Introduction to Knowledge Graphs

Computer Science Dept, Yonsei University

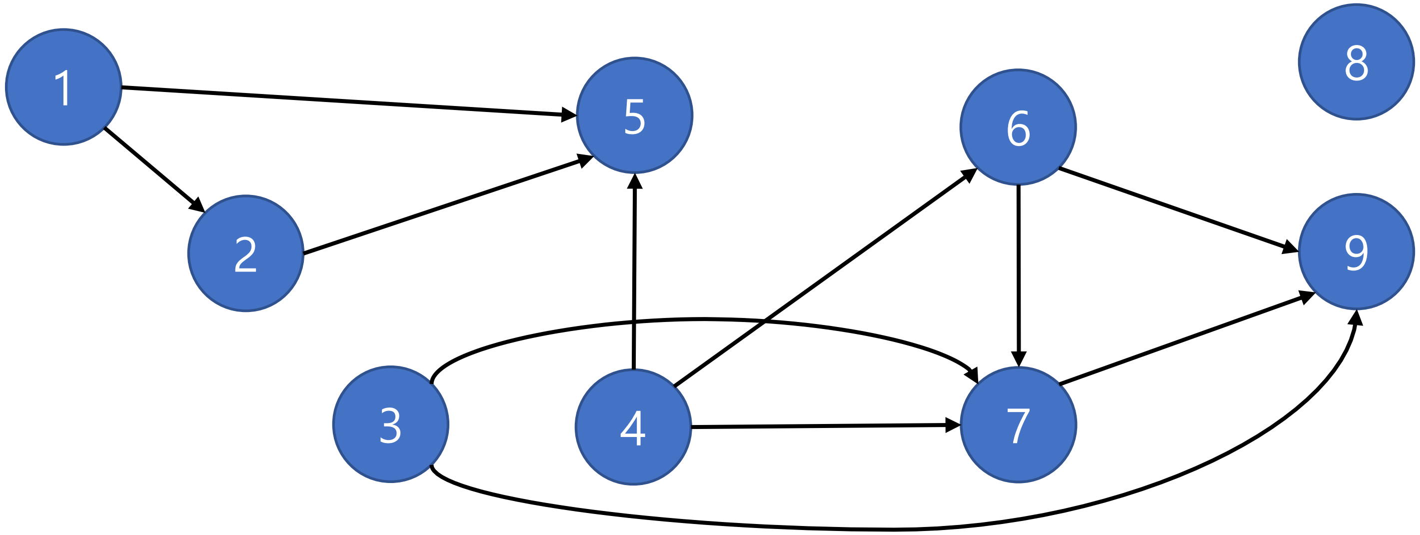
Seungone Kim

Referenced Papers

- Prerequisites
  - Knowledge Graph Embedding by Translating on Hyperplanes (AAAI 2014)
  - Learning Entity and Relation Embedding for Knowledge Graph Completion (AAAI 2015)
  - Commonsense Knowledge Base Completion (ACL 2016)
  - ConceptNet 5.5 : An Open Multilingual Graph of General Knowledge (AAAI 2017)
  - Open-World Knowledge Graph Completion (AAAI 2018)

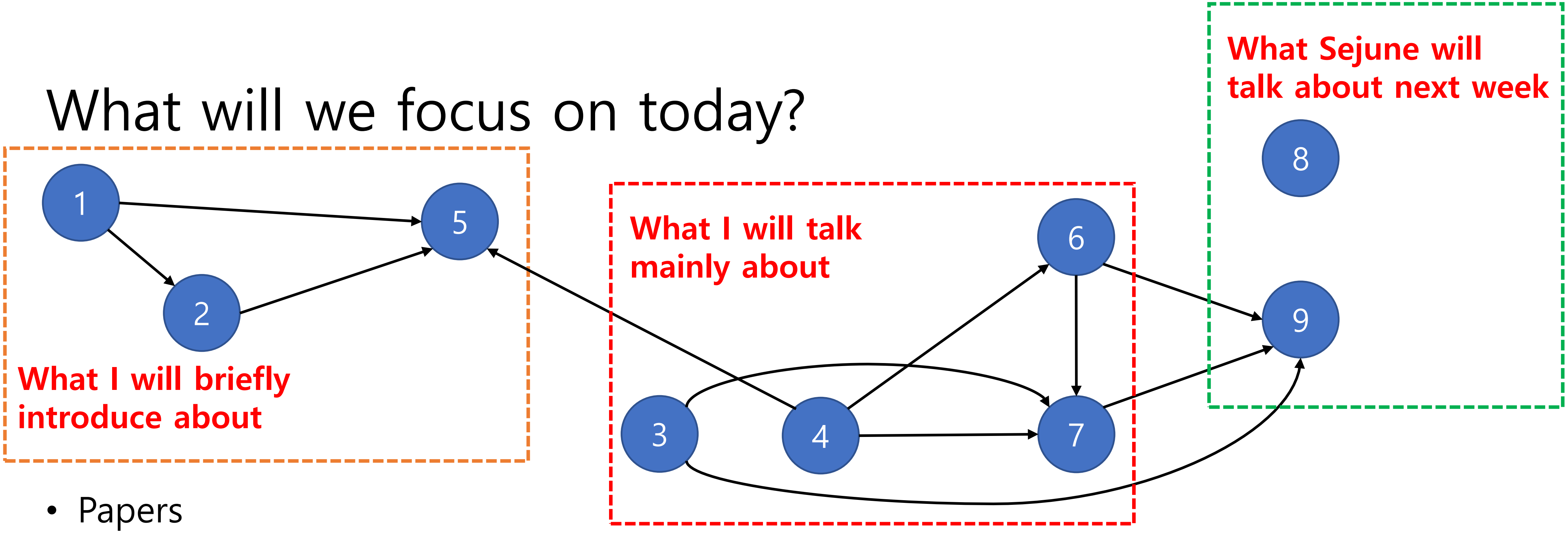
- Key Papers
  - ATOMIC : An Atlas of Machine Commonsense for If-Then Reasoning (AAAI 2019)
  - COMET : Commonsense Transformers for Automatic Knowledge Graph Construction (ACL 2019)
  - K-BERT : Enabling Language Representations with Knowledge Graph (AAAI 2020)
  - Dynamic Neuro-Symbolic Knowledge Graph Construction for Zero-shot Commonsense Question Answering (AAAI 2021)

Doesn’t Graph Representation give us more intuition?



- Papers
  - [1]Knowledge Graph Embedding by Translating on Hyperplanes (AAAI 2014)
  - [2]Learning Entity and Relation Embedding for Knowledge Graph Completion (AAAI 2015)
  - [3]Commonsense Knowledge Base Completion (ACL 2016)
  - [4]ConceptNet 5.5 : An Open Multilingual Graph of General Knowledge (AAAI 2017)
  - [5]Open-World Knowledge Graph Completion (AAAI 2018)
  - [6]ATOMIC : An Atlas of Machine Commonsense for If-Then Reasoning (AAAI 2019)
  - [7]COMET : Commonsense Transformers for Automatic Knowledge Graph Construction (ACL 2019)
  - [8]K-BERT : Enabling Language Representations with Knowledge Graph (AAAI 2020)
  - [9]Dynamic Neuro-Symbolic Knowledge Graph Construction for Zero-shot Commonsense Question Answering (AAAI 2021)

What will we focus on today?



- Papers
  - [1]Knowledge Graph Embedding by Translating on Hyperplanes (AAAI 2014)
  - [2]Learning Entity and Relation Embedding for Knowledge Graph Completion (AAAI 2015)
  - [3]Commonsense Knowledge Base Completion (ACL 2016)
  - [4]ConceptNet 5.5 : An Open Multilingual Graph of General Knowledge (AAAI 2017)
  - [5]Open-World Knowledge Graph Completion (AAAI 2018)
  - [6]ATOMIC : An Atlas of Machine Commonsense for If-Then Reasoning (AAAI 2019)
  - [7]COMET : Commonsense Transformers for Automatic Knowledge Graph Construction (ACL 2019)
  - [8]K-BERT : Enabling Language Representations with Knowledge Graph (AAAI 2020)
  - [9]Dynamic Neuro-Symbolic Knowledge Graph Construction for Zero-shot Commonsense Question Answering (AAAI 2021)

What are Knowledge Graphs / Bases

- Difference between KB / KG?
    - Here's an answer from StackOverFlow

1

Based on the related Wikipedia, a knowledge base (KB) is:

> a technology used to store complex structured and unstructured information used by a computer system. The initial use of the term was in connection with expert systems which were the first knowledge-based systems.

As there are different representation model for a KB, we can find different terminology in different domains. For example, in some AI articles, it's called **ontology**.

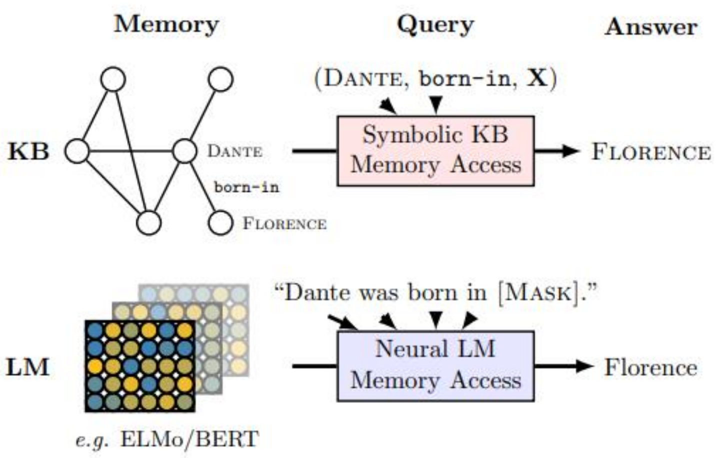
Knowledeg graph (KG) is another object model to KB realization which is introduced by Google for its search engine (as you have mentioned). Hence, KG is a specification of KB. You can find more information in the paper Knowledge Graphs, such as more history about the KG or a formal definition of that:

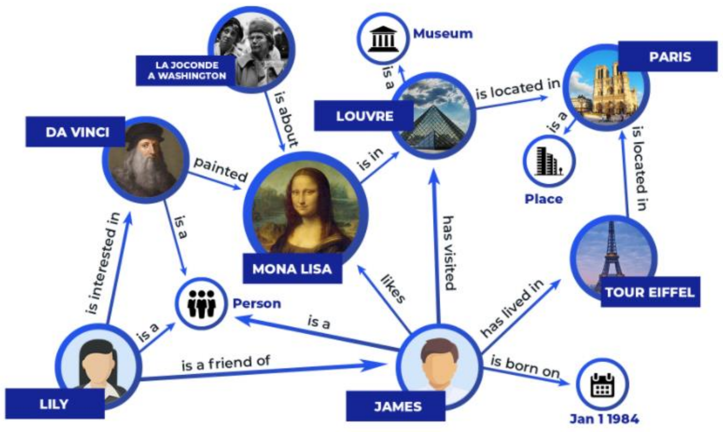
> knowledge graph is a graph of data intended to accumulate and convey knowledge of the real world, whose nodes represent entities of interest and whose edges represent relations between these entities.

