Common-sense integration and discovery for Natural Language Generation

ALKIVIADIS KATSALIS

PHD STUDENT

INTERNATIONAL HELLENIC UNIVERSITY, GREECE

UNIVERSITY OF ALICANTE, SPAIN

International Hellenic University, Greece
University of Alicante, Spain (STSM Host)

cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

GPLSI
GRUPO DE PROCESAMIENTO DEL LENGUAJE NATURAL
Y SISTEMAS DE INFORMACIÓN

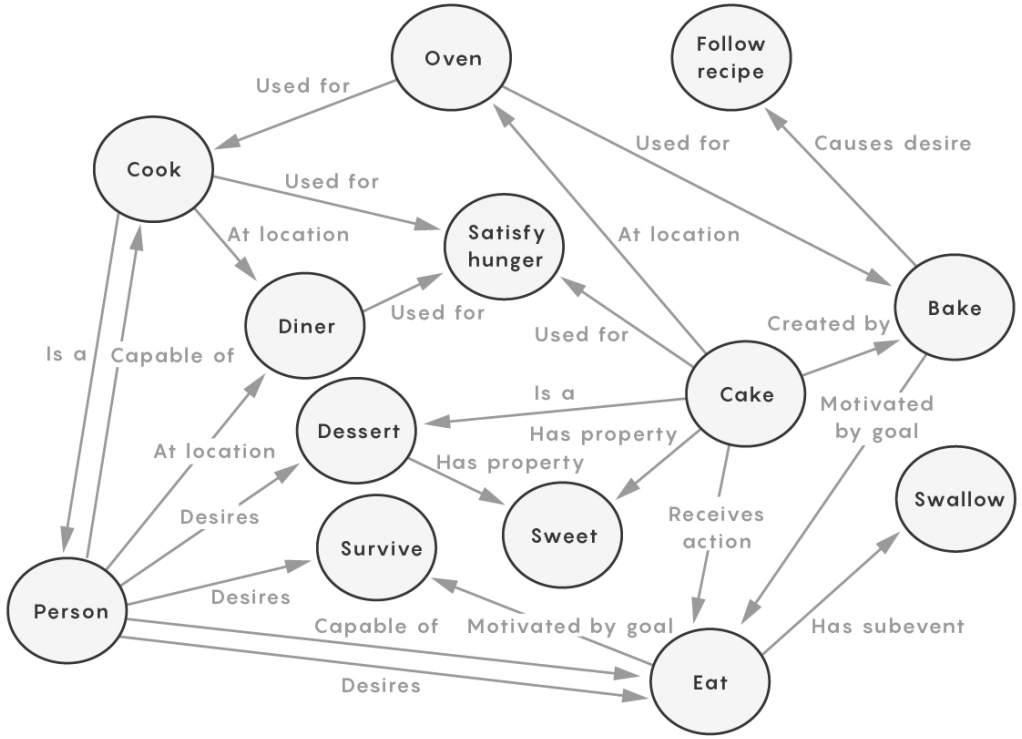
INTERNATIONAL
HELLENIC
UNIVERSITY

NLG M3
MULTI-TASK
MULTI-LINGUAL
MULTI-MODAL

Universitat d'Alacant
Universidad de Alicante

What is CommonSense?

- **CommonSense** can be represented as a Knowledge Graph. CommonSense is all the background knowledge we have about the physical and social world that we have absorbed during our lives. It includes such things as our understanding of physics, (causality, hot and cold), as well as **our expectations about how humans behave**.



