CS11-711 Advanced NLP

Learning From/For Knowledge Bases

Daniel Fried

Carnegie Mellon University
Language Technologies Institute

Site
cmu-anlp.github.io

With slides by Graham Neubig and Zhengbao Jiang

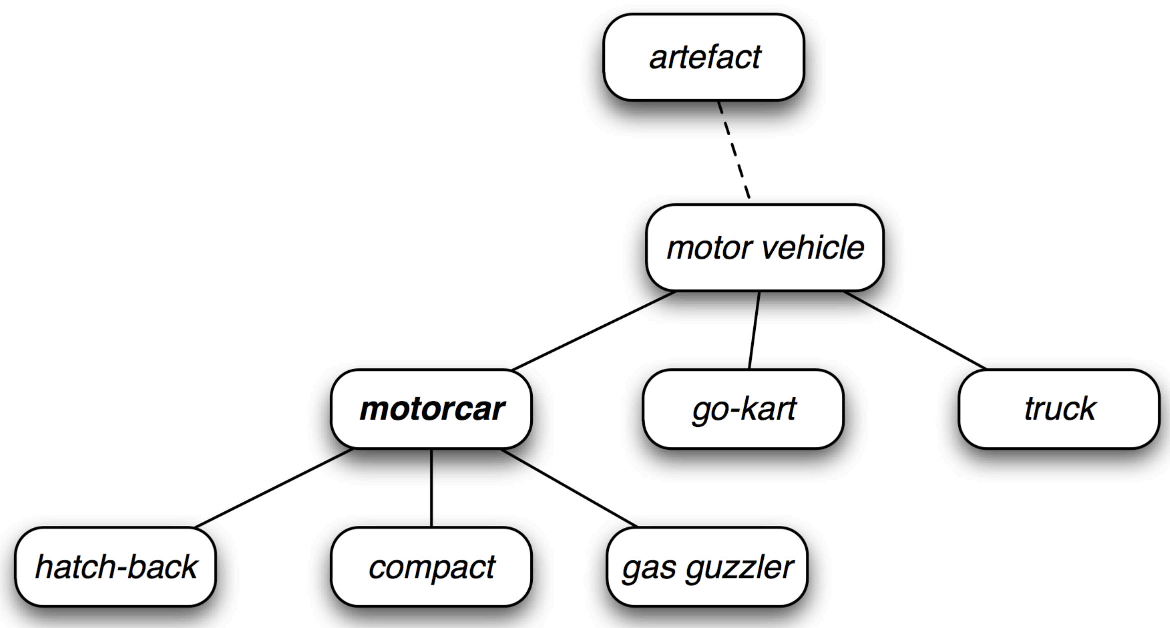
Knowledge Bases

- Structured databases of knowledge usually containing
  - Entities (nodes in a graph)
  - Relations (edges between nodes)
- How can we **learn to create/expand knowledge bases** with neural networks?
- How can we **learn from the information in knowledge bases** to improve neural representations?
- How can we use structured knowledge to answer questions (see also semantic parsing class)

Types of Knowledge Bases

WordNet (Miller 1995)

- WordNet is a large database of words including parts of speech, semantic relations



- Nouns: is-a relation (hatch-back/car), part-of (wheel/car), type/instance distinction
- Verb relations: ordered by specificity (communicate -> talk -> whisper)
- Adjective relations: antonymy (wet/dry)

Image Credit: NLTK

Cyc (Lenant 1995)

- A manually curated database attempting to encode all common sense knowledge, 30 years in the making

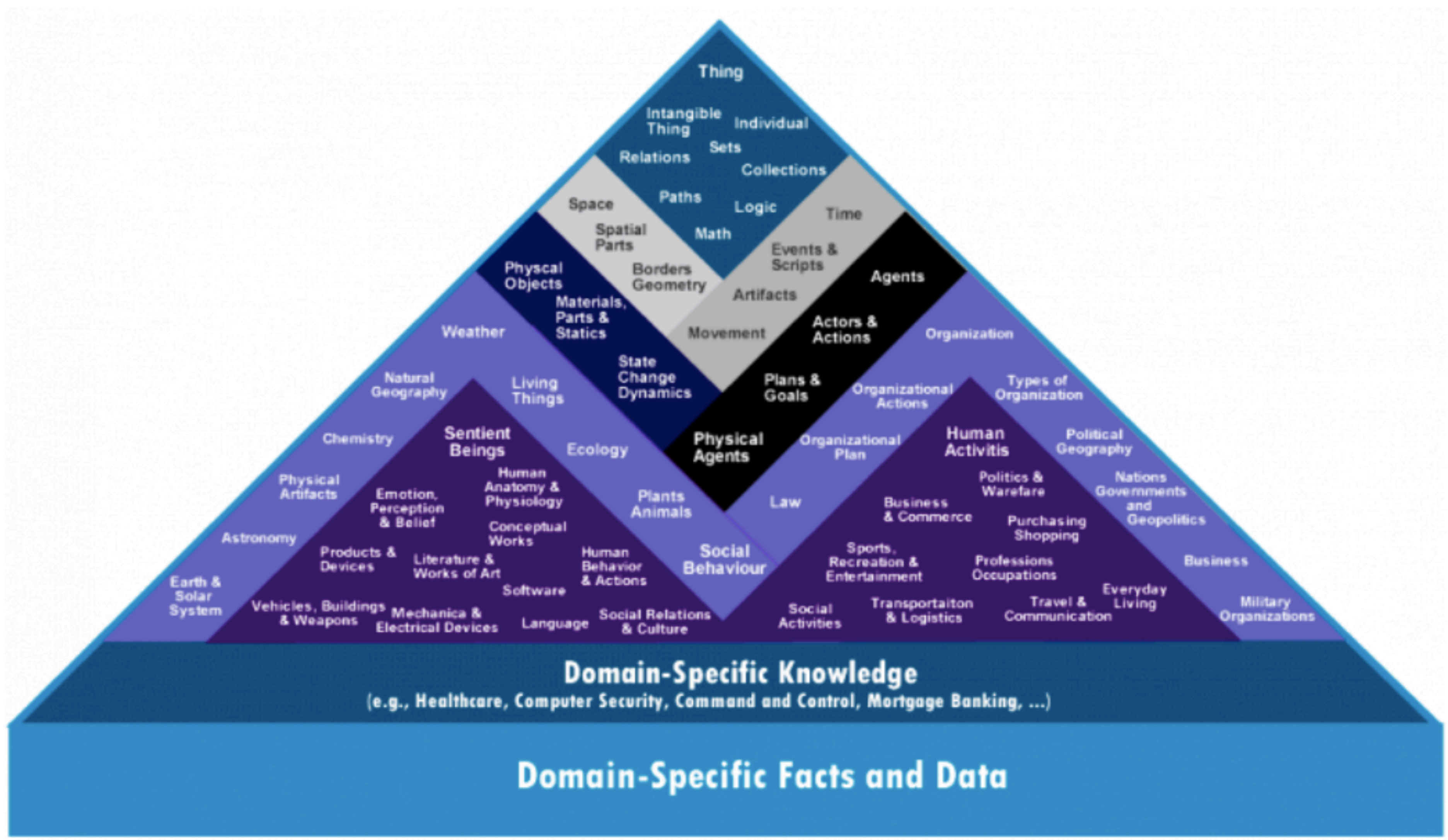
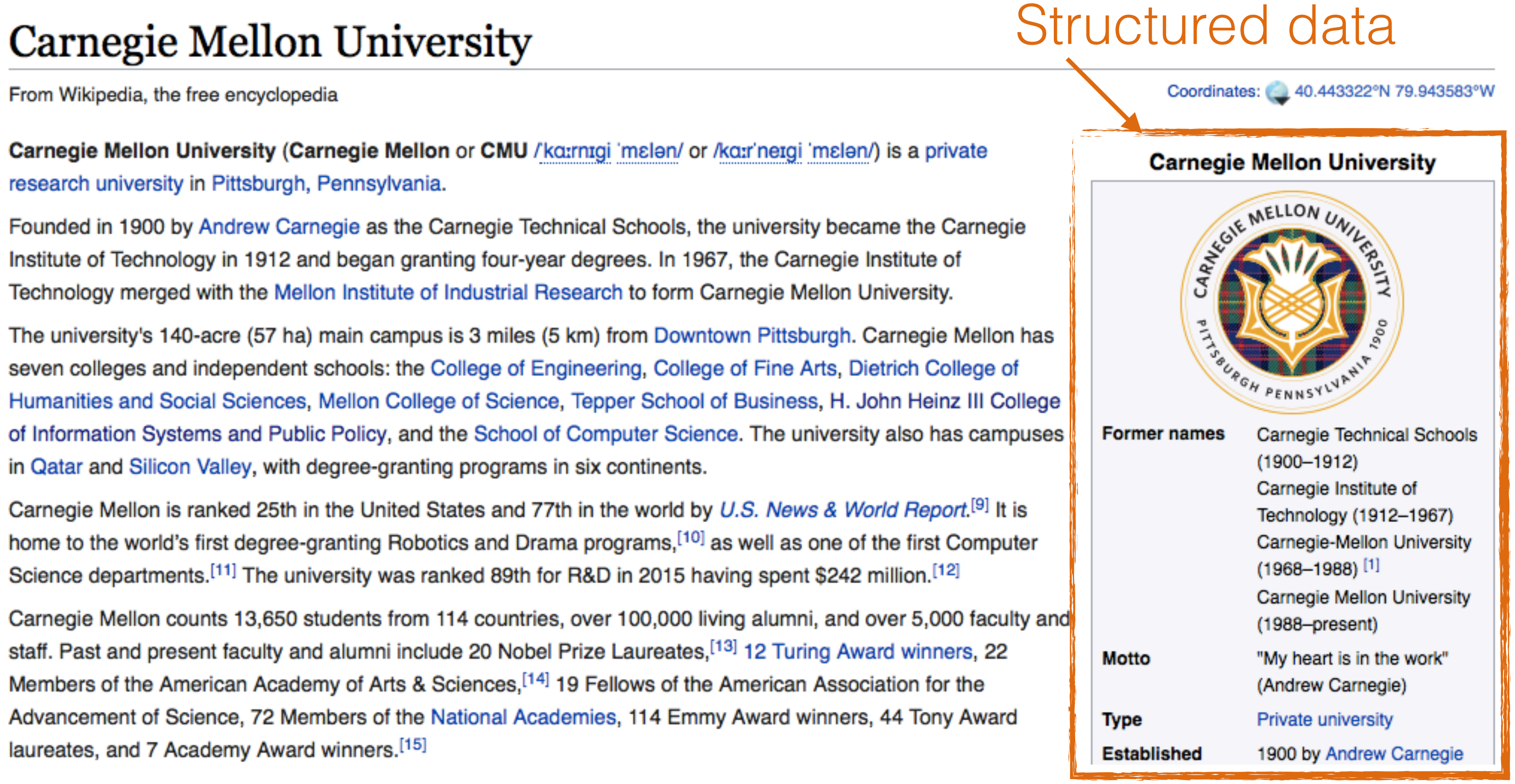


