

Part 3: Ranking

1. Creating scores

1.1) TF-IDF + cosine similarity

TF-IDF (Term Frequency-Inverse Document Frequency) is a well-established technique in information retrieval and text analysis. In this method, we calculate a numerical value for each term in a document to represent its importance in that document. The importance of a term is determined by both its frequency in the document (TF) and its rarity across the entire document collection (IDF). The cosine similarity is then used to measure the similarity between a query and documents based on their TF-IDF scores.

Advantages	Disadvantages
TF-IDF is widely used and understood in the field of information retrieval.	TF-IDF may not fully capture the nuanced aspects of document relevance, such as context or semantics.
It provides a numerical score that is easy to interpret, with higher scores indicating more relevance.	It can be sensitive to the choice of preprocessing techniques (e.g., stop-word removal, stemming, etc.).
It takes into account the importance of individual terms in a document and their rarity in the corpus, which can help in distinguishing relevant documents.	TF-IDF does not consider external factors like social media popularity, which can be essential for some applications.

These are the results obtained in each of the queries:

Query	Returned as relevant documents
Russian intervention Ukraine	<p>doc_id= doc_1984 - url_tweet= https://twitter.com/InGlore070/status/1575597493508710401</p> <p>doc_id= doc_3216 - url_tweet= https://twitter.com/InGlore070/status/1575271464021987332</p> <p>doc_id= doc_347 - url_tweet= https://twitter.com/infussambas/status/1575872044075724802</p> <p>doc_id= doc_3530 - url_tweet= https://twitter.com/ActForUA/status/1575207467390488576</p> <p>doc_id= doc_3944 - url_tweet= https://twitter.com/danish_greek/status/1575160028402954240</p> <p>doc_id= doc_934 - url_tweet= https://twitter.com/gulftoday/status/1575809946251239424</p> <p>doc_id= doc_3027 - url_tweet= https://twitter.com/infussambas/status/1575339603212771328</p>

	<p>doc_id= doc_2770 - url_tweet= https://twitter.com/infussambas/status/1575409000925261824</p> <p>doc_id= doc_1334 - url_tweet= https://twitter.com/infussambas/status/1575754204249194497</p> <p>doc_id= doc_1404 - url_tweet= https://twitter.com/infussambas/status/1575734900178956288</p>
NATO Ukraine	<p>doc_id= doc_54 - url_tweet= https://twitter.com/TrimiA98/status/1575911455806742528</p> <p>doc_id= doc_474 - url_tweet= https://twitter.com/_Nex3_/status/1575851734974578688</p> <p>doc_id= doc_800 - url_tweet= https://twitter.com/Ak1nc1Turk1071/status/1575819273078906881</p> <p>doc_id= doc_464 - url_tweet= https://twitter.com/_Nex3_/status/1575853653310976001</p> <p>doc_id= doc_324 - url_tweet= https://twitter.com/_Nex3_/status/1575879420300390403</p> <p>doc_id= doc_83 - url_tweet= https://twitter.com/elevotv/status/1575909051380432896</p> <p>doc_id= doc_3489 - url_tweet= https://twitter.com/GabeZZOZZ/status/1575213861359480833</p> <p>doc_id= doc_183 - url_tweet= https://twitter.com/NewsZapper/status/1575899185202724868</p> <p>doc_id= doc_409 - url_tweet= https://twitter.com/SGNewsAlerts/status/1575861833025478656</p> <p>doc_id= doc_100 - url_tweet= https://twitter.com/elevotv/status/1575908043891437569</p>
Dnipro war	<p>doc_id= doc_3292 - url_tweet= https://twitter.com/firamnews/status/1575250163882463251</p> <p>doc_id= doc_2066 - url_tweet= https://twitter.com/Chronology22/status/1575575566186852352</p> <p>doc_id= doc_1993 - url_tweet= https://twitter.com/Chronology22/status/1575595404044865536</p> <p>doc_id= doc_345 - url_tweet= https://twitter.com/Chronology22/status/1575873706668212224</p> <p>doc_id= doc_2000 - url_tweet= https://twitter.com/Chronology22/status/1575594105035378688</p> <p>doc_id= doc_2434 - url_tweet=</p>

