ASSIGNMENT 2

Instructions:

- Please download the provided IPython Notebook (ipynb) file and open it in Google Colab. Once opened, enter your code in the same file directly beneath the relevant question's code block.
- Insert a text block below your code to briefly explain it, mentioning any libraries or functions utilized. Answer the questions in brief with examples.

- Submit

1. The IPython Notebook (ipynb) file.
2. A PDF version of the notebook (converted from ipynb).

- The similarity score should be less than 15%

What is Sentiment Analysis?

Fundamentally, we can define sentiment analysis as the computational study of opinions, thoughts, evaluations, evaluations, interests, views, emotions, subjectivity, along with others, that are expressed in a text .

It involves classifying opinions found in text into categories like "positive" or "negative" or "neutral." Sentiment analysis is also known by different names, such as opinion mining, appraisal extraction, subjectivity analysis, and others.

For example:

"The story of the movie was boring and a waste of time."

The following terms can be extracted from the sentence above to perform sentiment analysis:

```
Opinion Owner: Audience
Object: Movie
Feature: Story
Opinion: Boring and a waste of time.
Polarity: Negative
```
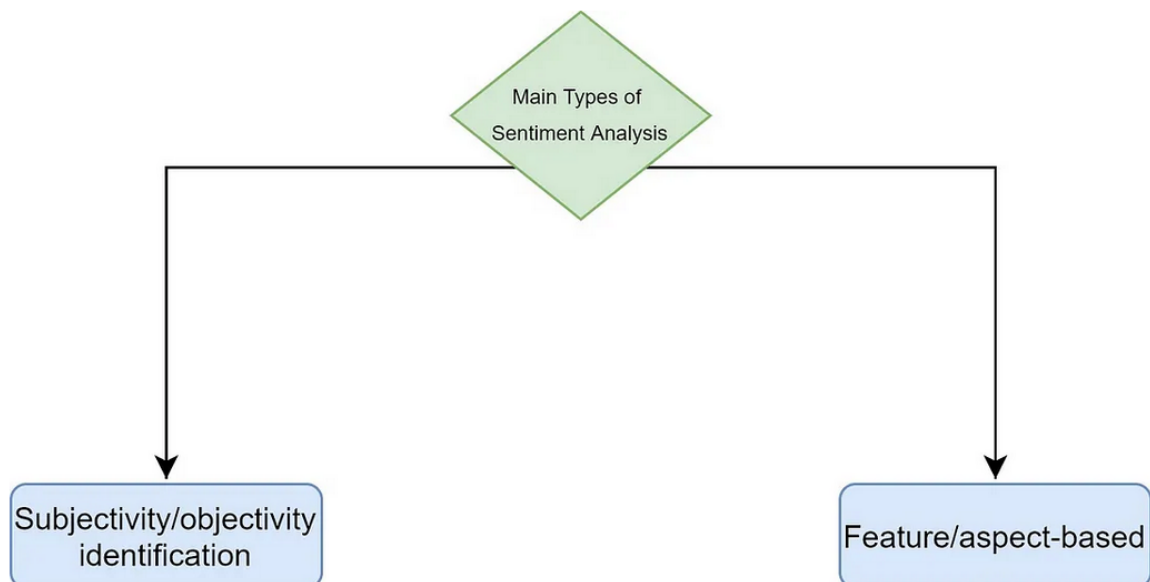

Figure 1: Main types of sentiment analysis.

Subjectivity/objectivity identification

In subjectivity or objectivity identification, a given text or sentence is classified into two different classes:

```
Subjectivity: It expresses an opinion that describes people's feelings towards a specific topic.
 e.g., The taste of this mango is good.

Objective: It expresses the fact.
 e.g., This mango is yellow.
```

The subjective sentence expresses personal feelings, views, or beliefs. Sentiment analysis works great on a text with a personal connection than on text with only an objective connection.

Feature/aspect-based

Feature or aspect-based sentiment analysis analyzes different features, attributes, or aspects of a product. Its main goal is to recognize the aspect of a given target and the sentiment shown towards each aspect.

For instance:

"Today, I purchased a Samsung phone, and my boyfriend purchased an iPhone. We called each other in the evening. The **voice** of my phone was not clear, but the **camera** was good. My girlfriend said the **sound** of her phone was very clear. So, I decided to buy a similar phone because its voice quality is very good. So, I bought an iPhone and returned the Samsung phone to the seller."

Applying aspect extraction to the sentences above:

```
Voice.
Camera.
Sound.
```

Task 1: Sentiment Analysis (50%)

Question 1

Explain different methodologies utilized in sentiment analysis, and which approach stands out as particularly effective? Note: Write atleast four different approaches, detailing their techniques and discussing their relative strengths and weaknesses in the context of sentiment analysis.

Your Answer:

- Some of the types of sentiment analysis are Emotion Detection, Aspect Sentiment analysis, Fine-Grained Sentiment analysis, and Multilingual Sentiment Analysis.

Fine Grain Sentiment Analysis:

- Done at text level and sentence level.
- Some methods that are used to understand the sentiment polarity in this analysis are stemming, a bag of words, bi-gram, tri-gram, etc.
- Categories of feedback in this are very positive, positive, neutral, negative, and very negative. ## Strengths:
- It provides detailed insights into the analysis.
- It can be applied to various types of text like social media posts, feedback, reviews, etc.
- Can capture polarity like very positive and very negative.
- Can be more effective for domain-related tasks. ## Weaknesses:
- It is limited by vocabulary
- Processing large amounts of data can lead to reduced the complexity.

Aspect Sentiment analysis

- In this type of sentiment analysis, the text is broken down into attributes or components called aspects.
- The sentiment categories here are positive, negative, or neutral which are allocated to each segment. ## Strengths:

- It provides detailed analysis for a specific aspect of features.
- It can capture multiple sentiments for the same text by separating different aspects.
- Very useful to improve a product or service based on customer reviews or feedback. ## Weaknesses:
- In order to perform aspect sentiment analysis we need to know the aspect accurately.
- The performance of the analysis depends on the complexity of the aspect.

Emotion Detection

- It detects the emotion by looking at the face of a person or from an audio or a written text. It is done based on probabilities of emotion. Whichever emotion is a high probability, the emotion of the person is that. ## Strengths:
- The emotion detection can identify the emotion from various sources like images, audio, video, text, etc.
- Can provide more detailed emotion instead of just positive or negative.
- Used in various places like customer service. ## Weaknesses:
- The accuracy sometimes may depend on the quality and quantity of training data.
- Cannot represent very complex emotional states.

Multilingual Sentiment Analysis

- This sentiment analysis is done in multiple languages using complex neural network architectures.
- There are several pre-trained models (Models trained on huge datasets) based on which the analysis is done. ## Strengths:
- It can identify the sentiment for different languages.
- It can improve because it has a large pre-trained data. ## Weaknesses:
- The performance may vary based on the quantity and quality of data present.
- Complex data can lead to performance and accuracy reduction.

Every approach has its own strengths and weaknesses. Which type of sentiment analysis to choose can depend on the where you are using that.

Tutorial 1

Rule Based Sentiment Analysis

```
In [1]:  #code block 1
         import re

         # Create a list of positive and negative words
         positive_words = ["good", "great", "excellent", "happy", "love"]
         negative_words = ["bad", "terrible", "horrible", "sad", "hate"]

         # Create a dictionary of negation words
         negation_words = ["not", "no", "never"]

         # Define a function to calculate the sentiment score of a text
         def calculate_sentiment_score(text):
             # Remove punctuation from the text and convert it to lowercase
             text = re.sub(r"[^\w\s]", "", text).lower()

             # Split the text into words
             words = text.split()

             # Initialize the sentiment score
             sentiment_score = 0

             # Initialize a variable to track if the current word is in a negation context
             in_negation = False

             # Iterate over the words in the text
             for word in words:
                 # If the word is a negation word, set the in_negation flag
                 if word in negation_words:
                     in_negation = True
                 else:
                     # If the word is in the positive words list and not in a negation context, increase the sentiment score
                     if word in positive_words and not in_negation:
                         sentiment_score += 1
                     # If the word is in the negative words list and not in a negation context, decrease the sentiment score
                     elif word in negative_words and not in_negation:
                         sentiment_score -= 1
                     # Reset the in_negation flag if the word is not a negation word
                     in_negation = False
```

```python
    # Calculate the sentiment score for the entire text
    return sentiment_score

# Example usage:

# Example texts of different sentiment sentences
sentences = ["I love this movie!",
             "This movie is terrible.",
             "I am not happy with this product.",
             "I am very happy with this product.",]

# Calculate the sentiment score for each sentence
sentiment_scores = []
for sentence in sentences:
    sentiment_score = calculate_sentiment_score(sentence)
    sentiment_scores.append(sentiment_score)

# Print the sentiment scores for each sentence
for i in range(len(sentences)):
    print(sentences[i], sentiment_scores[i])
```

```
I love this movie! 1
This movie is terrible. -1
I am not happy with this product. 0
I am very happy with this product. 1
```

- The above is a simple rule-based sentiment analysis approach.
- We have predefined the positive words, and negative words in an array.
- The sentences are the example text on which the sentiment analysis is performed.
- The idea here is, that we count the positive, and negative words for each sentence and the sentiment is based on the count.
- In the output, '1' is calssified as positive, '0' is classified as neutral and '-1' is classified as negative.

Question 2.A

Add atleast 2 or more words in each list of positive, negative and negations and show the sentiment score of sentences which has the words you added?

```python
In [2]: #code here
        #code here
        #code block 1
        import re

        # Create a list of positive and negative words
        positive_words = ["good", "great", "excellent", "happy", "love", "joy", "cheerful", "awesome", "superb", "delightful"]
        negative_words = ["bad", "terrible", "horrible", "sad", "hate", "annoying", "depressed", "upset", "regret", "disgusting"]

        # Create a dictionary of negation words
        negation_words = ["not", "no", "never", "can't", "wouldn't", "didn't", "won't", "don't"]

        # Define a function to calculate the sentiment score of a text
        def calculate_sentiment_score(text):
            # Remove punctuation from the text and convert it to lowercase
            text = re.sub(r"[^\w\s]", "", text).lower()

            # Split the text into words
            words = text.split()

            # Initialize the sentiment score
            sentiment_score = 0

            # Initialize a variable to track if the current word is in a negation context
            in_negation = False

            # Iterate over the words in the text
            for word in words:
                # If the word is a negation word, set the in_negation flag
                if word in negation_words:
                    in_negation = True
                else:
                    # If the word is in the positive words list and not in a negation context, increase the sentiment score
                    if word in positive_words and not in_negation:
                        sentiment_score += 1
                    # If the word is in the negative words list and not in a negation context, decrease the sentiment score
                    elif word in negative_words and not in_negation:
                        sentiment_score -= 1
                    # Reset the in_negation flag if the word is not a negation word
                    in_negation = False

            # Calculate the sentiment score for the entire text
            return sentiment_score

        # Example usage:
```

```python
# Example texts of different sentiment sentences
sentences = ["I love this movie!",
             "This movie is terrible.",
             "I am not happy with this product.",
             "I am very happy with this product.",
             "The food was superb and delightful",
             "I don't think this park is cheerful",
             "I won't recommend this place. It's disgusting.",
             "I haven't seen such a terrible movie in my life.",
             "I regret purchasing this phone.",
             "I can't make you upset because I love you"

             ]

# Calculate the sentiment score for each sentence
sentiment_scores = []
for sentence in sentences:
    sentiment_score = calculate_sentiment_score(sentence)
    sentiment_scores.append(sentiment_score)

# Print the sentiment scores for each sentence
for i in range(len(sentences)):
    print(sentences[i], sentiment_scores[i])
```

```
I love this movie! 1
This movie is terrible. -1
I am not happy with this product. 0
I am very happy with this product. 1
The food was superb and delightful 2
I don't think this park is cheerful 1
I won't recommend this place. It's disgusting. -1
I haven't seen such a terrible movie in my life. -1
I regret purchasing this phone. -1
I can't make you upset because I love you 0
```

- The above is a simple rule-based sentiment analysis approach using different new words added for this task.
- The idea here is, that we count the positive, and negative words for each sentence and the sentiment is based on the count.
- In the output, '1' is classified as positive, '0' is classified as neutral and '-1' is classified as negative.

- I have added new positive, negative and negation words and given some sentences to test the sentiment scores based on the new words.

- As you can see based on the words the sentimental score is been printed.

- For the 5th sentence, the score is 2 because there are 2 positive words.
- And for the sentences with -1, it means its a negative sentiment.

Tutorial 2

VADER Sentiment Analysis :

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon based sentiment analysis tool that is specifically attuned to sentiments expressed in social media. VADER uses a combination of A sentiment lexicon is a list of lexical features (e.g., words) which are generally labeled according to their semantic orientation as either positive or negative. VADER not only tells about the Positivity and Negativity score but also tells us about how positive or negative a sentiment is.

The **Compound score** is a metric that calculates the sum of all the lexicon ratings which have been normalized between -1(most extreme negative) and +1 (most extreme positive).

In [3]:
```python
!pip install vaderSentiment
```

```
Collecting vaderSentiment
  Downloading vaderSentiment-3.3.2-py2.py3-none-any.whl (125 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 126.0/126.0 kB 2.0 MB/s eta 0:00:00
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from vaderSentiment) (2.31.0)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests->vaderS
entiment) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests->vaderSentiment)
(3.6)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests->vaderSentime
nt) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests->vaderSentime
nt) (2024.2.2)
Installing collected packages: vaderSentiment
Successfully installed vaderSentiment-3.3.2
```

In [4]:
```python
#code block 2
# import SentimentIntensityAnalyzer class
# from vaderSentiment.vaderSentiment module.
```

```python
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer

# function to print sentiments
# of the sentence.
def sentiment_scores(sentence):

    # Create a SentimentIntensityAnalyzer object.
    sid_obj = SentimentIntensityAnalyzer()

    # polarity_scores method of SentimentIntensityAnalyzer
    # object gives a sentiment dictionary.
    # which contains pos, neg, neu, and compound scores.
    sentiment_dict = sid_obj.polarity_scores(sentence)

    print("Overall sentiment dictionary is : ", sentiment_dict)
    print("sentence was rated as ", sentiment_dict['neg']*100, "% Negative")
    print("sentence was rated as ", sentiment_dict['pos']*100, "% Positive")

    print("Sentence Overall Rated As", end = " ")

    # decide sentiment as positive, negative and neutral
    if sentiment_dict['compound'] >= 0.05 :
        print("Positive")

    elif sentiment_dict['compound'] <= - 0.05 :
        print("Negative")



# Driver code
if __name__ == "__main__" :

    print("\n1st statement :")
    sentence = "the movie is so good and i loved it"

    sentiment_scores(sentence)

    print("\n2nd Statement :")
    sentence = "study is going on as usual"
    #sentiment_scores(sentence)


    print("\n3rd Statement :")
    sentence = "I am so sad  today and please leave me alone."
    sentiment_scores(sentence)
```

```
1st statement :
Overall sentiment dictionary is :  {'neg': 0.0, 'neu': 0.478, 'pos': 0.522, 'compound': 0.8244}
sentence was rated as  0.0 % Negative
sentence was rated as  52.2 % Positive
Sentence Overall Rated As Positive

2nd Statement :

3rd Statement :
Overall sentiment dictionary is :  {'neg': 0.464, 'neu': 0.387, 'pos': 0.148, 'compound': -0.5982}
sentence was rated as  46.400000000000006 % Negative
sentence was rated as  14.799999999999999 % Positive
Sentence Overall Rated As Negative
```

- The above code uses Vader's sentimental analysis to give the sentiment of the given statements.
- Based on the percentages, sentence 1 is classified as positive, and statement 3 is classified as negative.

Question 2.B:

For the above code add neutral sentiment case where (compound score > -0.05) and (compound score < 0.05) and show the sentiment rate of the 2nd statement?

```python
In [5]: #code here
        #code here

        #code block 2
        # import SentimentIntensityAnalyzer class
        # from vaderSentiment.vaderSentiment module.
        from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer

        # function to print sentiments
        # of the sentence.
        def sentiment_scores(sentence):

            # Create a SentimentIntensityAnalyzer object.
            sid_obj = SentimentIntensityAnalyzer()
```

```python
    # polarity_scores method of SentimentIntensityAnalyzer
    # object gives a sentiment dictionary.
    # which contains pos, neg, neu, and compound scores.
    sentiment_dict = sid_obj.polarity_scores(sentence)

    print("Overall sentiment dictionary is : ", sentiment_dict)
    print("sentence was rated as ", sentiment_dict['neg']*100, "% Negative")
    print("sentence was rated as ", sentiment_dict['pos']*100, "% Positive")

    print("Sentence Overall Rated As", end = " ")

    # decide sentiment as positive, negative and neutral
    if sentiment_dict['compound'] >= 0.05 :
        print("Positive")

    elif sentiment_dict['compound'] <= - 0.05 :
        print("Negative")

    else:
        print("Neutral")



# Driver code
if __name__ == "__main__" :

    print("\n1st statement :")
    sentence = "the movie is so good and i loved it"

    sentiment_scores(sentence)

    print("\n2nd Statement :")
    sentence = "study is going on as usual"
    sentiment_scores(sentence)


    print("\n3rd Statement :")
    sentence = "I am so sad  today and please leave me alone."
    sentiment_scores(sentence)
```

```
1st statement :
Overall sentiment dictionary is :  {'neg': 0.0, 'neu': 0.478, 'pos': 0.522, 'compound': 0.8244}
sentence was rated as  0.0 % Negative
sentence was rated as  52.2 % Positive
Sentence Overall Rated As Positive

2nd Statement :
Overall sentiment dictionary is :  {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
sentence was rated as  0.0 % Negative
sentence was rated as  0.0 % Positive
Sentence Overall Rated As Neutral

3rd Statement :
Overall sentiment dictionary is :  {'neg': 0.464, 'neu': 0.387, 'pos': 0.148, 'compound': -0.5982}
sentence was rated as  46.400000000000006 % Negative
sentence was rated as  14.799999999999999 % Positive
Sentence Overall Rated As Negative
```

- Import the Intensity Analyzer which analyses the sentiment score for the sentence which we provided.
- Create a method which calculates the sentiment score.
- In that we create a dictonary wihch stores the scores of positivem negative and neutral sentiment.
- Compount score is a metric which aggregates the summilative score.
- As seen in the above code, if the compunt is greater than 0.05 then the sentence is classified as Positive, similarly if the score is less then 0.05, the sentence is classified as negative.
- To use this method we have given a few examples and calculated the sentiment.
- For sentence one the positive percentage is more so the sentence is classified as positive.
- Similarly, Statement 3 is classified as negative.
- For the above task I have added an else condition, which means that if the score is not more than or equal to 0.05 and not less that or equal to -0.05, which means if it is between -0.05 and 0.05 then it is a neutral statement.
- Since the score of Sentence 2 is 0, which means it is a neutral statement.