Image Credit: NLTK

DBPedia (Auer et al. 2007)

- Extraction of structured data from Wikipedia

Structured data



Carnegie Mellon University

From Wikipedia, the free encyclopedia

Coordinates: 40.443322°N 79.943583°W

Carnegie Mellon University (**Carnegie Mellon** or **CMU** /ˈkɑːrnɪɡi ˈmɛlən/ or /kɑːrˈneɪɡi ˈmɛlən/) is a private research university in Pittsburgh, Pennsylvania.

Founded in 1900 by Andrew Carnegie as the Carnegie Technical Schools, the university became the Carnegie Institute of Technology in 1912 and began granting four-year degrees. In 1967, the Carnegie Institute of Technology merged with the Mellon Institute of Industrial Research to form Carnegie Mellon University.

The university's 140-acre (57 ha) main campus is 3 miles (5 km) from Downtown Pittsburgh. Carnegie Mellon has seven colleges and independent schools: the College of Engineering, College of Fine Arts, Dietrich College of Humanities and Social Sciences, Mellon College of Science, Tepper School of Business, H. John Heinz III College of Information Systems and Public Policy, and the School of Computer Science. The university also has campuses in Qatar and Silicon Valley, with degree-granting programs in six continents.

Carnegie Mellon is ranked 25th in the United States and 77th in the world by *U.S. News & World Report*.[9] It is home to the world's first degree-granting Robotics and Drama programs,[10] as well as one of the first Computer Science departments.[11] The university was ranked 89th for R&D in 2015 having spent $242 million.[12]

Carnegie Mellon counts 13,650 students from 114 countries, over 100,000 living alumni, and over 5,000 faculty and staff. Past and present faculty and alumni include 20 Nobel Prize Laureates,[13] 12 Turing Award winners, 22 Members of the American Academy of Arts & Sciences,[14] 19 Fellows of the American Association for the Advancement of Science, 72 Members of the National Academies, 114 Emmy Award winners, 44 Tony Award laureates, and 7 Academy Award winners.[15]

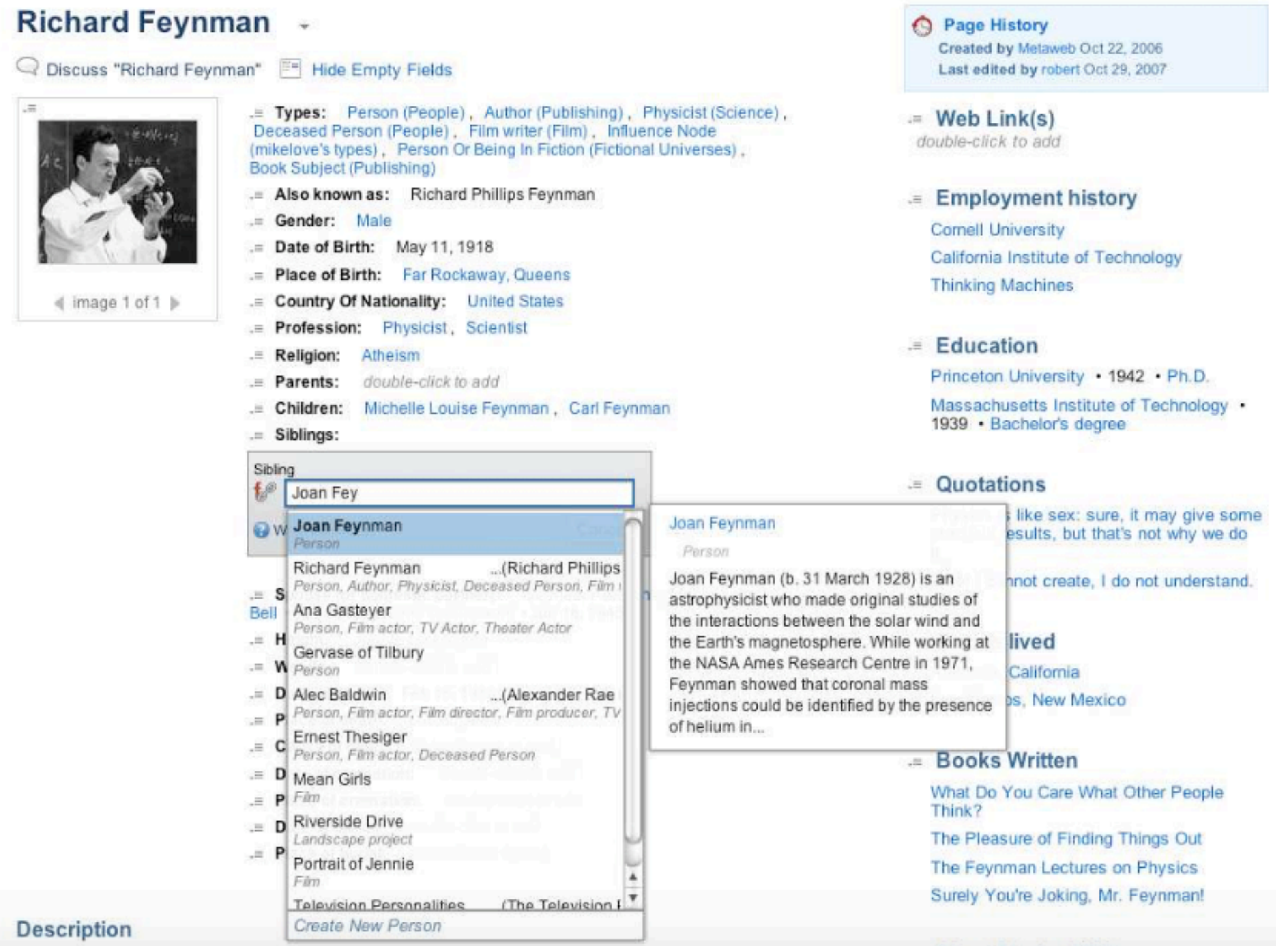
Carnegie Mellon University

Former names	Carnegie Technical Schools (1900–1912) Carnegie Institute of Technology (1912–1967) Carnegie-Mellon University (1968–1988) [1] Carnegie Mellon University (1988–present)
Motto	"My heart is in the work" (Andrew Carnegie)
Type	Private university
Established	1900 by Andrew Carnegie

- owl:Thing
- dul:Agent
- dul:SocialPerson
- wikidata:Q24229398
- wikidata:Q3918
- wikidata:Q43229
- dbo:Agent
- dbo:EducationalInstitution
- dbo:Organisation
- dbo:University
- geo:SpatialThing
- schema:CollegeOrUniversity
- schema:EducationalOrganization
- schema:Organization
- umbel-rc:Business
- umbel-rc:EducationalOrganization
- umbel-rc:Organization
- umbel-rc:University

WikiData (Bollacker et al. 2008)

- *Curated* database of entities, linked, and extremely large scale, multilingual



