

```
In [1]: d1 = '''Formula One, commonly known as Formula 1 or F1, is the highest class of international motorsport. A point-system is used at Grands Prix to determine two annual World Championships: one for drivers and one for constructors.'''
```

```
In [2]: d2 = '''On the track, the McLaren and Williams teams dominated the 1980s and 1990s. Brabham and Williams won the 1980 Constructors' Championship, while Williams won the 1996 Constructors' Championship.'''
```

```
In [3]: d3 = '''A major rule shake-up in 2014 saw the 2.4-litre naturally aspirated V8 engines replaced by 1.6-litre turbocharged engines.'''
```

```
In [4]: d4 = '''The race begins with a warm-up lap, after which the cars assemble on the starting grid. The race is run over 70 laps, with a safety car period in the middle.'''
```

```
In [5]: d5 = '''A Formula One constructor is the entity credited for designing the chassis and the engine, and for entering the cars and drivers in the race.'''
```

```
In [6]: d6 = '''The use of volunteers was integral in making and maintaining Wikipedia. However, even today, the project relies on the dedication of its community.'''
```

```
In [18]: import nltk
from nltk.corpus import stopwords
from collections import Counter
from nltk.tokenize import word_tokenize, sent_tokenize
```

```
In [9]: documents = [d1, d2, d3, d4, d5, d6]
```

```
In [26]: # Function to calculate TF scores for words in a paragraph
def calculate_tf(paragraph):
    words = word_tokenize(paragraph.lower())
    word_counts = Counter(words)
    total_words = len(words)
    return {word: count / total_words for word, count in word_counts.items()}
```

```
In [27]: # Function to find sentences containing the top word with the highest TF score
def find_sentences_with_top_word(paragraph, top_word):
    sentences_containing_top_word = []
    for sentence in sent_tokenize(paragraph):
        if top_word in word_tokenize(sentence.lower()):
            sentences_containing_top_word.append(sentence)
    return sentences_containing_top_word
```

```
In [28]: # Iterate through each paragraph
for idx, paragraph in enumerate(documents):
    # Calculate TF scores for words in the paragraph
    tf_scores = calculate_tf(paragraph)
    # Find the word with the highest TF score
    top_word = max(tf_scores, key=tf_scores.get)
    # Find sentences containing the top word
    sentences_with_top_word = find_sentences_with_top_word(paragraph, top_word)
    # Print summary and sentences containing the top word
    print(f"Paragraph {idx + 1}:")
    print("The word with the highest TF score is:", top_word)
    print("Sentences containing the top word:")
    for sentence in sentences_with_top_word:
        print("-", sentence)
    print()
```

Paragraph 1:

The word with the highest TF score is: the

Sentences containing the top word:

- Formula One, commonly known as Formula 1 or F1, is the highest class of international racing for open-wheel single-seater formula racing cars sanctioned by the Fédération Internationale de l'Automobile (FIA).
- The FIA Formula One World Championship has been one of the premier forms of racing around the world since its inaugural running in 1950.
- The word formula in the name refers to the set of rules to which all participants' cars must conform.
- Grands Prix take place in multiple countries and continents around the world on either purpose-built circuits or closed public roads.
- A point-system is used at Grands Prix to determine two annual World Championships: one for the drivers, and one for the constructors (the teams).
- Each driver must hold a valid Super Licence, the highest class of racing licence issued by the FIA, and the races must be held on grade one tracks, the highest grade-rating issued by the FIA for tracks.

Paragraph 2:

The word with the highest TF score is: the

Sentences containing the top word:

- On the track, the McLaren and Williams teams dominated the 1980s and 1990s.
- Brabham were also being competitive during the early part of the 1980s, winning two Drivers' Championships with Nelson Piquet.
- The rivalry between racers Ayrton Senna and Alain Prost became F1's central focus during 1988 and continued until Prost retired at the end of 1993.
- Senna died at the 1994 San Marino Grand Prix after crashing into a wall on the exit of the notorious curve Tamburello.
- The FIA worked to improve the sport's safety standards since that weekend, during which Roland Ratzenberger also died in an accident during Saturday qualifying.
- No driver died of injuries sustained on the track at the wheel of a Formula One car for 20 years until the 2014 Japanese Grand Prix, where Jules Bianchi collided with a recovery vehicle after aquaplaning off the circuit, dying nine months later from his injuries.
- Since 1994, three track marshals have died, one at the 2000 Italian Grand Prix,[31] the second at the 2001 Australian Grand Prix[31] and the third at the 2013 Canadian Grand Prix.