	<p>https://twitter.com/Chronology22/status/1575479438141280256</p> <p>doc_id= doc_2965 - url_tweet= https://twitter.com/Ukrinform_News/status/1575357208996478976</p> <p>doc_id= doc_2785 - url_tweet= https://twitter.com/Ukrinform_News/status/1575405596337135617</p> <p>doc_id= doc_1881 - url_tweet= https://twitter.com/Chronology22/status/1575617614394449920</p> <p>doc_id= doc_978 - url_tweet= https://twitter.com/KVladyslavka/status/1575805179730882561</p>
Zelenski leadership Ukraine war	<p>doc_id= doc_1283 - url_tweet= https://twitter.com/swapnilzs/status/1575761181674856448</p> <p>doc_id= doc_557 - url_tweet= https://twitter.com/have_fun_X/status/1575839188741951489</p> <p>doc_id= doc_71 - url_tweet= https://twitter.com/PMsincel994/status/1575909819810484224</p> <p>doc_id= doc_417 - url_tweet= https://twitter.com/Chronology22/status/1575860740556460032</p> <p>doc_id= doc_2784 - url_tweet= https://twitter.com/AlbaneseAl/status/1575405656772935681</p> <p>doc_id= doc_2782 - url_tweet= https://twitter.com/AGCsegreteria/status/1575575405725936934912</p> <p>doc_id= doc_2781 - url_tweet= https://twitter.com/pka71/status/1575405776260268032</p> <p>doc_id= doc_1162 - url_tweet= https://twitter.com/epirus_son/status/1575787462060052480</p> <p>doc_id= doc_1803 - url_tweet= https://twitter.com/monitor_ukraine/status/15751575633511557435393</p> <p>doc_id= doc_3879 - url_tweet= https://twitter.com/djarman/status/1575169420661760000</p>
Stop Putin's aggression Ukraine	<p>doc_id= doc_2157 - url_tweet= https://twitter.com/2536luis/status/1575548179218472960</p> <p>doc_id= doc_142 - url_tweet= https://twitter.com/GenevaSummit/status/1575904086448173057</p> <p>doc_id= doc_1344 - url_tweet= https://twitter.com/Viper_Toffi/status/1575751722017779713</p> <p>doc_id= doc_3941 - url_tweet= https://twitter.com/Txxani/status/15751604</p>

	<p><u>66980352000</u></p> <p>doc_id= doc_3076 - url_tweet= https://twitter.com/rybanova171195/status/1575321074627215364</p> <p>doc_id= doc_123 - url_tweet= https://twitter.com/Truecubbyblue15/status/1575906036678291456</p> <p>doc_id= doc_964 - url_tweet= https://twitter.com/bowen_dilwyn/status/1575806511317913600</p> <p>doc_id= doc_897 - url_tweet= https://twitter.com/UATV_en/status/1575812703070715905</p> <p>doc_id= doc_1209 - url_tweet= https://twitter.com/ArmchairOpinio1/status/1575778970389598209</p> <p>doc_id= doc_1017 - url_tweet= https://twitter.com/An_ObserverX/status/1575800773669257216</p>
--	--

1.2) Our-Score + cosine similarity

We wanted to improve the TF-IDF (Term Frequency-Inverse Document Frequency) model, so we came up with a cool idea. We thought about using retweets and likes on Twitter as important factors. Our thinking was that in social media, the number of retweets and likes a tweet gets can show how important and influential it is.

To make this work, we created a strict system to work out how much these factors should count in our model. We decided that tweets with more retweets and likes should be considered more important. We gave 30% importance to retweets and 15% importance to likes. We did this because retweeting a tweet shows a stronger endorsement than just liking it, which makes the tweet more relevant in a given context. We also wanted to give the highest weight to tf-idf because it's a more reliable model, and we didn't want the impact of likes/retweets to be enormous. So the intention was to mix TF-IDF and this measure.