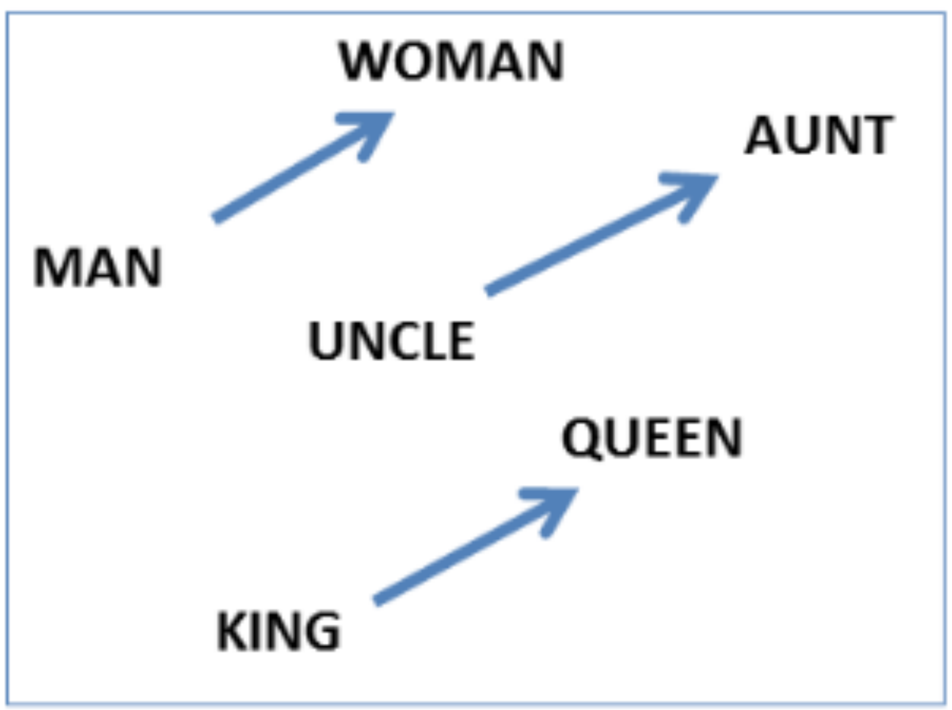
Learning Representations for Knowledge Bases

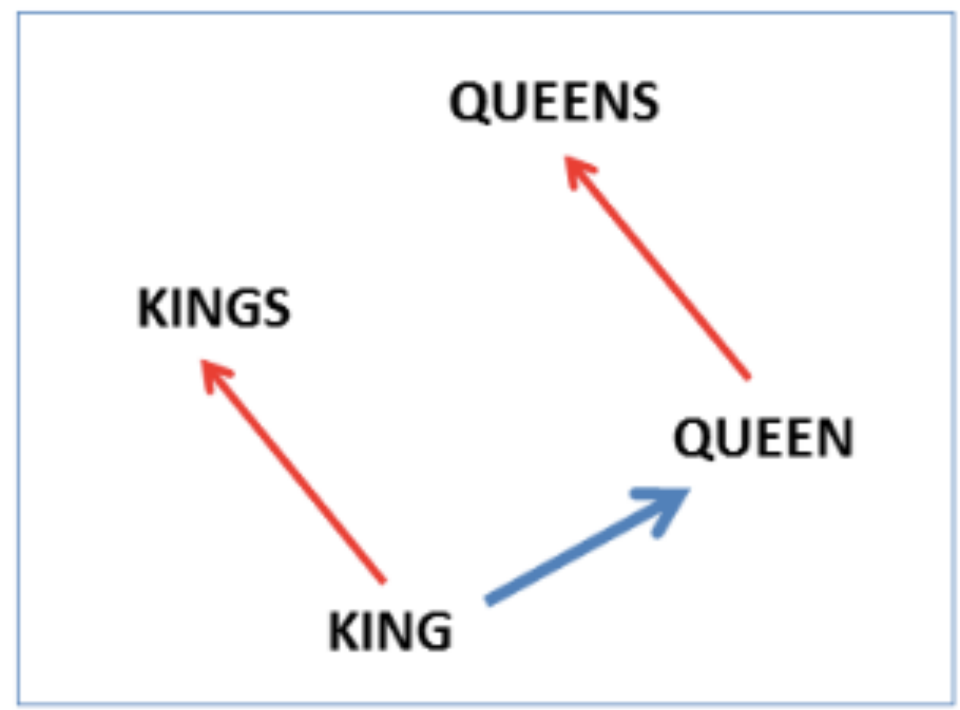
Knowledge Base Incompleteness

- Even w/ extremely large scale, knowledge bases are by nature incomplete
- e.g. in FreeBase 71% of humans were missing “date of birth” (West et al. 2014)
- Can we perform “relation extraction” to extract information for knowledge bases?

Consistency in Embeddings

e.g. king-man+woman = queen (Mikolov et al. 2013)



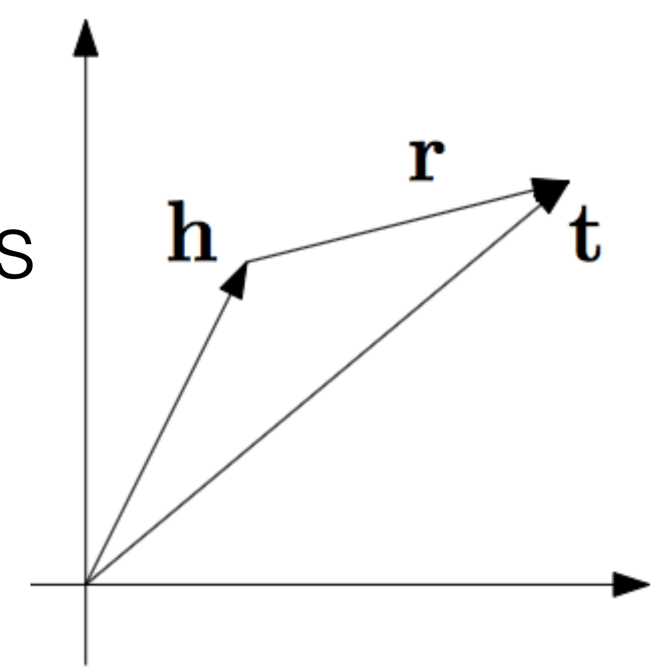


Learning Knowledge Graph Embeddings (Bordes et al. 2013)

- Motivation: express triples as additive transformation

- Method: minimize the distance of existing triples with a margin-based loss

$$\sum_{(h,\ell,t)\in S} \sum_{(h',\ell,t')\in S'_{(h,\ell,t)}} \left[\gamma + d(\boldsymbol{h}+\boldsymbol{\ell},\boldsymbol{t}) - d(\boldsymbol{h'}+\boldsymbol{\ell},\boldsymbol{t'})\right]_+$$



(a) TransE

Relation Extraction w/ Neural Tensor Networks (Socher et al. 2013)

- A first attempt at predicting relations: a multi-layer perceptron that predicts whether a relation exists

$$u_R^T f\left(W_{R,1}e_1 + W_{R,2}e_2\right)$$

- Neural Tensor Network: Adds bi-linear feature extractors, equivalent to projections in space

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

- Powerful model, but perhaps overparameterized!

Learning from Text Directly

Distant Supervision for Relation Extraction (Mintz et al. 2009)

- Given an entity-relation-entity triple, extract all text that matches this and use it to train

[Steven Spielberg]'s film [Saving Private Ryan] is loosely based on the brothers' story.

Allison co-produced the Academy Award-winning [Saving Private Ryan], directed by [Steven Spielberg]...

- Creates a large corpus of (noisily) labeled text to train a system

Relation Classification w/ Neural Nets (Zeng et al. 2014)

- Extract features and classify
  - Lexical features of the entities themselves
  - Features of the whole span

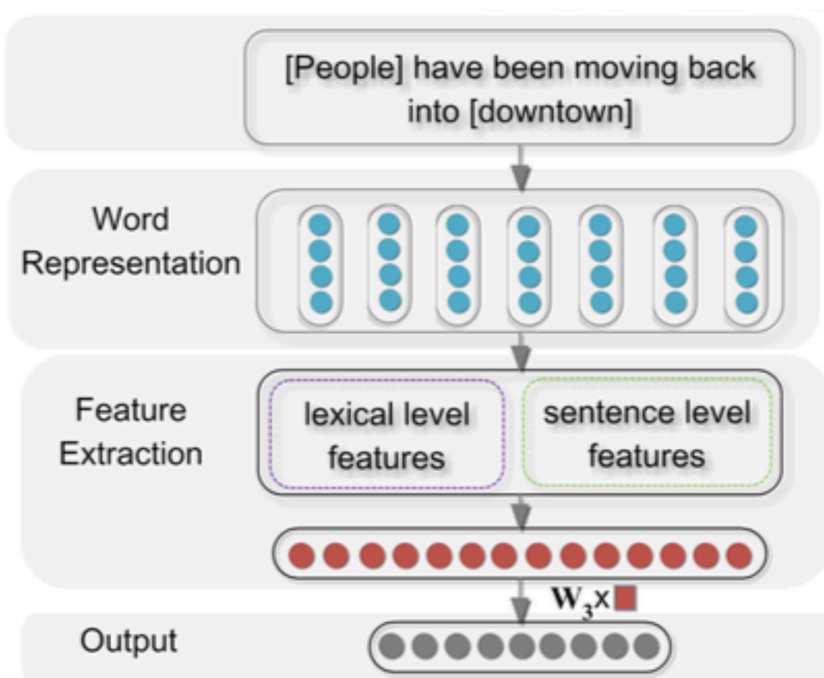


Figure 1: Architecture of the neural network used for relation classification.

