



Eastern University
A Leader in quality Education

Thesis Paper

A Thesis on “**Corpus Analyzer – Machine Learning and Deep Learning Based Virtual Assistant**”

Submitted by:

Mohammad Abdullah Al Salman

Student ID: 183400026

Md. Reaj Uddin

Student ID: 183400043

Jahanara Begum

Student ID: 181400026

Md. Mostofa Jalal

Student ID: 183400042

Md. Rifat Hasan

Student ID: 182400056

Emdad Hossain

Student ID: 181400102

In fulfillment of the requirements for the degree of Bachelor Science in Computer Science and Engineering

CANDIDATE’S DECLARATION

This is hereby declared that the work titled “**Corpus Analyzer – Machine Learning and Deep Learning Based Virtual Assistant**” is the outcome of research carried out by us under the supervision of Prof. Dr. Md. Mahfuzur Rahman, in the Department of Computer Science and Engineering, Eastern University. It is also declared that this thesis or any part of it has not been submitted elsewhere for the award of any degree or diploma.

Mohammad Abdullah Al Salman

Md. Reaj Uddin

Jahanara Begum

Md. Mostofa Jalal

Md. Rifat Hasan

Emdad Hossain

APPROVAL

This project titled "**Corpus Analyzer – Machine Learning and Deep Learning Based Virtual Assistant**" submitted by Mohammad Abdullah Al Salman (ID: 181400026), Md. Reaj Uddin (ID: 183400043), Jahanara Begum (ID: 181400026), Md. Mostofa Jalal (ID: 183400042), Md. Rifat Hasan (ID: 182400056), Emdad Hossain (ID: 1814000102) has been accepted satisfactory in partial fulfillment of the requirement for the degree Bachelor of Science in Computer Science and Engineering.

Board of Examiners

ACKNOWLEDGEMENT

Firstly, all praise to the Great Almighty for whom our thesis have been completed without any major interruption.

This thesis has been supervised by Prof. Dr. Md. Mahfuzur Rahman Dean, Faculty of Engineering and Technology, & Chairperson, Dept. of Computer Science and Engineering, Eastern University, Dhaka, Bangladesh. We are very much grateful and indebted to him for his kind suggestion, guidance, instructions and overall supervision that he offered us during our thesis investigation. We also thank him for reviewing the preliminary versions of this thesis and making useful corrections as well as providing valuable suggestions.

We would also like to express our gratitude to our parents and teachers for bringing us up where we are today. We are thankful to the Department of Computer Science & Engineering, Eastern University, for providing us with an excellent educational environment and computing facility. We also offer our best regards Eastern University administration, and all other esteemed teachers of the department for their affectionate feelings and encouragement throughout the period of our research work.

Finally, warm thanks to all of our well-wishers and friends for moral support and inspiration.

ABSTRACT

The modern world is progressing quickly along with technology and one of the major sectors is Artificial Intelligence. Artificial Intelligence enhances the speed, precision and effectiveness of human efforts. People use artificial intelligence in many sectors. One of the sectors is assistant. In the market lot's of artificial intelligence assistance is available. Some are google assistant, alexa and siri. They all are artificial intelligence based virtual assistants. They give answers to the questions that people ask and also do other stuff. When they answer the question, they do not follow the context of the question. Our thesis is based on this issue. We make such artificial intelligence that follows the context of the question and then it will answer that question. Our thesis is inspired by the above virtual assistants. We make such artificial intelligence that makes us answer based on our discussion context. And also it helps us to find the summary of the discussion and find the best article on the web. And it also has a self learn capability feature. And we are very glad to say that, we build such artificial intelligence that gives us the answer based on our discussion context. And we are very happy with the result of our artificial intelligence.

Table of Content

CANDIDATE’S DECLARATION	i
APPROVAL	ii
ACKNOWLEDGEMENT	iii
ABSTRACT	iv
TABLE OF CONTENT	v
Chapter 1 : Introduction	1-3
1.1 : Objectives	2
1.2 : Motivation	2-3
1.3 : Technology used	3
Chapter 2 : Problem Statement	4-5
2.1 : Problem Statement	5
Chapter 3 : Literature review	6-10
3.1 : Previous related work	7
3.1.1 : BERT	7-9
3.1.2 : BiDAF	9-10
Chapter 4 : Frameworks, Libraries and API	11-17
4.1 : Language	12-13
4.2 : Framework	13-14
4.3 : Libraries	14-16
4.4 : API	17
Chapter 5 : Implementation	18-30
5.1 : Data Collection	19-21
5.2 : Categorize Data	22
5.3 : Convert to tf-idf	22-27
5.4 : Apply ML for all category	27-28

5.5	: Apply ML for sub category-----	28-29
5.6	: Send data in BiDAF model-----	29-30
Chapter 6 : Results-----		31-53
6.1	: Topic detect and context question's answer-----	32-39
6.2	: Any article and answer from that article-----	39-42
6.3	: Search article in offline-----	42-46
6.4	: Get summary from the article-----	46-47
6.5	: Search online and get related article-----	48-50
6.6	: AI learn by its own-----	51-53
Chapter 7 : Conclusion-----		54-55
7.1	: Conclusion-----	55

Chapter One

Introduction

1.1 Objective

We make such an artificial intelligence that makes our life easier. Our artificial intelligence has below features.

- Determine the discussion topic and give the question's answer based on discussion context.
- Give answers from any given topic.
- Find the best article based on the search query in the offline dataset.
- Give a summary for any article.
- Find the best article from online based on a search query.
- Learn about the unknown topic through online and give question's answer about that topic.
- And multi language features.

1.2 Motivation

None of the services from the major tech companies was close to perfect. If I ask Google Assistant “What is DC meaning?”, It says about Direct Current but I need to know about Debit Cards. When I asked Alexa earlier “This week who was playing in the Super Bowl”, she responded, somewhat monotonously, “Super Bowl 49’s winner is New England Patriots.” But the match was played some year before.

Search Engines like Google, Yahoo, Bing, etc. stay in business by selling ads with access to large audiences. The product they sell is not search results but rather access to searchers. With this in mind, the goal of every Search Engine is to grow their search audience, expanding their product offering to companies willing to pay for more ads. They also understand that searchers must trust them to provide the results they are looking for, for them to come back and search again.

But we want to provide vetted, trusted and valuable content to our searchers. We want to provide the exact thing that they want, not any irrelevant thing. The original idea behind us was to allow users to get answers to questions posed in everyday, natural language, as well as by traditional keyword searching.

1.3 Technology Used

❖ Language

- Python
- HTML
- CSS
- JavaScript

❖ Framework

- Flask
- Selenium

❖ Libraries

- Scikit-learn
- Pickle
- Time
- OS
- AllenNLP
- Newspaper3k
- PyTorch
- NLTK(Natural Language Toolkit)

❖ API

- Free Google Translate API for Python
- Free Google Search API for Python
- ngrok (Local port to public url)

Chapter Two

Problem Statement

2.1 Problem Statement

There are lots of virtual assistants available. Those help people to get answers. But they do not follow the question's context. Our main goal of this thesis is to make such an artificial intelligence that gives answers based on context's discussion.

Chapter Three

Literature Review

3.1 Previous Related Work

This section includes the work already done on this thesis by various researchers using different methodologies and algorithms. Following is the brief description of them:

3.1.1 BERT

Pre-training of Deep Bidirectional Transformers for Language Understanding. *[Submitted on 11 Oct 2018 (v1), last revised 24 May 2019 (this version, v2)]*

In paper they introduce a new language representation model called BERT, which stands for Bidirectional Encoder Representations from Transformers. Unlike recent language representation models, BERT is designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers. As a result, the pre-trained BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, such as question answering and language inference, without substantial task-specific architecture modifications.

BERT is conceptually simple and empirically powerful. It obtains new state-of-the-art results on eleven natural language processing tasks, including pushing the GLUE score to 80.5% (7.7% point absolute improvement), MultiNLI accuracy to 86.7% (4.6% absolute improvement), SQuAD v1.1 question answering Test F1 to 93.2 (1.5 point absolute improvement) and SQuAD v2.0 Test F1 to 83.1 (5.1 point absolute improvement).

They have shown that the standard BERT recipe (including model architecture and training objective) is effective on a wide range of model sizes, beyond BERT-Base and BERT-Large. The smaller BERT models are intended for environments with restricted computational

resources. They can be fine-tuned in the same manner as the original BERT models. However, they are most effective in the context of knowledge distillation, where the fine-tuning labels are produced by a larger and more accurate teacher.

BERT is a method of pre-training language representations, meaning that we train a general-purpose "language understanding" model on a large text corpus (like Wikipedia), and then use that model for downstream NLP tasks that we care about (like question answering). BERT outperforms previous methods because it is the first unsupervised, deeply bidirectional system for pre-training NLP.

