Host : Srivatsan Srinivasan
(bit.ly/AIEngineering)

Question Generation in edtech using Natural Language Processing

Sat, March 6th, 2021
8 PM IST/9.30 AM EST

Colab Code and slides:
bit.ly/talkbyramsri

Talk by
Ramsri Goutham
ramsri.ai

Who am I ?

- Lead Data Scientist with 8+ years of work experience across Silicon Valley, Singapore, and India.

- Most recently been a Co-founder and CTO of an AI-assisted assessments startup.

- Spent the last 2 years on question generation using Natural Language processing in edtech.

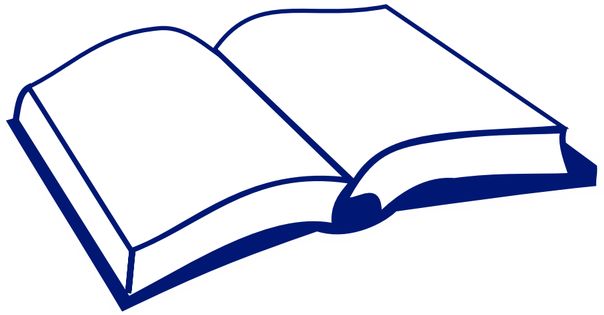
Ramsri Goutham Golla

ramsri.ai

The problem

Assessment creation (Quizzes and Tests) is manual and not scalable for online transformation due to Covid.

A textbook publisher hired you as a data scientist to create a tool that automates the assessment creation process .



The solution

Chapter 6 Social Studies

The Amazon River

The Amazon in South America is the largest river by discharge volume of water in the world, and the second-longest river in the world.

The total length of the river as measured from the headwaters of the Ucayali-Apurímac river system in southern Peru is at least 4,000 miles (6,400 km), which makes it slightly shorter than the Nile River but still the equivalent of the distance from New York City to Rome. Its westernmost source is high in the Andes Mountains, within 100 miles (160 km) of the Pacific Ocean, and its mouth is in the Atlantic Ocean, on the northeastern coast of Brazil. However, both the length of the Amazon and its ultimate source have been subjects of debate since the mid-20th century, and some claim that the Amazon is longer than the Nile.

1. Multiple Choice Questions (MCQs)
2. True or False Questions
3. Fill in the Blanks
4. Match the following

Open Source library for question generation

Questgen

https://github.com/ramsrigouthamg/Questgen.ai

Let's get started!

Talk Outline

We will use Google Colab notebooks as the development environment for our demos

1. Generate **MCQs** from any content using Wordnet & T5 transformer
2. Generate **Fill-in-the-blank** Questions using keyword extraction.
3. Generate **Match-the-following** Questions using BERT WSD.
4. Generate **True/False** Questions using constituency parsing & GPT-2.

1. Generate MCQs high level pipeline

Chapter 6 Social Studies

The Amazon River

The Amazon in South America is the largest river by discharge volume of water in the world, and the second-longest river in the world.

The total length of the river as measured from the headwaters of the Ucayali-Apurímac river system in southern Peru is at least 4,000 miles (6,400 km), which makes it slightly shorter than the Nile River but still the equivalent of the distance from New York City to Rome. Its westernmost source is high in the Andes Mountains, within 100 miles (160 km) of the Pacific Ocean, and its mouth is in the Atlantic Ocean, on the northeastern coast of Brazil. However, both the length of the Amazon and its ultimate source have been subjects of debate since the mid-20th century, and some claim that the Amazon is longer than the Nile.

Find important Sentences (Abstractive or Extractive Summarization)

The Amazon in South America is the largest river by discharge volume of water in the world, and the second-longest river in the world.

Sentence2MCQ

Which is the world's second longest river ?

- a Mississippi
- b Amazon
- c Nile
- d Yangtze

1.1 Generate wrong choices for MCQ options

Chapter 6 Social Studies

The Amazon River

TThe Amazon in South America is the largest river by discharge volume of water in the world, and the second-longest river in the world.

The total length of the river as measured from the headwaters of the Ucayali-Apurímac river system in southern Peru is at least 4,000 miles (6,400 km), which makes it slightly shorter than the Nile River but still the equivalent of the distance from New York City to Rome. Its westernmost source is high in the Andes Mountains, within 100 miles (160 km) of the Pacific Ocean, and its mouth is in the Atlantic Ocean, on the northeastern coast of Brazil. However, both the length of the Amazon and its ultimate source have been subjects of debate since the mid-20th century, and some claim that the Amazon is longer than the Nile.