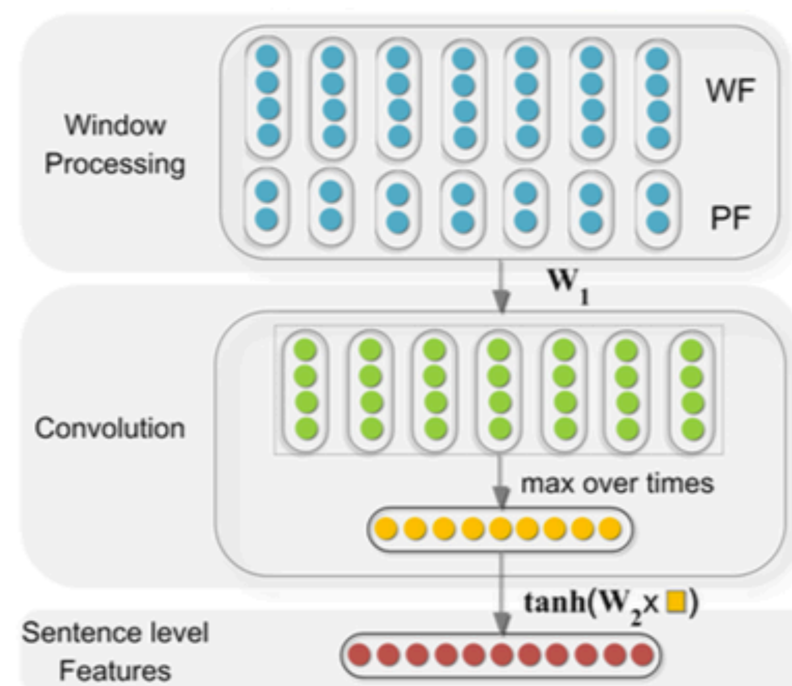


Figure 2: The framework used for extracting sentence level features.

QA on Tables and Knowledge Bases

Semantic Parsing

- Parse questions to logical forms which can be executed on a table or database
- Representative approaches: Wong and Mooney 2007; Zettlemoyer and Collins 2007; Liang et al. 2011
- See https://github.com/allenai/acl2018-semantic-parsing-tutorial for a nice overview

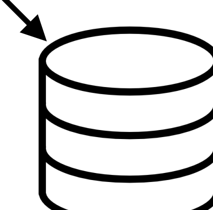
What	states	border	Texas
$(S/(S\backslash NP))/N$	N	$(S\backslash NP)/NP$	NP
$\lambda f.\lambda g.\lambda x.f(x) \wedge g(x)$	$\lambda x.state(x)$	$\lambda x.\lambda y.borders(y,x)$	$texas$

$$\xrightarrow{>}\; S/(S\backslash NP) \quad \lambda g.\lambda x.state(x) \wedge g(x) \qquad\qquad \xrightarrow{>}\; (S\backslash NP) \quad \lambda y.borders(y, texas)$$

$$\xrightarrow{>}\; S \quad \lambda x.state(x) \wedge borders(x, texas)$$

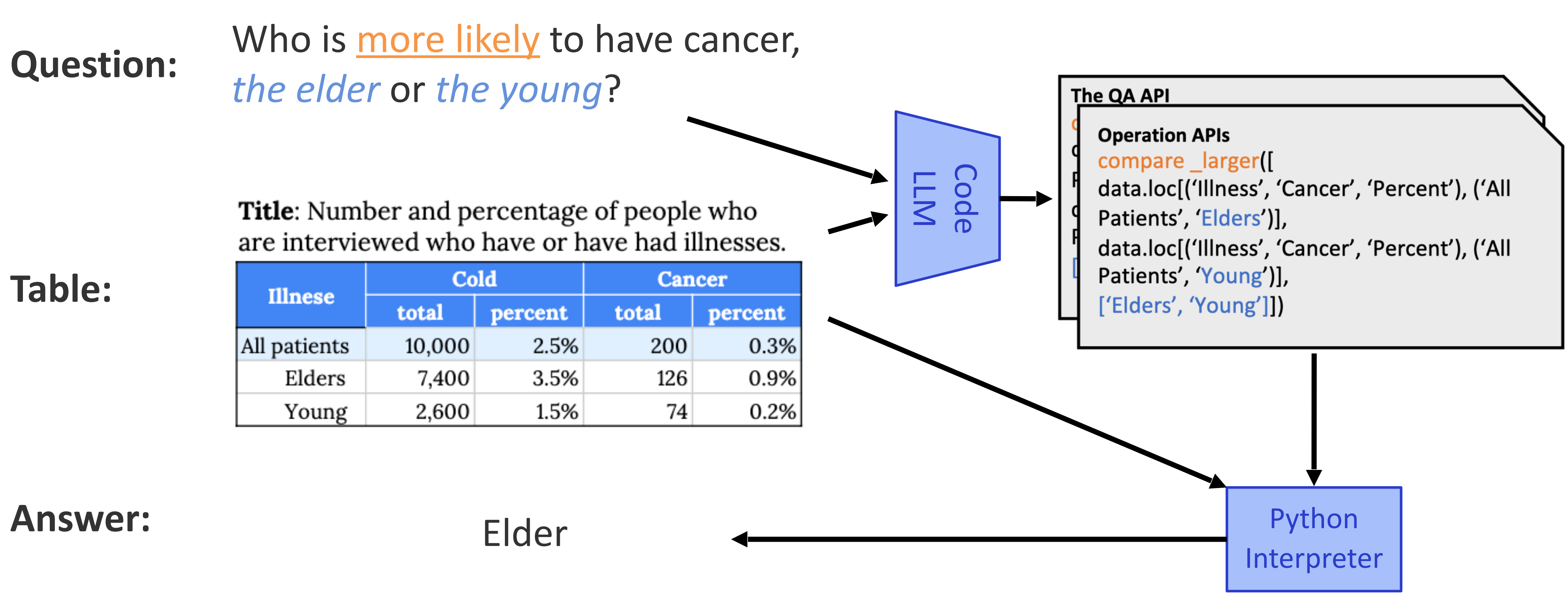
How many people live in Seattle?

```
SELECT Population from CityData
where City=="Seattle"
```



Code LLMs for Semantic Parsing

- Rather than train a language->query model, use a code LLM.
- Representative approaches: Scholak et al. 2021, Shin and Van Durme 2021, Cheng et al. 2023



Illnese	Cold		Cancer	
	total	percent	total	percent
All patients	10,000	2.5%	200	0.3%
Elders	7,400	3.5%	126	0.9%
Young	2,600	1.5%	74	0.2%

Cao et al. 2023 (ANLP project!)

Hybrid (“Neuro-Symbolic”) QA Approaches

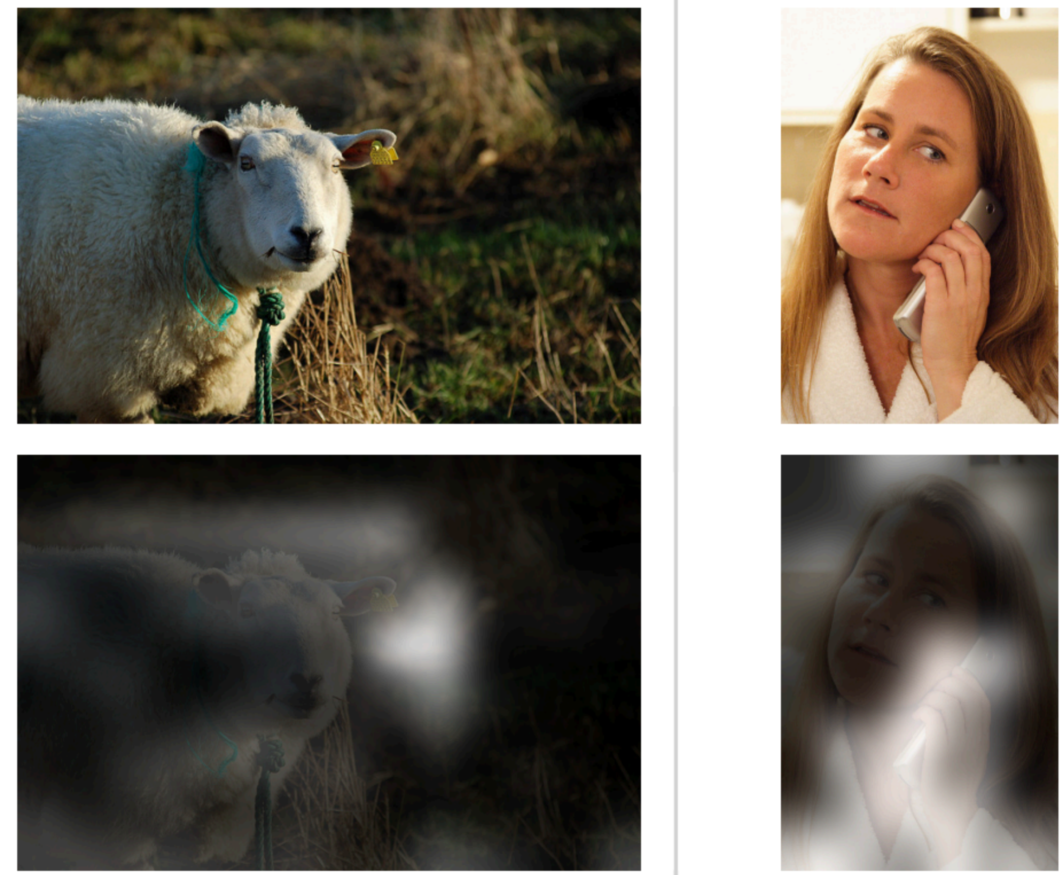
Machine Reading with Symbolic Operations