Unsupervised means that BERT was trained using only a plain text corpus, which is important because an enormous amount of plain text data is publicly available on the web in many languages.

Pre-trained representations can also either be context-free or contextual, and contextual representations can further be unidirectional or bidirectional. Context-free models such as word2vec or GloVe generate a single "word embedding" representation for each word in the vocabulary, so banks would have the same representation in bank deposit and river bank. Contextual models instead generate a representation of each word that is based on the other words in the sentence.

BERT was built upon recent work in pre-training contextual representations — including Semi-supervised Sequence Learning, Generative Pre-Training, ELMo, and ULMFit — but crucially these models are all unidirectional or shallowly bidirectional. This means that each word is only contextualized using the words to its left (or right). For example, in the sentence I made a bank deposit the unidirectional representation of the bank is only based on I made a but not deposit.

Some previous work does combine the representations from separate left-context and right-context models, but only in a "shallow" manner. BERT represents "bank" using both its left and right context — I made a ... deposit — starting from the very bottom of a deep neural network, so it is deeply bidirectional.

3.1.2 BiDAF

Bidirectional Attention Flow for Machine Comprehension.*[Submitted on 5 Nov 2016 ([v1](#)), last revised 21 Jun 2018 (this version, v6)]*

Machine comprehension (MC), answering a query about a given context paragraph, requires modeling complex interactions between the context and the query. Recently, attention mechanisms have been successfully extended to MC. Typically these methods use attention to focus on a small portion of the context and summarize it with a fixed-size vector, couple attention temporally, and/or often form a uni-directional attention. In this paper we introduce the Bi-Directional Attention Flow (BiDAF) network, a multi-stage hierarchical process that represents the context at different levels of granularity and uses bi-directional attention flow mechanism to obtain a query-aware context representation without early summarization. Our experimental evaluations show that our model achieves the state-of-the-art results in Stanford Question Answering Dataset (SQuAD) and CNN/DailyMail cloze test.

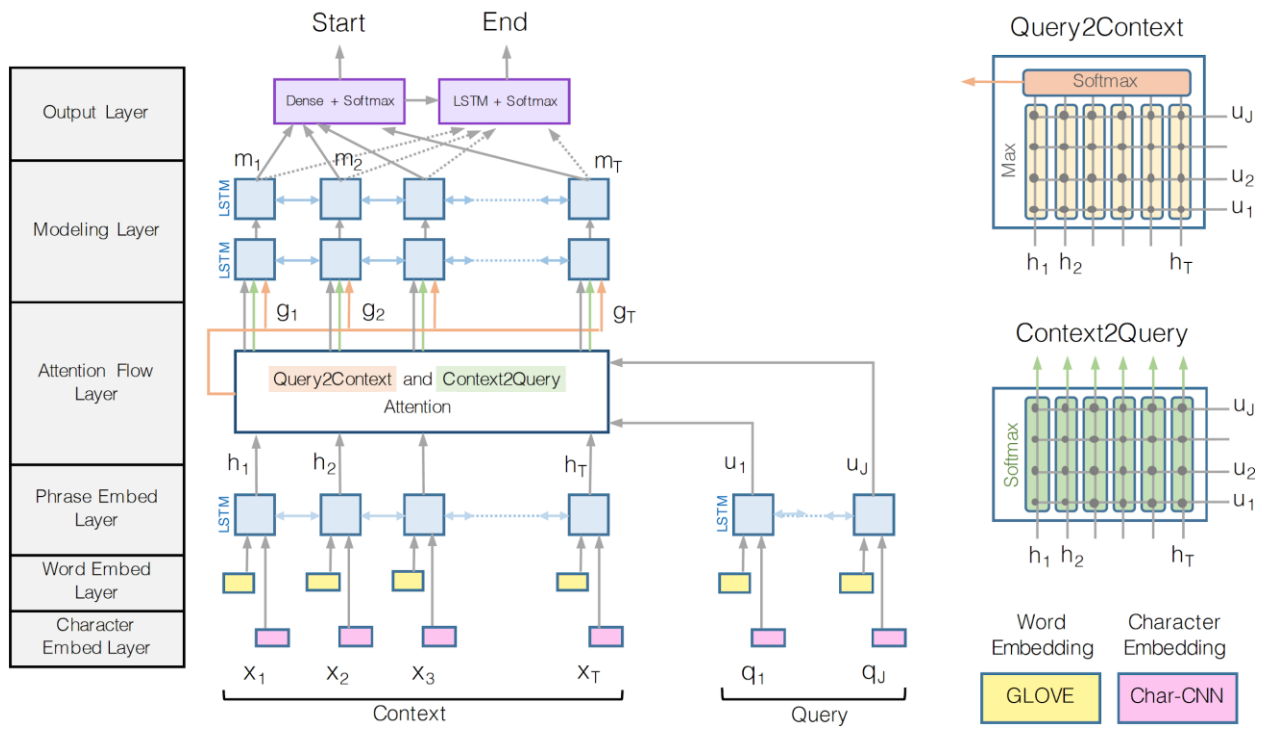


Figure: Bi-Directional Attention Flow for Machine Comprehension

Chapter Four

Frameworks, Libraries and API

For our thesis we use three languages, two frameworks, eight libraries and three api. The languages are Python, HTML, CSS, JavaScript. The frameworks are Flask and Selenium. The libraries are scikit-learn, pickle, time, os, AllenNLP, Newspaper3k, PyTorch, nltk(Natural Language Toolkit). And API's are Google Translate API, Google Search API and ngrok API.

4.1 Language

❑ Python

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Python has a standard library in development, and a few for AI. It has an intuitive syntax, basic control flow, and data structures. It also supports interpretive run-time, without standard compiler languages. This makes Python especially useful for prototyping algorithms for AI.

❑ HTML

Hypertext Markup Language is the standard markup language for documents designed to be displayed in a web browser. It can be assisted by technologies such as Cascading Style Sheets and scripting languages such as JavaScript and VBScript.

❑ CSS

CSS stands for Cascading Style Sheets. CSS describes how HTML elements are to be displayed on screen, paper, or in other media. CSS saves a lot of work. It can control the layout of multiple web pages all at once.

Cascading Style Sheets is a style sheet language used for describing the presentation of a document written in a markup language like HTML. CSS is a cornerstone technology of the World Wide Web, alongside HTML and JavaScript.

❑ JavaScript

JavaScript, often abbreviated as JS, is a programming language that conforms to the ECMAScript specification. JavaScript is high-level, often just-in-time compiled, and multi-paradigm. It has curly-bracket syntax, dynamic typing, prototype-based object-orientation, and first-class functions.

It's an object-oriented computer programming language commonly used to create interactive effects within web browsers.

4.2 Framework

❑ Flask

Flask is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions.

Flask is a web framework. This means flask provides you with tools, libraries and technologies that allow you to build a web application. This web application can be some web pages, a blog, a wiki or go as big as a web-based calendar application or a commercial website.

❑ **Selenium**

Selenium is a portable framework for testing web applications. Selenium provides a playback tool for authoring functional tests without the need to learn a test scripting language.

Selenium is an open-source web-based automation tool. Python language is used with Selenium for testing. It has far less verbose and easy to use than any other programming language. Selenium can send the standard Python commands to different browsers, despite variation in their browser's design.

4.3 Libraries

❑ **Scikit-learn**

Scikit-learn (formerly scikits.learn and also known as sklearn) is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

❑ **Pickle**

Python pickle module is used for serializing and de-serializing a Python object structure. Any object in Python can be pickled so that it can be saved on disk. What pickle does is that it “serializes” the object first before writing it to file. Pickling is a way to convert a python object (list, dict, etc.)

❑ **Time**

The Python time module provides many ways of representing time in code, such as objects, numbers, and strings. It also provides functionality other than representing time, like waiting during code execution and measuring the efficiency of any code.

❑ **OS**

The OS module in Python provides a way of using operating system dependent functionality. The functions that the OS module provides allows you to interface with the underlying operating system that Python is running on – be that Windows, Mac or Linux.

❑ **AllenNLP**

An open-source NLP research library, built on PyTorch. AllenNLP makes it easy to design and evaluate new deep learning models for nearly any NLP problem, along with the infrastructure to easily run them in the cloud or on your laptop. AllenNLP includes reference implementations of high quality models for both core NLP problems (e.g. semantic role labeling) and NLP applications (e.g. textual entailment).

❑ **Newspaper3k**

Newspaper is a Python3 library. Article scraping & curation. Inspired by requests for its simplicity and powered by lxml for its speed. Newspaper is an amazing python library for extracting & curating articles. Newspaper delivers Instapaper style article extraction. It is a very fast and easy way to scrape articles from url. It uses a python request library.

