Beatles

The "Fab Four" Beatles lineup in 1964. Clockwise from top left: John Lennon, Paul McCartney, Ringo Starr and George Harrison

Background information

Origin	Liverpool, England, United Kingdom
Genres	Rock · pop
Years active	1960–1970
Labels	EMI · Polyd Swan · Vee United Artis
Associated acts	The Quarry Preston · P
Website	thebeatles.
Past members	John Lenn Paul McCa George Ha Ringo Starr

© National Oceanography Centre, Liverpool

Mon 30th	00:18	07:06	12:36	19:32
Jan 2017	9.15m H	1.34m L	9.50m H	1.20m L
Tue 31st	00:55	07:43	13:14	20:10
	9.18m H	1.36m L	9.49m H	1.25m L
Wed 1st	01:33	08:21	13:53	20:47
Feb 2017	9.10m H	1.51m L	9.37m H	1.42m L
Thu 2nd	02:14	08:59	14:36	21:27
	8.91m H	1.76m L	9.15m H	1.70m L
Fri 3rd	03:00	09:42	15:24	22:12
	8.63m H	2.08m L	8.84m H	2.04m L
Sat 4th	03:52	10:34	16:21	23:09
	8.27m H	2.43m L	8.45m H	2.39m L
Sun 5th	04:59	11:42	17:34	
	7.95m H	2.71m L	8.13m H	
Mon 6th	00:24	06:20	13:09	18:57
	2.63m L	7.82m H	2.73m L	8.06m H
Tue 7th	01:49	07:39	14:31	20:13
	2.56m L	8.03m H	2.42m L	8.29m H
Wed 8th	03:03	08:49	15:43	21:18
	2.23m L	8.46m H	1.93m L	8.69m H
Thu 9th	04:08	09:47	16:45	22:14
	1.82m L	8.94m H	1.41m L	9.07m H
Fri 10th	05:03	10:36	17:38	23:01
	1.44m L	9.34m H	0.99m L	9.35m H
Sat 11th	05:51	11:21	18:24	23:44
	1.17m L	9.61m H	0.75m L	9.47m H

The Beatles Total Album Sales Statistics — Data

Total number of Beatles albums sold	2,303,500,000
Total Albums Sold on iTunes	785,000
Total Singles Sold on iTunes	3,800,000

Sales By Available Markets

United States	209.1 Million
Canada	13.6 Million
United Kingdom	7.5 Million
Germany	7.3 Million
France	3.1 Million
Australia	2.8 Million
Japan	1.9 Million
Argentina	1.6 Million
Brazil	600,000
Sweden	584,000
Austria	570,000
Switzerland	450,000

ADVERTISEMENT

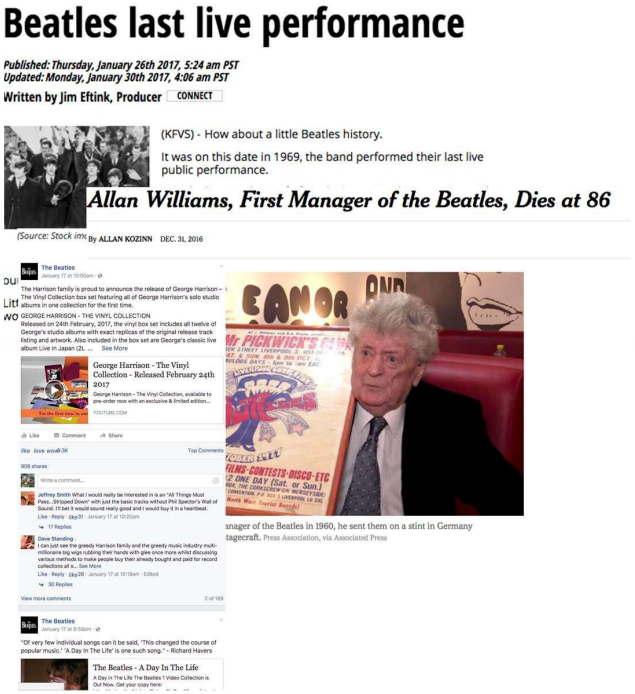
Beatles Billboard Chart Statistics

Total weeks on chart	1,278 weeks
Total number ones	15
Total weeks at number one	175 weeks
Album with longest time spent at number one ("Please Please Me")	30 weeks

PREV DATA SET — NEXT DATA SET

Where do knowledge graphs comefrom?

- Structured Text
  - Wikipedia Infoboxes, tables, databases, social nets
- Unstructured Text
  - WWW, news, social media, reference articles

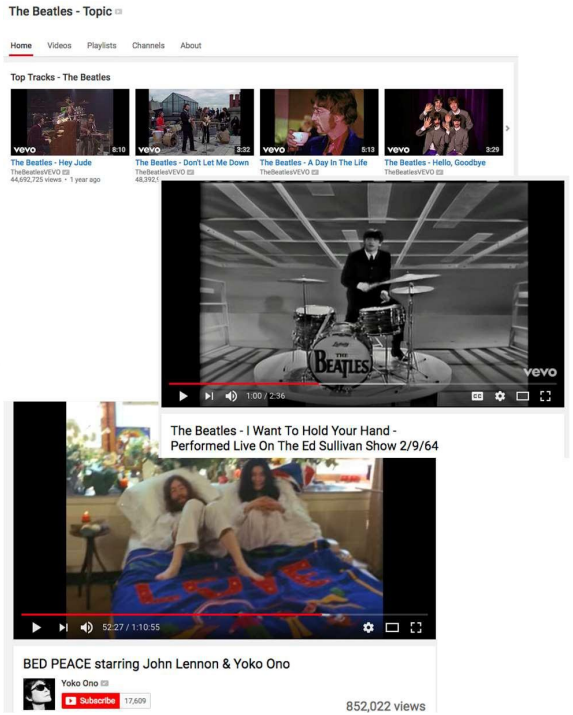


Where do knowledge graphs comefrom?

- Structured Text
  - Wikipedia Infoboxes, tables, databases, social nets
- Unstructured Text
  - WWW, news, social media, reference articles
- Images

Where do knowledge graphs comefrom?

- Structured Text
  - Wikipedia Infoboxes, tables, databases, social nets
- Unstructured Text
  - WWW, news, social media, reference articles
- Images
- Video
  - YouTube, video feeds

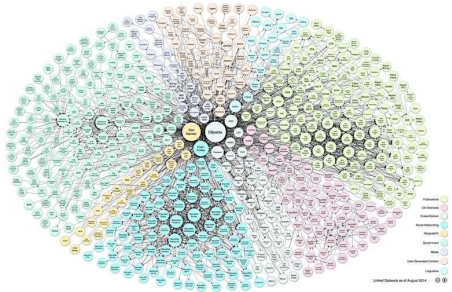


Knowledge Representation

- Decades of research into knowledge representation
- Most knowledge graph implementations use RDF triples
  - <rdf:subject, rdf:predicate, rdf:object> : r(s,p,o)
  - Temporal scoping, reification, and skolemization…
- ABox (assertions) versus TBox (terminology)
- Common ontological primitives
  - rdfs:domain, rdfs:range, rdf:type, rdfs:subClassOf, rdfs:subPropertyOf, …
  - owl:inverseOf, owl:TransitiveProperty, owl:FunctionalProperty, …

Semantic Web

- Standards for defining and exchanging knowledge
  - RDF, RDFa, JSON-LD, schema.org
  - RDFS, OWL, SKOS, FOAF
- Annotated data provide critical resource for automation
- Major weakness: annotate everything?



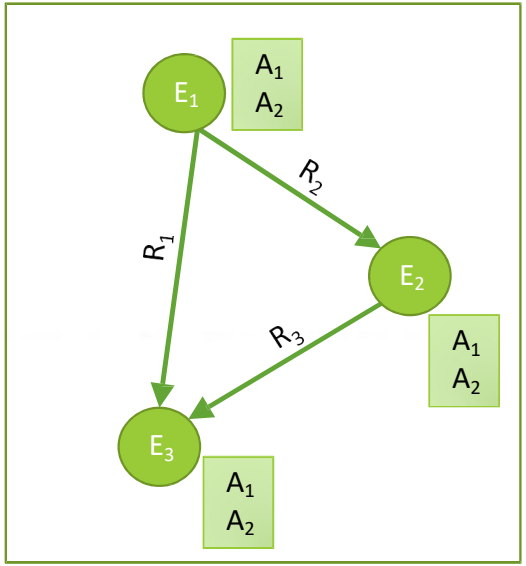
Information Extraction from Text

- Answer to the knowledge acquisition bottleneck

- Many challenges:
  - chunking
  - polysemy/word sense disambiguation
  - entity coreference
  - relational extraction

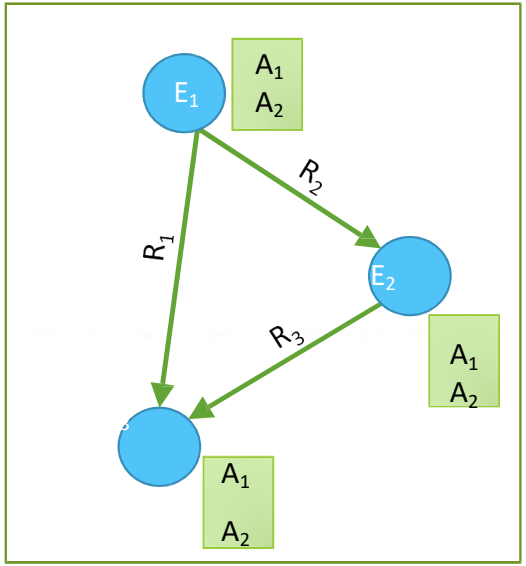
What is a knowledge graph?

- Knowledge in graph form!
- Captures entities, attributes, and relationships
- Nodes are entities
- Nodes are labeled with attributes (e.g., types)
- Typed edges between two nodes capture a relationship between entities



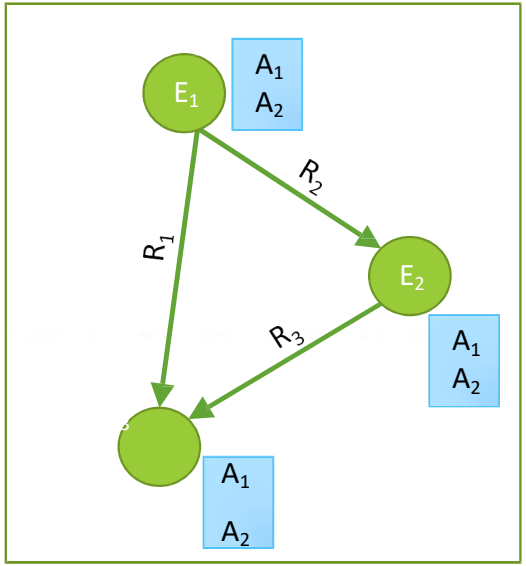
Basic problems

- **Who** are the entities (nodes) in the graph?



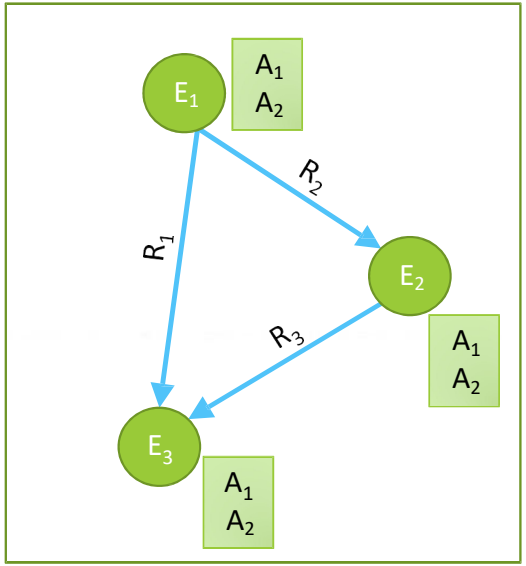
Basic problems

- **Who** are the entities (nodes) in the graph?
- **What** are their attributes and types (labels)?



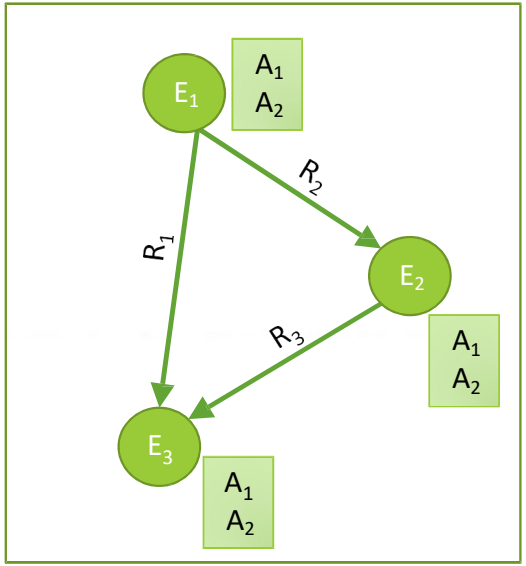
Basic problems

- **Who** are the entities (nodes) in the graph?
- **What** are their attributes and types (labels)?
- **How** are they related (edges)?