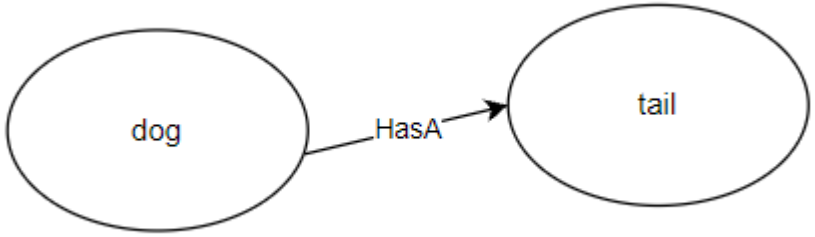
Popular CommonSense Resources

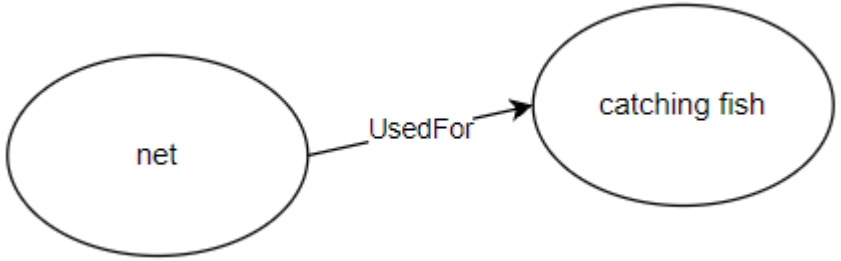
- **ConceptNet**[1] Knowledge Base (Semantic Knowledge)
- **Atomic**[2] Knowledge Base (Inferential Knowledge)

[1] Speer, Robyn, Joshua Chin, and Catherine Havasi. "Conceptnet 5.5: An open multilingual graph of general knowledge." *Thirty-first AAAI conference on artificial intelligence*. 2017.
[2] Sap, Maarten, et al. "Atomic: An atlas of machine commonsense for if-then reasoning." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 33. No. 01. 2019.

ConceptNet

- ConceptNet is a multilingual (83 languages) commonsense knowledge graph that connects words and phrases of natural language (*terms*) with labeled, weighted edges (*assertions*).
- Represents assertions as triples of their start node, relation label, and end node.
- Examples:
  - "a dog has a tail" can be represented as (*dog*, *HasA*, *tail*).
  - "A net is used for catching fish" can be represented as (net, *UsedFor, catching fish)*





ConceptNet

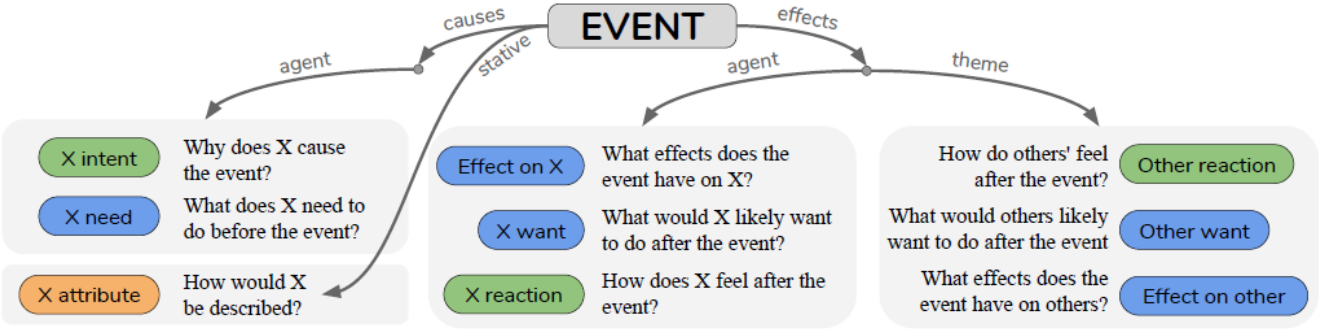
36 Relations

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

Dataset publicly available at: https://github.com/commonsense/conceptnet5/wiki/Downloads

Atomic

- Atlas of everyday commonsense reasoning
- ATOMIC focuses on inferential knowledge organized as typed if-then relations with variables (e.g., "if X pays Y a compliment, then Y will likely return the compliment")
- Nine if-then relation types:



