Knowledge Representation and Extraction at Scale

Christos Christodoulopoulos

Amit Gurdasani, Andrew Hopkinson, Arpit Mittal, Dave Palfrey, James Thorne, Andreas Vlachos

Amazon Research, Cambridge UK

SemDeep-3 @ COLING

20 August 2018

research cambridge

Amazon Alexa

Entertainment: Music, Books, Video…

Household: Timers, ToDo Lists, Shopping, Calendar…

Information: Weather, Traffic, News, Q/A…

Smart Home: Lights, Thermostats, Switches…

45,000+ skills: Developed by 3rd Parties

Amazon Alexa

Entertainment: Music, Books, Video…

Household: Timers, ToDo Lists, Shopping, Calendar…

Information: Weather, Traffic, News, Q/A…

Smart Home: Lights, Thermostats, Switches…

45,000+ skills: Developed by 3rd Parties

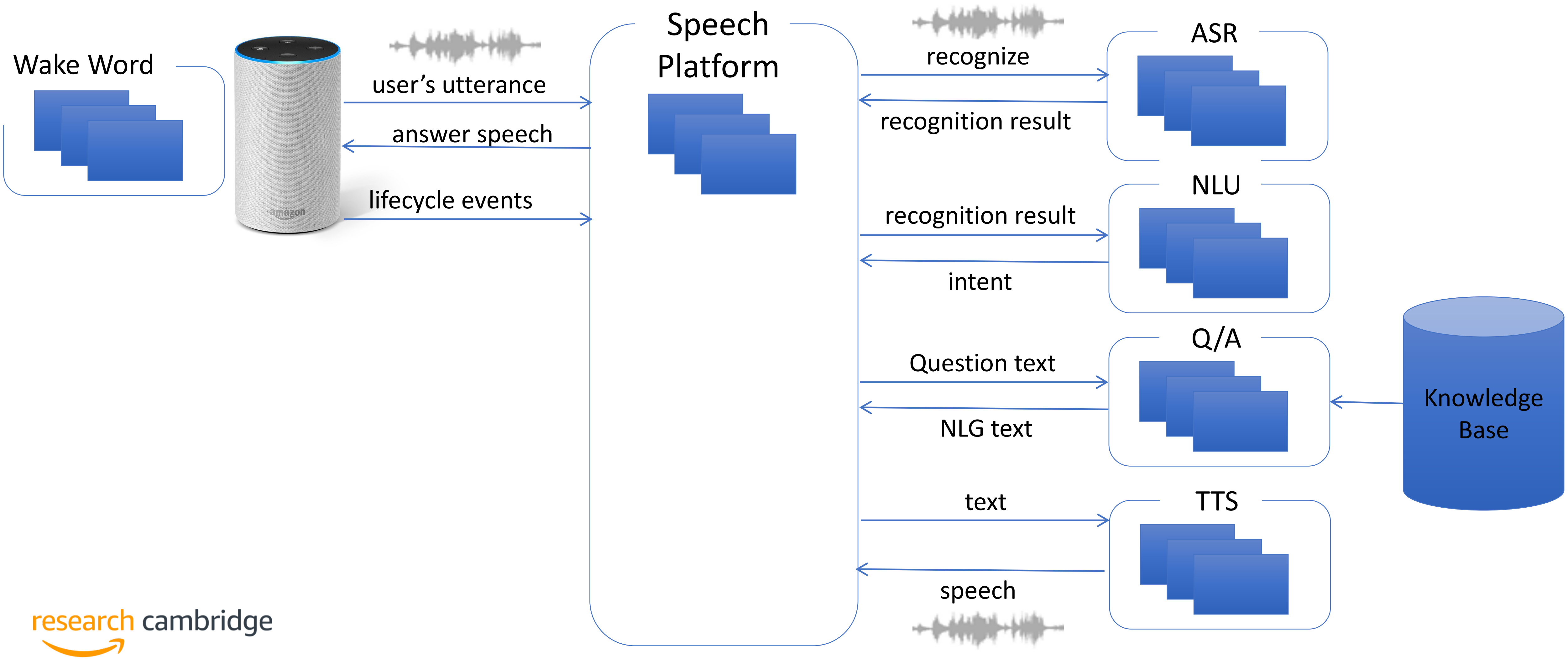
Alexa Question Answering

“Alexa, what books did Carrie Fisher write?”

“The books that Carrie Fisher is an author of are Delusions of Grandma, Shockaholic, Surrender the Pink, Postcards from the Edge, The Best Awful There Is and Wishful Drinking.”

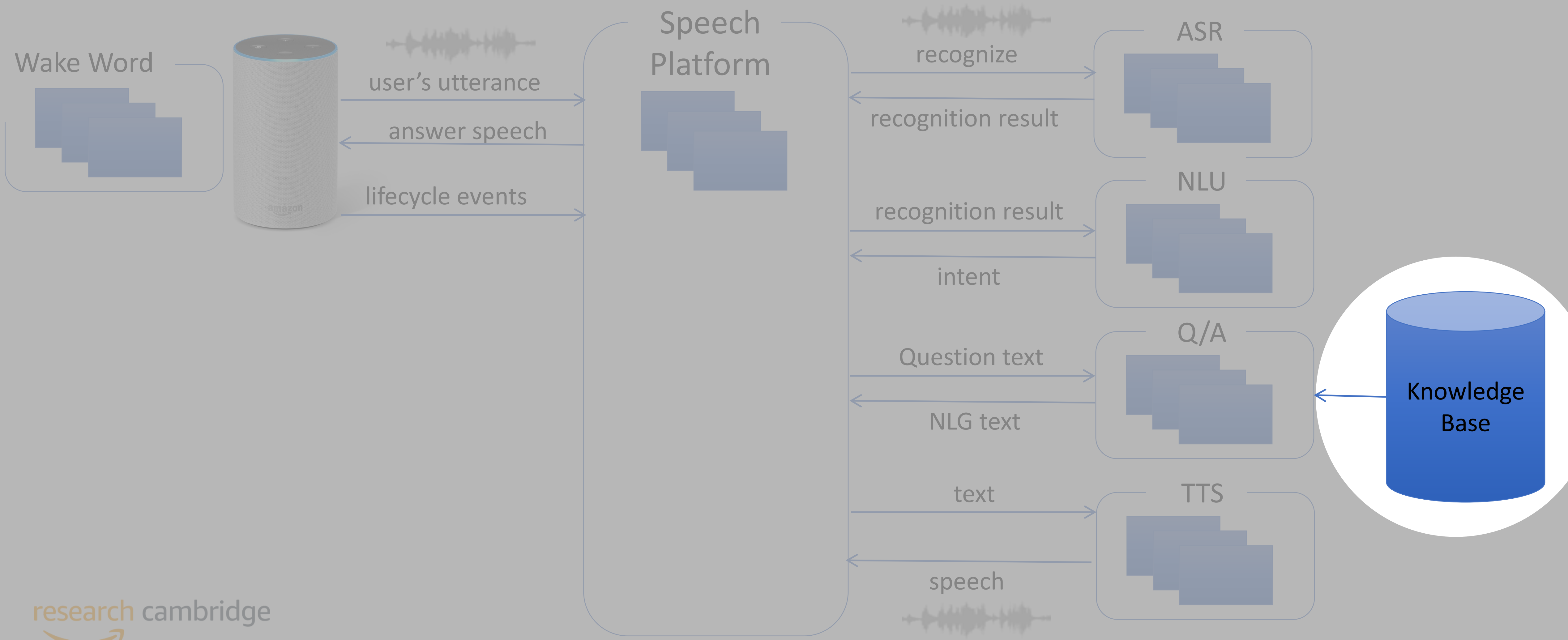
Alexa Question Answering

Alexa, what books did Carrie Fisher write?



Alexa Question Answering

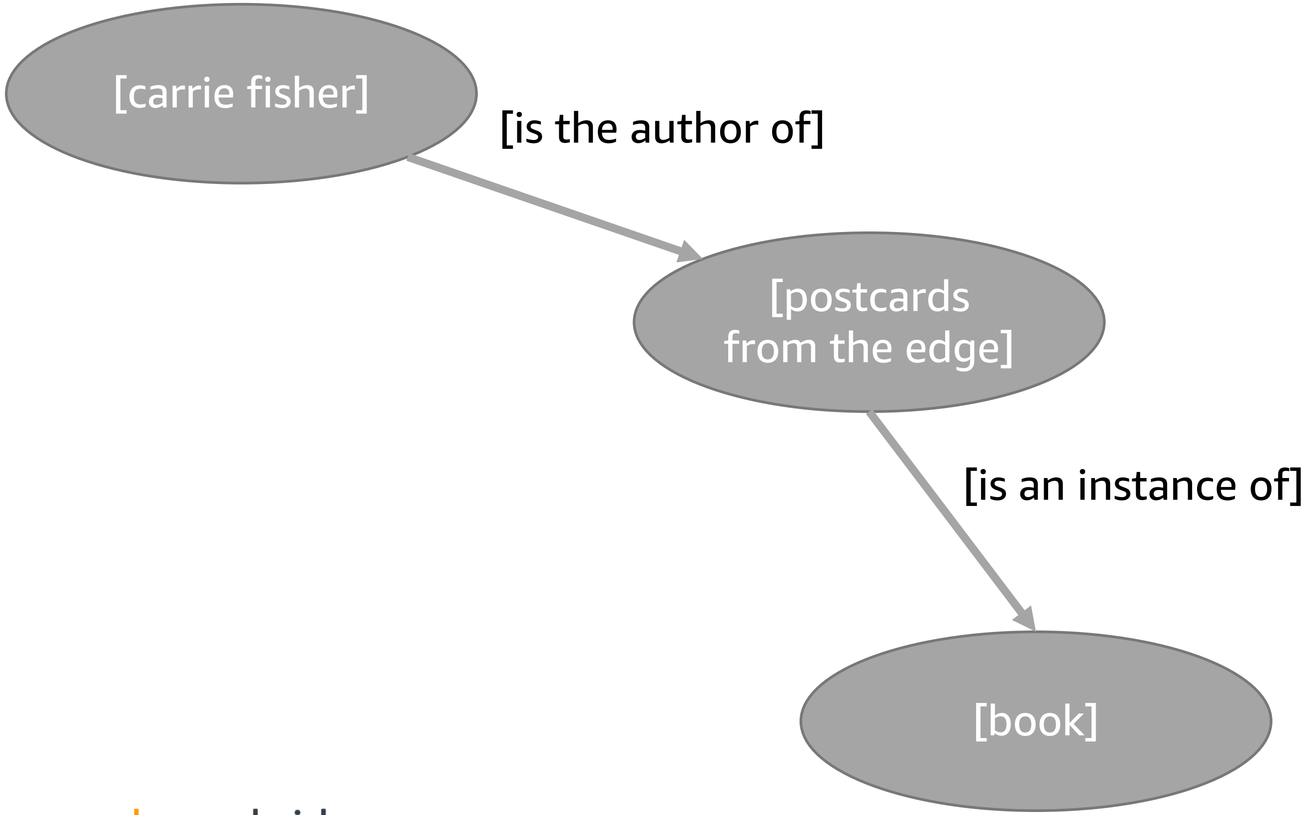
Alexa, what books did Carrie Fisher write?