To do this, we needed to normalise the likes and retweets because, as we showed in part 1, they had a very skewed distribution with many outliers. To do this, we applied a logarithmic scale (+1 to avoid zeros) to likes and retweets and then did a min-max normalisation to put these values from 0 to 1 once they were evenly distributed.

By giving more weight to retweets, we're recognising that they show a deeper level of engagement and therefore more relevance. This update to our TF-IDF model keeps pace with the fast-changing world of social media and helps us better understand the importance of tweets in the digital world.

Advantages	Disadvantages
By customising the TF-IDF model with retweets and likes, we better align with the particular requirements of social media	The decision to allocate 30% importance to retweets and 15% to likes is subjective and may not universally apply to all

content ranking.	contexts. Different users and applications might have different perspectives on the importance of these engagement metrics.
By incorporating retweets and likes as factors in the ranking model, we are likely to improve the relevance of search results, especially in the context of social media.	Depending heavily on external engagement metrics like retweets and likes can make the model vulnerable to changes in the social media platform's behaviour and limited to a specific platform (Twitter).
The model prioritises content that has gained popularity on the social media platform. Popularity often correlates with relevance, as content that resonates with a broad audience is more likely to address common interests or topics.	While retweets and likes are valuable signals, they do not capture all aspects of content relevance (complex content, sarcasm, irony...).

The results of using our score algorithm and then apply cosine similarity is the following:

Query	Returned as relevant documents
Russian intervention Ukraine	<p>doc_id= doc_2770 - url_tweet= https://twitter.com/infussambas/status/1575409000925261824 - retweet= 22 - likes=49</p> <p>doc_id= doc_2526 - url_tweet= https://twitter.com/infussambas/status/1575465229756301312 - retweet= 7 - likes=19</p> <p>doc_id= doc_2777 - url_tweet= https://twitter.com/infussambas/status/1575406964078956544 - retweet= 36 - likes=226</p> <p>doc_id= doc_1954 - url_tweet= https://twitter.com/Militarylandnet/status/1575605153704837120 - retweet= 63 - likes=375</p> <p>doc_id= doc_347 - url_tweet= https://twitter.com/infussambas/status/1575872044075724802 - retweet= 2 - likes=12</p> <p>doc_id= doc_2876 - url_tweet= https://twitter.com/UATV_en/status/1575375069206888448 - retweet= 21 - likes=37</p> <p>doc_id= doc_2295 - url_tweet= https://twitter.com/infussambas/status/1575515970860650497 - retweet= 4 - likes=23</p> <p>doc_id= doc_440 - url_tweet= https://twitter.com/infussambas/status/1575857540268982273 - retweet= 6 - likes=21</p> <p>doc_id= doc_934 - url_tweet= https://twitter.com/gulftoday/status/1575809946251239424 - retweet= 2 - likes=7</p> <p>doc_id= doc_1328 - url_tweet= https://twitter.com/infussambas/status/1575755533650300929 - retweet= 5 - likes=25</p>

