Intelligent Semantic Web Search

Jayaty Jayaty
North Carolina State University
jjayaty

Aishwarya Seth
North Carolina State University
aseth

I. Introduction

A vast majority of information has shifted online, web search has become one of the most important tools these days. However, just the possibility of having access to a humongous amount of data present online is not enough, if you are not able to extract relevant information. The ability to search across multitudes of web pages has advanced a lot but the ability to search intelligently and semantically is still relatively new. Users may not necessarily be well aware of the accurate keywords required to be used for giving them correct results. Therefore, a more lenient approach is required to provide relevant results to the users instead of a strict keyword search approach. In this project, we try to obtain the results of a web search query in a way that is not restricted to a strict keyword search. Thus, we title this type of search as an *Intelligent Semantic Web Search*.

Fig. 1. Keyword based web search

II. Usage Scenario

The intelligent and semantic web search is the next step towards making the user experience more seamless. This application can be integrated wherever a feature for user search is needed. For example, search engines, word meaning lookup on a website. This application can also be modified to integrate intelligence to semantic code search, for finding a specific piece of code in a huge project or code base or even online.

III. Special Features Of Application

The most special feature of the application is what we emphasize as the semantic search. The semantic search feature will empower the user to search what they want without exactly remembering the words or keywords for it. The feature will help in displaying the relevant results even in the absence of precise keywords as well. The feature is relatively recent and unorthodox from traditional tools. Search becomes easier by identifying entities and mapping unstructured data thereby removing the reliance on specific terminologies or precise memory/recall.

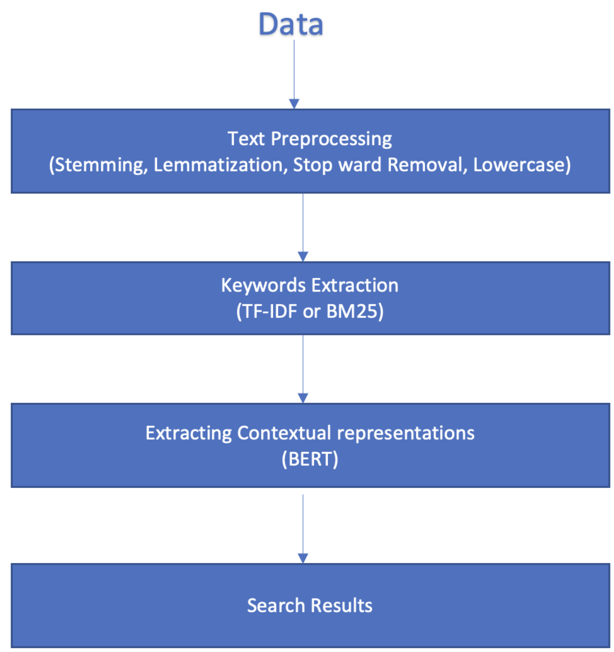


Fig. 2. Intelligent semantic web search

IV. Methodology

Our intelligent semantic web search consists of two phases: Text Pre-processing and Contextual Representation Extraction.

A. Text Pre-processing

The user query needs to be pre-processed first before proceeding with the task of searching. Our text-pre-processing techniques will include converting the query text into lowercase, stemming, lemmatization and finally stop words and punctuations removal(using the list of most common stop words available).

B. Contextual Representation Extraction

We will be extracting contextual representation in order to get the core meaning of the query which user intends to search for. Getting the core meaning helps in getting similar words which might mean the same thing as the query and lead to actual results which the user is searching for. Contextual representation extraction have been achieved by obtaining embeddings and finally extracting similarity using advanced models like BERT. We use a variant of the BERT model called Sentence-BERT or SBERT for this task as it does not require training from scratch and can produce results faster.

V. Implementation

A. Dataset

Datasets of different resources such as GeeksforGeeks, NY Times, Medium and covid-19 aritcles published on Kaggle were used. The original datasets had several columns such as authors, date of publication, title, content. We used only title and content columns for our project.