research cambridge

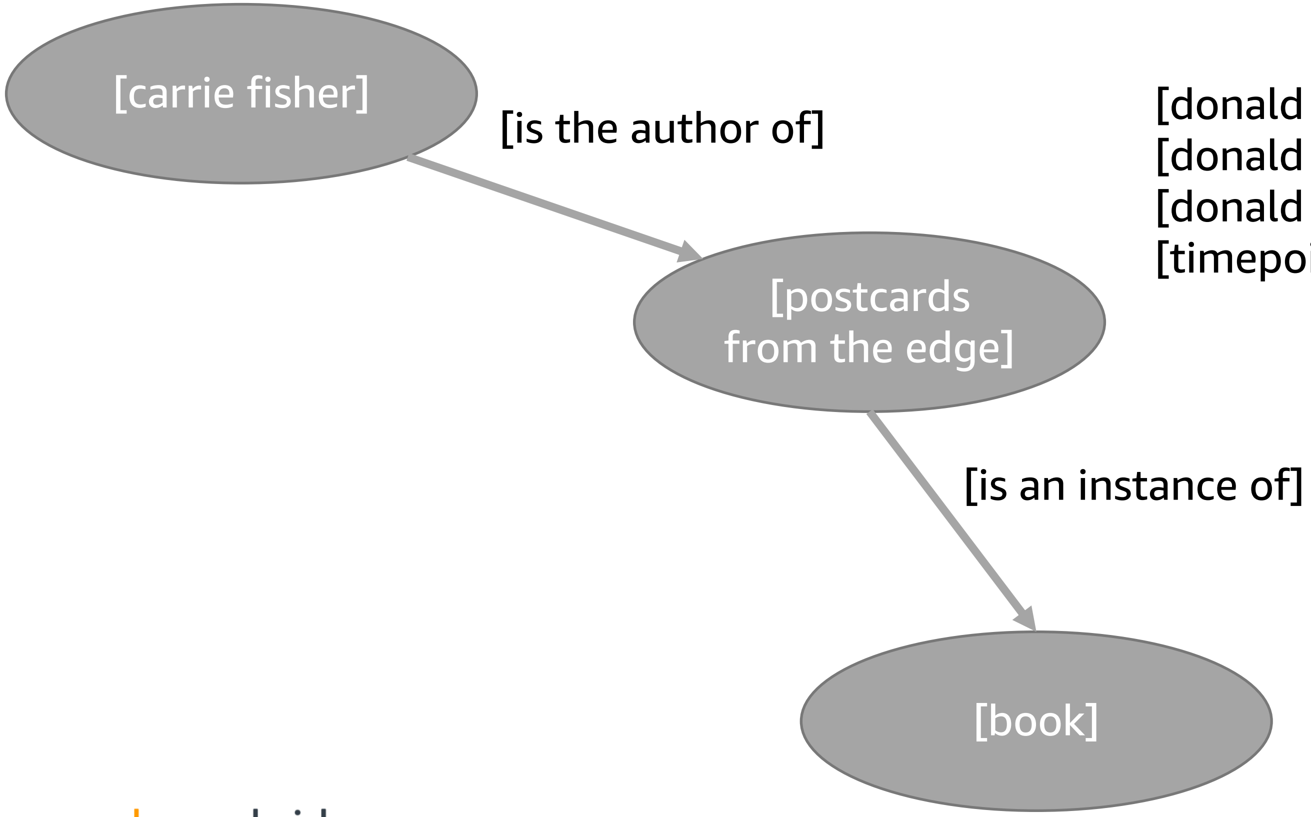
Alexa Knowledge Base

Named relations between entities



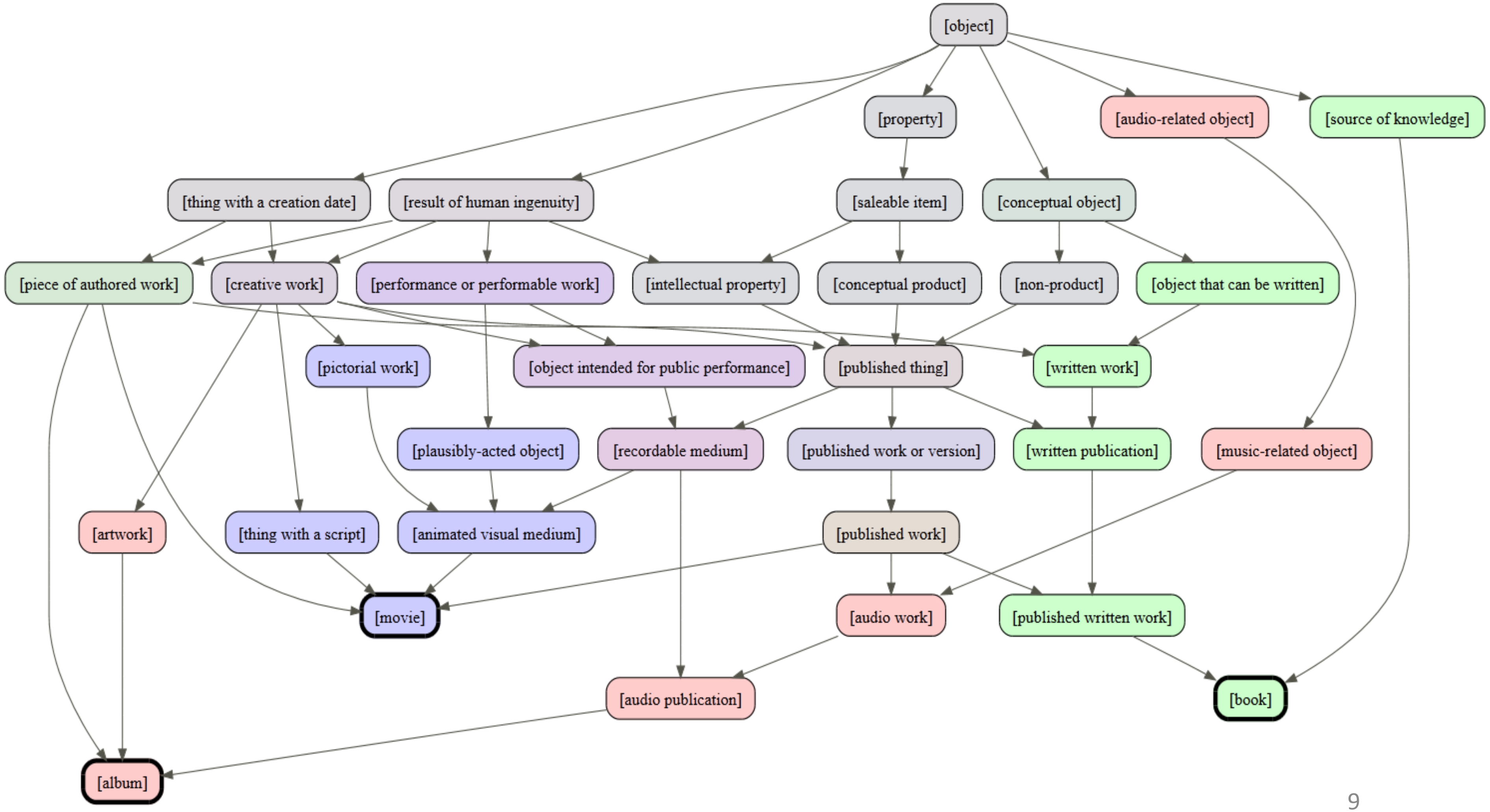
Alexa Knowledge Base

Named relations between entities

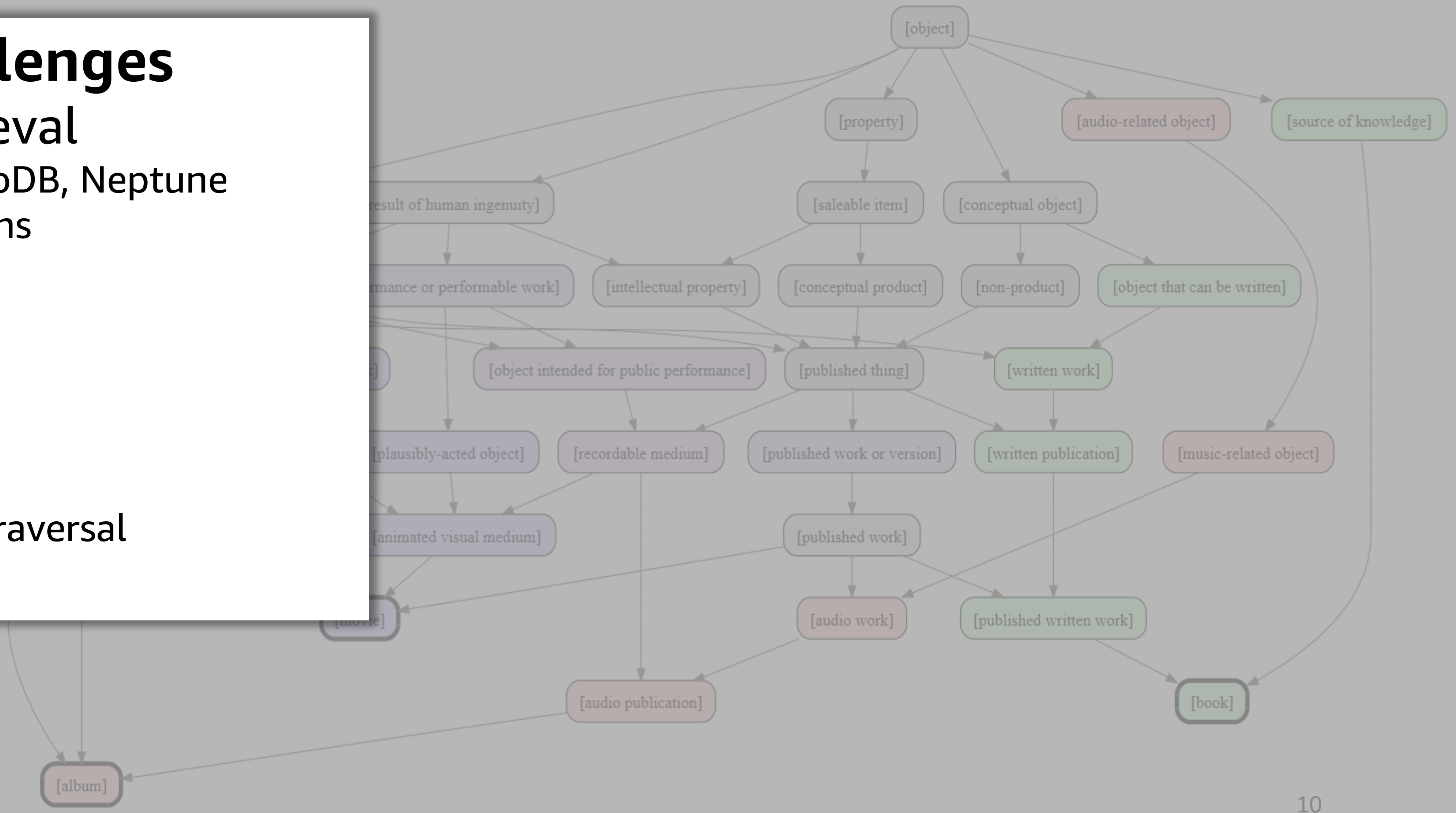


[donald trump] [is an instance of] [politician]
[donald trump] [is an instance of] [republican]
[donald trump] [is married to] [melania trump]
[timepoint: [“1946/6/14”]] [is the birthdate of] [donald trump]