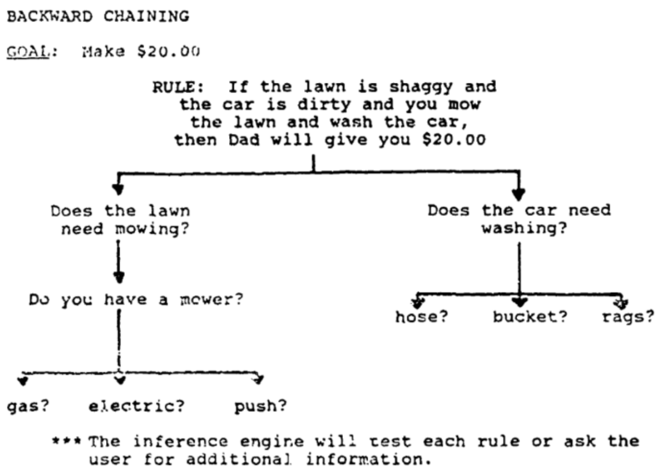
Moreover, you can find some articles about contextual KG (CKG) in the paper Learning Contextual Embeddings for Knowledge Graph Completion and KG^2: Learning to Reason Science Exam Questions with Contextual Knowledge Graph Embeddings.

What are Knowledge Graphs / Bases

- Difference between KB / KG?
  - Knowledge Base(KB) is a broader concept.
  - Knowledge Base is a technology to save KNOWLEDGE.
  - Knowledge Graph(KG) is a way to build a Knowledge Base.
  - Besides KGs, there might be other ways to build a Knowledge Base.
  - e.g. Language Models, Expert Systems …

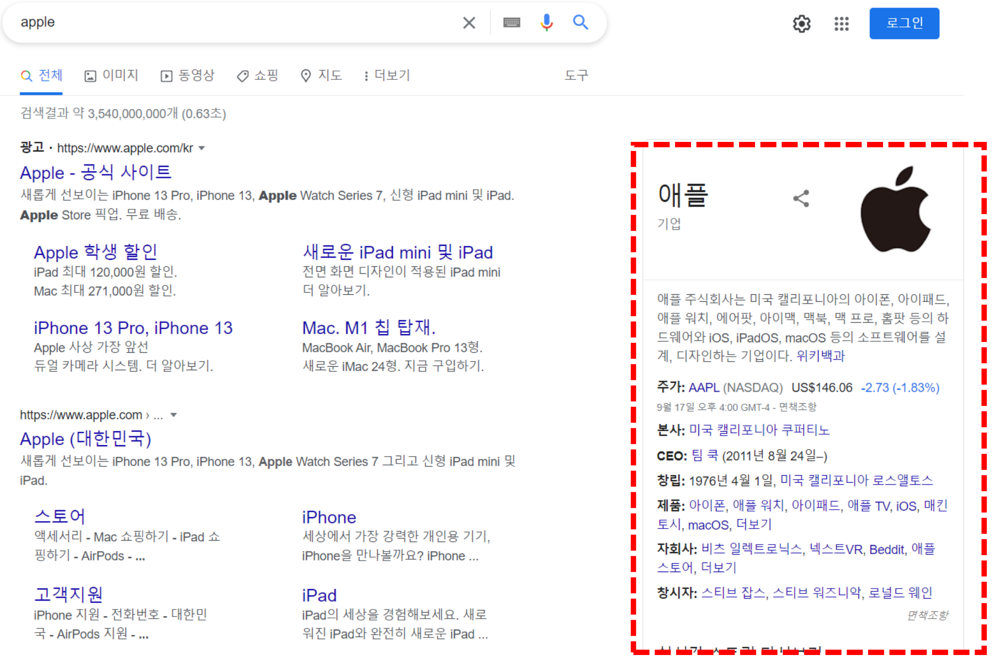




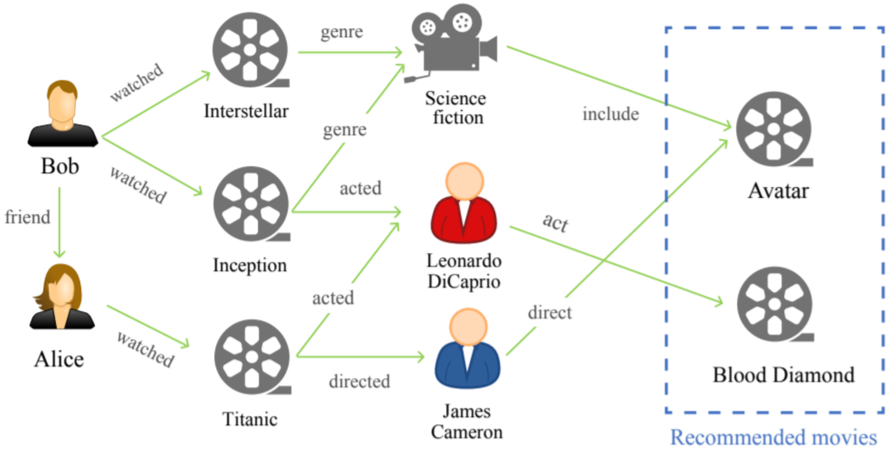


Why do we need Knowledge Graphs

- Advantages of Knowledge Graphs
  - Able to deal with Heterogenous Data
  - Models real-world information like a human brain does
  - Able to perform logical reasoning / query about complex information
  - Structured Representation compared to text data
  - Removes Redundancy



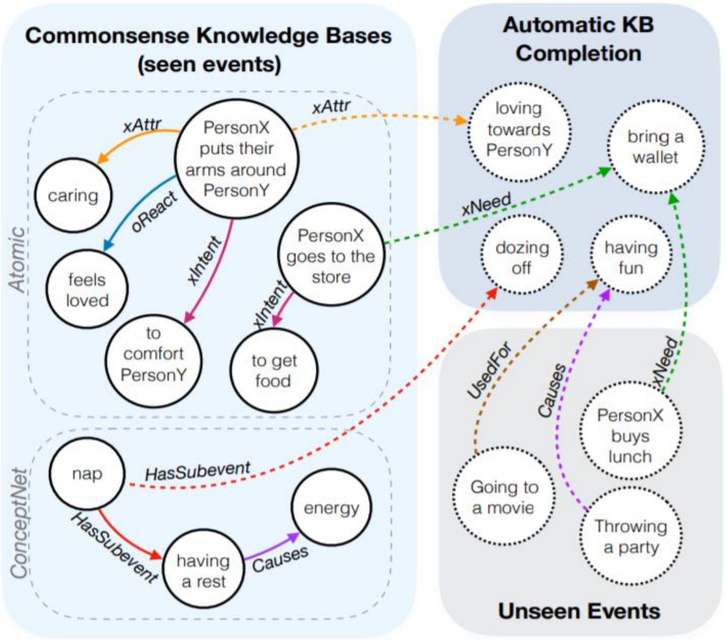
Query Types	Examples
One-hop Queries	Where did Hinton graduate?
Path Queries	Where did Turing Award winners graduate?
Conjunctive Queries	Where did Canadians with Turing Award graduate?
EPFO Queries	Where did Canadians with Turing Award or Nobel graduate?



Basic Introduction to Knowledge Graphs

- Knowledge Graph Completion
  - TransE => *Bordes, A. et al., "Translating embeddings for modeling multi-relational data" (NeurIPS 2013)*
  - TransH => *Wang, Zhen, et al., "Knowledge graph embedding by translating on hyperplanes." (AAAI 2014)*
  - TransR => *Lin, et al., "Learning entity and relation embeddings for knowledge graph completion" (AAAI 2015)*
  - DistMul => *Yang et al., "Embedding Entities and Relations for Learning and Inference in Knowledge Bases" (ICLR 2015)*
  - ComplEx => *Trouillon et al., 'Complex Embeddings for Simple Link Prediction" (ICLR 2016)*

- Examples of Knowledge Graphs
  - (Previous Methods) FreeBase, Wikidata, Dbpedia, YAGO, NELL
  - (Recent SOTA Methods; Commonsense KB) ConceptNet, Atomic



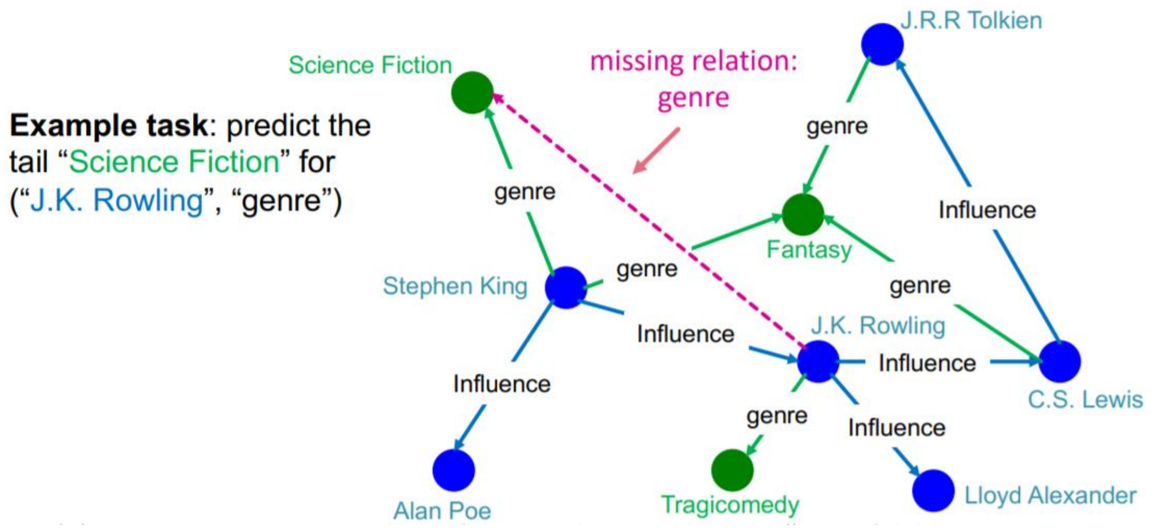
Freebase
- ~50 million **entities**
- ~38K **relation types**
- ~3 billion **facts/triples**

Freebase

93.8% of persons from Freebase have no place of birth and 78.5% have no nationality!