Types of relation

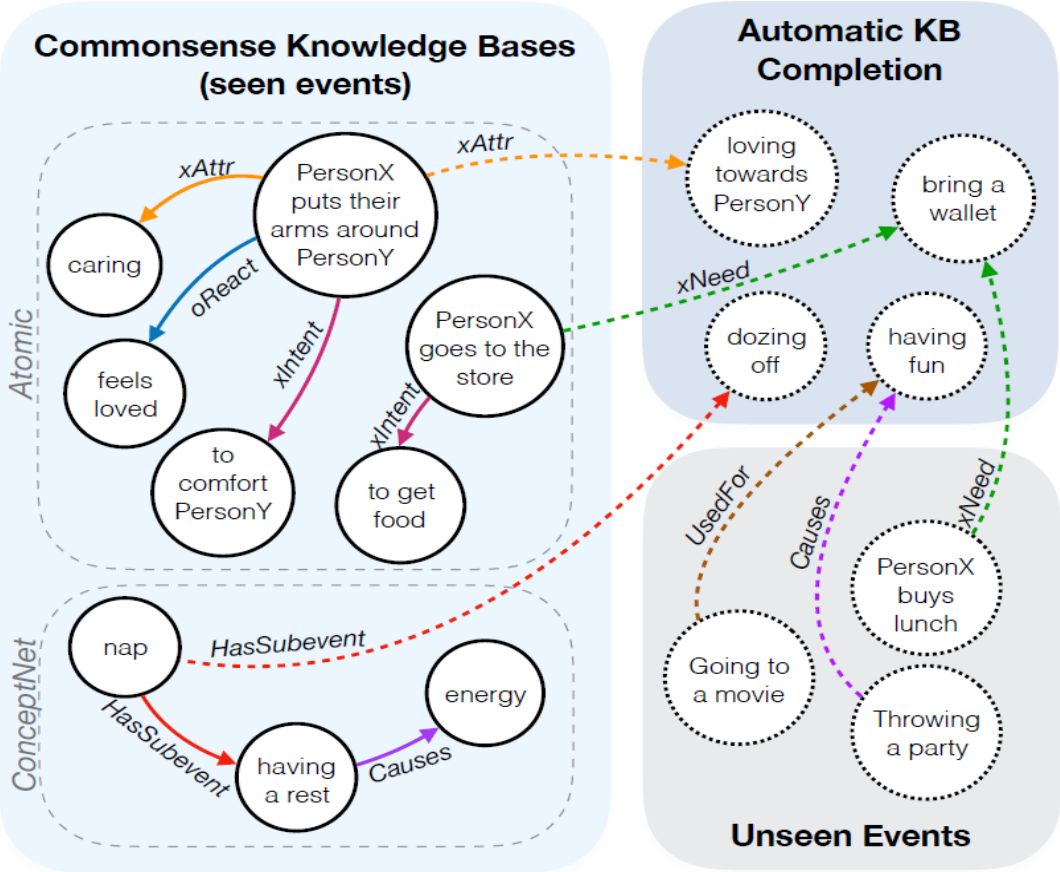
- If-Event-Then-Persona
- If-Event-Then-Event
- If-Event-Then-MentalState

Comet (CommonSense Transformers for Automatic Knowledge Graph Construction)

- Aim: Generation of new commonsense knowledge with Transformer model.
- **COMET**[3] is trained on ConceptNet and Atomic KGs and is able to **generate novel knowledge** that humans rate as high quality, with up to 77.5% (ATOMIC) and 91.7% (ConceptNet) precision at top 1, which approaches human performance for these resources.
- Task: given a training knowledge base of natural language tuples in {s, r, o} format, where **s is the phrase subject** of the tuple, **r** is the **relation** of the tuple, and **o is the phrase object** of the tuple. The **task** is to **generate o given s and r as inputs.**

[3] Bosselut, Antoine, et al. "Comet: Commonsense transformers for automatic knowledge graph construction." arXiv preprint arXiv:1906.05317 (2019).