Basic problems

- **Who** are the entities (nodes) in the graph?
- **What** are their attributes and types (labels)?
- **How** are they related (edges)?



Knowledge Graph Construction

Two perspectives

- Knowledge Extraction
- Who are the entities (nodes) in the graph?
  - Named Entity Recognition
  - Entity Coreference
- What are their attributes and types (labels)?
  - Named Entity Recognition
- How are they related (edges)?
  - Relation Extraction
  - Semantic Role Labeling

Two perspectives

- Graph Construction
- Who are the entities (nodes) in the graph?
  - Entity Linking
  - Entity Resolution
- What are their attributes and types (labels)?
  - Collective Classification
- How are they related (edges)?
  - Link Prediction

What is NLP?

Unstructured

Ambiguous

Lots and lots of it!

Humans can read them, but

… very slowly

… can’t remember all

… can’t answer questions

Structured

Precise, Actionable

Specific to the task

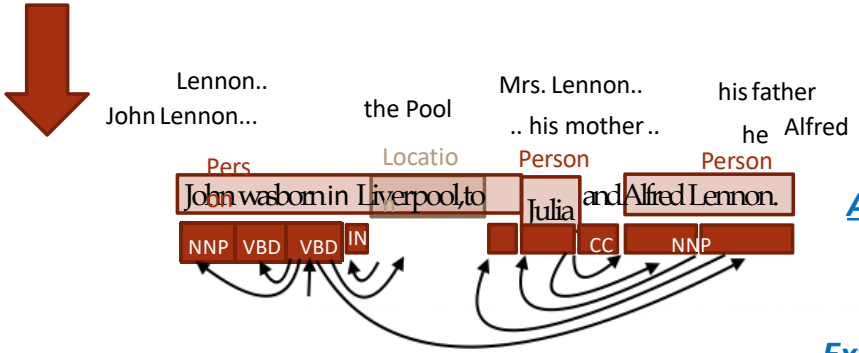
Can be used for downstream applications, such as creating Knowledge Graphs!

Knowledge Extraction

John was born in Liverpool, to Julia and Alfred Lennon.

Text

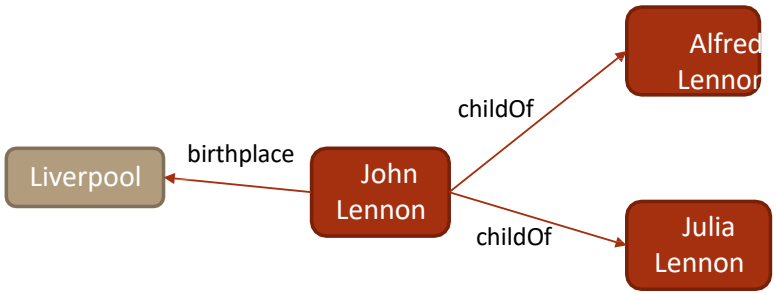
NLP



Annotated text

Extraction graph

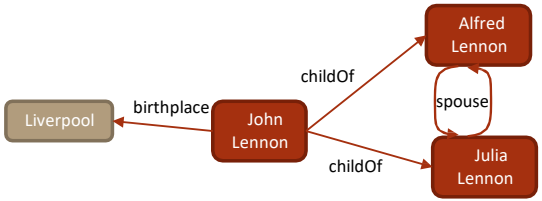
Information Extraction



Breaking it Down

Information Extraction

Entity resolution,
Entity linking,
Relation extraction…



Document

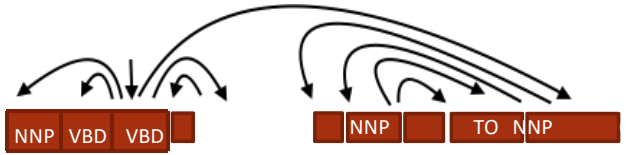
Coreference Resolution...

Lennon..
John Lennon...
the Pool
Mrs. Lennon..
.. his mother ..
his father
Alfred

Person: John — was born in — Location: Liverpool, to — Person: Julia — and — Person: Alfred Lennon.

Sentence

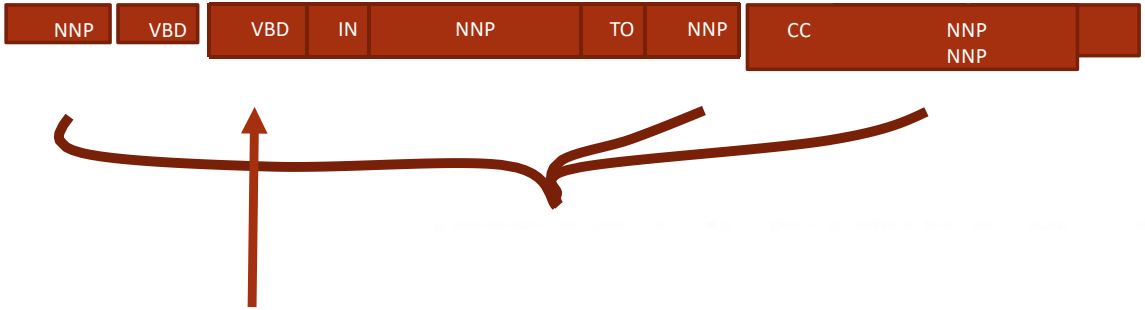
Dependency Parsing, Part of speech tagging,
Named entity recognition…



John was born in Liverpool, to Julia and Alfred Lennon.

Tagging the Parts of Speech

- John was born in Liverpool, to Julia and Alfred Lennon.



Nouns are entities Verbs are relations

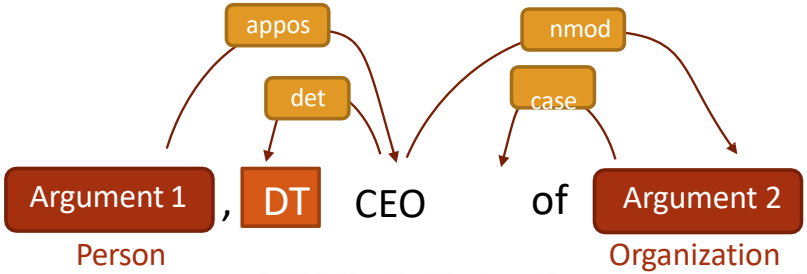
- Common approaches include CRFs, CNNs, LSTMs

Detecting Named Entities

- Structured prediction approaches
- Capture entity mentions and entity types

NLP annotations → features forIE

Combine tokens, dependency paths, and entity types to define rules.



Bill Gates, the CEO of Microsoft, said …

Mr. Jobs, the brilliant and charming CEO of Apple Inc., said …

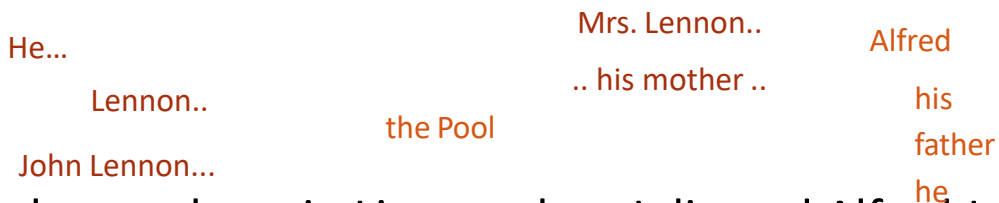
… announced by Steve Jobs, the CEO of Apple.

… announced by Bill Gates, the director and CEO of Microsoft.

… mused Bill, a former CEO of Microsoft.

and many other possible instantiations…

Within-document Coreference



John was born in Liverpool, to Julia and Alfred Lennon.

- Pairwise model for each noun/pronoun
- Can consolidate information, provide context

Entity Names: Two Main Problems

Entities with Same Name

Same type of entities share names

Kevin Smith, John Smith, Springfield, …

Things named after each other

Clinton, Washington, Paris, Amazon, Princeton, Kingston, …

Partial Reference

First names of people, Location instead of team name, Nick names

Different Names for Entities

Nick Names

Bam Bam, Drumpf, …

Typos/Misspellings

Baarak, Barak, Barrack, …

Inconsistent References

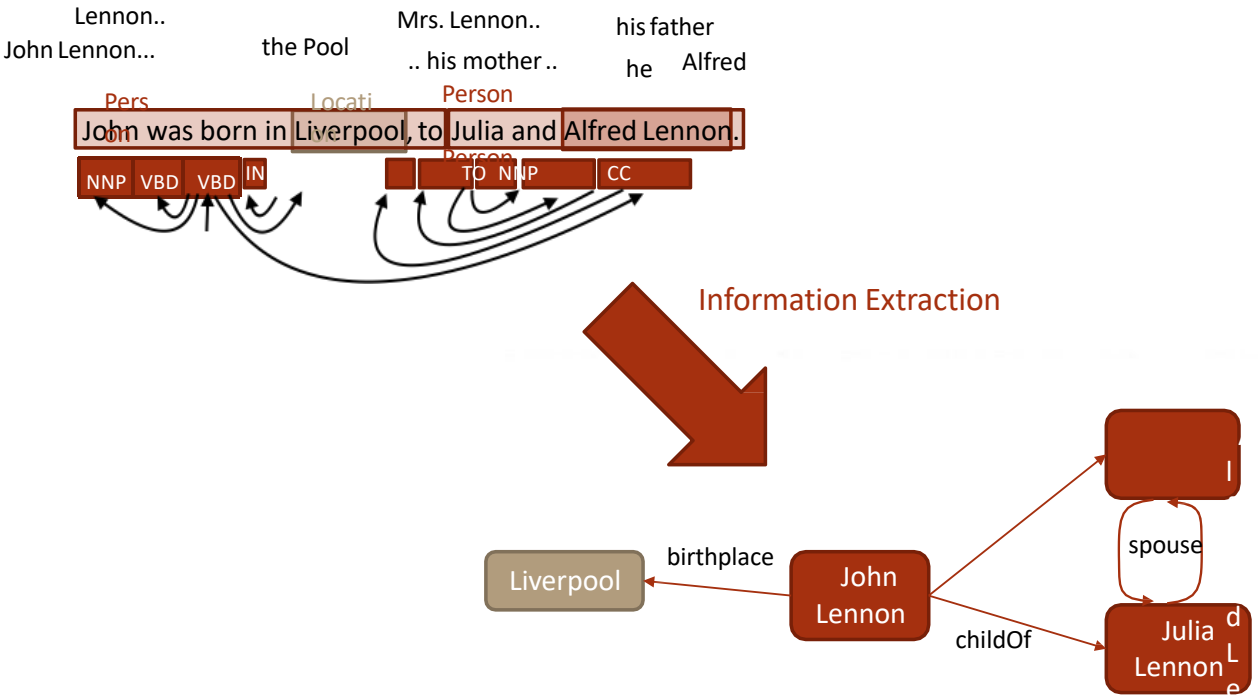
MSFT, APPL, GOOG…

Entity Linking Approach

Washington drops 10 points after game with UCLA Bruins.

