SAPIENZA
UNIVERSITÀ DI ROMA

Link Prediction on Knowledge Graphs for Service-Oriented Computing

Donatella Firmani

donatella.firmani@uniroma1.it

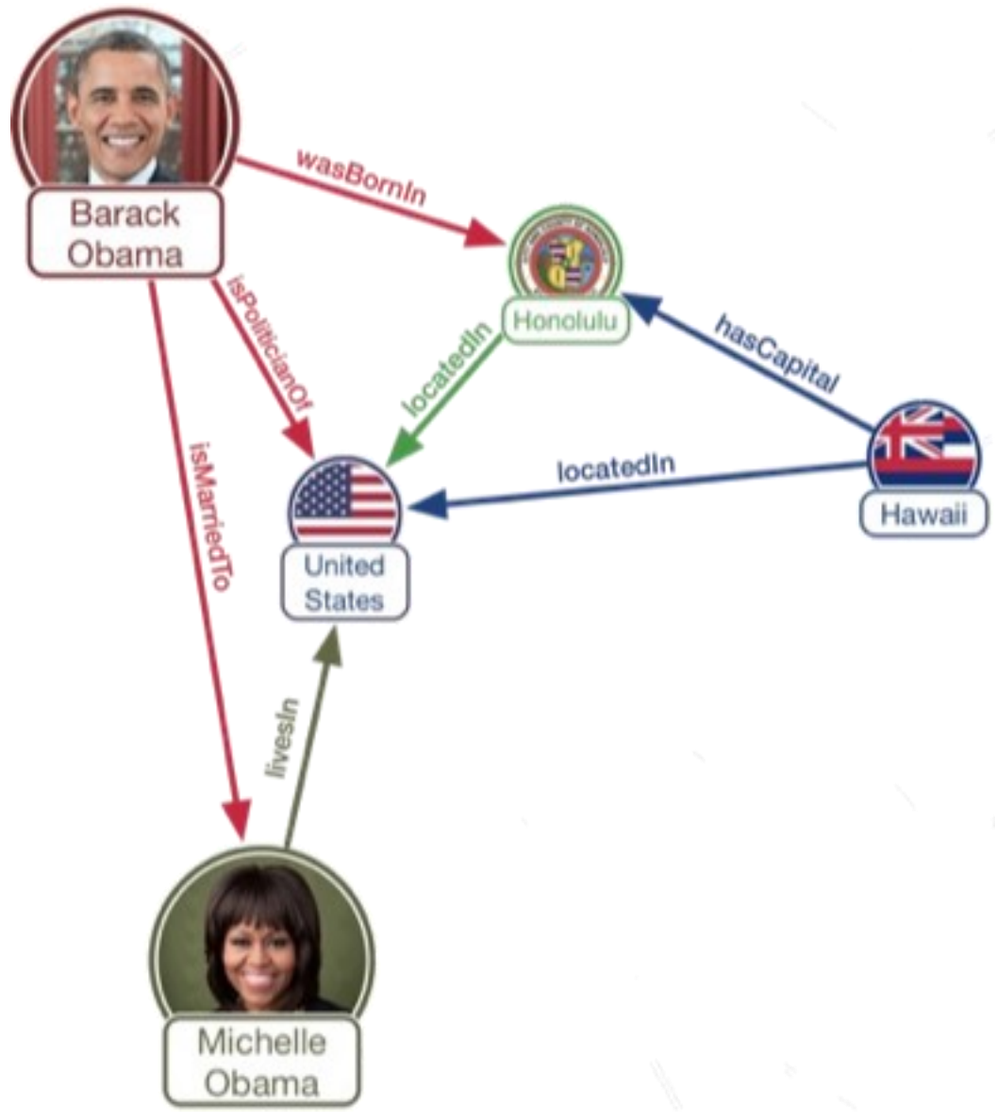
https://sites.google.com/uniroma1.it/donatellafirmani

SummerSOC
2026

Knowledge Graphs

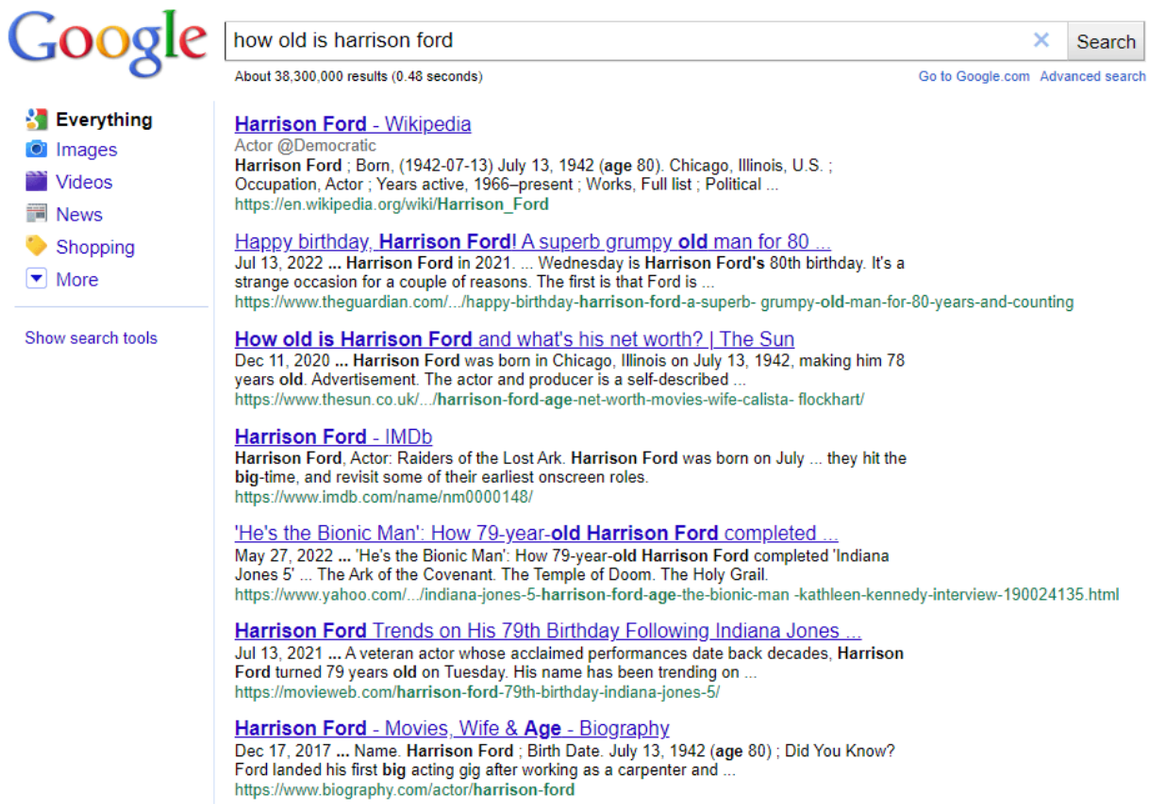
- Knowledge Graphs (KGs) are a widely used notion in many applications domains, sometimes with different terminology

- In this tutorial
  - KG → directed labelled graph
  - nodes → entities
  - labels → relations (names)
  - edge e → fact, triple, link
    - e = <h, r, t>
    - h → head
    - r → relation
    - t → tail



Popularity

- The Google search engine underwent a huge transformation that turned a google search from looking like this in 2012:



Popularity

- To this in 2023:

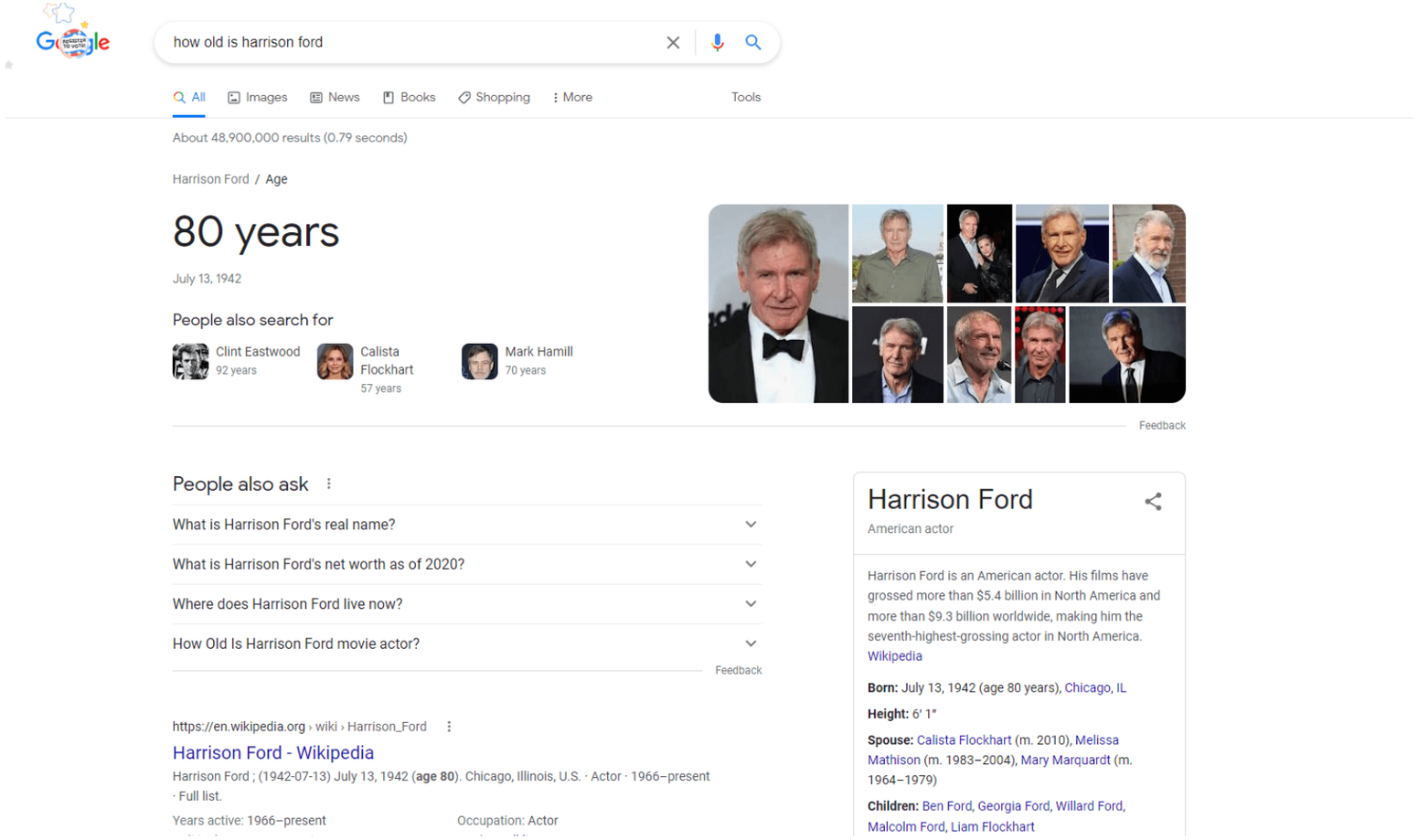


Image Source: https://blogs.cornell.edu/info2040/2022/09/20/the-impact-of-knowledge-graphs-on-search-engines/

Exciting novel developments

- Screenshot from 2026:

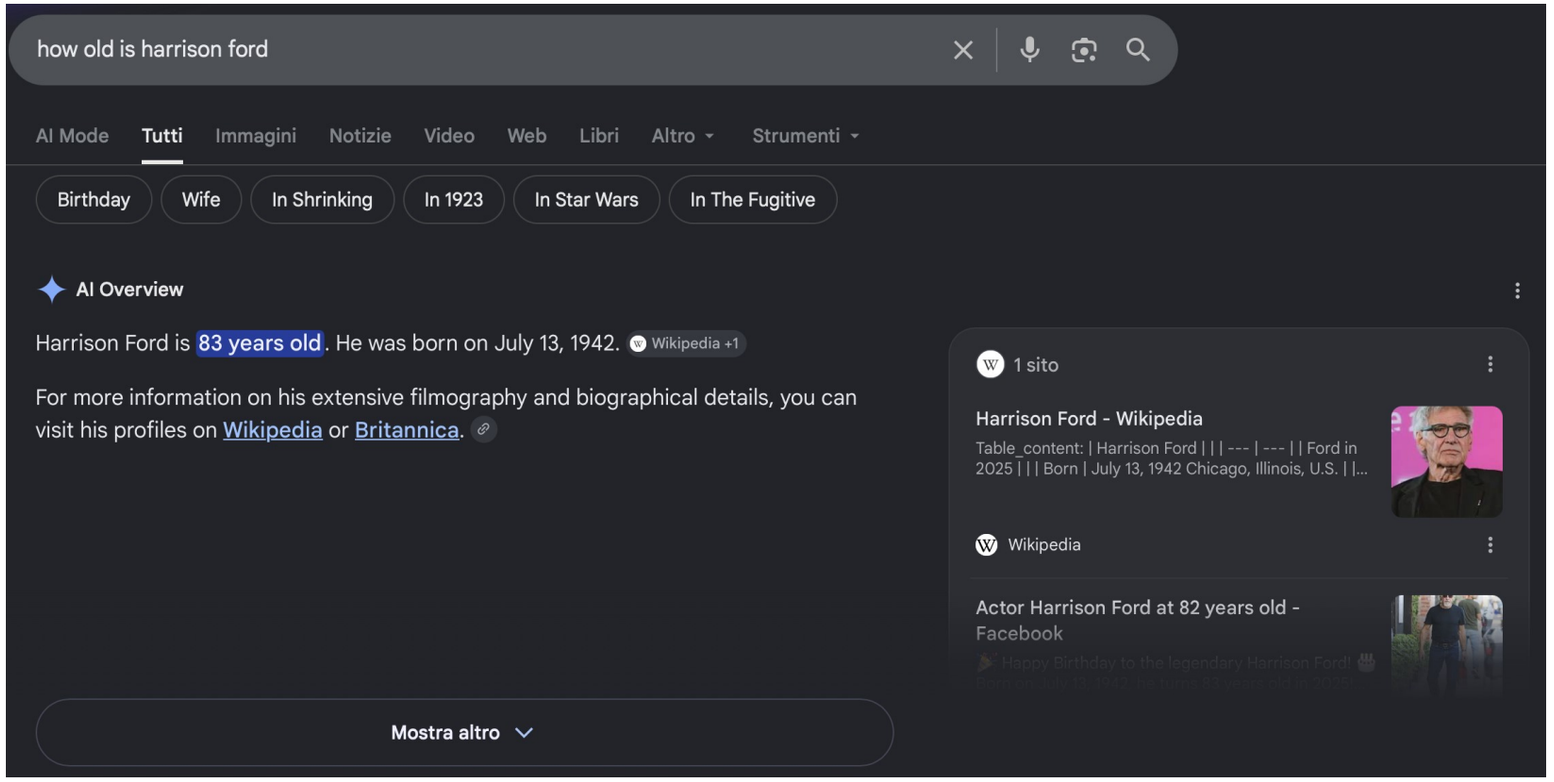


Image Source: https://blogs.cornell.edu/info2040/2022/09/20/the-impact-of-knowledge-graphs-on-search-engines/

What changed in 2012

- What changed was the implementation of a KG
- In this way a google search doesn't directly match words to other words, but matches the search to "things", i.e. entities

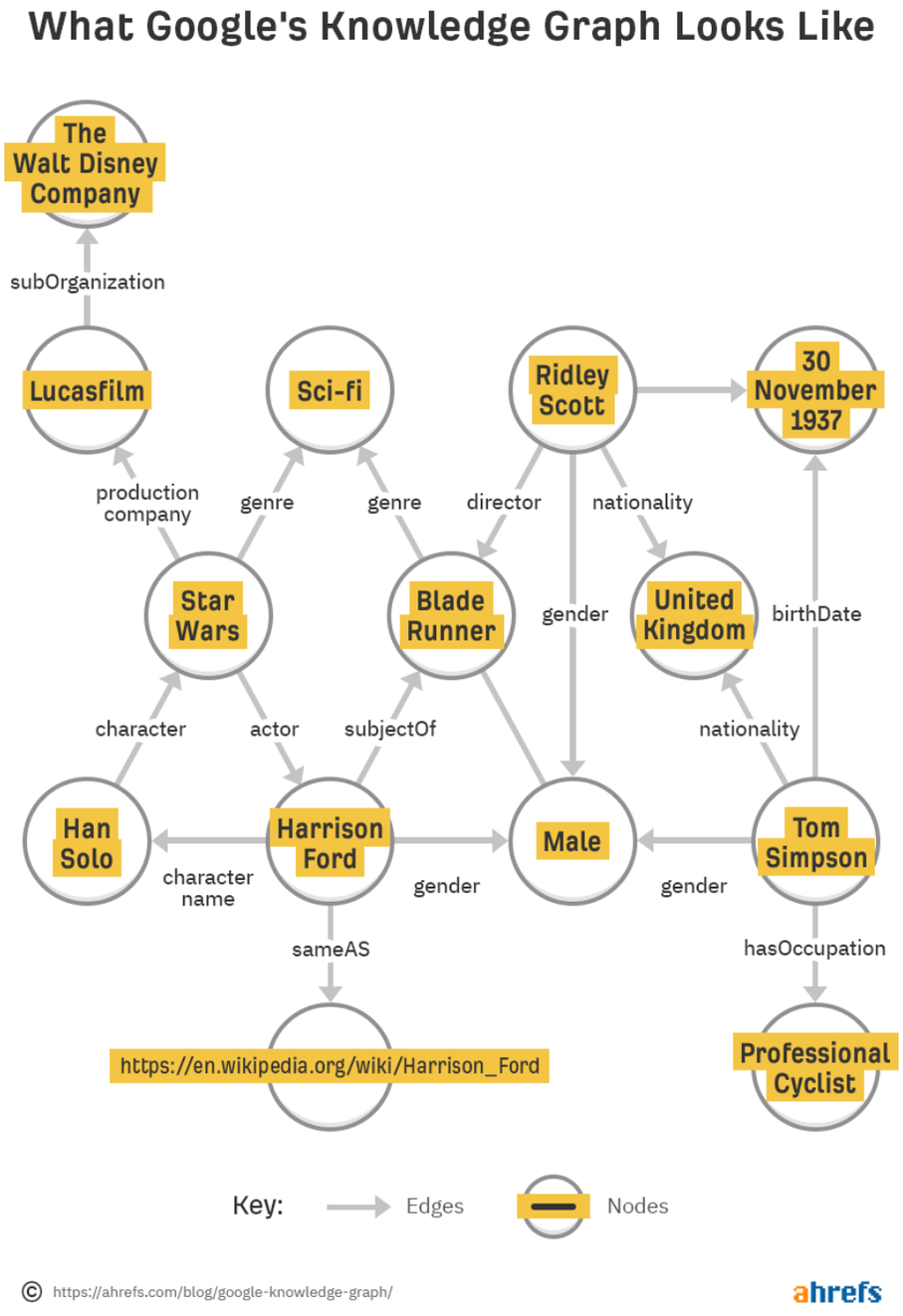
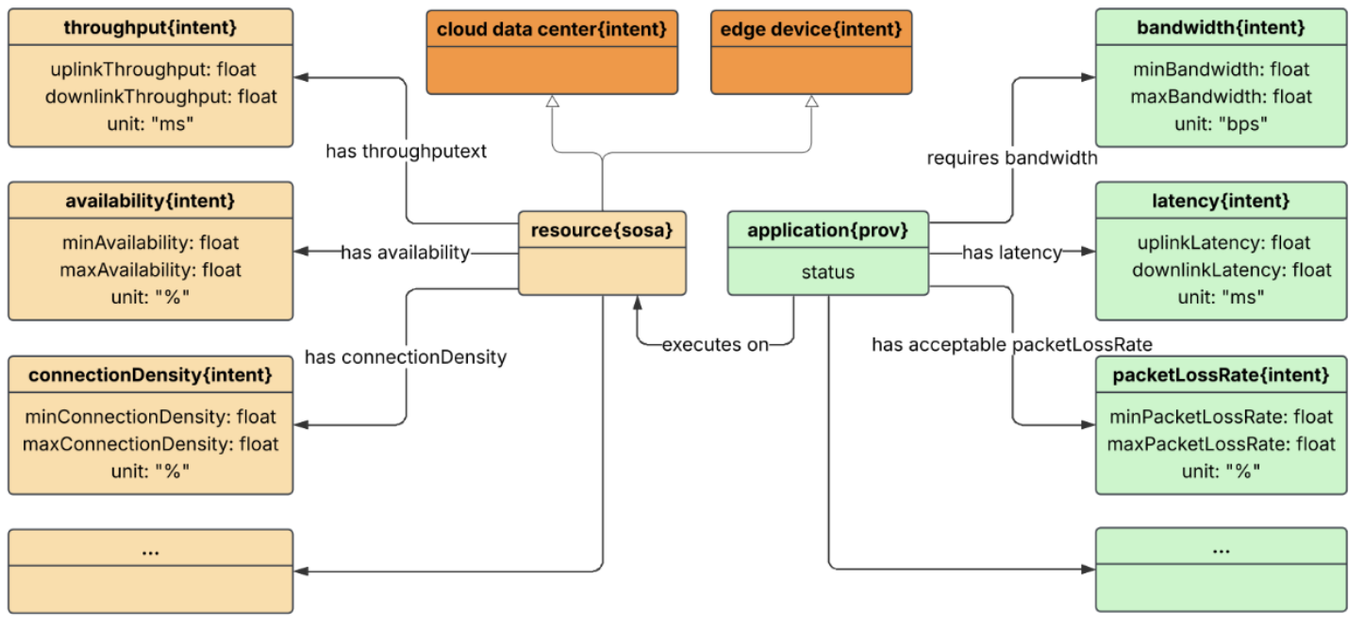


Image Source: https://ahrefs.com/blog/google-knowledge-graph/

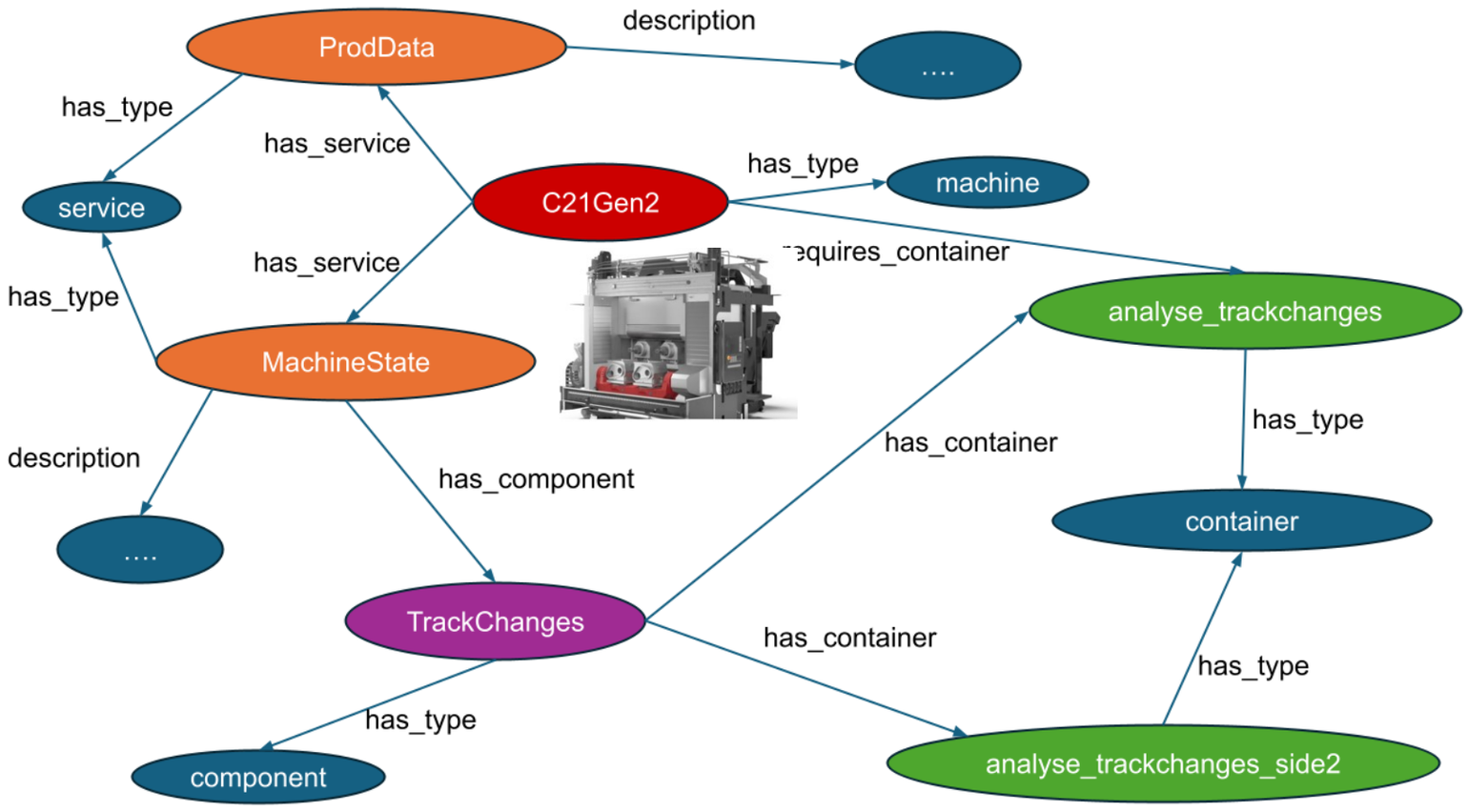
KGs and Service-Oriented Computing

- KGs can be used as data backbone for computing platform
- Domain data
  - Application data
  - Hardware data
  - Network properties
- Service-oriented data
  - Service dependencies
  - High-level computation intents