❑ **PyTorch**

PyTorch is an open source machine learning library based on the Torch library, used for applications such as computer vision and natural language processing, primarily developed by Facebook's AI Research lab. It is free and open-source software released under the Modified BSD license. PyTorch is a library for Python programs that facilitates building deep learning projects.

❑ **NLTK(Natural Language Toolkit)**

The Natural Language Toolkit, or more commonly NLTK, is a suite of libraries and programs for symbolic and statistical natural language processing (NLP) for English written in the Python programming language. It was developed by Steven Bird and Edward Loper in the Department of Computer and Information Science at the University of Pennsylvania. NLTK includes graphical demonstrations and sample data. It is accompanied by a book that explains the underlying concepts behind the language processing tasks supported by the toolkit, plus a cookbook.

4.4 API

❑ Google Translate API

Googletrans is a free and unlimited python library that implemented Google Translate API. This uses the Google Translate Ajax API to make calls to such methods as detect and translate. Compatible with Python 3.6+. It has this feature. It is Fast and reliable. It uses the same servers that translate.google.com uses. Auto language detection. Bulk translations. Customizable service URL. HTTP/2 support.

❑ Google Search API

googlesearch is a Python library for searching Google, easily. googlesearch uses requests and BeautifulSoup4 to scrape Google. It's free and very fast. It's a third party library.

❑ ngrok API

pyngrok is a Python wrapper for ngrok that manages its own binary and puts it on your path, making ngrok readily available from anywhere on the command line and via a convenient Python API.

ngrok is a reverse proxy tool that opens secure tunnels from public URLs to localhost, perfect for exposing local web servers, building webhook integrations, enabling SSH access, testing chatbots, demoing from your own machine, and more, and its made even more powerful with native Python integration through pyngrok.

Chapter Five

Implementation

The whole process can be divided into 6 major parts:-

1. Data Collection
2. Categorize Data
3. Convert to tf-idf
4. Apply ML for all category
5. Apply ML for sub category
6. Send data in BiDAF model

5.1 Data Collection

We collect data from google search engine. We are collecting data on a total of nine categories. The categories are:

1. Animation and Comics
2. Computers
3. Education
4. Food and Drink
5. Health
6. Jobs and Career
7. Pets and Animals
8. Politics
9. Religion

For collecting data we use selenium framework. Selenium is an open-source web-based automation tool. Python language is used with Selenium for testing. It has far less verbose and

easy to use than any other programming language. Selenium can send the standard Python commands to different browsers, despite variation in their browser's design.

We collect 1794 articles for the Animation and Comics category. Which has a total 2504098 words and 12742886 characters.

We collect 4891 articles for the Computers category. Which has a total 1255386 words and 7230500 characters.

We collect 4052 articles for the Education category. Which has a total 3127654 words and 17300185 characters.

We collect 4635 articles for the Food and Drink category. Which has a total 3121518 words and 15808709 characters.

We collect 4087 articles for the Health category. Which has a total of 3064762 words and 16554156 characters.

We collect 4875 articles for the Jobs and Career category. Which has a total 2871646 words and 15043733 characters.

We collect 4375 articles for the Pets and Animals category. Which has a total 3669088 words and 18162023 characters.

We collect 2625 articles for the Politics category. Which has a total 1215305 words and 5805350 characters.

We collect 2424 articles for the Religion category. Which has a total 934206 words and 4471838 characters.

In total we collect 33758 articles which contains total 21763663 words and total 113119380 characters .

Our data collection approach was below,

First we search the category in google keyword planner. Google keyword planner gives us keyword ideas about that category. For example when we search “Health” in google keyword planner, they gives below keywords:

health, apple cider vinegar, smoking, optometrist, healthy snacks,nutrition, healthy food near me, healthy food, disease, health department, hipaa, covered california, epidemiology, healthy, healthy breakfast, nurse practitioner, clinic, junk food, tricare online, world health organization, health equity, healthy meals, healthcare, green tea benefits, mental etc.

Google keyword planner gives us a total 140 related keywords for the health category.

Then we search every keyword in google search engine and collect articles from them.

Then we found a total 4087 articles for the health category.

We repeat the above process for all nine categories. So this is a repeated task. Every time we search a keyword in google search engine and then collect articles from google search results. That’s why we take help from the selenium framework and Newspaper3k librarie. Selenium framework collects links from google search results then Newspaper3k collects articles from that links. In this way we collect 4087 articles for the health category.

We repeat this process for all our nine categories. So it makes our work fast.

5.2 Categorize Data

So now we have a total of 33758 articles. Now we will put each article in its own category. We use python list data structure for holding our articles. We have a totally nice giant list whose individual list has lots of articles. Then we convert that giant list to pickle file. Because we have to use this list lots of time. Pickle file converts any kind of python objects (list, dict, etc.) into byte streams (0s and 1s) is called pickling or serialization or flattening or marshalling. So we can use this pickle file easily anywhere.

5.3 Convert to tf-idf

So now we have lots of articles. Which all are string. We all know, computers can't understand strings. It only understands numbers. So now our challenge is to convert those articles to meaningful numbers.

There are lots of techniques available for converting strings to meaningful numbers. Some are Bag of Words (BoW), Term Frequency-Inverse Document Frequency (TF-IDF) and Word2vec technique.

We use the TF-IDF technique for our thesis. TF-IDF short for term frequency-inverse document frequency, is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus.

$$tf_{t,d} = \frac{n_{t,d}}{\text{Number of terms in the document}}$$

Here, in the numerator, n is the number of times the term “t” appears in the document “d”. Thus, each document and term would have its own TF value. For better understanding see the below example.

Suppose we have three articles.

Review 1: This movie is very scary and long

Review 2: This movie is not scary and is slow

Review 3: This movie is spooky and good

We will first build a vocabulary from all the unique words in the above three reviews.

The vocabulary consists of these 11 words: 'This', 'movie', 'is', 'very', 'scary', 'and', 'long', 'not', 'slow', 'spooky', 'good'.

For finding TF, let's consider Review 2

Number of words in Review 2 = 8

TF for the word 'this' = (number of times 'this' appears in review 2)/(number of terms in review 2) = 1/8

Similarly,

$TF('movie') = 1/8$

$TF('is') = 2/8 = 1/4$

$TF('very') = 0/8 = 0$

$TF('scary') = 1/8$

$TF('and') = 1/8$

$TF('long') = 0/8 = 0$

$TF('not') = 1/8$

$TF('slow') = 1/8$

$TF('spooky') = 0/8 = 0$

$TF('good') = 0/8 = 0$

We can calculate the term frequencies for all the terms and all the reviews in this manner:

Term	Review 1	Review 2	Review 3	TF (Review 1)	TF (Review 2)	TF (Review 3)
This	1	1	1	1/7	1/8	1/6
movie	1	1	1	1/7	1/8	1/6
is	1	2	1	1/7	1/4	1/6
very	1	0	0	1/7	0	0
scary	1	1	0	1/7	1/8	0
and	1	1	1	1/7	1/8	1/6
long	1	0	0	1/7	0	0
not	0	1	0	0	1/8	0
slow	0	1	0	0	1/8	0
spooky	0	0	1	0	0	1/6
good	0	0	1	0	0	1/6

So we have to find IDF.

IDF is a measure of how important a term is. We need the IDF value because computing just the TF alone is not sufficient to understand the importance of words:

$$idf_t = \log \frac{\text{number of documents}}{\text{number of documents with term 't'}}$$

We can calculate the IDF values for the all the words in Review 2:

$$\begin{aligned} \text{IDF}(\text{'this'}) &= \log(\text{number of documents} / \text{number of documents containing the word 'this'}) = \\ \log(3/3) &= \log(1) = 0 \end{aligned}$$

Similarly,

$$\text{IDF}(\text{'movie', }) = \log(3/3) = 0$$

$$\text{IDF}(\text{'is'}) = \log(3/3) = 0$$

$$\text{IDF}(\text{'not'}) = \log(3/1) = \log(3) = 0.48$$

$$\text{IDF}(\text{'scary'}) = \log(3/2) = 0.18$$

$$\text{IDF}(\text{'and'}) = \log(3/3) = 0$$

$$\text{IDF}(\text{'slow'}) = \log(3/1) = 0.48$$

We can calculate the IDF values for each word like this. Thus, the IDF values for the entire vocabulary would be:

Term	Review 1	Review 2	Review 3	IDF
This	1	1	1	0.00
movie	1	1	1	0.00
is	1	2	1	0.00
very	1	0	0	0.48
scary	1	1	0	0.18
and	1	1	1	0.00
long	1	0	0	0.48
not	0	1	0	0.48
slow	0	1	0	0.48
spooky	0	0	1	0.48
good	0	0	1	0.48

Hence, we see that words like “is”, “this”, “and”, etc., are reduced to 0 and have little importance; while words like “scary”, “long”, “good”, etc. are words with more importance and thus have a higher value.