Alexa Knowledge Base



Alexa Knowledge Base

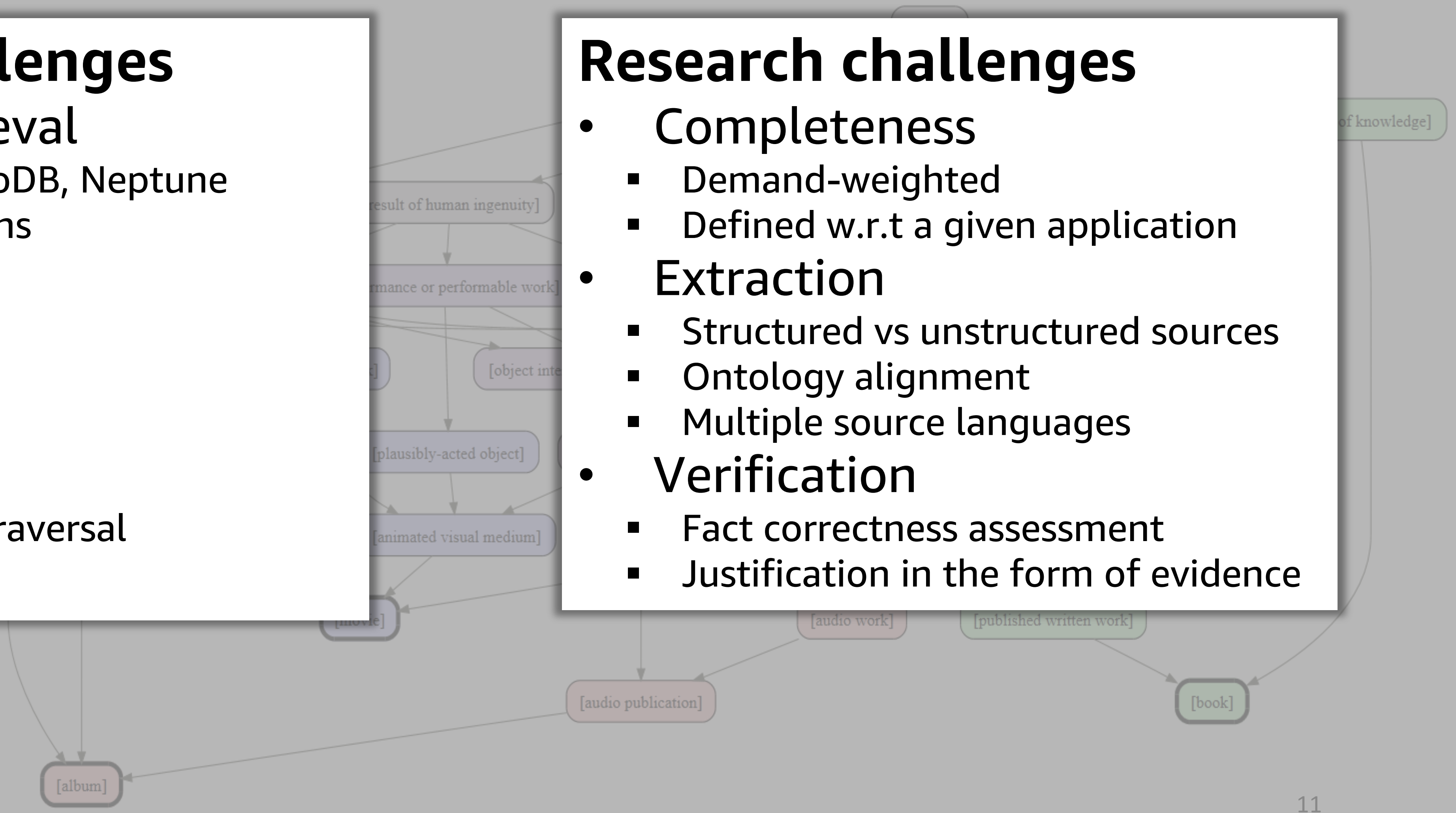


KB usage challenges

- Storage/retrieval
  - AWS S3, DynamoDB, Neptune
  - In-memory graphs
  - Backups
- Consistency
  - Ingestion checks
  - Stale facts
- Querying
  - Efficient graph traversal
  - Generated facts

research cambridge

Alexa Knowledge Base



KB usage challenges

- Storage/retrieval
  - AWS S3, DynamoDB, Neptune
  - In-memory graphs
  - Backups
- Consistency
  - Ingestion checks
  - Stale facts
- Querying
  - Efficient graph traversal
  - Generated facts

Research challenges

- Completeness
  - Demand-weighted
  - Defined w.r.t a given application
- Extraction
  - Structured vs unstructured sources
  - Ontology alignment
  - Multiple source languages
- Verification
  - Fact correctness assessment
  - Justification in the form of evidence

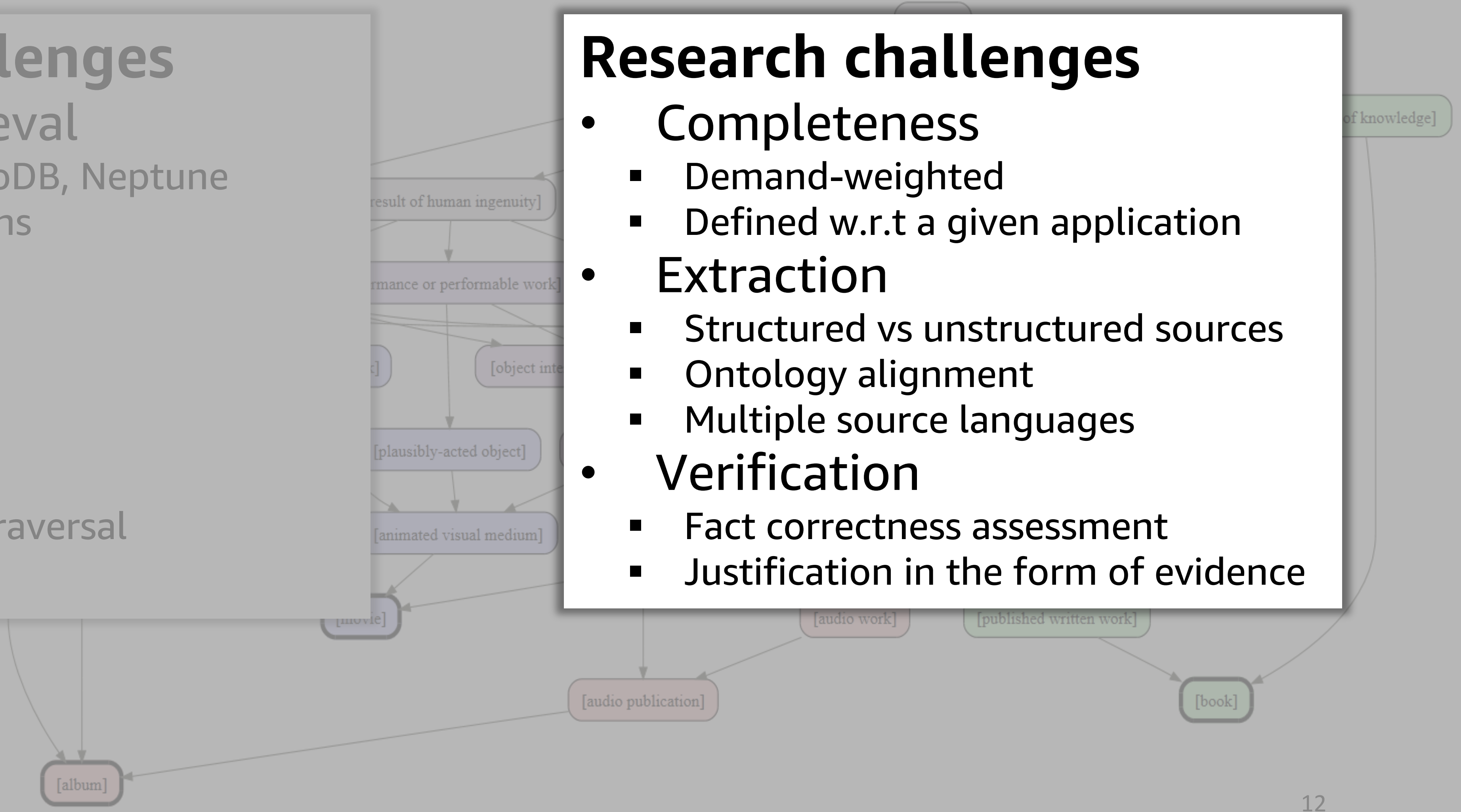
Alexa Knowledge Base

KB usage challenges

- Storage/retrieval
  - AWS S3, DynamoDB, Neptune
  - In-memory graphs
  - Backups
- Consistency
  - Ingestion checks
  - Stale facts
- Querying
  - Efficient graph traversal
  - Generated facts

Research challenges

- Completeness
  - Demand-weighted
  - Defined w.r.t a given application
- Extraction
  - Structured vs unstructured sources
  - Ontology alignment
  - Multiple source languages
- Verification
  - Fact correctness assessment
  - Justification in the form of evidence



Demand-Weighted Completeness Prediction for a Knowledge Base

Hopkinson et al. (2018) – NAACL

Demand-Weighted Completeness

- A KB can be complete for an entity if it contains all possible statements about that entity
- Practically this is not a very useful definition
  - Not all KB applications require all of the data equally
  - For each application, completeness can have a different meaning
- Given an application how can we determine KB completeness?
  - Can come from hand-written archetypes, or KB statistics
  - Can focus on **relation existence**, or include cardinality of relations