Multiple Choice Question

Which is the world's longest river ?

a Mississippi

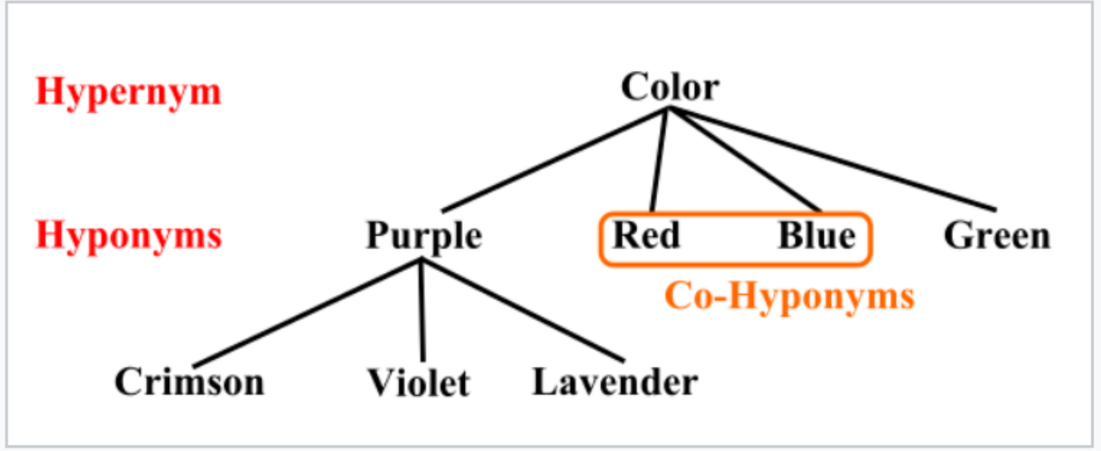
b Amazon

c Nile

d Yangtze

Given the **Amazon** as input, we will see how we can generate **distractors** (wrong choices) like **Mississippi, Nile, Yangtze,** using **Wordnet**.

1.2 Generate distractors using Wordnet



An example of the relationship between hyponyms and hypernym

Source: https://en.wikipedia.org/wiki/Hyponymy_and_hypernymy

https://wordnet.princeton.edu/

- WordNet® captures relations.
- A hyponym is in a type-of relationship with its hypernym.
- A hypernym is an umbrella term and blanket term. Eg: Color is the hypernym for Purple.
- Red, Blue, Green are the hyponyms of "Color". Since Red is a type of color.
- Our goal is to extract Co-Hyponyms as distractors.

1.3 Word Sense Disambiguation (WSD)

Amazon (Wordnet)

- A south American River.
- Green tropical America Parrot
- (Greek mythology) A nation of women warriors of Scythia
- An e-commerce company

How to disambiguate in our story? A word's meaning is defined by the group it keeps.

Collect sentences that contain **"amazon"** from the story.

Eg : ["Once upon a time, there lived a lion in the dense **amazon** rainforest."]

1.4 BERT Word Sense Disambiguation (WSD)

https://github.com/BPYap/BERT-WSD

1.5 Generate a question using T5 Transformer model

Context: The **Amazon** in South America is the largest river by discharge volume of water in the world, and the second-longest river in the world.

Answer: Amazon

Text to text transfer transformer

Question Which is the world's second longest river ?