We can now compute the TF-IDF score for each word in the corpus. Words with a higher score are more important, and those with a lower score are less important:

$$(tf_idf)_{t,a} = tf_{t,a} * idf_t$$

We can now calculate the TF-IDF score for every word in Review 2:

$$TF-IDF('this', Review 2) = TF('this', Review 2) * IDF('this') = 1/8 * 0 = 0$$

Similarly,

$$\text{TF-IDF}(\text{'movie'}, \text{Review 2}) = 1/8 * 0 = 0$$

$$\text{TF-IDF}(\text{'is'}, \text{Review 2}) = 1/4 * 0 = 0$$

$$\text{TF-IDF}(\text{'not'}, \text{Review 2}) = 1/8 * 0.48 = 0.06$$

$$\text{TF-IDF}(\text{'scary'}, \text{Review 2}) = 1/8 * 0.18 = 0.023$$

$$\text{TF-IDF}(\text{'and'}, \text{Review 2}) = 1/8 * 0 = 0$$

$$\text{TF-IDF}(\text{'slow'}, \text{Review 2}) = 1/8 * 0.48 = 0.06$$

Similarly, we can calculate the TF-IDF scores for all the words with respect to all the reviews:

Term	Review 1	Review 2	Review 3	IDF	TF-IDF (Review 1)	TF-IDF (Review 2)	TF-IDF (Review 3)
This	1	1	1	0.00	0.000	0.000	0.000
movie	1	1	1	0.00	0.000	0.000	0.000
is	1	2	1	0.00	0.000	0.000	0.000
very	1	0	0	0.48	0.068	0.000	0.000
scary	1	1	0	0.18	0.025	0.022	0.000
and	1	1	1	0.00	0.000	0.000	0.000
long	1	0	0	0.48	0.068	0.000	0.000
not	0	1	0	0.48	0.000	0.060	0.000
slow	0	1	0	0.48	0.000	0.060	0.000
spooky	0	0	1	0.48	0.000	0.000	0.080
good	0	0	1	0.48	0.000	0.000	0.080

We have now obtained the TF-IDF scores for our vocabulary. TF-IDF also gives larger values for less frequent words and is high when both IDF and TF values are high i.e the word is rare in all the documents combined but frequent in a single document.

In this above way we convert our all articles to Term Frequency-Inverse Document Frequency (TF-IDF).

So this technique works completely fine for our thesis. We didn't use the Bag-of-Words technique because Bag-of-Words has some drawbacks for our thesis. The drawbacks are:

- If the new sentences contain new words, then our vocabulary size would increase and thereby, the length of the vectors would increase too.
- Additionally, the vectors would also contain many 0s, thereby resulting in a sparse matrix (which is what we would like to avoid)
- We are retaining no information on the grammar of the sentences nor on the ordering of the words in the text.

5.4 Apply ML for all category

Now we have a huge articles list and we converted that articles to numbers. Now it's time to apply machine learning algorithms. There are lots of machine learning algorithms available. Like Linear Regression, Logistic Regression, Decision Tree, SVM, Naive Bayes, k NN, K-Means, Random Forest, Dimensionality Reduction Algorithms, Gradient Boosting algorithms, GBM, XGBoost, LightGBM, CatBoost etc.

For our dataset we use Naive Bayes MultinomialNB algorithm. The multinomial Naive Bayes classifier is suitable for classification with discrete features (e.g., word counts for text classification). The multinomial distribution normally requires integer feature counts. However, in practice, fractional counts such as tf-idf may also work.

Multinomial Naive Bayes is a specialized version of Naive Bayes that is designed more for text documents. Whereas simple naive Bayes would model a document as the presence and absence of particular words, multinomial naive Bayes explicitly models the word counts and adjusts the underlying calculations to deal with it.

Now we apply the MultinomialNB algorithm for our dataset. In Python there is a library called scikit-learn. Scikit-learn (formerly scikits.learn and also known as sklearn) is a free software machine learning library for the Python programming language. It features

various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and MultinomialNB.

So we just use this scikit-learn library to apply the MultinomialNB algorithm.

Our input was all nine giant lists. And output was integer 0 to 8. 0 means first category, 1 means second category and so on.

In this way we build such artificial intelligence that takes input and gives output about any nine categories. This artificial intelligence only can say, given string is which category. It can't detect the actual article where the given string belongs to. Our next section solves this issue.

5.5 Apply ML for sub category

In the above section we successfully determined the given string is which category. Now our task is to find the exact article where the given string belongs to.

Now again we apply Naive Bayes MultinomialNB algorithm. Because we are working with text. But this time our goal is to find the exact article where our article belongs to. That's why our input and output will be different.

We apply the MultinomialNB algorithm nine times for individual lists. For example if our category is health then our input will be all the 4087 articles. And our output will be 0 to 4086. 0 means first article, 1 means second article and so on.

We apply this process for all our nine categories. In this way our AI can determine which articles a given string belong to. We apply all the algorithms using scikit-learn library.

From the above section our AI determines the given string is which category and after finding the category it finds the article in that category where given input belongs to. In this way we get the exact article from our huge article lists.

5.6 Send data in BiDAF model

Now we have the exact article where given input belongs. Now we feed this article to BiDAF model. BiDAF is a pre-trained deep learning model by AllenNLP. This model is also called Bi-Directional Attention Flow for Machine Comprehension. This model is trained by SQuAD2.0 dataset. SQuAD2.0 is a Stanford Question Answering Dataset.

Stanford Question Answering Dataset (SQuAD) is a reading comprehension dataset, consisting of questions posed by crowdworkers on a set of Wikipedia articles, where the answer to every question is a segment of text, or span, from the corresponding reading passage, or the question might be unanswerable.

SQuAD2.0 combines the 100,000 questions in SQuAD1.1 with over 50,000 unanswerable questions written adversarially by crowdworkers to look similar to answerable ones. To do well on SQuAD2.0, systems must not only answer questions when possible, but also determine when no answer is supported by the paragraph and abstain from answering.

This model is capable of answering the question from the given article. So our task is just to send the article in this model with a question. This model gives us the possible answer for that question.

So we just send the article which we found from the above section with a question. The model will give us the answer.

So finally we found such an answer that belongs to our context.

Chapter Six

Results

We implement our whole thesis into a website. So anyone can use that website if the website is live. We make our website using the Flask framework. Flask is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions.

We have divided our AI into six sections. It's section has a unique feature.

- Section 1: Topic detect and context question's answer
- Section 2: Any article and answer from that article
- Section 3: Search article in offline
- Section 4: Get summary from the article.
- Section 5: Search online and get related article
- Section 6: AI learn by its own

6.1 Topic detect and context question's answer

In this section our artificial intelligence gives an answer based on the discussion context. Here our AI needs two inputs. First is discussion and second is question. In the website our first section looks like below picture.

Topic detect and context question's answer

Discussion

Give your discussion.....

Question

Give question....

Detect

Clear

People just give their discussion in the discussion box and ask a question in the question box. After giving the discussion and asking questions just press the Detect button. Then our AI will give the answer.

Topic detect and context question's answer

Discussion

Nowadays bank is important in our life.

Question

what is the bank

Detect

Topic: Bank Sector

Answer: a financial institution that accepts deposits and recurring accounts from the people and creates Demand Deposit

Clear

Here our AI determines two things. First is what topics are discussed and second is what is the answer of the question based on discussion context. Given I give some other examples, that can help to better understand.

Topic detect and context question's answer

Discussion

I have an account in Dutch-Bangla Bank. They have online transaction feature. You have to create an account in this bank. It's a great bank. I have an account in Dutch-Bangla Bank. They have online transaction feature. You have to create an account in this bank. It's a great bank.

Question

What is DC meaning

Detect

Topic: Bank Sector
Answer: Debit Card

Clear

In the above picture our question is “what is DC meaning”. Our AI answer that question is Debit Card. Because above our discussion topic is bank related. That’s why AI gives bank related answer.

Topic detect and context question's answer

Discussion

Social media is a great tool for communication. I have Facebook and Instagram account. Do you have any social media account?

Question

What is DC meaning

Detect

Topic: Social Network
Answer: In Tiktok DC meaning is Dance Challenge

Clear

Again our question is the same. “what is DC meaning”. But this time our AI answer is different. The AI gives DC meaning is Dance Challenge.

Topic detect and context question's answer

Discussion

The Marvel Cinematic Universe is an American media franchise and shared universe centered on a series of superhero films, independently produced by Marvel Studios and based on characters that appear in American comic books published by Marvel Comics.