- Can we explicitly incorporate numerical reasoning in machine reading?
- e.g. DROP dataset (Dua et al. 2019)

Reasoning	**Passage** (some parts shortened)	**Question**	**Answer**	**BiDAF**
Subtraction (28.8%)	That year, his **Untitled (1981)**, a painting of a haloed, black-headed man with a bright red skeletal body, depicted amid the artists signature scrawls, was **sold by Robert Lehrman for \$16.3 million, well above its \$12 million high estimate**.	How many more dollars was the Untitled (1981) painting sold for than the 12 million dollar estimation?	4300000	\$16.3 million
Comparison (18.2%)	In **1517, the seventeen-year-old King sailed to Castile**. There, his Flemish court .... **In May 1518, Charles traveled to Barcelona in Aragon**.	Where did Charles travel to first, Castile or Barcelona?	Castile	Aragon

Neural Module Networks

- Idea: semantic parsing, but the execution model is neural
- Have typically underperformed pre-trained end-to-end models, but using code LLMs are a promising avenue (more in a later lecture!)



What is in the sheep's ear?

```
(describe[what]
    (and find[sheep]
         find[ear]))
```

tag

What color is she wearing?

```
(describe[color]
    find[wear])
```

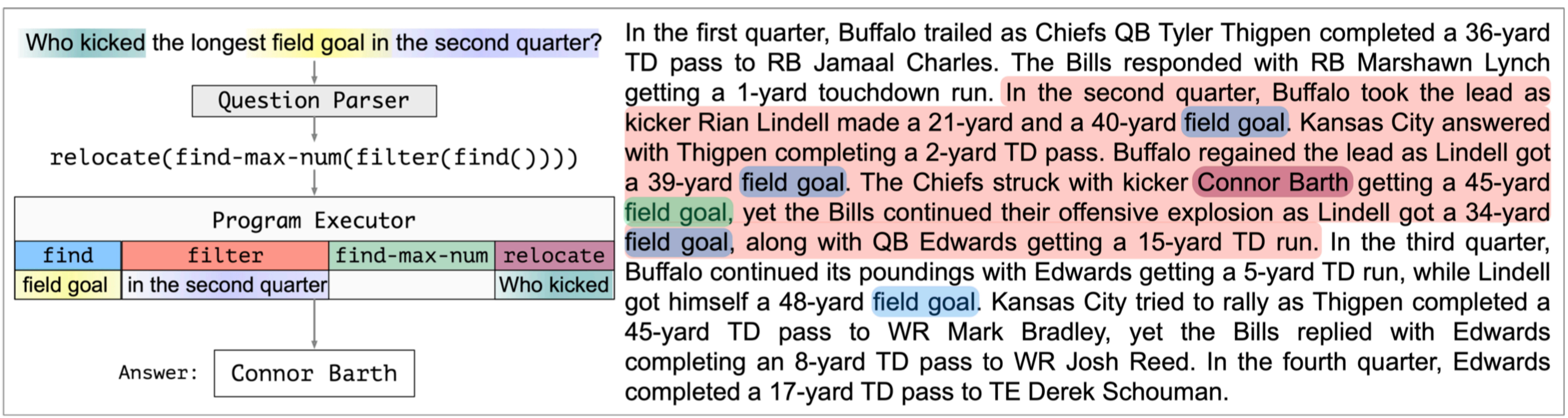
white

Andreas et al. (2016)

Solving Word Problems w/ Symbolic Reasoning

- Idea: NMNs for text combine semantic parsing (with explicit functions) and machine reading

- e.g. Gupta et al. (2020)



Using Knowledge Bases to Inform Neural Models

Retrofitting of Embeddings to Existing Lexicons (Faruqui et al. 2015)

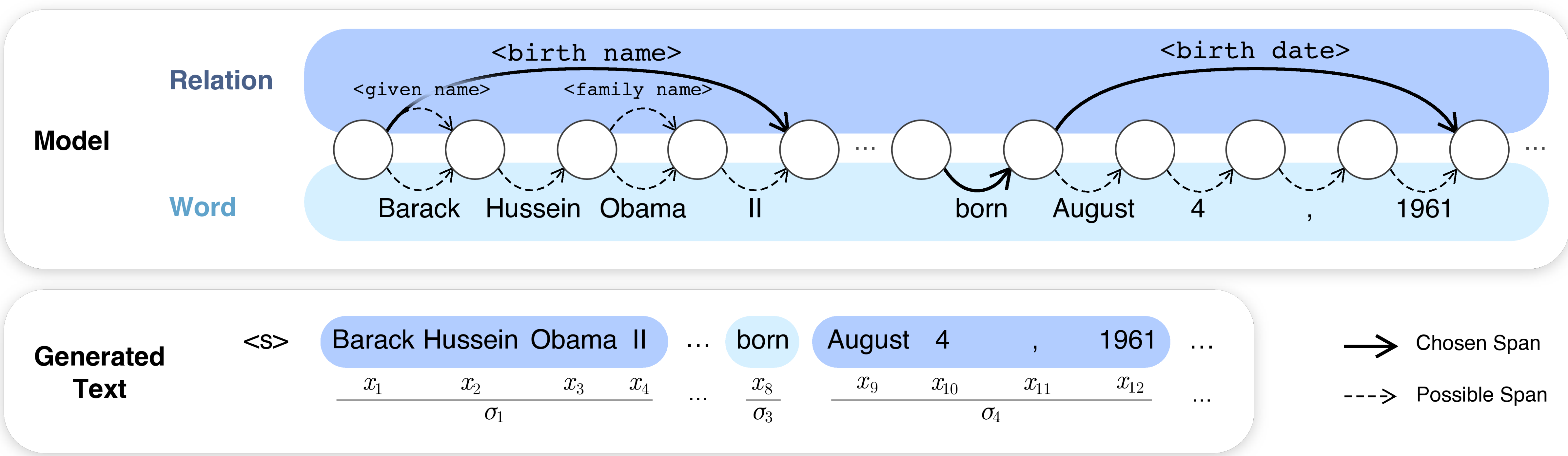
- Post-hoc transformation of embeddings, informed by relations in a Knowledge Base (e.g. WordNet)
  - Advantage of being usable with any pre-trained embeddings
- Double objective of making transformed embeddings close to neighbors, and close to original embedding

$$\Psi(Q) = \sum_{i=1}^{n} \left[\alpha_i \|q_i - \hat{q}_i\|^2 + \sum_{(i,j)\in E} \beta_{ij} \|q_i - q_j\|^2 \right]$$

- Can also force antonyms away from each-other (Mrksic et al. 2016)

Injecting Knowledge into Language Models (Hayashi et al. 2020)

- Provide LMs with topical knowledge in the form of copiable graphs
  - Each (Wiki) text is given relevant KB taken from Wikidata
- Examine all possible decoding "paths" and maximize the marginal probability



Schema-Free Extraction

Open Information Extraction
(Banko et al 2007)

- Basic idea: **the text is the relation**
- e.g. "United has a hub in Chicago, which is the headquarters of United Continental Holdings"
  - {United; has a hub in; Chicago}
  - {Chicago; is the headquarters of; United Continental Holdings}
- Can extract any variety of relations, but does not abstract