Candidate Generation		Washington DC, George Washington, Washington state, Lake Washington, Washington Huskies, Denzel Washington, University of Washington, Washington High School, …
Entity Types	LOC/ORG	Washington DC, ~~George Washington~~, Washington state, Lake Washington, Washington Huskies, ~~Denzel Washington~~, University of Washington, Washington High School, …
Coreference	UWashington, Huskies	~~Washington DC~~, ~~George Washington~~, ~~Washington state~~, ~~Lake Washington~~, Washington Huskies, ~~Denzel Washington~~, University of Washington, ~~Washington High School~~, …
Coherence	UCLA Bruins, USC Trojans	~~Washington DC~~, ~~George Washington~~, ~~Washington state~~, ~~Lake Washington~~, Washington Huskies, ~~Denzel Washington~~, ~~University of Washington~~, ~~Washington High School~~, …

Information Extraction



Information Extraction

3 CONCRETE SUB-PROBLEMS

Defining domain

Learning extractors

Scoring the facts

×

3 LEVELS OF SUPERVISION

Supervised

Semi-supervised

Unsupervised

Information Extraction

3 CONCRETE SUB-PROBLEMS

Defining domain

Learning extractors

Scoring the facts

×

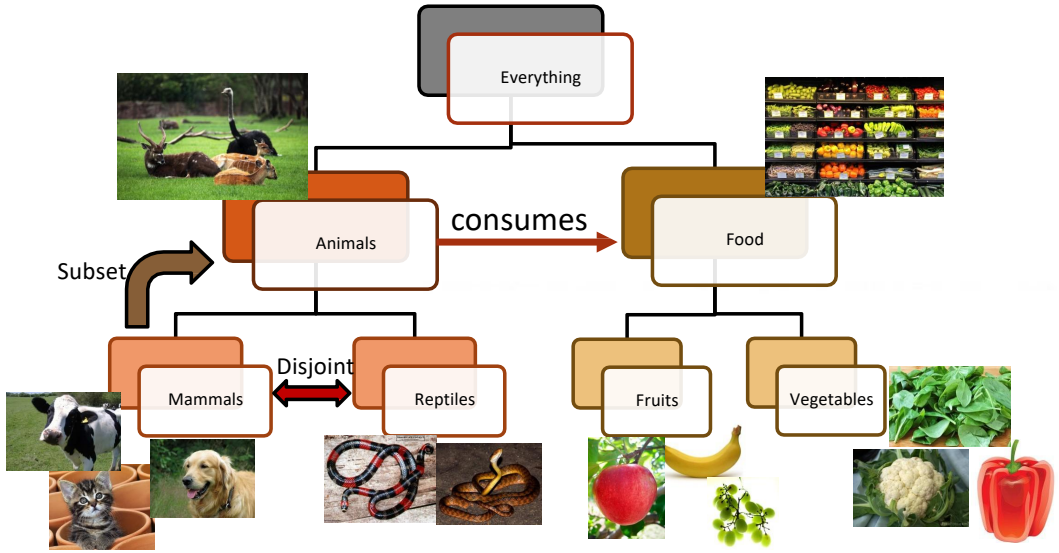
3 LEVELS OF SUPERVISION

Supervised

Semi-supervised

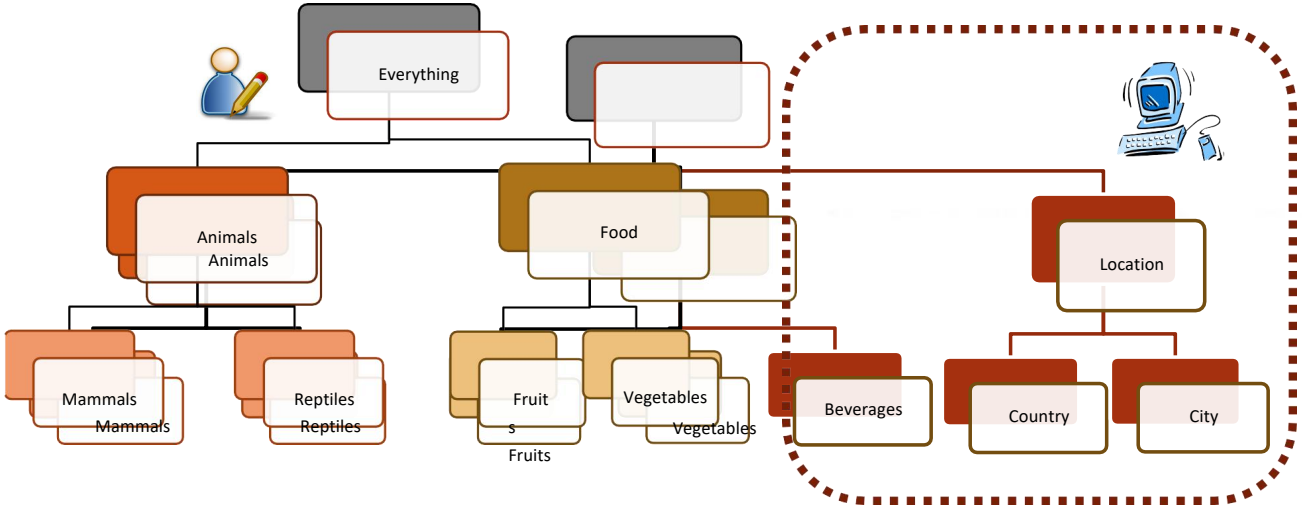
Unsupervised

Defining Domain: Manual



Defining Domain: Semi-automatic

- Subset of types are manually defined
- SSL methods discover new types from unlabeled data



Defining Domain: Automatic

- Any noun phrase is a candidate entity
  - Dog, cat, cow, reptile, mammal, apple, greens, mixed greens, lettuce, red leaf lettuce, romaine lettuce, iceberg lettuce…
- Any verb phrase is a candidate relation
  - Eats, feasts on, grazes, consumes,

Information Extraction

3 CONCRETE SUB-PROBLEMS

Defining domain

Learning extractors

Scoring candidate facts

×

3 LEVELS OF SUPERVISION

Supervised

Semi-supervised

Unsupervised

Learning Extractors

- Supervised: high precision patterns
  - <PERSON> plays in <BAND>

- Semi-supervised: Bootstrapping to learn patterns
  - Create examples (John Lennon, Beatles), find patterns
  - Manually correct incorrect patterns

- Unsupervised: cluster phrases with constraints
  - Identify candidate verb phrases, find candidate arguments, cluster by NER types

Information Extraction

3 CONCRETE SUB-PROBLEMS

Defining domain extractors

Learning

Scoring candidate facts

3 LEVELS OF SUPERVISION

Supervised

Semi-supervised

Unsupervised

Scoring the candidate facts

- Human defined scoring function or

  Scoring function learnt using supervised ML with large amount of training data

  {expensive, high precision}

- Small amount of training data is available

  scoring refined over multiple iterations using both labeled and unlabeled data

- Completely automatic (Self-training)

  Confidence(extraction pattern) ∝ (#unique instances it could extract)

  Score(candidate fact) ∝ (#distinct extraction patterns that support it)

  {cheap, leads to semantic drift}

IE systems in practice

	Defining domain	Learning extractors	Scoring candidate facts	Fusing extractors
ConceptNet				
NELL				Heuristic rules
Knowledge Vault				Classifier
OpenIE				

Knowledge Extraction: Key Points

- Built on the foundation of NLP techniques
  - Part-of-speech tagging, dependency parsing,

    named entity recognition, coreference resolution…
  - Challenging problems with very useful outputs
- Information extraction techniques use NLP to:
  - define the domain
  - extract entities and relations
  - score candidate outputs
- Trade-off between manual & automatic methods

Graph Construction Issues

Extracted knowledge is:

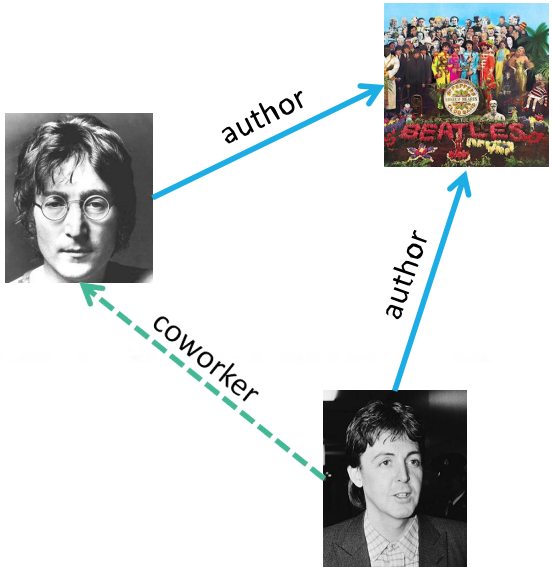
- ambiguous:
  - Ex: Beetles, beetles, Beatles
  - Ex: citizenOf, livedIn, bornIn

Graph Construction Issues

Extracted knowledge is:

- ambiguous

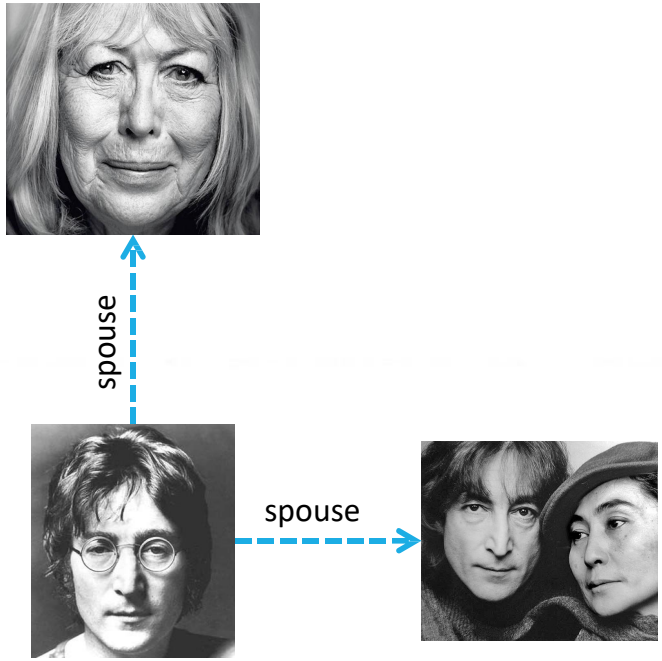
- incomplete
  - Ex: missing relationships
  - Ex: missing labels
  - Ex: missing entities



Graph Construction Issues

Extracted knowledge is:

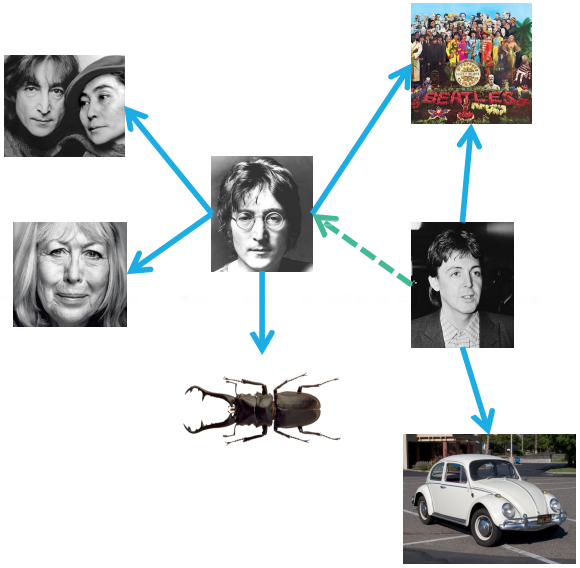
- ambiguous
- incomplete
- inconsistent
  - Ex: Cynthia Lennon, Yoko Ono
  - Ex: exclusive labels (alive, dead)
  - Ex: domain-range constraints



Graph Construction Issues

Extracted knowledge is:

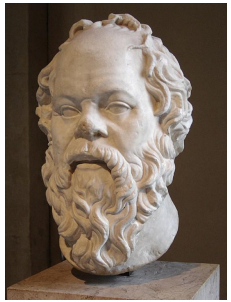
- ambiguous
- incomplete
- inconsistent



Graph Construction approach

- Graph construction **cleans** and **completes** extraction graph
- Incorporate ontological constraints and relational patterns
- Discover statistical relationships within knowledge graph

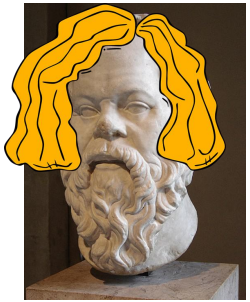
Beyond Pure Reasoning



- Classical AI approach to knowledge: reasoning

Lbl(Socrates, Man) & Sub(Man, Mortal) -> Lbl(Socrates, Mortal)

Beyond Pure Reasoning

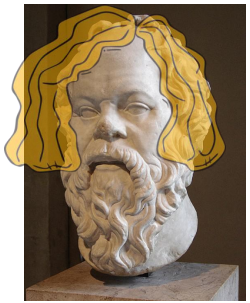


- Classical AI approach to knowledge: reasoning

Lbl(Socrates, Man) & Sub(Man, Mortal) -> Lbl(Socrates, Mortal)

- Reasoning difficult when extracted knowledge has errors

Beyond Pure Reasoning



- Classical AI approach to knowledge: reasoning

Lbl(Socrates, Man) & Sub(Man, Mortal) -> Lbl(Socrates, Mortal)

- Reasoning difficult when extracted knowledge has errors
- Solution: probabilistic models

P(Lbl(Socrates, Mortal)|Lbl(Socrates,Man)=0.9)

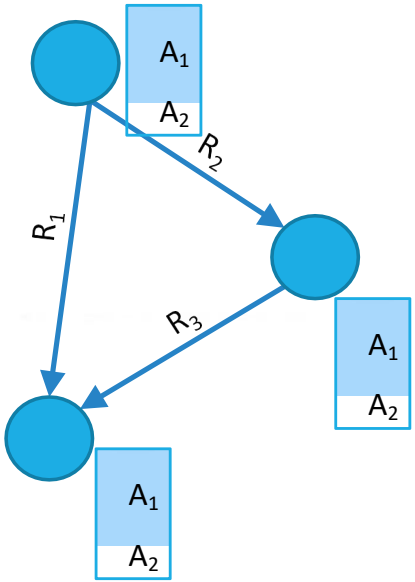
Graphical Models: Overview

- Define **joint probability distribution** on knowledge graphs
- Each candidate fact in the knowledge graph is a **variable**
- Statistical signals, ontological knowledge and rules parameterize the **dependencies** between variables
- Find most likely knowledge graph by **optimization/sampling**

Knowledge Graph Identification

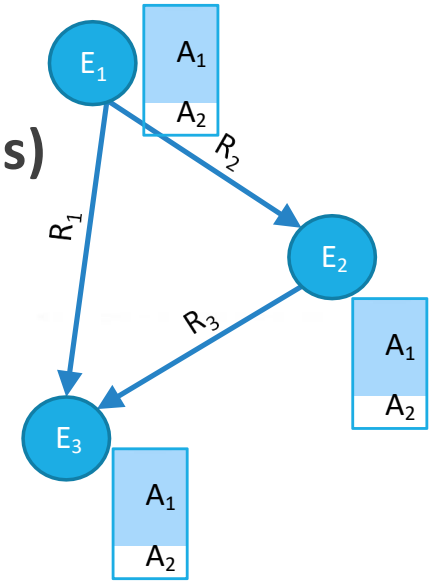
Define a graphical model to perform all three of these tasks simultaneously!