<p>NATO Ukraine</p>	<p>doc_id= doc_435 - url_tweet= https://twitter.com/I_Katchanovski/status/1575858386893099009 - retweet= 17 - likes=33</p> <p>doc_id= doc_464 - url_tweet= https://twitter.com/_Nex3_/status/1575853653310976001 - retweet= 1 - likes=5</p> <p>doc_id= doc_54 - url_tweet= https://twitter.com/TrimiA98/status/1575911455806742528 - retweet= 0 - likes=0</p> <p>doc_id= doc_474 - url_tweet= https://twitter.com/_Nex3_/status/1575851734974578688 - retweet= 0 - likes=1</p> <p>doc_id= doc_324 - url_tweet= https://twitter.com/_Nex3_/status/1575879420300390403 - retweet= 1 - likes=1</p> <p>doc_id= doc_800 - url_tweet= https://twitter.com/Ak1nc1Turk1071/status/1575819273078906881 - retweet= 0 - likes=3</p> <p>doc_id= doc_83 - url_tweet= https://twitter.com/elevotv/status/1575909051380432896 - retweet= 0 - likes=0</p> <p>doc_id= doc_3489 - url_tweet= https://twitter.com/GabeZZOZZ/status/1575213861359480833 - retweet= 0 - likes=1</p> <p>doc_id= doc_451 - url_tweet= https://twitter.com/Militarylandnet/status/1575855193547165696 - retweet= 96 - likes=1083</p> <p>doc_id= doc_409 - url_tweet= https://twitter.com/SGNewsAlerts/status/1575861833025478656 - retweet= 0 - likes=1</p>
<p>Dnipro war</p>	<p>doc_id= doc_3292 - url_tweet= https://twitter.com/firamnews/status/1575250163882463251 - retweet= 0 - likes=1</p> <p>doc_id= doc_2066 - url_tweet= https://twitter.com/Chronology22/status/1575575566186852352 - retweet= 0 - likes=1</p> <p>doc_id= doc_1993 - url_tweet= https://twitter.com/Chronology22/status/1575595404044865536 - retweet= 0 - likes=1</p> <p>doc_id= doc_345 - url_tweet= https://twitter.com/Chronology22/status/1575873706668212224 - retweet= 2 - likes=0</p> <p>doc_id= doc_2000 - url_tweet= https://twitter.com/Chronology22/status/1575594105035378688 - retweet= 0 - likes=0</p> <p>doc_id= doc_2434 - url_tweet= https://twitter.com/Chronology22/status/1575479438141280256 - retweet= 1 - likes=0</p> <p>doc_id= doc_2965 - url_tweet= https://twitter.com/Ukrinform_News/status/1575357208996478976 - retweet= 4 - likes=4</p> <p>doc_id= doc_2785 - url_tweet= https://twitter.com/Ukrinform_News/status/1575405596337135617 - retweet= 3 - likes=3</p>

	<p>doc_id= doc_1881 - url_tweet= https://twitter.com/Chronology22/status/1575617614394449920 - retweet= 1 - likes=0</p> <p>doc_id= doc_978 - url_tweet= https://twitter.com/KVladyslavka/status/1575805179730882561 - retweet= 0 - likes=0</p>
Zelenski leadership Ukraine war	<p>doc_id= doc_1283 - url_tweet= https://twitter.com/swapnilzs/status/1575761181674856448 - retweet= 0 - likes=0</p> <p>doc_id= doc_557 - url_tweet= https://twitter.com/have_fun_X/status/1575839188741951489 - retweet= 0 - likes=0</p> <p>doc_id= doc_1026 - url_tweet= https://twitter.com/georgephilipb/status/157575800143835561984 - retweet= 38 - likes=269</p> <p>doc_id= doc_451 - url_tweet= https://twitter.com/Militarylandnet/status/1575855193547165696 - retweet= 96 - likes=1083</p> <p>doc_id= doc_71 - url_tweet= https://twitter.com/PMsince1994/status/1575909819810484224 - retweet= 0 - likes=0</p> <p>doc_id= doc_417 - url_tweet= https://twitter.com/Chronology22/status/1575860740556460032 - retweet= 0 - likes=1</p> <p>doc_id= doc_2784 - url_tweet= https://twitter.com/AlbaneseAl/status/1575405656772935681 - retweet= 0 - likes=0</p> <p>doc_id= doc_2782 - url_tweet= https://twitter.com/AGCsegreteria/status/1575405725936934912 - retweet= 0 - likes=0</p> <p>doc_id= doc_2781 - url_tweet= https://twitter.com/pka71/status/1575405776260268032 - retweet= 0 - likes=0</p> <p>doc_id= doc_1162 - url_tweet= https://twitter.com/epirus_son/status/1575787462060052480 - retweet= 0 - likes=0</p>
Stop Putin's aggression Ukraine	<p>doc_id= doc_2157 - url_tweet= https://twitter.com/2536luis/status/1575548179218472960 - retweet= 1 - likes=4</p> <p>doc_id= doc_142 - url_tweet= https://twitter.com/GenevaSummit/status/1575904086448173057 - retweet= 0 - likes=0</p> <p>doc_id= doc_1344 - url_tweet= https://twitter.com/Viper_Toffi/status/1575751722017779713 - retweet= 0 - likes=1</p> <p>doc_id= doc_3941 - url_tweet= https://twitter.com/Txxani/status/1575160466980352000 - retweet= 0 - likes=1</p> <p>doc_id= doc_3076 - url_tweet= https://twitter.com/rybanova171195/status/1575321074627215364 - retweet= 0 - likes=1</p> <p>doc_id= doc_2091 - url_tweet= https://twitter.com/ukrbravery/status/1575</p>