1.6 Squad dataset used for training

https://rajpurkar.github.io/SQuAD-explorer/

Nikola_Tesla

The Stanford Question Answering Dataset

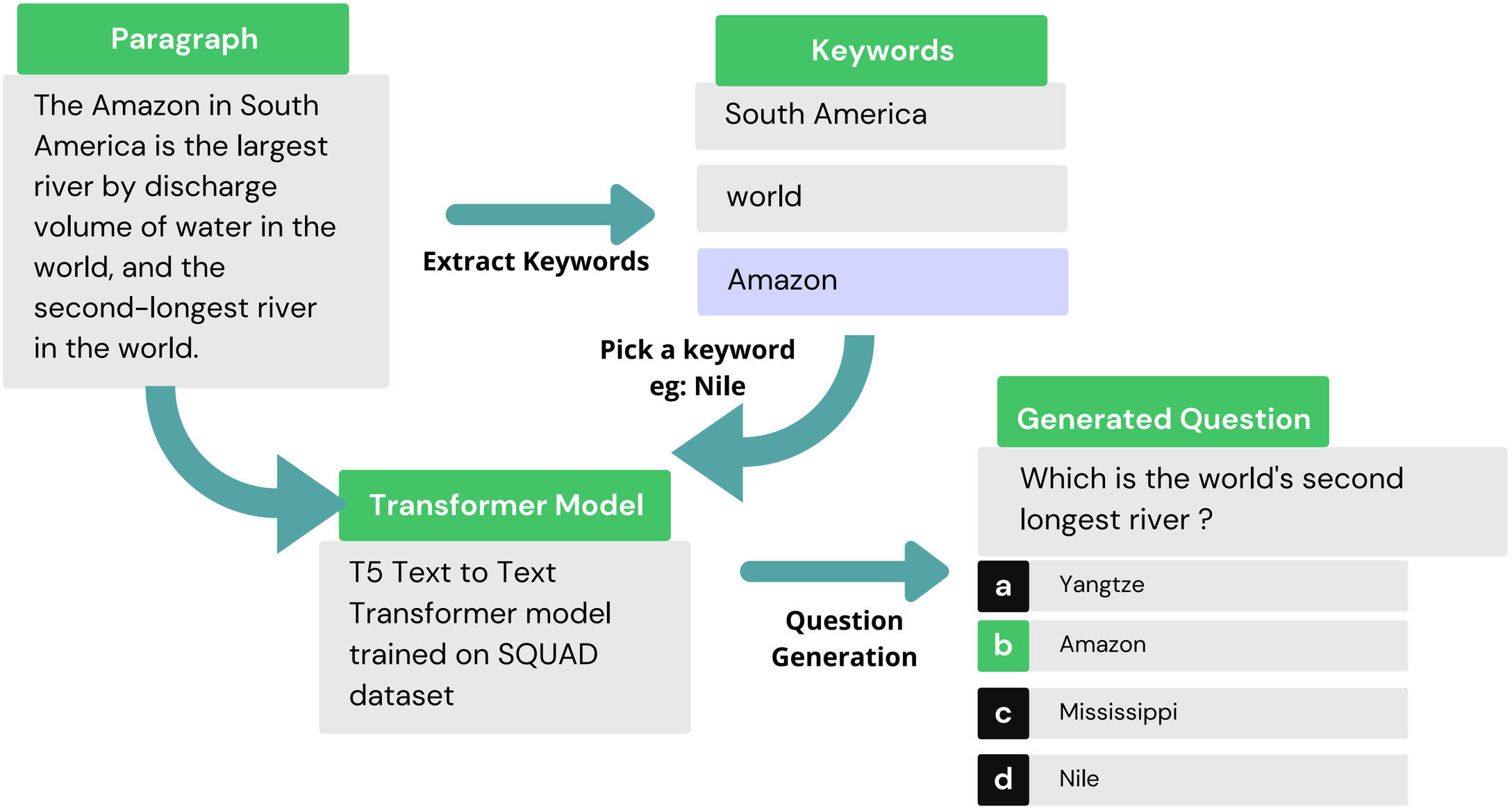
Nikola Tesla (Serbian Cyrillic: Никола Тесла; 10 July 1856 – 7 January 1943) was a Serbian American inventor, electrical engineer, mechanical engineer, physicist, and futurist best known for his contributions to the design of the modern alternating current (AC) electricity supply system.

In what year was Nikola Tesla born?
Ground Truth Answers: 1856 1856 1856
Prediction: 1856

What was Nikola Tesla's ethnicity?
Ground Truth Answers: Serbian Serbian Serbian
Prediction: Serbian

In what year did Tesla die?
Ground Truth Answers: 1943 1943 1943
Prediction: 1943

1.7 Generate MCQs from any content.



Sentence2MCQ
Google Colab Coding Session

Now, what you've just seen (MCQ generator) itself is a million-dollar edtech startup idea.

True story: We almost raised a million $ from VCs with the same tech.
Why we didn't continue is a story for later.

2. Fill in the blanks questions

Chapter 6 Social Studies

The Nile River

The Greek historian knew what he was talking about. The Nile River fed Egyptian civilization for hundreds of years.

The Longest River, The Nile is 4,160 miles long - the world's longest river. It begins near the equator in Africa and flows north to the Mediterranean Sea.

In the south it churns with cataracts. A cataract is a waterfall. Near the sea the Nile branches into a delta. A delta is an area near a river's mouth where the water deposits fine soil called silt. In the delta, the Nile divides into many streams.

The river is called the upper Nile in the south and the lower Nile in the north. For centuries, heavy rains in Ethiopia caused the Nile to flood every summer. The floods deposited rich soil along the Nile's shores. This soil was fertile, which means it was good for growing crops. Unlike the Tigris and Euphrates, the Nile River flooded at the same time every year, so farmers could predict when to plant their crops.

Fill in the blanks for these sentences with matching words at the top

Mediterranean Sea, rich soil, delta, Nile River, Ethiopia

1. The _____ fed Egyptian civilization for hundreds of years
2. It begins near the equator and flows north to the _____.
3. For centuries, heavy rains in the _____ caused the Nile to flood every summer.
4. The floods deposited _____ along the Nile's shores.
5. A _____ is an area near a river's mouth where he water deposits fine soil called silt.

2.1 Extract Keywords

Chapter 6 Social Studies

The Nile River

The Greek historian knew what he was talking about. The Nile River fed Egyptian civilization for hundreds of years.

The Longest River, The Nile is 4,160 miles long - the world's longest river. It begins near the equator in Africa and flows north to the Mediterranean Sea.