Task 2: Text Summarization (50%)

Text Summarization

Text summarization is the process of generating short, fluent, and most importantly accurate summary of a respectively longer text document. The main idea behind automatic text summarization is to be able to find a short subset of the most essential information from the entire set and present it in a human-readable format. As online textual data grows, automatic text summarization methods have the potential to be very helpful because more useful information can be read in a short time.

Type of summarization:
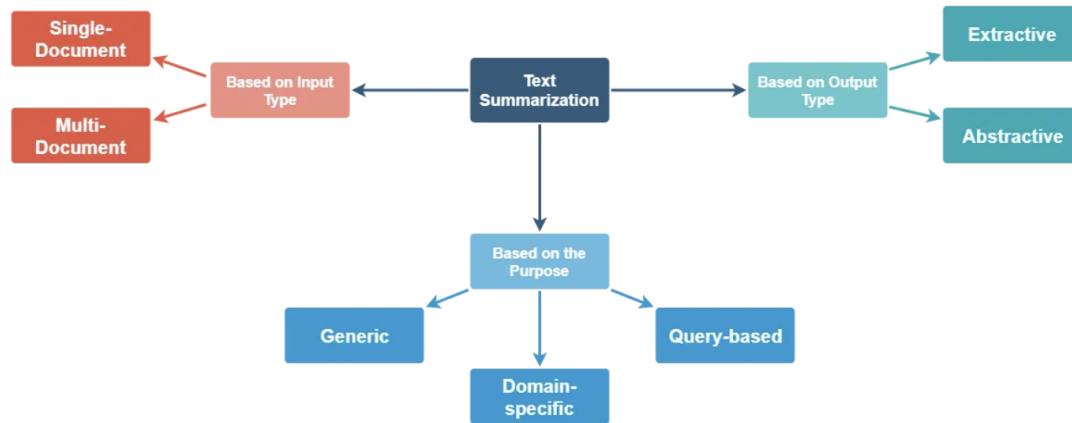
How to do text summarization

- Text cleaning
- Sentence tokenization
- Word tokenization
- Word-frequency table
- Summarization

- Text summarization is a process of generating a shorter version of a text document which contains the main points.
- The input for text summarization can be a news artical or a research paper, or a blog post and the output is a short text which connveys the essential information.

Question 1

Explain in detail how NLP's Text summarization is able to maintain the meaning and context of the initial corpus. Explain in detail how this is acheived.

Your Answer:

- Before we do actual summarization, we do preprocessing such as tokenization and removing stop words.

- Some of the ways that NLP's text summarization are Extractive summarization, Abstractive summarization, context prevention, and semantic understanding .

Extractive Summarization

- In this it extracts sentences thata are most important.
- Because it extracts the important sentence from the original text, it preservs the meaning of the original text. ## Abstract summarization
- This summarization techinique can contain sentences that are rephrased or paraphrased.
- It is done by understanding meansing which requires more advance NLP understnding. ### Contect Preservation
- There are few method which can retain pronouns, gramatical structure and maintain the tense of the original text.
- It also preservs the logical flow of the infomraiton given in original text. ### Semantic Understanding
- This contains name entity recognition, part-of speech and labeling of role, which is important for model to identify the replationship in the text.

Tutorial 1

Text summarization using NLP

How to do text summarization

- Text cleaning
- Sentence tokenization
- Word tokenization
- Word-frequency table
- Summarization

Input:

```
In [6]:  text = """
         Maria Sharapova has basically no friends as tennis players on the WTA Tour. The Russian player has no problems in openly
         I think everyone knows this is my job here. When I'm on the courts or when I'm on the court playing, I'm a competitor and
         So I'm not the one to strike up a conversation about the weather and know that in the next few minutes I have to go and t
         I'm a pretty competitive girl. I say my hellos, but I'm not sending any players flowers as well. Uhm, I'm not really frie
         I have not a lot of friends away from the courts.' When she said she is not really close to a lot of players, is that som
         I think just because you're in the same sport doesn't mean that you have to be friends with everyone just because you're
         I think every person has different interests. I have friends that have completely different jobs and interests, and I've
         I think everyone just thinks because we're tennis players we should be the greatest of friends. But ultimately tennis is
         There are so many other things that we're interested in, that we do.'
         """
```

Text cleaning:

```
In [7]:  # !pip install -U spacy
         # !python -m spacy download en_core_web_sm
         import spacy
         from spacy.lang.en.stop_words import STOP_WORDS
         from string import punctuation
         stopwords = list(STOP_WORDS)
         nlp = spacy.load('en_core_web_sm')
         doc = nlp(text)
```

Word tokenization:

```
In [8]:  tokens = [token.text for token in doc]
         print(tokens)
         punctuation = punctuation + '\n'   # Corrected quotation mark
         punctuation
         word_frequencies = {}
         for word in doc:
             if word.text.lower() not in stopwords:
                 if word.text.lower() not in punctuation:
                     if word.text not in word_frequencies.keys():
                         word_frequencies[word.text] = 1
                     else:
                         word_frequencies[word.text] += 1
         print(word_frequencies)
```

```
['\n', 'Maria', 'Sharapova', 'has', 'basically', 'no', 'friends', 'as', 'tennis', 'players', 'on', 'the', 'WTA', 'Tour',
'.', 'The', 'Russian', 'player', 'has', 'no', 'problems', 'in', 'openly', 'speaking', 'about', 'it', 'and', 'in', 'a', 'r
ecent', 'interview', 'she', 'said', ':', "'", 'I', 'do', 'n't', 'really', 'hide', 'any', 'feelings', 'too', 'much', '.',
'\n', 'I', 'think', 'everyone', 'knows', 'this', 'is', 'my', 'job', 'here', '.', 'When', 'I', ''m', 'on', 'the', 'court
s', 'or', 'when', 'I', ''m', 'on', 'the', 'court', 'playing', ',', 'I', ''m', 'a', 'competitor', 'and', 'I', 'want', 't
o', 'beat', 'every', 'single', 'person', 'whether', 'they', ''re', 'in', 'the', 'locker', 'room', 'or', 'across', 'the',
'net', '.', '\n', 'So', 'I', ''m', 'not', 'the', 'one', 'to', 'strike', 'up', 'a', 'conversation', 'about', 'the', 'weath
er', 'and', 'know', 'that', 'in', 'the', 'next', 'few', 'minutes', 'I', 'have', 'to', 'go', 'and', 'try', 'to', 'win',
'a', 'tennis', 'match', '.', '\n', 'I', ''m', 'a', 'pretty', 'competitive', 'girl', '.', 'I', 'say', 'my', 'hellos', ',',
'but', 'I', ''m', 'not', 'sending', 'any', 'players', 'flowers', 'as', 'well', '.', 'Uhm', ',', 'I', ''m', 'not', 'reall
y', 'friendly', 'or', 'close', 'to', 'many', 'players', '.', '\n', 'I', 'have', 'not', 'a', 'lot', 'of', 'friends', 'awa
y', 'from', 'the', 'courts', '.', ''', 'When', 'she', 'said', 'she', 'is', 'not', 'really', 'close', 'to', 'a', 'lot', 'o
f', 'players', ',', 'is', 'that', 'something', 'strategic', 'that', 'she', 'is', 'doing', '?', 'Is', 'it', 'different',
'on', 'the', 'men', ''s', 'tour', 'than', 'the', 'women', ''s', 'tour', '?', ''', 'No', ',', 'not', 'at', 'all', '.',
'\n', 'I', 'think', 'just', 'because', 'you', ''re', 'in', 'the', 'same', 'sport', 'does', 'n't', 'mean', 'that', 'you',
'have', 'to', 'be', 'friends', 'with', 'everyone', 'just', 'because', 'you', ''re', 'categorized', ',', 'you', ''re',
'a', 'tennis', 'player', ',', 'so', 'you', ''re', 'going', 'to', 'get', 'along', 'with', 'tennis', 'players', '.', '\n',
'I', 'think', 'every', 'person', 'has', 'different', 'interests', '.', 'I', 'have', 'friends', 'that', 'have', 'completel
y', 'different', 'jobs', 'and', 'interests', ',', 'and', 'I', ''ve', 'met', 'them', 'in', 'very', 'different', 'parts',
'of', 'my', 'life', '.', '\n', 'I', 'think', 'everyone', 'just', 'thinks', 'because', 'we', ''re', 'tennis', 'players',
'we', 'should', 'be', 'the', 'greatest', 'of', 'friends', '.', 'But', 'ultimately', 'tennis', 'is', 'just', 'a', 'very',
'small', 'part', 'of', 'what', 'we', 'do', '.', '\n', 'There', 'are', 'so', 'many', 'other', 'things', 'that', 'we', ''r
e', 'interested', 'in', ',', 'that', 'we', 'do', '.', "'", '\n']
{'Maria': 1, 'Sharapova': 1, 'basically': 1, 'friends': 5, 'tennis': 6, 'players': 6, 'WTA': 1, 'Tour': 1, 'Russian': 1,
'player': 2, 'problems': 1, 'openly': 1, 'speaking': 1, 'recent': 1, 'interview': 1, 'said': 2, 'hide': 1, 'feelings': 1,
'think': 4, 'knows': 1, 'job': 1, 'courts': 2, 'court': 1, 'playing': 1, 'competitor': 1, 'want': 1, 'beat': 1, 'single':
1, 'person': 2, 'locker': 1, 'room': 1, 'net': 1, 'strike': 1, 'conversation': 1, 'weather': 1, 'know': 1, 'minutes': 1,
'try': 1, 'win': 1, 'match': 1, 'pretty': 1, 'competitive': 1, 'girl': 1, 'hellos': 1, 'sending': 1, 'flowers': 1, 'Uhm':
1, 'friendly': 1, 'close': 2, 'lot': 2, 'away': 1, ''': 1, 'strategic': 1, 'different': 4, 'men': 1, 'tour': 2, 'women':
1, ''': 1, 'sport': 1, 'mean': 1, 'categorized': 1, 'going': 1, 'interests': 2, 'completely': 1, 'jobs': 1, 'met': 1, 'pa
rts': 1, 'life': 1, 'thinks': 1, 'greatest': 1, 'ultimately': 1, 'small': 1, 'things': 1, 'interested': 1}
```

- The above code extracts each word is a token and maintain the word frequency which means how many times the word has occored on the corpus.

Sentence tokenization:

In [9]:
```python
max_frequency = max(word_frequencies.values())
max_frequency
for word in word_frequencies.keys():
    word_frequencies[word] = word_frequencies[word] / max_frequency
print(word_frequencies)
sentence_tokens = [sent for sent in doc.sents]
print(sentence_tokens)
```

```
{'Maria': 0.16666666666666666, 'Sharapova': 0.16666666666666666, 'basically': 0.16666666666666666, 'friends': 0.833333333
3333334, 'tennis': 1.0, 'players': 1.0, 'WTA': 0.16666666666666666, 'Tour': 0.16666666666666666, 'Russian': 0.16666666666
666666, 'player': 0.3333333333333333, 'problems': 0.16666666666666666, 'openly': 0.16666666666666666, 'speaking': 0.16666
666666666666, 'recent': 0.16666666666666666, 'interview': 0.16666666666666666, 'said': 0.3333333333333333, 'hide': 0.1666
6666666666666, 'feelings': 0.16666666666666666, 'think': 0.6666666666666666, 'knows': 0.16666666666666666, 'job': 0.16666
666666666666, 'courts': 0.3333333333333333, 'court': 0.16666666666666666, 'playing': 0.16666666666666666, 'competitor':
0.16666666666666666, 'want': 0.16666666666666666, 'beat': 0.16666666666666666, 'single': 0.16666666666666666, 'person':
0.3333333333333333, 'locker': 0.16666666666666666, 'room': 0.16666666666666666, 'net': 0.16666666666666666, 'strike': 0.1
6666666666666666, 'conversation': 0.16666666666666666, 'weather': 0.16666666666666666, 'know': 0.16666666666666666, 'minu
tes': 0.16666666666666666, 'try': 0.16666666666666666, 'win': 0.16666666666666666, 'match': 0.16666666666666666, 'prett
y': 0.16666666666666666, 'competitive': 0.16666666666666666, 'girl': 0.16666666666666666, 'hellos': 0.16666666666666666,
'sending': 0.16666666666666666, 'flowers': 0.16666666666666666, 'Uhm': 0.16666666666666666, 'friendly': 0.166666666666666
66, 'close': 0.3333333333333333, 'lot': 0.3333333333333333, 'away': 0.16666666666666666, ''': 0.16666666666666666, 'strat
egic': 0.16666666666666666, 'different': 0.6666666666666666, 'men': 0.16666666666666666, 'tour': 0.3333333333333333, 'wom
en': 0.16666666666666666, ''': 0.16666666666666666, 'sport': 0.16666666666666666, 'mean': 0.16666666666666666, 'categoriz
ed': 0.16666666666666666, 'going': 0.16666666666666666, 'interests': 0.3333333333333333, 'completely': 0.166666666666666
6, 'jobs': 0.16666666666666666, 'met': 0.16666666666666666, 'parts': 0.16666666666666666, 'life': 0.16666666666666666, 't
hinks': 0.16666666666666666, 'greatest': 0.16666666666666666, 'ultimately': 0.16666666666666666, 'small': 0.1666666666666
6666, 'things': 0.16666666666666666, 'interested': 0.16666666666666666}
[
Maria Sharapova has basically no friends as tennis players on the WTA Tour., The Russian player has no problems in openly
speaking about it and in a recent interview she said: 'I don't really hide any feelings too much.
, I think everyone knows this is my job here., When I'm on the courts or when I'm on the court playing, I'm a competitor
and I want to beat every single person whether they're in the locker room or across the net.
, So I'm not the one to strike up a conversation about the weather and know that in the next few minutes I have to go and
try to win a tennis match.
, I'm a pretty competitive girl., I say my hellos, but I'm not sending any players flowers as well., Uhm, I'm not really
friendly or close to many players.
, I have not a lot of friends away from the courts.', When she said she is not really close to a lot of players, is that
something strategic that she is doing?, Is it different on the men's tour than the women's tour?, 'No, not at all.
, I think just because you're in the same sport doesn't mean that you have to be friends with everyone just because you'r
e categorized, you're a tennis player, so you're going to get along with tennis players.
, I think every person has different interests., I have friends that have completely different jobs and interests, and
I've met them in very different parts of my life.
, I think everyone just thinks because we're tennis players we should be the greatest of friends., But ultimately tennis
is just a very small part of what we do.
, There are so many other things that we're interested in, that we do.'
]
```

- The above code extracts sentences as tokens

Word frequency table:

```python
In [10]:   sentence_scores = {}
           for sent in sentence_tokens:
               for word in sent:
                   if word.text.lower() in word_frequencies.keys():
                       if sent not in sentence_scores.keys():
                           sentence_scores[sent] = word_frequencies[word.text.lower()]
                       else:
                           sentence_scores[sent] += word_frequencies[word.text.lower()]
           sentence_scores
```

```
Out[10]:  {
           Maria Sharapova has basically no friends as tennis players on the WTA Tour.: 3.3333333333333335,
           The Russian player has no problems in openly speaking about it and in a recent interview she said: 'I don't really hide
          any feelings too much.: 1.8333333333333333,
           I think everyone knows this is my job here.: 0.9999999999999999,
           When I'm on the courts or when I'm on the court playing, I'm a competitor and I want to beat every single person whether
          they're in the locker room or across the net.: 2.1666666666666665,
           So I'm not the one to strike up a conversation about the weather and know that in the next few minutes I have to go and
          try to win a tennis match.: 2.333333333333333,
           I'm a pretty competitive girl.: 0.5,
           I say my hellos, but I'm not sending any players flowers as well.: 1.5,
           Uhm, I'm not really friendly or close to many players.: 1.5,
           I have not a lot of friends away from the courts.': 1.8333333333333335,
           When she said she is not really close to a lot of players, is that something strategic that she is doing?: 2.16666666666
          66665,
           Is it different on the men's tour than the women's tour?: 1.6666666666666665,
           'No, not at all.: 0.16666666666666666,
           I think just because you're in the same sport doesn't mean that you have to be friends with everyone just because you're
          categorized, you're a tennis player, so you're going to get along with tennis players.: 5.5,
           I think every person has different interests.: 1.9999999999999998,
           I have friends that have completely different jobs and interests, and I've met them in very different parts of my life.:
          3.3333333333333326,
           I think everyone just thinks because we're tennis players we should be the greatest of friends.: 3.833333333333333,
           But ultimately tennis is just a very small part of what we do.: 1.3333333333333335,
           There are so many other things that we're interested in, that we do.': 0.3333333333333333}
```

- Here we create a word frequency table of each sentance.