Paragraph 3:

The word with the highest TF score is: the

Sentences containing the top word:

- A major rule shake-up in 2014 saw the 2.4-litre naturally aspirated V8 engines replaced by 1.6-litre turbocharged hybrid power units.
- This prompted Honda to return to the sport in 2015 as the championship's fourth power unit manufacturer.
- Mercedes emerged as the dominant force after the rule shake-up, with Lewis Hamilton winning the championship closely followed by his main rival and teammate, Nico Rosberg, with the team winning 16 out of the 19 races that season.
- The team continued this form in the following two seasons, again winning 16 races in 2015 before taking a record 19 wins in 2016, with Hamilton claiming the title in the former year and Rosberg winning it in the latter by five points.
- The 2016 season also saw a new team, Haas, join the grid, while Max Verstappen became the youngest-ever race winner at the age of 18 in Spain.

Paragraph 4:

The word with the highest TF score is: the

Sentences containing the top word:

- The race begins with a warm-up lap, after which the cars assemble on the starting grid in the order they qualified.
- This lap is often referred to as the formation lap, as the cars lap in formation with no overtaking (although a driver who makes a mistake may regain lost ground).
- The warm-up lap allows drivers to check the condition of the track and their car, gives the tyres a chance to warm up to increase traction and grip, and also gives the pit crews time to clear themselves and their equipment from the grid for the race start.

Paragraph 5:

The word with the highest TF score is: the

Sentences containing the top word:

- A Formula One constructor is the entity credited for designing the chassis and the engine.
- [97] If both are designed by the same company, that company receives sole credit as the constructor (e.g., Ferrari).
- If they are designed by different companies, both are credited, and the name of the chassis

sis designer is placed before that of the engine designer (e.g., McLaren-Mercedes).

Paragraph 6:

The word with the highest TF score is: the

Sentences containing the top word:

- The use of volunteers was integral in making and maintaining Wikipedia.
- However, even without the internet, huge complex projects of similar nature had made use of volunteers.
- Specifically, the creation of the Oxford English Dictionary was conceived with the speech at the London Library, on Guy Fawkes Day, 5 November 1857, by Richard Chenevix Trench.
- Dr. Trench envisioned a grand new dictionary of every word in the English language, and to be used democratically and freely.
- According to author Simon Winchester, "The undertaking of the scheme, he said, was beyond the ability of any one man.
- To peruse all of English literature - and to comb the London and New York newspapers and the most literate of the magazines and journals - must be instead 'the combined action of many.'

```
In [30]: # Concatenate all paragraphs into a single string
document = ' '.join(documents)

# Function to calculate TF scores for words in the document
def calculate_tf(document):
    words = word_tokenize(document.lower())
    word_counts = Counter(words)
    total_words = len(words)
    return {word: count / total_words for word, count in word_counts.items()}

# Function to find sentences containing the top word with the highest TF score
def find_sentences_with_top_word(document, top_word):
    sentences_containing_top_word = []
    for sentence in sent_tokenize(document):
        if top_word in word_tokenize(sentence.lower()):
            sentences_containing_top_word.append(sentence)
    return sentences_containing_top_word

# Calculate TF scores for words in the document
tf_scores = calculate_tf(document)

# Find the word with the highest TF score
top_word = max(tf_scores, key=tf_scores.get)

# Find sentences containing the top word
sentences_with_top_word = find_sentences_with_top_word(document, top_word)

# Print summary and sentences containing the top word
print("Summary:")
print("The word with the highest TF score is:", top_word)
print("Sentences containing the top word:")
for sentence in sentences_with_top_word:
    print("-", sentence)
```

Summary:

The word with the highest TF score is: the

Sentences containing the top word:

- Formula One, commonly known as Formula 1 or F1, is the highest class of international racing for open-wheel single-seater formula racing cars sanctioned by the Fédération Internationale de l'Automobile (FIA).
- The FIA Formula One World Championship has been one of the premier forms of racing around the world since its inaugural running in 1950.
- The word formula in the name refers to the set of rules to which all participants' cars must conform.
- Grands Prix take place in multiple countries and continents around the world on either purpose-built circuits or closed public roads.
- A point-system is used at Grands Prix to determine two annual World Championships: one for the drivers, and one for the constructors (the teams).
- Each driver must hold a valid Super Licence, the highest class of racing licence issued by the FIA, and the races must be held on grade one tracks, the highest grade-rating issued by the FIA for tracks.
- On the track, the McLaren and Williams teams dominated the 1980s and 1990s.
- Brabham were also being competitive during the early part of the 1980s, winning two Drivers' Championships with Nelson Piquet.
- The rivalry between racers Ayrton Senna and Alain Prost became F1's central focus during 1988 and continued until Prost retired at the end of 1993.
- Senna died at the 1994 San Marino Grand Prix after crashing into a wall on the exit of the notorious curve Tamburello.
- The FIA worked to improve the sport's safety standards since that weekend, during which Roland Ratzenberger also died in an accident during Saturday qualifying.
- No driver died of injuries sustained on the track at the wheel of a Formula One car for 20 years until the 2014 Japanese Grand Prix, where Jules Bianchi collided with a recovery vehicle after aquaplaning off the circuit, dying nine months later from his injuries.
- Since 1994, three track marshals have died, one at the 2000 Italian Grand Prix,[31] the second at the 2001 Australian Grand Prix[31] and the third at the 2013 Canadian Grand Prix.
- A major rule shake-up in 2014 saw the 2.4-litre naturally aspirated V8 engines replaced by 1.6-litre turbocharged hybrid power units.
- This prompted Honda to return to the sport in 2015 as the championship's fourth power unit manufacturer.
- Mercedes emerged as the dominant force after the rule shake-up, with Lewis Hamilton winning the championship closely followed by his main rival and teammate, Nico Rosberg, with the team winning 16 out of the 19 races that season.
- The team continued this form in the following two seasons, again winning 16 races in 2015 before taking a record 19 wins in 2016, with Hamilton claiming the title in the former year and Rosberg winning it in the latter by five points.
- The 2016 season also saw a new team, Haas, join the grid, while Max Verstappen became the youngest-ever race winner at the age of 18 in Spain.
- The race begins with a warm-up lap, after which the cars assemble on the starting grid in the order they qualified.
- This lap is often referred to as the formation lap, as the cars lap in formation with no overtaking (although a driver who makes a mistake may regain lost ground).
- The warm-up lap allows drivers to check the condition of the track and their car, gives the tyres a chance to warm up to increase traction and grip, and also gives the pit crews time to clear themselves and their equipment from the grid for the race start.
- A Formula One constructor is the entity credited for designing the chassis and the engine.
- [97] If both are designed by the same company, that company receives sole credit as the constructor (e.g., Ferrari).
- If they are designed by different companies, both are credited, and the name of the chassis designer is placed before that of the engine designer (e.g., McLaren-Mercedes).
- The use of volunteers was integral in making and maintaining Wikipedia.
- However, even without the internet, huge complex projects of similar nature had made use of volunteers.
- Specifically, the creation of the Oxford English Dictionary was conceived with the speech at the London Library, on Guy Fawkes Day, 5 November 1857, by Richard Chenevix Trench.
- Dr. Trench envisioned a grand new dictionary of every word in the English language, and to be used democratically and freely.
- According to author Simon Winchester, "The undertaking of the scheme, he said, was beyond the ability of any one man.
- To peruse all of English literature - and to comb the London and New York newspapers and the most literate of the magazines and journals - must be instead 'the combined action of many.'