Question

What is DC meaning

Detect

Topic: Animation and Comics

Answer: Detective Comics

Clear

This time our AI gives a different answer for the same question.

Topic detect and context question's answer

Discussion

Singapore is a beautiful country. One day I will visit that country. I also want to visit United State.

Question

What is DC meaning

Detect

Topic: Country and City

Answer: District of Columbia

Clear

This time the AI answer is District of Columbia. Because it detects discussion topics are Country and city. That's why AI detect DC means Washington, D.C in the United State of America. See some other examples.

Topic detect and context question's answer

Discussion

Time is very important in our life. Our life is very short but uncertain; so every moment of our life is very precious.

Question

what is date

Detect

Topic: Time Zone

Answer: Date indicates a specific day in a specific year.

Clear

Our question is what is date. See the answer.

Topic detect and context question's answer

Discussion

Ranveer Singh loves Deepika Padukone. Finally he married her. They are very happy now.

Question

what is date

Detect

Topic: Love and Relationships

Answer: date is a stage of romantic relationships in humans whereby two people meet socially with the aim of each assessing the other's suitability as a prospective partner in an intimate relationship.

Clear

Now again our question is what is date. Now see the answer. The answer is different. Because our context is different.

In this first section it has multi language features available. Below are given some examples.

Topic detect and context question's answer

Discussion

আজকাল ব্যাংক আমাদের জীবনে গুরুত্বপূর্ণ।

Question

ব্যাংক কি

Detect

বিষয়: ব্যাংক সেक्टर

উত্তর: একটি আর্থিক প্রতিষ্ঠান যা লোকদের কাছ থেকে আমানত এবং পুনরুদ্ভি অ্যাকাউন্ট গ্রহণ করে এবং ডিম্যান্ড ডিপোজিট তৈরি করে

Clear

Detect Bangla Language.

Topic detect and context question's answer

Discussion

如今，銀行在我們的生活中至關重要。

Question

銀行是什麼

Detect

主題：銀行部門

答案：一家接受人民存款和定期賬戶并創建活期存款的金融機構

Clear

Detect Chinese Language

Topic detect and context question's answer

Discussion

Nowadays bank is important in our life.

Question

what is the bank

Detect

Topic: Bank Sector

Answer: a financial institution that accepts deposits and recurring accounts from the people and creates Demand Deposit

Clear

Detect English Language

Topic detect and context question's answer

Discussion

आजकल बैंक हमारे जीवन में महत्वपूर्ण है।

Question

बैंक क्या है

Detect

विषय: बैंक सेक्टर

उत्तर: एक वित्तीय संस्थान जो लोगों से जमा और आवर्ती खातों को स्वीकार करता है और डिमांड डिपॉजिट बनाता है

Clear

Detect HindiLanguage

Topic detect and context question's answer

Discussion

Hoy en día el banco es importante en nuestra vida

Question

que es el banco

Detect

Tema: Sector bancario

Respuesta: una institución financiera que acepta depósitos y cuentas recurrentes de las personas y crea depósitos a la vista.

Clear

Detect SpanishLanguage

6.2 Any article and answer from that article

In this section artificial intelligence gives the answer based on the given article.

Any article and answer from that article

Article

Give your article.....

Question

Give question....

Detect

Clear

Section two looks like above.

Below given some examples of how this section actually works.

Any article and answer from that article

Article

It also gives you the option to identify different people in your photos by manually putting a name to the face. After that, pictures with those people are automatically sorted so you can later search for pictures featuring specific people.

In a similar vein, you can also set it to create "live albums," which automatically populate with photos of friends and family members.

For those looking to backup their physical photo prints, you can quickly 'scan' those to have them uploaded by taking a quick pic using your phone or other Photos-friendly device. And for those who photograph paper documents, Google Photos also lets you highlight desired sections of text, and even crop out backgrounds to make it easier to do things like upload and expense a receipt from a work trip.

Google Photos is a powerful and versatile tool that requires little effort to use to its fullest. And, given the fact that it provides free, unlimited storage without sacrificing too much on photo and video quality, it can be a solid option for backing up your media files.

Question

when google photos launched

Detect

May 2015

Clear

Any article and answer from that article

Article

It also gives you the option to identify different people in your photos by manually putting a name to the face. After that, pictures with those people are automatically sorted so you can later search for pictures featuring specific people.

In a similar vein, you can also set it to create "live albums," which automatically populate with photos of friends and family members.

For those looking to backup their physical photo prints, you can quickly 'scan' those to have them uploaded by taking a quick pic using your phone or other Photos-friendly device. And for those who photograph paper documents, Google Photos also lets you highlight desired sections of text, and even crop out backgrounds to make it easier to do things like upload and expense a receipt from a work trip.

Google Photos is a powerful and versatile tool that requires little effort to use to its fullest. And, given the fact that it provides free, unlimited storage without sacrificing too much on photo and video quality, it can be a solid option for backing up your media files.

Question

how many storage it's give

Detect

unlimited

Clear

Any article and answer from that article

Article

It also gives you the option to identify different people in your photos by manually putting a name to the face. After that, pictures with those people are automatically sorted so you can later search for pictures featuring specific people.

In a similar vein, you can also set it to create "live albums," which automatically populate with photos of friends and family members.

For those looking to backup their physical photo prints, you can quickly 'scan' those to have them uploaded by taking a quick pic using your phone or other Photos-friendly device. And for those who photograph paper documents, Google Photos also lets you highlight desired sections of text, and even crop out backgrounds to make it easier to do things like upload and expense a receipt from a work trip.

Google Photos is a powerful and versatile tool that requires little effort to use to its fullest. And, given the fact that it provides free, unlimited storage without sacrificing too much on photo and video quality, it can be a solid option for backing up your media files.

Question

is google photo free

Detect

Google Photos comes with free, unlimited storage

Clear

Any article and answer from that article

Article

It also gives you the option to identify different people in your photos by manually putting a name to the face. After that, pictures with those people are automatically sorted so you can later search for pictures featuring specific people.

In a similar vein, you can also set it to create "live albums," which automatically populate with photos of friends and family members.

For those looking to backup their physical photo prints, you can quickly 'scan' those to have them uploaded by taking a quick pic using your phone or other Photos-friendly device. And for those who photograph paper documents, Google Photos also lets you highlight desired sections of text, and even crop out backgrounds to make it easier to do things like upload and expense a receipt from a work trip.

Google Photos is a powerful and versatile tool that requires little effort to use to its fullest. And, given the fact that it provides free, unlimited storage without sacrificing too much on photo and video quality, it can be a solid option for backing up your media files.

Question

who is the president of united states

Detect

Google Photos

Clear

The last answer is wrong. Because AI doesn't know anything else instead of the above article. That's why it gives the wrong answer.

Any article and answer from that article

Article

My name is Donald Trump. I am the president of united states of america. I am very powerful.

Question

who is the president of united states of america

Detect

Donald Trump

Clear

Any article and answer from that article

Article

My name is Donald Trump. I am the president of united states of america. I am very powerful.

Question

What is my name

Detect

Donald Trump

Clear

This time AI gives the current answer.

6.3 Search article in offline

In this section our artificial intelligence gives an article based on search query. This section works offline. Because when we trained our AI, AI saw 33758 articles. So these 33758 articles are available in the AI database. So AI can easily find the search query related article in his database.

Search article in offline

Topic

Give topic....

Get Article

Clear

Section three looks like the picture above.

Now below given some examples of how this section actually works.