Seen and Unseen examples of COMET



Demo available at: https://mosaickg.apps.allenai.org

Augmenting BART with CommonSense Knowledge

- Knowledge graph augmented pre-trained language generation model KG-BART[4], **encompasses the complex relations of concepts through the knowledge graph** and **produces** more logical and **natural sentences** as output.
- KG-BART can leverage the graph attention[5] to aggregate the rich concept semantics that **enhances the model generalization on unseen concept sets.**
- Trained on CommonGen[6] Dataset

[4] Liu, Ye, et al. "KG-BART: Knowledge Graph-Augmented BART for Generative Commonsense Reasoning." arXiv preprint arXiv:2009.12677 (2020).
[5] Veličković, Petar, et al. "Graph attention networks." arXiv preprint arXiv:1710.10903 (2017).
[6] Lin, Bill Yuchen, et al. "Commongen: A constrained text generation dataset towards generative commonsense reasoning." (2019).

CommonGen Dataset

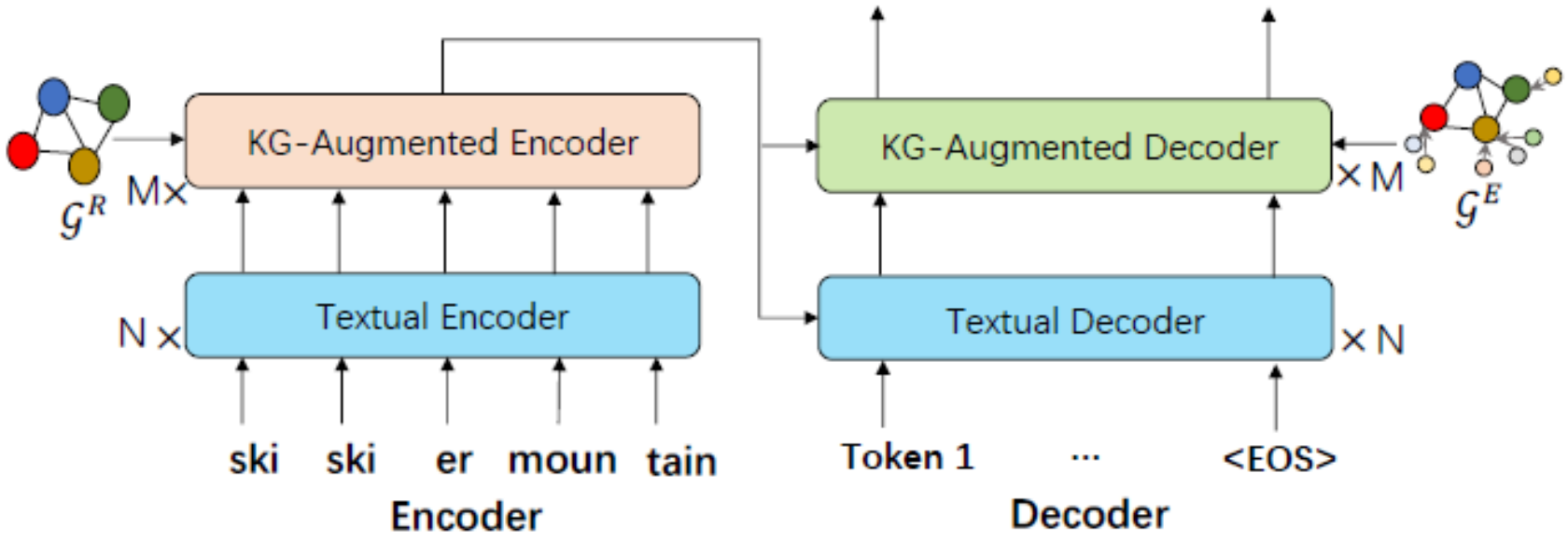
Train set examples:
['city' 'roof' 'window'] -> a **city** seen through a **window** over **roofs**.
living room with bay **windows** overlooking the **roofs** of the **city**

['passenger' 'let' 'airplane' 'airport' 'park'] -> An **airplane parked** in an **airport letting** off its **passengers**.

