Knowledge Representation in NLP

Sushma Akoju, Waad Alharthi

Contents

What is Knowledge?

- Defining Knowledge is a question in Philosophy.
- Knowledge is what we learn as a kid or a student or as our learning, understanding of the world surrounding us changes as we perceive it.
- **Knowledge is set of axioms or facts. (AIMA, 3rd edition)**
- **"...premise or starting point for reasoning"**
- Propositional Logic and Predicate Logic

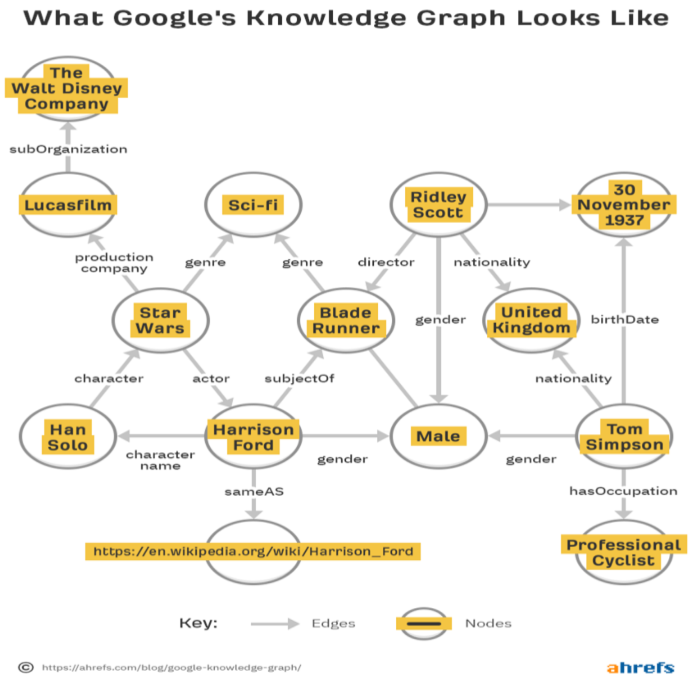
Duck test:

Looks like a duck, quacks like a duck - is a duck

Visual perception, auditory and locomotion

Everything we perceive is how we represent

- Students learn and move from middle school to high school
- Students ASK questions and (TELL) or ANSWER exams



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

Why Knowledge representation is required?

- Lot of knowledge is represented by various representations by human beings.
- Knowledge representation helps in reasoning.
- Knowledge representation also changes from reasoning.

Noun [edit]

knowledge (*usually uncountable*, *plural* **knowledges**)

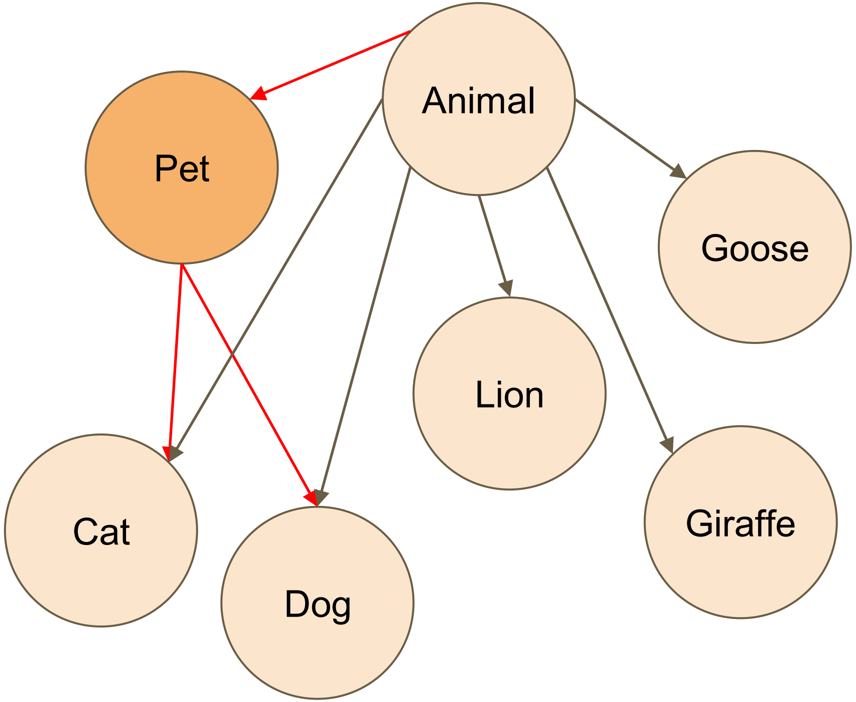
1. The fact of knowing about something; general understanding or familiarity with a subject, place, situation etc. [from 14th c.] [quotations ▼]

   *His **knowledge** of Iceland was limited to what he'd seen on the Travel Channel.*

2. Awareness of a particular fact or situation; a state of having been informed or made aware of something. [from 14th c.] [quotations ▼]
3. Intellectual understanding; the state of appreciating truth or information. [from 14th c.] [quotations ▼]

   ***Knowledge** consists in recognizing the difference between good and bad decisions.*

Image of change in entities from reasoning and new understanding of world around



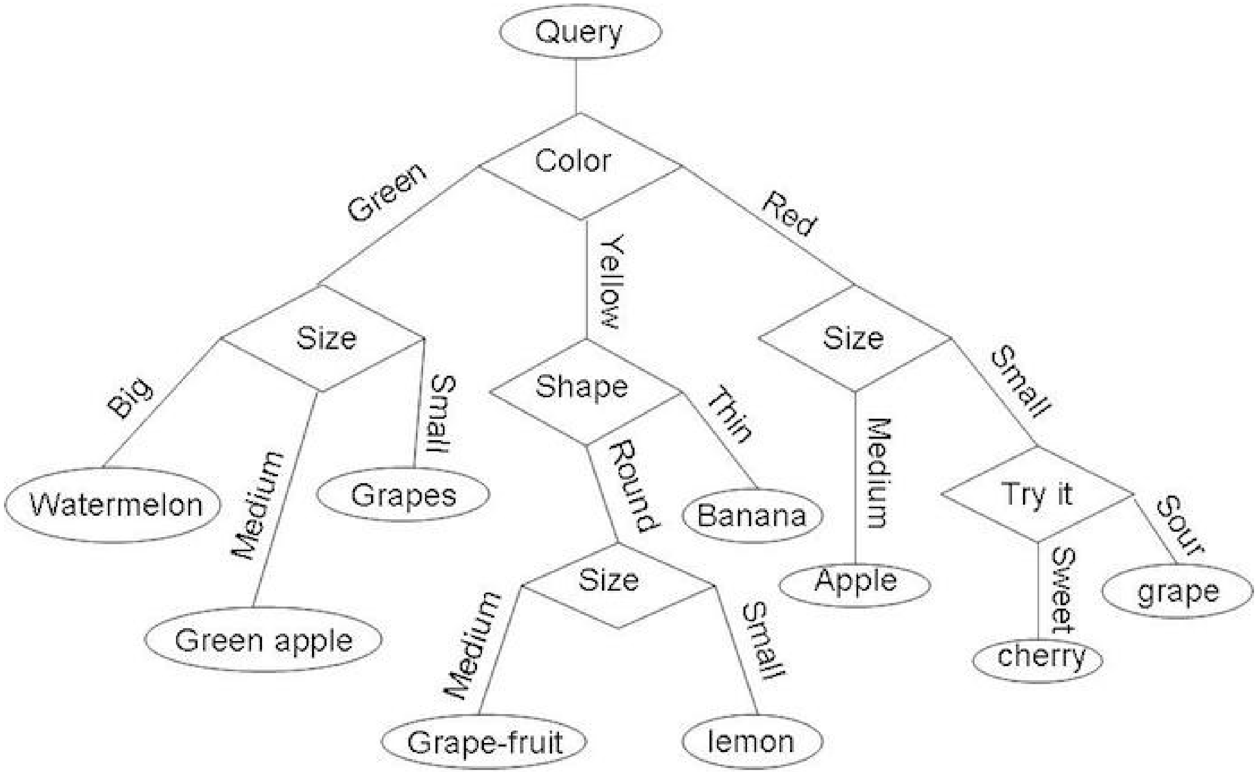
Where is Knowledge representation used?

ConversationalAI

Example:

Is tomato a fruit or vegetable?
Is avocado a fruit or vegetable?

- Flowering plants that disseminate seeds.
- Fruit families and Avocados belong to Berry family.



Example: Decision Trees: an Intuitive Form of Machine Learning

https://medium.com/@supernova22/decision-trees-4d614c6d9acf

Convserational AI example 2:

User: I want to get coffee from Starbucks. Well am hungry, am tired from homeworks, I might as well get a breakfast wrap.

ConversationalAI: Oh that's a great brunch from Starbucks! Do you like me to order for you to pick up?

User: yes, I'd love that. And I'd also like a Regular size coffee.

ConversationalAI: I don't mind being your genie. Alright, I placed the order.

Knowledge entity: Starbucks is a Entity, Starbucks "sells" coffee and breakfast wraps. Starbucks takes pickup orders. Model output resembles as follows:

entity(Coffeshops) -> entityName(Starbucks) ->sells(coffee) -> sells(breakfastWrap) -> pickupOrder(entity(FoodItemsSold))

entity(FoodItemsSold) -> coffee -> wrap -> greenTea -> chips -> cookies

Why do we want to achieve this goal?

Knowledge is everything for humans and machines

Knowledge is derived from facts, reasoning and represented from learning concepts. - holds true for AI.

ex, in ConversationalAI or an Q&A AI task

https://allenai.org/data/strategyqa

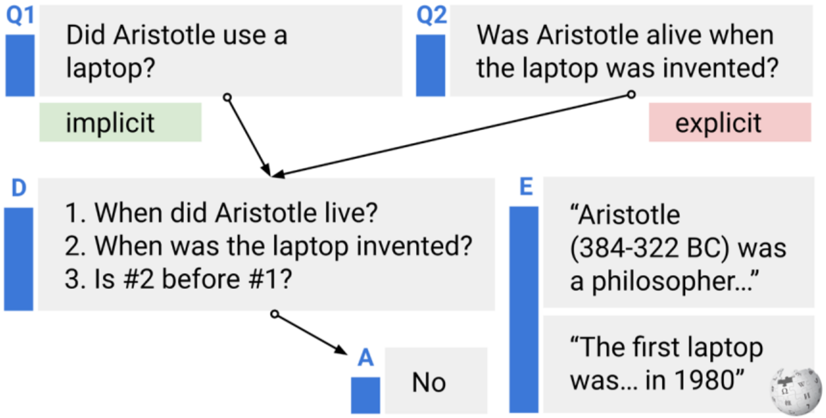
https://allenai.org/data/entailmentbank

Goal:
To understand reasoning for Knowledge Representation

1. Conduct Entity Extraction using NER, BERT, SpaCY libraries on 3 datasets.
    a. Entity Extraction is a text analysis process for representing Named entities in NLP.
2. Explore Implicit Reasoning technique in Knowledge representation
3. Explore Entailment analysis of sentences as facts in Knowledge representation of hypothesis.
4. Explore WikiQA dataset of Knowledge representation.