In the south it churns with cataracts. A cataract is a waterfall. Near the sea the Nile branches into a delta. A delta is an area near a river's mouth where the water deposits fine soil called silt. In the delta, the Nile divides into many streams.

The river is called the upper Nile in the south and the lower Nile in the north. For centuries, heavy rains in Ethiopia caused the Nile to flood every summer. The floods deposited rich soil along the Nile's shores. This soil was fertile, which means it was good for growing crops. Unlike the Tigris and Euphrates, the Nile River flooded at the same time every year, so farmers could predict when to plant their crops.

Extract Keyphrases using
Python Keyword Extraction Library

https://github.com/boudinfl/pke

Mediterranean Sea, rich soil, delta, Nile River, Ethiopia

MultipartiteRank

3. Match the following

Lion and the mouse

Moral Story

Once upon a time, there lived a lion in the dense Amazon rainforest. While he was sleeping by resting his big head on his paws, a tiny little mouse unexpectedly crossed by and ran across the lion's nose in haste. This woke up the lion and he laid his huge paw angrily on the tiny mouse to kill her.

The poor mouse begged the lion to spare her this time and she would pay him back on some other day. Hearing this, the lion was amused and wondered how could such a tiny creature ever help him. But he was in a good mood and in his generosity he finally let the mouse go.

A few days later, a hunter set a trap for the lion while the big animal was stalking for prey in the forest. Caught in the toils of a hunter's net, the lion found it difficult to free himself and roared loudly in anger.

As the mouse was passing by, she heard the roar and found the lion struggling hard to free himself from the hunter's net.

Match the following words with their corresponding meanings.

Mouse	large gregarious predatory feline of Africa and India ...
Amazon	a trap made of netting to catch fish or birds or insects ...
Net	any of numerous small rodents typically resembling rats ...
Lion	a major South American river ...

3.1 Extract keywords

Lion and the mouse

Moral Story

Once upon a time, there lived a lion in the dense Amazon rainforest. While he was sleeping by resting his big head on his paws, a tiny little mouse unexpectedly crossed by and ran across the lion's nose in haste. This woke up the lion and he laid his huge paw angrily on the tiny mouse to kill her.

The poor mouse begged the lion to spare her this time and she would pay him back on some other day. Hearing this, the lion was amused and wondered how could such a tiny creature ever help him. But he was in a good mood and in his generosity he finally let the mouse go.

A few days later, a hunter set a trap for the lion while the big animal was stalking for prey in the forest. Caught in the toils of a hunter's net, the lion found it difficult to free himself and roared loudly in anger.

As the mouse was passing by, she heard the roar and found the lion struggling hard to free himself from the hunter's net.

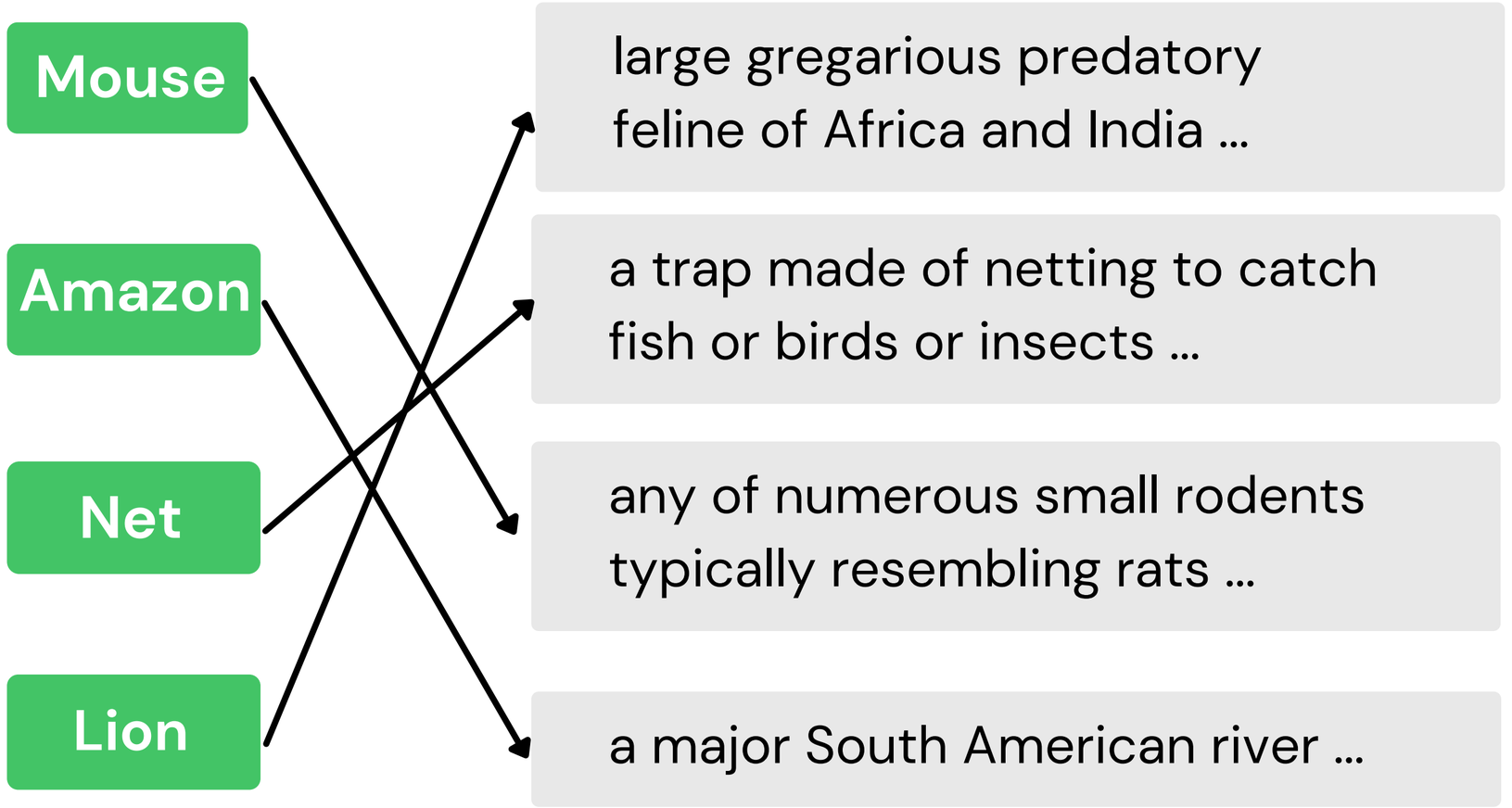
Extract Keyphrases using
Python Keyword Extraction Library