Search article in offline

Topic

big hero 6 marvel

Get Article

This fall, Walt Disney's famed feature animation division is adapting one of Marvel's comic series, Big Hero 6, into a full-length feature film. This film, which premieres November 7, follows as the next film from Walt Disney Animation Studios after the record-breaking success of Frozen. There's considerable interest in this movie given it's a melding of Marvel's characters and Walt Disney's famed animation division, but when movie-goers look to find the source material or find what's next – they won't find anything. Not from Marvel, at least. Despite the Big Hero 6 movie being synergistic product of Disney's acquisition of the company in 2009, Marvel tells Newsarama that there are no plans to feature the characters of the film as part of any upcoming comics projects – neither as guest stars, in their own series, or even reprinting earlier stories that inspired the movie. When asked for further comment on Big Hero 6 at Marvel Comics, the publisher declined to comment. Originally created by Steven T. Seagle and Duncan Rouleau during their tenure on Alpha Flight, Big Hero 6 actually debuted in a spin-off series by Scott Lobdell and Gus Vasquez titled Sunfire & Big Hero 6 in the fall of 1998. After that, the team popped up intermittently as guest stars of various secondary books in Marvel's comics line such as Thunderbolts and Defenders, but by and large was relegated to the back issue bins for ten years. In 2008 Marvel commissioned famous Uncanny X-men writer Chris Claremont to create a new Big Hero 6 miniseries with David Nakayama, and they were last seen in 2012's Amazing Spider-Man #685 as a Japan-based force to repel an assault by Doctor Octopus. But for fans interested in the upcoming Big Hero 6 movie and looking for its comics counterpart, those comic stories are hard to come by. The last collected edition of Big Hero 6 stories was a 2012 reprint of Claremont's 2008 series, but that has been out of print for some time (It is, however, available digitally on comiXology for \$8.99). The single issues of the original Sunfire & Big Hero 6 series are going on eBay for upwards of \$30, and their first intended appearance in Alpha Flight #17 has risen from being a title found in the dollar bin to fetching upwards of \$15 and \$20 online. Although Big Hero 6 comics neither new or old are planned from Marvel, in a surprising move the manga publisher Kodansha is publishing a Big Hero 6 comic inside it's Japanese anthology Magazine Special. Done by manga-ka Haruki Ueno, Big Hero 6's squishy mascot Baymax has his own serial which began in the magazine's August 6th issue. Kodanasha have done a number of manga adaptations of Walt Disney animated movies recently such as Frozen, and this actually isn't the first Kodanasha project of Marvel characters; back in 1970, Kodanasha published Japanese-inspired retelling of Spider-Man's origin by Ryoichi Ikegama that American audiences didn't see officially until an English reprint in 1997. In the past, Marvel have shown reticence publishing comics to capitalize on the mainstream popularity of Marvel-based movies if they weren't produced in-house by Marvel's film division, and earlier this month Marvel Senior Vice President of Publishing Tom Brevoort explained the reason. "If you had two things," Brevoort said on his Tumblr page, "and on one you earned 100% of the revenues from the efforts that you put into making it, and the other you earned a much smaller percentage for the same amount of time and effort, you'd be more likely to concentrate more heavily on the first, wouldn't you?" While the Big Hero 6 film is being created under the same Walt Disney umbrella that Marvel is under, the same logic Brevoort presented here is true: as Big Hero 6 is being produced by Walt Disney Animation Studios and not Marvel Studios, Marvel itself will receive far less returns on any success Big Hero 6 has than focusing on comics tying into, for example, Guardians of the Galaxy. In 2013, Marvel Television's Vice President of Animation Development and Production Cort Lane told Newsarama that "Big Hero 6 truly belongs to Disney Feature Animation Studio," which may be more true than you'd think. While it is true that in the build-up to this year's X-Men: Days of Future Past movie Marvel reprinted the original story-arc and did a thematic tie-in with the X-Men: Battle of the Atom event, the X-Men remain a huge pillar in Marvel's comics line despite its movie rights being owned by 20th Century Fox. Big Hero 6 has never been in that position at Marvel Comics, not even to the degree Guardians of the Galaxy had been prior to that movie's announcement. With Big Hero 6 coming out under the Walt Disney Animation Studios umbrella however, other Disney-owned publishing companies have been quick to create tie-in products for the movie; Random House has over a dozen Big Hero 6 books planned, from a junior novelization to an "Art of " book, and even two "Little Golden" books. Hiro and Baymax will appear as new characters in Disney Infinity 2.0, but their inclusion in the video game is under the "Disney Originals" category, not the "Marvel Super Heroes" category. Newsarama reached out to several comic creators who have worked on Big Hero 6 in the past, including co-creators Seagle and Rouleau, but as of press time have received no response.

Clear

Search article in offline

Topic

non alcoholic beer

Get Article

Non-alcoholic beer is quickly gaining traction today. But what is it and is it really safer to drink than alcohol? Today, there are a number of different brand-name and craft non-alcoholic beers to choose from, creating more options for anyone who drinks alcohol. Studies, however, have shown that this might not be the best option for some people, like alcoholics who are in recovery. Does Non-Alcoholic Beer Contain Alcohol? Non-alcoholic beer, also sometimes known as near beer is a misleading term. According to labeling regulations, non-alcoholic beer is not required to be alcohol-free and can contain some alcohol. Standard beer varies considerably in alcohol content. The average real beer contains around 5% alcohol per beer. Low-alcohol beer can contain between 0.5% and 1.2% alcohol, and non-alcoholic beer can actually contain up to 0.5 percent alcohol. There are very few, if any, non-alcoholic beers that are actually free of alcohol. In fact, some studies have shown that non-alcoholic beers may actually have more alcohol than what's listed on the label. Even if it does claim to be 5% alcohol, there may actually be quite a bit more alcohol in the non-alcoholic beer. Most will contain a small amount, which means they're not a good option for those who need to abstain from alcohol completely. Is Non-Alcoholic Beer Safe for Pregnancy? Studies have shown that alcohol can have serious impacts on the growing baby. The exact amount of alcohol that is dangerous to a fetus is unknown. What is known is that either frequent drinking of alcohol, or a single binge, can damage a growing baby. The truth is that there is no known amount of alcohol that is safe or not. A "safe" threshold has never been determined. Most women will, therefore, abstain from alcohol completely when they find out they're pregnant. However, with non-alcoholic beer, a lot of women may wonder if this is a safer option that enables them to enjoy a beer occasionally while they're pregnant without the worry that a typical beer might cause. Since non-alcoholic beers can contain some alcohol, and sometimes more than is stated on the label, it's best for anyone who is pregnant to avoid drinking non-alcoholic beer as well as regular beer. There is no known safe amount of alcohol, so those who want to stick with the rules will want to make sure they don't drink anything that could contain alcohol. A pregnant woman who is worried about not being able to drink should let their obstetrician know as soon as possible. Can I drink NA beer after I give birth? As much as you may really want a beer now that your pregnancy is over, the answer is still "probably not." After the baby is born, it's still not a good idea to drink any beer if the woman is breastfeeding. The amount of alcohol that is found in the bloodstream after drinking is the same amount that will be found in the breast milk. Long-term effects of this are still unknown, but even non-alcoholic beers can contain some alcohol and thus should be avoided while the mother is breastfeeding. Are there Risks of Drinking Non-Alcoholic Beer? There are risks associated with drinking alcohol, but are they lowered with a reduced alcohol intake? In some cases, they can be. Some studies have shown that non-alcoholic beer can help reduce the amount of time it takes for someone to fall asleep or help with anxiety. Non-alcoholic beer, nevertheless, can still contribute to liver damage. It's still not a safe option for those who are worried about liver-related medical conditions or who already suffering from medical issues with their liver. It is also dangerous to those suffering from pancreatitis. Since most alcohol is processed through the liver, even the small amount of alcohol in non-alcoholic beers can cause further damage for those who are already suffering from issues with their liver. This includes cirrhosis of the liver and a condition known as a fatty liver. Those who already have either of these conditions, other liver conditions, or who are at risk of developing these conditions will want to refrain from all alcohol, which includes non-alcoholic beers. Is Drinking Non-Alcoholic Beer Safe for Alcoholics? Once in a rare while, someone will tout non-alcoholic beer as an excellent alternative for recovering alcoholics, but this really is not the case. The claim is that the person can still enjoy a beer now and again, without worrying about actually drinking a lot of alcohol. They won't be able to get drunk, will not be at risk of driving drunk, and will significantly reduce the amount of alcohol they drink, which could help prevent medical issues they might be at risk for. The problem with this claim is that even non-alcoholic beer can trigger cravings, resulting in a relapse for many alcoholics. Triggers can occur because of the smell of the beer, the act of opening a bottle or a can, the taste of the beer, or just the idea that they're able to have a beer. These feelings remind them of when they used to drink and, without realizing it, they can easily switch back to drinking a standard beer and relapsing into their alcoholism. People in recovery Sobriety is difficult and many people have already lived for years trying to bend as many rules as possible so they can continue to drink or use drugs. If they have a non-alcoholic beer, it's just one more rule they've bent to be able to continue abusing alcohol. Though they might not be at risk for driving under the influence, they're not going to get drunk from the non-alcoholic beer, and they are significantly decreasing the amount of alcohol they're drinking, the fact is, they're still drinking alcohol. This can absolutely trigger a relapse into alcoholism or drug addiction. Another issue is that the amount of alcohol in the non-alcoholic beer might actually be more than they expect. This means they might end up drinking more than they thought they would. Even one or two drinks at night, while it won't cause them to be drunk, can be trigger cravings and a subsequent relapse. Those who want to make sure they recover from their addiction will want to make sure they avoid any alcohol at all. This means they'll want to stay away from non-alcoholic beers. Though non-alcoholic beers are quickly becoming more popular, they're not the best option for everyone. In fact, those who are pregnant, breastfeeding, dealing with liver ailments, or who are recovering from a substance abuse addiction should avoid non-alcoholic beers just like they avoid standard beers. These beers, though they're called non-alcoholic, do contain some alcohol, which means they're not an option for anyone who wants or needs to completely stop drinking alcoholic beverages. References: <https://www.drinkaware.co.uk/alcohol-facts/alcoholic-drinks-units/alcoholic-and-non-alcoholic-beers/> <https://www.thebump.com/a/drinking-alcohol-while-pregnant> <https://beerandhealth.eu/beer-and-health/health-aspects-of-non-alcoholic-beer/> <https://newliffehouse.com/non-alcoholic-beer-considered-relapse/>

Clear

Search article in offline

Topic

chicken recipes

Get Article

Here at BA, we publish a looooootttt of chicken recipes, and choosing which fall into the "very best chicken recipes" isn't easy. After all, you can prepare it loads of different ways, each with its own merits: There's fried, roasted, marinated, poached, sandwiched. Some of the best chicken recipes don't even need sides (looking at you chicken cordon bleu, you too, popcorn chicken)! And while we love all our chicken recipes, there's always going to be a pecking order. These are our very best chicken recipes—the dishes our staff and readers alike turn to again and again. In other words, this is our Chicken Recipes Hall of Fame.