	<pre>569862855823362 - retweet= 6 - likes=6 doc_id= doc_123 - url_tweet= https://twitter.com/Truecubbyblue15/status/1575906036678291456 - retweet= 0 - likes=0 doc_id= doc_964 - url_tweet= https://twitter.com/bowen_dilwyn/status/1575806511317913600 - retweet= 0 - likes=1 doc_id= doc_1209 - url_tweet= https://twitter.com/ArmchairOpinio1/status/1575778970389598209 - retweet= 0 - likes=1 doc_id= doc_897 - url_tweet= https://twitter.com/UATV_en/status/1575812703070715905 - retweet= 0 - likes=0</pre>
--	--

1.3) Comparing TF-IDF + cosine similarity and Our-Score + cosine similarity

The main difference between the two ranking methods is that the first uses a normal TF-IDF, which is a traditional text analysis technique, while the second also involves the design of a custom scoring mechanism that takes into account social media popularity metrics (likes and retweets), which may be more appropriate for ranking content in a social media context.

Most documents retrieved from all queries were different with our algorithm compared to the TF-IDF algorithm. This is due to the fact that we now take into account the retweets and likes of the tweet, and most of the tweets retrieved by the TF-IDF algorithm had 0 retweets and 0 likes.

Now that it takes them into account, most of the tweets have some likes or retweets but without a big impact, so we can still take the TF-IDF score into account. On the results of our score we can now see some documents with a really high number of likes or retweets and some with less but with a really strong TF-IDF.

2. Using word2vec + cosine similarity

Using Word2Vec and Cosine Similarity together for ranking leverages semantic understanding and can potentially provide more context-aware results compared to simple TF-IDF-based ranking. However, it requires careful preprocessing and consideration of computational resources and training data quality.

We used the same queries as in the previous section and obtained the following results:

Query	Returned as relevant documents (top 20)
Russian intervention Ukraine	<p>doc_id= 2271 - url_tweet= https://twitter.com/OlenaStarynets/status/1575522972495884288</p> <p>doc_id= 3705 - url_tweet= https://twitter.com/TheWilliams4405/status/1575188100166590465</p> <p>doc_id= 1161 - url_tweet= https://twitter.com/epirus_son/status/1575787462060052480</p> <p>doc_id= 974 - url_tweet= https://twitter.com/soreiabtc/status/1575805652945231873</p> <p>doc_id= 577 - url_tweet= https://twitter.com/Mark3Ds/status/1575836764929142784</p> <p>doc_id= 2461 - url_tweet= https://twitter.com/Mark3Ds/status/1575475318952951808</p> <p>doc_id= 125 - url_tweet= https://twitter.com/Kaan122143/status/1575905891978997767</p> <p>doc_id= 2931 - url_tweet= https://twitter.com/nojaydenx/status/1575364384729145345</p> <p>doc_id= 1852 - url_tweet= https://twitter.com/iduxking/status/1575624560787197952</p> <p>doc_id= 3835 - url_tweet= https://twitter.com/VoskopoulosPhd/status/1575175782132899841</p> <p>doc_id= 175 - url_tweet= https://twitter.com/chronosome/status/1575899801933185024</p> <p>doc_id= 799 - url_tweet= https://twitter.com/Ak1nc1Turk1071/status/1575819273078906881</p> <p>doc_id= 2813 - url_tweet= https://twitter.com/Militarylandnet/status/1575396903252025351</p> <p>doc_id= 2478 - url_tweet= https://twitter.com/Militarylandnet/status/1575471340391391232</p> <p>doc_id= 3711 - url_tweet= https://twitter.com/House_MD_PL/status/1575187570715246613</p>