Summarization:

```python
In [11]:   from heapq import nlargest
           select_length = int(len(sentence_tokens) * 0.3)
           select_length
           summary = nlargest(select_length, sentence_scores, key=sentence_scores.get)
           summary
           final_summary = [word.text for word in summary]
           summary = ' '.join(final_summary)
           print(summary)
```

```
I think just because you're in the same sport doesn't mean that you have to be friends with everyone just because you're
categorized, you're a tennis player, so you're going to get along with tennis players.
 I think everyone just thinks because we're tennis players we should be the greatest of friends.
Maria Sharapova has basically no friends as tennis players on the WTA Tour. I have friends that have completely different
jobs and interests, and I've met them in very different parts of my life.
 So I'm not the one to strike up a conversation about the weather and know that in the next few minutes I have to go and
try to win a tennis match.
```

- To summarize the text we use nlargest module.
- When calling the methods we also provide the sentense length whcih we watn as output.
- In the above example we have taken 30% of the origianl length.

Question 2.A

Now for this task select any news article url and extract the text from url and summarize the text like above and then perform sentiment analysis(from reference of VADER sentiment Analysis in tutorial 2 from task1:SENTIMENT ANALYSIS) on both the original text and its summary text, and then compare their sentiment scores to see if the summary accurately reflects the sentiment of the original text

```python
In [12]:   #code block 1
           #import the required libraries
```

```
from bs4 import BeautifulSoup
import urllib.request
import nltk
from nltk.corpus import stopwords
nltk.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
```

Out[12]: True

- Import all the required libraries

In [13]:
```
# code here
import urllib.request

#open and read the spacex data
response = urllib.request.urlopen('https://neuters.de/world/europe/kharkivs-civilians-under-fire-ukraine-faces-catastroph
html = response.read()
print(html)
```

b'<!DOCTYPE html><html lang="en"><head><title>Ukraine\'s air defence shortages leave Kharkiv more exposed to Russian bomb s</title><link rel="stylesheet" href="/main.css?v=0"><meta name="viewport" content="width=device-width, initial-scale=1"> <meta property="og:title" content="Ukraine\'s air defence shortages leave Kharkiv more exposed to Russian bombs"><meta pr operty="og:type" content="article"><meta property="og:description" content="Kateryna Velnychuk was having an afternoon na p when an explosion shattered the windows of her ground-floor flat, spraying shrapnel that tore holes through her walls a nd cupboards."><meta property="og:url" content="/world/europe/kharkivs-civilians-under-fire-ukraine-faces-catastrophic-ai r-defence-shortage-2024-04-12/"></head><body><main><h1>Ukraine\'s air defence shortages leave Kharkiv more exposed to Rus sian bombs</h1><p class="byline">2024-04-12 10:43 - Tom Balmforth</p><p>Kateryna Ve lnychuk was having an afternoon nap when an explosion shattered the windows of her ground-floor flat, spraying shrapnel t hat tore holes through her walls and cupboards.</p><p>A Russian guided bomb had exploded in the courtyard outside the fiv e-storey Soviet-era building, killing a postman on his rounds. As her flat filled with thick, milky smoke, the 22-year-ol d turned to see blood pouring from her boyfriend Vladyslav\'s head.</p><p>\xe2\x80\x9cAs we\xe2\x80\x99ve been living\xe2 \x80\xa6in a state of war, there was no sense of fear in the moment,\xe2\x80\x9d Velnychuk said. \xe2\x80\x9cYou just und erstand there was an explosion. The only thought in your head is \xe2\x80\x98I hope we survive\xe2\x80\x99.\xe2\x80\x9d</ p><p>As Russia has intensified i ts air campaign against Ukraine in the last month, hammering its energy infrastructure and urban areas, no major city has been harder hit than Kharkiv.</p><p>Just 30 km (18 miles) from the Russian border in northeast Ukraine, Kharkiv was a lready the most exposed to missile attacks and bombardment.</p><p>But the drying up of Western military support in recent months \xe2\x80\x93 as a vital U.S. military aid package has been stuck in Congress amid Republican resistance \xe2\x80\x93 has left Kh arkiv even more dangerously unprotected.</p><p>"We have a catastrophic shortage of air defence systems," Governor Oleh Sy nehubov told Reuters, standing in the city\xe2\x80\x99s vast central plaza, Freedom Square. "Not only in the Kharkiv regi on, but throughout the entire country. Especially in the Kharkiv region."</p><p>The city is so near the border that Russi an missiles can reach their target in less than a minute. The deployment of Ukraine\xe2\x80\x99s precious air defences, s uch as the U.S.-made Patriot surface-to-air missile systems, which are high-value targets for Russian airstrikes, has to be done more cautiously so close to enemy lines, officials say.</p><p>Ukrainian President Volodymyr Zelenskiy, who has ur gently appealed for more air defence supplies from the West, said this week that almost a quarter of Kharkiv had been des troyed.</p><p>He accused Russia of seeking to reduce the city - which was home to 2 million people before the war - to ru bble, clearing the way for its troops to advance. He said Ukraine\'s military would repel any such offensive.</p><p>The b ombardments come as the momentum on the battlefield has shifted in Russia\'s favour, more than two years since it launche d its Feb. 2022 invasion.</p><p>Russia denies targeting civilians and says Ukraine\xe2\x80\x99s energy system is a legiti mate military target. The Russian defence ministry did not immediately respond to a request for comment for this story.</ p><p>Reuters interviewed 15 civilians in Kharkiv who expressed their determination to stay in their homes despite the att acks - though two of them flagged the bleak situation on the power front as a real concern.</p><p>At least 10 missiles ra ined down on Kharkiv on Thursday, triggering emergency blackouts for 200,000 people in the surrounding region, as Russia launched its third major air attack on energy infrastructure across Ukraine in recent weeks.</p><p>The region\'s top pros ecutor Oleksandr Filchakov told Reuters that all of the Kharkiv region\'s power facilities have been damaged or destroyed since Russia renewed its aerial assault last month, causing large-scale power cuts.</p><p>Russia had test-fired a new kin d of aircraft-launched guided bomb at least six times as of Tuesday, he said, like the one that struck the courtyard outs ide Velnychuk\'s home.</p><p>The weapon, which Filchakov called a \xe2\x80\x9cunified multi-purpose guided munition\xe2\x 80\x9d, weighs just 250 kilogrammes (550 pounds) and has a range of 90 km (56 miles), meaning aircraft have no need to ri sk getting close to the city\xe2\x80\x99s defences.</p><p>While the guided bombs are less accurate and destructive than o ther missiles used by Russia, such as S-300s and Iskanders, they are much cheaper for Russia to produce, he said.</p><p >"The (attacks) are mainly aimed at intimidating the civilian population," Filchakov told Reuters in his offices. "They \'re trying to make people leave the city, leave their buildings, homes, apartments...To sow panic in the city."</p><p>St rikes and shelling have killed 97 civilians in the region this year, he said, adding that nearly all the recent attacks h ad hit civilian targets.</p><p>Velnychuk was shaken but suffered no major injuries from the guided bomb that landed outsi de her building on March 27, blowing out all the windows along two rows of the red-brick residential buildings. But both she and her partner, who works as a courier, said they had no plans to leave the city.</p><p>"I always imagined I would g row up and have some kind of life, move from the village to the city, study. Now I live and ... I don\'t even know if I \'ll wake up tomorrow morning," said Velynchuk, a hairdresser.</p><p>"But, at the same time, you want to live in your own home. It\'s normal to want to live where you were born."</p><h2>RUSSIA GAINS MOMENTUM</h2><p>After months of attritional fighting, Russia is slowly advancing in Ukraine\'s eastern region of Donetsk this year. Kyiv\'s forces find themselves on the back foot, facing shortages of artillery shells and air defences, and grappling with manpower problems.</p><p>Ukraine \'s parliament passed legislation on Thursday to overhaul how the armed forces draft civilians into the ranks, in an effort to bolster the f rontlines. The final law, however, excluded clauses on draconian penalties for draft dodging that had caused public outcr y.</p><p>Much would depend on how well the new law \xe2\x80\x93 expected to take effect in mid-May - was implemented, ana lysts said.</p><p>"There are two issues now: the ammunition issue and the manpower issue. If they address them, I think U kraine can hold back Russian advances," said Rob Lee, a senior fellow at the Foreign Policy Research Institute, a think t ank in Philadelphia.</p><p>\xe2\x80\x9cBut if they don\'t get addressed, there is a potential that Russia will make great er gains this summer."</p><p>Ukraine has tried to find a pressure point against Russia by bombing oil facilities far behi nd the front lines using long-range drones that have taken out 14% of energy giant Russia\'s oil refining capacity, accor ding to Reuters calculations.</p><p>Zelenskiy, who inspected Ukrainian defensive fortifications in the Kharkiv region on Tuesday, has said Russia may be preparing a big offensive in late May or in June. He did not say where.</p><p>Russia, whi ch captured the eastern town of Avdiivka in February and controls 18% of Ukrainian territory, has inched forward in the D onetsk region, keeping up pressure on the fronts west of Avdiivka and the city of Bakhmut.</p><p>Lee said Russia had fixed its own manpower prob lems and managed to recruit a large number of volunteers, allowing it to sustain losses in assaults, but that it faced eq uipment limitations that could become a problem next year.</p><p>Russia\'s decision on where to attack, he said, would in part depend on where it thought Ukraine looked weakest, although Moscow would likely maintain its focus on the eastern Do nbas region.</p><p>Putin said last month he did not rule out Russia trying to establish a buffer zone inside Ukrainian te rritory along the Russian border.</p><p>Oleksandr Kovalenko, an independent military analyst based in Odesa, said the str ikes on Kharkiv looked aimed at setting in motion such a plan by trying to scare people to leave the city, laying the gro und for a possible ground operation at a later date.</p><p>"For the moment, Russia does not have the forces and equipment to seize the city, but in the medium term they can terrorise the civilian population to prepare the corresponding conditi ons."</p><h2>\'HOW COULD I LEAVE?\'</h2><p>Kharkiv, an industrial hub that once served as the capital of Soviet Ukraine, is a jarring contrast of 1.3 million people going about their lives amid regular air raid sirens and the sound of machine guns downing drones at night.</p><p>Schools in the city have been closed because of the threat of attacks and children st udy online. But authorities have opened underground classrooms in a metro station to allow some pupils to come to classes in person.</p><p>The city\xe2\x80\x99s population plunged to 300,000 after the invasion but, after Ukraine recaptured occ upied areas of the region in two military offensives in 2022, it returned to around 1.3 million, where it has remained si nce.</p><p>Viktoria Zaremba, 37, a web designer and mother of a 10-year-old boy, said more than two years of war had chan ged her perception of risk.</p><p>"There is no fear," she said. She would only consider leaving Kharkiv if there were no central heating or electricity this winter, or a looming threat of occupation.</p><p>The number of attacks on the city an d region began increasing in October, Filchakov said, rising more than 35% in the first three months this year to 130 fro m 95 in the last quarter of 2023.</p><p>They have intensified again this month, he said.</p><p>Rolling blackouts last up

to 12 hours a day and the traffic lights don\'t work. Mobile coverage is patchy, online GPS maps don\'t function properl
y, and the street lights stay off at night.</p><p>But Synehubov, the regional governor, said there was no sign people wer
e abandoning the city.</p><p>"I\'ll never leave," said Borys Nosov, 63, a pensioner walking his dog in the city centre. N
osov said he was a veteran of the 1979–1989 Soviet–Afghan war.</p><p>"This is my city. How could I leave and abandon it?
I served in Afghanistan. That was terrifying. I think everything will be okay."</p></main><footer><div><a href="/">Home</
a> – Search – About</div></footer></body></html>'

- Read the html data from a website

```
In [14]:   #code block 3
           #apply beautifulsoap(used for web scraping) and split the text
           soup = BeautifulSoup(html,"html5lib") #"html5lib" parser is a robust parser that can handle poorly formatted HTML and pro
           text = soup.get_text(strip=True)
           print(text)
```

Ukraine's air defence shortages leave Kharkiv more exposed to Russian bombsUkraine's air defence shortages leave Kharkiv more exposed to Russian bombs2024-04-12 10:43 -Tom BalmforthKateryna Velnychuk was having an afternoon nap when an explosion shattered the windows of her ground-floor flat, spraying shrapnel that tore holes through her walls and cupboards.A Russian guided bomb had exploded in the courtyard outside the five-storey Soviet-era building, killing a postman on his rounds. As her flat filled with thick, milky smoke, the 22-year-old turned to see blood pouring from her boyfriend Vladyslav's head."As we've been living…in a state of war, there was no sense of fear in the moment," Velnychuk said. "You just understand there was an explosion. The only thought in your head is 'I hope we survive'."As Russia hasintensified its air campaignagainst Ukraine in the last month, hammering its energy infrastructure and urban areas, no major city has been harder hit than Kharkiv.Just 30 km (18 miles) from the Russian border in northeast Ukraine, Kharkiv was already the most exposed to missile attacks and bombardment.But the drying up of Western military support in recent months — as a vital U.S. military aid package has been stuck in Congressamid Republican resistance— has left Kharkiv even more dangerously unprotected."We have a catastrophic shortage of air defence systems," Governor Oleh Synehubov told Reuters, standing in the city's vast central plaza, Freedom Square. "Not only in the Kharkiv region, but throughout the entire country. Especially in the Kharkiv region."The city is so near the border that Russian missiles can reach their target in less than a minute. The deployment of Ukraine's precious air defences, such as the U.S.-made Patriot surface-to-air missile systems, which are high-value targets for Russian airstrikes, has to be done more cautiously so close to enemy lines, officials say.Ukrainian President Volodymyr Zelenskiy, who has urgently appealed for more air defence supplies from the West, said this week that almost a quarter of Kharkiv had been destroyed.He accused Russia of seeking to reduce the city — which was home to 2 million people before the war — to rubble, clearing the way for its troops to advance. He said Ukraine's military would repel any such offensive.The bombardments come as the momentum on the battlefield has shifted in Russia's favour, more than two years since it launched its Feb. 2022 invasion.Russia denies targeting civilians and says Ukraine's energy system is a legitimate military target. The Russian defence ministry did not immediately respond to a request for comment for this story.Reuters interviewed 15 civilians in Kharkiv who expressed their determination to stay in their homes despite the attacks — though two of them flagged the bleak situation on the power front as a real concern.At least 10 missiles rained down on Kharkiv on Thursday, triggering emergency blackouts for 200,000 people in the surrounding region, as Russia launched its third major air attack on energy infrastructure across Ukraine in recent weeks.The region's top prosecutor Oleksandr Filchakov told Reuters that all of the Kharkiv region's power facilities have been damaged or destroyed since Russia renewed its aerial assault last month, causing large-scale power cuts.Russia had test-fired a new kind of aircraft-launched guided bomb at least six times as of Tuesday, he said, like the one that struck the courtyard outside Velnychuk's home. The weapon, which Filchakov called a "unified multi-purpose guided munition", weighs just 250 kilogrammes (550 pounds) and has a range of 90 km (56 miles), meaning aircraft have no need to risk getting close to the city's defences.While the guided bombs are less accurate and destructive than other missiles used by Russia, such as S-300s and Iskanders, they are much cheaper for Russia to produce, he said."The (attacks) are mainly aimed at intimidating the civilian population," Filchakov told Reuters in his offices. "They're trying to make people leave the city, leave their buildings, homes, apartments...To sow panic in the city."Strikes and shelling have killed 97 civilians in the region this year, he said, adding that nearly all the recent attacks had hit civilian targets.Velnychuk was shaken but suffered no major injuries from the guided bomb that landed outside her building on March 27, blowing out all the windows along two rows of the red-brick residential buildings. But both she and her partner, who works as a courier, said they had no plans to leave the city."I always imagined I would grow up and have some kind of life, move from the village to the city, study. Now I live and ... I don't even know if I'll wake up tomorrow morning," said Velynchuk, a hairdresser."But, at the same time, you want to live in your own home. It's normal to want to live where you were born."RUSSIA GAINS MOMENTUMAfter months of attritional fighting, Russia is slowly advancing in Ukraine's eastern region of Donetsk this year. Kyiv's forces find themselves on the back foot, facing shortages of artillery shells and air defences, and grappling with manpower problems.Ukraine's parliament passed legislation on Thursdayto overhaul how the armed forces draft civiliansinto the ranks, in an effort to bolster the frontlines. The final law, however, excluded clauses on draconian penalties for draft dodging that had caused public outcry. Much would depend on how well the new law — expected to take effect in mid-May — was implemented, analysts said."There are two issues now: the ammunition issue and the manpower issue. If they address them, I think Ukraine can hold back Russian advances," said Rob Lee, a senior fellow at the Foreign Policy Research Institute, a think tank in Philadelphia."But if they don't get addressed, there is a potential that Russia will make greater gains this summer."Ukraine has tried to find a pressure point against Russia by bombing oil facilities far behind the front lines using long-range drones that have taken out 14% of energy giant Russia's oil refining capacity, according to Reuters calculations.Zelenskiy, who inspected Ukrainian defensive fortifications in the Kharkiv region on Tuesday, has said Russia may be preparing a big offensive in late May or in June. He did not say where.Russia, which captured the eastern town of Avdiivka in February and controls 18% of Ukrainian territory, has inched forward in the Donetsk region, keeping up pressure on the fronts west of Avdiivka andthe city of Bakhmut.Lee said Russia had fixed its own manpower problems and managed to recruit a large number of volunteers, allowing it to sustain losses in assaults, but that it faced equipment limitations that could become a problem next year.Russia's decision on where to attack, he said, would in part depend on where it thought Ukraine looked weakest, although Moscow would likely maintain its focus on the eastern Donbas region.Putin said last month he did not rule out Russia trying to establish a buffer zone inside Ukrainian territory along the Russian border.Oleksandr Kovalenko, an independent military analyst based in Odesa, said the strikes on Kharkiv looked aimed at setting in motion such a plan by trying to scare people to leave the city, laying the ground for a possible ground operation at a later date."For the moment, Russia does not have the forces and equipment to seize the city, but in the medium term they can terrorise the civilian population to prepare the corresponding conditions."'HOW COULD I LEAVE?'Kharkiv, an industrial hub that once served as the capital of Soviet Ukraine, is a jarring contrast of 1.3 million people going about their lives amid regular air raid sirens and the sound of machine guns downing drones at night.Schools in the city have been closed because of the threat of attacks and children study online. But authorities have opened underground classrooms in a metro station to allow some pupils to come to classes in person.The city's population plunged to 300,000 after the invasion but, after Ukraine recaptured occupied areas of the region in two military offensives in 2022, it returned to around 1.3 million, where it has remained since.Viktoria Zaremba, 37, a web designer and mother of a 10-year-old boy, said more than two years of war had changed her perception of risk."There is no fear," she said. She would only consider leaving Kharkiv if there were no central heating or electricity this winter, or a looming threat of occupation.The number of attacks on the city and region began increasing in October, Filchakov said, rising more than 35% in the first three months this year to 130 from 95 in the last quarter of 2023.They have intensified again this month, he said.Rolling blackouts last up to 12 hours a day and the traffic lights don't work. Mobile coverage is patchy, online GPS maps don't function properly, and the street lights stay off at night.But Synehubov, the regional governor, said there was no sign people were abandoning the city."I'll never leave," said Borys Nosov, 63, a pensioner walking his dog in the city centre. Nosov said he was a veteran of the 1979-1989 Soviet-Afghan war."This is my city. How could I leave and abandon it? I served in Afghanistan. That was terrifying. I think everything will be okay."Home-Search-About