- **Who** are the entities (nodes) in the graph?
- **What** are their attributes and types (labels)?
- **How** are they related (edges)?



Knowledge Graph Identification

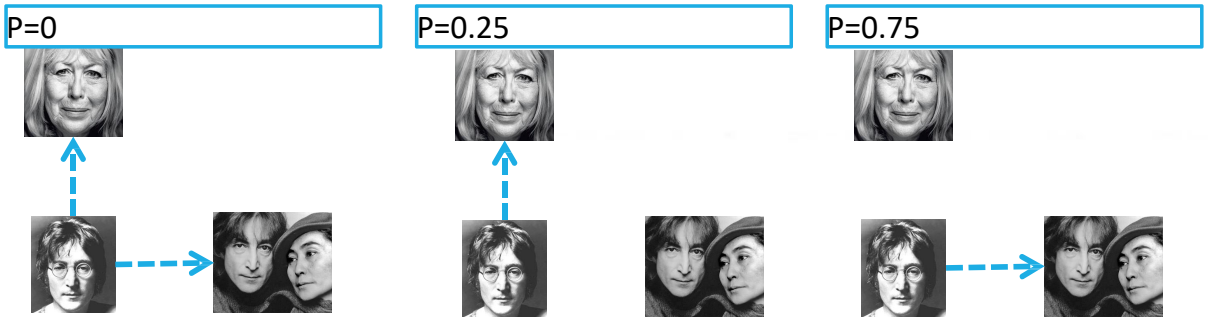
P(Who, What, How | Extractions)



Probabilistic Models

- Use dependencies between facts in KG

- Probability defined *jointly* over facts



What determines probability?

- **Statistical signals from text extractors and classifiers**

What determines probability?

- **Statistical signals from text extractors and classifiers**
  - P(R(John,Spouse,Yoko))=0.75;
    P(R(John,Spouse,Cynthia))=0.25
  - LevenshteinSimilarity(Beatles, Beetles) = 0.9

What determines probability?

- Statistical signals from text extractors and classifiers

- **Ontological knowledge about domain**

What determines probability?

- Statistical signals from text extractors and classifiers

- **Ontological knowledge about domain**
  - Functional(Spouse) & R(A,Spouse,B) -> !R(A,Spouse,C)
  - Range(Spouse, Person) & R(A,Spouse,B) -> Type(B, Person)

What determines probability?

- Statistical signals from text extractors and classifiers
- Ontological knowledge about domain
- **Rules and patterns mined from data**

What determines probability?

- Statistical signals from text extractors and classifiers
- Ontological knowledge about domain
- **Rules and patterns mined from data**
  - R(A, Spouse, B) & R(A, Lives, L) -> R(B, Lives, L)
  - R(A, Spouse, B) & R(A, Child, C) -> R(B, Child, C)

What determines probability?

- **Statistical signals from text extractors and classifiers**
  - P(R(John,Spouse,Yoko))=0.75;
    P(R(John,Spouse,Cynthia))=0.25
  - LevenshteinSimilarity(Beatles, Beetles) = 0.9

- **Ontological knowledge about domain**
  - Functional(Spouse) & R(A,Spouse,B) -> !R(A,Spouse,C)
  - Range(Spouse, Person) & R(A,Spouse,B) -> Type(B, Person)

- **Rules and patterns mined from data**
  - R(A, Spouse, B) & R(A, Lives, L) -> R(B, Lives, L)
  - R(A, Spouse, B) & R(A, Child, C) -> R(B, Child, C)

Graphical Models: Pros/Cons

BENEFITS	DRAWBACKS
• Define probability distribution over KGs	• Requires optimization over all KG facts - overkill
• Easily specified via rules	• Dependent on rules from ontology/expert
• Fuse knowledge from many different sources	• Require probabilistic semantics - unavailable

Random Walk Overview

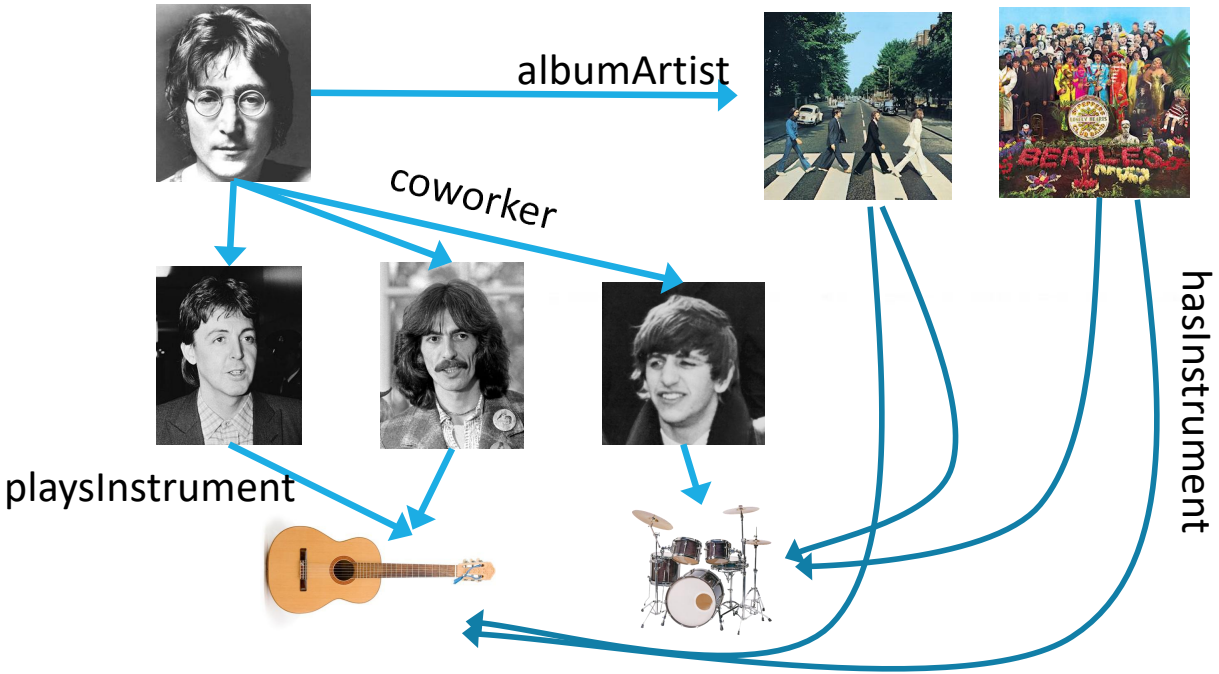
- Given: a query of an **entity** and **relation**
- Starting at the entity, **randomly walk** the KG
- Random walk ends when reaching an appropriate **goal**
- Learned **parameters** bias choices in the random walk
- Output **relative probabilities** of goal states

Random Walk Illustration

Query: R(Lennon, PlaysInstrument, ?)

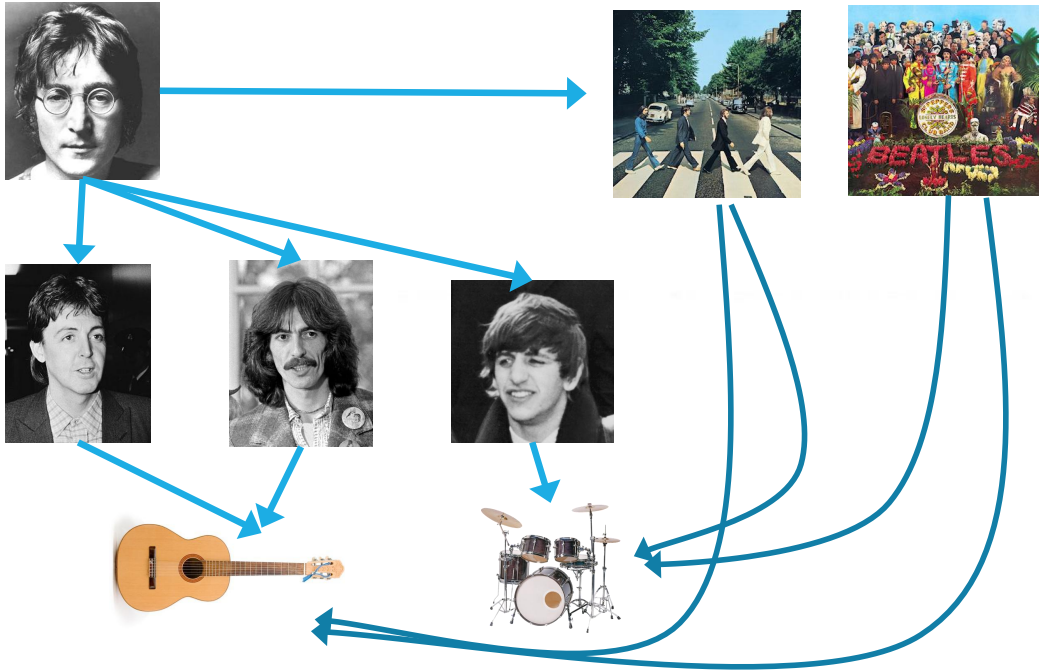
Random Walk Illustration

Query: R(Lennon, PlaysInstrument, ?)



Random Walk Illustration

Query: R(Lennon, PlaysInstrument, ?)

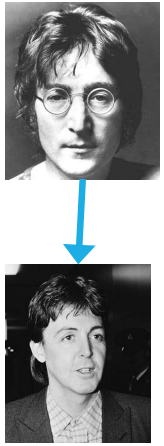


Random Walk Illustration

Query Q: R(Lennon, PlaysInstrument, ?)

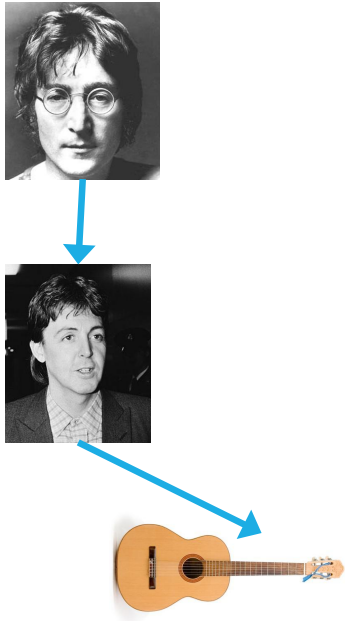
Random Walk Illustration

Query Q: R(Lennon, PlaysInstrument, ?)



Random Walk Illustration

Query Q: R(Lennon, PlaysInstrument, ?)



Random Walk Illustration

Query Q: R(Lennon, PlaysInstrument, ?)

Path

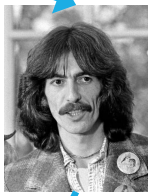
Weight of path

P(Q|θ=<coworker,playsInstrument>) W

Random Walk Illustration

Query Q: R(Lennon, PlaysInstrument, ?)