B. Data Pre-processing

To make our data diverse we gathered data from different resources such as geeks for geeks, NY Times News, Medium and covid-19 articles. We performed preprocessing to refine our datasets.
The datasets were loaded into a Pandas dataframe and following forms of text pre-processing techniques were used to polish the datasets:

- Removing duplicates

```
df_1 = df_1.drop_duplicates()
df_2 = df_2.drop_duplicates()
df_3 = df_3.drop_duplicates()
df_4 = df_4.drop_duplicates()
df_5 = df_5.drop_duplicates()
```

Fig. 3. Text Pre-processing: Removing Duplicate values

- Dropping NAN values

```
df_1 = df_1.loc[:,:].dropna(axis=1)
df_2 = df_2.loc[:,:].dropna(axis=1)
df_3 = df_3.loc[:,:].dropna(axis=1)
df_4 = df_4.loc[:,:].dropna(axis=1)
df_5 = df_5.loc[:,:].dropna(axis=1)
```

Fig. 4. Text Pre-processing: Dropping NAN values

- Removal of unrequited columns such as 'published date', 'author','category','filename'

```
df_5 = df_5.drop(columns = ['category', 'filename'], axis = 1)
df_5.head()
```

Fig. 5. Text Pre-processing: Dropping additional columns

The last step was done on the basis of general observation that the published date, authors, URLs of an article would not lead to any significant improvement in the extraction of semantic similarity. Hence, it was decided to drop such columns from the dataset.

C. Data Augmentation

After pre-processing the datasets, we concatenated 300 rows from all the datasets to form the final dataset. We used this data to train our model.

```
final_data = pd.concat([df_1[:300], df_2[:300], df_3[:300], df_4[:300], df_5[:300]])
```

Fig. 6. Data Augmentation

D. Model

We have used two different approaches to generate results for our semantic search. In the first approach we have used a pre-trained model of SBERT and in the second approach we trained the model on our custom dataset for adapting the pre-trained model to our dataset.

1) Pre-trained Model: We have used a variation of the BERT model, Sentence Bert aka SBERT model, for extracting inferences and semantic meaning from the texts of the articles. The process requires to first generate the word embeddings of the texts. The embeddings are then used for determining the semantic similarity between the articles and the searched query. We have used cosine similarity for the task of determining semantic similarity between the articles and the search query provided by the user.

```
sentences = df['headline_text'].values.tolist()
model = SentenceTransformer('bert-base-nli-mean-tokens')
```

```
sentence_embeddings = model.encode(sentences)
```

Fig. 7. Pre-trained Model execution

2) Trained Model: We have used a similar approach to the original SBERT model for training. We have generated embeddings of the titles and the contents. We have then proceeded to calculate the dot scores of the embeddings. We have used dot scores because it is optimal for asymmteric semantic textual similarity. The dot scores are between the titles and the contents. This has been done to simulate the task of calculating scores between a short search query and the dataset. The dot scores have also been calculated for each title with all contents in order to augment the training data. The final number of rows have been limited to 80000 due to constraint of resources.

```python
# Get embeddings of desired columns and dataframe
def get_embed(df, columns, model):
  embed_df = pd.DataFrame(columns=columns)

  for col in columns:
      sentences = df[col].values.tolist()
      colname = col+"Embed"
      embed_df[colname] = [(model.encode(sentences)) for sentences in df[col]]

  embed_df.drop(columns=columns, axis=1)
  return embed_df
```

```python
embed_df = get_embed(df, ['title', 'content'], model)
embed_df.head()
```

Fig. 8. Trained Model Embeddings Generation

```python
# Calculate similarity scores based on desired metric
def get_similarity(similarity, embed_df):
  sim_df = pd.DataFrame(columns=['titleEmbed', 'contentEmbed', 'dot'])

  i=0
  for title in embed_df['titleEmbed']:
      for content in embed_df['contentEmbed']:
        sim_df.at[i, 'titleEmbed'] = title
        sim_df.at[i, 'contentEmbed'] = content
        if similarity == "dot":
          sim_df.at[i, similarity] = float(util.dot_score([title], [content])[0])
        elif similarity == "cosine":
          sim_df.at[i, similarity] = float(util.cos_sim([title], [content])[0])
        i += 1
        if i > 80000:
          break
  return sim_df
```

```python
dot_df = get_similarity("dot", embed_df)
```