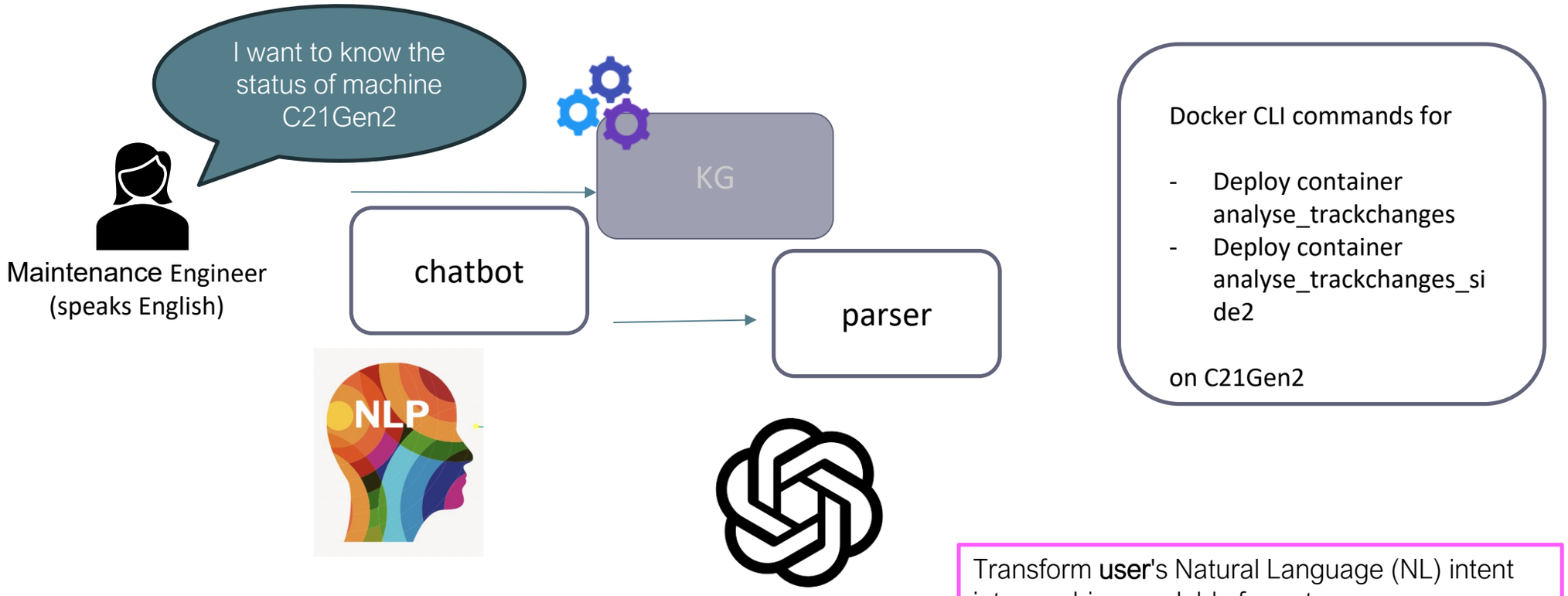
Example from a real-world scenario

- A manufacturing company providing a platform to collect and analyze sensor data for monitoring of machine health

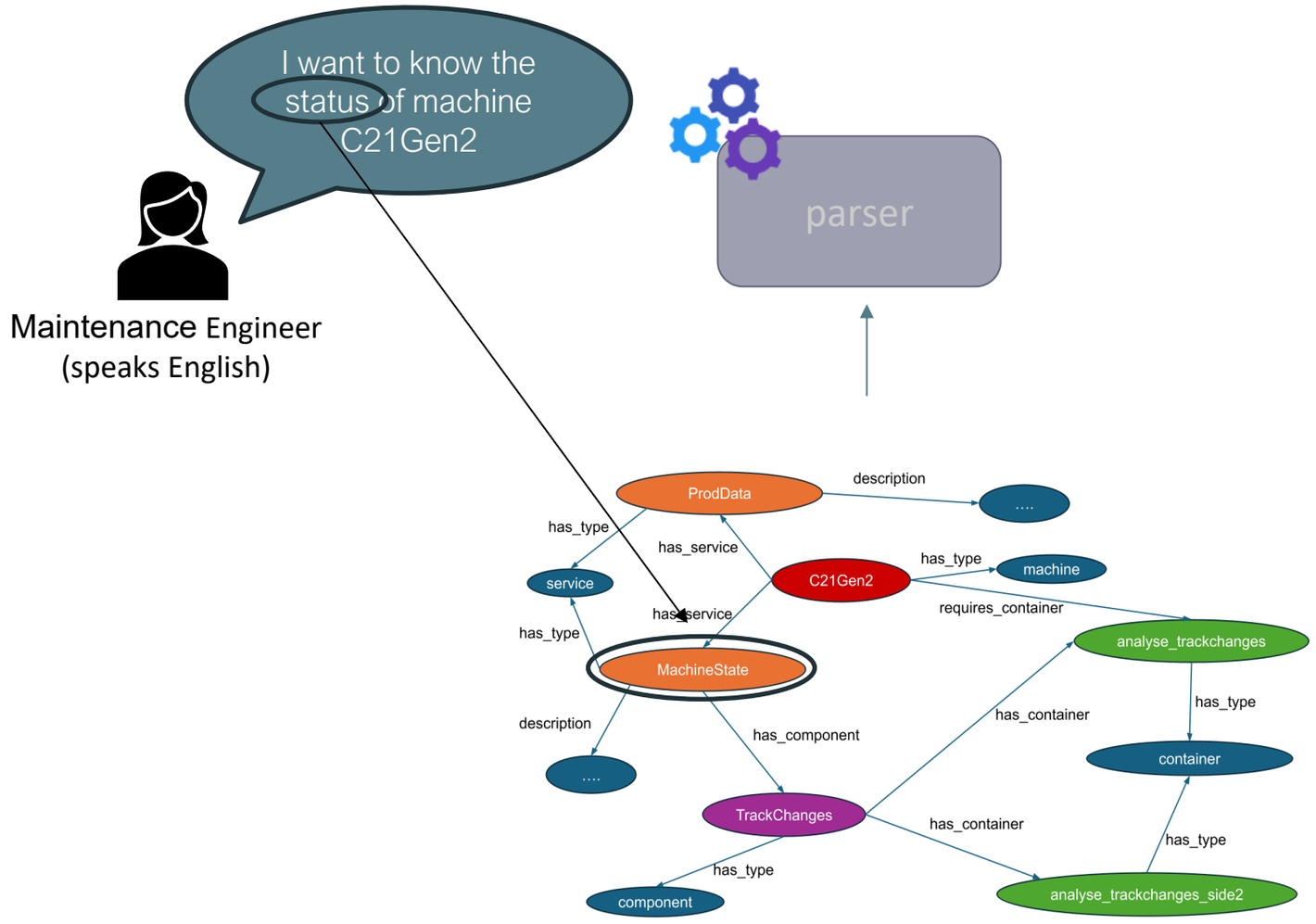


Example from a real-world scenario

- INTEND project: Intent-based interface for data operation



Reasoning about services



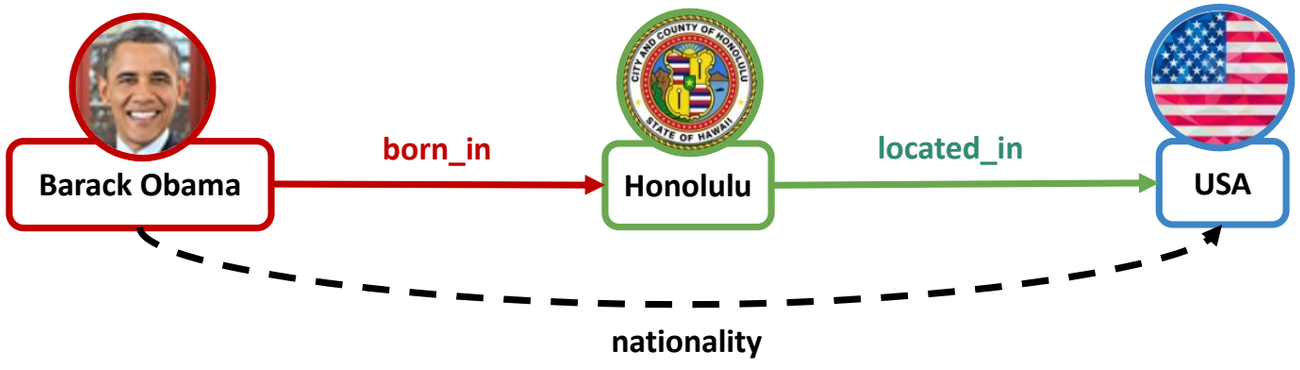
Promising results with the complete KG

- 10 intents from experts corresponding to a container set
- Max precision and recall in all cases, including empty

Intent	Set	Containers	|*Containers*|	Prec.	Rec.
Intent 1	Ground Truth	C1,C2,C3,C4,C5,C23,C25,C26	8	-	-
	Workflow Output	C1,C2,C3,C4,C5,C23,C25,C26	8	1.0	1.0
Intent 2	Ground Truth	C1,C2,C5,C6,C7,C10,C23,C24,C25,C26,C27,C28	12	-	-
	Workflow Output	C1,C2,C5,C6,C7,C10,C23,C24,C25,C26,C27,C28	12	1.0	1.0
Intent 3	Ground Truth	C1,C2,C5,C6,C7,C10,C11,C12,C15,C16,C17,C18,C30	13	-	-
	Workflow Output	C1,C2,C5,C6,C7,C10,C11,C12,C15,C16,C17,C18,C30	13	1.0	1.0
Intent 4	Ground Truth	C11,C12,C15,C16,C17,C18,C21,C22	8	-	-
	Workflow Output	C11,C12,C15,C16,C17,C18,C21,C22	8	1.0	1.0
Intent 5	Ground Truth	C11,C12,C17,C21	4	-	-
	Workflow Output	C11,C12,C17,C21	4	1.0	1.0
Intent 6	Ground Truth	C31,C32,C33,C34,C39	5	-	-
	Workflow Output	C31,C32,C33,C34,C39	5	1.0	1.0
Intent 7	Ground Truth	∅	0	-	-
	Workflow Output	∅	0	1.0	1.0
Intent 8	Ground Truth	C23,C24	2	-	-
	Workflow Output	C23,C24	2	1.0	1.0
Intent 9	Ground Truth	∅	0	-	-
	Workflow Output	∅	0	1.0	1.0
Intent 10	Ground Truth	C29	1	-	-
	Workflow Output	C29	1	1.0	1.0