Datasets: strategyQA by AllenAI, Aristo

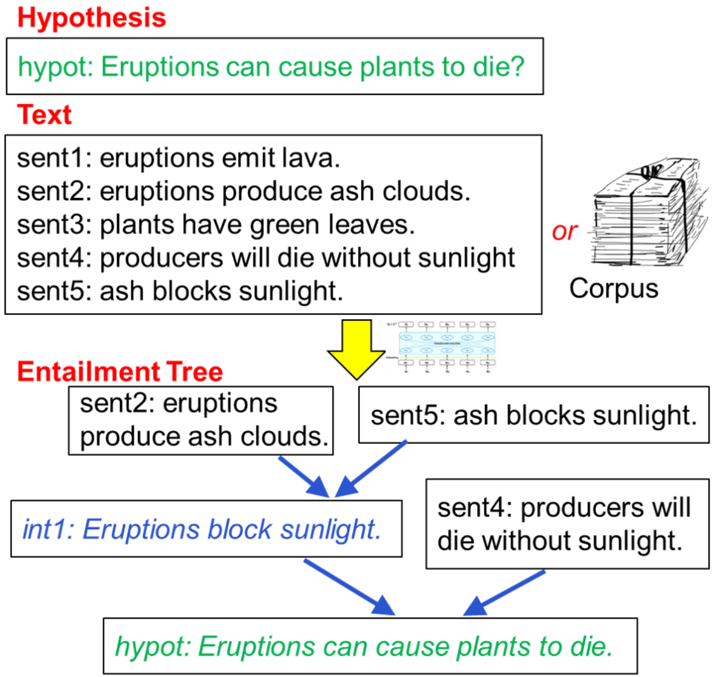
- Question
- Set of facts (i.e. Axioms)
- Decomposition
- Evidence is a representation (concepts/entities)
- And there are True/False questions
- Topic wise Paragraph texts.



[Demo: https://aristo-demo.allenai.org/ask?q=Which%20object%20is%20the%20best%20conductor%20of%20electricity%3F%20(A)%20metal%20fork%20(B)%20rubber%20boot%20(C)%20plastic%20spoon%20(D)%20wooden%20bat](https://aristo-demo.allenai.org/ask?q=Which%20object%20is%20the%20best%20conductor%20of%20electricity%3F%20(A)%20metal%20fork%20(B)%20rubber%20boot%20(C)%20plastic%20spoon%20(D)%20wooden%20bat)

Datasets: EntailmentBank by AllenAI, Aristo

- Hypothesis (single line text)
- ID
- Text or corpus with facts on each topic
- The model will derive entailment Tree as output



Reference:

https://www.semanticscholar.org/paper/Explaining-Answers-with-Entailment-Trees-Dalvi-Jansen/cf68637f53b0107a93dfefc4fe52bd596cceb017

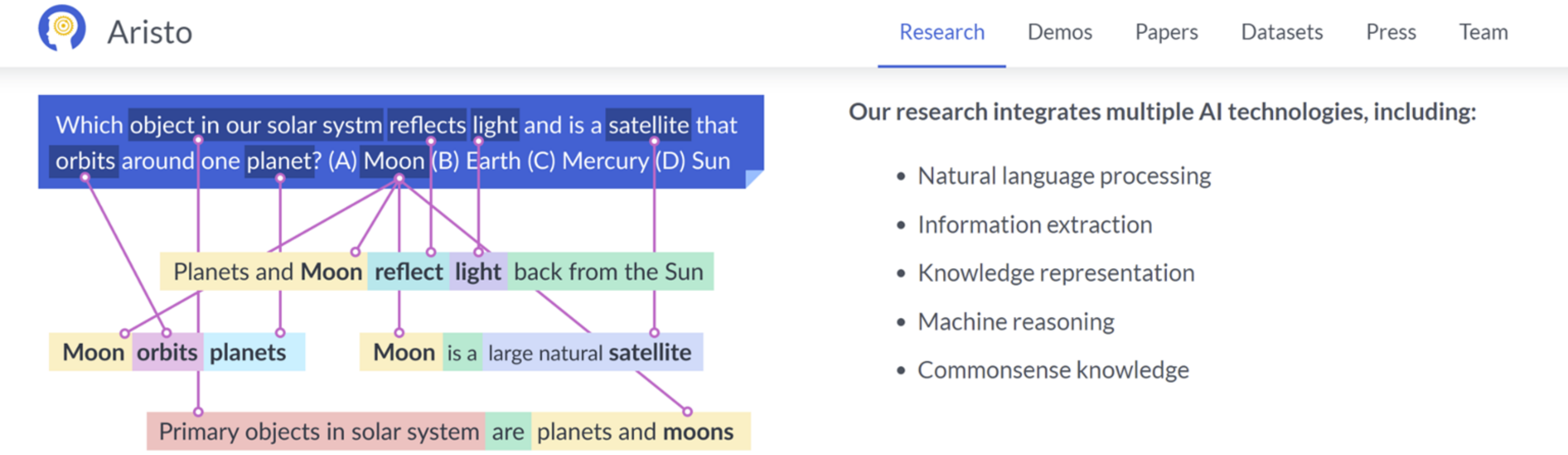
Datasets : WikiQA

- Question
- Choices
- 0 or 1 marked for each answer (1 - for right answer, 0 - for right answer)

how much is 1 tablespoon of water	This tablespoon has a capacity of about 15 mL .	1
how much is 1 tablespoon of water	In the US and parts of Canada , a tablespoon is the largest type of spoon used for eating from a bowl .	0
how much is 1 tablespoon of water	In countries where a tablespoon is a serving spoon , the nearest equivalent to the US tablespoon is either the dessert spoon or the soup spoon .	0
how much is 1 tablespoon of water	The capacity of ordinary tablespoons is not regulated by law and is subject to considerable variation .	0
how much is 1 tablespoon of water	In the USA one tablespoon (measurement unit) is approximately 15 mL ; the capacity of an actual tablespoon (dining utensil) ranges from 7 mL to 14 mL .	1
how much is 1 tablespoon of water	In Australia one tablespoon (measurement unit) is 20 mL .	1

Reference: https://www.microsoft.com/en-us/research/publication/wikiqa-a-challenge-dataset-for-open-domain-question-answering/#!related_info

Project Aristo



Credits: https://allenai.org/aristo

Approach:

Understand Research Papers on Knowledge Representation and analysis

- Study each dataset and its corresponding research paper:
- https://www.researchgate.net/publication/50194189_Comparative_Study_of_Three_Declarative_Knowledge_Representation_Techniques
- **Implicit Reasoning Dataset for Knowledge representation**
- **Entailment dataset for Knowledge representation**
- **Microsoft Open Data : WikiQA corpus**
- Test with known NLP models BERT, SPaCY, NLTK, CNNs, Transformers.
- Papers use different Approaches and Evaluation methods.

Evaluation of metrics and performance of Knowledge

- Expressiveness of the Knowledge (in formal logic)
- Ambiguity of the Knowledge graph - is it strictly unique or not.
- Number of Entailments given a sentence from test set or corpus
- Number of Incorrect Entailments given a sentence from test set or corpus

“Entailments: the deduction or implication, that is, something that follows logically from or is implied by something else.”

https://vocab.com/d/entailment

Milestones

Nov 16	Evaluate of metrics
Nov 2	Explore traditional knowledge graphs for knowledge representation of one of the datasets.
Oct 27	Explore Entailment analysis of sentences as facts in Knowledge representation of hypothesis.
Oct 20	Explore Implicit Reasoning technique in Knowledge representation
Oct 11	Conduct Entity Extraction using NER, BERT, SpaCY libraries on 3 datasets
Oct 4	Literature review

Questions?

KRR: Project Checkpoint

Sushma Akoju, Waad Alharthi

Team 10

Table of Contents

Approach

Goal : Understand existing approaches towards knowledge representation.

1. Reach out to Professors and researchers who published the datasets and seeking guidance on current methods.
2. To learn the approaches that best suit the knowledge representation.
3. To implement, test and analyze each of the methods.

Recap of our project

- Knowledge is set of axioms or facts. (AIMA, 3rd edition)
- “...premise or starting point for reasoning”
- Conduct Entity Extraction using NLTK, SpaCY libraries and BERT model on the 3 datasets.
- Explore Implicit Reasoning technique in Knowledge representation
- Explore Entailment analysis of sentences as facts in Knowledge representation of hypothesis.
- Explore WikiQA dataset of Knowledge representation.

What progress have you made for your project?

- [x] Literature review of the published papers for each data set.
- [x] Understanding recent works on knowledge representation, semantic web and knowledge graphs.
- [x] Learning NLP models and architectures BERT, Transformer models.
- [x] Replicated an Entity Extraction code using NLTK, SpACY and BERT
- [x] Replicated the code for StrategyQA and EntailmentBank and ran the training and prediction results.

Guidance from Professors and Researchers:

- Prof. Di Wu on SemanticWeb representation.
- Prof. Martha Palmer on Abstract meaning representation
- Prof. Tom Mitchell on recent Knowledge representation and reasoning approaches
- Dr. Peter Clark (Author for the two research papers on StrategyQA and Entailment datasets)

What modifications have you made for your goal?

- We added more literature review.
- We added more NLP models and learning exercises
- We learnt that Knowledge representation and reasoning involves various NLP tasks over a knowledge base.
- We fine-grained our goals towards more current methods
- We added sub-goals, sub-tasks for our learning to each of tasks

Literature Review

For 3 Datasets

StrategyQA

EntailmentBank

Abstract Meaning Representation

WikiQ&A

StrategyQA

Sushma Akoju

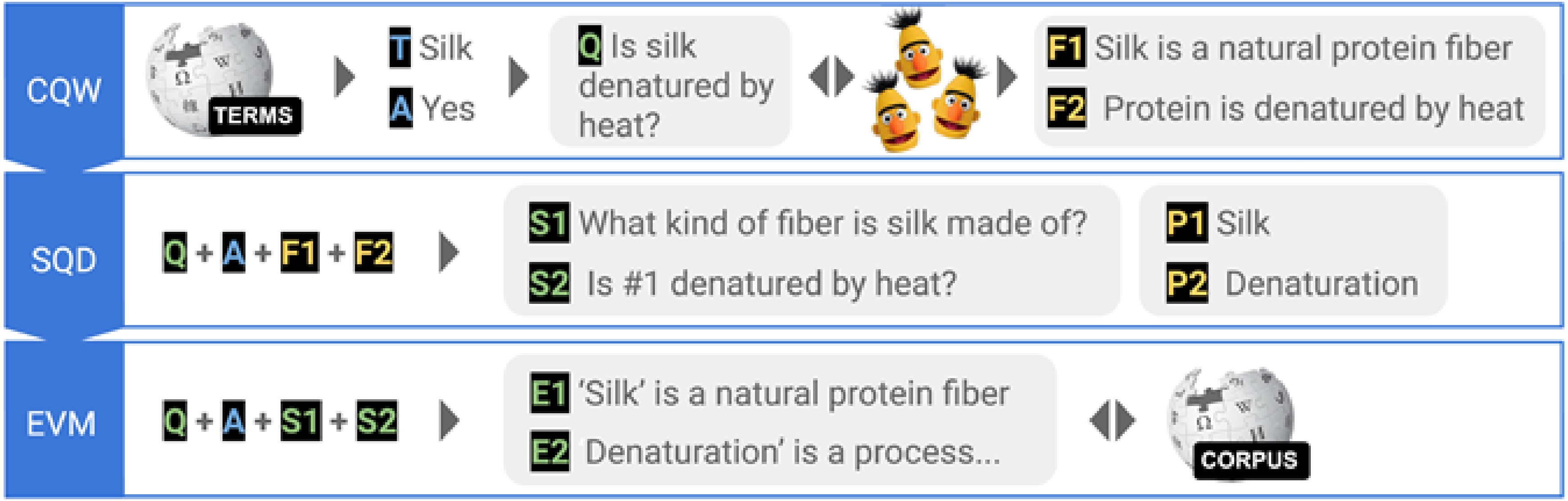
StrategyQA Dataset

Goal of the Paper: to create Strategy dataset for Multi-hop reasoning.