Knowledge Graph Completion

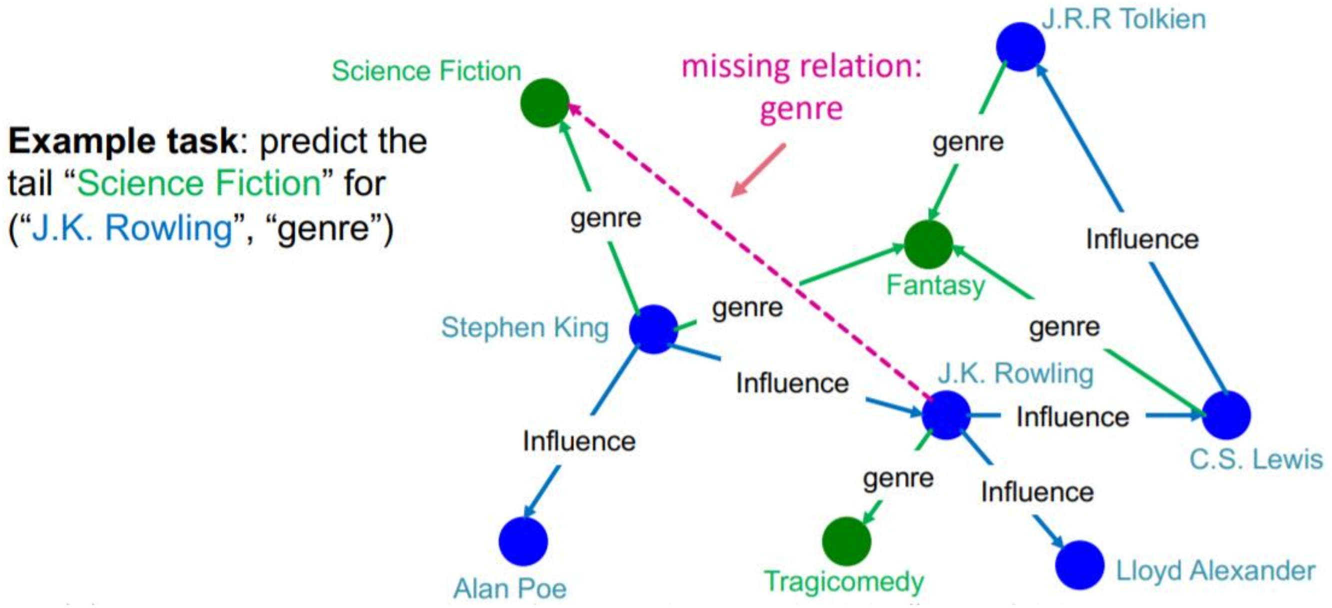
- Definition of KG Completion
  - For a given (head, relation), we predict missing tails.
  - In an enormous KG, there might be missing relations, which we ought to fill in.
  - KG Completion is slightly different from link prediction task.
  - KG Completion also differs from One-Hop Query(using KG).



- **KG completion:** Is link (h, r, t) in the KG?
- **One-hop query:** Is t an answer to query $(h, (r))$?

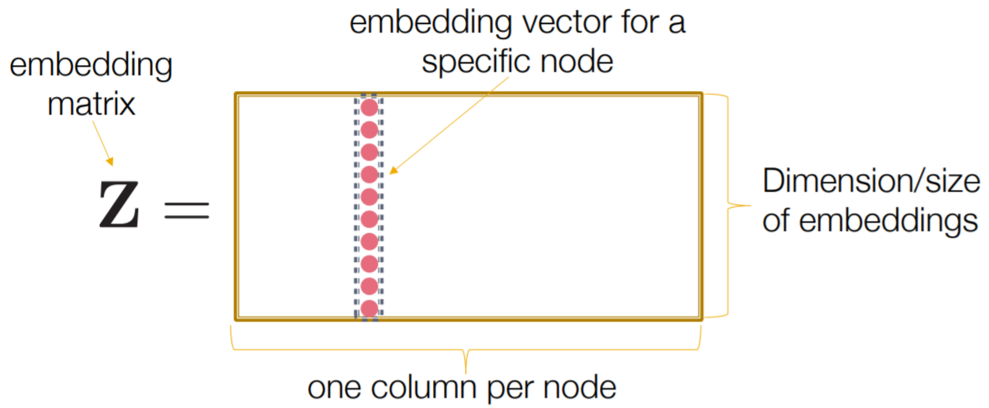
Knowledge Graph Completion

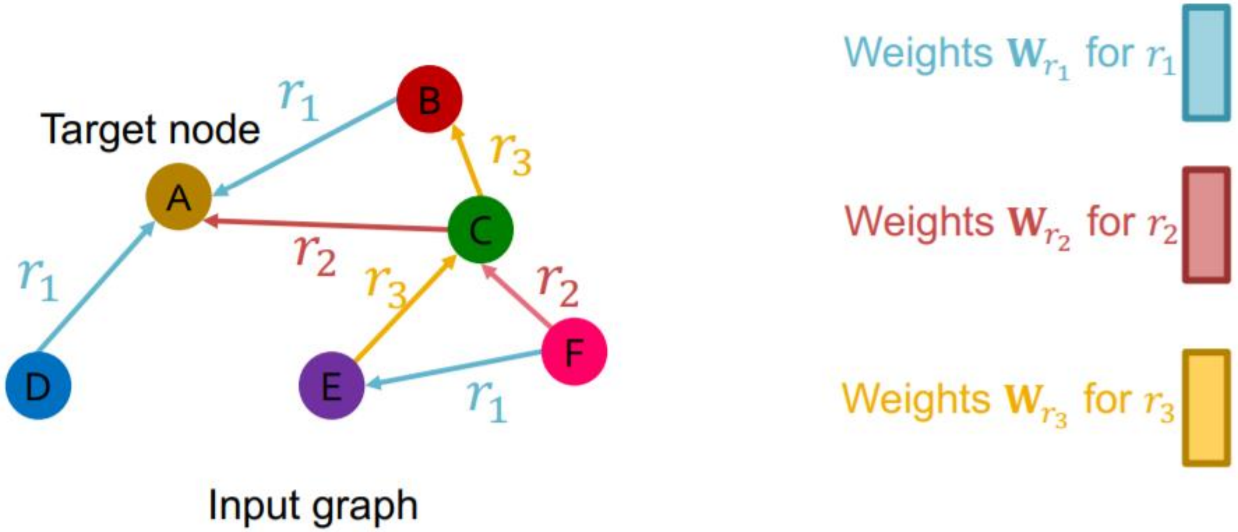
- Definition of KG Completion
  - Edges in KG are represented as triples (h,r,t).
  - That is, head(h) has relation(r) with tail(t).
  - Key idea is to model entities and relations in the embedding / vector space $\mathbb{R}^d$.
  - The goal is that the embedding of (h,r) should be close to the embedding of t.
  - We should define closeness.



Knowledge Graph Completion

- Embeddings in KG
  - Use different weights for different relation types.
  - Each node has an embedding vector.
  - In KG Completion, note that we do not learn a GNN!





Knowledge Graph Completion

- TransE (NeurIPS 2013)
  - For a triple (h,r,t), h, r, t ∈ $\mathbb{R}^d$, if given fact is true, h+r is similar to t, else h+r is not similar to t.

Algorithm 1 Learning TransE

input Training set $S = \{(h, \ell, t)\}$, entities and rel. sets E and L, margin γ, embeddings dim. k.

initialize $\boxed{\ell \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}}) \text{ for each } \ell \in L}$
$\boxed{\boldsymbol{\ell} \leftarrow \boldsymbol{\ell} / \|\boldsymbol{\ell}\| \text{ for each } \ell \in L}$
$\boxed{\mathbf{e} \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}}) \text{ for each entity } e \in E}$

Entities and relations are initialized uniformly, and normalized

loop
$\mathbf{e} \leftarrow \mathbf{e} / \|\mathbf{e}\|$ for each entity $e \in E$
$S_{batch} \leftarrow \text{sample}(S, b)$ // sample a minibatch of size b
$T_{batch} \leftarrow \emptyset$ // initialize the set of pairs of triplets
for $(h, \ell, t) \in S_{batch}$ **do**
$\boxed{(h', \ell, t') \leftarrow \text{sample}(S'_{(h,\ell,t)}) \text{ // sample a corrupted triplet}}$

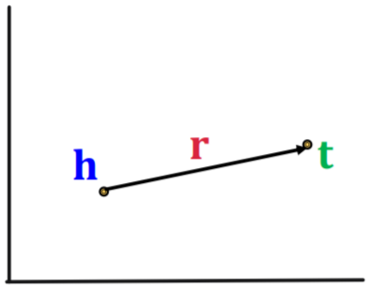
Negative sampling with triplet that does not appear in the KG

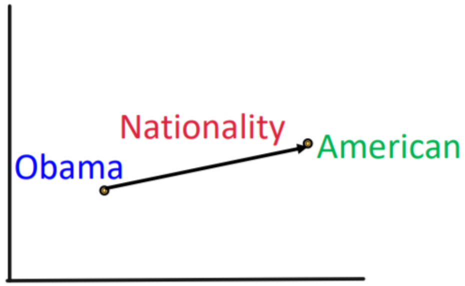
$T_{batch} \leftarrow T_{batch} \cup \{((h, \ell, t), (h', \ell, t'))\}$
end for
Update embeddings w.r.t. $\boxed{\sum_{((h,\ell,t),(h',\ell,t')) \in T_{batch}} \nabla \big[\gamma + \underbrace{d(\boldsymbol{h} + \boldsymbol{\ell}, \boldsymbol{t})}_{\text{positive sample}} - \underbrace{d(\boldsymbol{h'} + \boldsymbol{\ell}, \boldsymbol{t'})}_{\text{negative sample}}\big]_+}$
end loop

d represents distance (negative of score)

