

INFORMATION RETRIEVAL

OUTPUTS:

```
Enter your search query (or type 'exit' to quit): tech
Ranked Documents:
Document ID: canva, Score: 0.0743
Document ID: Ola, Score: 0.0697
Document ID: yahoo, Score: 0.0597
Document ID: skype, Score: 0.0574
-----
Enter your search query (or type 'exit' to quit): information
Ranked Documents:
Document ID: nokia, Score: 0.0772
Document ID: messenger, Score: 0.0764
Document ID: bing, Score: 0.0747
Document ID: zomato, Score: 0.0673
Document ID: samsung, Score: 0.0606
Document ID: whatsapp, Score: 0.0606
Document ID: Dell, Score: 0.0602
Document ID: yahoo, Score: 0.0597
Document ID: spotify, Score: 0.0588
Document ID: telegram, Score: 0.0542
```

When we enter a query like **"tech"** or **"information"**, the system returns documents ranked by relevance score. For example, for the query **"tech"**, the top document is **"canva"** with a score of **0.0743**, meaning it is the most relevant. Other documents like **"Ola"** (0.0697) and **"yahoo"** (0.0597) follow in decreasing order of relevance.

Similarly, for **"information"**, **"nokia"** scores highest at **0.0772**, followed by **"messenger"** at **0.0764**, showing their strong match to the query.

Higher scores indicate stronger relevance. The system displays up to 10 documents sorted by score and then by document ID.

```
Enter your search query (or type 'exit' to quit): Developing your Zomato business account and profile is a great way to boost your restaurant's
Ranked Documents:
Document ID: zomato, Score: 0.2147
Document ID: swiggy, Score: 0.1311
Document ID: instagram, Score: 0.0605
Document ID: messenger, Score: 0.0592
Document ID: youtube, Score: 0.0584
Document ID: Discord, Score: 0.0534
Document ID: bing, Score: 0.0518
Document ID: paypal, Score: 0.0471
Document ID: reddit, Score: 0.0441
Document ID: flipkart, Score: 0.0407
```

```
Enter your search query (or type 'exit' to quit): Warwickshire, came from an ancient family and was the heiress to some land
Ranked Documents:
Document ID: shakespeare, Score: 0.1200
Document ID: levis, Score: 0.0241
Document ID: Adobe, Score: 0.0227
Document ID: google, Score: 0.0207
Document ID: nike, Score: 0.0192
Document ID: zomato, Score: 0.0177
Document ID: huawei, Score: 0.0137
Document ID: skype, Score: 0.0117
Document ID: blackberry, Score: 0.0109
Document ID: Dell, Score: 0.0108
-----
Enter your search query (or type 'exit' to quit): Sodexo
No matching documents found.
```

```
Enter your search query (or type 'exit' to quit): Information Retrieval
Ranked Documents:
Document ID: nokia, Score: 0.0772
Document ID: messenger, Score: 0.0764
Document ID: bing, Score: 0.0747
Document ID: zomato, Score: 0.0673
Document ID: samsung, Score: 0.0606
Document ID: whatsapp, Score: 0.0606
Document ID: Dell, Score: 0.0602
Document ID: yahoo, Score: 0.0597
Document ID: spotify, Score: 0.0588
Document ID: telegram, Score: 0.0542
-----
Enter your search query (or type 'exit' to quit): Narendra Modi
No matching documents found.
-----
Enter your search query (or type 'exit' to quit): exit
```

NOVELTY:

#Highlight Query Matches

```
Document ID: microsoft
Snippet: court overturn breakup order still found compani guilti illeg tri maintain monopoli compani legal woe continu 2004 european union eu **le
-----
Document ID: levis
Snippet: **levis** walter haa born may 11 1889 san francisco calif u die dec 7 1979 san francisco american busi execut credit ... born may 11 1889
-----
```

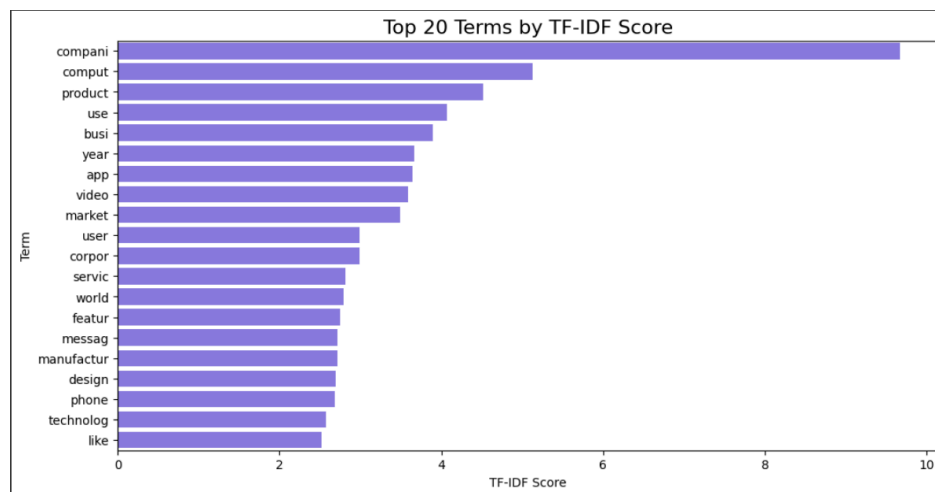
- **Document ID** → which document matched.
- **Snippet** → a preview showing your query terms in context, highlighted.
- You can now **see where and how your query term appears** without reading the whole document.

#Synonym Extraction

```
Original Query: food
Expanded Query: ['solid food', 'nutrient', 'intellectual nourishment', 'food']
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data] Package wordnet is already up-to-date!
```

- For each word in the query, WordNet finds **synonyms**.
- Returns a **list of original + synonym words** for your search

#TF-IDF Vectorization



This visualization displays the top 20 terms in the document corpus ranked by their TF-IDF scores. The TF-IDF (Term Frequency-Inverse Document Frequency) metric highlights words that are most important or distinctive across the collection, helping to reveal the key topics and themes.