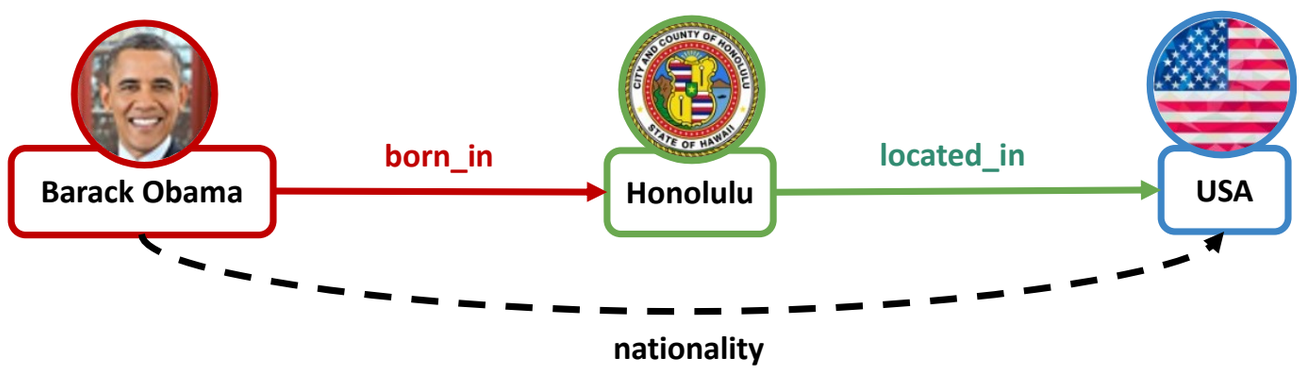
Bad news: KG Incompleteness

- KGs are often incomplete
  - the absence of a fact does not mean that the fact is false
  - E.g. the absence of a service availability on a certain machine could be due to data processing errors, to observation incompleteness and so on.
- Examples about service availability might be difficult so let's go back to Barack Obama sometimes ☺



Link Prediction

- Many types of incompleteness, such as
  - missing entities
  - missing relations
  - missing facts
  - …
- Link Prediction (LP): inferring new facts from known ones

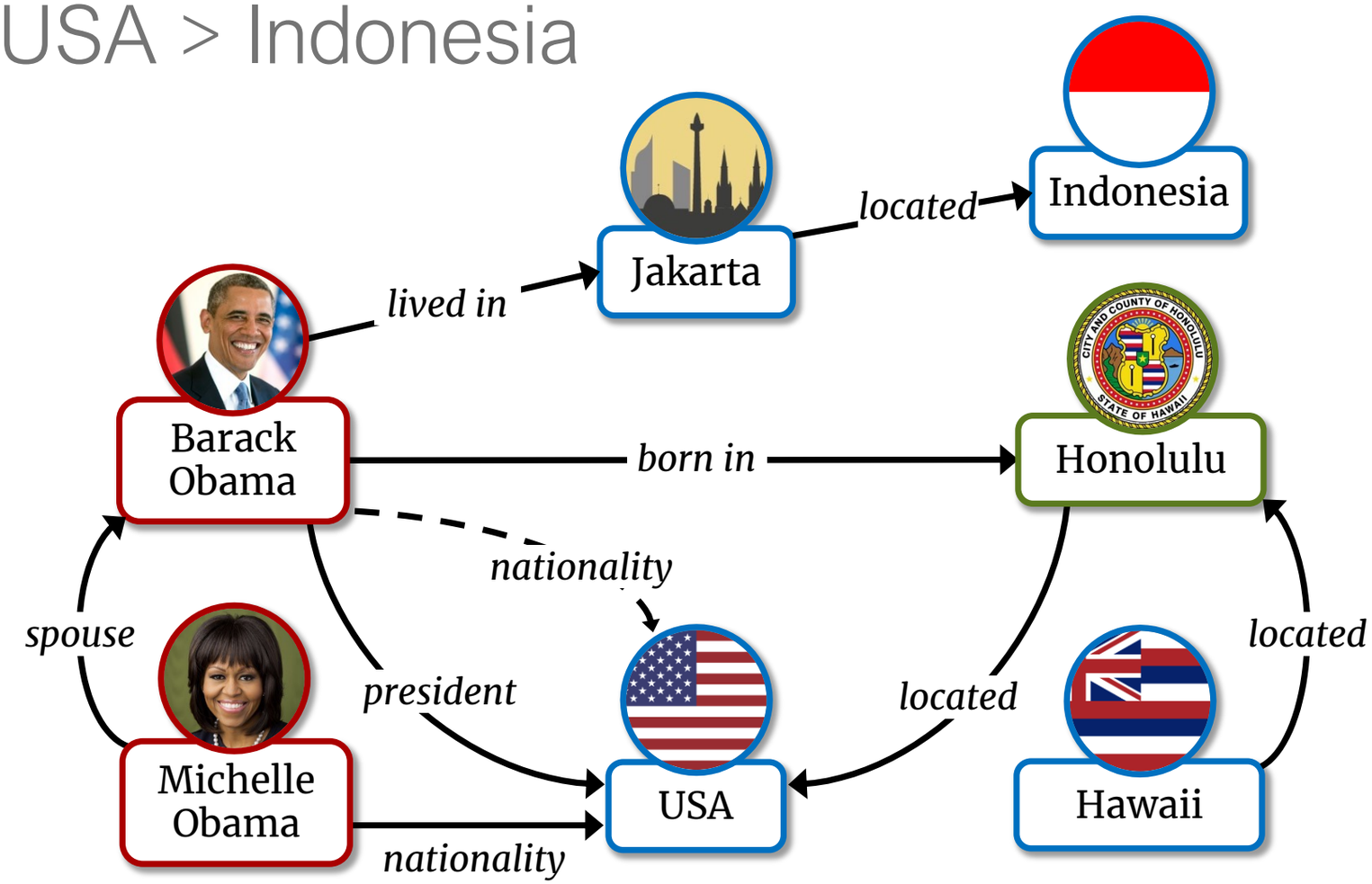


Outline of this Tutorial

- Introduction to methods for LP
  - Two families of approaches
    - rule-based methods → Interpretable but sub-opt
    - embedding-based methods → Accurate but opaque
  - We focus on embedding-based
- Known challenges and research directions
- Explainable AI for Link Prediction
- Concluding remarks with open directions

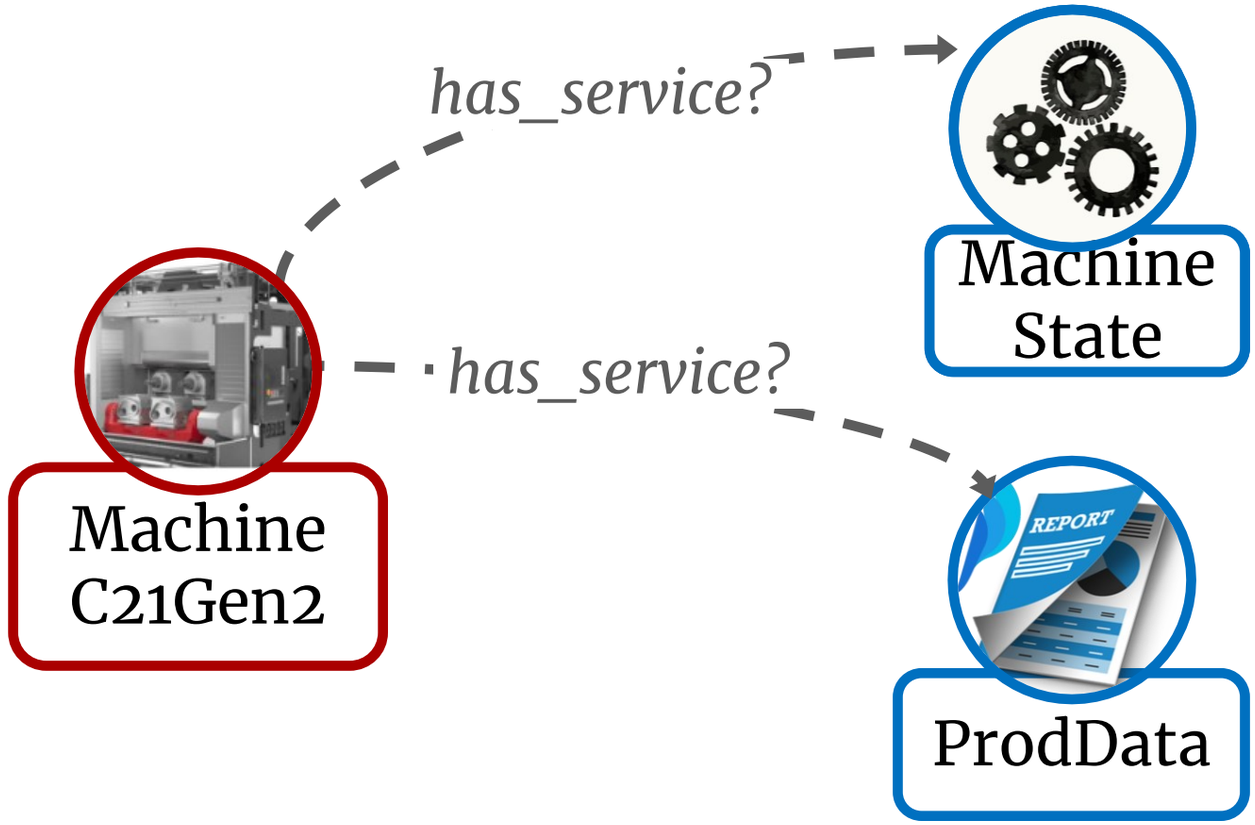
Rationale for LP methods

- Given a query <Barack Obama, Nationality, ?>
- Use available facts for computing plausibility of answers
- Predict USA > Indonesia

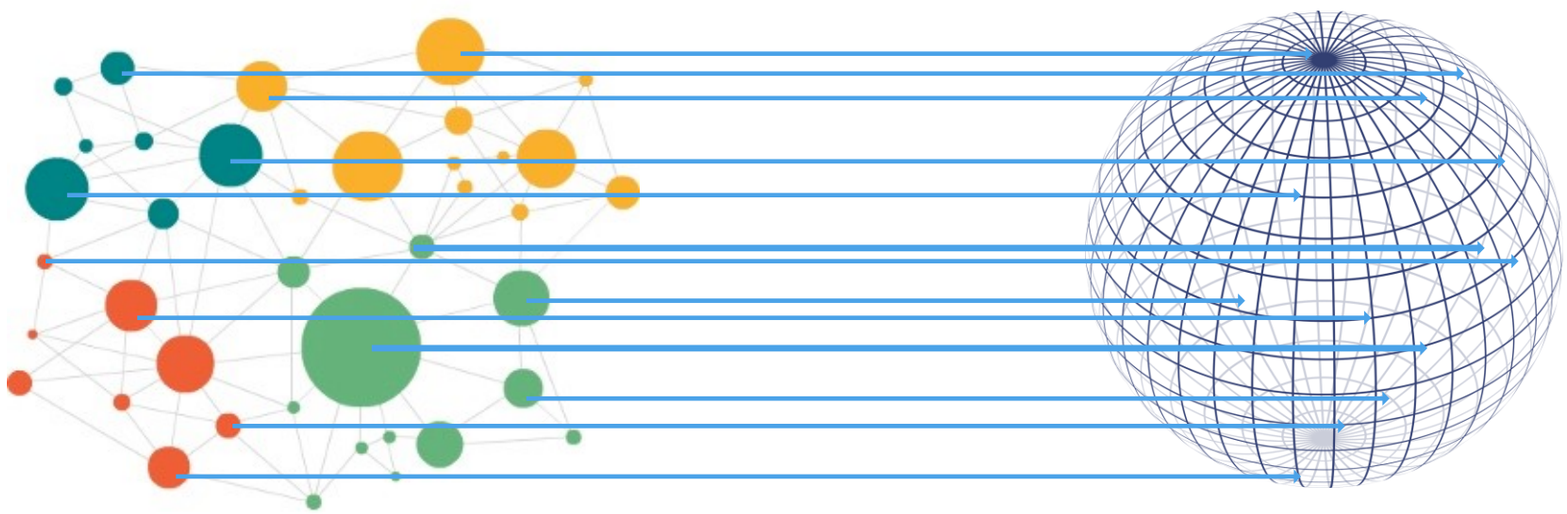


Rationale for LP methods

- Same example for machine data



The power of Embedding



Knowledge Graph

Latent Space

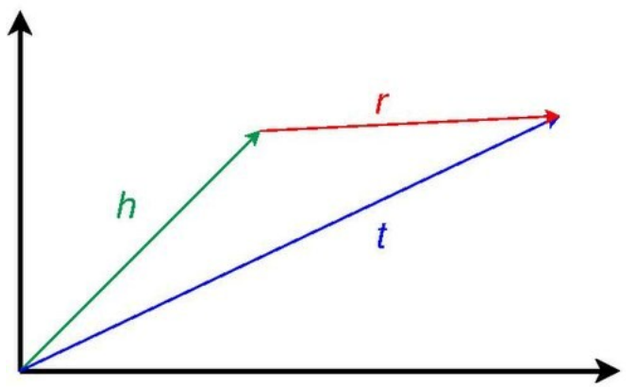
Analogously to word embeddings, entities and relations are mapped to points (vectors) or matrices in the latent space

KG Embedding

- Vectors or matrices for entities and relations are used to compute a scoring function ϕ

- ϕ estimates the plausibility of a triple <h,r,t> by combining the embeddings of h, r and t

- Example: In plausible facts the head plus the relation should be equal to the vector representation of the tail entity.