https://github.com/boudinfl/pke

hunter, big, net, mouse, amazon, lion, tiny, free

3.2 Match the following

Match the following words with their corresponding meanings.



4. Generate True or False Questions

Chapter 6 Social Studies

The Nile River

The Greek historian knew what he was talking about. The Nile River fed Egyptian civilization for hundreds of years.

The Longest River, The Nile is 4,160 miles long - the world's longest river. It begins near the equator in Africa and flows north to the Mediterranean Sea.

In the south it churns with cataracts. A cataract is a waterfall. Near the sea the Nile branches into a delta. A delta is an area near a river's mouth where the water deposits fine soil called silt. In the delta, the Nile divides into many streams.

The river is called the upper Nile in the south and the lower Nile in the north. For centuries, heavy rains in Ethiopia caused the Nile to flood every summer. The floods deposited rich soil along the Nile's shores. This soil was fertile, which means it was good for growing crops. Unlike the Tigris and Euphrates, the Nile River flooded at the same time every year, so farmers could predict when to plant their crops.

True or False Question

The Nile River begins near the equator in Africa and flows north to the Atlantic Ocean.

a. True

b. False

4.1 Ways to give a True sentence –

Split compound or complex sentence to a simple sentence

Original Sentence	True Sentence
He wiped off the water and **turned the boat upside down.**	He turned the boat upside down.
Subject pronouns which acn be found only in the beginning of sentencesand have no bearing on the word order, **are used to replace nouns.**	Subject pronouns are used to replace nouns.

4.2 Ways to Falsify a sentence -

Add or remove negation of a verb phrase or noun phrase

Original Sentence	Falsified Sentence
The amount of matter in a chemical reaction **doesn't change.**	The amount of matter in a chemical reaction **changes**. [Negate Verb phrase]

Change a Named Entity

Augustus was the first Roman Emperor.	**Nero** was the first Roman Emperor.

4.3 Ways to Falsify a sentence -

Change adjective

Original Sentence	Falsified Sentence
The **scariest** villain of all time is Darth Vader.	The **softest** villain of all time is Darth Vader.