- Using beautifulSoup convert html to text

In [15]:
```python
# !pip install -U spacy
# !python -m spacy download en_core_web_sm
import spacy
from spacy.lang.en.stop_words import STOP_WORDS
from string import punctuation
```

```
stopwords = list(STOP_WORDS)
nlp = spacy.load('en_core_web_sm')
doc = nlp(text)
```

- Remove stop words.

In [16]:
```
tokens = [token.text for token in doc]
print(tokens)
punctuation = punctuation + '\n'  # Corrected quotation mark
punctuation
word_frequencies = {}
for word in doc:
    if word.text.lower() not in stopwords:
        if word.text.lower() not in punctuation:
            if word.text not in word_frequencies.keys():
                word_frequencies[word.text] = 1
            else:
                word_frequencies[word.text] += 1
print(word_frequencies)
```

```
['Ukraine', "'s", 'air', 'defence', 'shortages', 'leave', 'Kharkiv', 'more', 'exposed', 'to', 'Russian', 'bombsUkraine',
"'s", 'air', 'defence', 'shortages', 'leave', 'Kharkiv', 'more', 'exposed', 'to', 'Russian', 'bombs2024', '–', '04', '–',
'12', '10:43', '–Tom', 'BalmforthKateryna', 'Velnychuk', 'was', 'having', 'an', 'afternoon', 'nap', 'when', 'an', 'explos
ion', 'shattered', 'the', 'windows', 'of', 'her', 'ground', '–', 'floor', 'flat', ',', 'spraying', 'shrapnel', 'that', 't
ore', 'holes', 'through', 'her', 'walls', 'and', 'cupboards', '.', 'A', 'Russian', 'guided', 'bomb', 'had', 'exploded',
'in', 'the', 'courtyard', 'outside', 'the', 'five', '–', 'storey', 'Soviet', '–', 'era', 'building', ',', 'killing', 'a',
'postman', 'on', 'his', 'rounds', '.', 'As', 'her', 'flat', 'filled', 'with', 'thick', ',', 'milky', 'smoke', ',', 'the',
'22', '–', 'year', '–', 'old', 'turned', 'to', 'see', 'blood', 'pouring', 'from', 'her', 'boyfriend', 'Vladyslav', "'s",
'head', '.', '“As', 'we', "'ve", 'been', 'living', '…', 'in', 'a', 'state', 'of', 'war', ',', 'there', 'was', 'no', 'sens
e', 'of', 'fear', 'in', 'the', 'moment', ',', '"', 'Velnychuk', 'said', '.', '"', 'You', 'just', 'understand', 'there',
'was', 'an', 'explosion', '.', 'The', 'only', 'thought', 'in', 'your', 'head', 'is', '‘', 'I', 'hope', 'we', 'survive’',
'.', '”As', 'Russia', 'hasintensified', 'its', 'air', 'campaignagainst', 'Ukraine', 'in', 'the', 'last', 'month', ',', 'h
ammering', 'its', 'energy', 'infrastructure', 'and', 'urban', 'areas', ',', 'no', 'major', 'city', 'has', 'been', 'harde
r', 'hit', 'than', 'Kharkiv', '.', 'Just', '30', 'km', '(', '18', 'miles', ')', 'from', 'the', 'Russian', 'border', 'in',
'northeast', 'Ukraine', ',', 'Kharkiv', 'was', 'already', 'the', 'most', 'exposed', 'to', 'missile', 'attacks', 'and', 'b
ombardment', '.', 'But', 'the', 'drying', 'up', 'of', 'Western', 'military', 'support', 'in', 'recent', 'months', '—', 'a
s', 'a', 'vital', 'U.S.', 'military', 'aid', 'package', 'has', 'been', 'stuck', 'in', 'Congressamid', 'Republican', 'resi
stance', '—', 'has', 'left', 'Kharkiv', 'even', 'more', 'dangerously', 'unprotected', '.', '"We', 'have', 'a', 'catastrop
hic', 'shortage', 'of', 'air', 'defence', 'systems', ',', '"', 'Governor', 'Oleh', 'Synehubov', 'told', 'Reuters', ',',
'standing', 'in', 'the', 'city', "'s", 'vast', 'central', 'plaza', ',', 'Freedom', 'Square', '.', '"', 'Not', 'only', 'i
n', 'the', 'Kharkiv', 'region', ',', 'but', 'throughout', 'the', 'entire', 'country', '.', 'Especially', 'in', 'the', 'Kh
arkiv', 'region', '.', '"The', 'city', 'is', 'so', 'near', 'the', 'border', 'that', 'Russian', 'missiles', 'can', 'reac
h', 'their', 'target', 'in', 'less', 'than', 'a', 'minute', '.', 'The', 'deployment', 'of', 'Ukraine', "'s", 'precious',
'air', 'defences', ',', 'such', 'as', 'the', 'U.S.–made', 'Patriot', 'surface', '–', 'to', '–', 'air', 'missile', 'system
s', ',', 'which', 'are', 'high', '–', 'value', 'targets', 'for', 'Russian', 'airstrikes', ',', 'has', 'to', 'be', 'done',
'more', 'cautiously', 'so', 'close', 'to', 'enemy', 'lines', ',', 'officials', 'say', '.', 'Ukrainian', 'President', 'Vol
odymyr', 'Zelenskiy', ',', 'who', 'has', 'urgently', 'appealed', 'for', 'more', 'air', 'defence', 'supplies', 'from', 'th
e', 'West', ',', 'said', 'this', 'week', 'that', 'almost', 'a', 'quarter', 'of', 'Kharkiv', 'had', 'been', 'destroyed',
'.', 'He', 'accused', 'Russia', 'of', 'seeking', 'to', 'reduce', 'the', 'city', '–', 'which', 'was', 'home', 'to', '2',
'million', 'people', 'before', 'the', 'war', '–', 'to', 'rubble', ',', 'clearing', 'the', 'way', 'for', 'its', 'troops',
'to', 'advance', '.', 'He', 'said', 'Ukraine', "'s", 'military', 'would', 'repel', 'any', 'such', 'offensive', '.', 'Th
e', 'bombardments', 'come', 'as', 'the', 'momentum', 'on', 'the', 'battlefield', 'has', 'shifted', 'in', 'Russia', "'s",
'favour', ',', 'more', 'than', 'two', 'years', 'since', 'it', 'launched', 'its', 'Feb.', '2022', 'invasion', '.', 'Russi
a', 'denies', 'targeting', 'civilians', 'and', 'says', 'Ukraine', "'s", 'energy', 'system', 'is', 'a', 'legitimate', 'mil
itary', 'target', '.', 'The', 'Russian', 'defence', 'ministry', 'did', 'not', 'immediately', 'respond', 'to', 'a', 'reque
st', 'for', 'comment', 'for', 'this', 'story', '.', 'Reuters', 'interviewed', '15', 'civilians', 'in', 'Kharkiv', 'who',
'expressed', 'their', 'determination', 'to', 'stay', 'in', 'their', 'homes', 'despite', 'the', 'attacks', '–', 'though',
'two', 'of', 'them', 'flagged', 'the', 'bleak', 'situation', 'on', 'the', 'power', 'front', 'as', 'a', 'real', 'concern',
'.', 'At', 'least', '10', 'missiles', 'rained', 'down', 'on', 'Kharkiv', 'on', 'Thursday', ',', 'triggering', 'emergenc
y', 'blackouts', 'for', '200,000', 'people', 'in', 'the', 'surrounding', 'region', ',', 'as', 'Russia', 'launched', 'it
s', 'third', 'major', 'air', 'attack', 'on', 'energy', 'infrastructure', 'across', 'Ukraine', 'in', 'recent', 'weeks',
'.', 'The', 'region', "'s", 'top', 'prosecutor', 'Oleksandr', 'Filchakov', 'told', 'Reuters', 'that', 'all', 'of', 'the',
'Kharkiv', 'region', "'s", 'power', 'facilities', 'have', 'been', 'damaged', 'or', 'destroyed', 'since', 'Russia', 'renew
ed', 'its', 'aerial', 'assault', 'last', 'month', ',', 'causing', 'large', '–', 'scale', 'power', 'cuts', '.', 'Russia',
'had', 'test', '–', 'fired', 'a', 'new', 'kind', 'of', 'aircraft', '–', 'launched', 'guided', 'bomb', 'at', 'least', 'si
x', 'times', 'as', 'of', 'Tuesday', ',', 'he', 'said', ',', 'like', 'the', 'one', 'that', 'struck', 'the', 'courtyard',
'outside', 'Velnychuk', "'s", 'home', '.', 'The', 'weapon', ',', 'which', 'Filchakov', 'called', 'a', '“', 'unified', 'mu
lti', '–', 'purpose', 'guided', 'munition', '”', ',', 'weighs', 'just', '250', 'kilogrammes', '(', '550', 'pounds', ')',
'and', 'has', 'a', 'range', 'of', '90', 'km', '(', '56', 'miles', ')', ',', 'meaning', 'aircraft', 'have', 'no', 'need',
'to', 'risk', 'getting', 'close', 'to', 'the', 'city', "'s", 'defences', '.', 'While', 'the', 'guided', 'bombs', 'are',
'less', 'accurate', 'and', 'destructive', 'than', 'other', 'missiles', 'used', 'by', 'Russia', ',', 'such', 'as', 'S–300
s', 'and', 'Iskanders', ',', 'they', 'are', 'much', 'cheaper', 'for', 'Russia', 'to', 'produce', ',', 'he', 'said', '.',
'"The', '(', 'attacks', ')', 'are', 'mainly', 'aimed', 'at', 'intimidating', 'the', 'civilian', 'population', ',', '"',
'Filchakov', 'told', 'Reuters', 'in', 'his', 'offices', '.', '"', 'They', "'re", 'trying', 'to', 'make', 'people', 'leav
e', 'the', 'city', ',', 'leave', 'their', 'buildings', ',', 'homes', ',', 'apartments', '...', 'To', 'sow', 'panic', 'i
n', 'the', 'city', '.', '"Strikes', 'and', 'shelling', 'have', 'killed', '97', 'civilians', 'in', 'the', 'region', 'thi
s', 'year', ',', 'he', 'said', ',', 'adding', 'that', 'nearly', 'all', 'the', 'recent', 'attacks', 'had', 'hit', 'civilia
n', 'targets', '.', 'Velnychuk', 'was', 'shaken', 'but', 'suffered', 'no', 'major', 'injuries', 'from', 'the', 'guided',
'bomb', 'that', 'landed', 'outside', 'her', 'building', 'on', 'March', '27', ',', 'blowing', 'out', 'all', 'the', 'window
s', 'along', 'two', 'rows', 'of', 'the', 'red', '–', 'brick', 'residential', 'buildings', '.', 'But', 'both', 'she', 'an
d', 'her', 'partner', ',', 'who', 'works', 'as', 'a', 'courier', ',', 'said', 'they', 'had', 'no', 'plans', 'to', 'leav
e', 'the', 'city', '.', '"I', 'always', 'imagined', 'I', 'would', 'grow', 'up', 'and', 'have', 'some', 'kind', 'of', 'lif
e', ',', 'move', 'from', 'the', 'village', 'to', 'the', 'city', ',', 'study', '.', 'Now', 'I', 'live', 'and', '...', 'I',
'do', "n't", 'even', 'know', 'if', 'I', "'ll", 'wake', 'up', 'tomorrow', 'morning', ',', '"', 'said', 'Velynchuk', ',',
'a', 'hairdresser', '.', '"But', ',', 'at', 'the', 'same', 'time', ',', 'you', 'want', 'to', 'live', 'in', 'your', 'own',
'home', '.', 'It', "'s", 'normal', 'to', 'want', 'to', 'live', 'where', 'you', 'were', 'born', '.', '"RUSSIA', 'GAINS',
'MOMENTUMAfter', 'months', 'of', 'attritional', 'fighting', ',', 'Russia', 'is', 'slowly', 'advancing', 'in', 'Ukraine',
"'s", 'eastern', 'region', 'of', 'Donetsk', 'this', 'year', '.', 'Kyiv', "'s", 'forces', 'find', 'themselves', 'on', 'th
e', 'back', 'foot', ',', 'facing', 'shortages', 'of', 'artillery', 'shells', 'and', 'air', 'defences', ',', 'and', 'grapp
ling', 'with', 'manpower', 'problems', '.', 'Ukraine', "'s", 'parliament', 'passed', 'legislation', 'on', 'Thursdayto',
'overhaul', 'how', 'the', 'armed', 'forces', 'draft', 'civiliansinto', 'the', 'ranks', ',', 'in', 'an', 'effort', 'to',
'bolster', 'the', 'frontlines', '.', 'The', 'final', 'law', ',', 'however', ',', 'excluded', 'clauses', 'on', 'draconia
n', 'penalties', 'for', 'draft', 'dodging', 'that', 'had', 'caused', 'public', 'outcry', '.', 'Much', 'would', 'depend',
'on', 'how', 'well', 'the', 'new', 'law', '—', 'expected', 'to', 'take', 'effect', 'in', 'mid', '–', 'May', '—', 'was',
'implemented', ',', 'analysts', 'said', '.', '"There', 'are', 'two', 'issues', 'now', ':', 'the', 'ammunition', 'issue',
'and', 'the', 'manpower', 'issue', '.', 'If', 'they', 'address', 'them', ',', 'I', 'think', 'Ukraine', 'can', 'hold', 'ba
ck', 'Russian', 'advances', ',', '"', 'said', 'Rob', 'Lee', ',', 'a', 'senior', 'fellow', 'at', 'the', 'Foreign', 'Polic
y', 'Research', 'Institute', ',', 'a', 'think', 'tank', 'in', 'Philadelphia', '.', '“But', 'if', 'they', 'do', "n't", 'ge
t', 'addressed', ',', 'there', 'is', 'a', 'potential', 'that', 'Russia', 'will', 'make', 'greater', 'gains', 'this', 'sum
mer', '.', '”Ukraine', 'has', 'tried', 'to', 'find', 'a', 'pressure', 'point', 'against', 'Russia', 'by', 'bombing', 'oi
l', 'facilities', 'far', 'behind', 'the', 'front', 'lines', 'using', 'long', '–', 'range', 'drones', 'that', 'have', 'tak
en', 'out', '14', '%', 'of', 'energy', 'giant', 'Russia', "'s", 'oil', 'refining', 'capacity', ',', 'according', 'to', 'R
euters', 'calculations', '.', 'Zelenskiy', ',', 'who', 'inspected', 'Ukrainian', 'defensive', 'fortifications', 'in', 'th
e', 'Kharkiv', 'region', 'on', 'Tuesday', ',', 'has', 'said', 'Russia', 'may', 'be', 'preparing', 'a', 'big', 'offensiv
e', 'in', 'late', 'May', 'or', 'in', 'June', '.', 'He', 'did', 'not', 'say', 'where', '.', 'Russia', ',', 'which', 'captu
red', 'the', 'eastern', 'town', 'of', 'Avdiivka', 'in', 'February', 'and', 'controls', '18', '%', 'of', 'Ukrainian', 'ter
ritory', ',', 'has', 'inched', 'forward', 'in', 'the', 'Donetsk', 'region', ',', 'keeping', 'up', 'pressure', 'on', 'th
```