Contrastive loss: favors lower distance (or higher score) for valid triplets, high distance (or lower score) for corrupted ones

Scoring function: $f_r(h,t) = -||\mathbf{h} + \mathbf{r} - \mathbf{t}||$





Knowledge Graph Completion

- Is TransE the perfect algorithm for KG Completion?
  - Relations in heterogeneous KG have different properties.
  - If 형주 is roommate of 세준, 세준 is roommate of 형주.
  - If 규진 is mentor of 승원, 승원 is mentee of 규진, not mentor of 규진.
  - 승원's mother's husband is 승원's father

- **Symmetric (Antisymmetric) Relations:**

  $$r(h,t) \Rightarrow r(t,h) \;\; (r(h,t) \Rightarrow \neg r(t,h)) \quad \forall h,t$$

  - **Example**:
    - Symmetric: Family, Roommate
    - Antisymmetric: Hypernym
- **Inverse Relations:**

  $$r_2(h,t) \Rightarrow r_1(t,h)$$

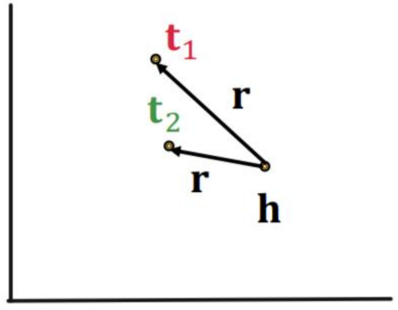
  - **Example** : (Advisor, Advisee)
- **Composition (Transitive) Relations:**

  $$r_1(x,y) \wedge r_2(y,z) \Rightarrow r_3(x,z) \quad \forall x,y,z$$

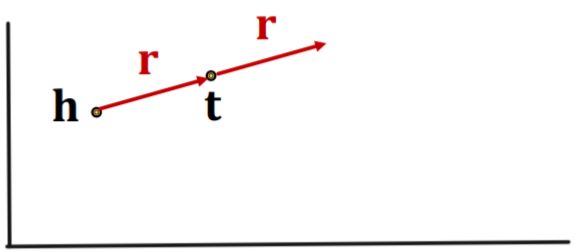
  - **Example**: My mother's husband is my father.
- **1-to-N relations:**

  $r(h,t_1), r(h,t_2), \ldots, r(h,t_n)$ are all True.

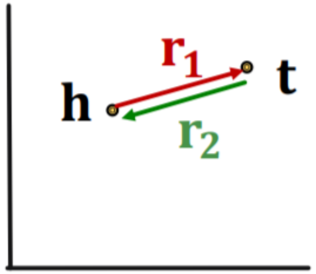
  - **Example**: r is "StudentsOf"



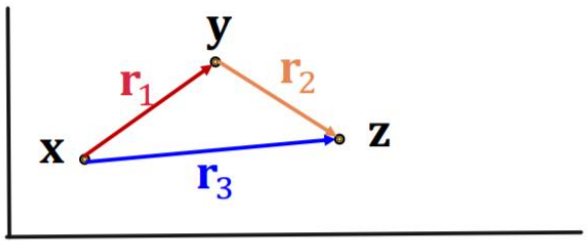
1-to-N



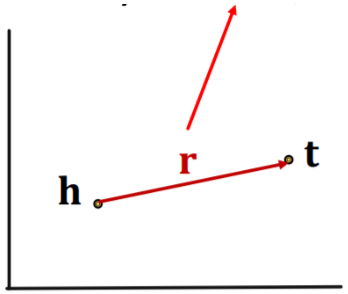
Antisymmetric



Inverse



Composition (Transitive)



Symmetric

Knowledge Graph Completion

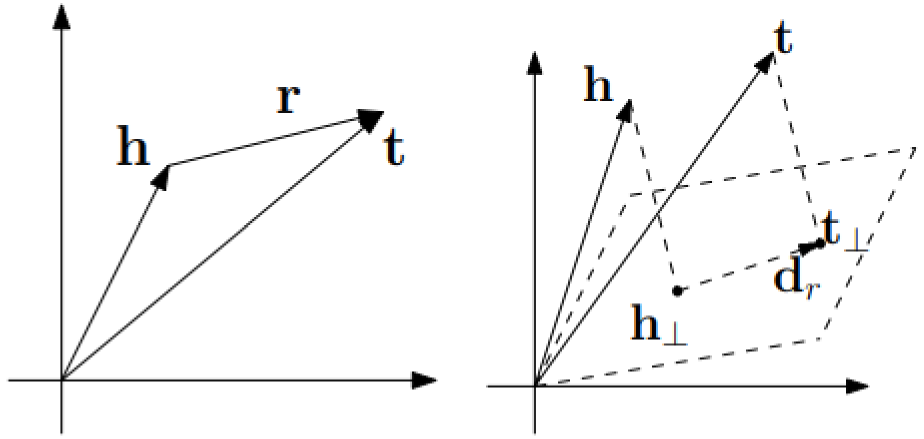
- TransH (AAAI 2014)
  - For a triple (h,r,t), h, t $\in \mathbb{R}^d$, r, $h_\perp, t_\perp \in \mathbb{R}^k$ (Each relation has it's own space)
  - if given fact is true, $h_\perp + r$ is similar to $t_\perp$, else $h_\perp + r$ is not similar to $t_\perp$.
  - Using a relation-specific space, we can add more expressivity.
  - However, given a relation, each entity is projected to the same spot on the relation space.
  - However, it cannot model composition relations.

$$\|\mathbf{h}_\perp + \mathbf{d}_r - \mathbf{t}_\perp\|_2^2$$

Score function of TransH

$$\mathbf{h}_\perp = \mathbf{h} - \mathbf{w}_r^\top \mathbf{h} \mathbf{w}_r, \quad \mathbf{t}_\perp = \mathbf{t} - \mathbf{w}_r^\top \mathbf{t} \mathbf{w}_r.$$

TransH (this paper)	$\|(\mathbf{h} - \mathbf{w}_r^\top \mathbf{h} \mathbf{w}_r) + \mathbf{d}_r - (\mathbf{t} - \mathbf{w}_r^\top \mathbf{t} \mathbf{w}_r)\|_2^2$ $\mathbf{w}_r, \mathbf{d}_r \in \mathbb{R}^k$	$O(n_e k + 2 n_r k)$

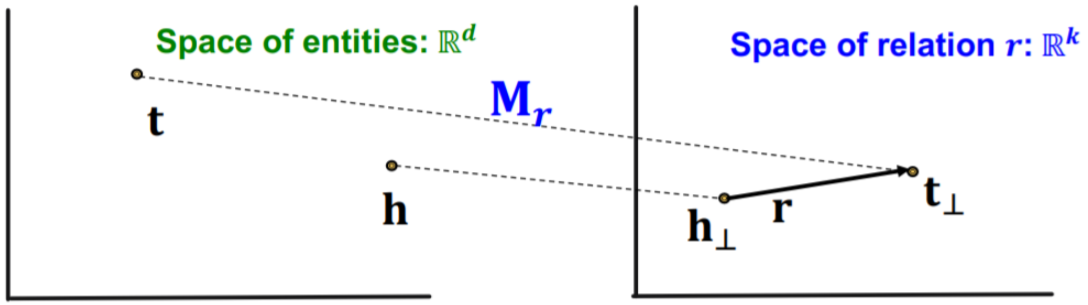


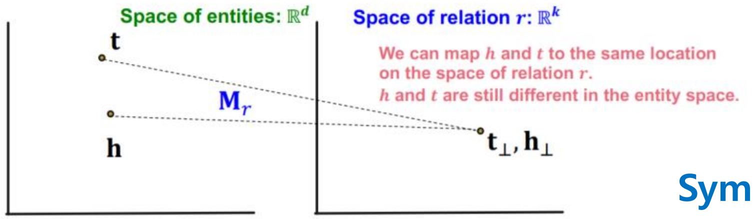
(a) TransE (b) TransH

Knowledge Graph Completion

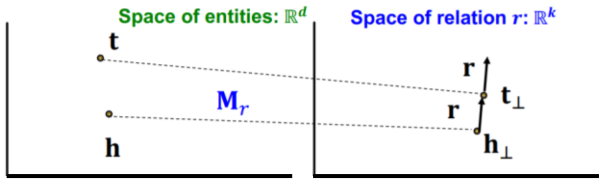
- TransR (AAAI 2015)
  - For a triple (h,r,t), h, t ∈ $\mathbb{R}^d$, r ∈ $\mathbb{R}^k$ with $M_r \in \mathbb{R}^{k \times d}$ as projection matrix (Each relation has it's own space)
  - if given fact is true, $M_r h + r$ is similar to $M_r t$, else $M_r h + r$ is not similar to $M_r t$.
  - Using a separate vector space for entities and relations, we can also express Symmetric and 1-N relations.
  - However, it cannot model composition relations.