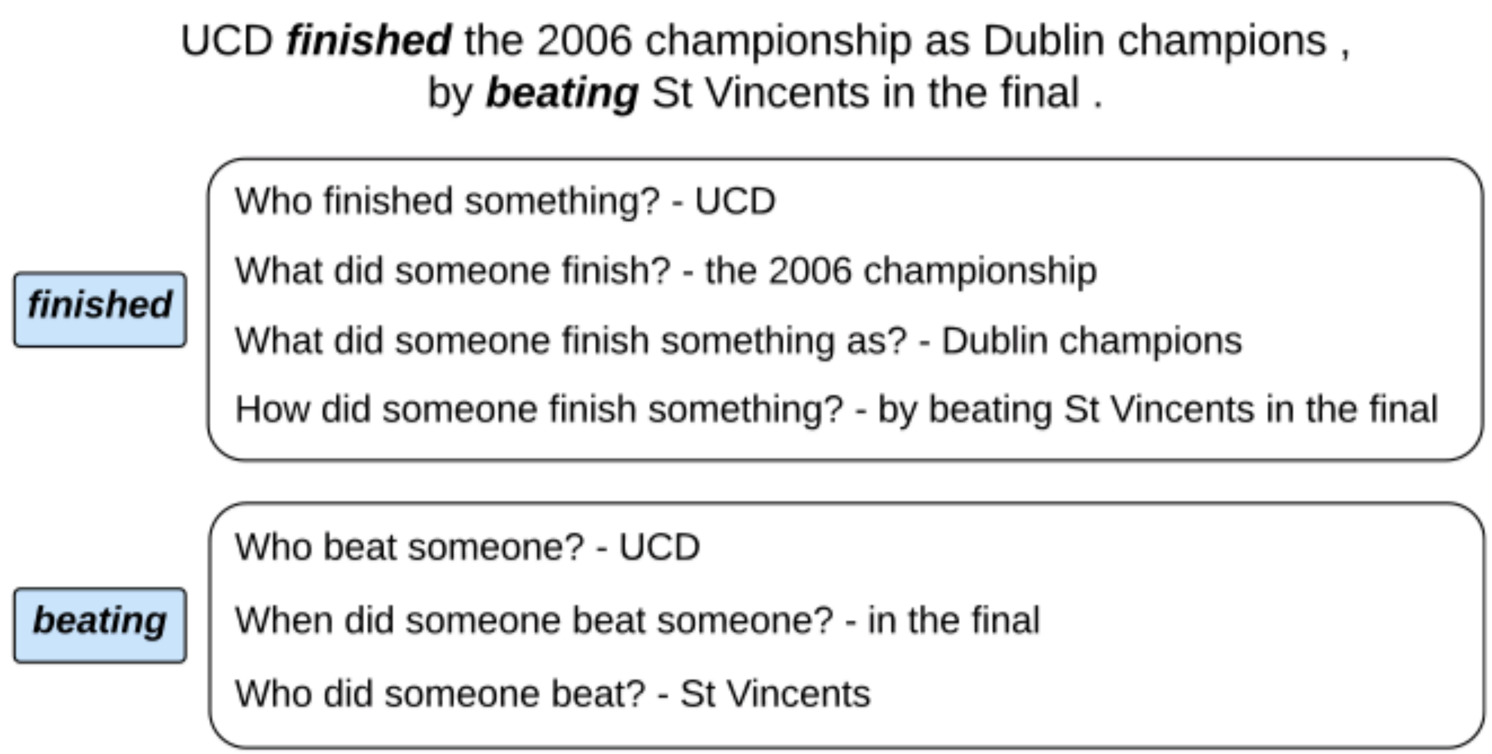
Rule-based Open IE

- e.g. TextRunner (Banko et al. 2007), ReVerb (Fader et al. 2011)
- Use parser to extract according to rules
    - e.g. relation must contain a predicate, subject object must be noun phrases, etc.
- Train a fast model to extract over large amounts of data
- Aggregate multiple pieces of evidence (heuristically) to find common, and therefore potentially reliable, extractions

Crowdsourcing + Neural Models for Open IE

- Unfortunately, heuristics are still not perfect

- Possible to create relatively large datasets by asking simple questions (QA-SRL; He et al. 2015):



- Can be converted into OpenIE extractions, for use in supervised neural BIO tagger (Stanovsky et al. 2018)

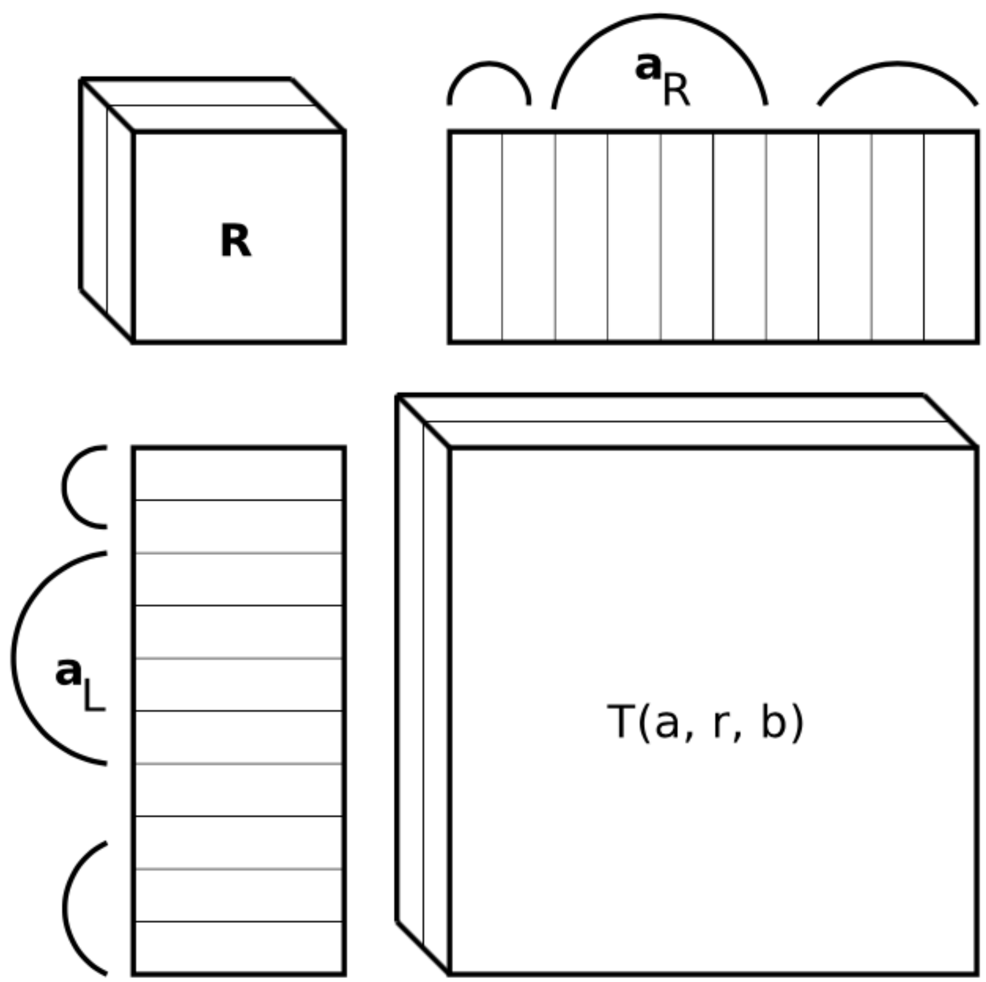
Learning Relations from Relations

Modeling Word Embeddings vs. Modeling Relations

- Word embeddings give information of the word in context, which is indicative of KB traits
- However, other relations (or combinations thereof) are also indicative
    - This is a *link prediction* problem in graphs

Tensor Decomposition
(Sutskever et al. 2009)