Test set example:
['look' 'map' 'find']-> **look** what I **found** on **maps** .
Both are **looking** at the **map** to **find** out the direction.
tourists stop to **look** at a **map** while trying to **find** their way

KG-BART MODEL

- It has two major steps: **1)knowledge graph grounding** and **2)graph-based encoder-decoder modeling.**

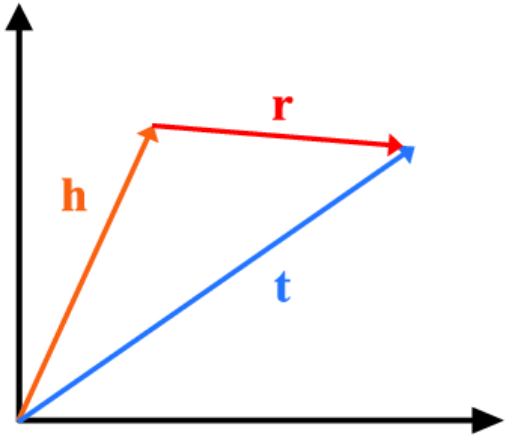


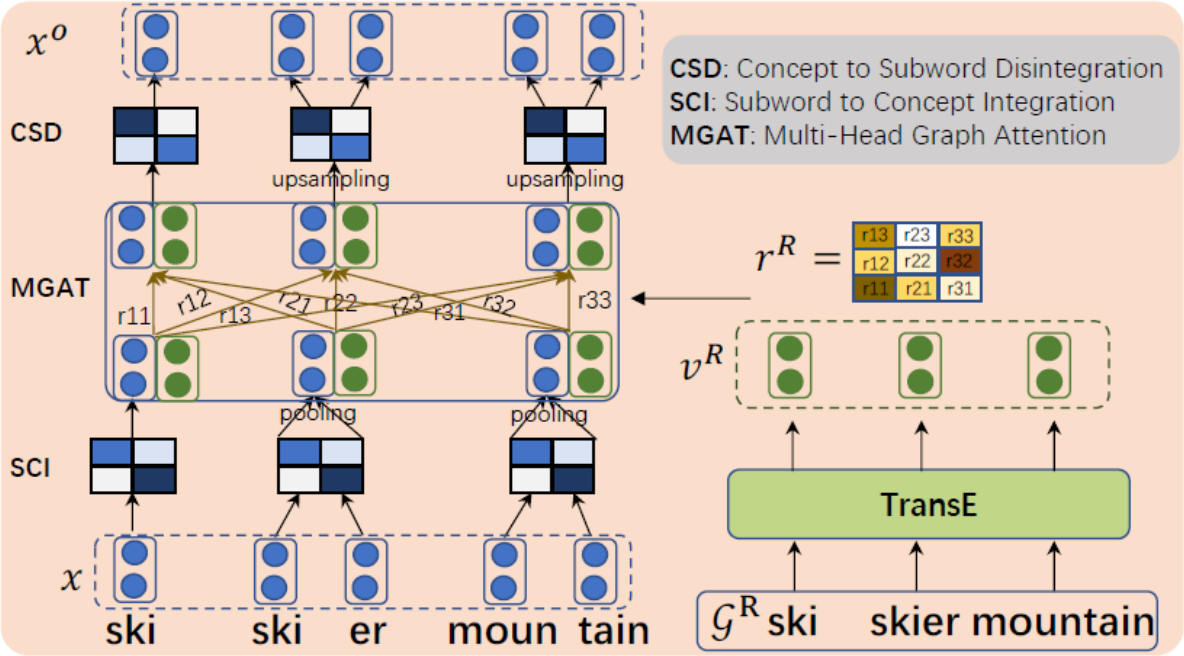
KG Grounding

- **KG Grounding**: Construct and learn the embeddings representation of **Concept-reasoning Graph** (G^R) and the **hierarchical concept-expanding graph** (G^E) from the Conceptnet KG.
- Each concept corresponds to a KG's unigram entity and with this way G^R is generated. G^E is generated by coupling G^R and selection of neighborhood.
- Example: Concept-set: {ski, skier, mountain}. Adjunct concept for "mountain" will be influenced by ski, skier -> Glove embeddings with cosine similarity in order to get top-k neighborhood.
- When there are no direct connection between concept pairs then TransE is used.

Graph-Based Encoder Decoder Modeling

- **KG-Augmented Encoder**: Integrates the input token embeddings (textual encoders) and the embeddings (obtained using TransE[7] of concept-reasoning graph G^R.