Fig. 9. Trained Model Semantic Score Calculation

```python
# Compiling training dataset
train_examples = []

ctr = 0

for index, row in dot_df.iterrows():
    input = InputExample(texts=[row[0], row[1]], label = row[2]) # texts = [title, content], label = (dot/cosine)score
    if ctr%2 == 0:
        train_examples.append(input)
    ctr += 1
```

```python
train_dataset = SentencesDataset(train_examples, model)
train_dataloader = DataLoader(train_examples, shuffle=True, batch_size=16)
train_loss = losses.CosineSimilarityLoss(model)


#Tune the model
model.fit(train_objectives=[(train_dataloader, train_loss)], epochs=1, warmup_steps=100, output_path=model_save_path)
```

Fig. 10. Trained Model Training

VI. Results

A. Processed Data

After executing the data pre-processing steps mentioned above, the resultant dataset is as shown below. The SBERT model chosen to generate embeddings and contextual representations and the model training is executed on this dataset.

```python
final_data[1:100]
```

	title	content
1	if you've tasted 21/52 of these international ...	food I if you've tasted 21/52 of these interna...
2	16 twenties vs. thirties tweets that are so ac...	16 twenties vs. thirties tweets that are so ac...
3	the 19 most tone-deaf things celebrities have ...	celebrity I the 19 most tone-deaf things celeb...
4	if you're bored, try matching these disney pri...	tv and movies I if you're bored, try matching ...
5	kendall jenner responded to a fan who suggeste...	kendall jenner responded to a fan who suggeste...

Fig. 11. Processed Data

B. Embedding Vectors generated by SBERT model

After running the pre-trained SBERT model, the embedding vectors generated from the articles are as shown below in the snapshot.

```python
print('\n Sample BERT embedding vector - length', len(sentence_embeddings[0]))

print('\n Sample BERT embedding vector - note includes negative values \n', sentence_embeddings[0])
```

```
Sample BERT embedding vector - length 768

Sample BERT embedding vector - note includes negative values
[ 4.39588100e-01  6.60045862e-01  1.78375101e+00 -4.98062581e-01
-1.98180834e-03  2.65921533e-01  1.14297855e+00 -2.31351927e-01
 1.77076697e-01  1.98336288e-01 -8.13299596e-01  7.71076024e-01
 1.68776929e-01  8.83808970e-01  1.51478261e-01  5.27982950e-01
-3.49639416e-01  1.78840458e-02  9.56008434e-02 -3.90725583e-01
-4.73431438e-01  3.07736814e-01  4.05728012e-01  4.15130258e-01
 7.11948395e-01  1.08209401e-01  7.34861612e-01 -2.65426457e-01
```

Fig. 12. Embedding Vectors generated from pre-trained model

C. Output

To compare both the models we input six search queries to each model to find the most semantically related articles using the cosine similarity score. The outputs after searching the respective queries in each model have been shown below along with the respective queries and the snapshots:

```
Semantic Search Results in pre-trained models

----------------

Query: Employees are upset

Top 8 most similar news headlines:

Kristen Bell And Dax Shepard Said They're "At Each Other's Throats" In Isolation, And Things Got Awk (Cosine Score: 0.5016)
Abnormalities of serum and plasma components in patients with multiple sclerosis Qualitative and qua (Cosine Score: 0.4731)
Kim Kardashian Dragged Kendall And Kourtney's Lack Of Work Ethic In A Comment That Caused Their Huge (Cosine Score: 0.4572)
The 'ticking budget' facing the US The budget proposals laid out by the administration of US Preside (Cosine Score: 0.4519)
Blockchain is not only crappy technology but a bad vision for the futureBlockchain is not only crapp (Cosine Score: 0.4391)
Blockchain is not only crappy technology but a bad vision for the futureBlockchain is not only crapp (Cosine Score: 0.4391)
Alterations in Pulmonary Function Following Respiratory Viral Infection Respiratory viral illness is (Cosine Score: 0.4188)
18 People Who Are Having A Way, Wayyyy Worse Time Stuck Indoors Than You18 People Who Are Having A W (Cosine Score: 0.4171)
```