e', 'fronts', 'west', 'of', 'Avdiivka', 'andthe', 'city', 'of', 'Bakhmut', '.', 'Lee', 'said', 'Russia', 'had', 'fixed', 'its', 'own', 'manpower', 'problems', 'and', 'managed', 'to', 'recruit', 'a', 'large', 'number', 'of', 'volunteers', ',', 'allowing', 'it', 'to', 'sustain', 'losses', 'in', 'assaults', ',', 'but', 'that', 'it', 'faced', 'equipment', 'limitatio ns', 'that', 'could', 'become', 'a', 'problem', 'next', 'year', '.', 'Russia', "'s", 'decision', 'on', 'where', 'to', 'at tack', ',', 'he', 'said', ',', 'would', 'in', 'part', 'depend', 'on', 'where', 'it', 'thought', 'Ukraine', 'looked', 'wea kest', ',', 'although', 'Moscow', 'would', 'likely', 'maintain', 'its', 'focus', 'on', 'the', 'eastern', 'Donbas', 'regio n', '.', 'Putin', 'said', 'last', 'month', 'he', 'did', 'not', 'rule', 'out', 'Russia', 'trying', 'to', 'establish', 'a', 'buffer', 'zone', 'inside', 'Ukrainian', 'territory', 'along', 'the', 'Russian', 'border', '.', 'Oleksandr', 'Kovalenko', ',', 'an', 'independent', 'military', 'analyst', 'based', 'in', 'Odesa', ',', 'said', 'the', 'strikes', 'on', 'Kharkiv', 'looked', 'aimed', 'at', 'setting', 'in', 'motion', 'such', 'a', 'plan', 'by', 'trying', 'to', 'scare', 'people', 'to', 'leave', 'the', 'city', ',', 'laying', 'the', 'ground', 'for', 'a', 'possible', 'ground', 'operation', 'at', 'a', 'late r', 'date', '.', '"For', 'the', 'moment', ',', 'Russia', 'does', 'not', 'have', 'the', 'forces', 'and', 'equipment', 't o', 'seize', 'the', 'city', ',', 'but', 'in', 'the', 'medium', 'term', 'they', 'can', 'terrorise', 'the', 'civilian', 'po pulation', 'to', 'prepare', 'the', 'corresponding', 'conditions', '.', '"\'HOW', 'COULD', 'I', "LEAVE?'Kharkiv", ',', 'a n', 'industrial', 'hub', 'that', 'once', 'served', 'as', 'the', 'capital', 'of', 'Soviet', 'Ukraine', ',', 'is', 'a', 'ja rring', 'contrast', 'of', '1.3', 'million', 'people', 'going', 'about', 'their', 'lives', 'amid', 'regular', 'air', 'rai d', 'sirens', 'and', 'the', 'sound', 'of', 'machine', 'guns', 'downing', 'drones', 'at', 'night', '.', 'Schools', 'in', 'the', 'city', 'have', 'been', 'closed', 'because', 'of', 'the', 'threat', 'of', 'attacks', 'and', 'children', 'study', 'online', '.', 'But', 'authorities', 'have', 'opened', 'underground', 'classrooms', 'in', 'a', 'metro', 'station', 'to', 'allow', 'some', 'pupils', 'to', 'come', 'to', 'classes', 'in', 'person', '.', 'The', 'city', "'s", 'population', 'plunge d', 'to', '300,000', 'after', 'the', 'invasion', 'but', ',', 'after', 'Ukraine', 'recaptured', 'occupied', 'areas', 'of', 'the', 'region', 'in', 'two', 'military', 'offensives', 'in', '2022', ',', 'it', 'returned', 'to', 'around', '1.3', 'mill ion', ',', 'where', 'it', 'has', 'remained', 'since', '.', 'Viktoria', 'Zaremba', ',', '37', ',', 'a', 'web', 'designer', 'and', 'mother', 'of', 'a', '10', '–', 'year', '–', 'old', 'boy', ',', 'said', 'more', 'than', 'two', 'years', 'of', 'wa r', 'had', 'changed', 'her', 'perception', 'of', 'risk', '.', '"There', 'is', 'no', 'fear', ',', '"', 'she', 'said', ',', 'She', 'would', 'only', 'consider', 'leaving', 'Kharkiv', 'if', 'there', 'were', 'no', 'central', 'heating', 'or', 'elect ricity', 'this', 'winter', ',', 'or', 'a', 'looming', 'threat', 'of', 'occupation', '.', 'The', 'number', 'of', 'attack s', 'on', 'the', 'city', 'and', 'region', 'began', 'increasing', 'in', 'October', ',', 'Filchakov', 'said', ',', 'risin g', 'more', 'than', '35', '%', 'in', 'the', 'first', 'three', 'months', 'this', 'year', 'to', '130', 'from', '95', 'in', 'the', 'last', 'quarter', 'of', '2023.They', 'have', 'intensified', 'again', 'this', 'month', ',', 'he', 'said', '.', 'Ro lling', 'blackouts', 'last', 'up', 'to', '12', 'hours', 'a', 'day', 'and', 'the', 'traffic', 'lights', 'do', "n't", 'wor k', '.', 'Mobile', 'coverage', 'is', 'patchy', ',', 'online', 'GPS', 'maps', 'do', "n't", 'function', 'properly', ',', 'a nd', 'the', 'street', 'lights', 'stay', 'off', 'at', 'night', '.', 'But', 'Synehubov', ',', 'the', 'regional', 'governo r', ',', 'said', 'there', 'was', 'no', 'sign', 'people', 'were', 'abandoning', 'the', 'city', '.', '"I\'ll', 'never', 'le ave', ',', '"', 'said', 'Borys', 'Nosov', ',', '63', ',', 'a', 'pensioner', 'walking', 'his', 'dog', 'in', 'the', 'city', 'centre', '.', 'Nosov', 'said', 'he', 'was', 'a', 'veteran', 'of', 'the', '1979', '–', '1989', 'Soviet', '–', 'Afghan', 'war', '.', '"This', 'is', 'my', 'city', '.', 'How', 'could', 'I', 'leave', 'and', 'abandon', 'it', '?', 'I', 'served', 'in', 'Afghanistan', '.', 'That', 'was', 'terrifying', '.', 'I', 'think', 'everything', 'will', 'be', 'okay', '.', '"Hom e', '–', 'Search', '–', 'About']
{'Ukraine': 13, 'air': 10, 'defence': 5, 'shortages': 3, 'leave': 8, 'Kharkiv': 14, 'exposed': 3, 'Russian': 9, 'bombsUkr aine': 1, 'bombs2024': 1, '04': 1, '12': 2, '10:43': 1, '–Tom': 1, 'BalmforthKateryna': 1, 'Velnychuk': 4, 'having': 1, 'afternoon': 1, 'nap': 1, 'explosion': 2, 'shattered': 1, 'windows': 2, 'ground': 3, 'floor': 1, 'flat': 2, 'spraying': 1, 'shrapnel': 1, 'tore': 1, 'holes': 1, 'walls': 1, 'cupboards': 1, 'guided': 5, 'bomb': 3, 'exploded': 1, 'courtyard': 2, 'outside': 3, 'storey': 1, 'Soviet': 3, 'era': 1, 'building': 2, 'killing': 1, 'postman': 1, 'rounds': 1, 'filled': 1, 'thick': 1, 'milky': 1, '22': 1, 'year': 6, 'old': 2, 'turned': 1, 'blood': 1, 'pouring': 1, 'boyfriend': 1, 'Vladyslav': 1, 'head': 2, '"As': 1, 'living': 1, '…': 1, 'state': 1, 'war': 4, 'sense': 1, 'fear': 2, 'moment': 2, '"': 2, 'said': 22, '"': 2, 'understand': 1, 'thought': 2, ''': 1, 'hope': 1, 'survive''': 1, '"As': 1, 'Russia': 19, 'hasinten sified': 1, 'campaignagainst': 1, 'month': 4, 'hammering': 1, 'energy': 4, 'infrastructure': 2, 'urban': 1, 'areas': 2, 'major': 3, 'city': 18, 'harder': 1, 'hit': 2, '30': 1, 'km': 2, '18': 2, 'miles': 2, 'border': 3, 'northeast': 1, 'missi le': 2, 'attacks': 6, 'bombardment': 1, 'drying': 1, 'Western': 1, 'military': 6, 'support': 1, 'recent': 3, 'months': 3, '–': 3, 'vital': 1, 'U.S.': 1, 'aid': 1, 'package': 1, 'stuck': 1, 'Congressamid': 1, 'Republican': 1, 'resistance': 1, 'left': 1, 'dangerously': 1, 'unprotected': 1, '"We': 1, 'catastrophic': 1, 'shortage': 1, 'systems': 2, 'Governor': 1, 'Oleh': 1, 'Synehubov': 2, 'told': 3, 'Reuters': 5, 'standing': 1, 'vast': 1, 'central': 2, 'plaza': 1, 'Freedom': 1, 'Sq uare': 1, 'region': 12, 'entire': 1, 'country': 1, 'Especially': 1, '"The': 2, 'near': 1, 'missiles': 3, 'reach': 1, 'tar get': 2, 'minute': 1, 'deployment': 1, 'precious': 1, 'defences': 3, 'U.S.–made': 1, 'Patriot': 1, 'surface': 1, 'high': 1, 'value': 1, 'targets': 2, 'airstrikes': 1, 'cautiously': 1, 'close': 2, 'enemy': 1, 'lines': 2, 'officials': 1, 'Ukrai nian': 4, 'President': 1, 'Volodymyr': 1, 'Zelenskiy': 2, 'urgently': 1, 'appealed': 1, 'supplies': 1, 'West': 1, 'week': 1, 'quarter': 2, 'destroyed': 2, 'accused': 1, 'seeking': 1, 'reduce': 1, 'home': 3, '2': 1, 'million': 3, 'people': 6, 'rubble': 1, 'clearing': 1, 'way': 1, 'troops': 1, 'advance': 1, 'repel': 1, 'offensive': 2, 'bombardments': 1, 'come': 2, 'momentum': 1, 'battlefield': 1, 'shifted': 1, 'favour': 1, 'years': 2, 'launched': 3, 'Feb.': 1, '2022': 2, 'invasio n': 2, 'denies': 1, 'targeting': 1, 'civilians': 3, 'says': 1, 'system': 1, 'legitimate': 1, 'ministry': 1, 'immediatel y': 1, 'respond': 1, 'request': 1, 'comment': 1, 'story': 1, 'interviewed': 1, '15': 1, 'expressed': 1, 'determination': 1, 'stay': 2, 'homes': 2, 'despite': 1, 'flagged': 1, 'bleak': 1, 'situation': 1, 'power': 3, 'real': 1, 'concern': 1, '1 0': 2, 'rained': 1, 'Thursday': 1, 'triggering': 1, 'emergency': 1, 'blackouts': 2, '200,000': 1, 'surrounding': 1, 'atta ck': 2, 'weeks': 1, 'prosecutor': 1, 'Oleksandr': 2, 'Filchakov': 4, 'facilities': 2, 'damaged': 1, 'renewed': 1, 'aeria l': 1, 'assault': 1, 'causing': 1, 'large': 2, 'scale': 1, 'cuts': 1, 'test': 1, 'fired': 1, 'new': 2, 'kind': 2, 'aircra ft': 2, 'times': 1, 'Tuesday': 2, 'like': 1, 'struck': 1, 'weapon': 1, 'called': 1, 'unified': 1, 'multi': 1, 'purpose': 1, 'munition': 1, 'weighs': 1, '250': 1, 'kilogrammes': 1, '550': 1, 'pounds': 1, 'range': 2, '90': 1, '56': 1, 'meanin g': 1, 'need': 1, 'risk': 2, 'getting': 1, 'bombs': 1, 'accurate': 1, 'destructive': 1, 'S–300s': 1, 'Iskanders': 1, 'che aper': 1, 'produce': 1, 'mainly': 1, 'aimed': 2, 'intimidating': 1, 'civilian': 3, 'population': 3, 'offices': 1, 'tryin g': 3, 'buildings': 2, 'apartments': 1, '...': 2, 'sow': 1, 'panic': 1, '"Strikes': 1, 'shelling': 1, 'killed': 1, '97': 1, 'adding': 1, 'nearly': 1, 'shaken': 1, 'suffered': 1, 'injuries': 1, 'landed': 1, 'March': 1, '27': 1, 'blowing': 1, 'rows': 1, 'red': 1, 'brick': 1, 'residential': 1, 'partner': 1, 'works': 1, 'courier': 1, 'plans': 1, '"I': 1, 'imagine d': 1, 'grow': 1, 'life': 1, 'village': 1, 'study': 2, 'live': 3, 'know': 1, 'wake': 1, 'tomorrow': 1, 'morning': 1, 'Vel ynchuk': 1, 'hairdresser': 1, '"But': 1, 'time': 1, 'want': 2, 'normal': 1, 'born': 1, '"RUSSIA': 1, 'GAINS': 1, 'MOMENTU MAfter': 1, 'attritional': 1, 'fighting': 1, 'slowly': 1, 'advancing': 1, 'eastern': 3, 'Donetsk': 2, 'Kyiv': 1, 'force s': 3, 'find': 2, 'foot': 1, 'facing': 1, 'artillery': 1, 'shells': 1, 'grappling': 1, 'manpower': 3, 'problems': 2, 'par liament': 1, 'passed': 1, 'legislation': 1, 'Thursdayto': 1, 'overhaul': 1, 'armed': 1, 'draft': 2, 'civiliansinto': 1, 'ranks': 1, 'effort': 1, 'bolster': 1, 'frontlines': 1, 'final': 1, 'law': 2, 'excluded': 1, 'clauses': 1, 'draconian': 1, 'penalties': 1, 'dodging': 1, 'caused': 1, 'public': 1, 'outcry': 1, 'depend': 2, 'expected': 1, 'effect': 1, 'mid': 1, 'implemented': 1, 'analysts': 1, '"There': 2, 'issues': 1, 'ammunition': 1, 'issue': 2, 'address': 1, 'think': 3, 'hol d': 1, 'advances': 1, 'Rob': 1, 'Lee': 2, 'senior': 1, 'fellow': 1, 'Foreign': 1, 'Policy': 1, 'Research': 1, 'Institut e': 1, 'tank': 1, 'Philadelphia': 1, '"But': 1, 'addressed': 1, 'potential': 1, 'greater': 1, 'gains': 1, 'summer': 1, '"Ukraine': 1, 'tried': 1, 'pressure': 2, 'point': 1, 'bombing': 1, 'oil': 2, 'far': 1, 'long': 1, 'drones': 2, 'taken': 1, '14': 1, 'giant': 1, 'refining': 1, 'capacity': 1, 'according': 1, 'calculations': 1, 'inspected': 1, 'defensive': 1, 'fortifications': 1, 'preparing': 1, 'big': 1, 'late': 1, 'June': 1, 'captured': 1, 'town': 1, 'Avdiivka': 2, 'February': 1, 'controls': 1, 'territory': 2, 'inched': 1, 'forward': 1, 'keeping': 1, 'fronts': 1, 'west': 1, 'andthe': 1, 'Bakhmu

t': 1, 'fixed': 1, 'managed': 1, 'recruit': 1, 'number': 2, 'volunteers': 1, 'allowing': 1, 'sustain': 1, 'losses': 1, 'assaults': 1, 'faced': 1, 'equipment': 2, 'limitations': 1, 'problem': 1, 'decision': 1, 'looked': 2, 'weakest': 1, 'Moscow': 1, 'likely': 1, 'maintain': 1, 'focus': 1, 'Donbas': 1, 'Putin': 1, 'rule': 1, 'establish': 1, 'buffer': 1, 'zone': 1, 'inside': 1, 'Kovalenko': 1, 'independent': 1, 'analyst': 1, 'based': 1, 'Odesa': 1, 'strikes': 1, 'setting': 1, 'motion': 1, 'plan': 1, 'scare': 1, 'laying': 1, 'possible': 1, 'operation': 1, 'later': 1, 'date': 1, '"For': 1, 'seize': 1, 'medium': 1, 'term': 1, 'terrorise': 1, 'prepare': 1, 'corresponding': 1, 'conditions': 1, '"\'HOW': 1, "LEAVE?'Kharkiv": 1, 'industrial': 1, 'hub': 1, 'served': 2, 'capital': 1, 'jarring': 1, 'contrast': 1, '1.3': 2, 'going': 1, 'lives': 1, 'amid': 1, 'regular': 1, 'raid': 1, 'sirens': 1, 'sound': 1, 'machine': 1, 'guns': 1, 'downing': 1, 'night': 2, 'Schools': 1, 'closed': 1, 'threat': 2, 'children': 1, 'online': 2, 'authorities': 1, 'opened': 1, 'underground': 1, 'classrooms': 1, 'metro': 1, 'station': 1, 'allow': 1, 'pupils': 1, 'classes': 1, 'person': 1, 'plunged': 1, '300,000': 1, 'recaptured': 1, 'occupied': 1, 'offensives': 1, 'returned': 1, 'remained': 1, 'Viktoria': 1, 'Zaremba': 1, '37': 1, 'web': 1, 'designer': 1, 'mother': 1, 'boy': 1, 'changed': 1, 'perception': 1, 'consider': 1, 'leaving': 1, 'heating': 1, 'electricity': 1, 'winter': 1, 'looming': 1, 'occupation': 1, 'began': 1, 'increasing': 1, 'October': 1, 'rising': 1, '35': 1, '130': 1, '95': 1, '2023.They': 1, 'intensified': 1, 'Rolling': 1, 'hours': 1, 'day': 1, 'traffic': 1, 'lights': 2, 'work': 1, 'Mobile': 1, 'coverage': 1, 'patchy': 1, 'GPS': 1, 'maps': 1, 'function': 1, 'properly': 1, 'street': 1, 'regional': 1, 'governor': 1, 'sign': 1, 'abandoning': 1, '"I\'ll': 1, 'Borys': 1, 'Nosov': 2, '63': 1, 'pensioner': 1, 'walking': 1, 'dog': 1, 'centre': 1, 'veteran': 1, '1979': 1, '1989': 1, 'Afghan': 1, '"This': 1, 'abandon': 1, 'Afghanistan': 1, 'terrifying': 1, 'okay': 1, '"Home': 1, 'Search': 1}