Change main verb

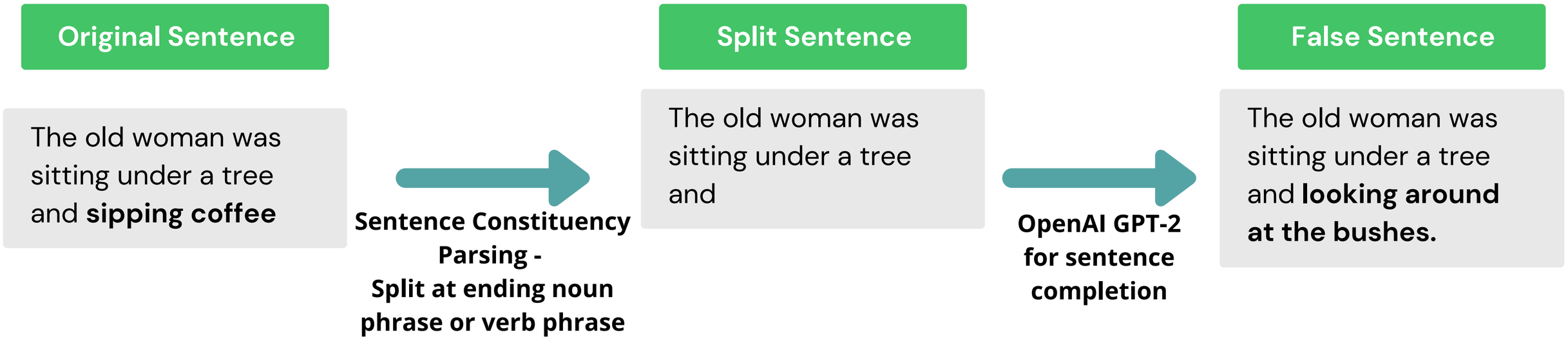
When electrons are **shared** between two atoms, a covalent bond is formed.	When electrons are **transferred** between two atoms, a covalent bond is formed.

4.4 Ways to Falsify a sentence -
Alternate verb phrase or noun phrase

Original Sentence	Falsified Sentence
Aunt Avanti was sitting **in an armchair watching the rain-washed trees and blue sky.**	Aunt Avanti was sitting **on the sofa watching Television.**

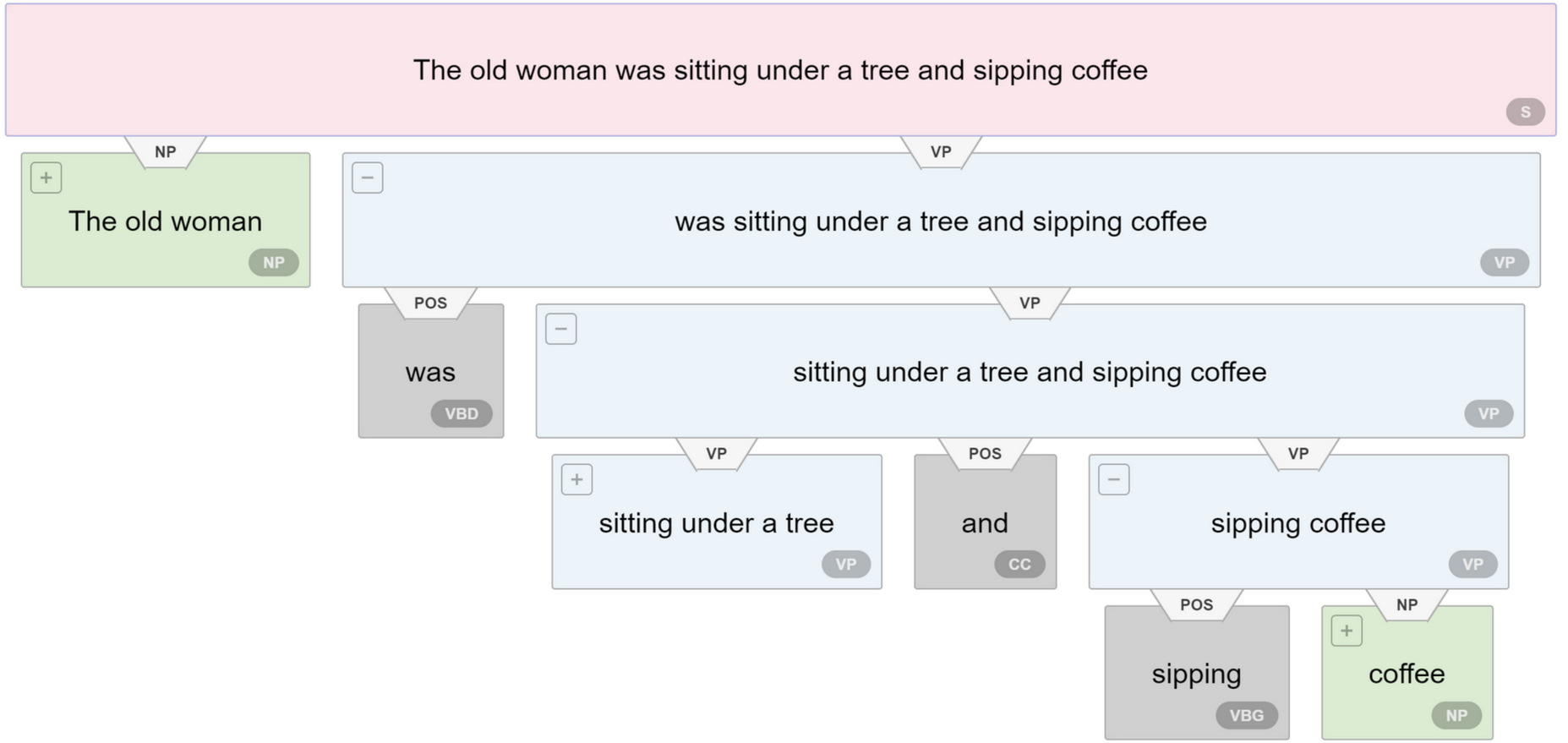
4.5 Generate Falsified sentence with alternate verb phrase or noun phrase.