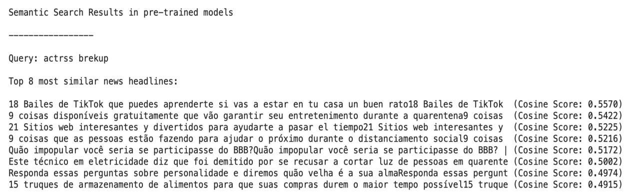
Fig. 13. Pre-trained model output of Query:"Employees are upset"

```
Semantic Search Results in pre-trained models

----------------

Query: difficult time

Top 8 most similar news headlines:

Kristen Bell And Dax Shepard Said They're "At Each Other's Throats" In Isolation, And Things Got Awk (Cosine Score: 0.4421)
C Quiz ? 111 | Question 4Pick the best statement for the below program:                              (Cosine Score: 0.3851)
Segregating negative and positive maintaining order and O(1) spaceSegregation of negative and positi (Cosine Score: 0.3563)
Blockchain is not only crappy technology but a bad vision for the futureBlockchain is not only crapp (Cosine Score: 0.3526)
Blockchain is not only crappy technology but a bad vision for the futureBlockchain is not only crapp (Cosine Score: 0.3526)
MP de Bolsonaro que suspende contrato de trabalho por 4 meses é inconstitucional, dizem juristasMP d (Cosine Score: 0.3524)
Modular Exponentiation in PythonGiven three numbers x, y and p, compute (x^y) % pExamples: To show s (Cosine Score: 0.3502)
Abnormalities of serum and plasma components in patients with multiple sclerosis Qualitative and qua (Cosine Score: 0.3413)
```

Fig. 14. Pre-trained model output of Query:"Difficult time"

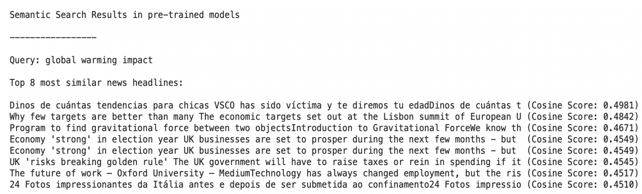
```
Semantic Search Results in pre-trained models

----------------

Query: crickt sprts

Top 8 most similar news headlines:

Este técnico em eletricidade diz que foi demitido por se recusar a cortar luz de pessoas em quarente (Cosine Score: 0.4973
lineto() function in CThe header file graphics.h contains lineto() function which draws a line from  (Cosine Score: 0.4554
9 coisas que as pessoas estão fazendo para ajudar o próximo durante o distanciamento social9 coisas  (Cosine Score: 0.4437
9 coisas disponíveis gratuitamente que vão garantir seu entretenimento durante a quarentena9 coisas  (Cosine Score: 0.4433
18 Bailes de TikTok que puedes aprenderte si vas a estar en tu casa un buen rato18 Bailes de TikTok  (Cosine Score: 0.4351
Estas são as músicas mais injustiçadas da Demi LovatoEstas são as músicas mais injustiçadas da Demi  (Cosine Score: 0.4349
Quão impopular você seria se participasse do BBB?Quão impopular você seria se participasse do BBB? | (Cosine Score: 0.4346
Le macrophage alveolaire de porc: Revue bibliographique Résumé Après une présentation des techniques (Cosine Score: 0.4313
```

Fig. 15. Pre-trained model output of Query:"crickt sprts"



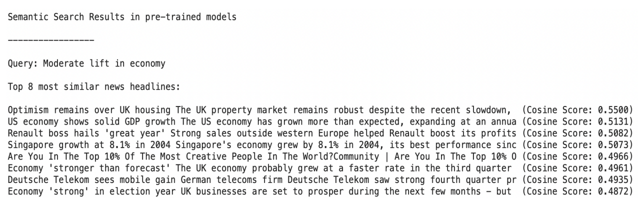
```
Semantic Search Results in pre-trained models

-----------------

Query: actrss brekup

Top 8 most similar news headlines:

18 Bailes de TikTok que puedes aprenderte si vas a estar en tu casa un buen rato18 Bailes de TikTok  (Cosine Score: 0.5570)
9 coisas disponíveis gratuitamente que vão garantir seu entretenimento durante a quarentena9 coisas  (Cosine Score: 0.5422)
21 Sitios web interesantes y divertidos para ayudarte a pasar el tiempo21 Sitios web interesantes y  (Cosine Score: 0.5225)
9 coisas que as pessoas estão fazendo para ajudar o próximo durante o distanciamento social9 coisas  (Cosine Score: 0.5216)
Quão impopular você seria se participasse do BBB?Quão impopular você seria se participasse do BBB? | (Cosine Score: 0.5172)
Este técnico em eletricidade diz que foi demitido por se recusar a cortar luz de pessoas em quarente (Cosine Score: 0.5002)
Responda essas perguntas sobre personalidade e diremos quão velha é a sua almaResponda essas pergunt (Cosine Score: 0.4974)
15 truques de armazenamento de alimentos para que suas compras durem o maior tempo possível15 truque (Cosine Score: 0.4915)
```