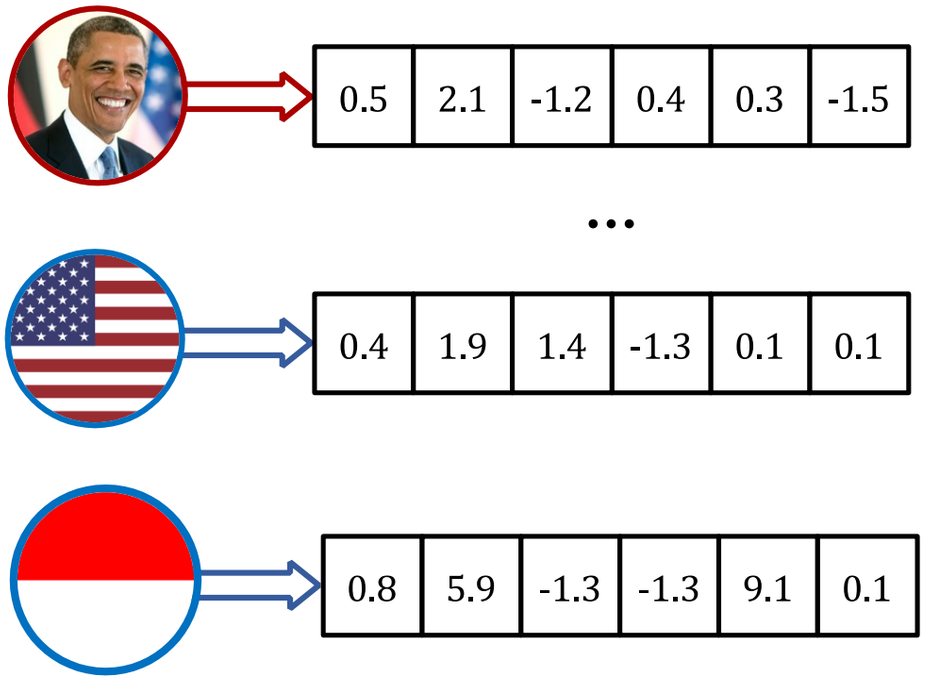


Image Source: Wikipedia page on Knowledge graph embedding

Training

- Split KG in train & test facts
- Loss: scoring function error for triples in the training set
  - E.g., Negative Log-Likelihood

Measuring accuracy

- De facto benchmark datasets, such as WN18 or FB15k (from Freebase, the backbone of the Google KG)

- HITS@k
  - the ratio of predictions for which the correct entity is in the top K

$$\text{Hits@K} = \frac{|\{q \in Q : tq < k\}|}{|Q|}$$

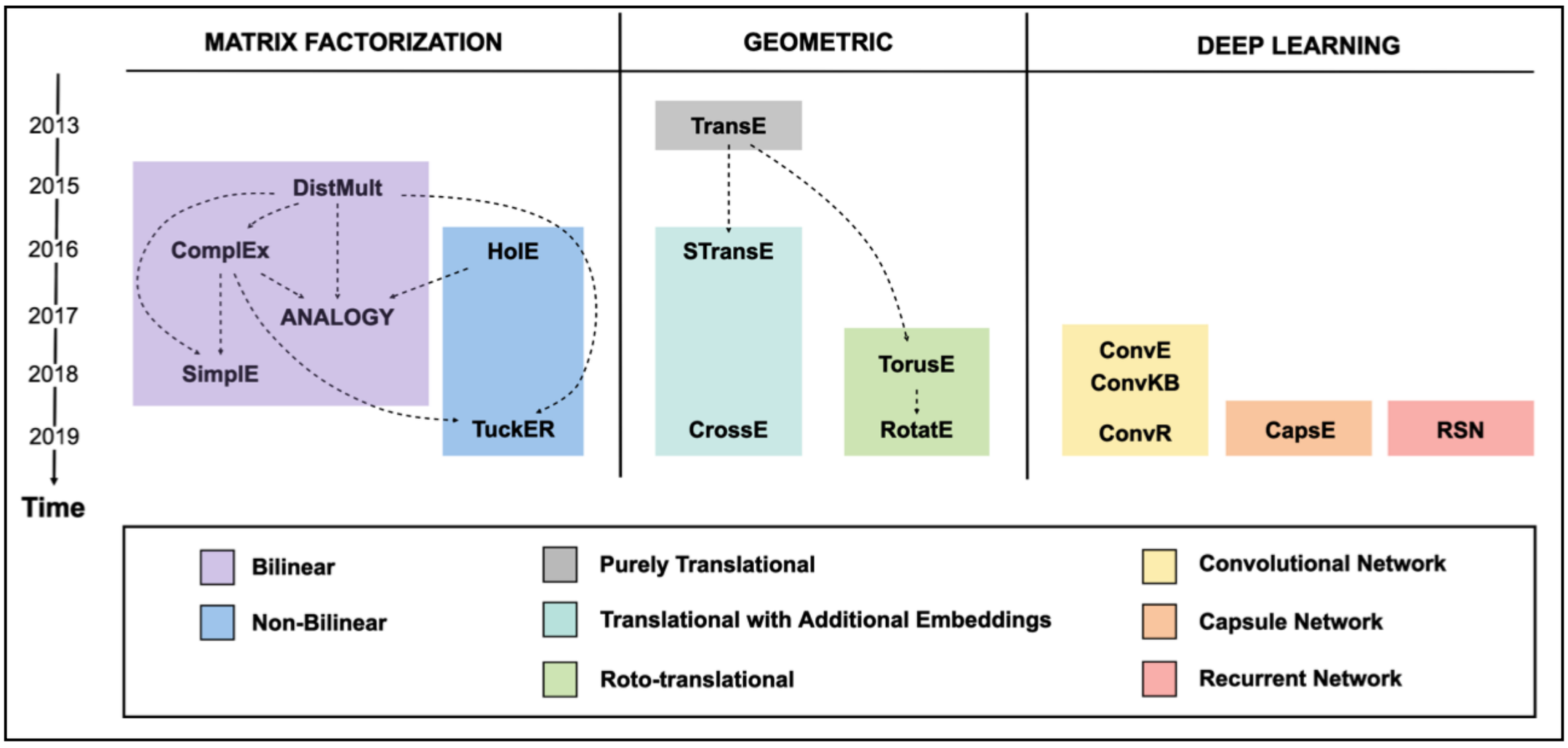
- MRR (Mean Reciprocal Rank)
  - inverse of the rank of the correct entity, average

- Both metrics are in [0,1], the higher the better

Ali, M., Berrendorf, M., Hoyt, C. T., Vermue, L., Galkin, M., Sharifzadeh, S., ... & Lehmann, J. Bringing light into the dark: A large-scale evaluation of knowledge graph embedding models under a unified framework. IEEE TPAMI, 2021.

Taxonomy of methods pre Large Language Models

- Different methods use different scoring and loss functions
- Many novel methods are proposed every year



Rossi, A., Barbosa, D., Firmani, D., Matinata, A., & Merialdo, P.. Knowledge graph embedding for link prediction: A comparative analysis. ACM TKDD 2021.

Large Language Models

- In-context learning → NL description of the KG
- It reduces training overhead with SOTA performances ☺
- It works mostly for common-sense KGs
  - Does not apply to service data ☹
- Outside the scope of this tutorial

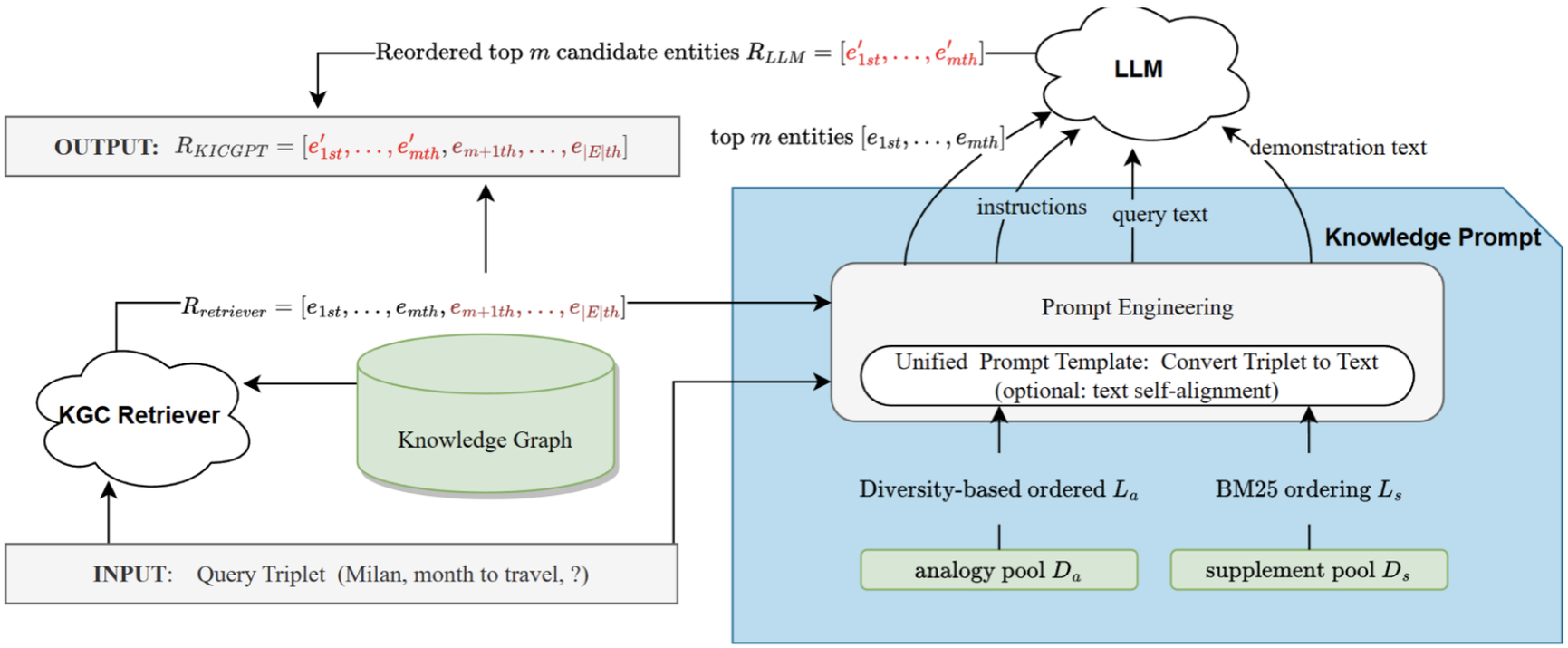


Image Source: Wei, Yanbin, et al. "KICGPTv2: Large Language Model with Knowledge in Context for Knowledge Graph Completion.» IEEE Transactions on Knowledge and Data Engineering (2026).