Clear

Some of the articles are not good. Because it's totally offline.

6.4 Get summary from the article

In this section AI gives the summary from the given article.

Get summary from the article

Article

Give your article OR URL

Summary

Clear

Section four looks at the picture above.

Below given some examples of how this section works.

Get summary from *the article*

Article

I didn't get YouTube Premium by subscribing to it, but instead through my existing subscription to Google Play Music, the least-buzzed-about Spotify competitor, which is just as good as the industry leader or Apple Music. That's right, all Google Play Music subscribers get YouTube Premium, and all YouTube Premium subscribers get Google Play Music -- and YouTube Music Premium.

The Google Play Music library offers just as many songs as Spotify, plus it includes a cloud locker service, so MP3 collectors can access their rarities anywhere, with up to 50,000 uploads. The service isn't limited to just songs, though, as Google Play Music's recently added podcasts, so you can pack all of your listening needs into a single app.

In fact, Google plans to bridge the gap between YouTube Premium and Google Play Music in the near future. During a panel session for the New Music Seminar conference in New York, YouTube's head of music [Lyor Cohen](#) told audiences that the company is looking into "combining YouTube Premium and

Summary

YouTube is so much better without the ads If you're like me and the other billion-plus YouTube video watchers, you've come to rely on the service as a super-channel that provides an amazingly wide variety of content on your schedule.

Image: DoublePHOTO studio / Shutterstock.comImage: DoublePHOTO studio / Shutterstock.com This way, you can keep a handful of your favorite music videos on hand for your next dance break, or do what I do, and save the video versions of your favorite podcasts at home so you're not limited to the audio only versions on the go.

You get YouTube Music Premium and Google Play Music for free too I didn't get YouTube Premium by subscribing to it, but instead through my existing subscription to Google Play Music, the least-buzzed-about Spotify competitor, which is just as good as the industry leader or Apple Music.

That's right, all Google Play Music subscribers get YouTube Premium, and all YouTube Premium subscribers get Google Play Music -- and YouTube Music Premium.

Clear

Get summary from *the article*

Article

<https://www.businessinsider.com/how-does-google-photos-work>

Summary

Launched in 2015, Google Photos is a tool that can store photos, videos and screenshots taken by your phone.

For those who don't have a Google device (like the Pixel phone), you can opt to automatically backup and sync your photos and videos as you take them.

You can also set your Google Drive to keep a copy of your photos and videos by going to your Drive's settings, turning on "Create a Google Photos Folder," and clicking save.

For those who prefer to go with original file sizes, it's free up to 15GB, and that includes everything from Photos to Gmail to everything else in the Drive.

Clear

If we just put the url of the article. It is able to give the summary from that article.

6.5 Search online and get related article

This is section five. In this section users give a search query. Then our AI searches online and finds the best article for the query.



The image shows a web form with a light pink background. At the top, the title "Search online and get related article" is centered. Below the title is a text input field with the placeholder text "Give topic.....". Above the input field, the word "Topic" is centered. Below the input field is a blue button with the text "Get Article". Below the button is a large, empty rectangular box with a dashed border. At the bottom of the form is a black button with the text "Clear".

Our section five looks at the above picture.

Now below give an example how this section works.