Problem Statement

Given an entity E in a KB, and
query usage data of the KB,
predict the distribution of relations
that E must have in order for 95%
of (usage) queries about E to be answered
successfully

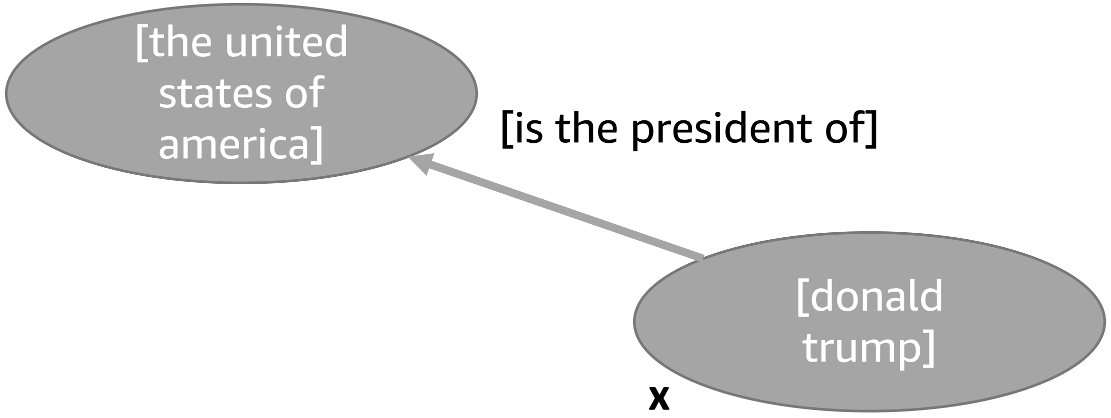
Problem Statement

Given an entity E in a KB, and
query usage data of the KB,
predict the **distribution of relations**
that E must have in order for 95%
of (usage) queries about E to be answered
successfully

Relation Distributions

query y
x [is the president of] [the united states of america]
x [is married to] y

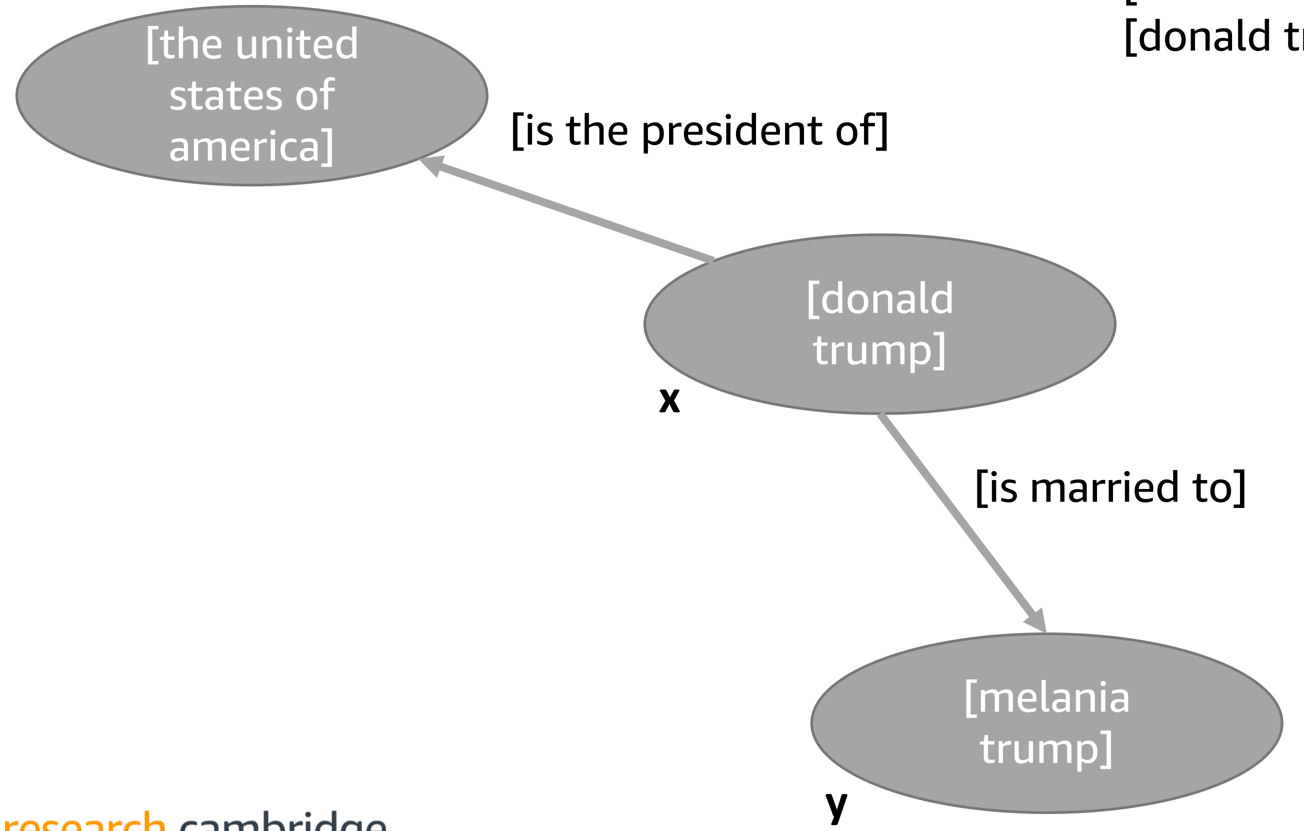
Relation Distributions



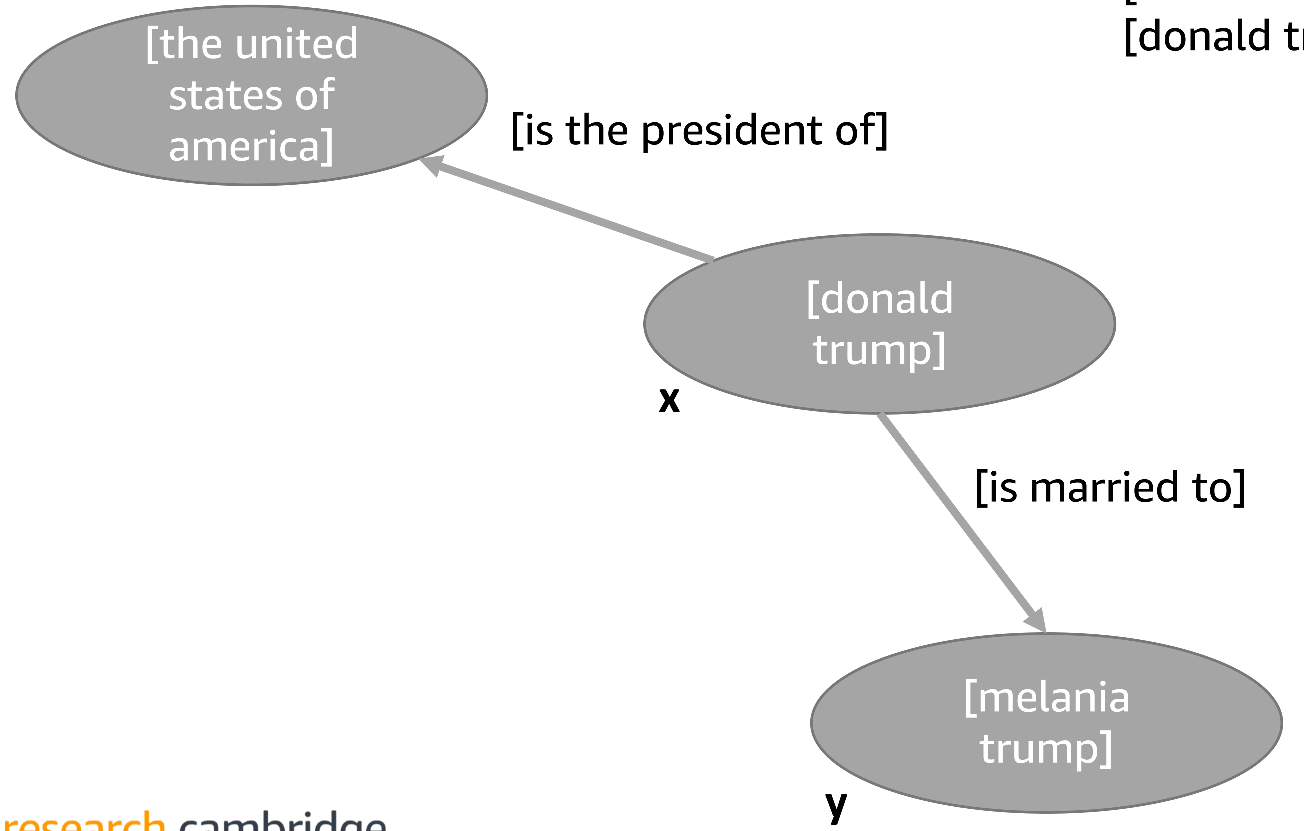
query y
x **[is the president of]** [the united states of america]
x [is married to] y

Relation Distributions

query y
[donald trump] **[is the president of]** [the united states of america]
[donald trump] **[is married to]** y



Relation Distributions



query y
[donald trump] **[is the president of]** [the united states of america]
[donald trump] **[is married to]** y

[the united states of america]	
hasPresident	13
hasCapital	8
hasPopulation	6

[donald trump]	
hasSpouse	16
hasBirthdate	12
hasBirthplace	5

research cambridge

Representing Entities

- We may have little usage information for individual entities
- Some relations may have an incidence lower than 1 per entity
- In order to generalise, we need a representation that allows grouping of similar entities
- Additionally, the representation should be interpretable

- We represent entities by their membership of different classes or **class signatures**

Representing Entities

- We may have little usage information for individual entities
- Some relations may have an incidence lower than 1 per entity
- In order to generalise, we need a representation that allows grouping of similar entities
- Additionally, the representation should be interpretable

- We represent entities by their membership of different classes or **class signatures**

[donald trump]	
person	1
politician	1
democrat	0
republican	1
…	

Models

- Baseline – Frequency Based model
  - Statistics for each class
  - Combine and normalize the relation counts for an entity’s classes
- Linear Regression
  - Least squares linear regression
- Neural network
  - Feed-forward network with n-hot input vector for the class signature
  - Single 10 node ReLU hidden layer, softmax output layer
  - Objective function: KL Divergence

Datasets

Dataset	Classes	Relations	Signatures
$D1_{small}$	4,400	1,300	12,000
$D2_{medium}$	8,000	2,000	25,000
$D3_{large}$	9,400	2,100	37,000

Results

Model	Jaccard	False Negative	False Positive
$D1_{small}$			
Frequency	0.779	0.066	0.123
Regression	0.667	0.090	0.242
NN	**0.808**	0.032	0.159
$D2_{medium}$			
Frequency	0.816	0.059	0.094
Regression	0.703	0.077	0.220
NN	**0.840**	0.037	0.123
$D3_{large}$			
Frequency	0.819	0.062	0.088
Regression	0.720	0.069	0.210
NN	**0.850**	0.038	0.113

Summary

- For each entity in a KB:
  - Get its class signature
  - Predict the relation distribution the entity ‘should’ have
  - Compare the existing facts for the entity to the predicted distribution
  - Record the missing relations, and their weight
- Do an entity-weighted aggregation over all entities

- This gives a demand-weighted list of missing facts in the KB

Alexa Knowledge Base

KB usage challenges

- Storage/retrieval
  - AWS S3, DynamoDB, Neptune
  - In-memory graphs
  - Backups
- Consistency
  - Ingestion checks
  - Stale facts
- Querying
  - Efficient graph traversal
  - Generated facts

Research challenges

- Completeness
  - Demand-weighted
  - Defined w.r.t a given application
- Extraction
  - Structured vs unstructured sources
  - Ontology alignment
  - Multiple source languages
- Verification
  - Fact correctness assessment
  - Justification in the form of evidence

Simple Large-scale Relation Extraction from Unstructured Text

Christodoulopoulos and Mittal (2018) – LREC

Knowledge from Unstructured Text

The Goal:

Carrie Fisher wrote several semi-autobiographical novels, including Postcards from the Edge.