Fig. 16. Pre-trained model output of Query:"actrss brkup"



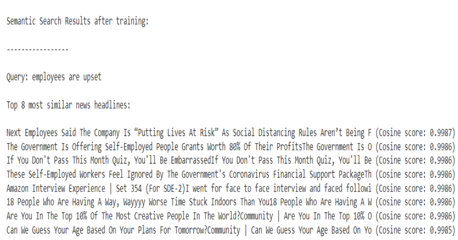
```
Semantic Search Results in pre-trained models

-----------------

Query: global warming impact

Top 8 most similar news headlines:

Dinos de cuántas tendencias para chicas VSCO has sido víctima y te diremos tu edadDinos de cuántas t (Cosine Score: 0.4981)
Why few targets are better than many The economic targets set out at the Lisbon summit of European U (Cosine Score: 0.4842)
Program to find gravitational force between two objectsIntroduction to Gravitational ForceWe know th (Cosine Score: 0.4671)
Economy 'strong' in election year UK businesses are set to prosper during the next few months - but  (Cosine Score: 0.4549)
Economy 'strong' in election year UK businesses are set to prosper during the next few months - but  (Cosine Score: 0.4549)
UK 'risks breaking golden rule' The UK government will have to raise taxes or rein in spending if it (Cosine Score: 0.4545)
The future of work — Oxford University — MediumTechnology has always changed employment, but the ris (Cosine Score: 0.4517)
24 Fotos impressionantes da Itália antes e depois de ser submetida ao confinamento24 Fotos impressio (Cosine Score: 0.4510)
```

Fig. 17. Pre-trained model output of Query:"global warming impact"



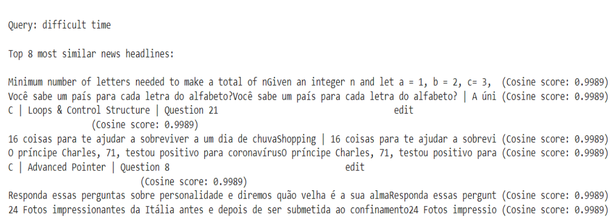
```
Semantic Search Results in pre-trained models

-----------------

Query: Moderate lift in economy

Top 8 most similar news headlines:

Optimism remains over UK housing The UK property market remains robust despite the recent slowdown,  (Cosine Score: 0.5500)
US economy shows solid GDP growth The US economy has grown more than expected, expanding at an annua (Cosine Score: 0.5131)
Renault boss hails 'great year' Strong sales outside western Europe helped Renault boost its profits (Cosine Score: 0.5082)
Singapore growth at 8.1% in 2004 Singapore's economy grew by 8.1% in 2004, its best performance sinc (Cosine Score: 0.5073)
Are You In The Top 10% Of The Most Creative People In The World?Community | Are You In The Top 10% O (Cosine Score: 0.4966)
Economy 'stronger than forecast' The UK economy probably grew at a faster rate in the third quarter  (Cosine Score: 0.4961)
Deutsche Telekom sees mobile gain German telecoms firm Deutsche Telekom saw strong fourth quarter pr (Cosine Score: 0.4935)
Economy 'strong' in election year UK businesses are set to prosper during the next few months - but  (Cosine Score: 0.4872)
```