	<p>doc_id= 3208 - url_tweet= https://twitter.com/jimmyrails1/status/1575273715390156800</p> <p>doc_id= 3280 - url_tweet= https://twitter.com/NEWS_ALL_TIME/status/1575253795352809472</p> <p>doc_id= 3123 - url_tweet= https://twitter.com/Shaya_7/status/1575304263282089984</p> <p>doc_id= 3274 - url_tweet= https://twitter.com/UKikaski/status/1575255073617334275</p> <p>doc_id= 3851 - url_tweet= https://twitter.com/Cicke69/status/1575174068353249281</p>
NATO Ukraine	<p>doc_id= 2271 - url_tweet= https://twitter.com/OlenaStarynets/status/1575522972495884288</p> <p>doc_id= 974 - url_tweet= https://twitter.com/soreiabtc/status/1575805652945231873</p> <p>doc_id= 3705 - url_tweet= https://twitter.com/TheWilliams4405/status/1575188100166590465</p> <p>doc_id= 1161 - url_tweet= https://twitter.com/epirus_son/status/1575787462060052480</p> <p>doc_id= 1749 - url_tweet= https://twitter.com/RAWONE420/status/1575640389670776832</p> <p>doc_id= 3308 - url_tweet= https://twitter.com/temafey/status/1575246948478734338</p> <p>doc_id= 577 - url_tweet= https://twitter.com/Mark3Ds/status/1575836764929142784</p> <p>doc_id= 2461 - url_tweet= https://twitter.com/Mark3Ds/status/1575475318952951808</p> <p>doc_id= 249 - url_tweet= https://twitter.com/0xethmaxii/status/1575890381027147781</p> <p>doc_id= 85 - url_tweet= https://twitter.com/lets_gobrandon_/status/1575908722077229056</p> <p>doc_id= 1869 - url_tweet= https://twitter.com/UkrainianNews24/status/1575621491785289729</p> <p>doc_id= 1097 - url_tweet= https://twitter.com/Ukrinform_News/status/1575793780350889984</p> <p>doc_id= 1486 - url_tweet= https://twitter.com/evg__10/status/1575697914046791681</p> <p>doc_id= 980 - url_tweet= https://twitter.com/IWPR/status/1575804828</p>

	<p><u>562890753</u></p> <p>doc_id= 1649 - url_tweet= https://twitter.com/L_Team10/status/1575655876882354177</p> <p>doc_id= 1654 - url_tweet= https://twitter.com/L_Team10/status/1575653787426922496</p> <p>doc_id= 1439 - url_tweet= https://twitter.com/Ukrinform_News/status/1575716269063639041</p> <p>doc_id= 2927 - url_tweet= https://twitter.com/nojaydenx/status/1575365303696908290</p> <p>doc_id= 3680 - url_tweet= https://twitter.com/annamarie_36/status/1575190323424133120</p> <p>doc_id= 2777 - url_tweet= https://twitter.com/DusharaDusharaa/status/1575406945485422592</p>
Dnipro war	<p>doc_id= 2271 - url_tweet= https://twitter.com/OlenaStarynets/status/1575522972495884288</p> <p>doc_id= 1161 - url_tweet= https://twitter.com/epirus_son/status/1575787462060052480</p> <p>doc_id= 3705 - url_tweet= https://twitter.com/TheWilliams4405/status/1575188100166590465</p> <p>doc_id= 974 - url_tweet= https://twitter.com/soreiabtc/status/1575805652945231873</p> <p>doc_id= 577 - url_tweet= https://twitter.com/Mark3Ds/status/1575836764929142784</p> <p>doc_id= 2461 - url_tweet= https://twitter.com/Mark3Ds/status/1575475318952951808</p> <p>doc_id= 125 - url_tweet= https://twitter.com/Kaan122143/status/1575905891978997767</p> <p>doc_id= 2931 - url_tweet= https://twitter.com/nojaydenx/status/1575364384729145345</p> <p>doc_id= 1852 - url_tweet= https://twitter.com/iduxking/status/1575624560787197952</p> <p>doc_id= 3835 - url_tweet= https://twitter.com/VoskopoulosPhd/status/1575175782132899841</p> <p>doc_id= 2478 - url_tweet= https://twitter.com/Militarylandnet/status/1575471340391391232</p> <p>doc_id= 2813 - url_tweet= https://twitter.com/Militarylandnet/status/1575396903252025351</p> <p>doc_id= 799 - url_tweet=</p>