- **Score function**: $f_r(h,t) = -||\mathbf{h}_\perp + \mathbf{r} - \mathbf{t}_\perp||$

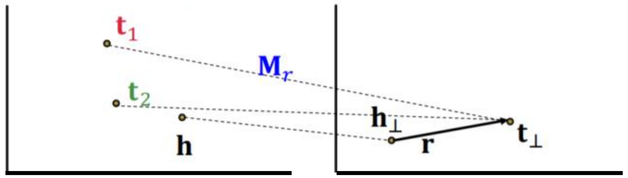




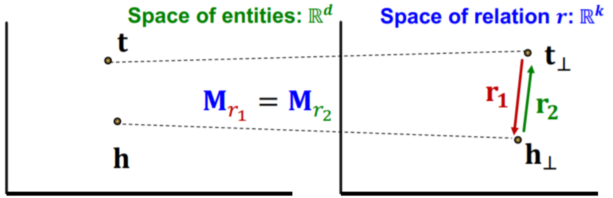
Symmetric



Antisymmetric



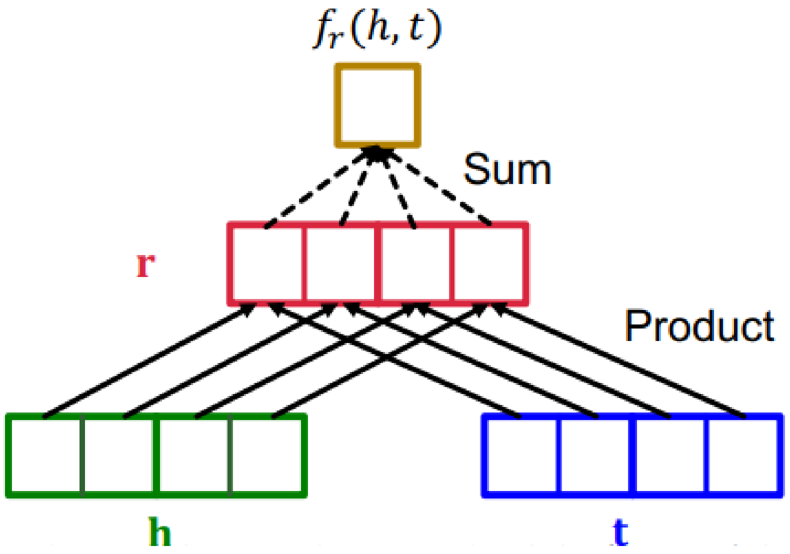
1-to-N



Inverse

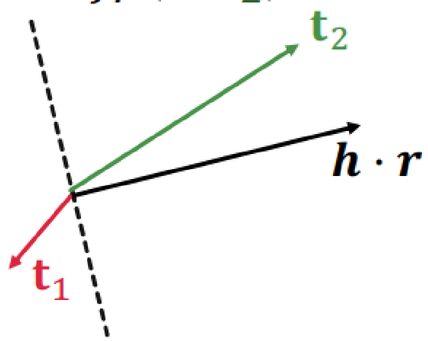
Knowledge Graph Completion

- DistMult (ICLR 2015)
  - For a triple (h,r,t), h, r, t $\in \mathbb{R}^d$, if given fact is true, $\sum_i h_i * r_i * t_i$ is a large value, else $\sum_i h_i * r_i * t_i$ is a small value.
  - The intuition of the score function can be viewed as a cosine similarity between h*r and t.
  - However, DistMult cannot model compositional relations.
  - Also, DistMult cannot model antisymmetric relations because $r(h,t)$ and $r(t,h)$ have the same score.
  - Also, DistMult cannot model inverse relations because $r(h,t)$ and $r(t,h)$ have the same score.



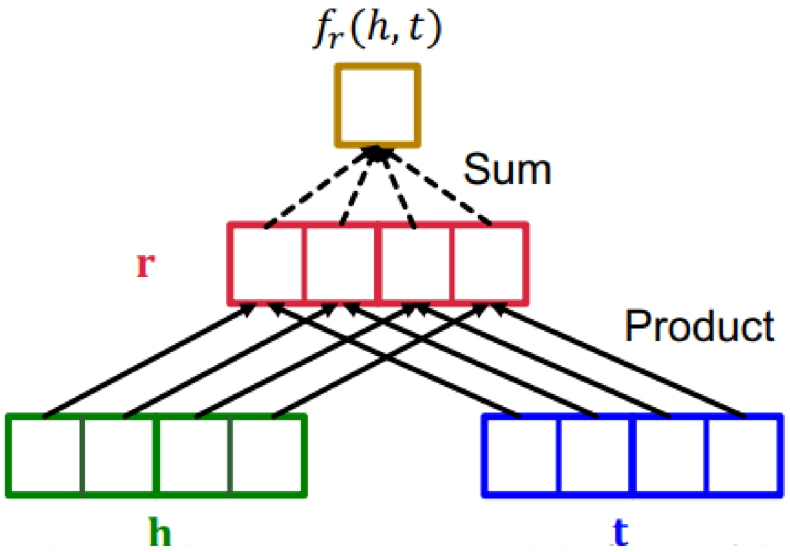
Score function: $f_r(h,t) =< \mathbf{h}, \mathbf{r}, \mathbf{t} > = \sum_i \mathbf{h}_i \cdot \mathbf{r}_i \cdot \mathbf{t}_i$

$f_r(h,t_1) < 0, \qquad f_r(h,t_2) > 0$



Knowledge Graph Completion

- ComplEx (ICLR 2016)
  - For a triple (h,r,t), h, r, t ∈ $\mathbb{R}^d$, if given fact is true, $\mathrm{Re}(\sum_i h_i * r_i * t_i)$ is a large value, else $\mathrm{Re}(\sum_i h_i * r_i * t_i)$ is a small value.
  - The difference with DistMult is that ComplEx uses a Complex Vector Space.
  - Using Complex Vector Space, results are improved in many situations that involve non-compositionality.
  - ComplEx can model antisymmetric, symmetric, inverse relations due to using complex conjugate.



Score function: $f_r(h,t) =< \mathbf{h}, \mathbf{r}, \mathbf{t} > = \sum_i \mathbf{h}_i \cdot \mathbf{r}_i \cdot \mathbf{t}_i$

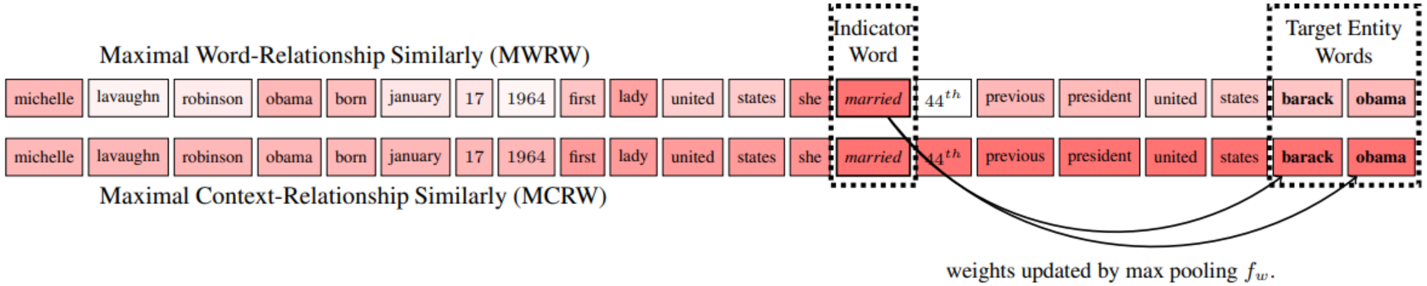
Knowledge Graph Completion

- Overview of all the methods for KG Completion

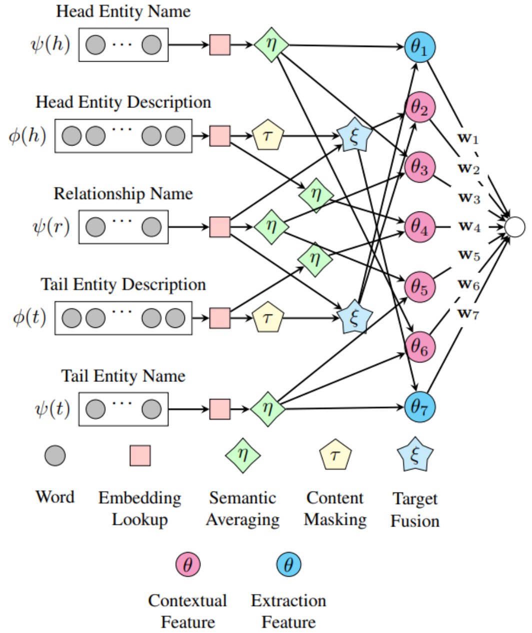
Model	Score	Embedding	Sym.	Antisym.	Inv.	Compos.	1-to-N
TransE	$-\|\mathbf{h}+\mathbf{r}-\mathbf{t}\|$	$\mathbf{h},\mathbf{t},\mathbf{r}\in\mathbb{R}^k$	✗	✓	✓	✓	✗
TransR	$-\|\mathbf{W}_r\mathbf{h}+\mathbf{r}-\mathbf{W}_r\mathbf{t}\|$	$\mathbf{h},\mathbf{t},\mathbf{r}\in\mathbb{R}^k$, $\mathbf{W}_r\in\mathbb{R}^k$	✓	✓	✓	✗	✓
DistMult	$<\mathbf{h},\mathbf{r},\mathbf{t}>$	$\mathbf{h},\mathbf{t},\mathbf{r}\in\mathbb{R}^k$	✓	✗	✗	✗	✓
ComplEx	$\mathrm{Re}(<\mathbf{h},\mathbf{r},\bar{\mathbf{t}}>)$	$\mathbf{h},\mathbf{t},\mathbf{r}\in\mathbb{C}^k$	✓	✓	✓	✗	✓

Open World Knowledge Graph

- Open-World Knowledge Graph Completion (AAAI 2018)
  - *What is the difference between Closed World KG and Open World KG?*
  - Most Knowledge Base Completion(KBC) focuses on situation where KG is fixed, and new **entities** cannot be easily added.
  - Open World Knowledge Base relax the assumption above.
  - Goal of the paper was to build an open world KGC Model that can learn embeddings of entity's name to connect unseen entities in KG.
  - In Open World KBC, while adding a new entity to the KG, the embedding of the existing entities that are connected should also be updated.
  - Given tuple $< t_1,\ r,\ ? >$, and an entity description, ConMask extract embeddings and compare the similarity score between target entity candidates in the KG.



Example Task: Complete triple ⟨Ameen Sayani, residence, ? ⟩, where Ameen Sayani is absent from the KG.
Snippet of Entity Description: "... ***Ameen Sayani*** *was introduced to All India Radio,* ***Bombay****, by his brother Hamid Sayani. Ameen participated in English programmes there for ten years ...*".
Predicted Target Entity: Mumbai.