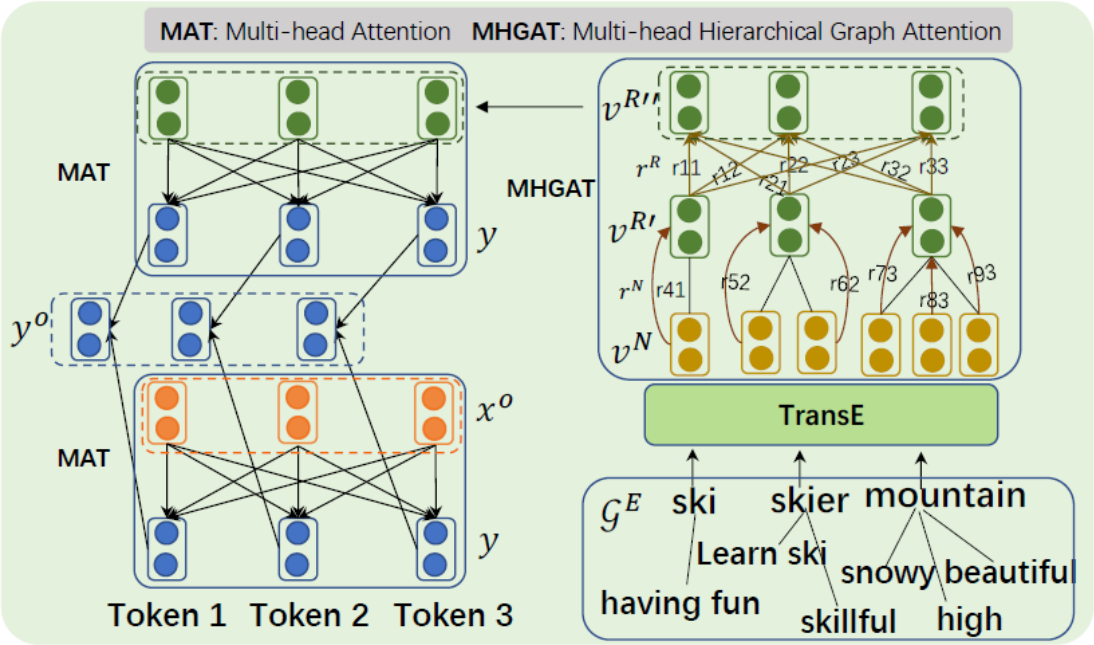




[7]Bordes, Antoine, et al. "Translating embeddings for modeling multi-relational data." Advances in neural information processing systems 26 (2013).

Graph-Based Encoder Decoder Modeling

- **KG-Augmented Decoder**: Incorporates hierarchical graph structure into the decoding process to capture the relations between concepts and thein neighborhood nodes -> precise and natural output.



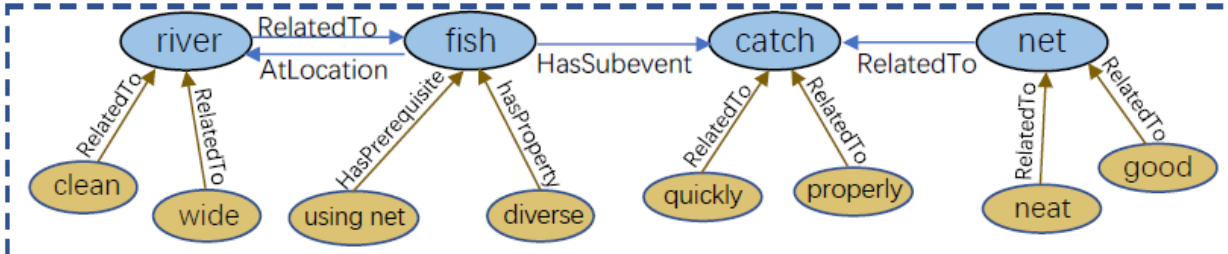
KG-BART

Concept Set: {river, fish, net, catch}

[Expected Output]: everyday scenarios covering all given concepts.