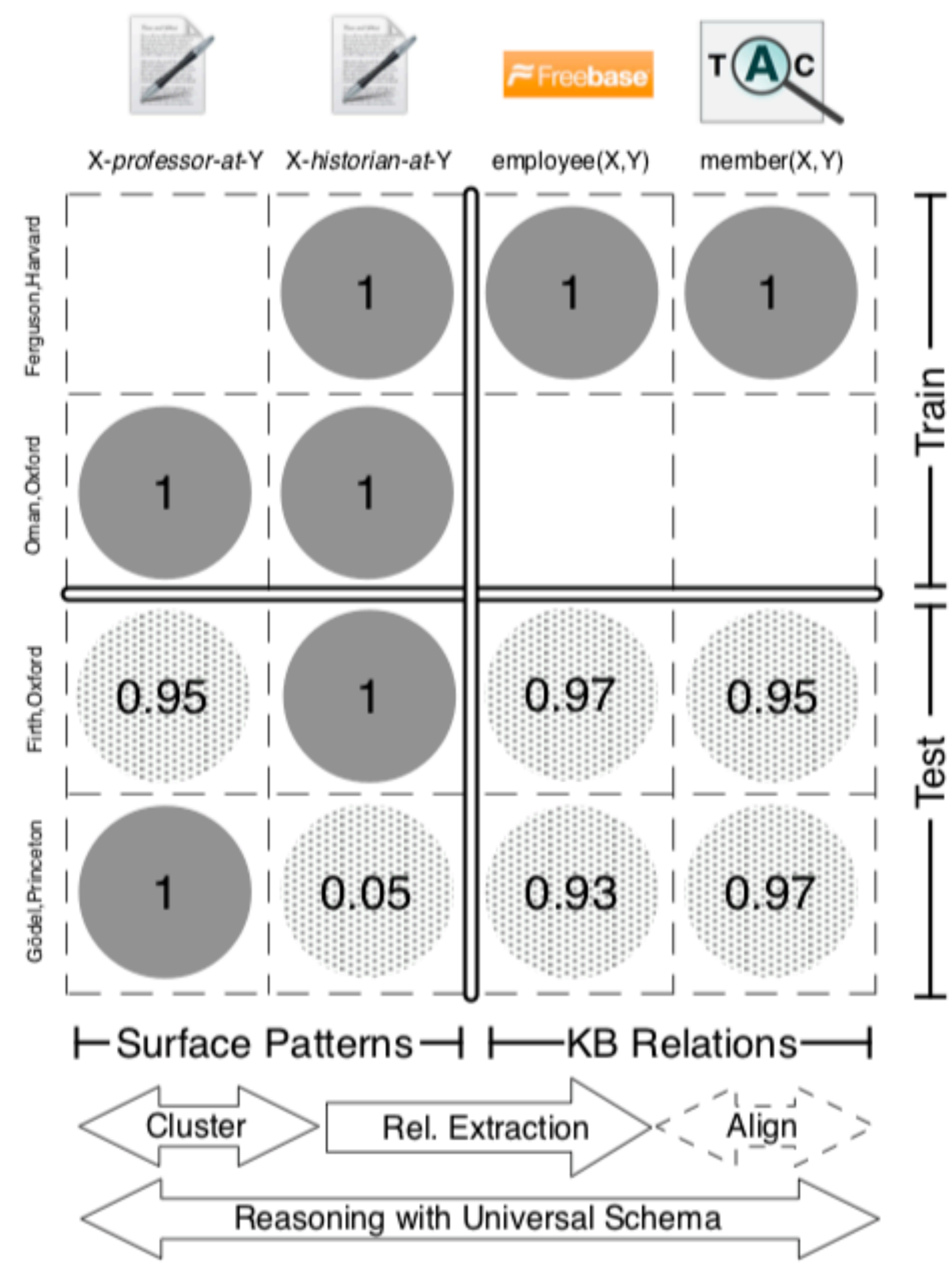
- Can model relations by decomposing a tensor containing entity/relation/entity tuples



Matrix Factorization to Reconcile Schema-based and Open IE Extractions
(Riedel et al. 2013)

- What to do when we have a knowledge base, and text from OpenIE extractions?

- **Universal schema:** embed relations from multiple schema in the same space



Probing Knowledge in LMs

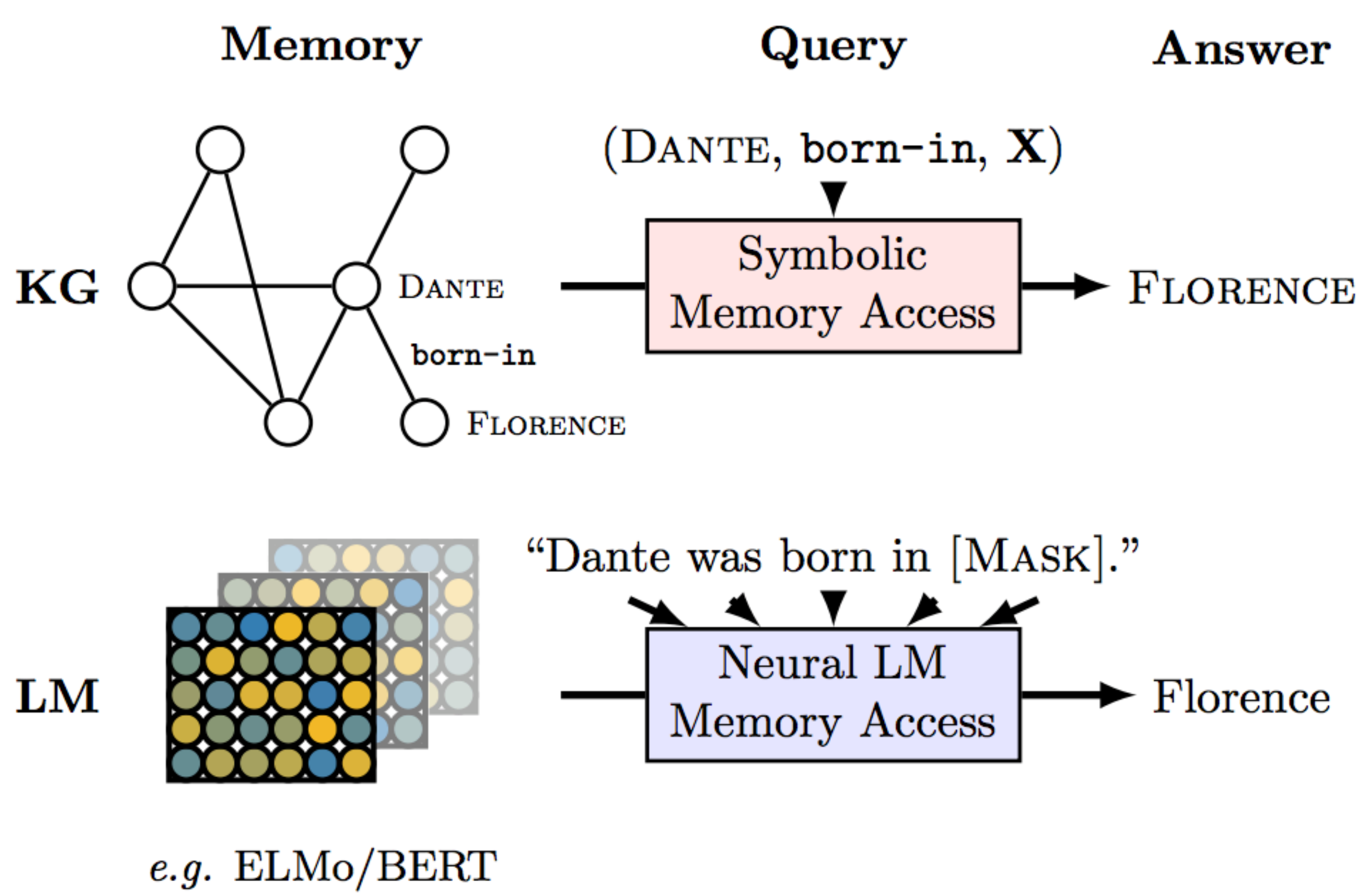
Probing Knowledge in LMs

- Traditional QA/MRC models usually refer to external resources to answer questions, e.g., Wikipedia articles or KGs.
- Do LMs pre-trained on a large text corpus already capture those knowledge?

LMs as KBs?

(Petroni et al. 2019)

- Structured queries (e.g., SQL) to query KBs.
- Natural language prompts to query LMs.



LMs as KBs?
(Petroni et al. 2019)

- LAMA benchmark
  - Manual prompts for 41 relations: “[X] was founded in [Y].”
  - Fill in subjects and have LMs (e.g., BERT) predict objects: “Bloomberg L.P. was founded in [MASK].”
  - Accuracy: ELMo 7.1%, Transformer-XL 18.3%, BERT-base 31.1%

Mask 1 Predictions:

5.2% **Chicago**

4.1% **London**

2.8% **Toronto**

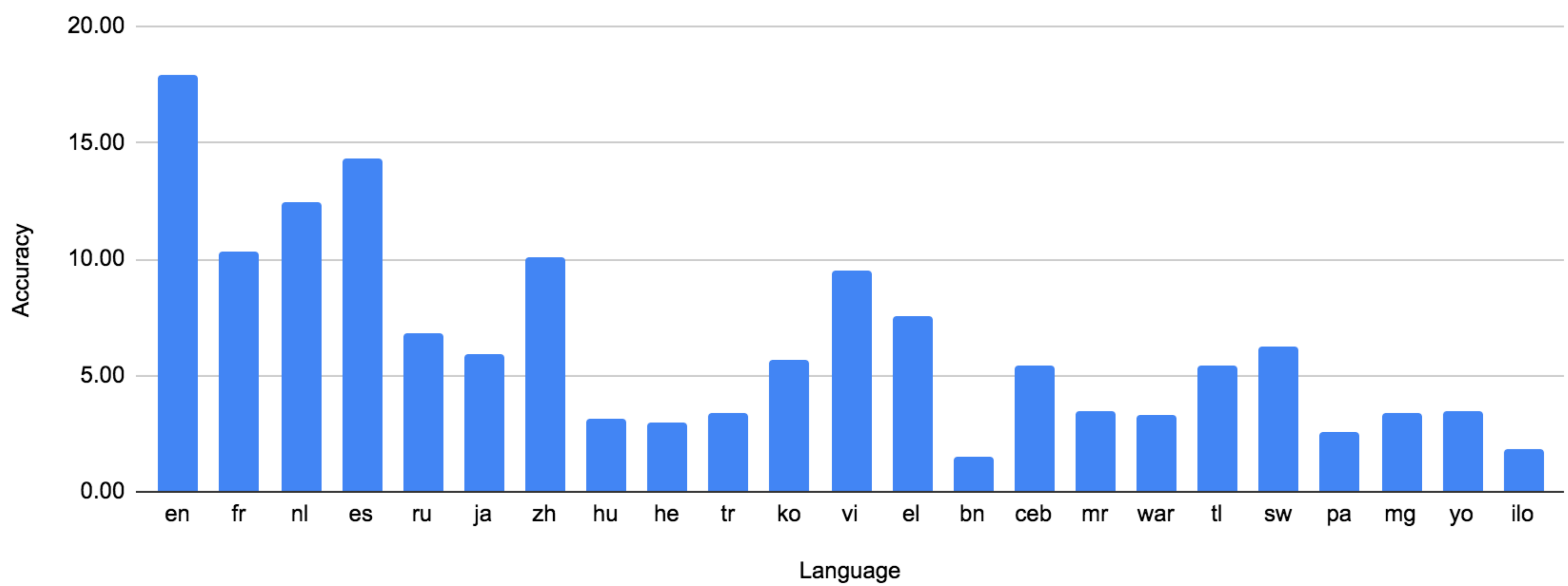
2.3% **c**

1.6% **India**

https://demo.allennlp.org/masked-lm/s/bloomberg-lp-was-founded-mask/I5Q1P2T5Z0

X-FACTR: Multilingual Factual Knowledge Probing (Jiang et al. 2020)

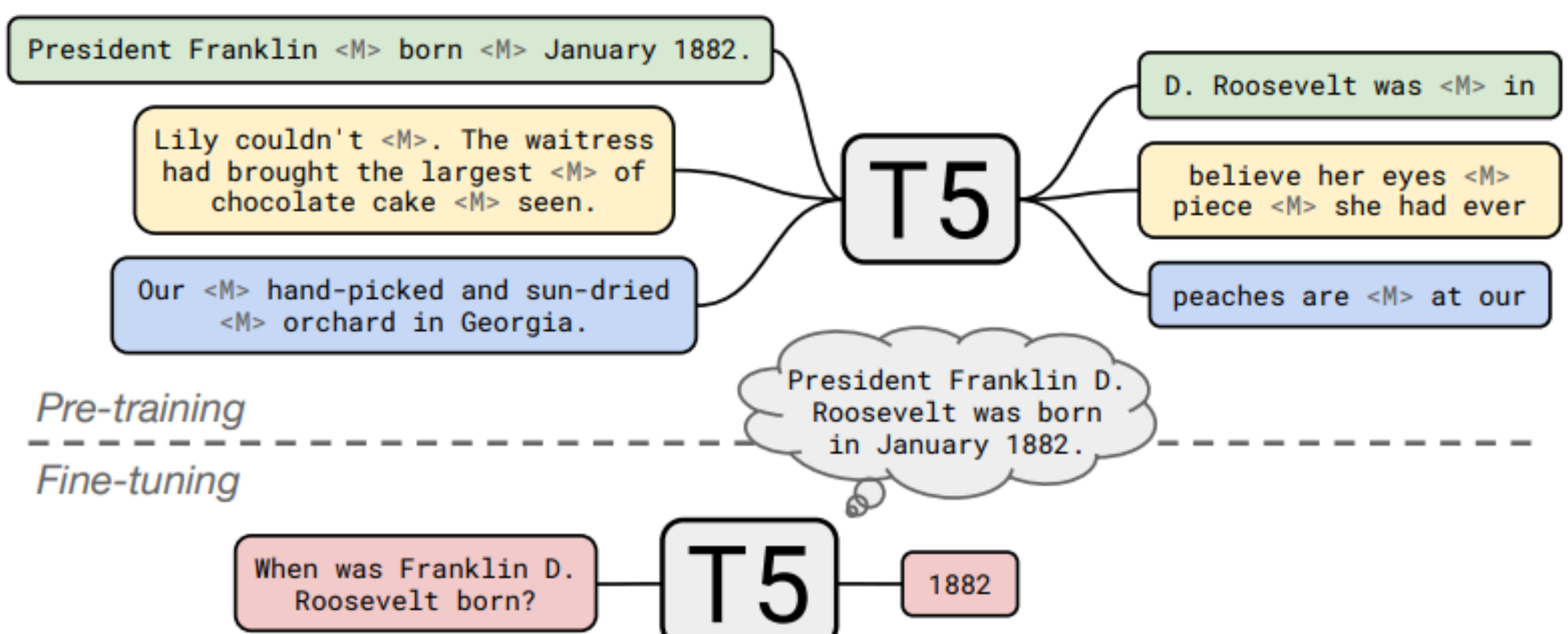
- Overall, factual knowledge in LMs is still limited, especially for low-resource languages.



Max performance of M-BERT, XLM, XLM-R

Close-book T5: Directly Fine-tune with QA Pairs (Roberts et al. 2020)

- Generate answers given questions without additional context.
- Underperforms retrieval-based models, but shows there is a lot of knowledge in LLMs

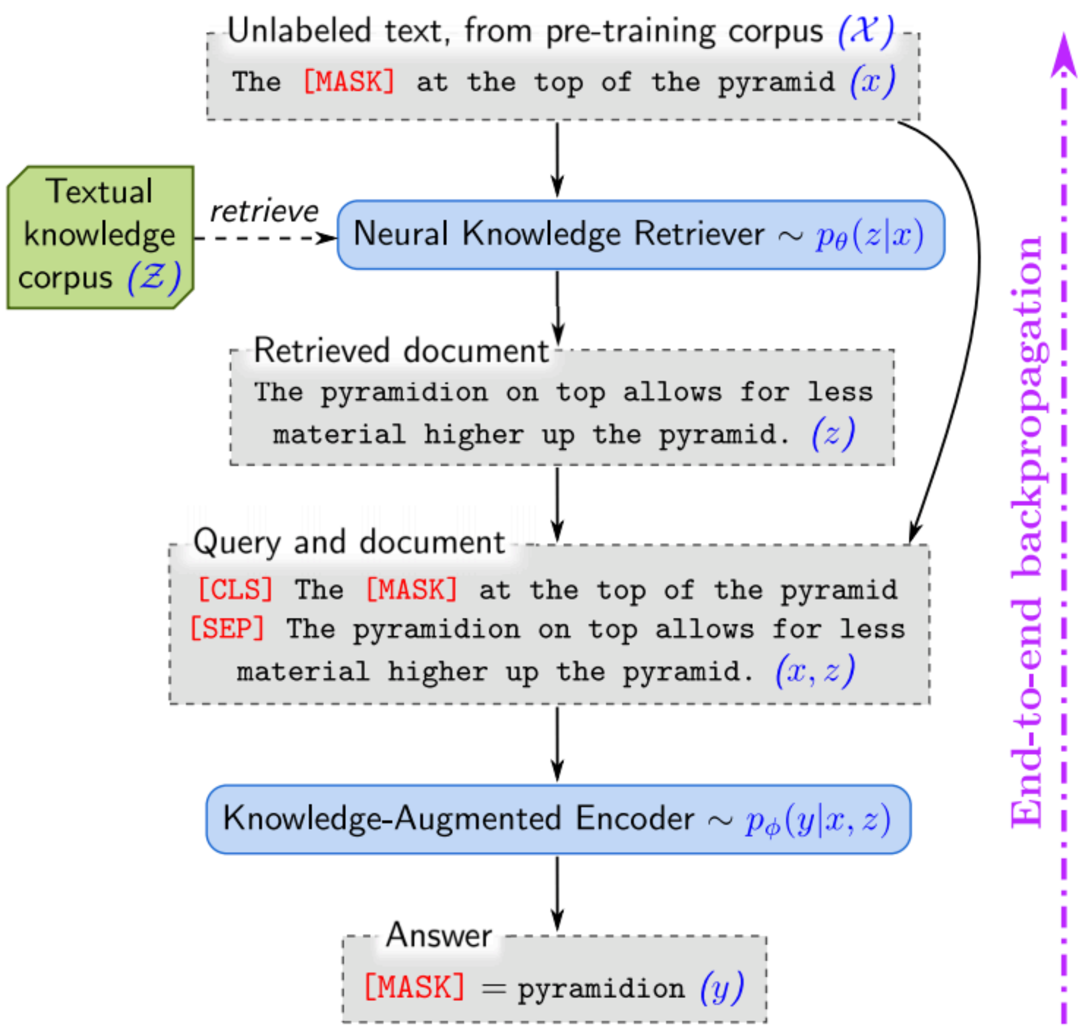


Nonparametric Models Outperform Parametric Models

- For knowledge-intensive tasks like QA, nonparametric models (w/ retrieved context) outperform parametric models (w/o context) by a large margin.

- For example, REALM (Guu et al. 2020), RAG (Lewis et al. 2020) on the NaturalQuestion datasets.

Close-book T5	34.5
REALM	40.4
RAG	44.5



Questions?