Basic Introduction to Knowledge Graphs

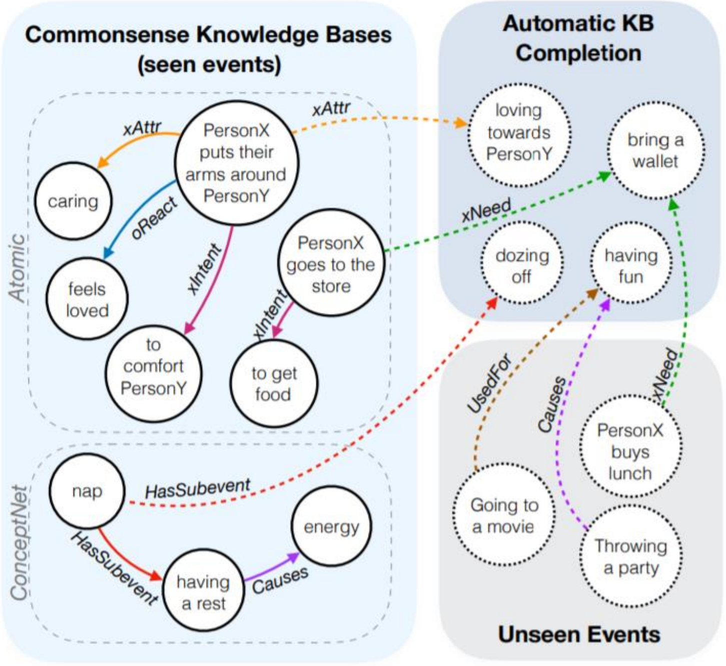
- Knowledge Graph Completion
  - TransE => *Bordes, A. et al., "Translating embeddings for modeling multi-relational data" (NeurIPS 2013)*
  - TransH => *Wang, Zhen, et al., "Knowledge graph embedding by translating on hyperplanes." (AAAI 2014)*
  - TransR => *Lin, et al., "Learning entity and relation embeddings for knowledge graph completion" (AAAI 2015)*
  - DistMul => *Yang et al., "Embedding Entities and Relations for Learning and Inference in Knowledge Bases" (ICLR 2015)*
  - ComplEx => *Trouillon et al., 'Complex Embeddings for Simple Link Prediction" (ICLR 2016)*

- Examples of Knowledge Graphs
  - (Previous Methods) FreeBase, Wikidata, Dbpedia, YAGO, NELL
  - (Recent SOTA Methods; Commonsense KB) ConceptNet, Atomic

- **Freebase**
  - ~50 million **entities**
  - ~38K **relation types**
  - ~3 billion **facts/triples**

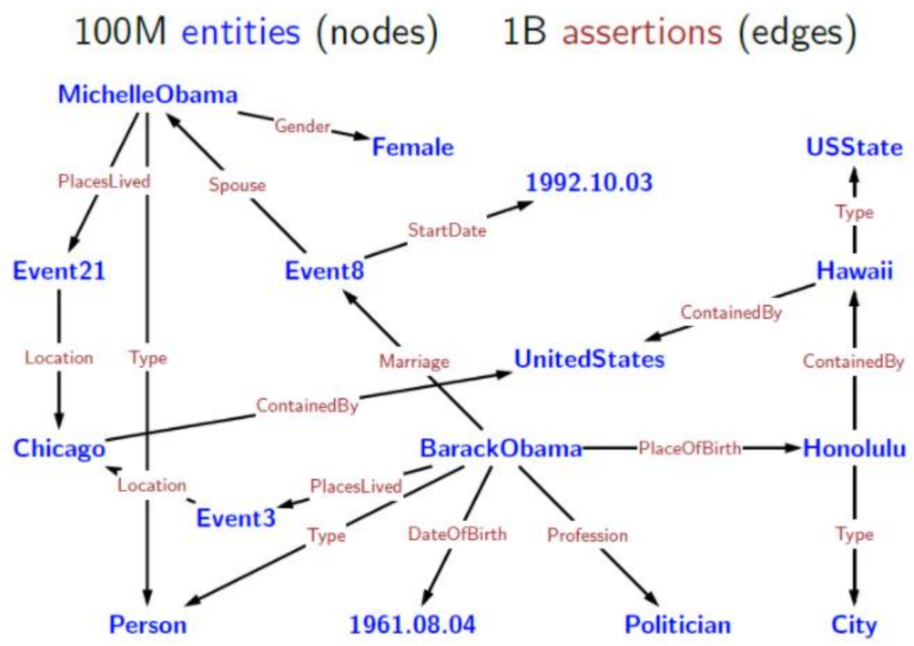
Freebase

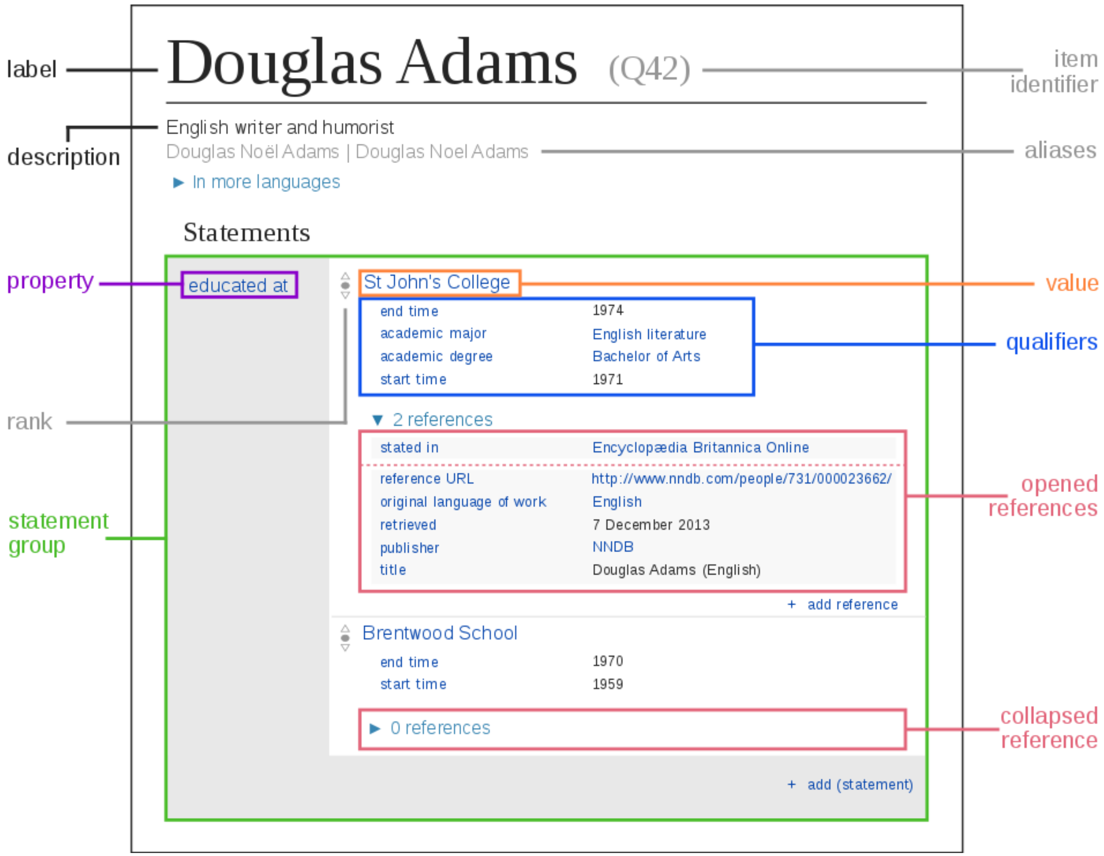
93.8% of persons from Freebase have no place of birth and 78.5% have no nationality!



Examples of KGs

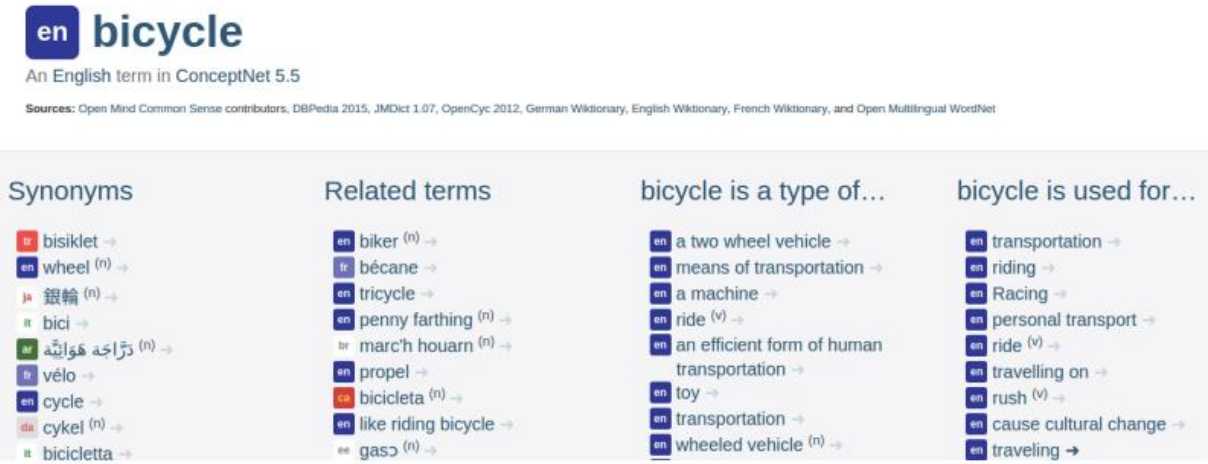
- Previous Famous KGs
  - FreeBase, Wikidata, Dbpedia, YAGO, NELL