Fig. 18. Pre-trained model output of Query:"Moderate impact on economy"



```
Semantic Search Results after training:

-----------------

Query: employees are upset

Top 8 most similar news headlines:

Next Employees Said The Company Is "Putting Lives At Risk" As Social Distancing Rules Aren't Being F (Cosine score: 0.9987)
The Government Is Offering Self-Employed People Grants Worth 80% Of Their ProfitsThe Government Is O (Cosine score: 0.9986)
If You Don't Pass This Month Quiz, You'll Be EmbarrassedIf You Don't Pass This Month Quiz, You'll Be (Cosine score: 0.9986)
These Self-Employed Workers Feel Ignored By The Government's Coronavirus Financial Support PackageTh (Cosine score: 0.9986)
Amazon Interview Experience | Set 354 (For SDE-2)I went for face to face interview and faced followi (Cosine score: 0.9986)
18 People Who Are Having A Way, Wayyyy Worse Time Stuck Indoors Than You18 People Who Are Having A W (Cosine score: 0.9986)
Are You In The Top 10% Of The Most Creative People In The World?Community | Are You In The Top 10% O (Cosine score: 0.9986)
Can We Guess Your Age Based On Your Plans For Tomorrow?Community | Can We Guess Your Age Based On Yo (Cosine score: 0.9985)
```

Fig. 19. Trained model output of Query:"Employees are upset"

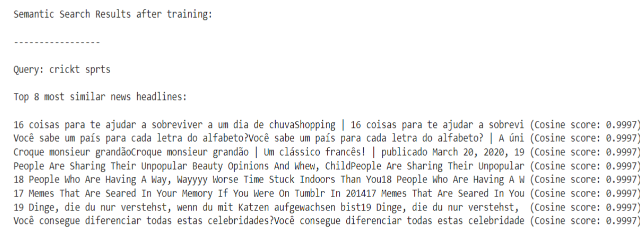


```
Query: difficult time

Top 8 most similar news headlines:

Minimum number of letters needed to make a total of nGiven an integer n and let a = 1, b = 2, c= 3,  (Cosine score: 0.9989)
Você sabe um país para cada letra do alfabeto?Você sabe um país para cada letra do alfabeto? | A úni (Cosine score: 0.9989)
C | Loops & Control Structure | Question 21                                    edit
                 (Cosine score: 0.9989)
16 coisas para te ajudar a sobreviver a um dia de chuvaShopping | 16 coisas para te ajudar a sobrevi (Cosine score: 0.9989)
O príncipe Charles, 71, testou positivo para coronavírusO príncipe Charles, 71, testou positivo para (Cosine score: 0.9989)
C | Advanced Pointer | Question 8                                    edit
                            (Cosine score: 0.9989)
Responda essas perguntas sobre personalidade e diremos quão velha é a sua almaResponda essas pergunt (Cosine score: 0.9989)
24 Fotos impressionantes da Itália antes e depois de ser submetida ao confinamento24 Fotos impressio (Cosine score: 0.9989)
```

Fig. 20. Trained model output of Query:"Difficult time"

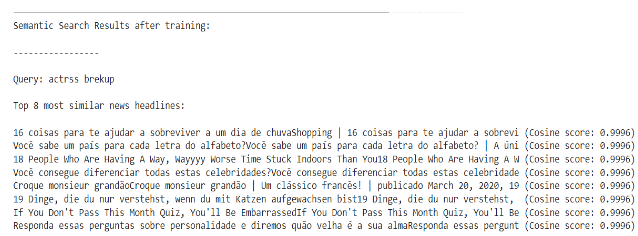
VII. Conclusion

We have compared the performances of SBERT model with our custom trained SBERT model for semantic web search. While our custom trained SBERT model retrieves human perceivable better results for most of the input search queries. There is a comparable performance when a search query consists of any typos. In case there exists any typos in the search query, both models returned results of languages other than english. Thus bringing a shortcoming in sight in the pre-trained model, SBERT.