Search online and get related article

Topic

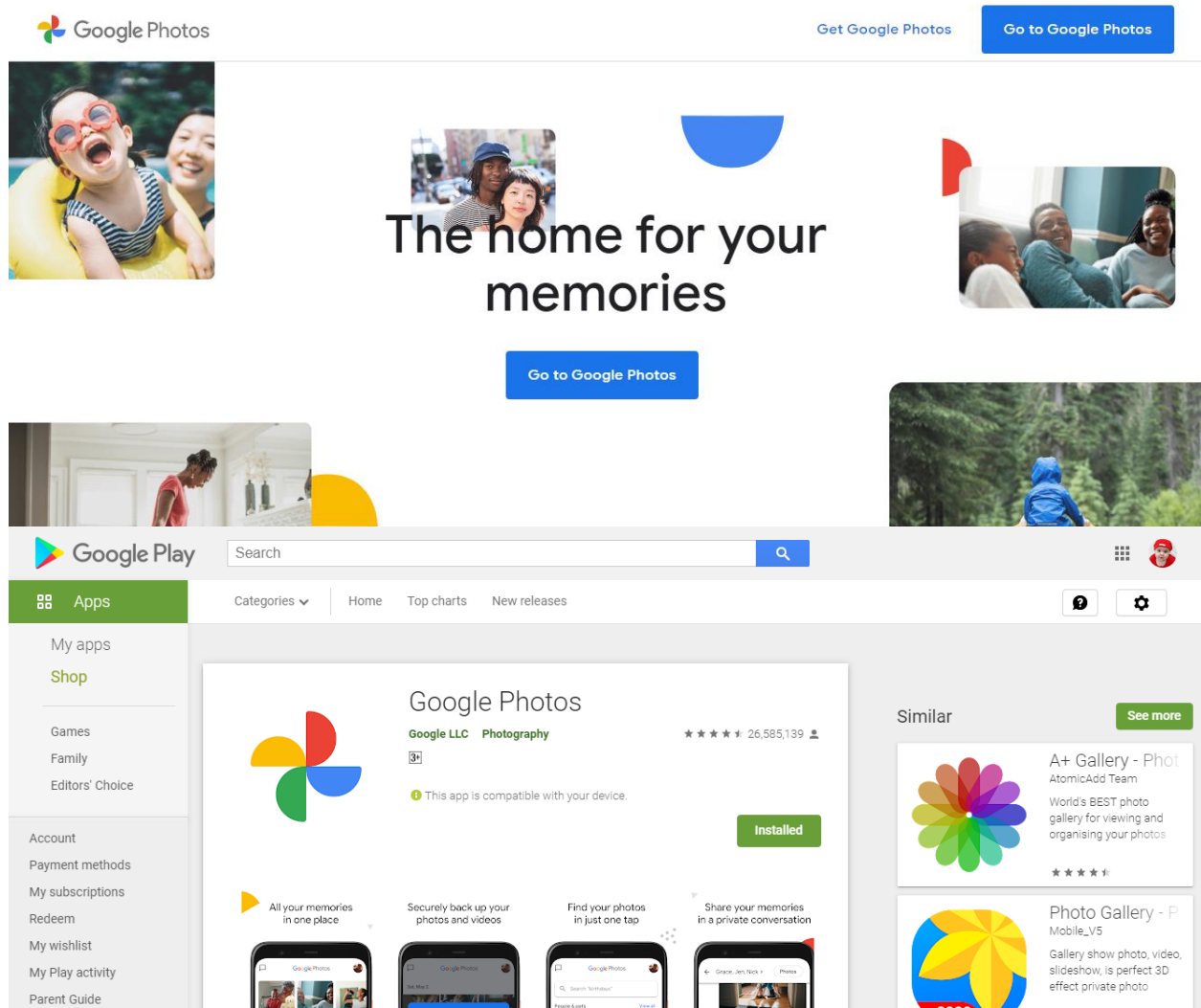
what is google photos

Get Article

Google Photos may be one of the best online services for people who love taking lots of pictures and videos with their smartphones. It combines cloud-based storage with its AI-based organization of images and clips. With these features, it's a must for mobile photographers. Let's take a quick look at how to get started using Google Photos, including a look at its major features, and some tips that should make using the service easier for you. What is Google Photos? Google Photos officially launched in May 2015 as a standalone spin-off of sorts from Google+ Photos, which was based in the Google+ social network (Google would later retire Google+ Photos). It retained many of the features of the older Google+ Photos, as well as some new features. The apps and the service were immediate hits with smartphone owners, hitting 500 million users by in May 2017, two years after it launched. Those users upload 1.2 billion photos every day. In June 2017, the company announced Google Photos had reached over 1 billion app downloads. You can download the Google Photos app on both Android and iOS, and you can also go to the photos.google.com site to view your stored images and clips on a PC or mobile web browser. What's the difference between High Quality, Original Quality, and Express options in Google Photos? Google Photos has three storage settings to choose from, with some rather confusing labels. The "High Quality" setting is actually the middle pick of the two modes. It allows you to back up photos from your smartphone up to 16MP each or video clips at up to 1080p resolution to your Google Photos cloud account. Using the High Quality setting lets you upload an unlimited number of photos and videos for free, if they conform to those limits. Google also recently updated its terms of service for videos stored on High Quality in Google Photos. The following video formats are supported: mpeg, mod, mmv, tod, wmv, asf, avi, divx, mov, m4v, 3gp, 3g2, mp4, m2t, m2ts, mts, and mkv. If you upload videos in other formats for High Quality, including the popular RAW format, they will be counted as part of your Google One storage limits. Related Google One vs the competition: Dropbox, OneDrive, iCloud, and more Google Google's updated subscription plans for cloud storage now fall under its new Google One program. Announced back in May 2018, Google One first launched in the U.S. and rolled out to users across ... The "Original Quality" setting for Google Photos is the highest pick of the three modes. It will preserve the original quality, megapixels, and resolution of any photo or video you upload, which sometimes exceed the limits of the High Quality setting. However, for most smartphones with Google Photos installed, there is an upper limit of cloud storage for the Original Quality setting. They will be stored on the user's Google One cloud account, and they will share storage space with the user's Google Drive documents and Gmail email messages. The good news is there are a set of smartphones can get around the storage limitations of the Original Quality setting in Google Photos. If you own the original Google Pixel or the Google Pixel XL, you can upload photos and videos at the Original Quality setting on Google Photos for free and never hit a cloud storage limit until the end of 2020. If you own the Pixel 2 or Pixel 2 XL, you can also upload as many photos or videos you want at the Original Quality setting until the end of 2021. After that timeframe, any new photos or videos you take with those phones will be compressed down to the High Quality resolution and megapixels for Google Photos storage. Credit: Kris Carlon / Android Authority The most recent Google smartphones, the Google Pixel 3 and Pixel 3 XL, have a similar deal for Google Photos in terms of storage. Owners can upload as many photos and videos at the Original Quality setting on Google Photos for free until the end of 2021. After that, any more photos and videos they upload will be converted down to the High Quality settings. For people with a limited amount of data on their smartphone plans, particularly in developing countries and markets such as India, Google Photos recently added an Express backup option for some users. This will compress any photo down to just 3MP before it is uploaded to Google Photos, and videos are also cut down to the standard 480p definition. The Google Photos app also recently added the option to cap the amount of mobile data used by the app. Users can limit it to 5MB, 10MB, or 30MB, or they can choose for Google Photos to not use any mobile data at all when uploading content. Google Photos AI grouping of images into groups and albums Google Photos also uses artificial intelligence and machine learning to identify items in your pictures. Those items can then be placed in their own groups and albums. If you take a lot of photos of your dog with your smartphone, Google Photos will see it, then group all those dog photos into one album on your app or the Google Photos website. In fact, an update in 2017 added the ability to identify the specific dog in the pictures when uploaded. When you tap on the albums section of the Google Photos app, it shows you three categories on top to search for the image you want to find: People, Places, and Things. The People category shows images of faces, and the Places category shows locations, based on both geotagging information and also identifying specific well-known landmarks in the photo. The Things category can not only show non-human subjects but also photos and videos taken for specific events, like birthdays or vacations. Photo books support in Google Photos If you see a group of images in Google Photos and want to bring them into the physical world, the service also lets you order Photo Books from your collected images. In fact, using AI, Google Photos arrange the book, adding the photos it think are best. Of course, you can still manually select which images go in the photo book as well. Pricing for these books begin at \$9.99 and they can be ordered on the website and in the Google Photos apps. Editing images in Google Photos Google Photos also lets you edit your photos, either in the apps or on its website. There's an auto selection which allows Google Photos to create what it believes is the best looking version of that photo, and a number of other filters you can select like Vista, which turns color images into black and white photos. You can also manually adjust the light, color, and pop options on your images with sliders, and you can also change the aspect ratio and angles of your pictures within Google Photos as well. Sharing Google Photos content with others Another nice thing about Google Photos is you can actually share images you upload with others, even if they don't have the Google Photos app. All you have to do is select a photo, video, or album, tap on the Share icon in the Google Photos app and type in who you want to share an album with. You can also type in a phone number or email address. Then just type in a personal message if you want and then tap on Send. Shared albums can support up to 20,000 total items. You can also use the Share feature to upload photos and videos to your social networking accounts like Facebook and Twitter. What is Google Photos Assistant? Google Photos Assistant (not to be confused with the Google Assistant AI digital helper) is designed to help you keep track of your library of photos and videos. It will generate cards with suggestions on which photos can be turned into collages, animations, and movies. It will even use machine learning to show you pictures in your account that you might want to delete. Assistant will also offer notifications for alerts like if you are using up your storage space. More Google Photos tips and tricks Here are some more tips and tricks you can try out while using Google Photos: Live Albums is a recent feature addition to Google Photos. Just make an album, select the people you want to see in that album, and Google Photos will automatically put pictures of those folks in that album. Another recent addition lets you see information about a specific photo, such as its date, its file size, and where it was taken, by just swiping up on that photo. You can automatically create movies from your Google Photos pictures by tapping on Assistant, and then select Movie. You have 10 different categories to choose from, including Love Story, Selfie Movie, and Doggie Movie. If you share a photo or album with others in Google Photos, they can now "like" images by tapping on a heart icon. There's a Color Pop feature that lets you keep the color on the subject of an image in Google Photos, but then turns the background into black and white. If you make motion pictures with your Pixel 2 or Pixel 3 phone, Google Photos lets you turn them into GIF images. Conclusion That's just the tip of the iceberg when it comes to using Google Photos. Stay tuned, we will update this post with more information and new features as they release.

Clear

Here search query is “what is google photos”. If we search this query in google. We found the results below.



In the above two pictures is google first and second result. This result is not relevant to our question. Our question was “what is google photos”. But our AI gave a relevant article.

6.6 AI learn by its own

This is the last and six sections in our thesis. In this section our AI learns by own and then is able to answer the question from what AI learns.

The screenshot shows a web interface titled "Learn by own" on a light green background. It features two input fields: the top one is labeled "Topics" with the placeholder text "Give your topics using comma"; the bottom one is labeled "Question" with the placeholder text "Give question....". Between these fields is a purple button labeled "LEARN IT". Below the "Question" field is a blue button labeled "Detect". At the bottom is a black button labeled "Clear". A large, empty dashed rectangular box is positioned below the "Detect" button.

This section looks like the picture above.

First we have to learn our AI. Otherwise AI can't give the answer.

This screenshot shows the same "Learn by own" interface, but with updated content. The "Topics" field now contains the text "what is mask". The "Question" field is empty. The "LEARN IT" button is still purple, but the "Detect" button is now blue. The large dashed rectangular box now contains the text "I didn't learn anything." followed by "So first learn me then ask question." on a new line. The "Clear" button remains at the bottom.

We have to give some topics in the topic box. Then AI takes some time to learn.

The screenshot shows a web interface titled "Learn by own". At the top, there is a "Topics" input box containing the text "what is google photos, covid 19, lose weight". Below this box is a purple button labeled "LEARN IT". Underneath the button, the text "Learning complete. You may ask questions" is displayed in purple. Below this text is a "Question" input box with the placeholder text "Give question....". Below the question box is a blue button labeled "Detect". At the bottom, there is a dashed rectangular box for the answer, which is currently empty, and a black button labeled "Clear".

Now we can ask questions.

This screenshot shows the same "Learn by own" interface, but with a question and answer. The "Topics" box remains the same. The "Question" box now contains the text "what is google photos". The "Detect" button is still present. The dashed answer box now contains the text "a photo sharing and storage service". The "Clear" button is at the bottom.

Learn by own

Topics

what is google photos, covid 19, lose weight

LEARN IT

Learning complete. You may ask questions

Question

what is mask

Detect

Going Out & Daily Life

Clear

AI can only answer the question from a given topic. It doesn't know anything else.

Chapter Seven

Conclusions

Conclusions

Artificial intelligence is being used in almost every sector nowadays. Artificial intelligence has made people's lives much easier. Most of the people use artificial intelligence in their daily life. We hope our AI will help people in their daily life. We make our artificial intelligence free for everyone. We hope that our research will be useful to people and that people will take advantage of it. And in the future, we will create more artificial intelligence that will make people's daily lives easier.