Knowledge from Unstructured Text

The Goal:

Carrie Fisher wrote several semi-autobiographical novels, including **Postcards from the Edge**.

Knowledge from Unstructured Text

The Goal:

Carrie Fisher wrote several semi-autobiographical novels, including **Postcards from the Edge**.

Entity Recognition

↓

Entity Resolution

↓

Relation Extraction

Knowledge from Unstructured Text

The Goal:

Carrie Fisher wrote several semi-autobiographical novels, including **Postcards from the Edge**.

[carrie fisher] [postcards from the edge]

Entity Recognition

↓

Entity Resolution

↓

Relation Extraction

Knowledge from Unstructured Text

The Goal:

Carrie Fisher **wrote** several semi-autobiographical novels, including **Postcards from the Edge**.

[carrie fisher] [is the author of] [postcards from the edge]

Entity Recognition

↓

Entity Resolution

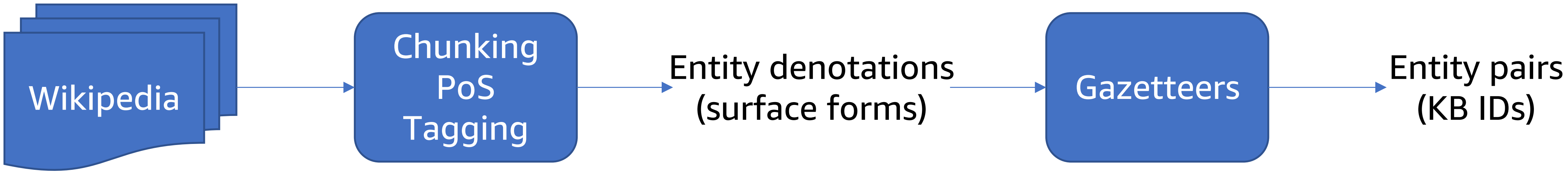
↓

Relation Extraction

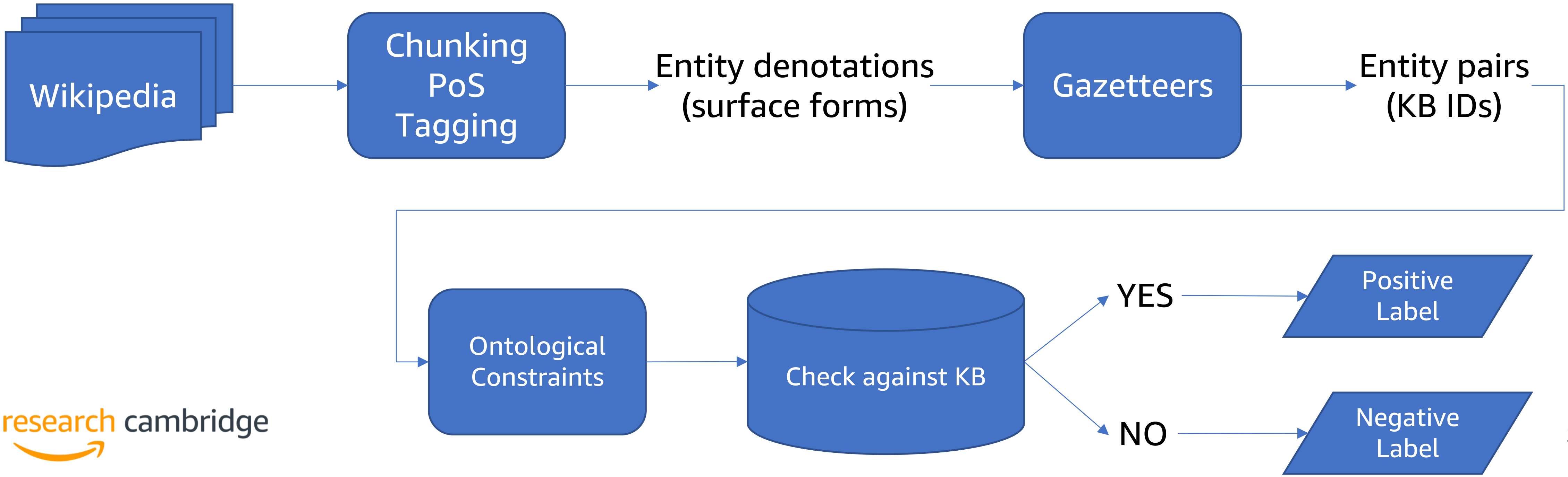
Relation Extraction Approaches

- Rule-based
- Fully supervised
- Unsupervised
- **Distant/weakly supervised**
  - Snow, Jurafsky, Ng, 2005
  - Main assumption: if two entities are linked by a relation, any sentence containing both entities is *likely* to express that relation
    - [steven spielberg] [is the director of] [saving private ryan]
    - "Spielberg's film Saving Private Ryan is based on…"

Distant supervision label generation



Distant supervision label generation

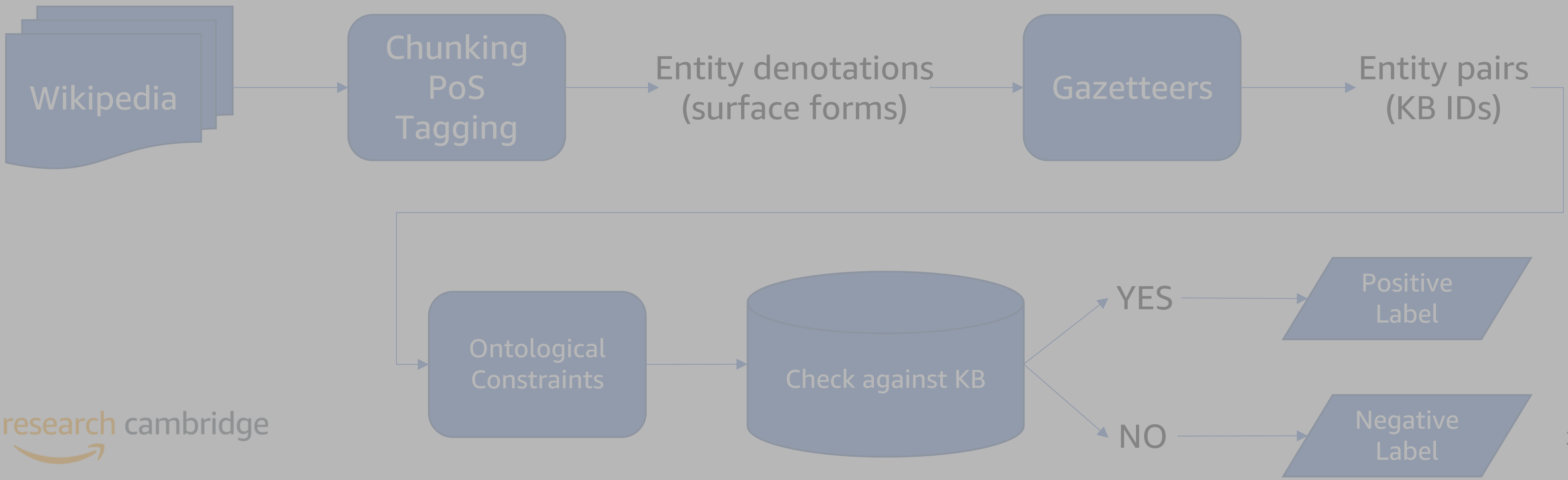


Distant supervision label generation

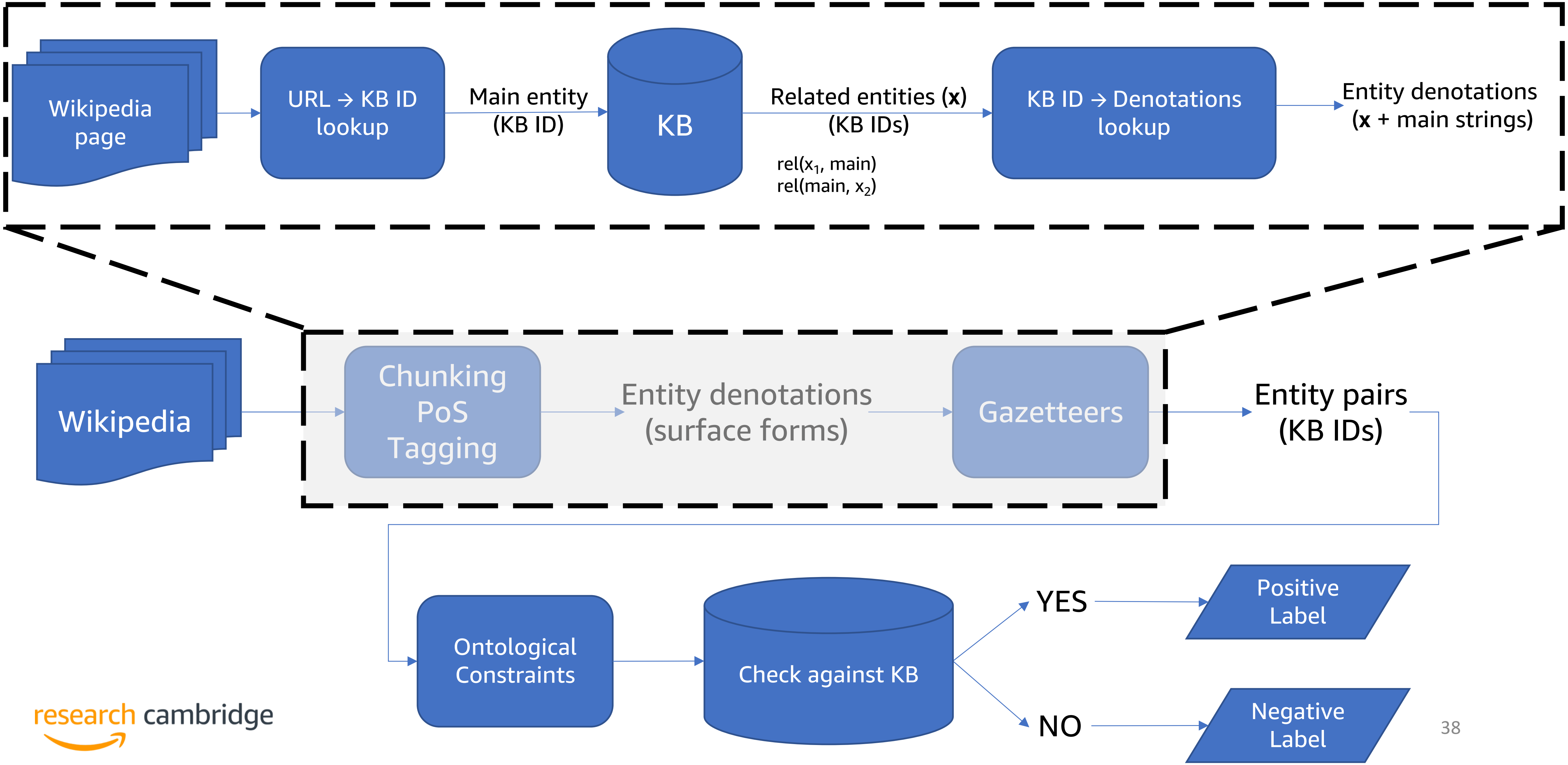
His studies were interrupted by army service and at the *end* of the *war* he was forced to return. . .
`[the second world war] [is an instance of] [cause of death]`