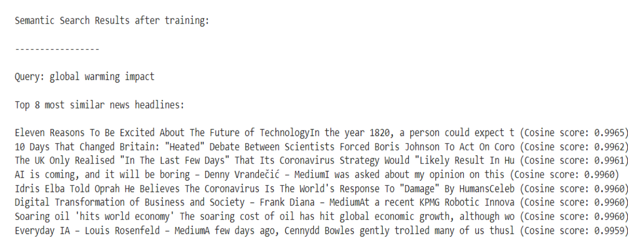
```
Semantic Search Results after training:

-----------------

Query: crickt sprts

Top 8 most similar news headlines:

16 coisas para te ajudar a sobreviver a um dia de chuvaShopping | 16 coisas para te ajudar a sobrevi (Cosine score: 0.9997)
Você sabe um país para cada letra do alfabeto?Você sabe um país para cada letra do alfabeto? | A úni (Cosine score: 0.9997)
Croque monsieur grandãoCroque monsieur grandão | Um clássico francês! | publicado March 20, 2020, 19 (Cosine score: 0.9997)
People Are Sharing Their Unpopular Beauty Opinions And Whew, ChildPeople Are Sharing Their Unpopular (Cosine score: 0.9997)
18 People Who Are Having A Way, Wayyyy Worse Time Stuck Indoors Than You18 People Who Are Having A W (Cosine score: 0.9997)
17 Memes That Are Seared In Your Memory If You Were On Tumblr In 201417 Memes That Are Seared In You (Cosine score: 0.9997)
19 Dinge, die du nur verstehst, wenn du mit Katzen aufgewachsen bist19 Dinge, die du nur verstehst,  (Cosine score: 0.9997)
Você consegue diferenciar todas estas celebridades?Você consegue diferenciar todas estas celebridade (Cosine score: 0.9997)
```

Fig. 21. Trained model output of Query:"crickt sprts"



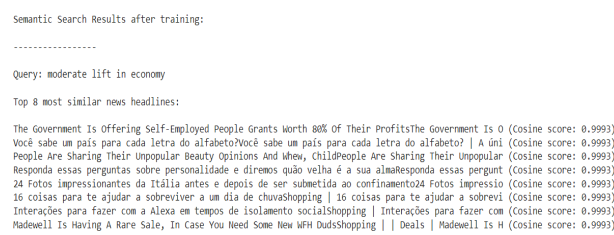
```
Semantic Search Results after training:

-----------------

Query: actrss brekup

Top 8 most similar news headlines:

16 coisas para te ajudar a sobreviver a um dia de chuvaShopping | 16 coisas para te ajudar a sobrevi (Cosine score: 0.9996)
Você sabe um país para cada letra do alfabeto?Você sabe um país para cada letra do alfabeto? | A úni (Cosine score: 0.9996)
18 People Who Are Having A Way, Wayyyy Worse Time Stuck Indoors Than You18 People Who Are Having A W (Cosine score: 0.9996)
Você consegue diferenciar todas estas celebridades?Você consegue diferenciar todas estas celebridade (Cosine score: 0.9996)
Croque monsieur grandãoCroque monsieur grandão | Um clássico francês! | publicado March 20, 2020, 19 (Cosine score: 0.9996)
19 Dinge, die du nur verstehst, wenn du mit Katzen aufgewachsen bist19 Dinge, die du nur verstehst,  (Cosine score: 0.9996)
If You Don't Pass This Month Quiz, You'll Be EmbarrassedIf You Don't Pass This Month Quiz, You'll Be (Cosine score: 0.9996)
Responda essas perguntas sobre personalidade e diremos quão velha é a sua almaResponda essas pergunt (Cosine score: 0.9996)
```

Fig. 22. Trained model output of Query:"actrss brkup"



```
Semantic Search Results after training:

-----------------

Query: global warming impact

Top 8 most similar news headlines:

Eleven Reasons To Be Excited About The Future of TechnologyIn the year 1820, a person could expect t (Cosine score: 0.9965)
10 Days That Changed Britain: "Heated" Debate Between Scientists Forced Boris Johnson To Act On Coro (Cosine score: 0.9962)
The UK Only Realised "In The Last Few Days" That Its Coronavirus Strategy Would "Likely Result In Hu (Cosine score: 0.9961)
AI is coming, and it will be boring - Denny Vrandečić - MediumI was asked about my opinion on this (Cosine score: 0.9960)
Idris Elba Told Oprah He Believes The Coronavirus Is The World's Response To "Damage" By HumansCeleb (Cosine score: 0.9960)
Digital Transformation of Business and Society - Frank Diana - MediumAt a recent KPMG Robotic Innova (Cosine score: 0.9960)
Soaring oil 'hits world economy' The soaring cost of oil has hit global economic growth, although wo (Cosine score: 0.9960)
Everyday IA - Louis Rosenfeld - MediumA few days ago, Cennydd Bowles gently trolled many of us thusl (Cosine score: 0.9959)
```