- Extract each words and calculate the frequencies.

```
In [17]: max_frequency = max(word_frequencies.values())
max_frequency
for word in word_frequencies.keys():
    word_frequencies[word] = word_frequencies[word] / max_frequency
print(word_frequencies)
sentence_tokens = [sent for sent in doc.sents]
print(sentence_tokens)
```

{'Ukraine': 0.5909090909090909, 'air': 0.45454545454545453, 'defence': 0.22727272727272727, 'shortages': 0.13636363636363
635, 'leave': 0.3636363636363636365, 'Kharkiv': 0.6363636363636364, 'exposed': 0.13636363636363635, 'Russian': 0.4090909090
909091, 'bombsUkraine': 0.045454545454545456, 'bombs2024': 0.045454545454545456, '04': 0.045454545454545456, '12': 0.0909
0909090909091, '10:43': 0.045454545454545456, '—Tom': 0.045454545454545456, 'BalmforthKateryna': 0.045454545454545456, 'V
elnychuk': 0.18181818181818182, 'having': 0.045454545454545456, 'afternoon': 0.045454545454545456, 'nap': 0.0454545454545
45456, 'explosion': 0.09090909090909091, 'shattered': 0.045454545454545456, 'windows': 0.09090909090909091, 'ground': 0.1
3636363636363635, 'floor': 0.045454545454545456, 'flat': 0.09090909090909091, 'spraying': 0.045454545454545456, 'shrapne
l': 0.045454545454545456, 'tore': 0.045454545454545456, 'holes': 0.045454545454545456, 'walls': 0.045454545454545456, 'cu
pboards': 0.045454545454545456, 'guided': 0.22727272727272727, 'bomb': 0.13636363636363635, 'exploded': 0.045454545454545
456, 'courtyard': 0.09090909090909091, 'outside': 0.13636363636363635, 'storey': 0.045454545454545456, 'Soviet': 0.136363
63636363635, 'era': 0.045454545454545456, 'building': 0.09090909090909091, 'killing': 0.045454545454545456, 'postman': 0.
045454545454545456, 'rounds': 0.045454545454545456, 'filled': 0.045454545454545456, 'thick': 0.045454545454545456, 'milk
y': 0.045454545454545456, 'smoke': 0.045454545454545456, '22': 0.045454545454545456, 'year': 0.2727272727272727, 'old':
0.09090909090909091, 'turned': 0.045454545454545456, 'blood': 0.045454545454545456, 'pouring': 0.045454545454545456, 'boy
friend': 0.045454545454545456, 'Vladyslav': 0.045454545454545456, 'head': 0.09090909090909091, '"As': 0.045454545454545
6, 'living': 0.045454545454545456, '…': 0.045454545454545456, 'state': 0.045454545454545456, 'war': 0.18181818181818182,
'sense': 0.045454545454545456, 'fear': 0.09090909090909091, 'moment': 0.09090909090909091, '"': 0.09090909090909091, 'sai
d': 1.0, '"': 0.09090909090909091, 'understand': 0.045454545454545456, 'thought': 0.09090909090909091, '': 0.04545454545
4545456, 'hope': 0.045454545454545456, 'survive"': 0.045454545454545456, '"As': 0.045454545454545456, 'Russia': 0.8636363
636363636, 'hasintensified': 0.045454545454545456, 'campaignagainst': 0.045454545454545456, 'month': 0.18181818181818182,
'hammering': 0.045454545454545456, 'energy': 0.18181818181818182, 'infrastructure': 0.09090909090909091, 'urban': 0.04545
4545454545456, 'areas': 0.09090909090909091, 'major': 0.13636363636363635, 'city': 0.8181818181818182, 'harder': 0.045454
545454545456, 'hit': 0.09090909090909091, '30': 0.045454545454545456, 'km': 0.09090909090909091, '18': 0.0909090909090909
1, 'miles': 0.09090909090909091, 'border': 0.13636363636363635, 'northeast': 0.045454545454545456, 'missile': 0.090909090
90909091, 'attacks': 0.2727272727272727, 'bombardment': 0.045454545454545456, 'drying': 0.045454545454545456, 'Western':
0.045454545454545456, 'military': 0.2727272727272727, 'support': 0.045454545454545456, 'recent': 0.13636363636363635, 'mo
nths': 0.13636363636363635, '—': 0.13636363636363635, 'vital': 0.045454545454545456, 'U.S.': 0.045454545454545456, 'aid':
0.045454545454545456, 'package': 0.045454545454545456, 'stuck': 0.045454545454545456, 'Congressamid': 0.04545454545454545
6, 'Republican': 0.045454545454545456, 'resistance': 0.045454545454545456, 'left': 0.045454545454545456, 'dangerously':
0.045454545454545456, 'unprotected': 0.045454545454545456, 'catastrophic': 0.045454545454545
456, 'shortage': 0.045454545454545456, 'systems': 0.09090909090909091, 'Governor': 0.045454545454545456, 'Oleh': 0.045454
545454545456, 'Synehubov': 0.09090909090909091, 'told': 0.13636363636363635, 'Reuters': 0.22727272727272727, 'standing':
0.045454545454545456, 'vast': 0.045454545454545456, 'central': 0.09090909090909091, 'plaza': 0.045454545454545456, 'Freed
om': 0.045454545454545456, 'Square': 0.045454545454545456, 'region': 0.5454545454545454, 'entire': 0.045454545454545456,
'country': 0.045454545454545456, 'Especially': 0.045454545454545456, '"The': 0.09090909090909091, 'near': 0.04545454545
456, 'missiles': 0.13636363636363635, 'reach': 0.045454545454545456, 'target': 0.09090909090909091, 'minute': 0.0454545
45454545456, 'deployment': 0.045454545454545456, 'precious': 0.045454545454545456, 'defences': 0.13636363636363635, 'U.S.
—made': 0.045454545454545456, 'Patriot': 0.045454545454545456, 'surface': 0.045454545454545456, 'high': 0.045454545454545
456, 'value': 0.045454545454545456, 'targets': 0.09090909090909091, 'airstrikes': 0.045454545454545456, 'cautiously': 0.0
45454545454545456, 'close': 0.09090909090909091, 'enemy': 0.045454545454545456, 'lines': 0.09090909090909091, 'official
s': 0.045454545454545456, 'Ukrainian': 0.18181818181818182, 'President': 0.045454545454545456, 'Volodymyr': 0.04545454545
4545456, 'Zelenskiy': 0.09090909090909091, 'urgently': 0.045454545454545456, 'appealed': 0.045454545454545456, 'supplie
s': 0.045454545454545456, 'West': 0.045454545454545456, 'week': 0.045454545454545456, 'quarter': 0.09090909090909091, 'de
stroyed': 0.09090909090909091, 'accused': 0.045454545454545456, 'seeking': 0.045454545454545456, 'reduce': 0.045454545454
545456, 'home': 0.13636363636363635, '2': 0.045454545454545456, 'million': 0.13636363636363635, 'people': 0.2727272727272
727, 'rubble': 0.045454545454545456, 'clearing': 0.045454545454545456, 'way': 0.045454545454545456, 'troops': 0.045454545
454545456, 'advance': 0.045454545454545456, 'repel': 0.045454545454545456, 'offensive': 0.09090909090909091, 'bombardment
s': 0.045454545454545456, 'come': 0.09090909090909091, 'momentum': 0.045454545454545456, 'battlefield': 0.045454545454545
456, 'shifted': 0.045454545454545456, 'favour': 0.045454545454545456, 'years': 0.09090909090909091, 'launched': 0.1363636
3636363635, 'Feb.': 0.045454545454545456, '2022': 0.09090909090909091, 'invasion': 0.09090909090909091, 'denies': 0.04545
4545454545456, 'targeting': 0.045454545454545456, 'civilians': 0.13636363636363635, 'says': 0.045454545454545456, 'syste
m': 0.045454545454545456, 'legitimate': 0.045454545454545456, 'ministry': 0.045454545454545456, 'immediately': 0.04545454
5454545456, 'respond': 0.045454545454545456, 'request': 0.045454545454545456, 'comment': 0.045454545454545456, 'story':
0.045454545454545456, 'interviewed': 0.045454545454545456, '15': 0.045454545454545456, 'expressed': 0.045454545454545456,
'determination': 0.045454545454545456, 'stay': 0.09090909090909091, 'homes': 0.09090909090909091, 'despite': 0.0454545454
54545456, 'flagged': 0.045454545454545456, 'bleak': 0.045454545454545456, 'situation': 0.045454545454545456, 'power': 0.1
3636363636363635, 'real': 0.045454545454545456, 'concern': 0.045454545454545456, '10': 0.09090909090909091, 'rained': 0.0
45454545454545456, 'Thursday': 0.045454545454545456, 'triggering': 0.045454545454545456, 'emergency': 0.04545454545454545
6, 'blackouts': 0.09090909090909091, '200,000': 0.045454545454545456, 'surrounding': 0.045454545454545456, 'attack': 0.09
090909090909091, 'weeks': 0.045454545454545456, 'prosecutor': 0.045454545454545456, 'Oleksandr': 0.09090909090909091, 'Fi
lchakov': 0.18181818181818182, 'facilities': 0.09090909090909091, 'damaged': 0.045454545454545456, 'renewed': 0.045454545
454545456, 'aerial': 0.045454545454545456, 'assault': 0.045454545454545456, 'causing': 0.045454545454545456, 'large': 0.0
9090909090909091, 'scale': 0.045454545454545456, 'cuts': 0.045454545454545456, 'test': 0.045454545454545456, 'fired': 0.0
45454545454545456, 'new': 0.09090909090909091, 'kind': 0.09090909090909091, 'aircraft': 0.09090909090909091, 'times': 0.0
45454545454545456, 'Tuesday': 0.09090909090909091, 'like': 0.045454545454545456, 'struck': 0.045454545454545456, 'weapo
n': 0.045454545454545456, 'called': 0.045454545454545456, 'unified': 0.045454545454545456, 'multi': 0.045454545454545456,
'purpose': 0.045454545454545456, 'munition': 0.045454545454545456, 'weighs': 0.045454545454545456, '250': 0.0454545454545
45456, 'kilogrammes': 0.045454545454545456, '550': 0.045454545454545456, 'pounds': 0.045454545454545456, 'range': 0.09090
909090909091, '90': 0.045454545454545456, '56': 0.045454545454545456, 'meaning': 0.045454545454545456, 'need': 0.04545454
5454545456, 'risk': 0.09090909090909091, 'getting': 0.045454545454545456, 'bombs': 0.045454545454545456, 'accurate': 0.04
545454545454545456, 'destructive': 0.045454545454545456, 'S—300s': 0.045454545454545456, 'Iskanders': 0.045454545454545456,
'cheaper': 0.045454545454545456, 'produce': 0.045454545454545456, 'mainly': 0.045454545454545456, 'aimed': 0.090909090909
09091, 'intimidating': 0.045454545454545456, 'civilian': 0.13636363636363635, 'population': 0.13636363636363635, 'office
s': 0.045454545454545456, 'trying': 0.13636363636363635, 'buildings': 0.09090909090909091, 'apartments': 0.04545454545454
5456, '...': 0.09090909090909091, 'sow': 0.045454545454545456, 'panic': 0.045454545454545456, '"Strikes': 0.0454545454545
45456, 'shelling': 0.045454545454545456, 'killed': 0.045454545454545456, '97': 0.045454545454545456, 'adding': 0.04545454
5454545456, 'nearly': 0.045454545454545456, 'shaken': 0.045454545454545456, 'suffered': 0.045454545454545456, 'injuries':
0.045454545454545456, 'landed': 0.045454545454545456, 'March': 0.045454545454545456, '27': 0.045454545454545456, 'blowin
g': 0.045454545454545456, 'rows': 0.045454545454545456, 'red': 0.045454545454545456, 'brick': 0.045454545454545456, 'resi
dential': 0.045454545454545456, 'partner': 0.045454545454545456, 'works': 0.045454545454545456, 'courier': 0.045454545454
545456, 'plans': 0.045454545454545456, '"I': 0.045454545454545456, 'imagined': 0.045454545454545456, 'grow': 0.045454545
454545456, 'life': 0.045454545454545456, 'village': 0.045454545454545456, 'study': 0.09090909090909091, 'live': 0.13636363
636363635, 'know': 0.045454545454545456, 'wake': 0.045454545454545456, 'tomorrow': 0.045454545454545456, 'morning': 0.045
454545454545456, 'Velynchuk': 0.045454545454545456, 'hairdresser': 0.045454545454545456, '"But': 0.045454545454545456, 't
ime': 0.045454545454545456, 'want': 0.09090909090909091, 'normal': 0.045454545454545456, 'born': 0.045454545454545456,
'"RUSSIA': 0.045454545454545456, 'GAINS': 0.045454545454545456, 'MOMENTUMAfter': 0.045454545454545456, 'attritional': 0.0
45454545454545456, 'fighting': 0.045454545454545456, 'slowly': 0.045454545454545456, 'advancing': 0.045454545454545456,