In the *intro* to the *song*, Fred Durst makes reference to. . .
`[intro 15367][is an instance of] [song]`

Turner also released one *album* and several *singles* under the moniker Repeat.
`[the singles the 2011 album] [is an instance of] [album]`

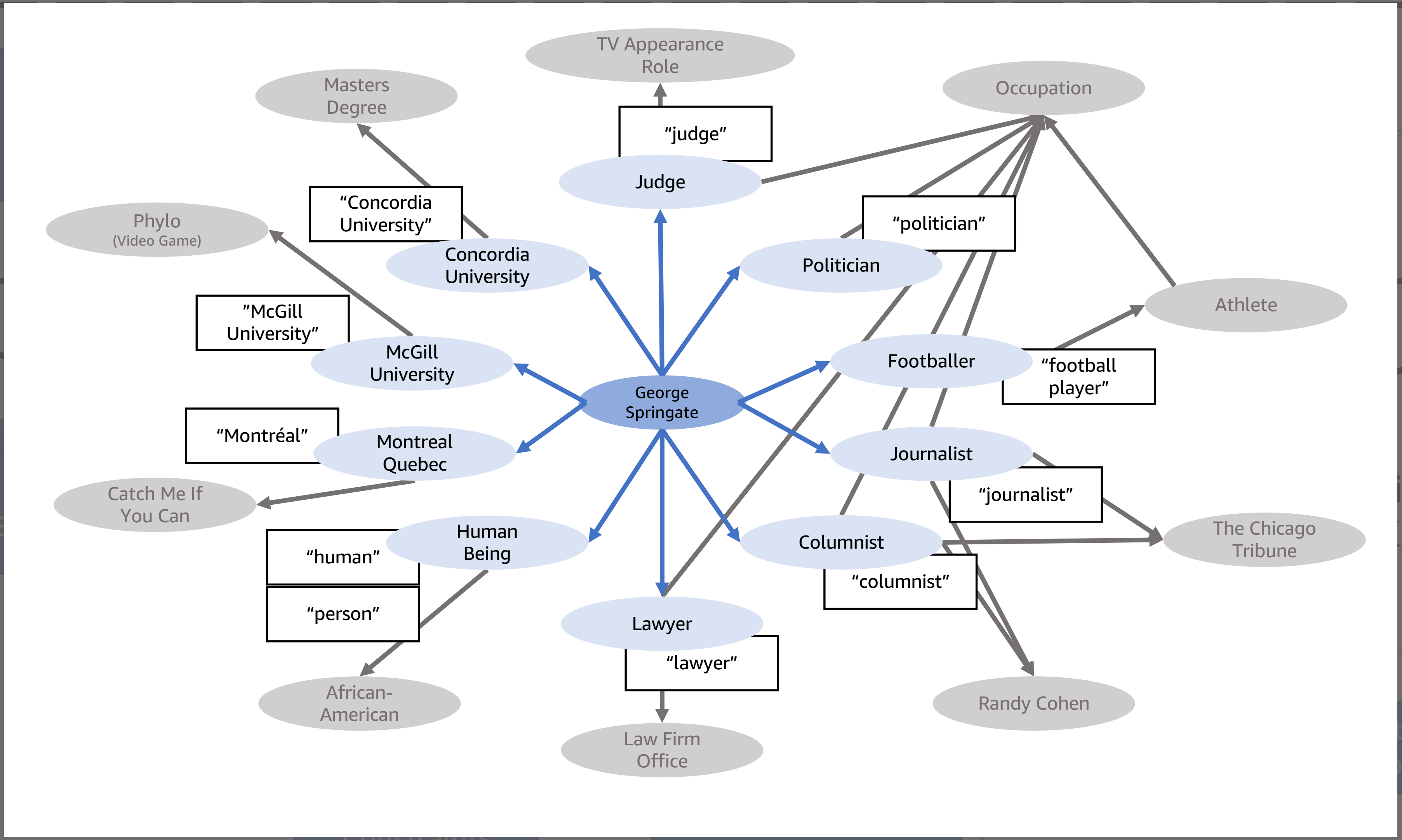


Distant supervision label generation



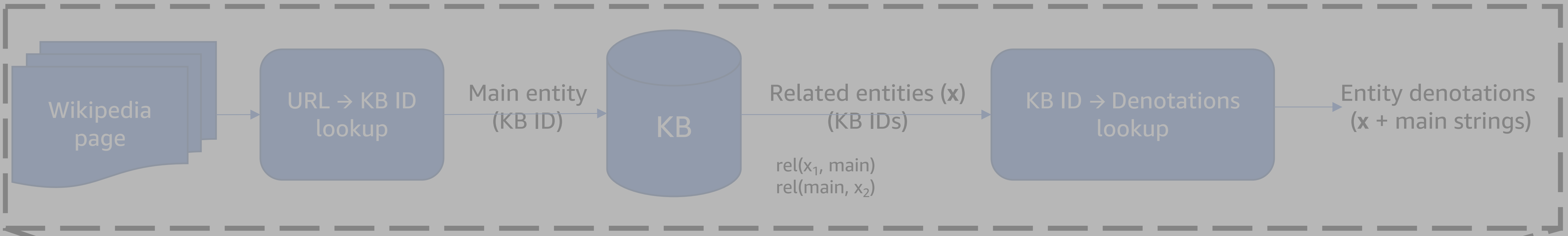
research cambridge

Distant supervision label generation



research cambridge

Distant supervision label generation



Call Your Girlfriend was written by Robyn, Alexander Kronlund and Klas Åhlund, with the latter producing the *song*.

`[call your girlfriend 3] [is an instance of] [song]`

Forget Her is a *song* by Jeff Buckley.

`[forget her] [is an instance of] [song]`

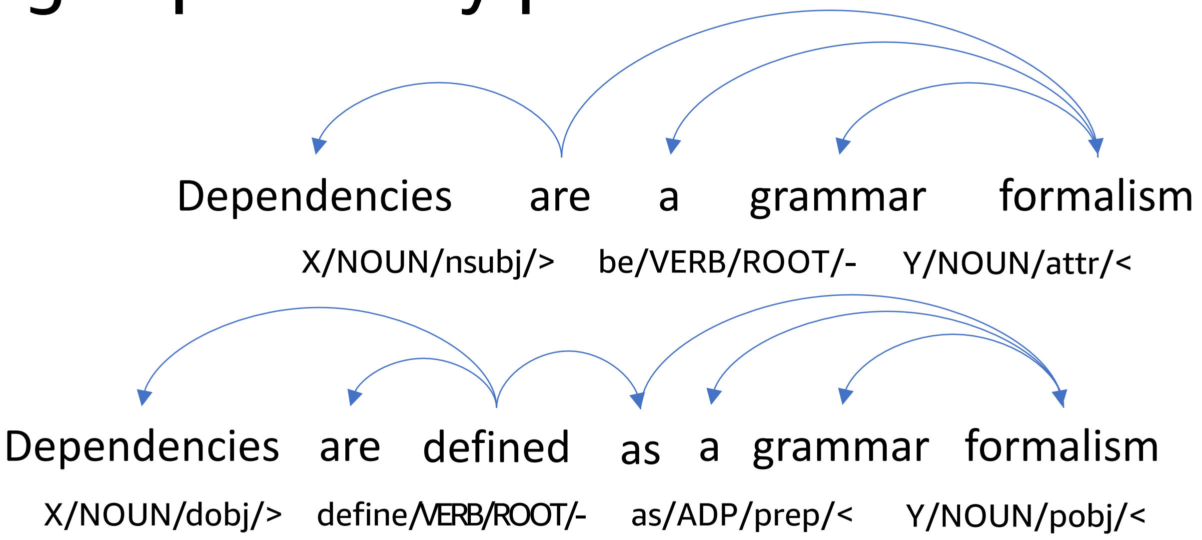
The *Subei Mongol Autonomous County* is an autonomous *county* within the prefecture-level city of Jiuquan in the northwestern Chinese province of Gansu.