Algorithm



4.6 Constituency parsing of a sentence

Constituency parsing is the task of breaking a text into sub-phrases or constituents.
Non-terminals in the parse tree are types of phrases, the terminals are the words in the sentence.

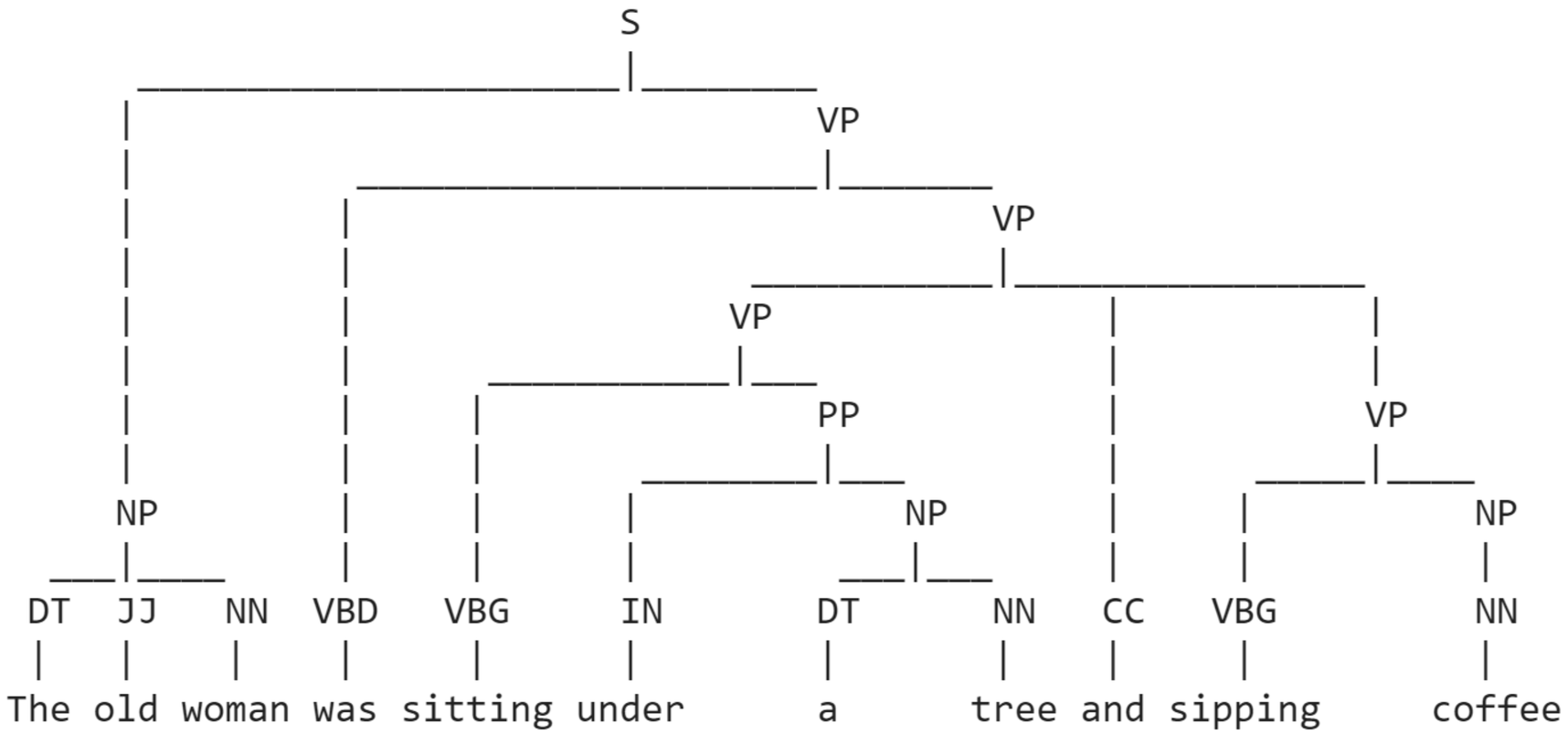


https://demo.allennlp.org/constituency-parsing/

https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html

4.7 Constituency parsing of a sentence

Constituency parsing is the task of breaking a text into sub-phrases or constituents.
Non-terminals in the parse tree are types of phrases, the terminals are the words in the sentence.