'eastern': 0.13636363636363635, 'Donetsk': 0.09090909090909091, 'Kyiv': 0.045454545454545456, 'forces': 0.13636363636363635, 'find': 0.09090909090909091, 'foot': 0.045454545454545456, 'facing': 0.045454545454545456, 'artillery': 0.045454545454545456, 'shells': 0.045454545454545456, 'grappling': 0.045454545454545456, 'manpower': 0.13636363636363635, 'problems': 0.09090909090909091, 'parliament': 0.045454545454545456, 'passed': 0.045454545454545456, 'legislation': 0.045454545454545456, 'Thursdayto': 0.045454545454545456, 'overhaul': 0.045454545454545456, 'armed': 0.045454545454545456, 'draft': 0.09090909090909091, 'civiliansinto': 0.045454545454545456, 'ranks': 0.045454545454545456, 'effort': 0.045454545454545456, 'bolster': 0.045454545454545456, 'frontlines': 0.045454545454545456, 'final': 0.045454545454545456, 'law': 0.09090909090909091, 'excluded': 0.045454545454545456, 'clauses': 0.045454545454545456, 'draconian': 0.045454545454545456, 'penalties': 0.045454545454545456, 'dodging': 0.045454545454545456, 'caused': 0.045454545454545456, 'public': 0.045454545454545456, 'outcry': 0.045454545454545456, 'depend': 0.09090909090909091, 'expected': 0.045454545454545456, 'effect': 0.045454545454545456, 'mid': 0.045454545454545456, 'implemented': 0.045454545454545456, 'analysts': 0.045454545454545456, '"There': 0.09090909090909091, 'issues': 0.045454545454545456, 'ammunition': 0.045454545454545456, 'issue': 0.09090909090909091, 'address': 0.045454545454545456, 'think': 0.13636363636363635, 'hold': 0.045454545454545456, 'advances': 0.045454545454545456, 'Rob': 0.045454545454545456, 'Lee': 0.09090909090909091, 'senior': 0.045454545454545456, 'fellow': 0.045454545454545456, 'Foreign': 0.045454545454545456, 'Policy': 0.045454545454545456, 'Research': 0.045454545454545456, 'Institute': 0.045454545454545456, 'tank': 0.045454545454545456, 'Philadelphia': 0.045454545454545456, '"But': 0.045454545454545456, 'addressed': 0.045454545454545456, 'potential': 0.045454545454545456, 'greater': 0.045454545454545456, 'gains': 0.045454545454545456, 'summer': 0.045454545454545456, '"Ukraine': 0.045454545454545456, 'tried': 0.045454545454545456, 'pressure': 0.09090909090909091, 'point': 0.045454545454545456, 'bombing': 0.045454545454545456, 'oil': 0.09090909090909091, 'far': 0.045454545454545456, 'long': 0.045454545454545456, 'drones': 0.09090909090909091, 'taken': 0.045454545454545456, '14': 0.045454545454545456, 'giant': 0.045454545454545456, 'refining': 0.045454545454545456, 'capacity': 0.045454545454545456, 'according': 0.045454545454545456, 'calculations': 0.045454545454545456, 'inspected': 0.045454545454545456, 'defensive': 0.045454545454545456, 'fortifications': 0.045454545454545456, 'preparing': 0.045454545454545456, 'big': 0.045454545454545456, 'late': 0.045454545454545456, 'June': 0.045454545454545456, 'captured': 0.045454545454545456, 'town': 0.045454545454545456, 'Avdiivka': 0.09090909090909091, 'February': 0.045454545454545456, 'controls': 0.045454545454545456, 'territory': 0.09090909090909091, 'inched': 0.045454545454545456, 'forward': 0.045454545454545456, 'keeping': 0.045454545454545456, 'fronts': 0.045454545454545456, 'west': 0.045454545454545456, 'andthe': 0.045454545454545456, 'Bakhmut': 0.045454545454545456, 'fixed': 0.045454545454545456, 'managed': 0.045454545454545456, 'recruit': 0.045454545454545456, 'number': 0.09090909090909091, 'volunteers': 0.045454545454545456, 'allowing': 0.045454545454545456, 'sustain': 0.045454545454545456, 'losses': 0.045454545454545456, 'assaults': 0.045454545454545456, 'faced': 0.045454545454545456, 'equipment': 0.09090909090909091, 'limitations': 0.045454545454545456, 'problem': 0.045454545454545456, 'decision': 0.045454545454545456, 'looked': 0.09090909090909091, 'weakest': 0.045454545454545456, 'Moscow': 0.045454545454545456, 'likely': 0.045454545454545456, 'maintain': 0.045454545454545456, 'focus': 0.045454545454545456, 'Donbas': 0.045454545454545456, 'Putin': 0.045454545454545456, 'rule': 0.045454545454545456, 'establish': 0.045454545454545456, 'buffer': 0.045454545454545456, 'zone': 0.045454545454545456, 'inside': 0.045454545454545456, 'Kovalenko': 0.045454545454545456, 'independent': 0.045454545454545456, 'analyst': 0.045454545454545456, 'based': 0.045454545454545456, 'Odesa': 0.045454545454545456, 'strikes': 0.045454545454545456, 'setting': 0.045454545454545456, 'motion': 0.045454545454545456, 'plan': 0.045454545454545456, 'scare': 0.045454545454545456, 'laying': 0.045454545454545456, 'possible': 0.045454545454545456, 'operation': 0.045454545454545456, 'later': 0.045454545454545456, 'date': 0.045454545454545456, '"For': 0.045454545454545456, 'seize': 0.045454545454545456, 'medium': 0.045454545454545456, 'term': 0.045454545454545456, 'terrorise': 0.045454545454545456, 'prepare': 0.045454545454545456, 'corresponding': 0.045454545454545456, 'conditions': 0.045454545454545456, '"\'HOW': 0.045454545454545456, "LEAVE?'Kharkiv": 0.045454545454545456, 'industrial': 0.045454545454545456, 'hub': 0.045454545454545456, 'served': 0.09090909090909091, 'capital': 0.045454545454545456, 'jarring': 0.045454545454545456, 'contrast': 0.045454545454545456, '1.3': 0.09090909090909091, 'going': 0.045454545454545456, 'lives': 0.045454545454545456, 'amid': 0.045454545454545456, 'regular': 0.045454545454545456, 'raid': 0.045454545454545456, 'sirens': 0.045454545454545456, 'sound': 0.045454545454545456, 'machine': 0.045454545454545456, 'guns': 0.045454545454545456, 'downing': 0.045454545454545456, 'night': 0.09090909090909091, 'Schools': 0.045454545454545456, 'closed': 0.045454545454545456, 'threat': 0.09090909090909091, 'children': 0.045454545454545456, 'online': 0.09090909090909091, 'authorities': 0.045454545454545456, 'opened': 0.045454545454545456, 'underground': 0.045454545454545456, 'classrooms': 0.045454545454545456, 'metro': 0.045454545454545456, 'station': 0.045454545454545456, 'allow': 0.045454545454545456, 'pupils': 0.045454545454545456, 'classes': 0.045454545454545456, 'person': 0.045454545454545456, 'plunged': 0.045454545454545456, '300,000': 0.045454545454545456, 'recaptured': 0.045454545454545456, 'occupied': 0.045454545454545456, 'offensives': 0.045454545454545456, 'returned': 0.045454545454545456, 'remained': 0.045454545454545456, 'Viktoria': 0.045454545454545456, 'Zaremba': 0.045454545454545456, '37': 0.045454545454545456, 'web': 0.045454545454545456, 'designer': 0.045454545454545456, 'mother': 0.045454545454545456, 'boy': 0.045454545454545456, 'changed': 0.045454545454545456, 'perception': 0.045454545454545456, 'consider': 0.045454545454545456, 'leaving': 0.045454545454545456, 'heating': 0.045454545454545456, 'electricity': 0.045454545454545456, 'winter': 0.045454545454545456, 'looming': 0.045454545454545456, 'occupation': 0.045454545454545456, 'began': 0.045454545454545456, 'increasing': 0.045454545454545456, 'October': 0.045454545454545456, 'rising': 0.045454545454545456, '35': 0.045454545454545456, '130': 0.045454545454545456, '95': 0.045454545454545456, '2023.They': 0.045454545454545456, 'intensified': 0.045454545454545456, 'Rolling': 0.045454545454545456, 'hours': 0.045454545454545456, 'day': 0.045454545454545456, 'traffic': 0.045454545454545456, 'lights': 0.09090909090909091, 'work': 0.045454545454545456, 'Mobile': 0.045454545454545456, 'coverage': 0.045454545454545456, 'patchy': 0.045454545454545456, 'GPS': 0.045454545454545456, 'maps': 0.045454545454545456, 'function': 0.045454545454545456, 'properly': 0.045454545454545456, 'street': 0.045454545454545456, 'regional': 0.045454545454545456, 'governor': 0.045454545454545456, 'sign': 0.045454545454545456, 'abandoning': 0.045454545454545456, "'I\'ll": 0.045454545454545456, 'Borys': 0.045454545454545456, 'Nosov': 0.09090909090909091, '63': 0.045454545454545456, 'pensioner': 0.045454545454545456, 'walking': 0.045454545454545456, 'dog': 0.045454545454545456, 'centre': 0.045454545454545456, 'veteran': 0.045454545454545456, '1979': 0.045454545454545456, '1989': 0.045454545454545456, 'Afghan': 0.045454545454545456, '"This': 0.045454545454545456, 'abandon': 0.045454545454545456, 'Afghanistan': 0.045454545454545456, 'terrifying': 0.045454545454545456, 'okay': 0.045454545454545456, '"Home': 0.045454545454545456, 'Search': 0.045454545454545456}

[Ukraine's air defence shortages leave Kharkiv more exposed to Russian bombsUkraine's air defence shortages leave Kharkiv more exposed to Russian bombs2024-04-12 10:43 -Tom, BalmforthKateryna Velnychuk was having an afternoon nap when an explosion shattered the windows of her ground-floor flat, spraying shrapnel that tore holes through her walls and cupboards., A Russian guided bomb had exploded in the courtyard outside the five-storey Soviet-era building, killing a postman on his rounds., As her flat filled with thick, milky smoke, the 22-year-old turned to see blood pouring from her boyfriend Vladyslav's head., "As we've been living…in a state of war, there was no sense of fear in the moment," Velnychuk said., "You just understand there was an explosion., The only thought in your head is 'I hope we survive'."As, Russia hasintensified its air campaignagainst Ukraine in the last month, hammering its energy infrastructure and urban areas, no major city has been harder hit than Kharkiv., Just 30 km (18 miles) from the Russian border in northeast Ukraine, Kharkiv was already the most exposed to missile attacks and bombardment., But the drying up of Western military support in recent months – as a vital U.S. military aid package has been stuck in Congressamid Republican resistance– has left Kharkiv even more dangerously unprotected., "We have a catastrophic shortage of air defence systems," Governor Oleh Synehubov told Reuters, standing in the city's vast central plaza, Freedom Square., "Not only in the Kharkiv region, but throughout the entire country., Especially in the Kharkiv region., "The city is so near the border that Russian missiles can reach their target in less than a minute., The deployment of Ukraine's precious air defences, such as the U.S.-made Patriot surface-to-air missile systems, which are high-value targets for Russian airstrikes, has to be done more cautiously so close to enemy lines, officials say., Ukrainian President Volodymyr Zelenskiy, who has urgently appealed for more air defence supplies from the West, said this week that almost a quarter of Kharkiv had been destroyed., He accused Russia of seeking to reduce the city –

which was home to 2 million people before the war — to rubble, clearing the way for its troops to advance., He said Ukraine's military would repel any such offensive., The bombardments come as the momentum on the battlefield has shifted in Russia's favour, more than two years since it launched its Feb. 2022 invasion., Russia denies targeting civilians and says Ukraine's energy system is a legitimate military target., The Russian defence ministry did not immediately respond to a request for comment for this story., Reuters interviewed 15 civilians in Kharkiv who expressed their determination to stay in their homes despite the attacks — though two of them flagged the bleak situation on the power front as a real concern., At least 10 missiles rained down on Kharkiv on Thursday, triggering emergency blackouts for 200,000 people in the surrounding region, as Russia launched its third major air attack on energy infrastructure across Ukraine in recent weeks., The region's top prosecutor Oleksandr Filchakov told Reuters that all of the Kharkiv region's power facilities have been damaged or destroyed since Russia renewed its aerial assault last month, causing large-scale power cuts., Russia had test-fired a new kind of aircraft-launched guided bomb at least six times as of Tuesday, he said, like the one that struck the courtyard outside Velnychuk's home., The weapon, which Filchakov called a "unified multi-purpose guided munition", weighs just 250 kilogrammes (550 pounds) and has a range of 90 km (56 miles), meaning aircraft have no need to risk getting close to the city's defences., While the guided bombs are less accurate and destructive than other missiles used by Russia, such as S-300s and Iskanders, they are much cheaper for Russia to produce, he said., "The (attacks) are mainly aimed at intimidating the civilian population," Filchakov told Reuters in his offices., "They're trying to make people leave the city, leave their buildings, homes, apartments...To sow panic in the city., "Strikes and shelling have killed 97 civilians in the region this year, he said, adding that nearly all the recent attacks had hit civilian targets., Velnychuk was shaken but suffered no major injuries from the guided bomb that landed outside her building on March 27, blowing out all the windows along two rows of the red-brick residential buildings., But both she and her partner, who works as a courier, said they had no plans to leave the city., "I always imagined I would grow up and have some kind of life, move from the village to the city, study., Now I live, and ..., I don't even know if I'll wake up tomorrow morning," said Velynchuk, a hairdresser., "But, at the same time, you want to live in your own home., It's normal to want to live where you were born., "RUSSIA GAINS MOMENTUMAfter months of attritional fighting, Russia is slowly advancing in Ukraine's eastern region of Donetsk this year., Kyiv's forces find themselves on the back foot, facing shortages of artillery shells and air defences, and grappling with manpower problems., Ukraine's parliament passed legislation on Thursdayto overhaul how the armed forces draft civiliansinto the ranks, in an effort to bolster the frontlines., The final law, however, excluded clauses on draconian penalties for draft dodging that had caused public outcry., Much would depend on how well the new law — expected to take effect in mid-May — was implemented, analysts said., "There are two issues now: the ammunition issue and the manpower issue., If they address them, I think Ukraine can hold back Russian advances," said Rob Lee, a senior fellow at the Foreign Policy Research Institute, a think tank in Philadelphia., "But if they don't get addressed, there is a potential that Russia will make greater gains this summer., "Ukraine has tried to find a pressure point against Russia by bombing oil facilities far behind the front lines using long-range drones that have taken out 14% of energy giant Russia's oil refining capacity, according to Reuters calculations., Zelenskiy, who inspected Ukrainian defensive fortifications in the Kharkiv region on Tuesday, has said Russia may be preparing a big offensive in late May or in June., He did not say where., Russia, which captured the eastern town of Avdiivka in February and controls 18% of Ukrainian territory, has inched forward in the Donetsk region, keeping up pressure on the fronts west of Avdiivka andthe city of Bakhmut., Lee said Russia had fixed its own manpower problems and managed to recruit a large number of volunteers, allowing it to sustain losses in assaults, but that it faced equipment limitations that could become a problem next year., Russia's decision on where to attack, he said, would in part depend on where it thought Ukraine looked weakest, although Moscow would likely maintain its focus on the eastern Donbas region., Putin said last month he did not rule out Russia trying to establish a buffer zone inside Ukrainian territory along the Russian border., Oleksandr Kovalenko, an independent military analyst based in Odesa, said the strikes on Kharkiv looked aimed at setting in motion such a plan by trying to scare people to leave the city, laying the ground for a possible ground operation at a later date., "For the moment, Russia does not have the forces and equipment to seize the city, but in the medium term they can terrorise the civilian population to prepare the corresponding conditions., "'HOW, COULD I LEAVE?'Kharkiv, an industrial hub that once served as the capital of Soviet Ukraine, is a jarring contrast of 1.3 million people going about their lives amid regular air raid sirens and the sound of machine guns downing drones at night., Schools in the city have been closed because of the threat of attacks and children study online., But authorities have opened underground classrooms in a metro station to allow some pupils to come to classes in person., The city's population plunged to 300,000 after the invasion but, after Ukraine recaptured occupied areas of the region in two military offensives in 2022, it returned to around 1.3 million, where it has remained since., Viktoria Zaremba, 37, a web designer and mother of a 10-year-old boy, said more than two years of war had changed her perception of risk., "There is no fear," she said., She would only consider leaving Kharkiv if there were no central heating or electricity this winter, or a looming threat of occupation., The number of attacks on the city and region began increasing in October, Filchakov said, rising more than 35% in the first three months this year to 130 from 95 in the last quarter of 2023.They have intensified again this month, he said., Rolling blackouts last up to 12 hours a day and the traffic lights don't work., Mobile coverage is patchy, online GPS maps don't function properly, and the street lights stay off at night., But Synehubov, the regional governor, said there was no sign people were abandoning the city., "I'll never leave," said Borys Nosov, 63, a pensioner walking his dog in the city centre., Nosov said he was a veteran of the 1979-1989 Soviet-Afghan war., "This is my city., How could I leave and abandon it?, I served in Afghanistan., That was terrifying., I think everything will be okay., "Home-Search-About]

- Print the word frequencies and generate sentence tokens

```
In [18]:  sentence_scores = {}
          for sent in sentence_tokens:
              for word in sent:
                  if word.text.lower() in word_frequencies.keys():
                      if sent not in sentence_scores.keys():
                          sentence_scores[sent] = word_frequencies[word.text.lower()]
                      else:
                          sentence_scores[sent] += word_frequencies[word.text.lower()]
          sentence_scores
```

Out[18]: {Ukraine's air defence shortages leave Kharkiv more exposed to Russian bombsUkraine's air defence shortages leave Kharkiv more exposed to Russian bombs2024-04-12 10:43 -Tom: 2.8636363636363633,

 BalmforthKateryna Velnychuk was having an afternoon nap when an explosion shattered the windows of her ground-floor flat, spraying shrapnel that tore holes through her walls and cupboards.: 0.9090909090909088,

 A Russian guided bomb had exploded in the courtyard outside the five-storey Soviet-era building, killing a postman on his rounds.: 0.9545454545454544,

 As her flat filled with thick, milky smoke, the 22-year-old turned to see blood pouring from her boyfriend Vladyslav's head.: 0.9545454545454545,

 "As we've been living…in a state of war, there was no sense of fear in the moment," Velnychuk said.: 1.6363636363636365,
 "You just understand there was an explosion.: 0.22727272727272727,

 The only thought in your head is 'I hope we survive'."As: 0.31818181818181823,

 Russia hasintensified its air campaignagainst Ukraine in the last month, hammering its energy infrastructure and urban areas, no major city has been harder hit than Kharkiv.: 2.2727272727272725,

 Just 30 km (18 miles) from the Russian border in northeast Ukraine, Kharkiv was already the most exposed to missile attacks and bombardment.: 1.0454545454545454,

 But the drying up of Western military support in recent months — as a vital U.S. military aid package has been stuck in Congressamid Republican resistance— has left Kharkiv even more dangerously unprotected.: 1.545454545454545,

 "We have a catastrophic shortage of air defence systems," Governor Oleh Synehubov told Reuters, standing in the city's vast central plaza, Freedom Square.: 2.090909090909091,

 "Not only in the Kharkiv region, but throughout the entire country.: 0.6363636363636362,

 Especially in the Kharkiv region.: 0.5454545454545454,

 "The city is so near the border that Russian missiles can reach their target in less than a minute.: 1.318181818181818,

 The deployment of Ukraine's precious air defences, such as the U.S.-made Patriot surface-to-air missile systems, which are high-value targets for Russian airstrikes, has to be done more cautiously so close to enemy lines, officials say.: 1.9090909090909083,

 Ukrainian President Volodymyr Zelenskiy, who has urgently appealed for more air defence supplies from the West, said this week that almost a quarter of Kharkiv had been destroyed.: 2.090909090909091,

 He accused Russia of seeking to reduce the city — which was home to 2 million people before the war — to rubble, clearing the way for its troops to advance.: 1.9545454545454541,

 He said Ukraine's military would repel any such offensive.: 1.409090909090909,

 The bombardments come as the momentum on the battlefield has shifted in Russia's favour, more than two years since it launched its Feb. 2022 invasion.: 0.7272727272727274,

 Russia denies targeting civilians and says Ukraine's energy system is a legitimate military target.: 0.9090909090909091,
 The Russian defence ministry did not immediately respond to a request for comment for this story.: 0.5,

 Reuters interviewed 15 civilians in Kharkiv who expressed their determination to stay in their homes despite the attacks — though two of them flagged the bleak situation on the power front as a real concern.: 1.1818181818181817,

 At least 10 missiles rained down on Kharkiv on Thursday, triggering emergency blackouts for 200,000 people in the surrounding region, as Russia launched its third major air attack on energy infrastructure across Ukraine in recent weeks.: 2.636363636363636,

 The region's top prosecutor Oleksandr Filchakov told Reuters that all of the Kharkiv region's power facilities have been damaged or destroyed since Russia renewed its aerial assault last month, causing large-scale power cuts.: 2.3181818181818175,

 Russia had test-fired a new kind of aircraft-launched guided bomb at least six times as of Tuesday, he said, like the one that struck the courtyard outside Velnychuk's home.: 2.3636363636363633,

 The weapon, which Filchakov called a "unified multi-purpose guided munition", weighs just 250 kilogrammes (550 pounds) and has a range of 90 km (56 miles), meaning aircraft have no need to risk getting close to the city's defences.: 2.636363636363636,