`[subei mongol autonomous county] [is an instance of] [chinese county]`

research cambridge

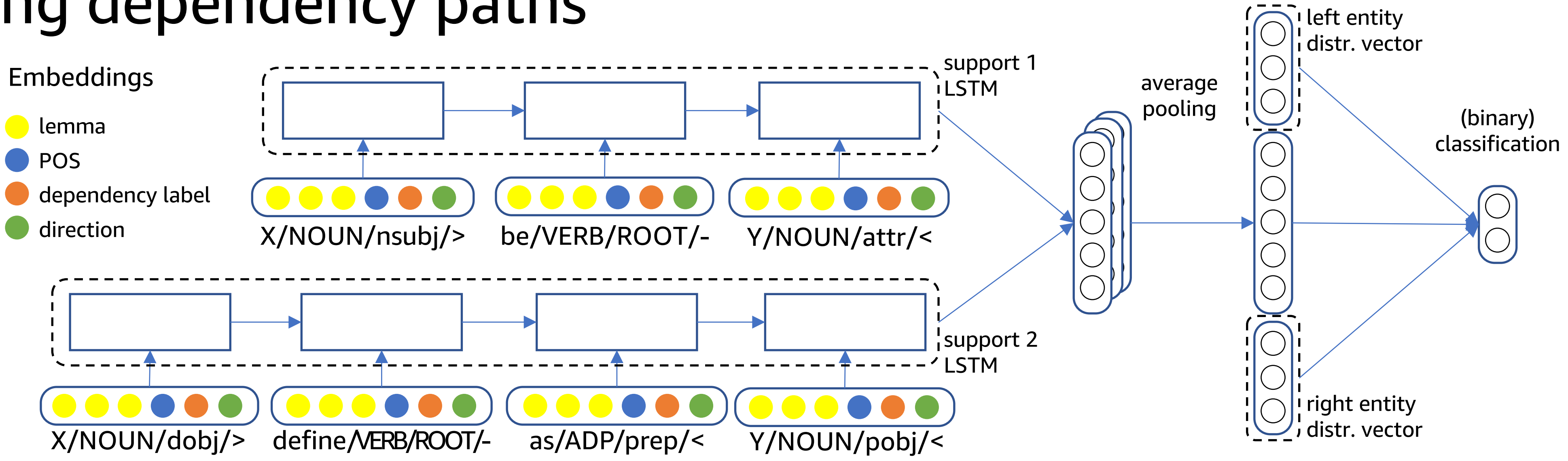
Relation extraction

- HypeNET (Shwartz and Goldberg, 2016)
- Hyponyms [is an instance of] only
  - LexNET extends to multiple relations
- Using dependency paths



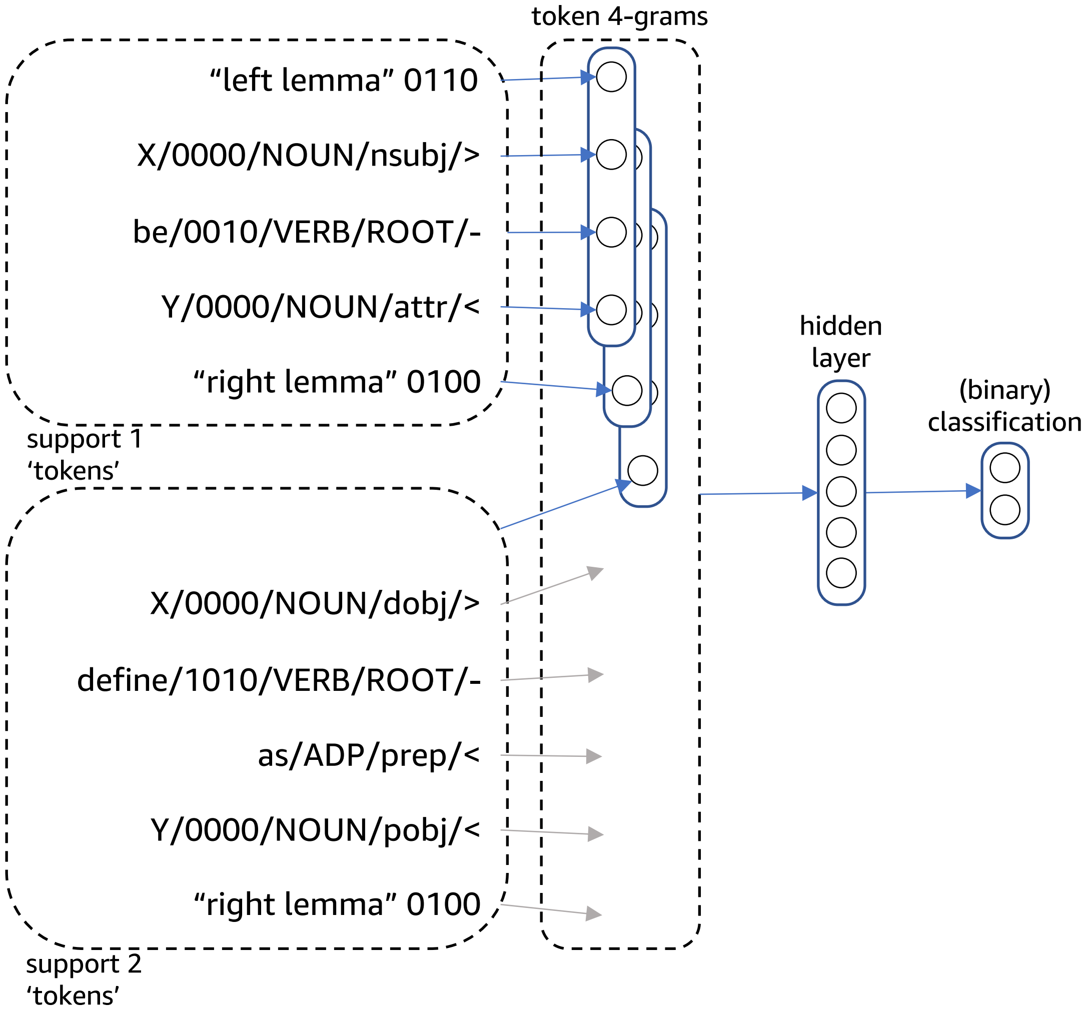
Relation extraction

- HypeNET (Shwartz and Goldberg, 2016)
- Hyponyms [is an instance of] only
  - LexNET extends to multiple relations
- Using dependency paths



Relation extraction

- fastText (Joulin et al., 2016)
- Linear model
  - One hidden layer
  - Rank constraint



Results

HypeNET **equally good** as the much simpler fastText with the **same input features**.

Alexa KB

Relation	HypeNET	fastText
[is an instance of]	94.29 (0.21)	94.31 (0.03)
[is the birthplace of]	85.57 (0.26)	**87.63** (0.01)
[applies to]	81.98 (1.78)	**86.17** (0.01)

Results

HypeNET **equally good** as the much simpler fastText with the **same input features**.

Alexa KB

Relation	HypeNET	fastText
[is an instance of]	94.29 (0.21)	94.31 (0.03)
[is the birthplace of]	85.57 (0.26)	**87.63** (0.01)
[applies to]	81.98 (1.78)	**86.17** (0.01)

Wikidata

Relation	HypeNET	fastText
instance of (P31)	93.90 (0.21)	**96.44** (0.01)
birthplace of (P19)	92.06 (0.90)	**93.05** (0.07)
part of (P527)	48.73 (2.59)	**72.87** (0.16)

Results

HypeNET **equally good** as the much simpler fastText with the **same input features**.

MaxEnt results show that **features alone are not enough.**
Need to create higher-dimensional representations of discrete features.

Alexa KB

Relation	HypeNET	fastText	MaxEnt
[is an instance of]	94.29 (0.21)	94.31 (0.03)	83.93
[is the birthplace of]	85.57 (0.26)	**87.63** (0.01)	80.83
[applies to]	81.98 (1.78)	**86.17** (0.01)	65.27

Wikidata

Relation	HypeNET	fastText	MaxEnt
instance of (P31)	93.90 (0.21)	**96.44** (0.01)	58.45
birthplace of (P19)	92.06 (0.90)	**93.05** (0.07)	66.72
part of (P527)	48.73 (2.59)	**72.87** (0.16)	45.13

Summary

- New method for entity resolution
  - Page-specific gazetteers
- Features are important
  - HypeNET vs fastText
- Feature representation is important
  - fastText vs MaxEnt

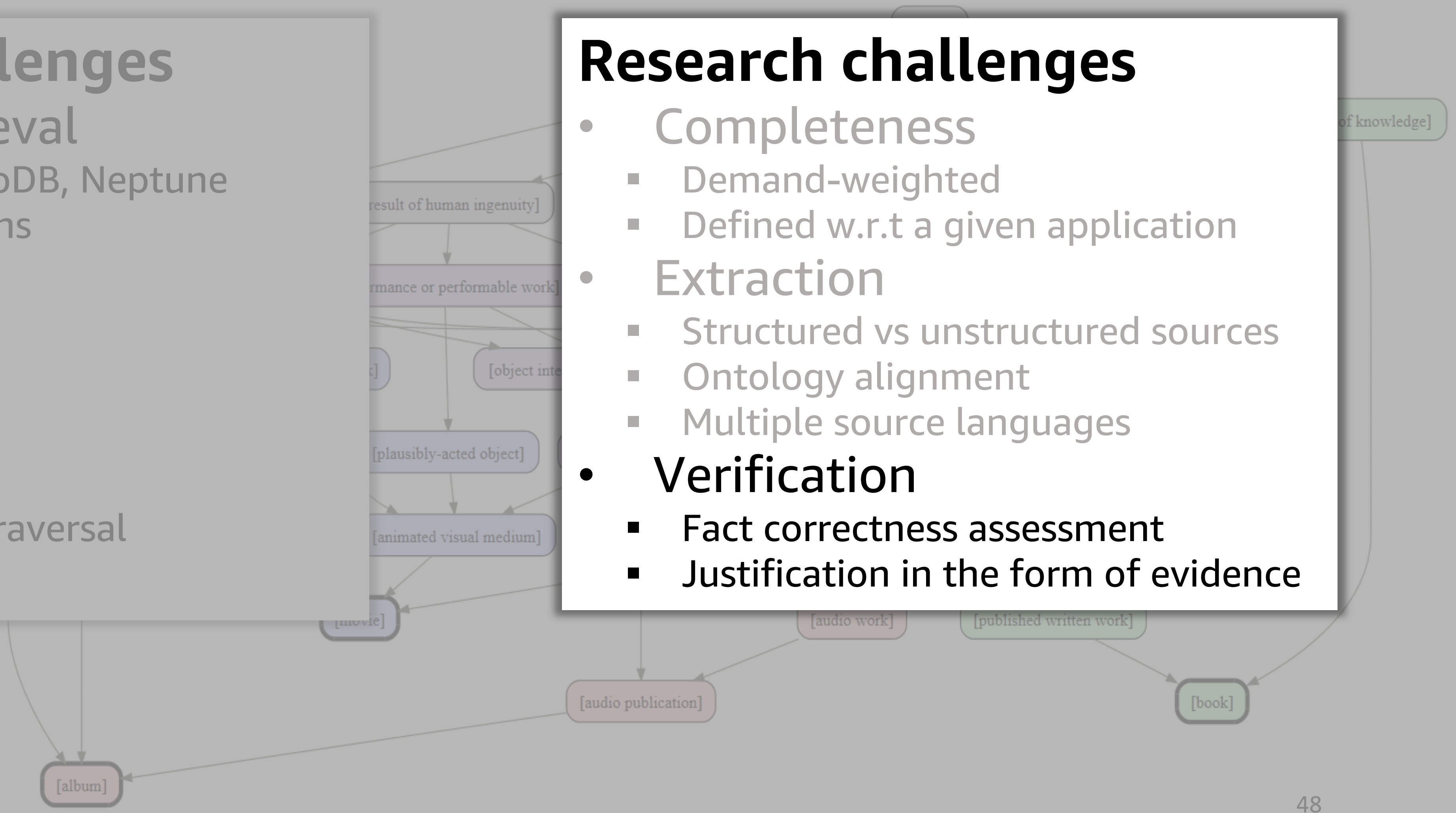
Alexa Knowledge Base

KB usage challenges

- Storage/retrieval
  - AWS S3, DynamoDB, Neptune
  - In-memory graphs
  - Backups
- Consistency
  - Ingestion checks
  - Stale facts
- Querying
  - Efficient graph traversal
  - Generated facts

Research challenges

- Completeness
  - Demand-weighted
  - Defined w.r.t a given application
- Extraction
  - Structured vs unstructured sources
  - Ontology alignment
  - Multiple source languages
- Verification
  - Fact correctness assessment
  - Justification in the form of evidence



FEVER: a Large-scale Dataset for Fact Extraction and VERification

Thorne et al. (2018) – NAACL

Why automate verification?