Repository: https://github.com/eladsegal/strategyqa

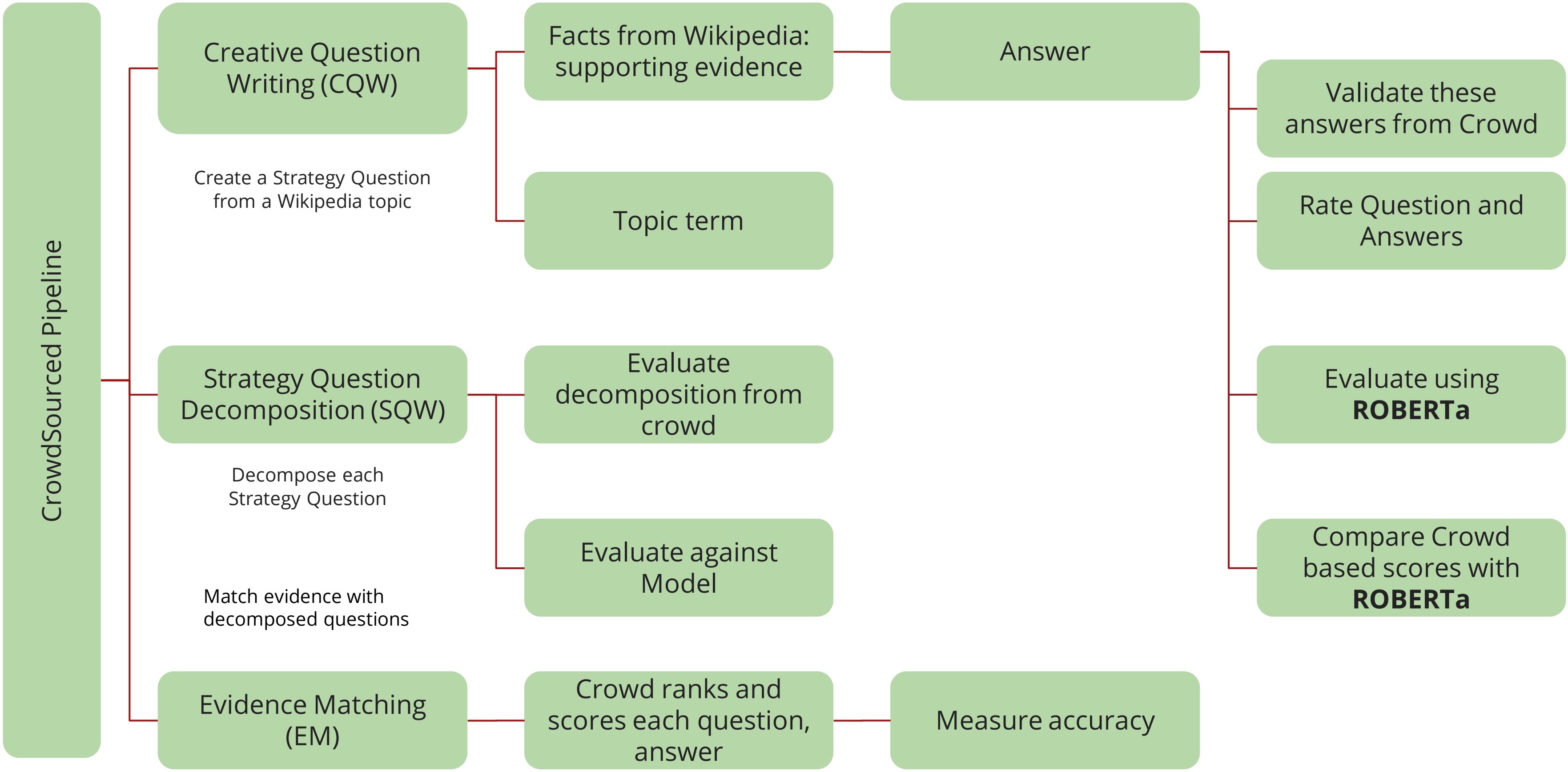
Multi-Hop/Implicit reasoning	Explicit reasoning
Did Aristotle have a laptop?	Was Aristotle Alive when laptop was invented?
Requires to decompose into steps to answer the question that is explicit within it.	Does not require decomposition.
Requires question decomposition and finding answers to each decomposed question.	Requires facts to directly answer the question.

StrategyQA : Data Collection Pipeline

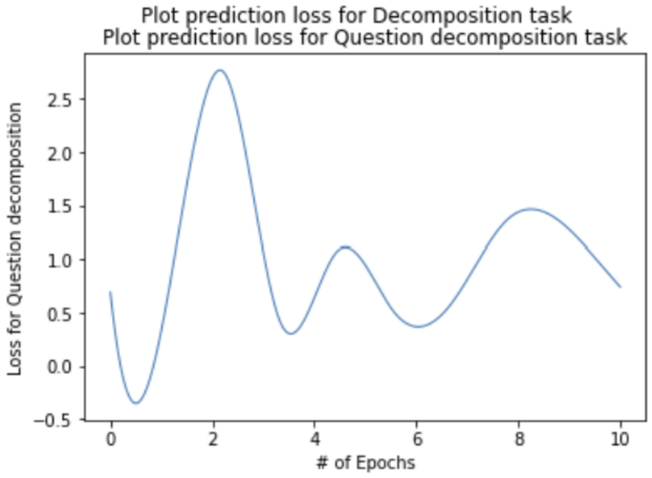


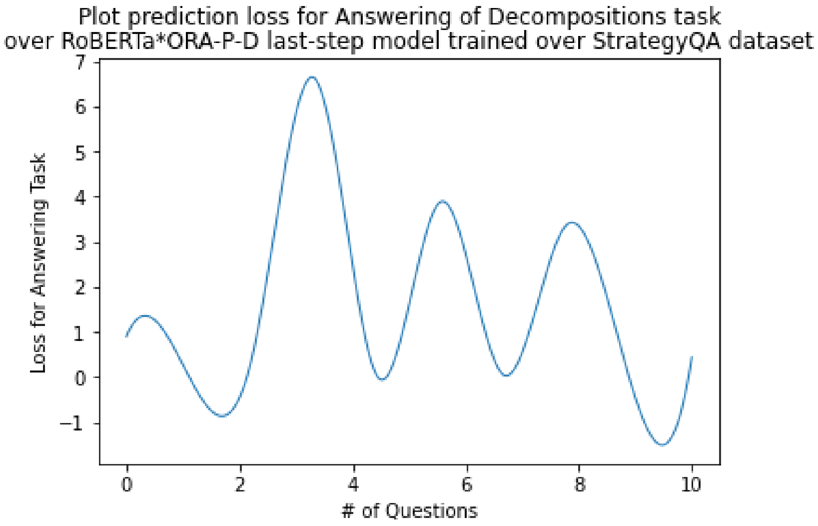
Creative Question Writing (CQW)
Strategy Question Decomposition (SQW)
Evidence Matching (EVW)

Crowdsourced Pipeline and ROBERTa



Cross-Entropy Loss over each of 3 tasks

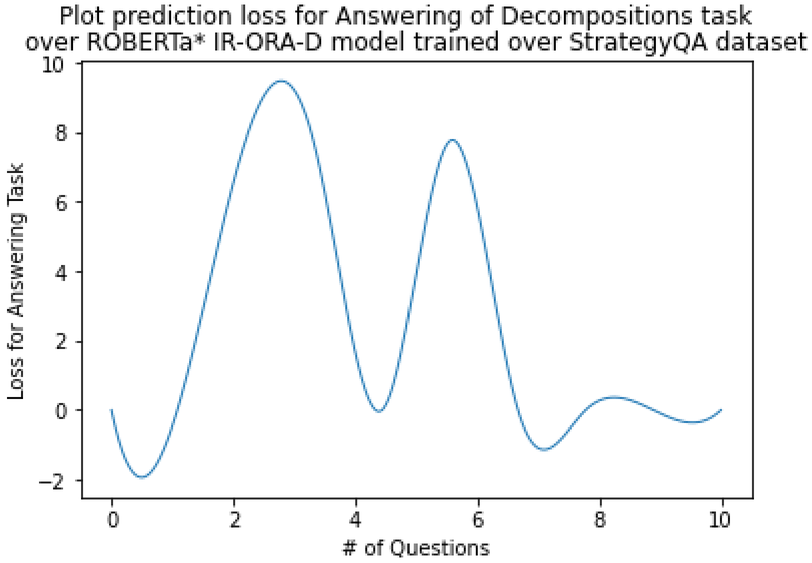





```
"rouge_ROUGE-L": 0.47535,
"bleu_BLEU": 0.2332
```

```
"accuracy": 0.5720524017467249,
"loss": 1.7002235415381501
```

Task 3 : Loss over RoBERTa* Iterative Retrieval




```
"accuracy": 0.6724890829694323,
  "loss": 2.038126806672612
```

Evaluation for Summarization and Quality of comparing Human judgement (crowdsource) vs Model results :

- **R**ecall-**O**riented **U**nderstudy for **G**isting **E**valuation **R**OUGE:

- **ROUGE-1** refers to the overlap of unigram (each word) between the system and reference summaries.

- **ROUGE-2** refers to the overlap of bigrams2 between the system and reference summaries.

- **ROUGE-L**: Longest Common Subsequence (LCS)[3] based statistics. Longest common subsequence problem takes into account sentence level structure similarity naturally and identifies longest co-occurring in sequence n-grams automatically.

- **BLEU** (bilingual evaluation understudy) is an algorithm for evaluating the quality of text.

StrategyQA evaluation results: Google colab

prediction: {"loss": 0.6949480175971985,

"decomposition": ["What is the population of Albany, Georgia?", "How many people are in the Albany, New York? Is #1 greater than or equal to #2?"], "qid": "e0044a7b4d146d611e73", "question": "Will the Albany in Georgia reach a hundred thousand occupants before the one in New York?", "gold_decomposition": ["What is the population of Albany, Georgia?", "What is the population of Albany, New York?", "What is the difference between 100,000 and #1?", "What is the difference between 100,000 and #2?", "Is #3 smaller than #4?"]}