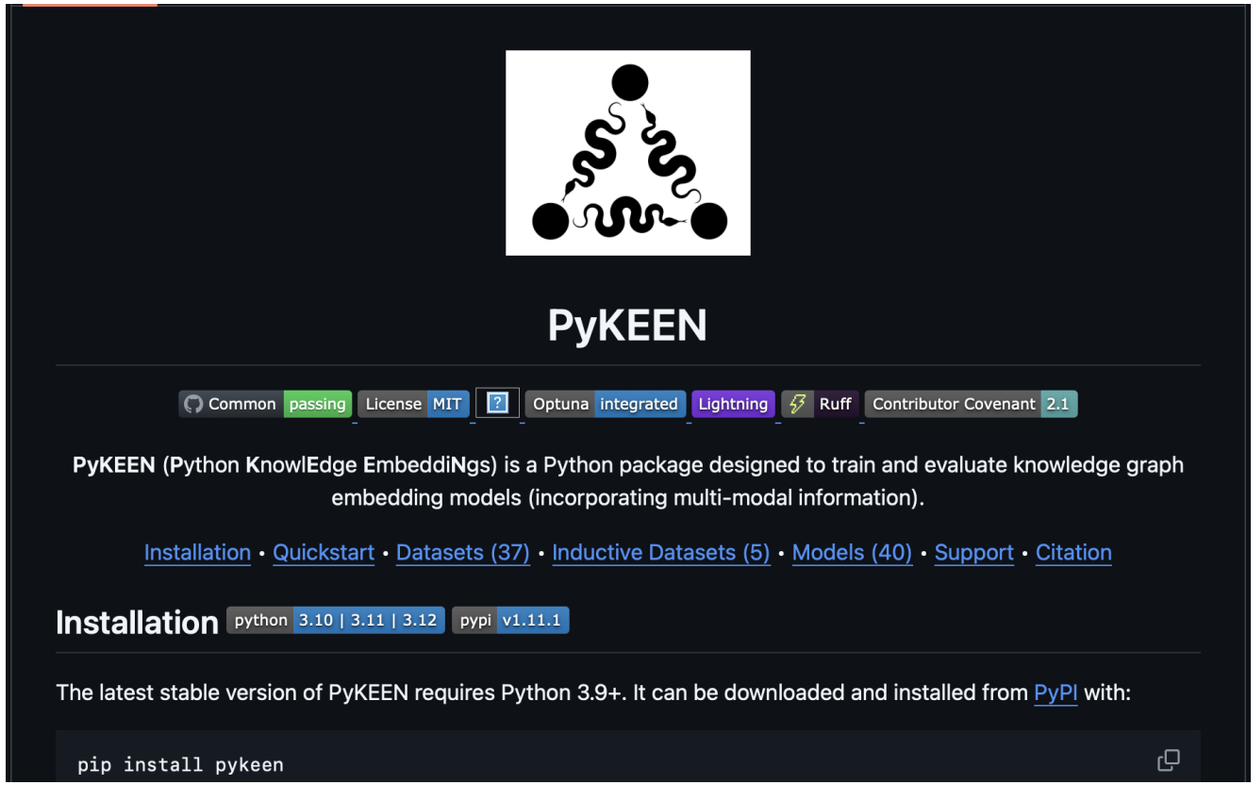
Outline of this Tutorial

- Introduction to methods for LP
  - Two families of approaches
    - rule-based methods → Interpretable but sub-opt
    - embedding-based methods → Accurate but opaque
  - We focus on embedding-based
- Known challenges and research directions
- Explainable AI for Link Prediction
- Concluding remarks with open directions

Reproducibility

- Difficult to reproduce papers' results
  - Different technologies and platforms
  - Strong dependence on hyper-parameters setting
  - Expensive training process (hundreds of GPU hours)

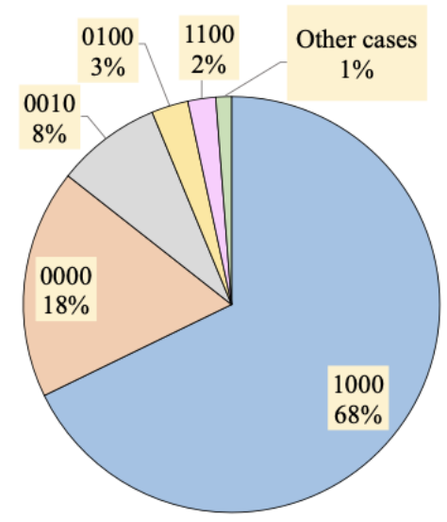
- Extremely difficult to apply models to your own data and workflow
  - There are efforts to put everything in the same package (E.g., PyKEEN)
  - Internals of the models are still difficult to access



Ali, M., Berrendorf, M., Hoyt, C. T., Vermue, L., Sharifzadeh, S., Tresp, V., & Lehmann, J.
PyKEEN 1.0: a python library for training and evaluating knowledge graph embeddings. JMLR, 2021.

Training Data Leakage

- Some datasets (e.g. FB15k and WN18) are affected by data redudancy, a form of training data leakage
  - reverse relations
    - e.g., <europe, has_part, republic_of_estonia> and <republic_of_estonia, part_of, europe>
  - duplicate relations
    - e.g., football_player/current_team and sports_team_roster/position
  - cartesian product relations
    - every triple in the cart. prod. of subjects and objects is true
  - …



- More robust and less common-sense benchmarks are needed
  - Service data would provide an excellent domain!

Akrami, F., Saeef, M. S., Zhang, Q., Hu, W., & Li, C. Realistic re-evaluation of knowledge graph completion methods: An experimental study. SIGMOD 2020.

Intrinsic evaluation metric issues

- Rank based methods do not consider score
- Need for a score-aware evaluation framework
  - direction: probability theory of proper scores

Model A		Model B		Model C	
Entity	Score	Entity	Score	Entity	Score
Canada	8.5	**Canada**	1.5	USA	1.5
USA	1.0	USA	1.3	Toronto	1.5
Toronto	0.8	Toronto	1.3	Tom	1.4
Rome	0.8	Rome	1.2	**Canada**	1.3
Tom	0.6	Tom	1.2	Rome	1.2

Figure 1: Predictions of models A, B and C for the query *(Tom, Nationality,?)*. The true tail is *Canada*. Rows of each table represent the entities in the KG. The line delimits the positions for Hits@3 where the true entity can appear.

Resin, Johannes. "From classification accuracy to proper scoring rules: Elicitability of probabilistic top list predictions." Journal of Machine Learning Research 24.173 (2023): 1-21.

Bias

- Given ⟨h, r, t⟩, we can identify different sources of bias
- Type 1 (source/target peers)
  - e.g., in the case of <Barack Obama, nationality, USA> there more persons that have nationality USA than Indonesia
  - Easier to predict USA citizens
- Type 2 (source/target peers with one-to-many/many-to-one)
  - e.g., <Cristiano_Ronaldo, language, English> is biased if most people, in addition to other languages, also speak English.
  - Relations that have a "default" correct answer.
- Type 3 (spurious correlations)
  - e.g., in the FB15k dataset the producer of a TV program is almost always its creator
  - this may lead to assume that creating a program implies being its producer

	Entities	Relations	Facts			Test Predictions				
			Train	Valid	Test	All	w/o B1	w/o B2	w/o B3	w/o B*
FB15k	14.951	1.345	483.142	50.000	59.071	118.142	115.165 (-2.5%)	108.705 (-8.0%)	101.406 (-14.2%)	93.005 (-21.3)
FB15k-237	14.541	237	272.115	17.535	20.466	40.932	39.145 (-4.4%)	36.482 (-5.5%)	40.932 (-0.0%)	36.912 (-9.8%)
WN18	40.943	18	141.442	5.000	5.000	10.000	10.000 (-0.0%)	10.000 (-0.0%)	10.000 (-0.0%)	10.000 (-0.0%)
WN18RR	40.943	11	86.835	3.034	3.134	6.268	6.268 (-0.0%)	6.268 (-0.0%)	6.268 (-0.0%)	6.268 (-0.0%)
YAGO3-10	123.155	37	1.078.868	5.000	5.000	10.000	9.725 (-2.8%)	9.992 (-0.1%)	4.876 (-51.2%)	4.593 (-54.1%)

Rossi, A., Firmani, & Merialdo, P. Knowledge Graph Embeddings or Bias Embeddings? A Study of Bias in Link Prediction Models. DL4KG@ISWC 2021

Consequences

- Perfomance worsens considerably after removal of biased relations.
  - embedding-based LP models can low up to 50% to 70% of their H@1 and MRR in worst cases such as YAGO3-10
  - analogous results holds also for rule-based approaches

- Even accurate predictions, if made for the wrong reasons, can lead to erroneous and harmful decisions.

Research questions

- Are LLM KG methods more or less prone to bias?
- Can we identify the «reason» of a prediction?

Outline of this Tutorial

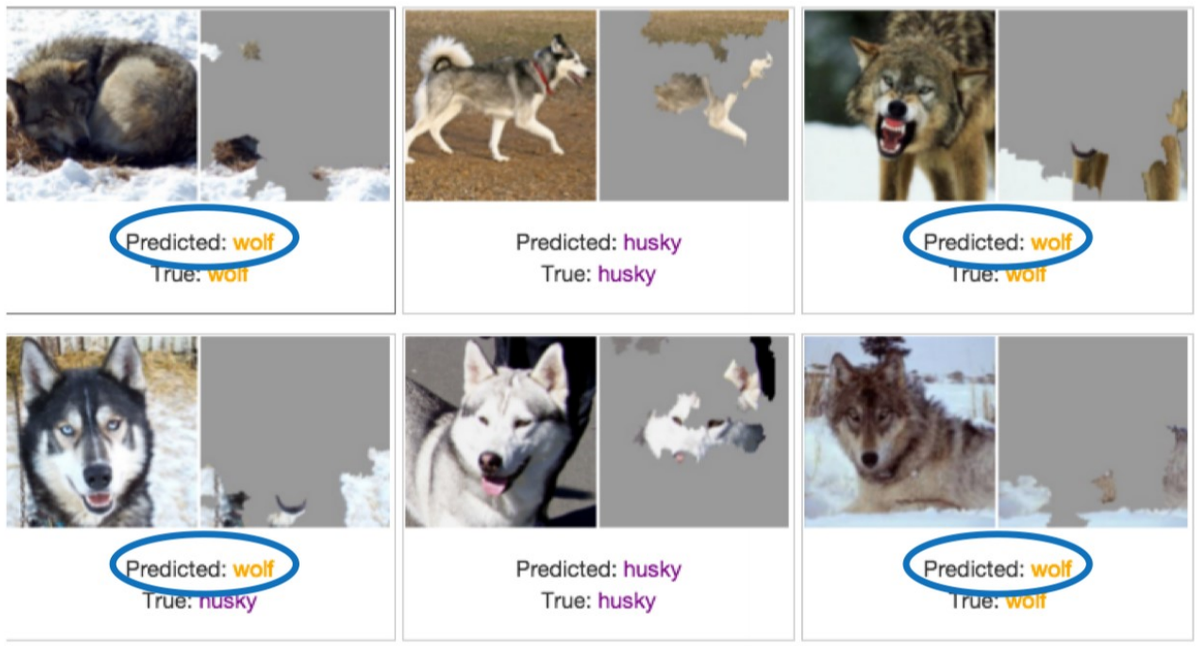
- Introduction to methods for LP
  - Two families of approaches
    - rule-based methods → Interpretable but sub-opt
    - embedding-based methods → Accurate but opaque
  - We focus on embedding-based
- Known challenges and research directions
- Explainable AI for Link Prediction
- Concluding remarks with open directions

Explainable AI in the wild

- Husky vs Wolf
- Only 1 prediction mistake: great accuracy
- Interested in “Why” questions, rather than “How Accurate”?