- More frequent updates
  - Match the scale/speed of fact extraction
- Increased number of facts checked
- Wider variety of sources
- Provide justification of answers

The FEVER dataset (fever.ai)

- 185,000 true and false claims
  - Written by human annotators
- For each claim
  - Evidence from multiple Wikipedia pages at a sentence level
  - `supported/refuted/not enough info` label given the evidence

- Both evidence and label must be correct for scoring

research cambridge

The University Of Sheffield.

The FEVER dataset (fever.ai)

Claim:

The Rodney King riots took place in the most populous county in the USA.

Evidence:

`[wiki/Los Angeles Riots]`: The 1992 Los Angeles riots, also known as the Rodney King riots were a series of riots, lootings, arsons, and civil disturbances that occurred in Los Angeles County, California in April and May 1992.

`[wiki/Los Angeles County]`: Los Angeles County, officially the County of Los Angeles, is the most populous county in the United States.

The FEVER dataset (fever.ai)

Claim:

The Rodney King riots took place in the most populous county in the USA.

SUPPORTED

Evidence:

[wiki/Los Angeles Riots]: The 1992 Los Angeles riots, **also known as the Rodney King riots** were a series of riots, lootings, arsons, and civil disturbances that occurred in **Los Angeles County**, California in April and May 1992.

[wiki/Los Angeles County]: Los Angeles County, officially the County of Los Angeles, **is the most populous county in the United States**.

Dataset construction

- Sentences
- Factoid Claims
- Mutated Claims
- Verification & Labeling

Dataset construction

- Sentences
- Factoid Claims
- Mutated Claims
- Verification & Labeling

- June 2017 English Wiki Dump
- Sentence-split, tokenized 5.4m pages (CoreNLP)
- Kept only the intro section of articles
- Used 50,000 popular pages for claim generation

Dataset construction

- Sentences
- Factoid Claims
- Mutated Claims
- Verification & Labeling

- Simple sentences from the original sentence
  - One fact (clause) per sentence
- Context from original Wikipedia article
- World knowledge introduced using dictionary of Wikipedia first sentences

Generating Claims About

Warren Buffett

Source Sentence

This is the sentence that is used to substantiate your claims about Warren Buffett

Buffett has been the chairman and largest shareholder of Berkshire Hathaway since 1970, and his business exploits have had him referred to as the "Wizard", "Oracle" or "Sage" of Omaha by global media outlets.

Show Context

Dictionary

Click the word for a definition. These definitions can be used to support the claims you write or make the claims more complex by making a deduction using the dictionary definitions

The dictionary comes from the blue links on Wikipedia. This may be empty if the passage from Wikipedia contains no links.

Berkshire Hathaway

Berkshire Hathaway Inc. is an American multinational conglomerate holding company headquartered in Omaha, Nebraska, United States.

List of assets owned by Berkshire Hathaway

Omaha, Nebraska

Omaha is the largest city in the state of Nebraska and the county seat of Douglas County.

shareholder

True Claims (one per line)

Aim to spend about 2 minutes generating **2-5** claims from this source sentence

If the source sentence is uninformative, press the skip button

Example

Warren Buffett is the chairman of an American multinational company.
Warren Buffet's company is based in the largest city in the state of Nebraska

Submit Claims | Skip | Home

Dataset construction

- Sentences
- Factoid Claims
- Mutated Claims
- Verification & Labeling

- 6 mutations akin to relations in NLI
  - paraphrasing
  - negation
  - generalization
  - specialization
  - substitution with similar entity/relation
  - substitution with dissimilar entity/relation
- Mutated claims may be true or false or unknown

Da

- Se
- Fa
- Mu
- Ve

own

Substitution for a similar entity and/or relation (Type 3)

Substitute either a relation, property and/or an attribute of Warren Buffett in the claim to something else from the **same set of things**.

AVOID rephrasing the original claim.

The mutated claim should not imply the original claim.

The mutated claim must be about **Warren Buffett**.

Examples

Warren Buffet is the chairman of a Canadian multinational corporation.
Warren Buffet is the president of an American multinational company.
Warren Buffet is chairman of McDonalds.

Substitution for a dissimilar entity and/or relation (Type 4)

Substitute either a relation, property and/or an attribute of Warren Buffett in the claim to something plausible **from a different set of things**.

AVOID rephrasing the original claim.

The mutated claim should not imply the original claim.

The mutated claim must be about **Warren Buffett**.

Examples

Warren Buffet is the chairman of an American football team.

Original Claim

Warren Buffett is the chairman of an American multinational company.

Generated From: *Buffett has been the chairman and largest shareholder of Berkshire Hathaway since 1970, and his business exploits have had him referred to as the "Wizard", "Oracle" or "Sage" of Omaha by global media outlets.*

Make the Claim more Specific (So that the new claim implies the original) (Type 5)

Modify the claim by replacing either a relation, property and/or an attribute of Warren Buffett to something **more specific** that implies the original claim.

AVOID rephrasing the original claim.

The mutated claim should imply the original claim.

The mutated claim must be about **Warren Buffett**.

Examples

Warren Buffet is chairman of Berkshire Hathaway

Make the Claim more General (So that the new claim is implied by the original) (Type 6)

Modify the claim by replacing either a relation, property and/or an attribute of Warren Buffett to something **more general** that is implied by the original claim.

AVOID rephrasing the original claim.

The mutated claim should be implied by the original claim.

The mutated claim must be about **Warren Buffett**.

Examples

Warren Buffet is the chairman of a company.

Dataset construction

- Sentences
- Factoid Claims
- Mutated Claims
- Verification & Labeling

- Sentence-level annotation for supporting/refuting information
- Can combine information from multiple pages
- Trade-off between time to find evidence vs. recall

Claim

The 1992 Los Angeles riots occurred in the most populous county in the United States.

Submit

Skip (opens menu) | Home | Guidelines

Wikipedia article for 1992 Los Angeles riots

The 1992 Los Angeles riots , also known as the Rodney King riots, the South Central riots, the 1992 Los Angeles civil disturbance , the 1992 Los Angeles civil unrest, and the Los Angeles uprising, were a series of riots, lootings, arsons , and civil disturbances that occurred in Los Angeles County , California in April and May 1992.

✔Supports ✖Refutes Cancel

The unrest began in South Central Los Angeles on April 29 after a trial jury acquitted four officers of the Los Angeles Police Department of the use of excessive force in the videotaped arrest and beating of Rodney King . Expand

It then spread throughout the Los Angeles metropolitan area as thousands of people rioted over a six-day period following the announcement of the verdict. Expand

Widespread looting , assault, arson, and killings occurred during the riots , and estimates of property damage were over $1 billion. Expand

Order was only restored after members of the California Army National Guard , the 7th Infantry Division , and the 1st Marine Division were called in to stop the rioting when local police could not control the situation. Expand

In total, 58 people were killed during the riots , more than 2,000 people were injured, and more than 11,000 were arrested. Expand

LAPD chief of police Daryl Gates , who had already announced his resignation by the time of the riots , took much of the institutional blame. Expand

Los Angeles County, California

- [x] Los Angeles County, officially the County of Los Angeles, is the most populous county in the United States.
- [ ] Its population is larger than that of 42 individual U.S. states.
- [ ] It has 88 incorporated cities and many unincorporated areas and at 4083 sqmi, it is larger than the combined areas of the U.S. states of Delaware and Rhode Island.
- [ ] The county is home to more than one-quarter of California residents and is one of the most ethnically diverse counties in the U.S. Its county seat, the City of Los Angeles, is also its most populous city at about four million.

Rodney King

- [ ] Rodney Glen King (April 2, 1965 -- June 17, 2012) was a taxi driver who became internationally known after being beaten by Los Angeles Police Department officers following a high-speed car chase on March 3, 1991.
- [ ] A witness, George Holliday, videotaped much of the beating from his balcony, and sent the footage to local news station KTLA.
- [ ] The footage shows four officers surrounding King, several of them striking him repeatedly, while other officers stood by.

Quality of annotation

- Precision: 95.4%, Recall: 72.4%
  - against super-annotators with no time restrictions
- 5-way IAA κ of 0.684 over 4% of claims (n=7506)
- We re-annotated 227 claims, 91.2% annotated correctly

- Lessons Learned:
  - Hard to remove annotator's world knowledge
  - Hard to come up with 'universal' definitions

The FEVER challenge

- 87 submissions from 23 teams
- Preliminary leaderboard – pending further annotation

Rank	Team Name	Evidence F1	Label Accuracy	FEVER score
1	UNC-NLP	0.5296	0.6821	0.6421
2	UCL Machine Reading Group	0.3497	0.6762	0.6252
3	Athene UKP TU Darmstadt	0.3697	0.6546	0.6158
…	…	…	…	…
20	Baseline	0.1826	0.4884	0.2745

Summary

- Fact verification is crucial for QA systems
  - Gauge of extracted fact quality
  - Provide justification of answers
- First large-scale dataset for fact verification
  - 185,000 human-generated claims
  - Labelled via evidence annotation
- Shared task results
  - Significant improvements over baselines – plenty of room for more
- FEVER 2?

Overall conclusions

- Knowledge at scale is difficult!
- Presented ideas on
  - **What**: demand-weighted coverage
  - **How**: entity/relation extraction from unstructured text
  - **Why**: verification through evidence
- Challenges
  - Different languages, sources, relations
  - Emerging trends

Overall conclusions

- Knowledge at scale is difficult!
- Presented ideas on
  - **What**: demand-weighted coverage
  - **How**: entity/relation extraction from unstructured text
  - **Why**: verification through evidence
- Challenges
  - Different languages, sources, relations
  - Emerging trends

Get more FEVER at the EMNLP workshop!

Thanks!