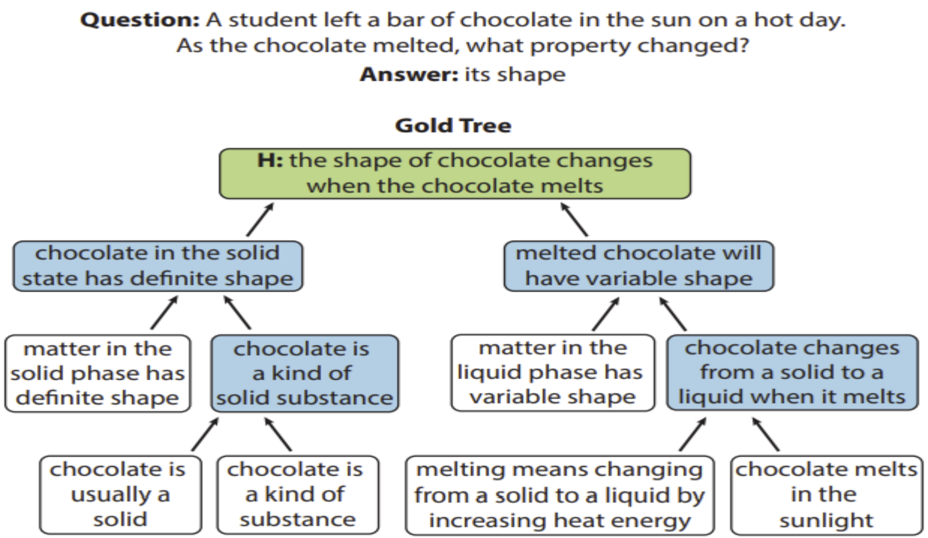
EntailmentBank

Sushma Akoju

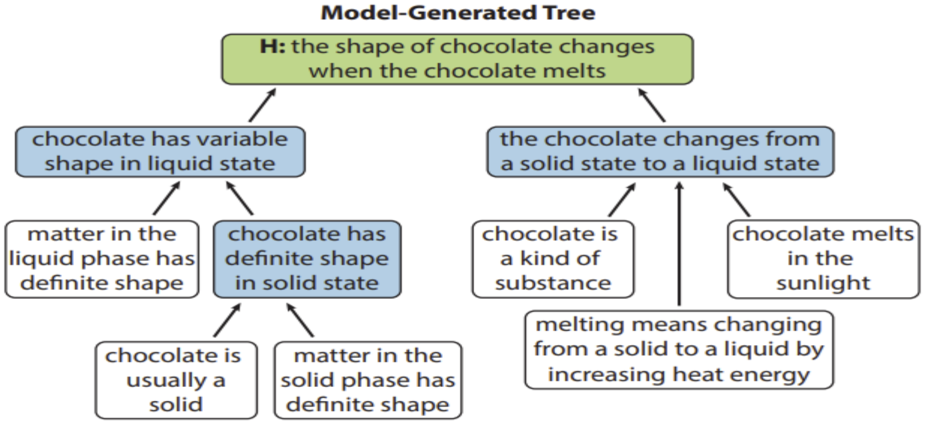
EntailmentBank Dataset

- Goal: to find *“line of reasoning”* for answering a natural question.
- The repository was made available recently a week ago once we reached out to the authors (was marked private): https://github.com/allenai/entailment_bank/
- Some parts of resources required to execute the scripts are missing.
- Algorithms used for comparing Human generated Gold tree with Model generated Entailment tree
    - Tree Alignment Algorithm - for Comparing Gold tree with Model generated Tree
    - Relevant Fact Retrieval Algorithm - Tensorflow-Ranking-BERT

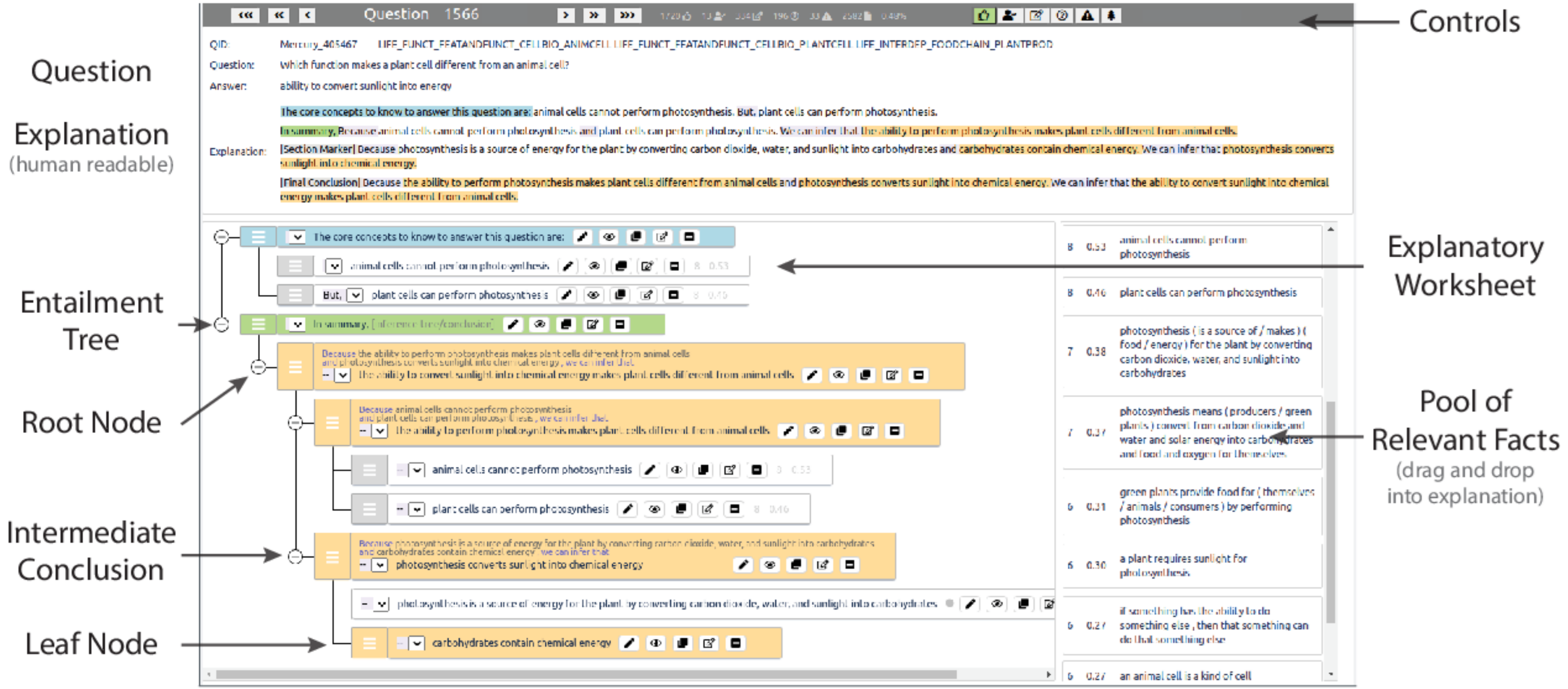
Gold Tree built by Crowdsourced authors



Model Generated Tree



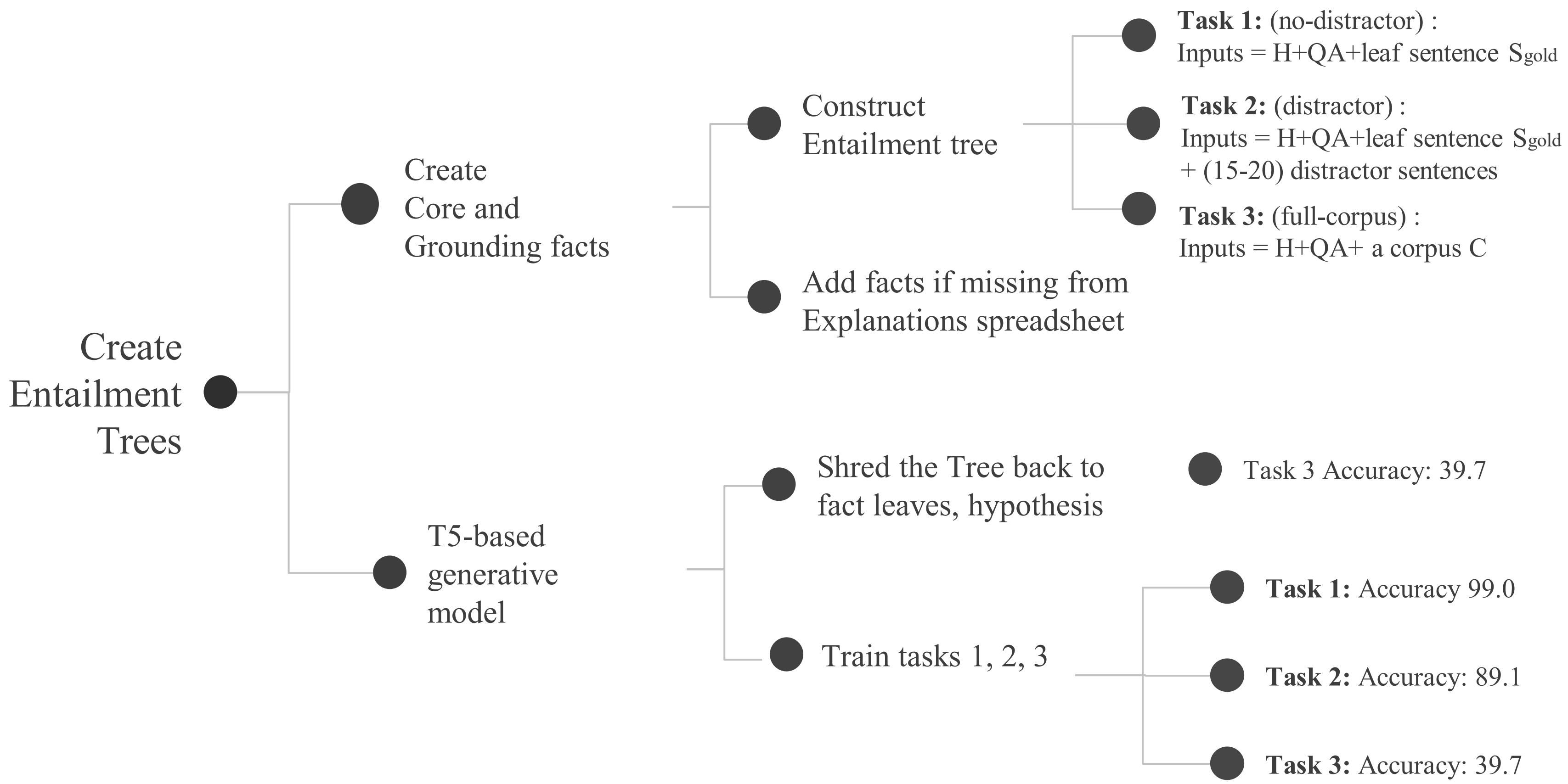
Crowdsourced Entailment Tree Authoring tool



Types of Inferences Addressed by the Dataset

Inference Type	Prop.		Example Entailment
Substitution	42%	s_1	when a light wave hits a **reflective object**, the light wave will be reflected
		s_2	a **mirror** is a kind of **reflective object**
		int	when a light wave hits a **mirror**, the light wave will be reflected
Inference from Rule	33%	s_1	if **two species have similar characteristics**, **they may share a common ancestor**
		s_2	**rhinoceroses and horses have similar characteristics**
		int	**rhinoceroses and horses might share a common ancestor**
Further Specification or Conjunction	15%	s_1	an animal requires **warmth** for survival **as the season changes to winter**
		s_2	**thick fur can be used for keeping warm**
		int	**thick fur can be used for keeping warm as the season changes to winter**
Infer Class from Properties	4%	s_1	**A compound** is **made of two or more elements chemically combined**
		s_2	**sodium chloride** is **made of two elements chemically combined**
		int	**sodium chloride** is a kind of **compound**
Property Inheritance	4%	s_1	**an animal's shell** is **usually hard**
		s_2	**something hard** can be used for **protection**
		int	**an animal's shell** is **usually hard** for **protection**
Sequential Inference	3%	s_1	In molecular biology, **translation** follows **transcription**
		s_2	**transcription** is when genetic information **flows from DNA to RNA**
		s_3	**translation** is when genetic information **flows from RNA to proteins**
		int	In molecular biology, genetic information **flows from DNA to RNA to proteins**