P(Q|θ=<coworker,playsInstrument>) W



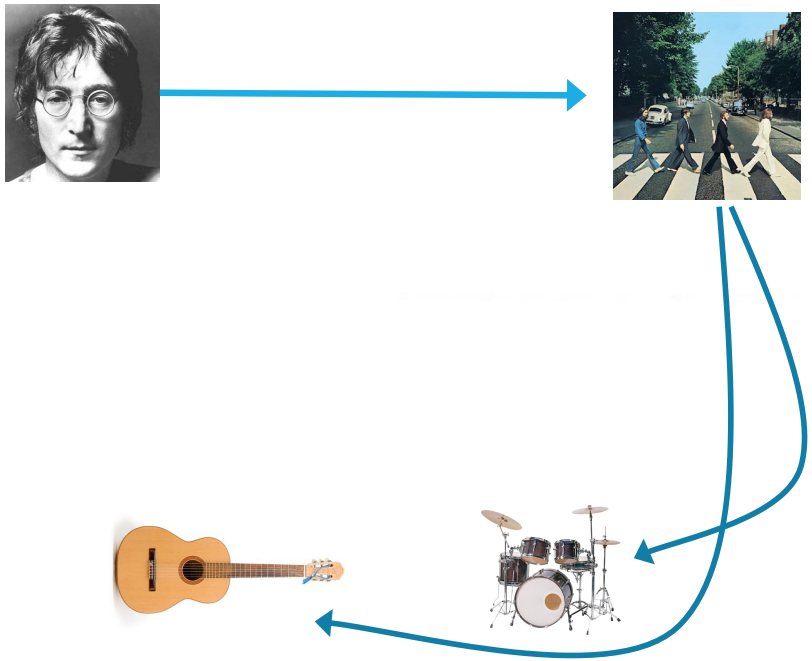
Random Walk Illustration

Query Q: R(Lennon, PlaysInstrument, ?)

P(Q|θ=<coworker,playsInstrument>) W

Random Walk Illustration

Query Q: R(Lennon, PlaysInstrument, ?)



Random Walk Illustration

Query Q: R(Lennon, PlaysInstrument, ?)

P(Q|θ=<albumArtist,hasInstrument>) W

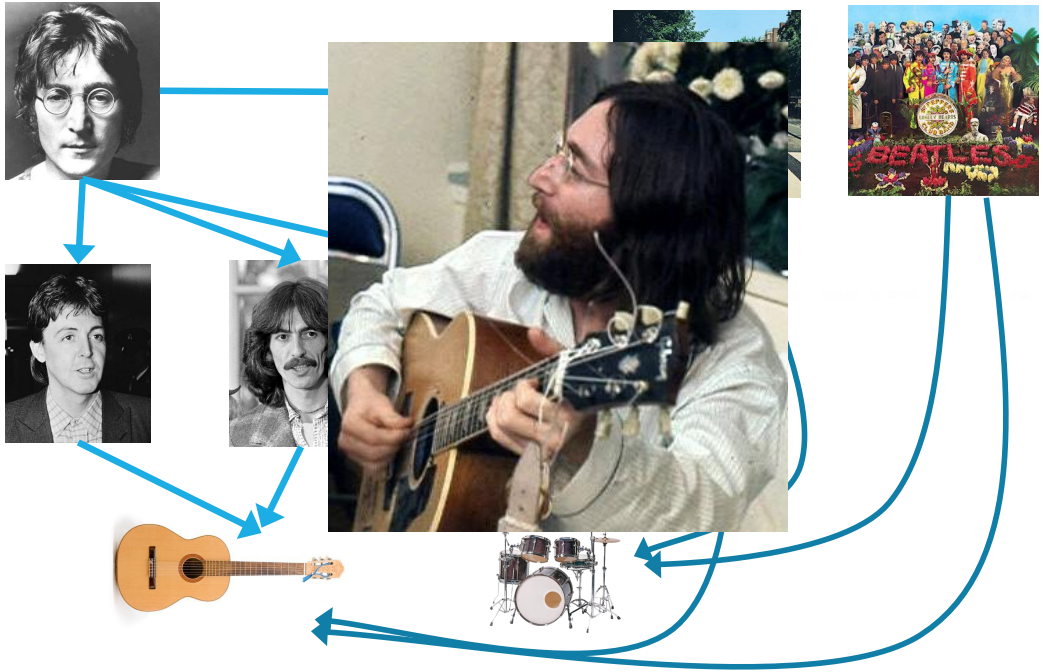
Random Walk Illustration

Query Q: R(Lennon, PlaysInstrument, ?)

P(Q|θ=<albumArtist,hasInstrument>) W

Random Walk Illustration

Query: R(Lennon, PlaysInstrument, ?)



Random Walks: Pros/Cons

BENEFITS	DRAWBACKS
• KG query estimation independent of KG size	• Full KG completion task inefficient
• Model training produces interpretable, logical rules	• Training data difficult to obtain at scale
• Robust to noisy extractions through probabilistic form	• Input must follow probabilistic semantics

Two classes of Probabilistic Models

- GRAPHICAL MODELS
  - Possible facts in KG are variables
  - Logical rules relate facts
  - Probability rules
  - Universally-quantified

Two classes of Probabilistic Models

- RANDOM WALK METHODS
  - Possible facts posed as queries
  - Random walks of the KG constitute “proofs”
  - Probability path lengths/transitions
  - Locally grounded

Probabilistic Models: Downsides

Limitation to Logical Relations

- Representation restricted by manual design
  - Clustering? Assymetric implications?
  - Information flows through these relations
- Difficult to generalize to unseen entities/relations

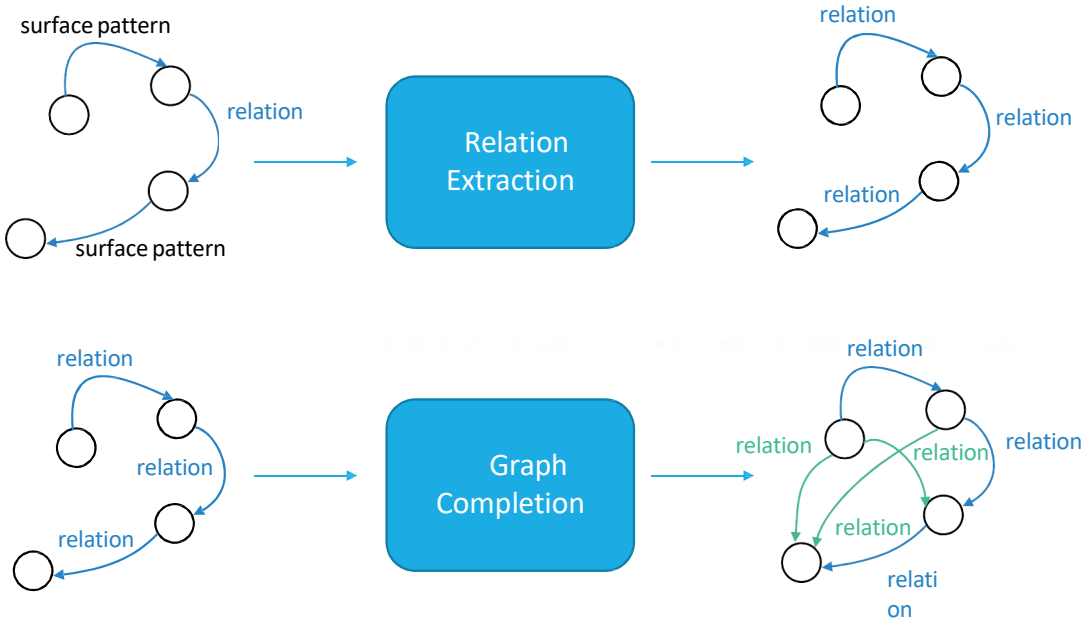
Computational Complexity of Algorithms

- Complexity depends on explicit dimensionality
  - Often NP-Hard, in size of data
  - More rules, more expensive inference
- Query-time inference is sometimes NP-Hard
- Not trivial to parallelize, or use GPUs

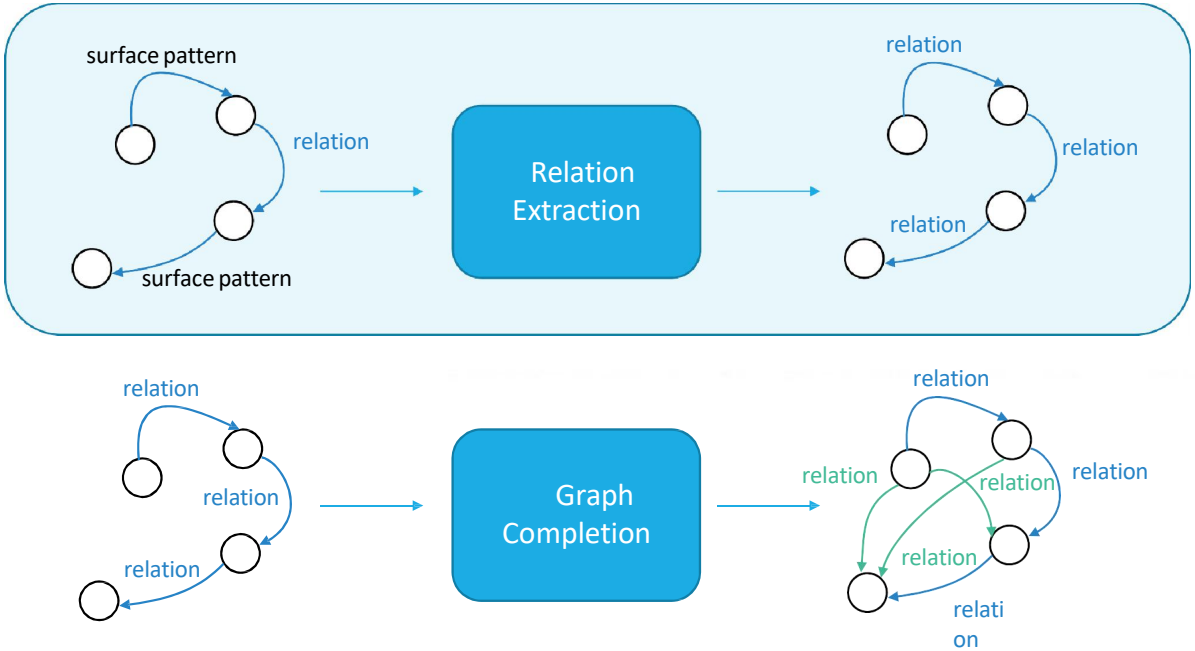
Embeddings

- Everything as dense vectors
- Can capture many relations
- Learned from data
- Complexity depends on latent dimensions
- Learning using stochastic gradient, back-propagation
- Querying is often cheap
- GPU-parallelism friendly

Two Related Tasks



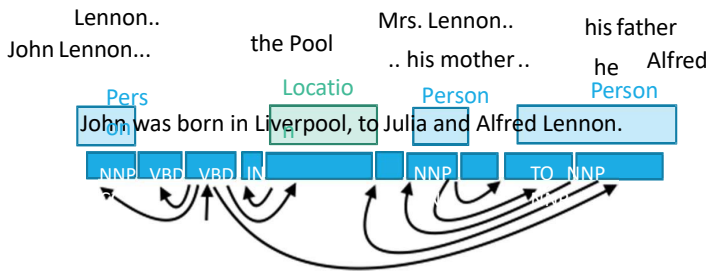
Two Related Tasks



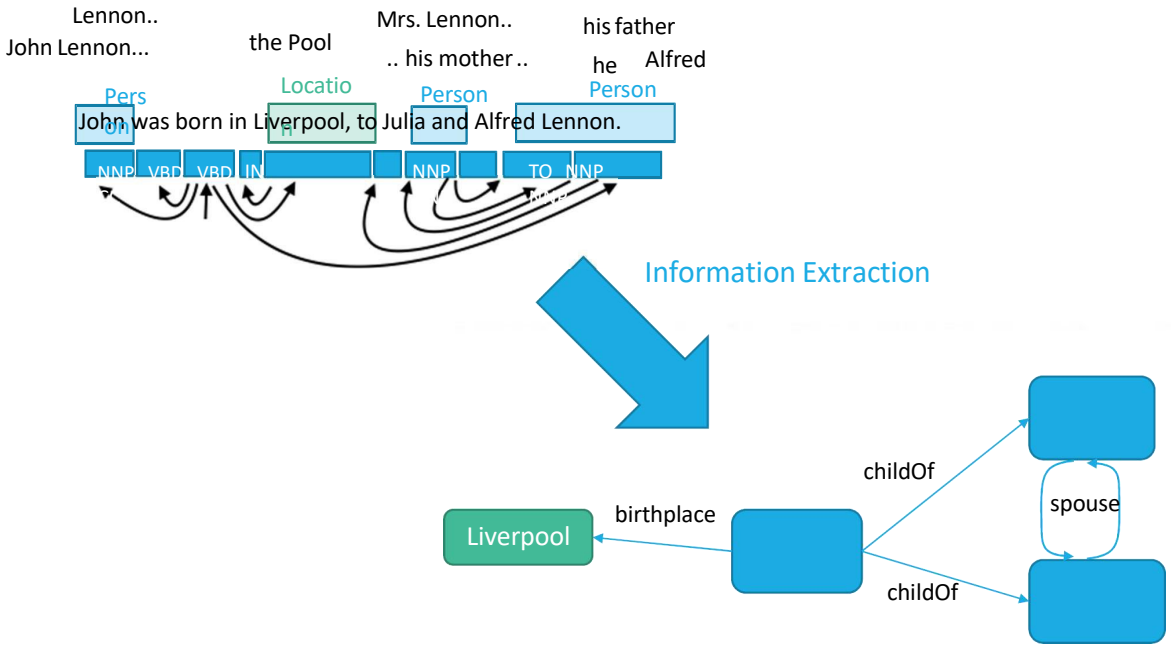
What is NLP?