Image Source: https://filene.org/assets/images-layout/Panel_Singh.pdf

Snow vs land



In 2018 Pixels on the right are experimentally shown to be the most relevant for prediction

The Explainable AI (XAI) field was born

Explaining LP methods

- Given a prediction <Barack Obama, Nationality, >
- What are the relevant facts?
- A wide variety of methods in XAI that we can use to define relevance in the context of LP

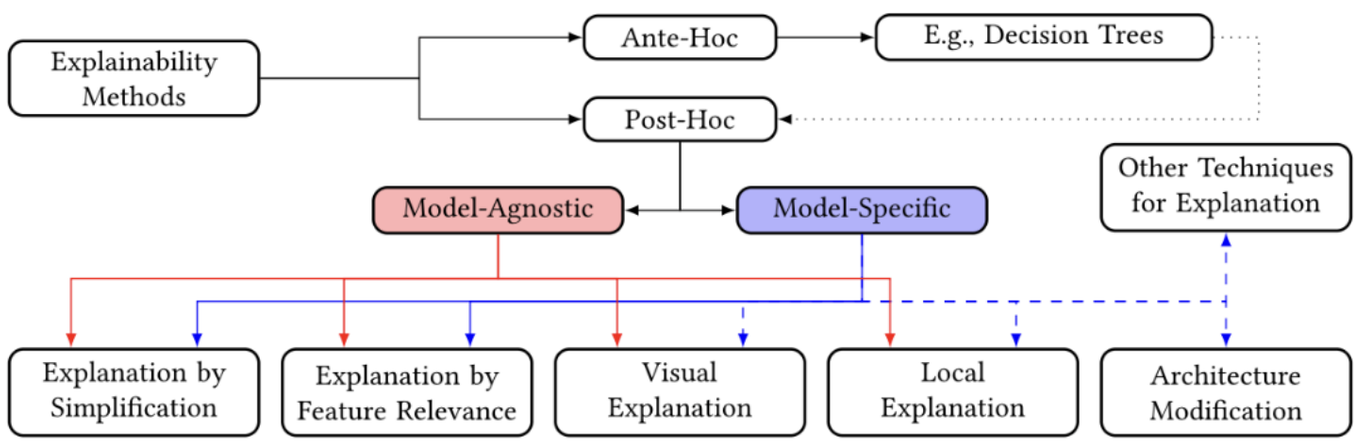
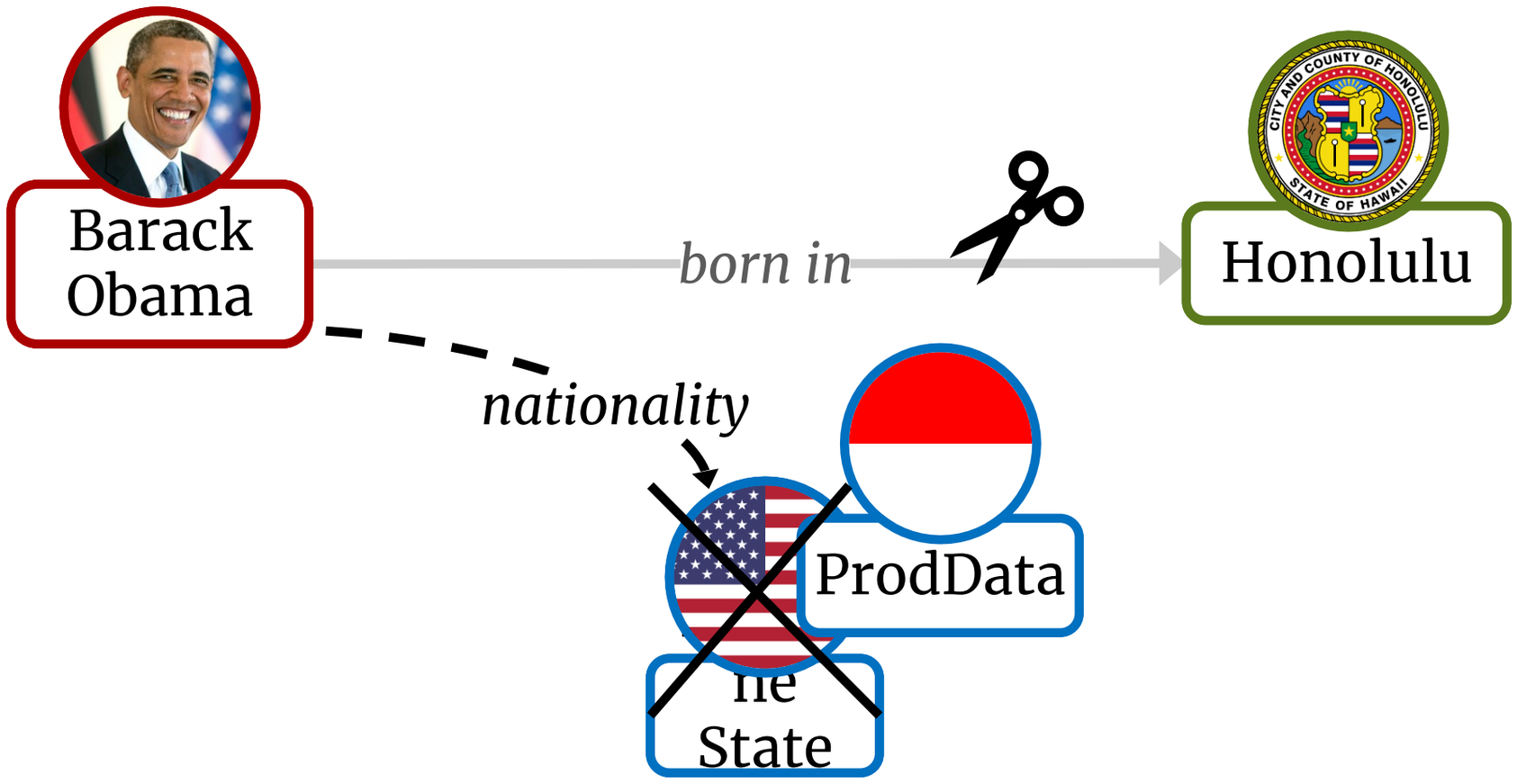


Image Source: Speith, Timo. "A review of taxonomies of explainable artificial intelligence (XAI) methods." Proceedings of the 2022 ACM conference on fairness, accountability, and transparency. 2022.

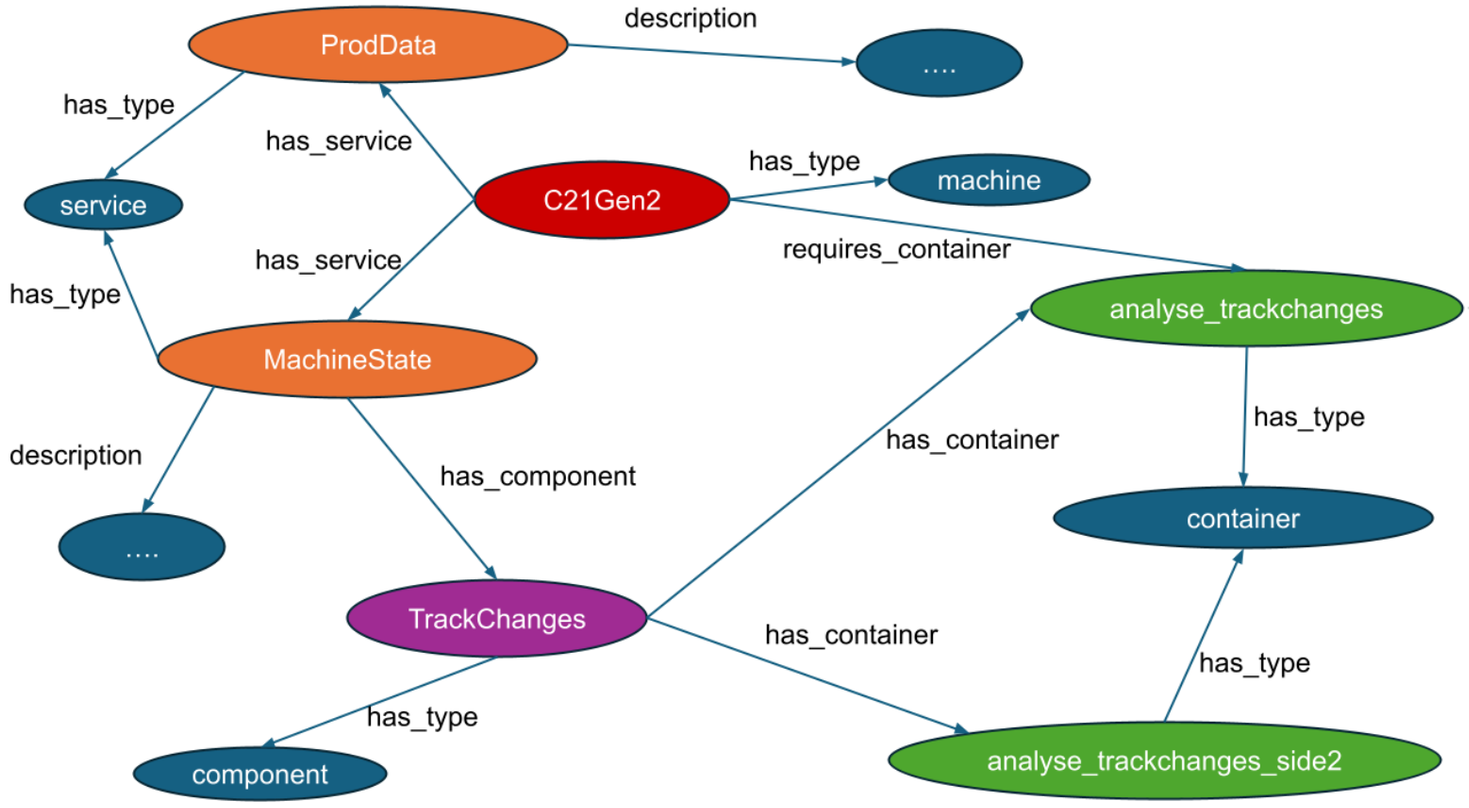
Necessary Explanation

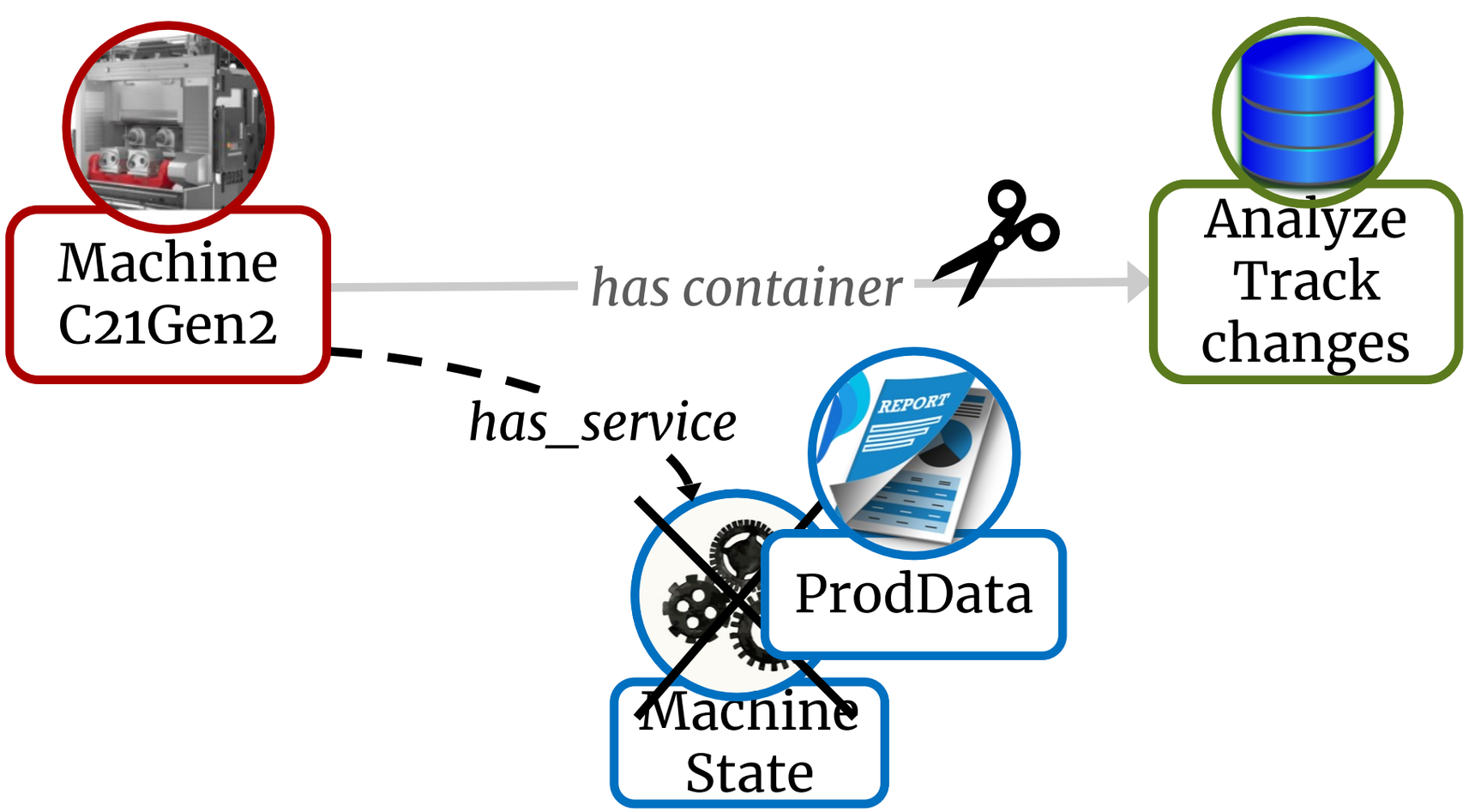
- Consider prediction <Barack Obama, Nationality, USA>
- Suppose that by removing <Barack Obama, Born_In, Honolulu> and recomputing embeddings, prediction for <Barack Obama, Nationality, ?> become Indonesia
- Then <Barack Obama, Born_In, Honolulu> is an example of necessary explanation