Entailment Bank research Pipeline



WikiQ&A

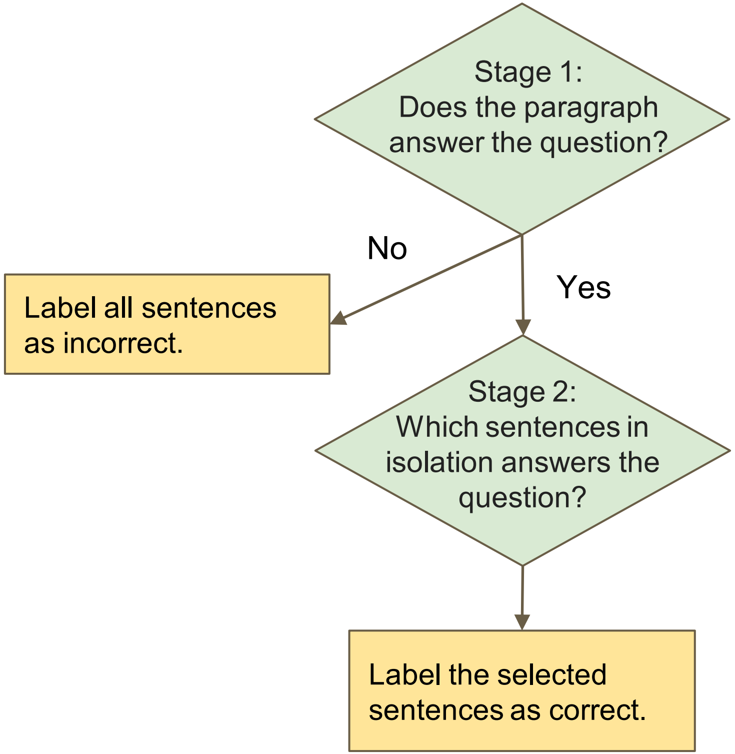
Waad Alharthi

WikiQ&A Corpus

- A set of question and sentence pairs, collected and annotated for research.
- Questions are collected from users Bing search queries from 5 unique users.
- Candidate sentences are collected from Wikipedia summary page.
- 62% of the questions don't have a correct answer.
- 36% are description questions which are harder to answer than numeric or location questions.
- 20.3% of the answers don't share words with the question to avoid the lexical overlap.

WikiQ&A - Labelling candidate sentences

- Each question candidate answers are labeled by three workers.
- Inconsistent labels are verified by a different group of workers.



WikiQ&A - Evaluation

- Evaluation: question level precision, recall, F1 scores (~30%).
- Sentence semantic approaches such as the convolutional neural networks performed better than lexical semantics.
- Adding the answer sentence length feature improved the results slightly,
- While adding the question class feature worsened the results slightly.

WikiQnABot

WikiQnABot using Multi-head attention

- We worked on implementing a Question and Answer Bot
- We need to evaluate Mean reciprocal rank

WikiQ&A replicating original code by the authors:

- We went through the code
- Code uses outdated libraries such as Theano which is not supported anymore
- We could not replicate this code.
- Paper evaluates Convolutional Neural Networks (CNN), and Logistic regression with CNN.

NLP Models

Language Models

Language Understanding

Self-Attention Transformers

BERT

Learning

- Due to the research focus of our project, we had a learning curve in the following:
  - Knowledge representation and reasoning involves Language Understanding
  - Entity recognition requires real world entities
  - NLP neural network Language models are the key to understanding Knowledge representation and reasoning
  - Reasoning in AI is specialized and complex Artificial Neural network

Transformer : Self-Attention Operation

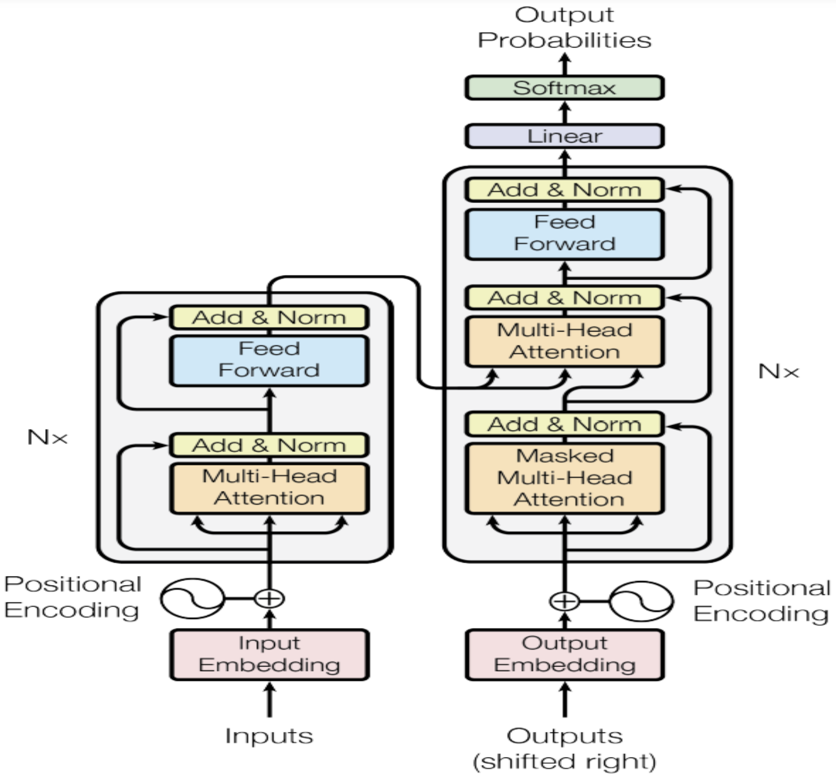


Figure 1: The Transformer - model architecture.

$$\mathrm{Attention}(Q,K,V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

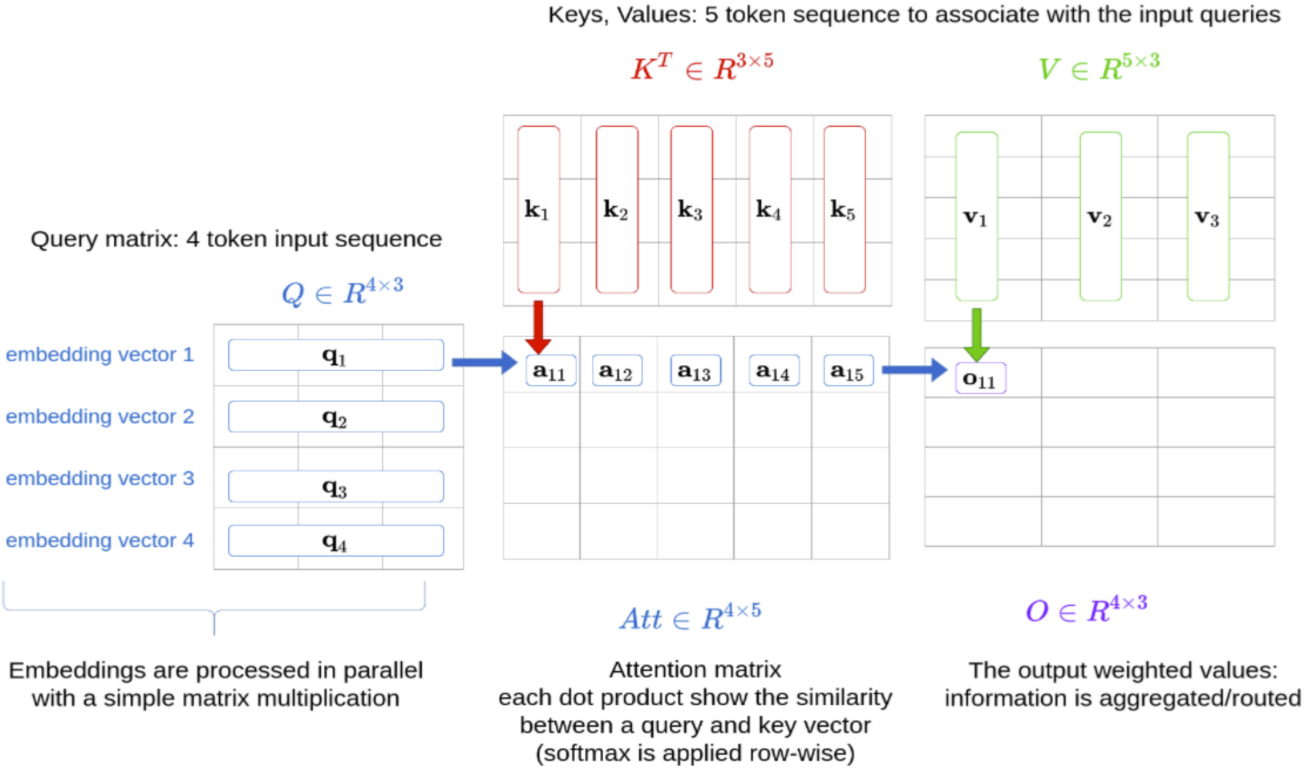
Attention is a weighted average of values.

Q - Query to search
K - Key to map Q to attributes
V - Value

Tells how related two words are in input sequence.

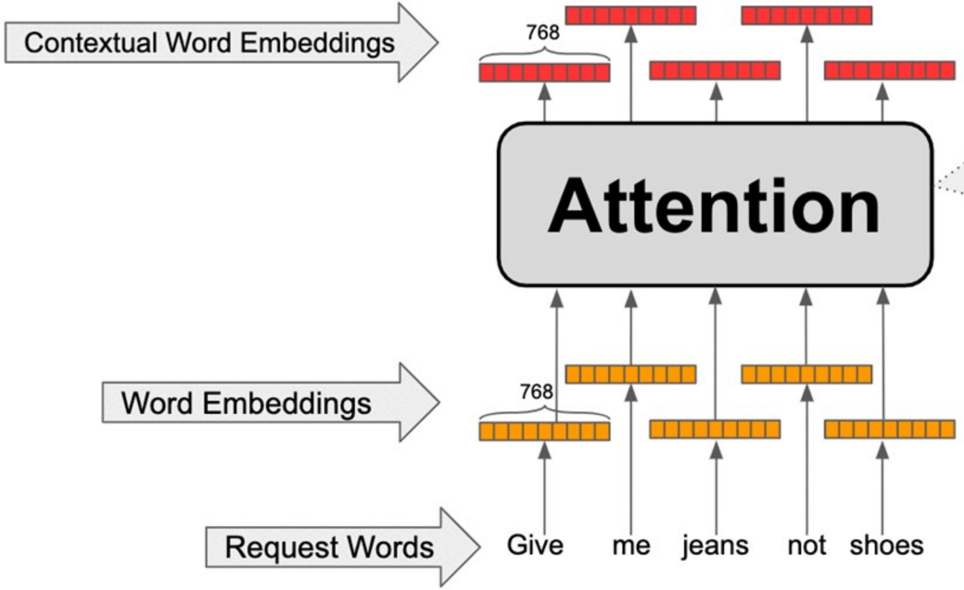
1) Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems.* 2017.
2) https://theaisummer.com/attention/

Self-Attention : Query, Key and Values.
Content-based attention has distinct representations