John was born in Liverpool, to Julia and Alfred Lennon.

Natural Language Processing

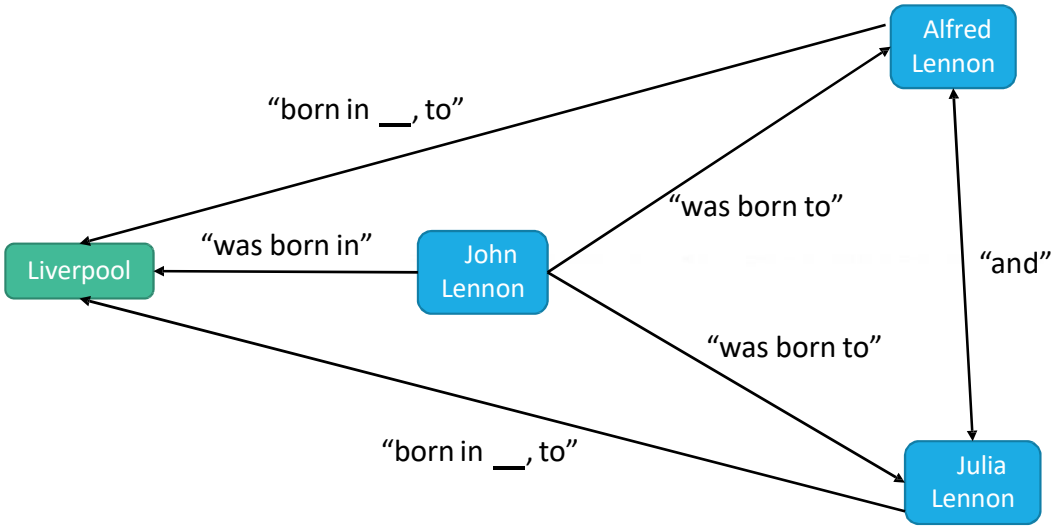


What is Information Extraction?



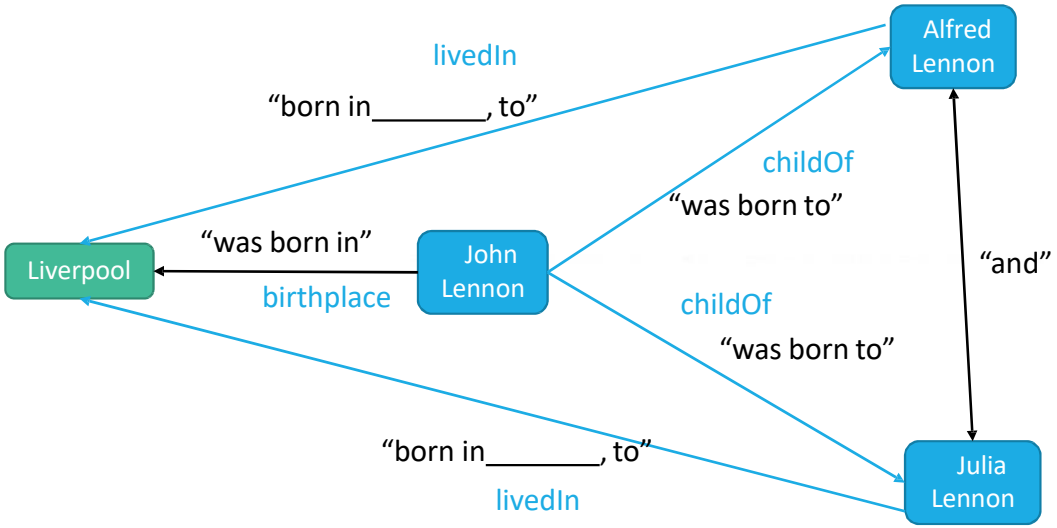
Relation Extraction From Text

John was born in Liverpool, to Julia and Alfred Lennon.

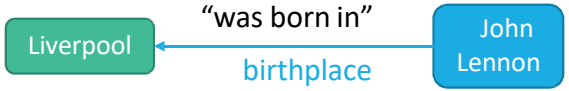


Relation Extraction From Text

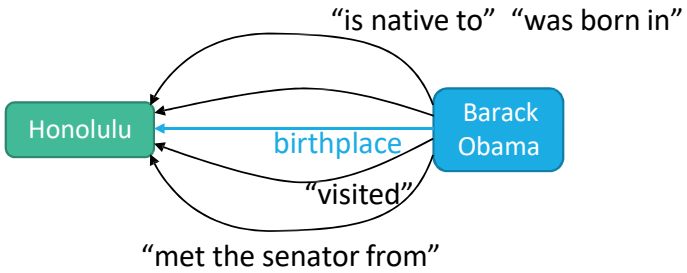
John was born in Liverpool, to Julia and Alfred Lennon.



“Distant” Supervision

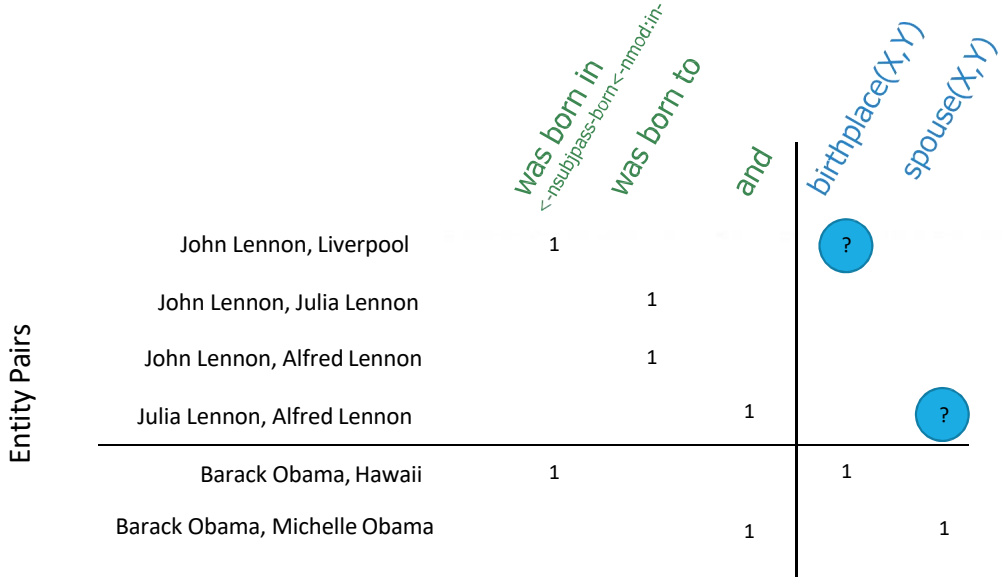


No direct supervision gives us this information. Supervised: Too expensive to label sentences Rule-based: Too much variety in language
Both only work for a small set of relations, i.e. 10s, not 100s

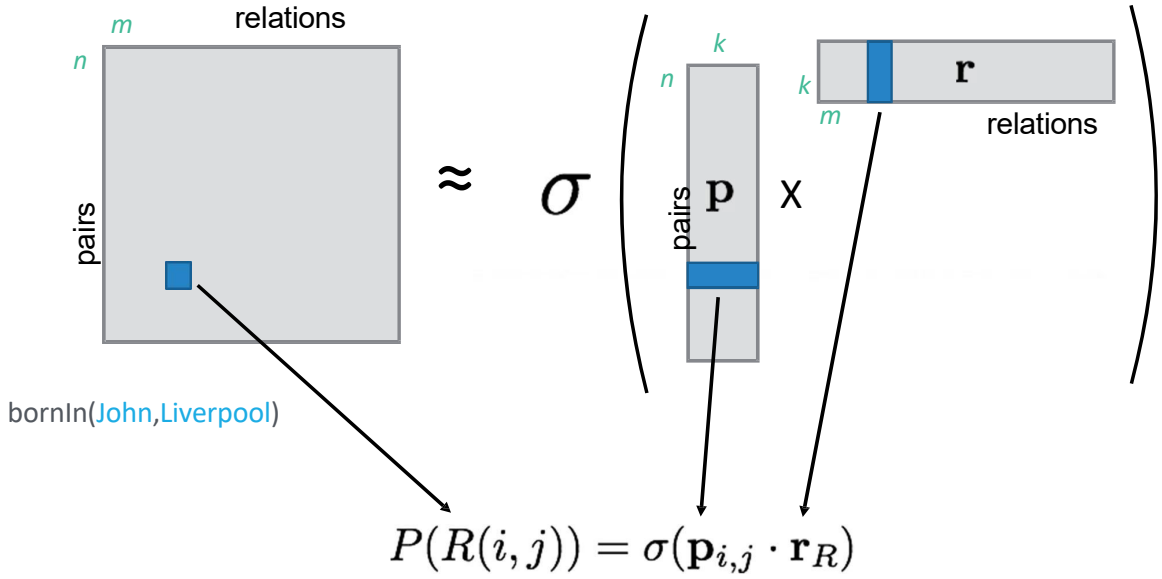


Relation Extraction as a Matrix

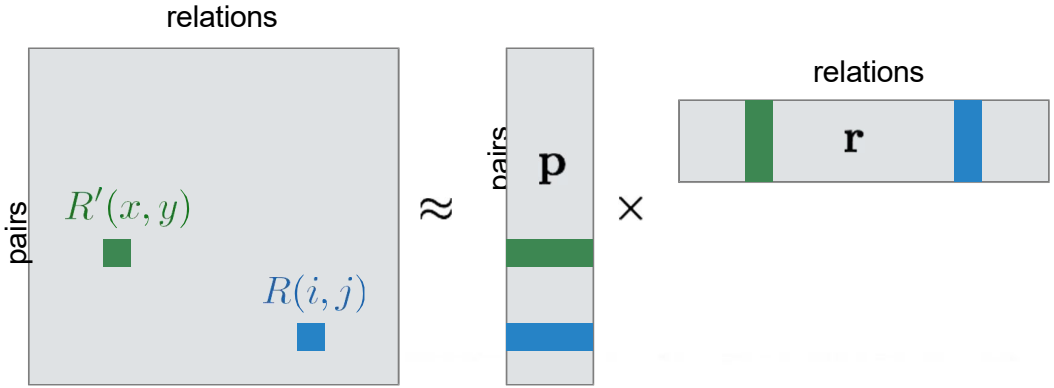
John was born in Liverpool, to Julia and Alfred Lennon.



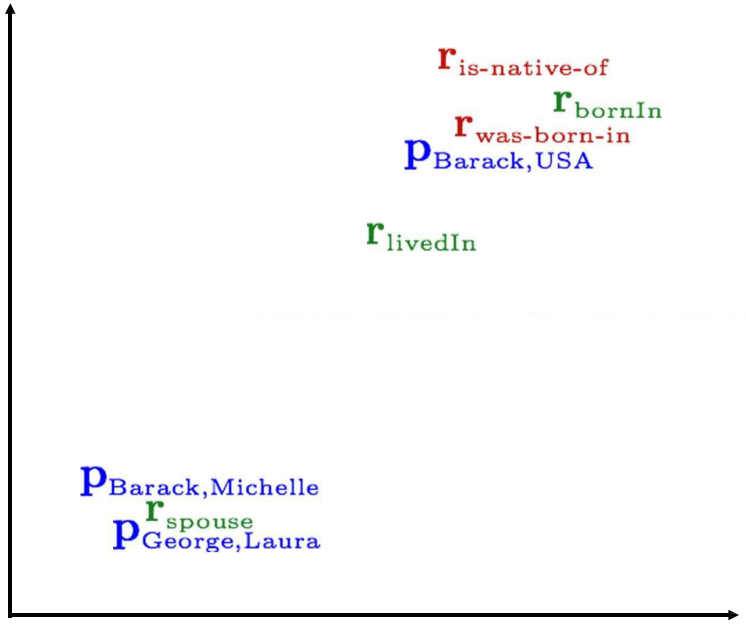
Matrix Factorization



Training: Stochastic Updates



Relation Embeddings



Embeddings ~ Logical Relations

Relation Embeddings, w

- Similar embedding for 2 relations denote they are paraphrases
  - is married to, spouseOf(X,Y), /person/spouse
- One embedding can be contained by another
  - w(topEmployeeOf) ⊂ w(employeeOf)
  - topEmployeeOf(X,Y) → employeeOf(X,Y)
- Can capture logical patterns, without needing to specify them!