1. Fisherman uses a strong net to catch plentiful fishes in the river.
2. Men like to catch fishes in the wide river with a net in the afternoon.

[GPT-2]: A fish is catching in a net
[UniLM]: A net catches fish in a river
[T5]: Fish are caught in a net in the river.
[BART]: A man catches a fish with a net in the river



[KG-BART]: A fisherman catches fishes by using good net in the clean river.

Concept Set: {stand, hold, street, umbrella }

[GPT-2]: A woman holding a umbrella in street
[BERT-Gen]: The woman stands on the street holding an umbrella.
[UniLM]: A man stands next to an umbrella on a street.
[T5]: A man holding an umbrella stands on a street.
[BART]: The woman holding an umbrella stands on the street and holds an umbrella.

1. A man held an umbrella while standing on the street.
2. People standing in the crowd street, many holding umbrellas.

[KG-BART]: A man holds an umbrella as he stands on the empty street.

Human Evaluation

Model	1	2	3	4	5	Rating
GPT-2	22%	16%	23%	20%	19%	2.98
UniLM	5%	17%	22%	24%	32%	3.61
T5-large	2%	15%	12%	32%	39%	3.91
BART	1%	10%	17%	30%	42%	4.02
KG-BART	0 %	8%	12%	25%	55%	4.27

Pre-trained LMs are great

- Captures world knowledge in parameters
- Strong results on loads of tasks
- Applicable for almost every task

BUT:

- Hallunicate (Generate non factual text)
- Struggle to access and apply knowledge
- Difficult to update

Retrieval is great

External Knowledge/text is useful for huge variety of NLP tasks:

- Precise and accurate knowledge
- Trivial to update at test time
- Dense retrieval starting to outperform traditional IR

BUT

- Need retrieval supervision
- Need some (usually task specific) way to integrate into downstream tasks

Retrieval-Augmented Generation (RAG) for Knowledge-Intensive NLP Tasks[7]

- **RAG looks and acts like a standard seq2seq model**, meaning it takes in one sequence and outputs a corresponding sequence. There is an **intermediary step** though, which differentiates and elevates RAG above the usual seq2seq methods. Rather than passing the input directly to the generator, **RAG instead uses the input to retrieve a set of relevant documents**, in our case from **Wikipedia**.
- Given the prompt **"When did the first mammal appear on Earth?"** for instance, RAG might **surface documents for "Mammal," "History of Earth," and "Evolution of Mammals."** These supporting **documents are then concatenated as context with the original input** and fed to the seq2seq model that produces the actual output. RAG thus has **two sources of knowledge:** the **knowledge** that **seq2seq** models store in their parameters (**parametric memory**) and the **knowledge** stored in the **corpus** from which RAG retrieves passages (**nonparametric memory**).
- Demo: https://ai.facebook.com/blog/retrieval-augmented-generation-streamlining-the-creation-of-intelligent-natural-language-processing-models

[7] Lewis, Patrick, et al. "Retrieval-augmented generation for knowledge-intensive nlp tasks." *arXiv preprint arXiv:2005.11401* (2020).

Off-the-She-Shelf Models Need 3 things

- A **Pre-trained Generator** model P(y/…) (**BART**, T5,…)
- A **Pre-trained retriever** model P(z/x) (Dense Passage Retrieval (**DPR**)[8])
- An **indexed (Non-Parametric) KB** of text documents (**Wikipedia**, tweets,…)

RAG models **combine parametric and non-parametric memory** and work well for knowledge intensive tasks (tasks that even humans need external knowledge).

[8] Karpukhin, Vladimir, et al. "Dense passage retrieval for open-domain question answering." *arXiv preprint arXiv:2004.04906* (2020).