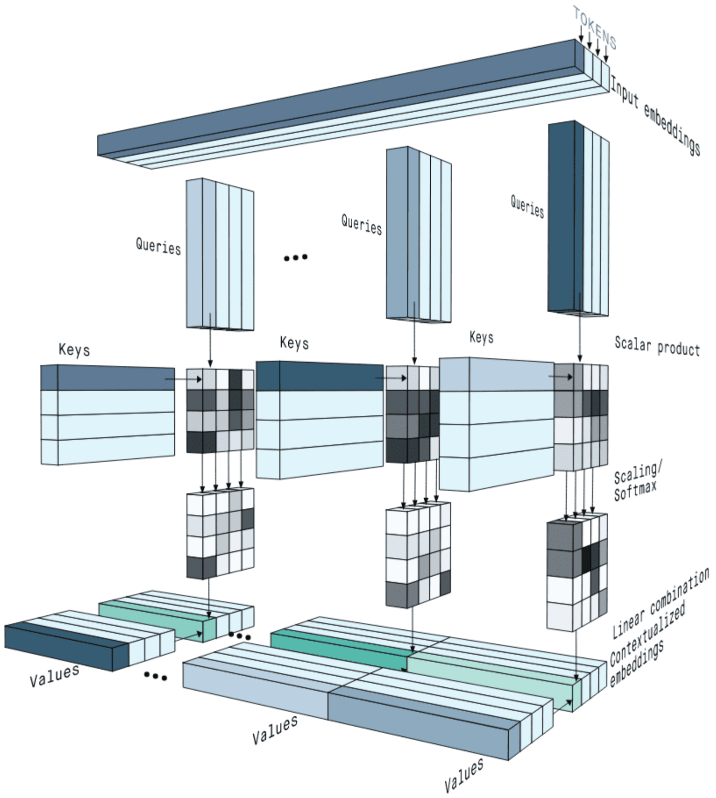


Query, Key and Value approach : BERT

1- Query and Keys Dot Product



Multihead Attention: words can mean different things to different neighbor words: Attention becomes a routing algorithm



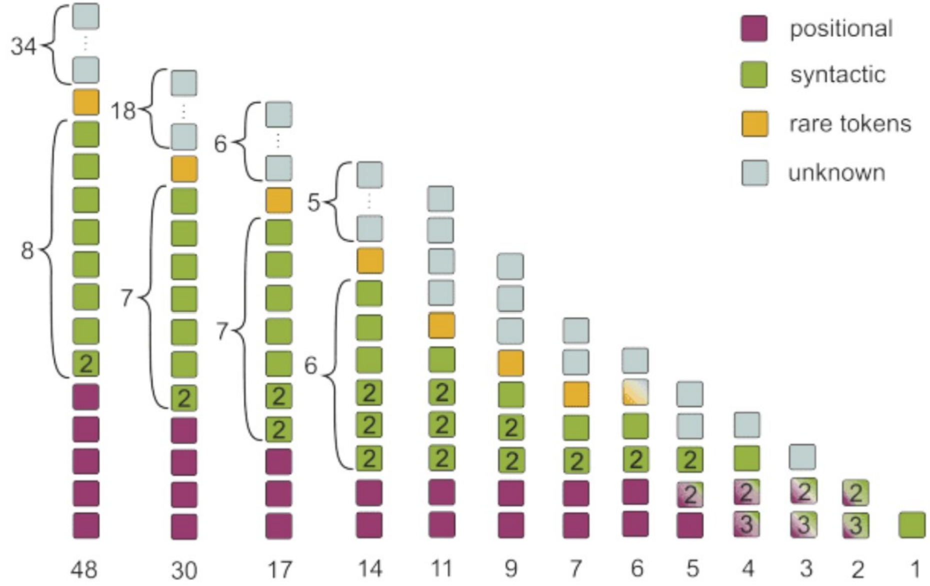
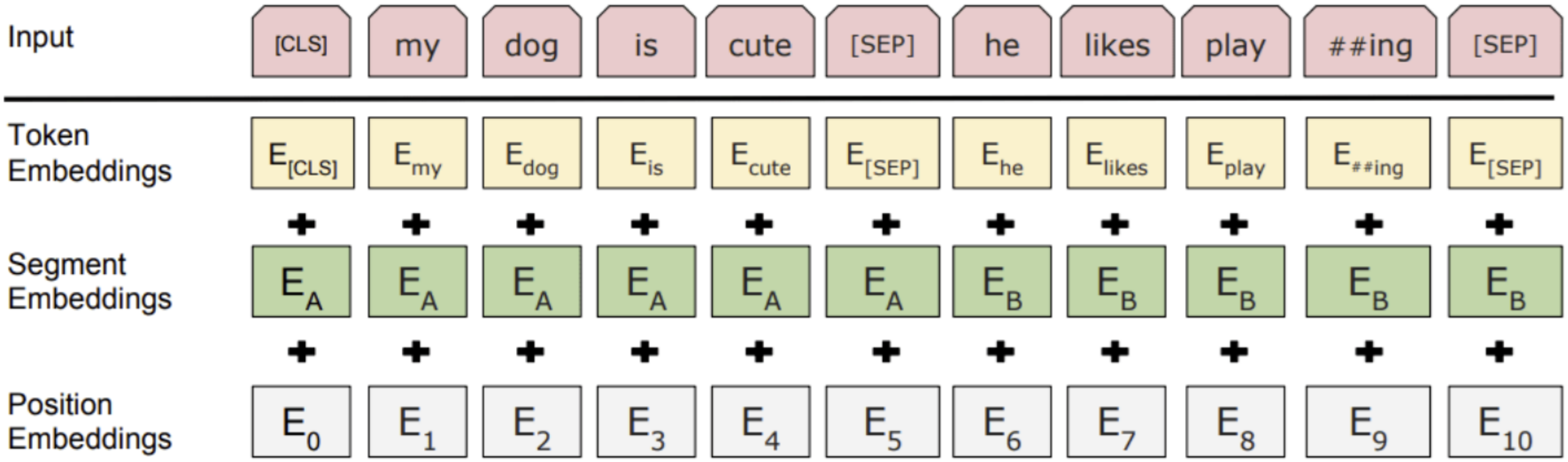


Image by Voita et al. Source: Analyzing Multi-Head Self-Attention: Specialized Heads Do the Heavy Lifting, the Rest Can Be Pruned

Embeddings for Understanding contexts



Ways to compute Attention

Name	Alignment score function	Citation
Content-base attention	$\text{score}(\boldsymbol{s}_t, \boldsymbol{h}_i) = \text{cosine}[\boldsymbol{s}_t, \boldsymbol{h}_i]$	Graves2014
Additive(*)	$\text{score}(\boldsymbol{s}_t, \boldsymbol{h}_i) = \mathbf{v}_a^\top \tanh(\mathbf{W}_a[\boldsymbol{s}_t; \boldsymbol{h}_i])$	Bahdanau2015
Location-Base	$\alpha_{t,i} = \text{softmax}(\mathbf{W}_a \boldsymbol{s}_t)$ Note: This simplifies the softmax alignment to only depend on the target position.	Luong2015
General	$\text{score}(\boldsymbol{s}_t, \boldsymbol{h}_i) = \boldsymbol{s}_t^\top \mathbf{W}_a \boldsymbol{h}_i$ where $\mathbf{W}_a$ is a trainable weight matrix in the attention layer.	Luong2015
Dot-Product	$\text{score}(\boldsymbol{s}_t, \boldsymbol{h}_i) = \boldsymbol{s}_t^\top \boldsymbol{h}_i$	Luong2015
Scaled Dot-Product(^)	$\text{score}(\boldsymbol{s}_t, \boldsymbol{h}_i) = \frac{\boldsymbol{s}_t^\top \boldsymbol{h}_i}{\sqrt{n}}$ Note: very similar to the dot-product attention except for a scaling factor; where n is the dimension of the source hidden state.	Vaswani2017

BERT

1. Learn a weight from each input state to each output state.
2. Each Weight represents a correlation of each input state to each other input state.
3. We want the weights to sum up to 1.
4. Query: compute meaning vector for your own output
5. Key: compute relationship between word_i in input and word_j in output
6. Value: compute relationship between word_i and all outputs

Entity Extraction

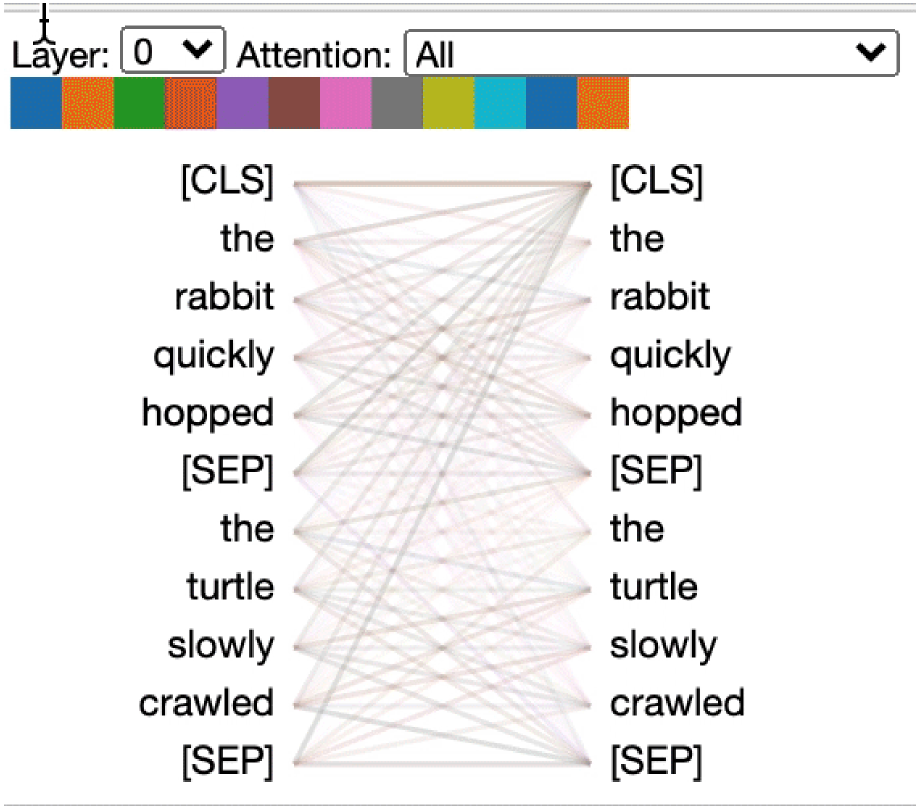
What is Entity extraction?

It is a subtask of Information Extraction that seeks to locate and classify named entities from unstructured textual data.