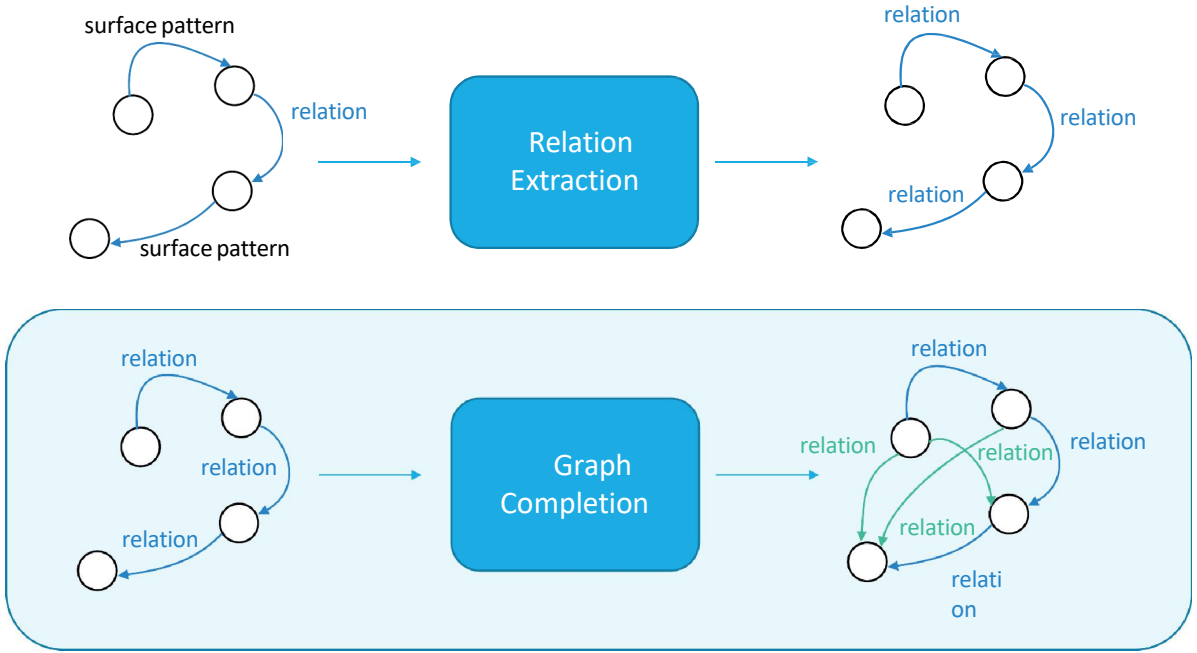
Entity Pair Embeddings, v

Similar entity pairs denote similar relations between them

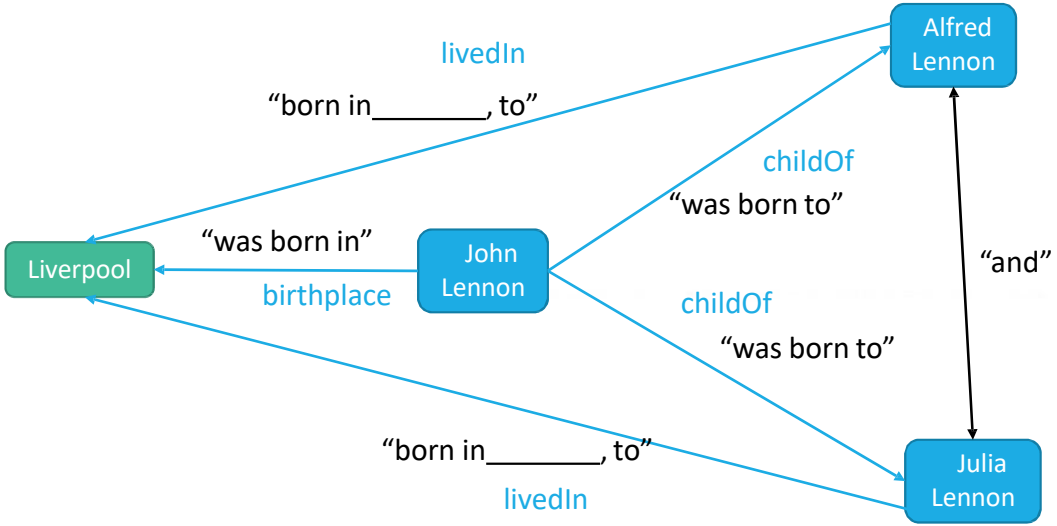
Entity pairs may describe multiple "relations"

independent foundedBy and employeeOf relations

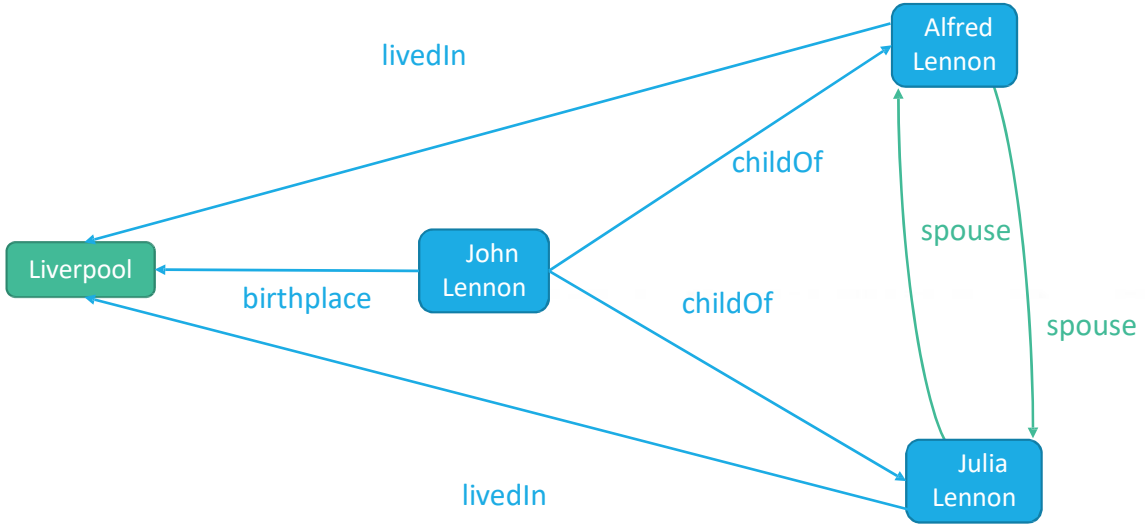
Two Related Tasks



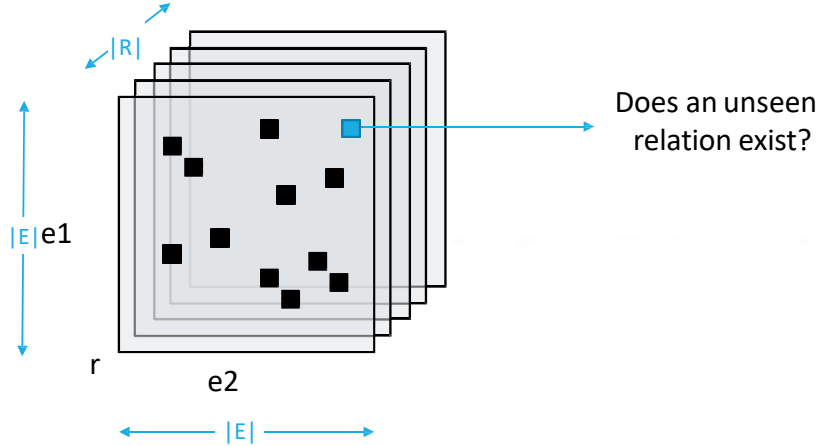
Graph Completion



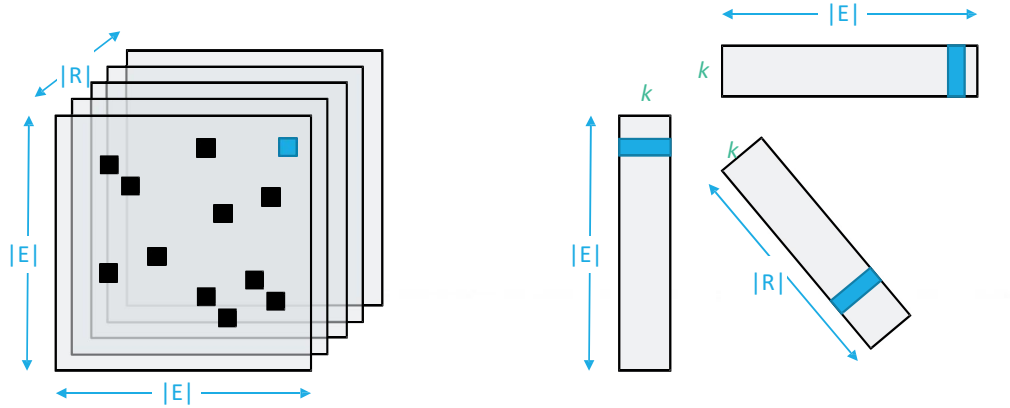
Graph Completion



Tensor Formulation of KG



Factorize that Tensor



$$S(r(a,b)) = f(\mathbf{v}_r, \mathbf{v}_a, \mathbf{v}_b)$$

Many Different Factorizations

CANDECOMP/PARAFAC-Decomposition

$$S\left(r(a,b)\right) = \sum_k R_{r,k} \cdot e_{a,k} \cdot e_{b,k}$$

Tucker2 and RESCAL Decompositions

$$S\left(r(a,b)\right) = \left(\mathbf{R}_r \times \mathbf{e}_a\right) \times \mathbf{e}_b$$

Model E

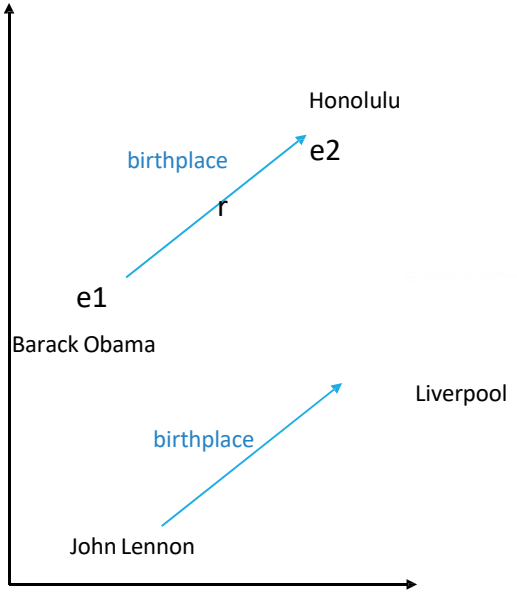
$$S\left(r(a,b)\right) = \mathbf{R}_{r,1} \cdot \mathbf{e}_a + \mathbf{R}_{r,2} \cdot \mathbf{e}_b$$

Holographic Embeddings

$$S(r(a,b)) = \mathbf{R}_r \times \left(\mathbf{e}_a \star \mathbf{e}_b\right)$$

Not tensor factorization (per se)

Translation Embeddings



TransE

$$S\left(r(a,b)\right) = -\|\mathbf{e}_a + \mathbf{R}_r - \mathbf{e}_b\|_2^2$$

TransH

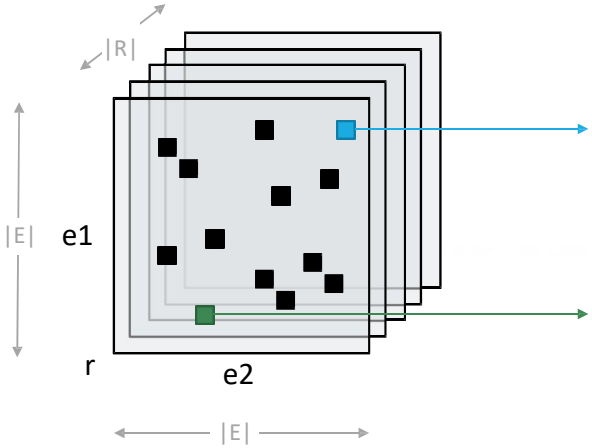
$$S\left(r(a,b)\right) = -\|\mathbf{e}_a^{\perp} + \mathbf{R}_r - \mathbf{e}_b^{\perp}\|_2^2$$

$$\mathbf{e}_a^{\perp} = \mathbf{e}_a - \mathbf{w}_r^T \mathbf{e}_a \mathbf{w}_r$$