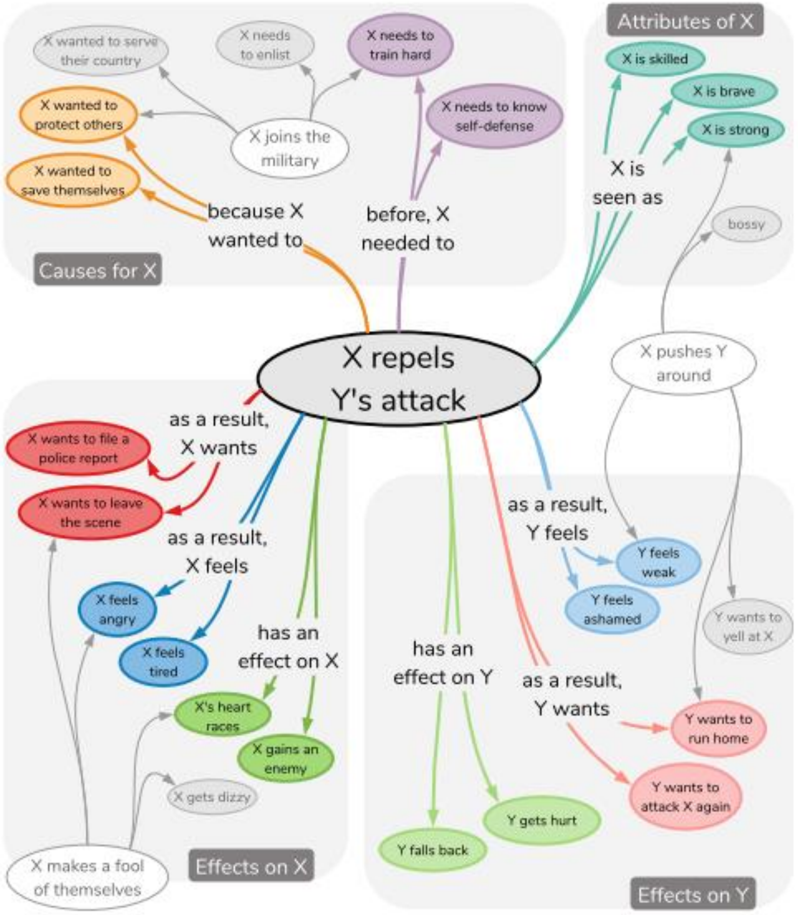




Examples of KGs

- SOTA Knowledge Graphs
  - Commonsense Knowledge Graph
  - ConceptNet 5.5, ATOMIC





Commonsense Knowledge Graph

- Commonsense Knowledge Base Completion (ACL 2016)
  - *What is the difference between previous KG and Commonsense KG?*
  - Most Knowledge Base Completion(KBC) focuses on knowledge bases like Freebase that relate **entities** drawn from a fixed set.
  - Commonsense Knowledge Base define relations between an unbounded set of phrases.
  - Commonsense Knowledge is rarely expressed explicitly in textual corpora.
  - Goal of the paper was to develop a parametric model that can provide a confidence score for new, unseen tuples.
  - Assume that embeddings for words is provided.
  - Given tuple $< t_1,\ r,\ t_2 >$, use negative sampling to get t_2 and train a Bilinear Model or a DNN to get a predicted score.

relation	right term	conf.
MotivatedByGoal	relax	3.3
UsedFor	relaxation	2.6
MotivatedByGoal	your muscle be sore	2.3
HasPrerequisite	go to spa	2.0
Causes	get pruny skin	1.6
HasPrerequisite	change into swim suit	1.6

Table 1: ConceptNet tuples with left term "soak in hotspring"; final column is confidence score.

t_1, R, t_2	score
bus, IsA, *public transportation*	0.95
bus, IsA, *public transit*	0.90
bus, IsA, *mass transit*	0.79
bus, AtLocation, *downtown area*	0.98
bus, AtLocation, *subway station*	0.98
bus, AtLocation, *city center*	0.94
bus, CapableOf, *low cost*	0.72
bus, CapableOf, *local service*	0.65
bus, CapableOf, *train service*	0.63

After nine years of primary school, students can go to the high school or to an educational institution.

t_1, R, t_2	score
school, HasProperty, *educational*	0.89
school, IsA, *educational institution*	0.80
school, IsA, *institution*	0.78
school, HasProperty, *high*	0.77
high school, IsA, *institution*	0.71

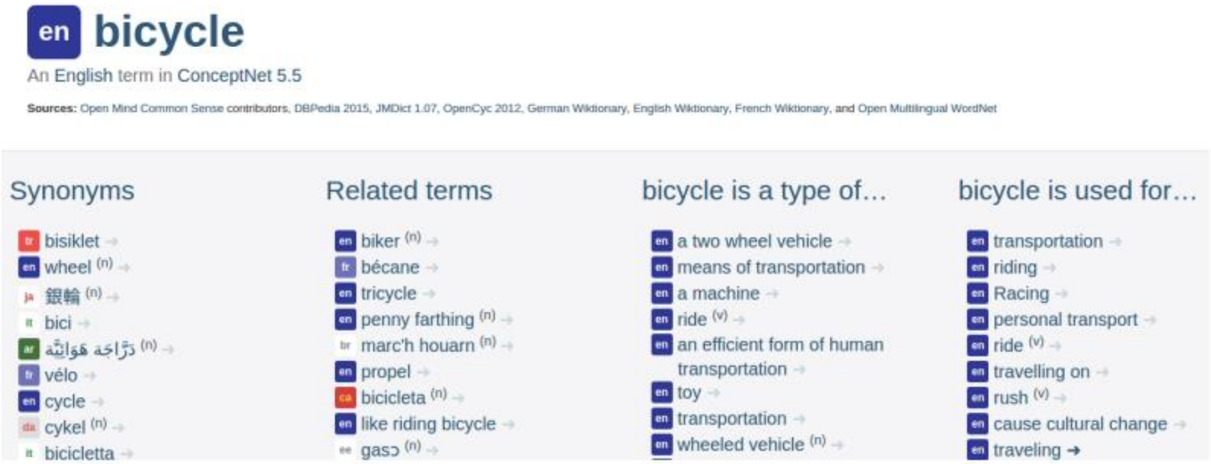
On March 14, 1964, Ruby was convicted of murder with malice, for which he received a death sentence.

t_1, R, t_2	score
murder, Causes, *death**	1.00
murder, Causes, *death sentence*	0.86
murder, HasSubevent, *death*	0.84
murder, CapableOf, *death*	0.51

Commonsense Knowledge Graph

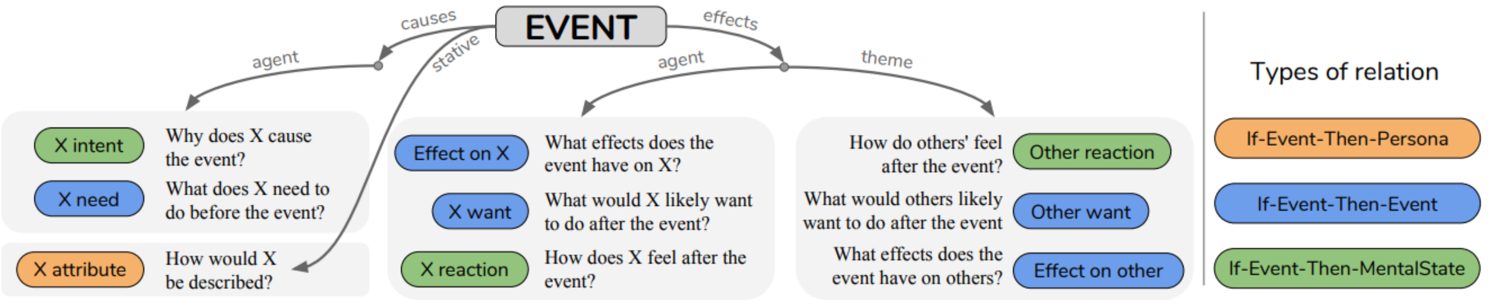
- ConceptNet 5.5 : An Open Multilingual Graph of General Knowledge (AAAI 2017)
  - ConceptNet is a freely-available semantic network designed to create word embeddings like word2vec, but better
  - The word embeddings are free, multilingual, aligned across languages, and designed to avoid representing harmful stereotypes.
  - Demo :
  - [ConceptNet]()

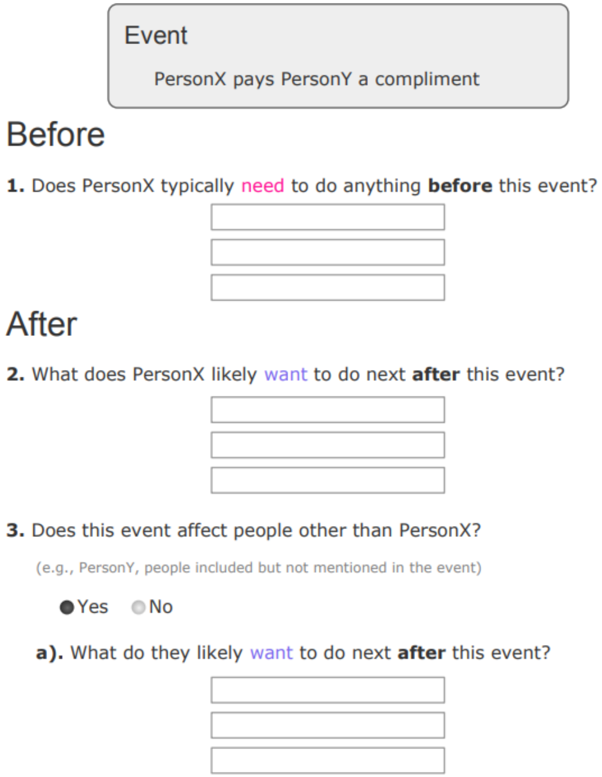
- **Symmetric relations**: *Antonym*, *DistinctFrom*, *EtymologicallyRelatedTo*, *LocatedNear*, *RelatedTo*, *SimilarTo*, and *Synonym*
- **Asymmetric relations**: *AtLocation*, *CapableOf*, *Causes*, *CausesDesire*, *CreatedBy*, *DefinedAs*, *DerivedFrom*, *Desires*, *Entails*, *ExternalURL*, *FormOf*, *HasA*, *HasContext*, *HasFirstSubevent*, *HasLastSubevent*, *HasPrerequisite*, *HasProperty*, *InstanceOf*, *IsA*, *MadeOf*, *MannerOf*, *MotivatedByGoal*, *ObstructedBy*, *PartOf*, *ReceivesAction*, *SenseOf*, *SymbolOf*, and *UsedFor*