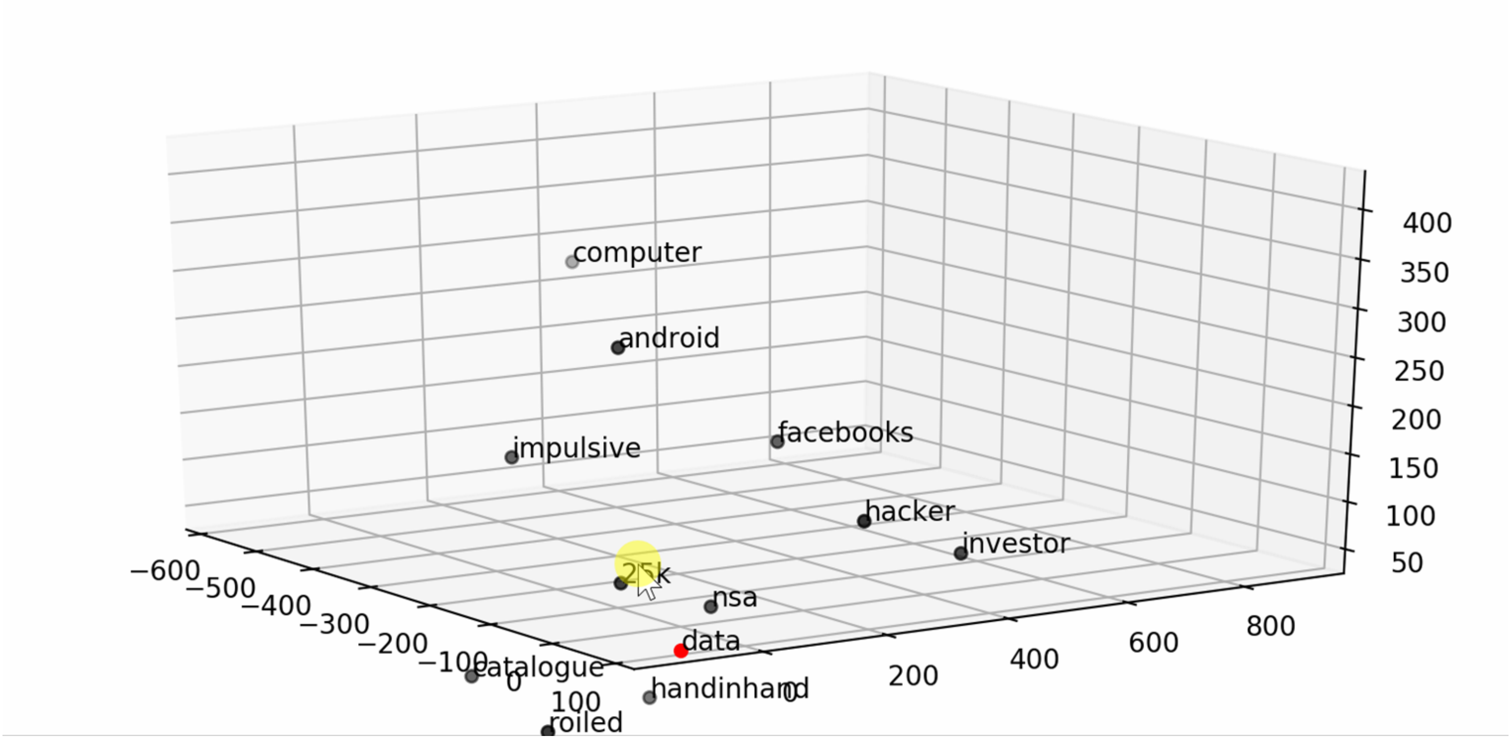
Example of BERT named entity recognition (NER) task using SpaCY:

Prime Minister Jacinda Ardern PERSON has claimed that New Zealand GPE had won a big battle over the spread of coronavirus.

BERT : for Language Understanding



BERT applications : Question and Answering system with SQuAD



BERT: Important steps

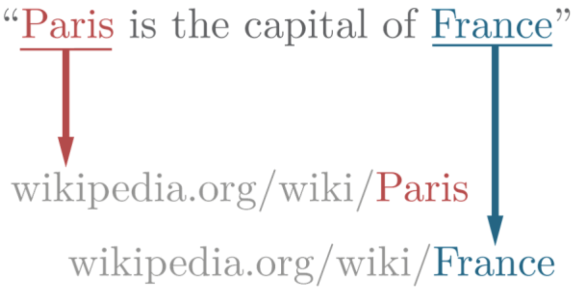
1) Masked LM (MLM): Randomly mask out 15% of the words in the input — replacing them with a [MASK] token — run the entire sequence through the BERT attention based encoder and predict [MASK]ed words
2) Next Sentence Prediction : NSP - a key step to BERT training process.

Input = [CLS] the man went to [MASK] store [SEP]
he bought a gallon [MASK] milk [SEP]
Label = IsNext

Input = [CLS] the man [MASK] to the store [SEP]
penguin [MASK] are flight ##less birds [SEP]
Label = NotNext

BERT: Named Entity Recognition & Question and Answering System

Prime Minister **Jacinda Ardern** PERSON has claimed that **New Zealand** GPE had won a big battle over the spread of coronavirus.



- **Input Question:**

  `Where do water droplets collide with ice crystals to form precipitation?`

- **Input Paragraph:**

  `... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals within a cloud. ...`

- **Output Answer:**

  `within a cloud`

Additional Milestones

Status

Milestones

Status

Learning

Task for final project

Tasks for final project completion

- [ ] Implement and test Bert with Squad Q&A system - StrategyQA and WikiQA
- [ ] Implement and test Roberta with Squad Q&A system - StrategyQA and WikiQA
- [ ] Theoretical aspect: Abstract Meaning representation with entity and relation representation.
- [ ] Complete project report with analysis

Future work: outside of project

1) Literature review: summaries for 20 papers

Questions?

Thank you!

Special thanks for the guidance from the Professors and researchers.

KRR: Final Project Presentation

Sushma Akoju,Waad Alharthi

Table of Contents

Named Entity Recognition (NER) Using Spacy

Text example from the book "James and the Giant Peach"

```
displacy.render(text, style = 'ent', jupyter=True)
```

Their names were Aunt Sponge **ORG** and Aunt Spiker **PERSON** , and I am sorry to say that they were both really horrible people. They were selfish and lazy and cruel, and right from the beginning they started beating poor James **PERSON** for almost no reason at all. They never called him by his real name, but always referred to him as "you disgusting little beast" or "you filthy nuisance" or "you miserable creature," and they certainly never gave him any toys to play with or any picture books to look at. His room was as bare as a prison cell. They lived -- Aunt Sponge **ORG** , Aunt Spiker **PERSON** , and now James **PERSON** as well -- in a queer ramshackle house on the top of a high hill in the south of England **GPE** . The hill was so high that from almost anywhere in the garden James **PERSON** could look down and see for miles and miles **QUANTITY** across a marvelous landscape of woods and fields; and on a very clear day **DATE** , if he looked in the right direction, he could see a tiny gray dot far away on the horizon, which was the house that he used to live in with his beloved mother and father. And just beyond that, he could see the ocean itself -- a long thin streak of blackish **NORP** -blue, like a line of ink, beneath the rim of the sky.

Customizing a NER model using Spacy

Text example from the book "James and the Giant Peach"

[65]:

```
test  = nlp_model("Because you are miserable, aren't you? You needn
displacy.render(test, style = 'ent', jupyter=True)
```

Because you are miserable **HURTFUL** , aren't you? You needn't tell me! I know all about it! Now, off you go and do exactly as I say. And don't whisper a word of this to those two horrible aunts of yours! Not a word! And don't let those green things in there get away from you either! Because if they do escape, then they will be working their magic upon somebody else instead of upon you !

WikiQA

[52]:

```
output = predict("how glacier caves are formed?", model)
```

```
Input: how glacier caves are formed?
Output: ice
```

[51]:

```
output = predict("Are glacier caves formed with ice?", model)
```

```
Input: Are glacier caves formed with ice?
Output: the
```

[61]:

```
output = predict("how glacier caves are made?", model)
```

```
Input: how glacier caves are made?
Output: ice
```

Backstage

- Which part is most difficult of your project?
  - Reading research papers
- Which part is most interesting of your project?
  - Visualizing the results
- If you could start it over, what would you modify?
  - Not stress too much..
- What is the most valuable thing you learned from conducting the project?
  - Labeling data is a project in itself.
  - Keeping up with the latest methods while learning the basics is hard.

Findings and work after Checkpoint

Q&A with BERT+ SqUAD

Finetuning BERT with decomposed questions from StrategyQA

Testing finetuned BERT + Squad with StrategyQA

Tasks for final project completion

- [x] Implement and test Pretrained Bert with Squad Q&A system - StrategyQA
- [ ] Implement and test Roberta with Squad Q&A system - StrategyQA
- [x] Fine-Tune and test BERT with Squad Q&A system - StrategyQA
- [x] Theoretical aspect: Abstract Meaning representation with entity and relation representation.

BERT with SQUaD vs Roberta with SQUaD

Did not work with Roberta with SQUAD due to change in direction for conducting BERT with SQUaD

First fine-tuned BERT with SQUaD over decomposed questions and answers paragraphs

Tested the fine-tuned BERT with SQUaD with new decomposed questions and answers

Pretrained BERT with SQUaD for Q&A task

Results: Using decomposed questions, facts for basic Q&A with BERT SQUAD

```
****************************************************************************************************
This is for decomposed questions with the facts without logical evidences. The answers could be one word or multiphrase answers
----------------------------------------------------------------------------------------------------------
Question"How many American troops were in Vietnam in 1965?"
Query has 103 tokens.

Answer: "125 , 000"
Question"Who cultivated the maize that Spaniards took to Europe from America in 1492?"
Query has 65 tokens.

Answer: "indigenous americans"
Question"What is Oscar Wilde's most famous book?"
Query has 96 tokens.

Answer: "what"
Question"What is disgust?"
Query has 143 tokens.

Answer: "the emotion"
Question"How many times was Kublai Khan married?"
Query has 124 tokens.

Answer: "one"
```

Results: Using decomposed questions, paragraphs for basic Q&A with BERT

```
Question"Would Donald Duck be allowed into most grocery stores?"
Query has 141 tokens.

Answer: "donald is an anthropomorphic white duck"
```

Drawbacks

Not able to answer questions for decomposed questions with facts and without paragraphs for multiple hops.

Fine-Tuned BERT with SQUaD over Q&A Task for StrategyQA dataset

Fine tuning BERT over StrategyQA Dataset

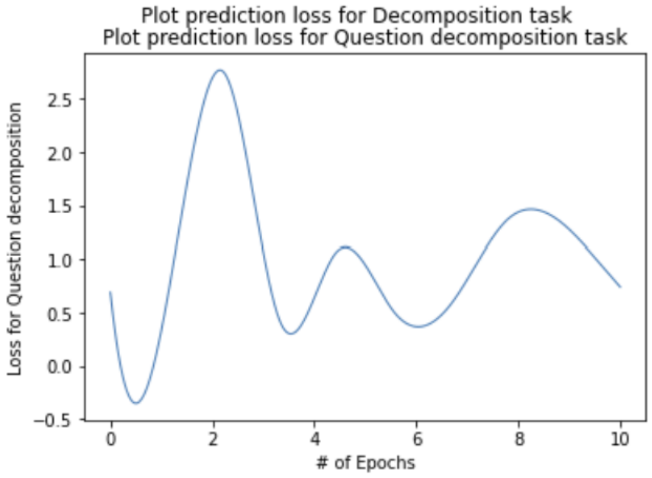
Decomposed questions (Our StrategyQA results from ROBERTa)

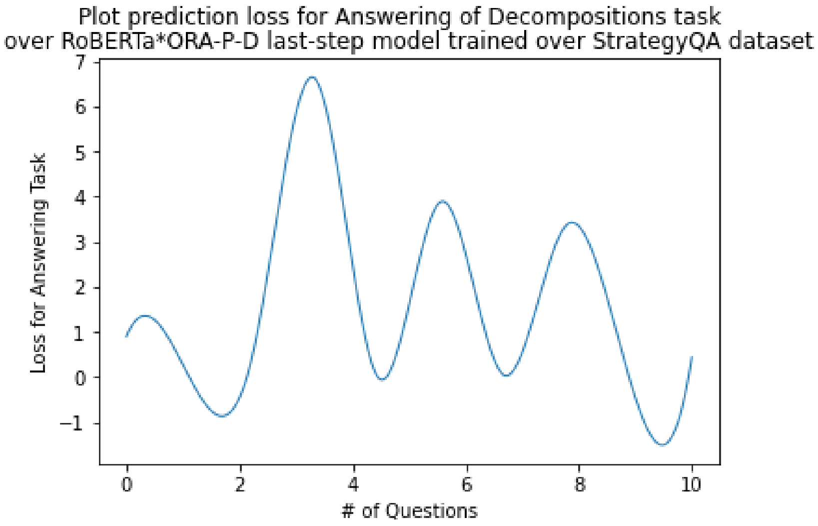
Facts with Paragraph Context as answers (SQUaD format)

This approach gave good results but using SQUAD approach is something prefer to avoid. Squad approach requires explicit start, end tags of answers.