 While the guided bombs are less accurate and destructive than other missiles used by Russia, such as S-300s and Iskanders, they are much cheaper for Russia to produce, he said.: 1.5909090909090908,

 "The (attacks) are mainly aimed at intimidating the civilian population," Filchakov told Reuters in his offices.: 0.909090909090909,

 "They're trying to make people leave the city, leave their buildings, homes, apartments...To sow panic in the city.: 3.1818181818181817,

 "Strikes and shelling have killed 97 civilians in the region this year, he said, adding that nearly all the recent attacks had hit civilian targets.: 2.9090909090909087,

 Velnychuk was shaken but suffered no major injuries from the guided bomb that landed outside her building on March 27, blowing out all the windows along two rows of the red-brick residential buildings.: 1.3636363636363633,

 But both she and her partner, who works as a courier, said they had no plans to leave the city.: 2.3636363636363638,

 "I always imagined I would grow up and have some kind of life, move from the village to the city, study.: 1.1818181818181819,

 Now I live: 0.13636363636363635,

 and ...: 0.09090909090909091,

 I don't even know if I'll wake up tomorrow morning," said Velynchuk, a hairdresser.: 1.2272727272727273,

 "But, at the same time, you want to live in your own home.: 0.40909090909090906,

 It's normal to want to live where you were born.: 0.3181818181818182,

 "RUSSIA GAINS MOMENTUMAfter months of attritional fighting, Russia is slowly advancing in Ukraine's eastern region of Donetsk this year.: 1.3181818181818181,

 Kyiv's forces find themselves on the back foot, facing shortages of artillery shells and air defences, and grappling with manpower problems.: 1.4090909090909087,

 Ukraine's parliament passed legislation on Thursdayto overhaul how the armed forces draft civiliansinto the ranks, in an effort to bolster the frontlines.: 0.6818181818181817,

 The final law, however, excluded clauses on draconian penalties for draft dodging that had caused public outcry.: 0.5909090909090909,

 Much would depend on how well the new law — expected to take effect in mid-May — was implemented, analysts said.: 1.6363636363636362,

 "There are two issues now: the ammunition issue and the manpower issue.: 0.40909090909090906,

 If they address them, I think Ukraine can hold back Russian advances," said Rob Lee, a senior fellow at the Foreign Policy Research Institute, a think tank in Philadelphia.: 1.5454545454545454,

 "But if they don't get addressed, there is a potential that Russia will make greater gains this summer.: 0.2272727272727273,

 "Ukraine has tried to find a pressure point against Russia by bombing oil facilities far behind the front lines using long-range drones that have taken out 14% of energy giant Russia's oil refining capacity, according to Reuters calculations.: 1.4545454545454541,

 Zelenskiy, who inspected Ukrainian defensive fortifications in the Kharkiv region on Tuesday, has said Russia may be preparing a big offensive in late May or in June.: 1.9090909090909087,

 Russia, which captured the eastern town of Avdiivka in February and controls 18% of Ukrainian territory, has inched forward in the Donetsk region, keeping up pressure on the fronts west of Avdiivka andthe city of Bakhmut.: 2.181818181818181

7,
 Lee said Russia had fixed its own manpower problems and managed to recruit a large number of volunteers, allowing it to sustain losses in assaults, but that it faced equipment limitations that could become a problem next year.: 2.27272727272 72716,
 Russia's decision on where to attack, he said, would in part depend on where it thought Ukraine looked weakest, although Moscow would likely maintain its focus on the eastern Donbas region.: 2.272727272727272,
 Putin said last month he did not rule out Russia trying to establish a buffer zone inside Ukrainian territory along the Russian border.: 1.7727272727272725,
 Oleksandr Kovalenko, an independent military analyst based in Odesa, said the strikes on Kharkiv looked aimed at setting in motion such a plan by trying to scare people to leave the city, laying the ground for a possible ground operation at a later date.: 3.9090909090909083,
 "For the moment, Russia does not have the forces and equipment to seize the city, but in the medium term they can terror ise the civilian population to prepare the corresponding conditions.: 1.727272727272727,
 COULD I LEAVE?'Kharkiv, an industrial hub that once served as the capital of Soviet Ukraine, is a jarring contrast of 1. 3 million people going about their lives amid regular air raid sirens and the sound of machine guns downing drones at nig ht.: 1.9090909090909085,
 Schools in the city have been closed because of the threat of attacks and children study online.: 1.4545454545454544,
 But authorities have opened underground classrooms in a metro station to allow some pupils to come to classes in perso n.: 0.5454545454545454,
 The city's population plunged to 300,000 after the invasion but, after Ukraine recaptured occupied areas of the region i n two military offensives in 2022, it returned to around 1.3 million, where it has remained since.: 2.59090909090909,
 Viktoria Zaremba, 37, a web designer and mother of a 10-year-old boy, said more than two years of war had changed her pe rception of risk.: 2.1363636363636362,
 "There is no fear," she said.: 1.0909090909090908,
 She would only consider leaving Kharkiv if there were no central heating or electricity this winter, or a looming threat of occupation.: 0.5,
 The number of attacks on the city and region began increasing in October, Filchakov said, rising more than 35% in the fi rst three months this year to 130 from 95 in the last quarter of 2023.They have intensified again this month, he said.: 4.727272727272727,
 Rolling blackouts last up to 12 hours a day and the traffic lights don't work.: 0.45454545454545464,
 Mobile coverage is patchy, online GPS maps don't function properly, and the street lights stay off at night.: 0.63636363 63636365,
 But Synehubov, the regional governor, said there was no sign people were abandoning the city.: 2.2727272727272725,
 "I'll never leave," said Borys Nosov, 63, a pensioner walking his dog in the city centre.: 2.409090909090909,
 Nosov said he was a veteran of the 1979-1989 Soviet-Afghan war.: 1.3181818181818181,
 "This is my city.: 0.8181818181818182,
 How could I leave and abandon it?: 0.4090909090909091,
 I served in Afghanistan.: 0.09090909090909091,
 That was terrifying.: 0.045454545454545456,
 I think everything will be okay.: 0.18181818181818182}

- Calculate the sentence score of each sentences

```
In [19]:   from heapq import nlargest
           select_length = int(len(sentence_tokens) * 0.3)
           select_length
           summary = nlargest(select_length, sentence_scores, key=sentence_scores.get)
           summary
           final_summary = [word.text for word in summary]
           summary = ' '.join(final_summary)
           print(summary)
```

The number of attacks on the city and region began increasing in October, Filchakov said, rising more than 35% in the fir st three months this year to 130 from 95 in the last quarter of 2023.They have intensified again this month, he said. Ole ksandr Kovalenko, an independent military analyst based in Odesa, said the strikes on Kharkiv looked aimed at setting in motion such a plan by trying to scare people to leave the city, laying the ground for a possible ground operation at a la ter date. "They're trying to make people leave the city, leave their buildings, homes, apartments...To sow panic in the c ity. "Strikes and shelling have killed 97 civilians in the region this year, he said, adding that nearly all the recent a ttacks had hit civilian targets. Ukraine's air defence shortages leave Kharkiv more exposed to Russian bombsUkraine's air defence shortages leave Kharkiv more exposed to Russian bombs2024-04-12 10:43 -Tom At least 10 missiles rained down on Kh arkiv on Thursday, triggering emergency blackouts for 200,000 people in the surrounding region, as Russia launched its th ird major air attack on energy infrastructure across Ukraine in recent weeks. The weapon, which Filchakov called a "unifi ed multi-purpose guided munition", weighs just 250 kilogrammes (550 pounds) and has a range of 90 km (56 miles), meaning aircraft have no need to risk getting close to the city's defences. The city's population plunged to 300,000 after the in vasion but, after Ukraine recaptured occupied areas of the region in two military offensives in 2022, it returned to arou nd 1.3 million, where it has remained since. "I'll never leave," said Borys Nosov, 63, a pensioner walking his dog in the city centre. But both she and her partner, who works as a courier, said they had no plans to leave the city. Russia had t est-fired a new kind of aircraft-launched guided bomb at least six times as of Tuesday, he said, like the one that struck the courtyard outside Velnychuk's home. The region's top prosecutor Oleksandr Filchakov told Reuters that all of the Khar kiv region's power facilities have been damaged or destroyed since Russia renewed its aerial assault last month, causing large-scale power cuts. Russia hasintensified its air campaignagainst Ukraine in the last month, hammering its energy inf rastructure and urban areas, no major city has been harder hit than Kharkiv. But Synehubov, the regional governor, said t here was no sign people were abandoning the city. Russia's decision on where to attack, he said, would in part depend on where it thought Ukraine looked weakest, although Moscow would likely maintain its focus on the eastern Donbas region. Le e said Russia had fixed its own manpower problems and managed to recruit a large number of volunteers, allowing it to sus tain losses in assaults, but that it faced equipment limitations that could become a problem next year. Russia, which cap tured the eastern town of Avdiivka in February and controls 18% of Ukrainian territory, has inched forward in the Donetsk region, keeping up pressure on the fronts west of Avdiivka andthe city of Bakhmut. Viktoria Zaremba, 37, a web designer a nd mother of a 10-year-old boy, said more than two years of war had changed her perception of risk. "We have a catastroph ic shortage of air defence systems," Governor Oleh Synehubov told Reuters, standing in the city's vast central plaza, Fre edom Square. Ukrainian President Volodymyr Zelenskiy, who has urgently appealed for more air defence supplies from the We st, said this week that almost a quarter of Kharkiv had been destroyed. He accused Russia of seeking to reduce the city — which was home to 2 million people before the war — to rubble, clearing the way for its troops to advance. Zelenskiy, who inspected Ukrainian defensive fortifications in the Kharkiv region on Tuesday, has said Russia may be preparing a big off ensive in late May or in June.

- Generate the summary which is 30% as long as the original text.

```python
In [20]:  # import SentimentIntensityAnalyzer class
          # from vaderSentiment.vaderSentiment module.
          from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer

          # function to print sentiments
          # of the sentence.
          def sentiment_scores(sentence):

              # Create a SentimentIntensityAnalyzer object.
              sid_obj = SentimentIntensityAnalyzer()

              # polarity_scores method of SentimentIntensityAnalyzer
              # object gives a sentiment dictionary.
              # which contains pos, neg, neu, and compound scores.
              sentiment_dict = sid_obj.polarity_scores(sentence)

              print("Overall sentiment dictionary is : ", sentiment_dict)
              print("sentence was rated as ", sentiment_dict['neg']*100, "% Negative")
              print("sentence was rated as ", sentiment_dict['pos']*100, "% Positive")

              print("Sentence Overall Rated As", end = " ")

              # decide sentiment as positive, negative and neutral
              if sentiment_dict['compound'] >= 0.05 :
                  print("Positive")

              elif sentiment_dict['compound'] <= - 0.05 :
                  print("Negative")
              else:
                  print("Neutral")



          # Driver code
          if __name__ == "__main__" :
              print("Original text:")
              sentiment_scores(text)
              print()
              print("Summarized text")
              sentiment_scores(summary)
```

```
Original text:
Overall sentiment dictionary is :  {'neg': 0.143, 'neu': 0.812, 'pos': 0.045, 'compound': -0.9995}
sentence was rated as  14.299999999999999 % Negative
sentence was rated as  4.5 % Positive
Sentence Overall Rated As Negative

Summarized text
Overall sentiment dictionary is :  {'neg': 0.176, 'neu': 0.786, 'pos': 0.038, 'compound': -0.9983}
sentence was rated as  17.599999999999998 % Negative
sentence was rated as  3.8 % Positive
Sentence Overall Rated As Negative
```

- The above is the sentiment analysis of orignal and summarized text.
- You can see that both the orignal and sumarizes text gives a negative overall rating.
- The negative score and positive score of both are almost the same in orignal and summarized text which means the summarization is been done correctly.
- The negative score of original text is 14.3% and summarized text is 17.6%.

Tutorial 2

What is ROUGE score ?:

Recall-Oriented Understudy for Gisting Evaluation, is a set of metrics used to evaluate the quality of text summarization and machine translation systems. It measures the similarity between the generated summary (candidate) and one or more reference summaries (manually created by humans). The goal of ROUGE is to provide a quantitative measure of how well a summarization system captures the important information from the source text.

ROUGE scores are widely used in natural language processing research and are often reported alongside other evaluation metrics to assess the performance of summarization systems objectively. They provide a standardized way to compare different systems and algorithms based on their ability to generate summaries that are similar to human-written references.

$$RECALL = \frac{Overlapping\ number\ of\ n-grams}{Number\ of\ n-grams\ in\ the\ reference}$$

$$PRECISION = \frac{Overlapping\ number\ of\ n-grams}{Number\ of\ n-grams\ in\ the\ candidate}$$

Recall and Precision Equations

To finalize calculation we also need to calculate F1 scores (Harmonic mean) :

$$F1 = \frac{2 \times Precision \times Recall}{(Precision + Recall)}$$

```
In [21]:  Canditate1 = "Summarization is cool"
          reference2 = "Summarization is beneficial and cool"
          reference3 = "Summarization saves time"
```

Rouge 1

Consider the first candidate and the reference set :

Overlapping words(unigrams) for reference 1 is more than reference 2. Basically we won't make any calculations based on reference 2 for this candidate

```
In [22]:  recall = 3 / 5   # Recall = 0.6
          precision = 3 / 3   # Precision = 1
          Rouge_1_1 = 2 * recall * precision / (recall + precision)   # ROUGE-1.1 score
          print("ROUGE-1 score for Candidate 1 :", Rouge_1_1)

          ROUGE-1 score for Candidate 1 : 0.7499999999999999
```

Recall

- 3 in the numerator means 3 tokens matches from reference2 in Candidaet1 which is "Summarization", "is", "cool"
- 5 in denominator means 5 tokens are there in reference2. ### Precision
- Means it has to match all the tokens in the candidate. ### Rouge
- Harmonica means for recall and precision.

Rouge1 score for candidate 1 is 0.75 we have to consider other candidate as well and calculate the mean of the each candidate's rouge scores.

```
In [23]:  Canditate2 = "I love Machine Learning"
          bestreference = "I think i love Machine Learning"
```

```
In [24]:  recall = 4 / 6
          precision = 4/4
          Rouge_1_2 = 2* recall * precision / (recall + precision)
          print("ROUGE-1 score for Candidate 2:", Rouge_1_2)

          ROUGE-1 score for Candidate 2: 0.8
```

```
In [25]:  Canditate3 = "Good night"
          bestreference = "Good night everyone!"
```

```
In [26]:  recall = 2/3
          precision = 2/2
          Rouge_1_3 = 2* recall * precision / (recall + precision)
          print("ROUGE-1 score for Candidate 3:", Rouge_1_3)

          ROUGE-1 score for Candidate 3: 0.8
```

Mean of the F1 scores will gives us the full ROUGE-1 score for dataset

```
In [27]: mean_rouge = (Rouge_1_1 + Rouge_1_2 + Rouge_1_3) / 3
         print("Total ROUGE score:", mean_rouge)

         Total ROUGE score: 0.7833333333333332
```

- The above is overall rouge score for while data.

Question 2.B

Create a code by using the above tutorial Candidate reference set to calculate the total ROUGE-2 score.

Explain in detail the differences between ROUGE-1 and ROUGE-2 as well.

```
In [28]: # code here
         Canditate1 = "Summarization is cool"
         reference2 = "Summarization is beneficial and cool"
```

```
In [29]: recall = 1 / 4
         precision = 1 / 2
         Rouge_2_1 = 2 * recall * precision / (recall + precision)
         print("ROUGE-2 score for Candidate 1 :", Rouge_2_1)

         ROUGE-2 score for Candidate 1 : 0.3333333333333333
```

- In case of ROUGE-2 we need to consider 2 word combination.
- In the above example The candidate and referance have only 1 two word combination matching which are "Summarization is"

```
In [30]: Canditate2 = "I love Machine Learning"
         bestreference = "I think i love Machine Learning"
```

```
In [31]: recall = 2/5
         precision = 2/3
         Rouge_2_2 = 2* recall * precision / (recall + precision)
         print("ROUGE-2 score for Candidate 2:", Rouge_2_2)

         ROUGE-2 score for Candidate 2: 0.5
```

```
In [32]: Canditate3 = "Good night"
         bestreference = "Good night everyone!"
```

```
In [33]: recall = 1/2
         precision = 1/1
         Rouge_2_3 = 2* recall * precision / (recall + precision)
         print("ROUGE-2 score for Candidate 3:", Rouge_2_3)

         ROUGE-2 score for Candidate 3: 0.6666666666666666
```

```
In [34]: mean_rouge = (Rouge_2_1 + Rouge_2_2 + Rouge_2_3) / 3
         print("Total ROUGE score:", mean_rouge)

         Total ROUGE score: 0.5
```

- Rouge 1 considers uni graph which means single words, where as Rouge 2 considers bi-gram words.
- Rouge 2 is the ration of number of overlapping bigrams over total number of bigrams in referance summaries
- In Rouge 1 precision is calculated as the ratio of number of uni-grapms in referance that also appears in candidate by the number of uni-grapms in Referance.
- The Rouge 2 precision is calculated as the ratio of number of bi-grapms in referance that also appears in candidate by the number of bi-grapms in Referance.
- In Rouge 1 precision is calculated as the ratio of number of uni-grapms in referance that also appears in candidate by the number of uni-grapms in Candidate.
- In Rouge 2 precision is calculated as the ratio of number of bi-grapms in referance that also appears in candidate by the number of bi-grapms in Candidate.

References

https://www.machinelearningplus.com/nlp/text-summarization-approaches-nlp-example/

https://www.datacamp.com/tutorial/simplifying-sentiment-analysis-python