Commonsense Knowledge Graph

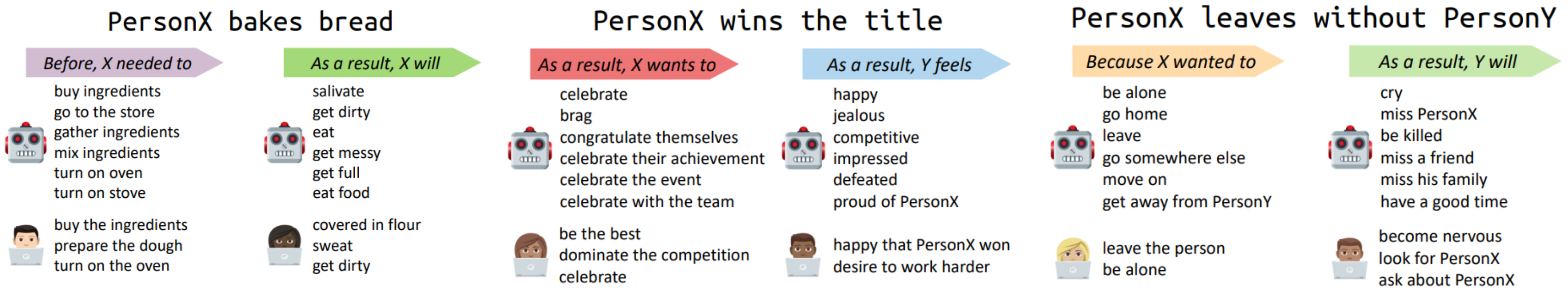
- ATOMIC : An Atlas of Machine Commonsense for If-Then Reasoning (AAAI 2019)
  - ATOMIC propose nine if-then relation types to distinguish causes vs effects, agents vs themes, voluntary vs involuntary events, and actions vs mental states
  - With ATOMIC, neural models can acquire simple commonsense capabilities and reason about previously unseen events.
  - ATOMIC is a crowdsourced dataset that may be used for commonsense reasoning.





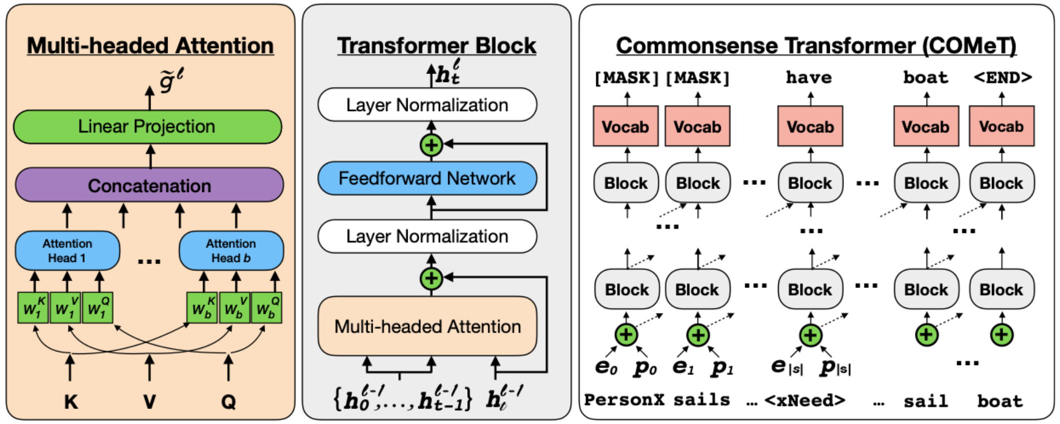
Commonsense Knowledge Graph

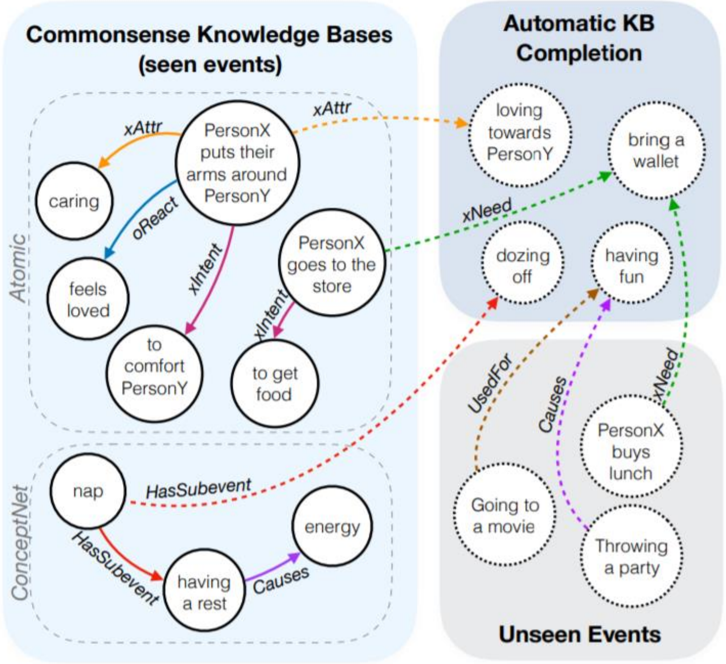
- ATOMIC : An Atlas of Machine Commonsense for If-Then Reasoning (AAAI 2019)
  - Day-to-day commonsense reasoning can be operationalized through a densely connected collection of inferential knowledge.
  - Vast majority of AI systems are trained for task-specific datasets and objectives, therefore **lack** simple and explainable commonsense reasoning abilities.
  - Given an event phrase e and an inference dimension c, the model generate the target t (Conditional Sequence Generation Problem)
  - While **ConceptNet** focuses to capture general commonsense knowledge, ATOMIC focuses on sequences of events and the social commonsense relating to them.



Commonsense Knowledge Graph

- COMET : Commonsense Transformers for Automatic Knowledge Graph Construction (ACL 2019)
  - COMET is a framework for adapting the weights of language models to learn to produce novel and diverse commonsense knowledge tuples.
  - Using two Commonsense Knowledge Bases, ATOMIC and ConceptNet, COMET produces novel commonsense knowledge.
  - COMET uses GPT as a baseline LM to train on with BLEU-2 as a metric.





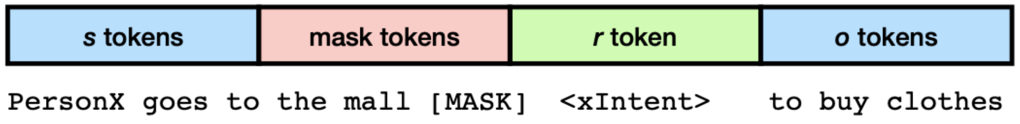
Seed	Relation	Completion	Plausible
piece	PartOf	machine	✓
bread	IsA	food	✓
oldsmobile	IsA	car	✓
happiness	IsA	feel	✓
math	IsA	subject	✓
mango	IsA	fruit	✓
maine	IsA	state	✓
planet	AtLocation	space	✓
dust	AtLocation	fridge	
puzzle	AtLocation	your mind	🤔
college	AtLocation	town	✓
dental chair	AtLocation	dentist	✓
finger	AtLocation	your finger	
sing	Causes	you feel good	✓
doctor	CapableOf	save life	✓
post office	CapableOf	receive letter	✓
dove	SymbolOf	purity	✓
sun	HasProperty	big	✓
bird bone	HasProperty	fragile	✓
earth	HasA	many plant	✓
yard	UsedFor	play game	✓
get pay	HasPrerequisite	work	✓
print on printer	HasPrerequisite	get printer	✓
play game	HasPrerequisite	have game	✓
live	HasLastSubevent	die	✓
swim	HasSubevent	get wet	✓
sit down	MotivatedByGoal	you be tire	✓
all paper	ReceivesAction	recycle	✓
chair	MadeOf	wood	✓
earth	DefinedAs	planet	✓

Commonsense Knowledge Graph

- COMET : Commonsense Transformers for Automatic Knowledge Graph Construction (ACL 2019)
  - Authors use MLM training objective to fill in the o tokens given the s and r tokens.
  - r tokens are learned during fine tuning.
  - Within the ablation studies, it was proved empirically that using pretrained weights outperforms randomly initialized weights.
  - Also, it was proved empirically that using Greedy Decoding outperforms Beam Search or Random Sampling during decoding.
  - Using COMET, we can explicitly extract knowledge from a pretrained LM and represent relations with language.
  - Demo :
  - https://mosaickg.apps.allenai.org/model-comet2020

Model	PPL[5]	BLEU-2	N/T *sro*[6]	N/T *o*	N/U *o*
9ENC9DEC (Sap et al., 2019)	-	10.01	100.00	8.61	40.77
NearestNeighbor (Sap et al., 2019)	-	6.61	-	-	-
Event2(IN)VOLUN (Sap et al., 2019)	-	9.67	100.00	9.52	45.06
Event2PERSONX/Y (Sap et al., 2019)	-	9.24	100.00	8.22	41.66
Event2PRE/POST (Sap et al., 2019)	-	9.93	100.00	7.38	41.99
COMET (- pretrain)	15.42	13.88	100.00	7.25	45.71
COMET	**11.14**	**15.10**	100.00	**9.71**	**51.20**

ATOMIC Input Template and ConceptNet Relation-only Input Template



ConceptNet Relation to Language Input Template

$$\mathcal{L} = -\sum_{t=|s|+|r|}^{|s|+|r|+|o|} \log P(x_t | x_{<t}) \qquad (11)$$

COMET Decoding method	oEffect	oReact	oWant	xAttr	xEffect	xIntent	xNeed	xReact	xWant	Avg
Top-5 random sampling (n=2500 per relation)	34.60	44.04	35.56	64.56	55.68	58.84	46.68	80.96	58.52	53.27
Top-10 random sampling (n=5000 per relation)	25.20	37.42	27.34	49.20	47.34	47.06	38.24	72.60	48.10	43.61
Beam search - 2 beams (n=1000 per relation)	43.70	54.20	47.60	**84.00**	51.10	73.80	50.70	85.80	78.70	63.29
Beam search - 5 beams (n=2500 per relation)	37.12	45.36	42.04	63.64	**61.76**	63.60	57.60	78.64	68.40	57.57
Beam search - 10 beams (n=5000 per relation)	29.02	37.68	44.48	57.48	55.50	68.32	64.24	76.18	75.16	56.45
Greedy decoding (n=500 per relation)	**61.20**	**69.80**	**80.00**	77.00	53.00	**89.60**	**85.60**	**92.20**	**89.40**	**77.53**
Human validation of gold ATOMIC	84.62	86.13	83.12	78.44	83.92	91.37	81.98	95.18	90.90	86.18

ANY QUESTIONS?