With Service data

- Our machine KG





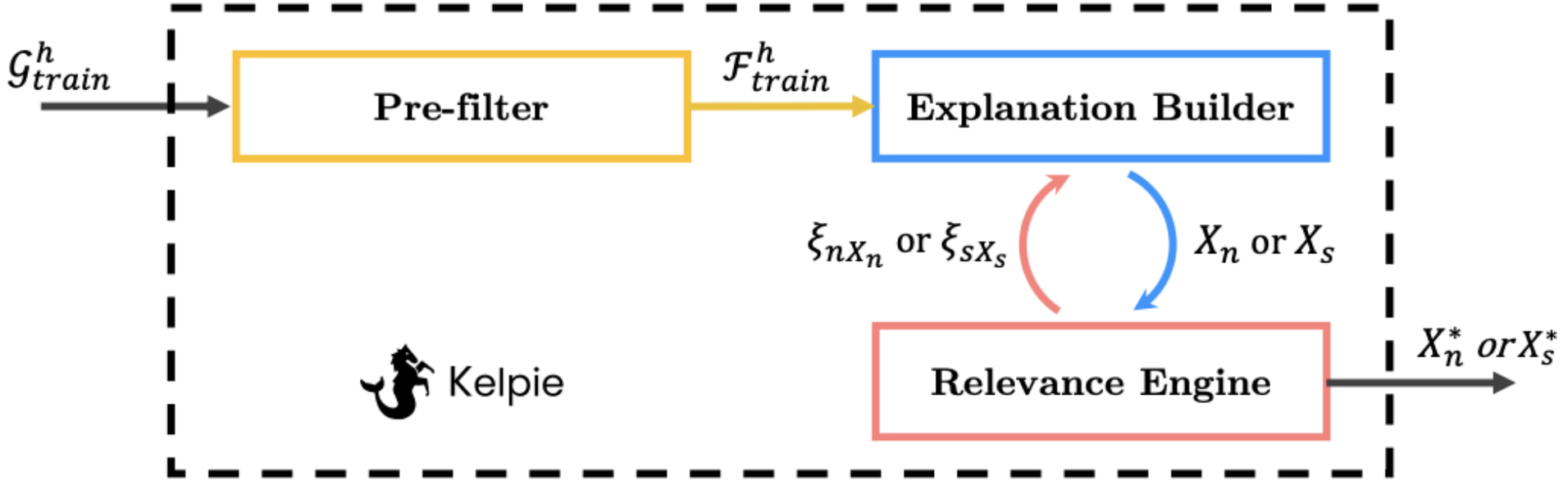
Sufficient Explanation

- Consider again <Barack Obama, Nationality, USA>
- Suppose that by adding facts such as <Emmanuel Macron, President_of, USA> and recomputing embeddings, predictions such as <Emmanuel Macron, Nationality, ?> systematically become USA
- Then <Barack Obama, President_of, USA> is an example of sufficient explanation

Kelpie

Knowledge Graph Embeddings for Link Prediction: Interpretable Explanations

- Given a tail prediction <h,r,t> Kelpie computes necessary and sufficient explanations
  - head predictions can be treated analogously



A. Rossi, D. Firmani, P. Merialdo, and T. Teofili. Explaining link prediction systems based on knowledge graph embeddings. In *SIGMOD*, 2022

Key intuition

- Given a tail prediction <h,r,t> Kelpie computes necessary and sufficient explanations

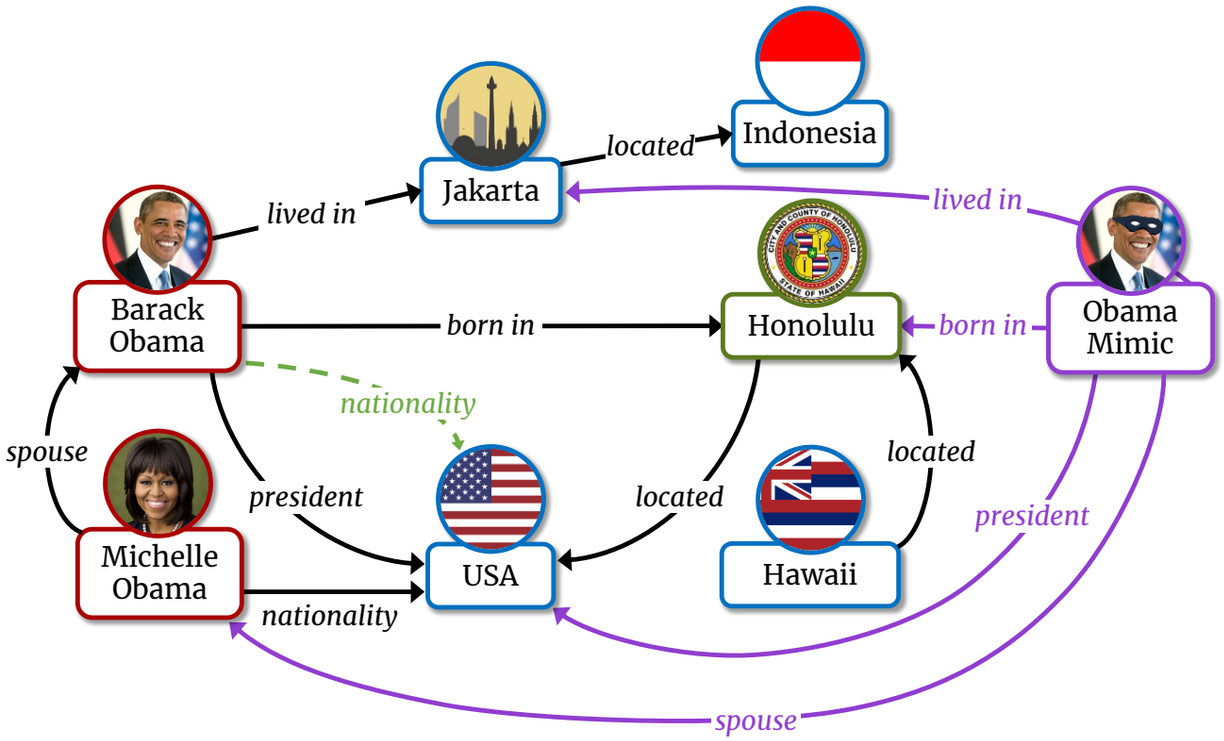
- Challenge: Perturbing the data and observing the changes in the model's predictions is a common approach in many applications (e.g., LIME, ANCHOR, SHAP, etc.)
  - in KGs you have to re-compute embeddings!

- Brute force solution: train from scratch upon every perturbation and re-compute all embeddinigs
  - Infeasible ☹

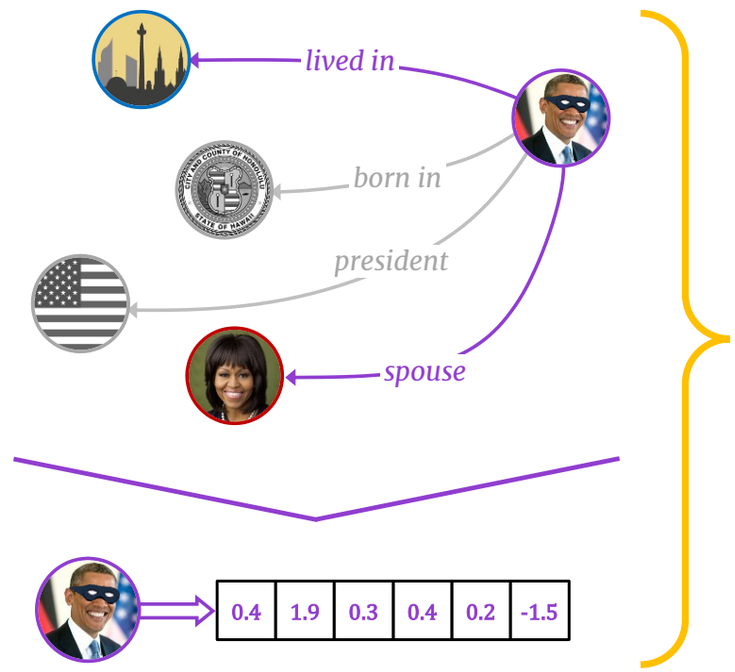
- Kelpie solution: post-training
  - generate one-time copies ("mimic") of the head and re-compute only their embeddings

Post-Training

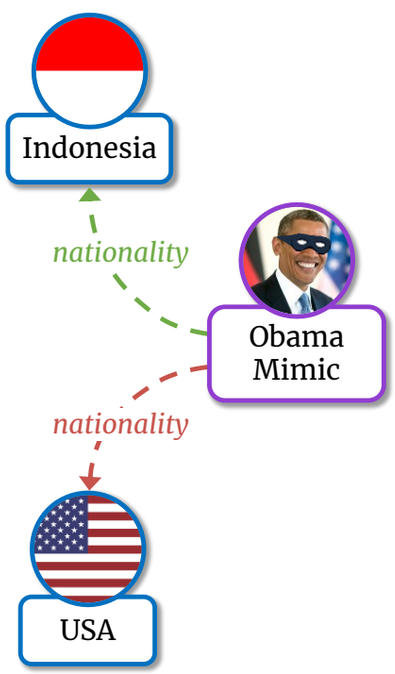
- Step 1: Freeze embeddings and create a mimic of the original head entity
- Step 2: Perturb incident facts and optimize the embedding of the mimic
  - Necessary explanations → Remove
  - Sufficient explanations → Add
- Step 3: Compare predictions of original and mimic



CREATE A MIMIC OF THE KNOWN ENTITY IN THE PREDICTION



ADD / REMOVE FACTS TO THE MIMIC AND POST-TRAIN ITS EMBEDDING



CHECK FOR CHANGES IN THE MIMIC PREDICTION

Limitations

- Change can be anything, whereas there is difference in changing to the second best or modifying completely the ranking

- Support Multiple explanations but Limited Interactivity

- Local explanations might miss large scale performance

Outline of this Tutorial

- Introduction to methods for LP
  - Two families of approaches
    - rule-based methods → Interpretable but sub-opt
    - embedding-based methods → Accurate but opaque
  - We focus on embedding-based
- Known challenges and research directions
- Explainable AI for Link Prediction
- Concluding remarks

Conclusions

- Discussed KGs completion methods and applied to emerging service-oriented applications
- Open directions
  - Incorporating LP models in computing platforms
  - Building benchmarks with service data
  - Evaluating bias and LLM performances in a no common-knowledge setup
  - Novel explanation methods to capture global patterns

DEPARTMENT
OF STATISTICAL SCIENCES

www.dss.uniroma1.it

Thanks for your Attention