We fine tuned BERT over StrategyQA without SQUAD, yet the results do not improve (better than Roberta).

Cross-Entropy Loss over each of 3 tasks: for more epochs and learning rate






```
"rouge_ROUGE-L": 0.47535,
"bleu_BLEU": 0.2332
```

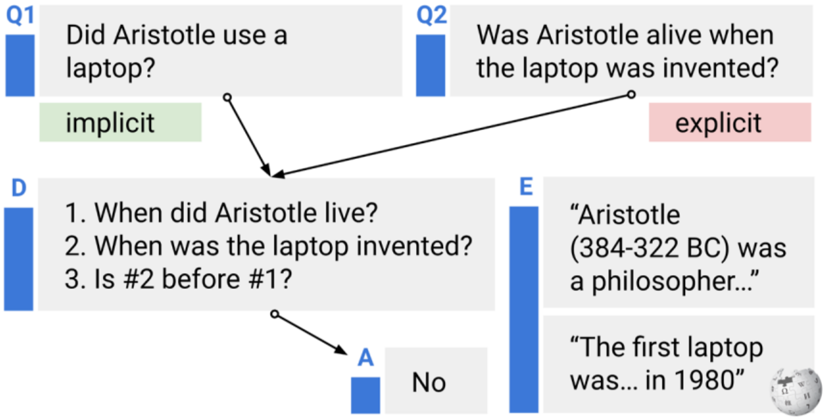
"accuracy": 0.5720524017467249,
"loss": 1.7002235415381501

Generalized Language Understanding Evaluation Tasks : Benchmark Tasks

Evaluating Fine-tuned BERT without SQUAD model over StrategyQA

The fine tuning of the model is executing and would take a while to run and evaluate it over GLUE task benchmarks for language understanding.

StrategyQA - Performance is ~66.5% accuracy



Solved puzzle in BERT Architecture

Before “The Annotated Transformer” :

RNN - *implemented and test to understand the RNN*

Bidirectional RNN

Gated Recurrent Unit (GRU)

Long Short Term memory - *implemented and test to understand the LSTM*

word2Vec and Glove - *implemented and test to understand the* word2Vec

Beam Search

Self-Attention

Multi-Head Attention

Transformer Network - *implemented and test to understand the* Transformer network

Exploring BERT

- ☑ Understanding the models from RNN to until Transformers, helped to understand BERT.

Abstract Meaning Representation and Semantic Parsing

Abstract Meaning representation (AMR) is a project at CU Boulder.

AMR is representation of English sentences with their whole-sentence, logical meanings.

AMR helps in Language understanding.

“Did Aristotle use a laptop?”

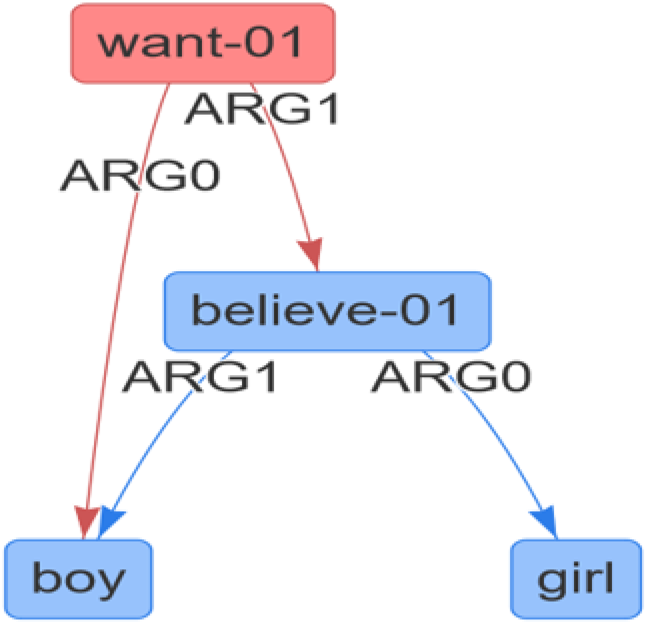
“Aristotle lived in 384-322 BC”

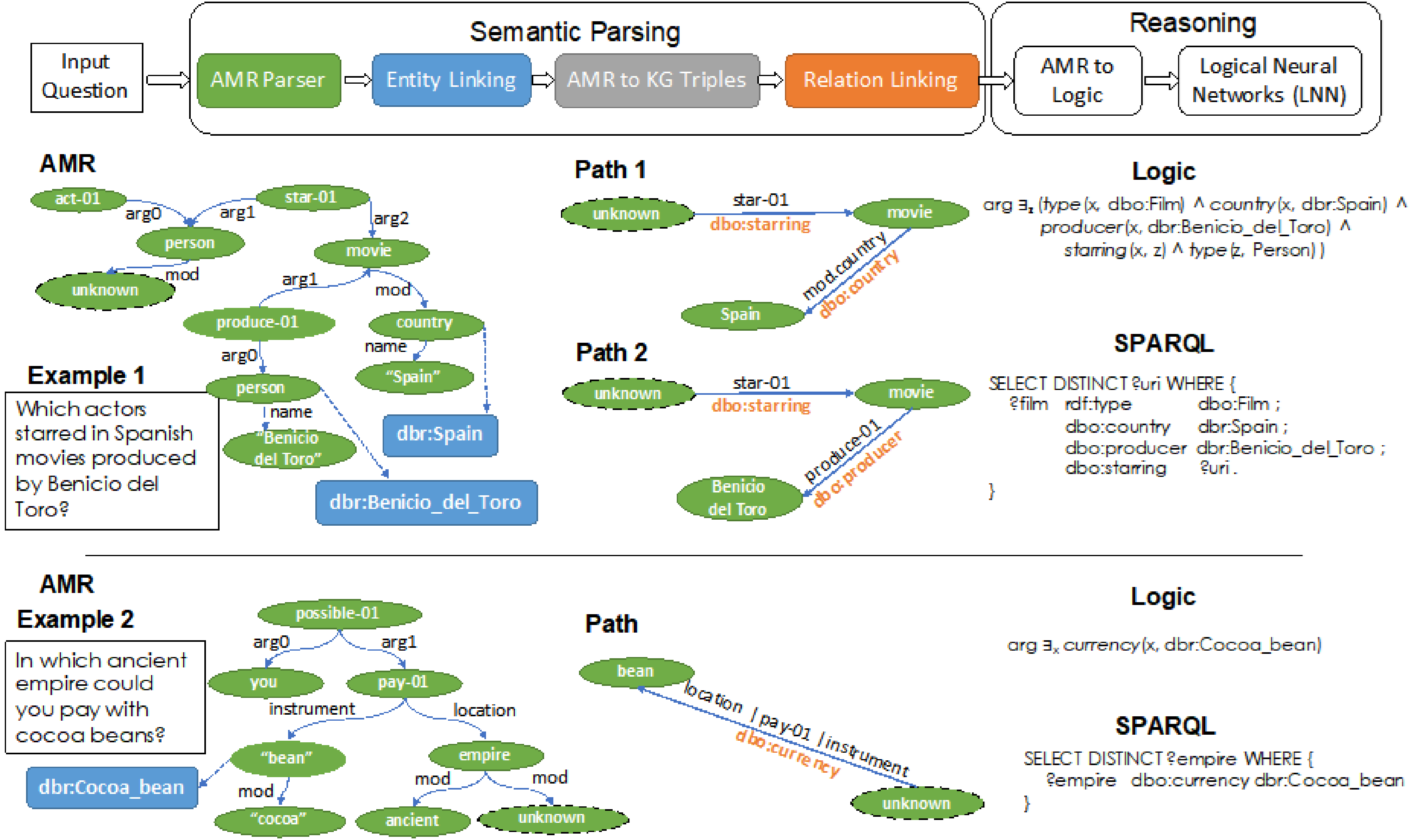
“The first laptop came in 1980.”

Explored Semantic Parsing and AMR

1. It has become essential to represent freeform sentences into logical forms.
2. AMR is research work by Prof. Martha Palmer, from CU Boulder.
3. We ran the test for AMR
4. https://github.com/IBM/transition-amr-parser
5. https://aclanthology.org/W13-2322.pdf

The boy wants the girl to believe him.
The boy wants to be believed by the girl.




Input Question
Semantic Parsing
AMR Parser
Entity Linking
AMR to KG Triples
Relation Linking
Reasoning
AMR to Logic
Logical Neural Networks (LNN)
AMR
act-01
star-01
arg0
arg1
arg2
person
movie
unknown
mod
arg1
mod
produce-01
country
name
arg0
"Spain"
Example 1
person
name
"Benicio del Toro"
dbr:Spain
dbr:Benicio_del_Toro
Which actors starred in Spanish movies produced by Benicio del Toro?
Path 1
unknown
star-01
dbo:starring
movie
mod.country
dbo:country
Spain
Path 2
unknown
star-01
dbo:starring
movie
produce-01
dbo:producer
Benicio del Toro
Logic
arg ∃z (type(x, dbo:Film) ∧ country(x, dbr:Spain) ∧ producer(x, dbr:Benicio_del_Toro) ∧ starring(x, z) ∧ type(z, Person))
SPARQL
SELECT DISTINCT ?uri WHERE {
?film rdf:type dbo:Film ;
dbo:country dbr:Spain ;
dbo:producer dbr:Benicio_del_Toro ;
dbo:starring ?uri .
}
AMR
Example 2
possible-01
arg0
arg1
you
pay-01
instrument
location
In which ancient empire could you pay with cocoa beans?
"bean"
empire
dbr:Cocoa_bean
mod
mod
mod
"cocoa"
ancient
unknown
Path
bean
location | pay-01 | instrument
dbo:currency
unknown
Logic
arg ∃x currency(x, dbr:Cocoa_bean)
SPARQL
SELECT DISTINCT ?empire WHERE {
?empire dbo:currency dbr:Cocoa_bean
}

Why your final finding matters?

- Reasoning adds additional “thoughtfulness” to an AI task.
- Reasoning from formal logic such as meaning Representation and generation of formal logical statements add to distributional hypothesis of Linguistics.
- I have more work to learn and implement

How did you reach the final findings? (Datasets, Methods, evaluations of your project)

- Knowledge representation and reasoning involves Language Understanding
- Entity recognition requires real world entities
- NLP neural network Language models are the key to understanding Knowledge representation and reasoning
- The Method followed: finetune BERT with SQUaD
- Evaluation: evaluate over unseen Test decomposed questions and answers
- Calculated loss over test question and answer paragraph pairs. It was bit fluctuating due to unseen words.
- Explored Abstract Meaning representation
- Explored Semantic parsing to convert natural language into formal logic

Future work (Winter break)

Build entire pipeline for StrategyQA as extension outside scope of the paper.

Train a Semantic parser for AMR

Train Logical neural networks for reasoning

Add Types of reasoning to the survey paper

Publish the paper

Where will your project impact the future? (Show your ambition)

The future work:

Since semantic parsing with AMR (abstract meaning representation) is the key to future AI tasks over Natural freeform texts, the project helped to learn and understand what new approaches can we add.

Reasoning is the thoughtfulness to NLP tasks- supported by Semantic parsing

Authors

Which part is most difficult of your project? - **Having being experienced, comes with lot more responsibilities.**

Which part is most interesting of your project? - **formal logic has deeper influence on Language understanding**

If you could start it over, what would you modify? - **None.**

What is the most valuable thing you learned from conducting the project? -

Who you are? and what is your role in the project?

Q&A

Thank you