TransR

$$S\left(r(a,b)\right) = -\|\mathbf{e}_a\mathbf{M}_r + \mathbf{R}_r - \mathbf{e}_b\mathbf{M}_r\|_2^2$$

Parameter Estimation



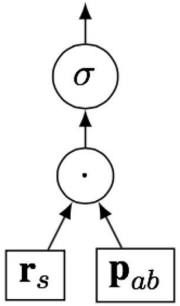
Observed cell: increase score

$$S\left(r(a,b)\right)$$

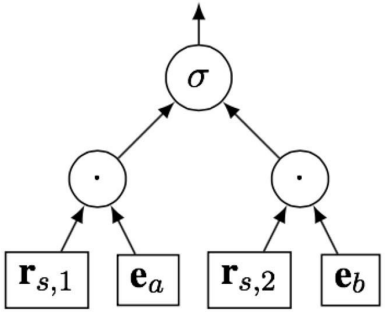
Unobserved cell: decrease score

$$S\left(r'(x,y)\right)$$

Matrix vs Tensor Factorization

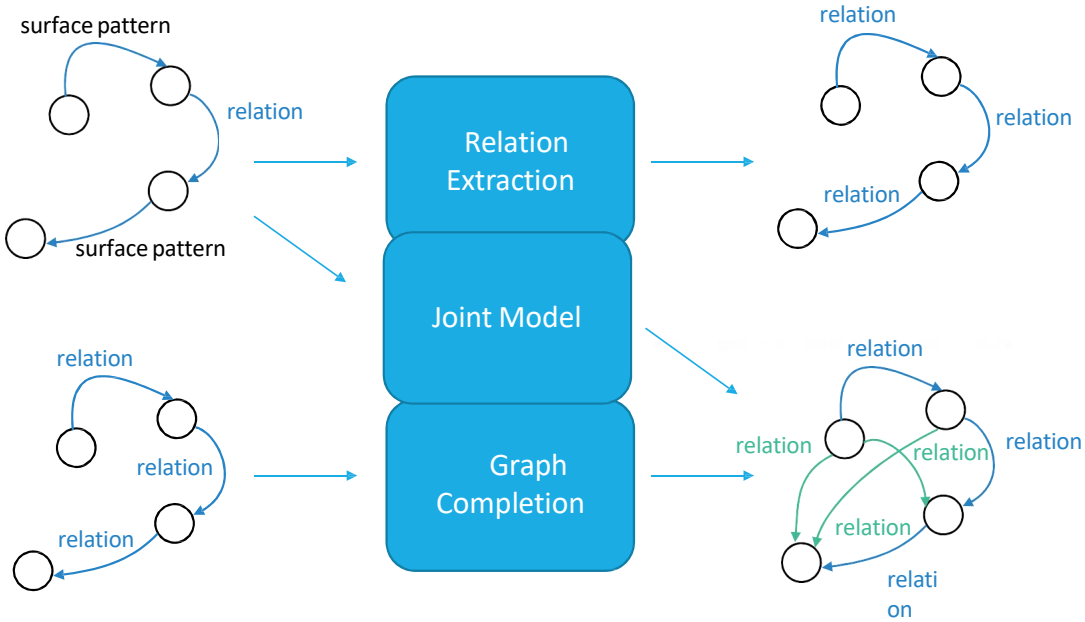


- Vectors for each entity pair
- Can only predict for entity pairs that appear in text together
- No sharing for same entity in different entity pairs



- Vectors for each entity
- Assume entity pairs are “low-rank”
  - But many relations are not!
  - Spouse: you can have only ~1
- Cannot learn pair specific information

Joint Extraction+Completion



Review: Embedding Techniques

Two Related Tasks:

- Relation Extraction from Text
- Graph (or Link) Completion

Relation Extraction:

- Matrix Factorization Approaches

Graph Completion:

- Tensor Factorization Approaches

Overview

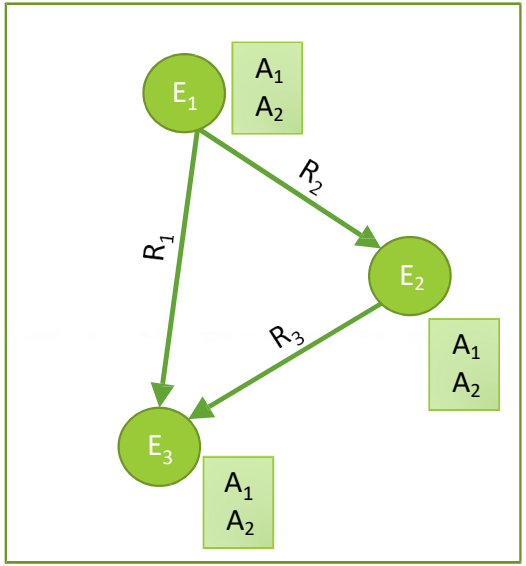
Why do we need Knowledgegraphs?

- Humans can explore large database in intuitive ways

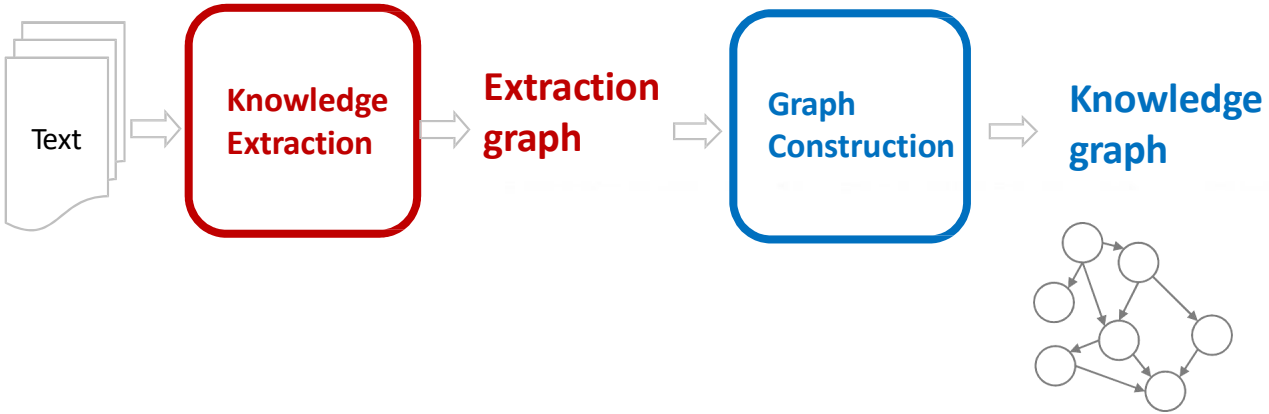
- AI agents get access to human common sense knowledge

Knowledge graph construction

- **Who** are the entities (nodes) in the graph?
- **What** are their attributes and types (labels)?
- **How** are they related (edges)?



Knowledge Graph Construction



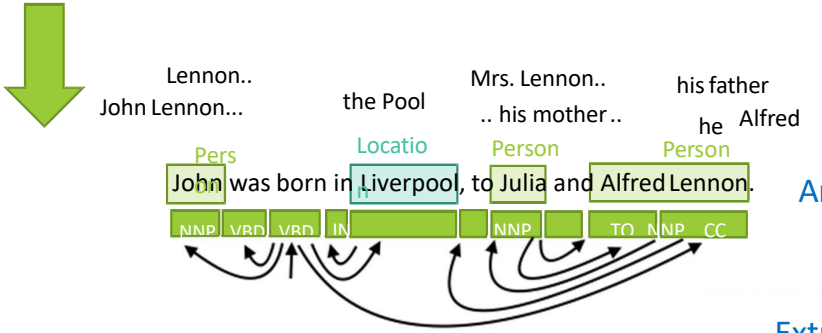
Two perspectives

	Extraction graph	Knowledge graph
Who are the entities? (nodes)	• Named Entity Recognition • Entity Coreference	• Entity Linking • Entity Resolution
What are their attributes? (labels)	• Entity Typing	• Collective classification
How are they related? (edges)	• Semantic role labeling • Relation Extraction	• Link prediction

Knowledge Extraction

John was born in Liverpool, to Julia and Alfred Lennon. Text

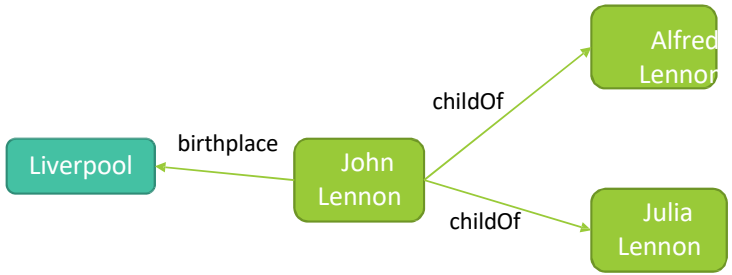
NLP



Annotated text

Information Extraction

Extraction graph



Information Extraction

Single extractor

Defining domain

Learning extractors

Scoring candidate facts

×

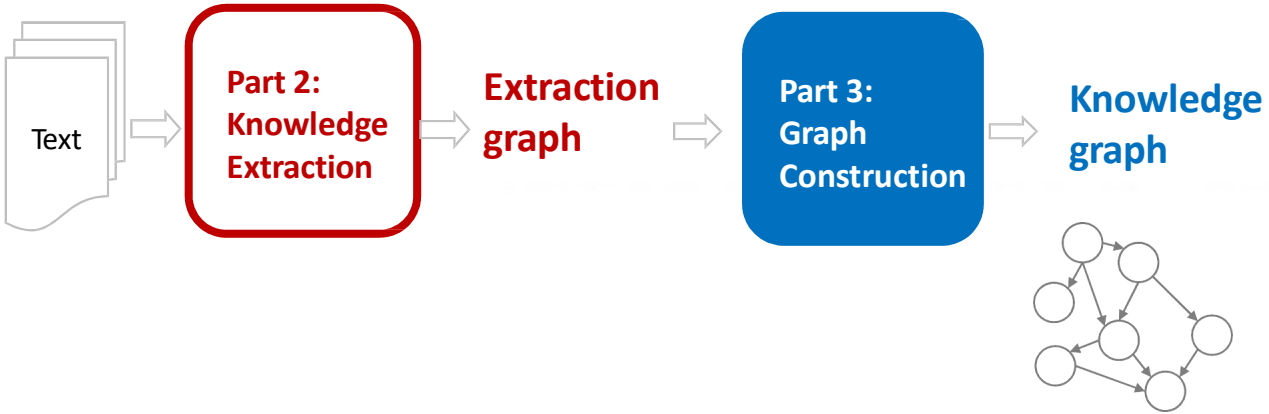
Supervised

Semi-supervised

Unsupervised

Fusing multiple extractors

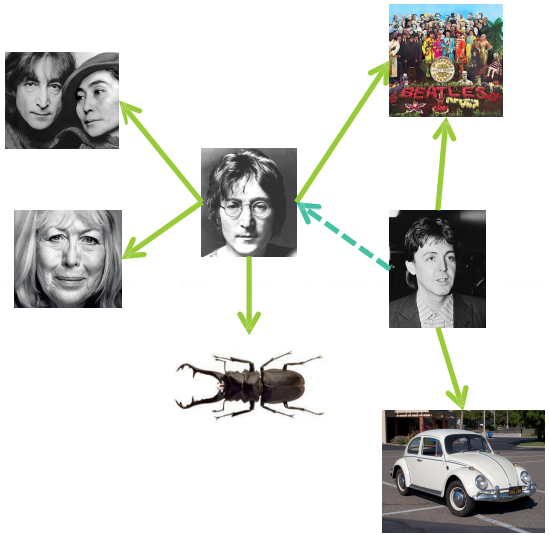
Knowledge Graph Construction



Issues with Extraction Graph

Extracted knowledge could be:

- ambiguous
- incomplete
- inconsistent



- Two approaches for KG construction

PROBABILISTIC MODELS

EMBEDDING BASED MODELS

- Two approaches for KG construction

PROBABILISTIC MODELS

EMBEDDING BASED MODELS

Relation Embeddings