Fig. 23. Trained model output of Query:"global warming impact"



```
Semantic Search Results after training:

-----------------

Query: moderate lift in economy

Top 8 most similar news headlines:

The Government Is Offering Self-Employed People Grants Worth 80% Of Their ProfitsThe Government Is O (Cosine score: 0.9993)
Você sabe um país para cada letra do alfabeto?Você sabe um país para cada letra do alfabeto? | A úni (Cosine score: 0.9993)
People Are Sharing Their Unpopular Beauty Opinions And Whew, ChildPeople Are Sharing Their Unpopular (Cosine score: 0.9993)
Responda essas perguntas sobre personalidade e diremos quão velha é a sua almaResponda essas pergunt (Cosine score: 0.9993)
24 Fotos impressionantes da Itália antes e depois de ser submetida ao confinamento24 Fotos impressio (Cosine score: 0.9993)
16 coisas para te ajudar a sobreviver a um dia de chuvaShopping | 16 coisas para te ajudar a sobrevi (Cosine score: 0.9993)
Interações para fazer com a Alexa em tempos de isolamento socialShopping | Interações para fazer com (Cosine score: 0.9993)
Madewell Is Having A Rare Sale, In Case You Need Some New WFH DudsShopping | | Deals | Madewell Is H (Cosine score: 0.9993)
```

Fig. 24. Trained model output of Query:"Moderate impact on economy"

References

[1] Jui-Ting Huang, Ashish Sharma, Shuying Sun, Li Xia, David Zhang, Philip Pronin, Janani Padmanabhan, Giuseppe Ottaviano, Linjun Yan, "Embedding-based Retrieval in Facebook Search", CoRR abs/2006.11632, 2020 https://arxiv.org/abs/2006.11632

[2] Pascal Hitzler, "A Review of the Semantic Web Field", Communications of the ACM Volume 64Issue 2February 2021 pp 76–83 https://doi.org/10.1145/3397512

[3] Hamel Husain, "How To Create Natural Language Semantic Search For Arbitrary Objects With Deep Learning", https://towardsdatascience.com/semantic-code-search-3cd6d244a39c

[4] Yusuf Sermet, Ibrahim Demir, "A Semantic Web Framework for Automated Smart Assistants: COVID-19 Case Study", arXiv:2007.00747, https://arxiv.org/abs/2007.00747

[5] "A decade of Semantic Web research through the lenses of a mixed methods approach", Semantic Web, vol. 11, no. 6, pp. 979-1005, 2020, https://content.iospress.com/articles/semantic-web/sw200371

[6] "The Semantic Web identity crisis: In search of the trivialities that never were", Semantic Web, vol. 11, no. 1, pp. 19-27, 2020, https://content.iospress.com/articles/semantic-web/sw190372
[7] "Information extraction meets the Semantic Web: A survey", Semantic Web, vol. 11, no. 2, pp. 255-335, 2020, https://content.iospress.com/articles/semantic-web/sw180333
[8] Nils Reimers, Iryna Gurevych,"Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks", Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, pages 3982–3992, Hong Kong, China, November 3–7, 2019, https://aclanthology.org/D19-1410.pdf
[9] Kexin Wang, Nils Reimers, Iryna Gurevych, "TSDAE: Using Transformer-based Sequential Denoising Auto-Encoder for Unsupervised Sentence Embedding Learning",arXiv:2104.06979v3 [cs.CL] 10 Sep 2021, https://arxiv.org/pdf/2104.06979.pdf
[10] Nils Reimers and Iryna Gurevych, "Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation"", arXiv:2004.09813v2 [cs.CL] 5 Oct 2020, https://arxiv.org/pdf/2004.09813.pdf
[11] https://www.kaggle.com/parsonsandrew1/nytimes-article-lead-paragraphs-18512017
[12] https://www.kaggle.com/hsankesara/medium-articles
[13] https://www.kaggle.com/kashnitsky/covid19-articles-by-elsevier?select=covidartilceselseviertrain.csv
[14] https://www.kaggle.com/naidukarthi2193/geeks-for-geeks-articles-dataset
[15] https://www.kaggle.com/promptcloud/articles-from-buzzfeed-2020