	<p>https://twitter.com/AkInclTurk1071/status/1575819273078906881 doc_id= 175 - url_tweet= https://twitter.com/chronosome/status/1575899801933185024 doc_id= 3123 - url_tweet= https://twitter.com/Shaya_7/status/1575304263282089984 doc_id= 3851 - url_tweet= https://twitter.com/Cicke69/status/1575174068353249281 doc_id= 3274 - url_tweet= https://twitter.com/UKikaski/status/1575255073617334275 doc_id= 3280 - url_tweet= https://twitter.com/NEWS_ALL_TIME/status/1575253795352809472 doc_id= 3711 - url_tweet= https://twitter.com/House_MD_PL/status/1575187570715246613 doc_id= 3208 - url_tweet= https://twitter.com/jimmyrails1/status/1575273715390156800</p>
Zelenski leadership Ukraine war	<p>doc_id= 2271 - url_tweet= https://twitter.com/OlenaStarynets/status/1575522972495884288 doc_id= 1161 - url_tweet= https://twitter.com/epirus_son/status/1575787462060052480 doc_id= 974 - url_tweet= https://twitter.com/soreiabtc/status/1575805652945231873 doc_id= 3705 - url_tweet= https://twitter.com/TheWilliams4405/status/1575188100166590465 doc_id= 2461 - url_tweet= https://twitter.com/Mark3Ds/status/1575475318952951808 doc_id= 577 - url_tweet= https://twitter.com/Mark3Ds/status/1575836764929142784 doc_id= 125 - url_tweet= https://twitter.com/Kaan122143/status/1575905891978997767 doc_id= 2931 - url_tweet= https://twitter.com/nojaydenx/status/1575364384729145345 doc_id= 3711 - url_tweet= https://twitter.com/House_MD_PL/status/1575187570715246613 doc_id= 829 - url_tweet= https://twitter.com/VDacicus/status/1575817352830074880 doc_id= 2777 - url_tweet= https://twitter.com/DusharaDusharaa/status/1575406945485422592</p>

	<p>doc_id= 2927 - url_tweet= https://twitter.com/nojaydenx/status/1575365303696908290</p> <p>doc_id= 3680 - url_tweet= https://twitter.com/annamarie_36/status/1575190323424133120</p> <p>doc_id= 1749 - url_tweet= https://twitter.com/RAWONE420/status/1575640389670776832</p> <p>doc_id= 175 - url_tweet= https://twitter.com/chronosome/status/1575899801933185024</p> <p>doc_id= 799 - url_tweet= https://twitter.com/AkInclTurk1071/status/1575819273078906881</p> <p>doc_id= 3835 - url_tweet= https://twitter.com/VoskopoulosPhd/status/1575175782132899841</p> <p>doc_id= 2813 - url_tweet= https://twitter.com/Militarylandnet/status/1575396903252025351</p> <p>doc_id= 2478 - url_tweet= https://twitter.com/Militarylandnet/status/1575471340391391232</p> <p>doc_id= 1852 - url_tweet= https://twitter.com/iduxking/status/1575624560787197952</p>
<p>Stop Putin's aggression Ukraine</p>	<p>doc_id= 2271 - url_tweet= https://twitter.com/OlenaStarynets/status/1575522972495884288</p> <p>doc_id= 3705 - url_tweet= https://twitter.com/TheWilliams4405/status/1575188100166590465</p> <p>doc_id= 1161 - url_tweet= https://twitter.com/epirus_son/status/1575787462060052480</p> <p>doc_id= 974 - url_tweet= https://twitter.com/soreiabtc/status/1575805652945231873</p> <p>doc_id= 577 - url_tweet= https://twitter.com/Mark3Ds/status/1575836764929142784</p> <p>doc_id= 2461 - url_tweet= https://twitter.com/Mark3Ds/status/1575475318952951808</p> <p>doc_id= 125 - url_tweet= https://twitter.com/Kaan122143/status/1575905891978997767</p> <p>doc_id= 2931 - url_tweet= https://twitter.com/nojaydenx/status/1575364384729145345</p> <p>doc_id= 2478 - url_tweet= https://twitter.com/Militarylandnet/status/1575471340391391232</p> <p>doc_id= 2813 - url_tweet= https://twitter.com/Militarylandnet/status/1575471340391391232</p>