RAG Architecture

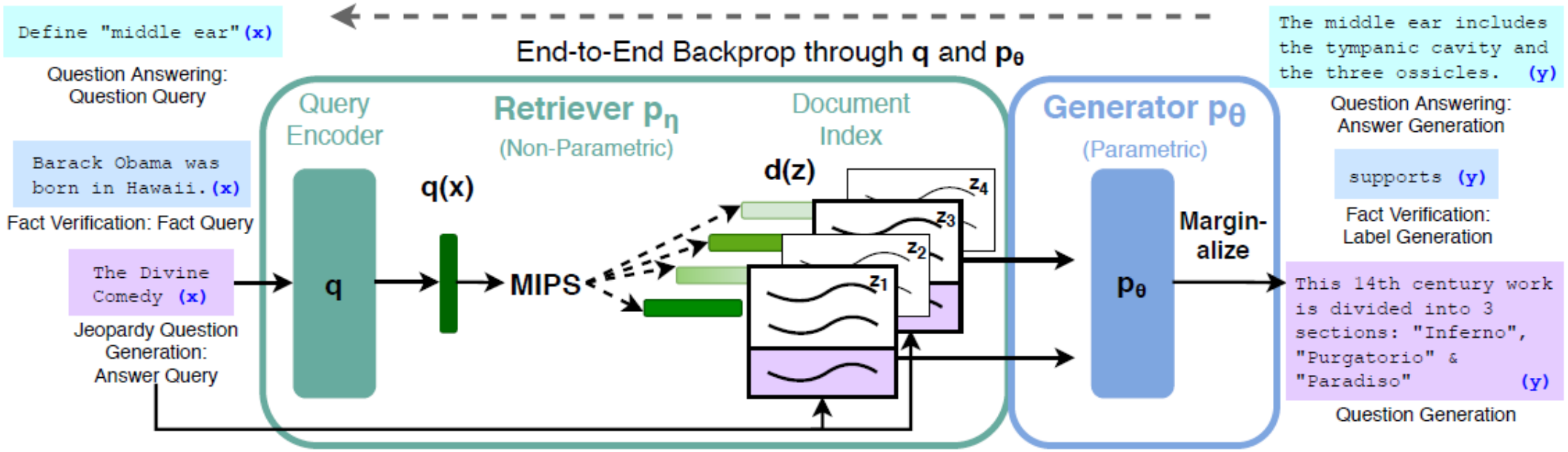


Figure 1: Overview of our approach. We combine a pre-trained retriever (*Query Encoder* + *Document Index*) with a pre-trained seq2seq model (*Generator*) and fine-tune end-to-end. For query x, we use Maximum Inner Product Search (MIPS) to find the top-K documents z_i. For final prediction y, we treat z as a latent variable and marginalize over seq2seq predictions given different documents.

Results

Outperforms all of the SOT in Open Domain QA and close enough to the SOT for the Fact Verification Problem(FEVER).

	Model	NQ	TQA	WQ	CT
Closed Book	T5-11B [52]	34.5	- /50.1	37.4	-
	T5-11B+SSM[52]	36.6	- /60.5	44.7	-
Open Book	REALM [20]	40.4	- / -	40.7	46.8
	DPR [26]	41.5	**57.9**/ -	41.1	50.6
	RAG-Token	44.1	55.2/66.1	**45.5**	50.0
	RAG-Seq.	**44.5**	56.8/**68.0**	45.2	**52.2**

Model	Jeopardy		MSMARCO		FVR3	FVR2
	B-1	QB-1	R-L	B-1	Label Acc.	
SotA	-	-	**49.8***	**49.9***	**76.8**	**92.2***
BART	15.1	19.7	38.2	41.6	64.0	81.1
RAG-Tok.	**17.3**	**22.2**	40.1	41.5	72.5	89.5
RAG-Seq.	14.7	21.4	40.8	44.2		

Plans accomplished from the STSM: Commonsense Knowledge Integration for NLG tasks during Training

- Use commonsense knowledge in natural language text (output of KGBART) in DPR part in RAG model.
- Make use of COMET and KG-BART to generate new knowledge.
- Tweak KG-BART by adding more knowledge(use Comet Transformer) information instead of just using the ConceptNet KG.
- Fine-Tune BERT/BART model in relation prediction task and with those updated embeddings use the BERT/BART to NLG tasks.
- After experiments the target is to make a multilingual framework.