4.8 OpenAI GPT-2 Text Generation

- GPT-2 is a transformer-based language model trained to predict the next word.
- The most famous use cases of transformer models are language translation, text generation, summarization, etc

The old woman was sitting under a tree and

OpenAI GPT-2 for sentence completion

The old woman was sitting under a tree and **looking around at the bushes.**

The old woman was sitting under a tree and **reading a book.**

The old woman was sitting under a tree and **sipping tea.**

http://jalammar.github.io/illustrated-gpt2/

4.9 Sorting the generated sentences

Split Sentence: The old woman was sitting under a tree and

Original Sentence: The old woman was sitting under a tree and **sipping coffee.**

OpenAI GPT-2 for sentence completion

3. The old woman was sitting under a tree and **looking around at the bushes.**

2.The old woman was sitting under a tree and **reading a book.**

1.The old woman was sitting under a tree and **sipping tea.**

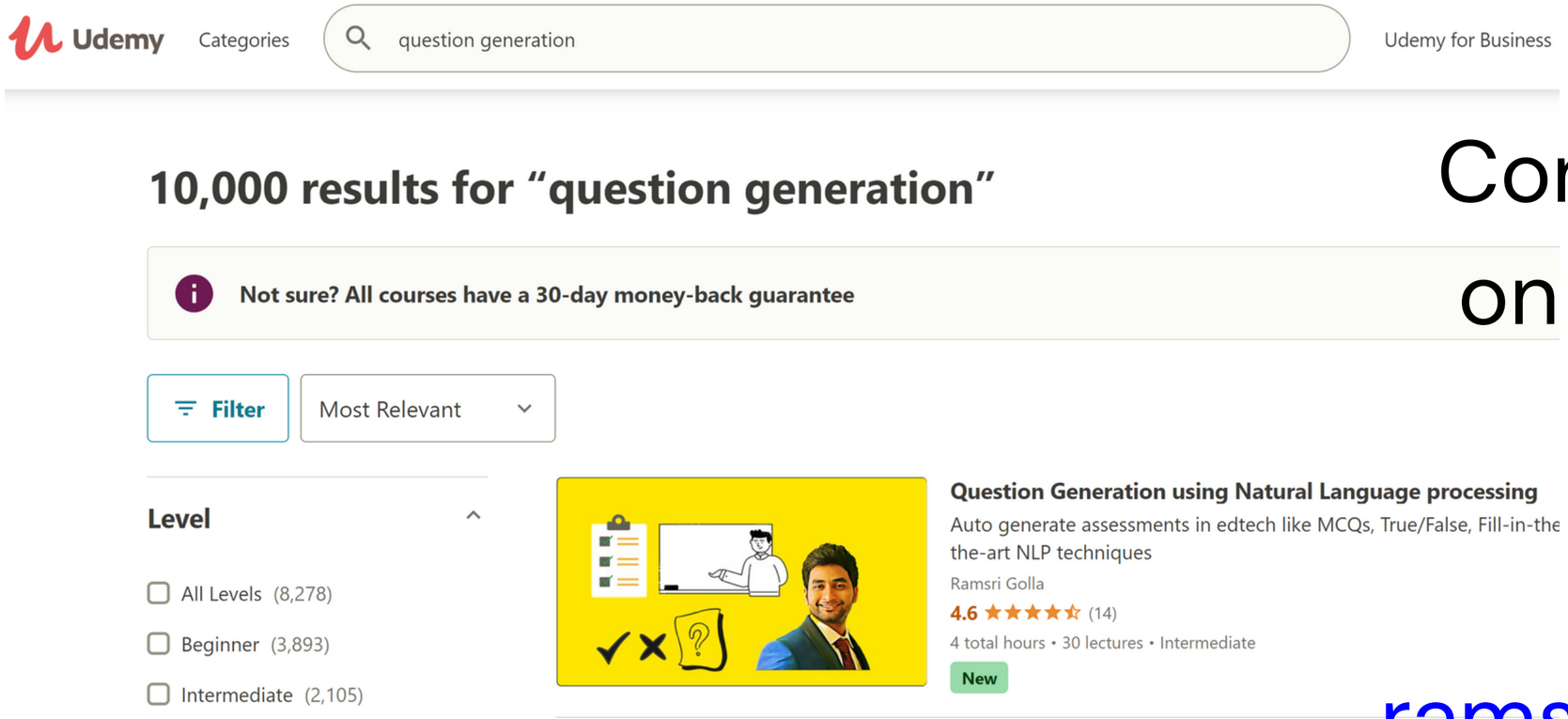
- Use sentence embeddings and cosine similarity to sort the sentences by order of similarity.
- Use sentence transformers library.

https://www.sbert.net/

True or False Generation
Google Colab Coding Session

Udemy Course: Question generation using Natural Language Processing (bit.ly/questiongeneration)

Special Coupon Code: SRIVATSAN



Connect with me on Social Media:

ramsri.ai

Email:

ramsri@questgen.ai