	<p><u>/1575396903252025351</u></p> <p>doc_id= 799 - url_tweet= https://twitter.com/Ak1nc1Turk1071/status/1575819273078906881</p> <p>doc_id= 175 - url_tweet= https://twitter.com/chronosome/status/1575899801933185024</p> <p>doc_id= 3835 - url_tweet= https://twitter.com/VoskopoulosPhd/status/1575175782132899841</p> <p>doc_id= 1749 - url_tweet= https://twitter.com/RAWONE420/status/1575640389670776832</p> <p>doc_id= 3711 - url_tweet= https://twitter.com/House_MD_PL/status/1575187570715246613</p> <p>doc_id= 3123 - url_tweet= https://twitter.com/Shaya_7/status/1575304263282089984</p> <p>doc_id= 3280 - url_tweet= https://twitter.com/NEWS_ALL_TIME/status/1575253795352809472</p> <p>doc_id= 3274 - url_tweet= https://twitter.com/UKikaski/status/1575255073617334275</p> <p>doc_id= 3851 - url_tweet= https://twitter.com/Cicke69/status/1575174068353249281</p> <p>doc_id= 3680 - url_tweet= https://twitter.com/annamarie_36/status/1575190323424133120</p>
--	--

3. How might transformer-based embeddings enhance or complicate the retrieval process compared to traditional embeddings like word2vec?

Transformer-based embeddings, such as those produced by models like BERT (Bidirectional Encoder Representations from Transformers) or GPT (Generative Pre-trained Transformer), have transformed the field of natural language processing. They offer significant advancements over traditional embeddings like Word2Vec.

However, their adoption for information retrieval and ranking tasks introduces both enhancements and complexities compared to traditional embeddings

Enhancements:

- Capture contextual understanding by considering the surrounding words in a text which can significantly improve the retrieval process by better understanding the meaning and nuances of words and phrases.
- Can perform semantic matching. They understand the context in which a word appears, allowing for better recognition of synonyms, antonyms, and context-related terms.
- Can process multi-modal data, such as text, images, and audio.
- Transformers are typically trained on vast amounts of data, making them more robust and capable of handling a wide range of topics and languages.

Complexities:

- Are computationally intensive, requiring powerful hardware for inference. This can be challenging for real-time applications or large-scale retrieval tasks.
- Are large in size, which can complicate deployment and increase memory and storage requirements.
- The complexity can lead to increased latency in retrieval processes, which may not be acceptable in some real-time applications.
- To fully leverage transformer-based embeddings, we may need access to a large amount of data, pre-trained models, and the computational resources to support them.

Although transformer-based embeddings offer substantial enhancements in terms of semantic understanding and contextual information, making them ideal for improving retrieval results in a wide range of applications, their computational intensity, size, and complexity can complicate their integration and may not be suitable for all retrieval scenarios. Choosing between traditional embeddings like Word2Vec and transformer-based embeddings should depend on the specific needs and resources of the retrieval application.