```
In [1]: import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize

def remove_stopwords(text):
    stop_words = set(stopwords.words('english'))
    word_tokens = word_tokenize(text)
    filtered_text = [word for word in word_tokens if word.lower() not in stop_words]
    return filtered_text

document = "On the track, the McLaren and Williams teams dominated the 1980s and 1990s. Brabham, also, competitive, early, part, 1980s, winning, two, Drivers, Championships, Nelson, Piquet, Powered, Porsche, Honda, Mercedes-Benz, McLaren, sixteen, championships, seven, constructors, nine, drivers, period, Williams, used, engines, Ford, Honda, Renault, also, win, sixteen, titles, nine, constructors, seven, drivers, rivalry, racers, Ayrton, Senna, Alain, Prost, became, F1, s, central, focus, 1988, continued, Prost, retired, end, 1993, Senna, died, 1994, San, Marino, Grand, Prix, crashing, wall, exit, notorious, curve, Tamburello, FIA, worked, improve, sport, s, safety, standards, since, weekend, Roland, Ratzenberger, also, died, accident, Saturday, qualifying, driver, died, injuries, sustained, track, wheel, Formula, One, car, 20, years, 2014, Japanese, Grand, Prix, Jules, Bianchi, collided, recovery, vehicle, aquaplaning, circuit, dying, nine, months, later, injuries, Since, 1994, three, track, marshals, died, one, 2000, Italian, Grand, Prix, 31, second, 2001, Australian, Grand, Prix, 31, third, 2013, Canadian, Grand, Prix, .]"
tokenized_document = remove_stopwords(document)
print(tokenized_document)
```

```
['track', ',', 'McLaren', 'Williams', 'teams', 'dominated', '1980s', '1990s', '.', 'Brabham', 'also', 'competitive', 'early', 'part', '1980s', ',', 'winning', 'two', 'Drivers', '"', 'Championships', 'Nelson', 'Piquet', '.', 'Powered', 'Porsche', ',', 'Honda', ',', 'Mercedes-Benz', ',', 'McLaren', 'sixteen', 'championships', '(', 'seven', 'constructors', '"', 'nine', 'drivers', '"', ')', 'period', ',', 'Williams', 'used', 'engines', 'Ford', ',', 'Honda', ',', 'Renault', 'also', 'win', 'sixteen', 'titles', '(', 'nine', 'constructors', '"', 'seven', 'drivers', '"', ')', '.', 'rivalry', 'racers', 'Ayrton', 'Senna', 'Alain', 'Prost', 'became', 'F1', "s", 'central', 'focus', '1988', 'continued', 'Prost', 'retired', 'end', '1993', '.', 'Senna', 'died', '1994', 'San', 'Marino', 'Grand', 'Prix', 'crashing', 'wall', 'exit', 'notorious', 'curve', 'Tamburello', '.', 'FIA', 'worked', 'improve', 'sport', "s", 'safety', 'standards', 'since', 'weekend', ',', 'Roland', 'Ratzenberger', 'also', 'died', 'accident', 'Saturday', 'qualifying', '.', 'driver', 'died', 'injuries', 'sustained', 'track', 'wheel', 'Formula', 'One', 'car', '20', 'years', '2014', 'Japanese', 'Grand', 'Prix', ',', 'Jules', 'Bianchi', 'collided', 'recovery', 'vehicle', 'aquaplaning', 'circuit', ',', 'dying', 'nine', 'months', 'later', 'injuries', '.', 'Since', '1994', ',', 'three', 'track', 'marshals', 'died', ',', 'one', '2000', 'Italian', 'Grand', 'Prix', ',', '31', ']', 'second', '2001', 'Australian', 'Grand', 'Prix', '31', ']', 'third', '2013', 'Canadian', 'Grand', 'Prix', '.']
```

```
In [2]: from collections import Counter

def calculate_tf(tokens):
    cleaned_tokens = [''.join(char for char in token if char.isalnum()) for token in tokens
    cleaned_tokens = [token for token in cleaned_tokens if token] # Remove empty tokens
    tf_scores = Counter(cleaned_tokens)
    total_tokens = len(cleaned_tokens)
    for token in tf_scores:
        tf_scores[token] /= total_tokens
    return tf_scores

tf_scores = calculate_tf(tokenized_document)
sorted_tf_scores = sorted(tf_scores.items(), key=lambda x: x[1], reverse=True)

def print_tf_scores(sorted_tf_scores):
    print("Term                TF Score")
    print("-----")
    for term, tf_score in sorted_tf_scores:
        print(f"{term.ljust(20)}{tf_score:.6f}")

print_tf_scores(sorted_tf_scores)
```

Term	TF Score
Grand	0.035461
Prix	0.035461
died	0.028369
track	0.021277
also	0.021277
nine	0.021277
McLaren	0.014184
Williams	0.014184
1980s	0.014184
Honda	0.014184
sixteen	0.014184
seven	0.014184
constructors	0.014184
drivers	0.014184
Senna	0.014184
Prost	0.014184
s	0.014184
...	...

```
In [6]: def get_sentence_tf(sentence, tf_scores):
    tokens = nltk.word_tokenize(sentence.lower())
    total_words = len(tokens)
    tf_score = sum(tf_scores[word] for word in tokens) / total_words
    return tf_score
```



```
In [7]: def get_top_sentences(document, tf_scores, num_sentences=3):
        sentences = nltk.sent_tokenize(document)
        sentence_tfs = [(sentence, get_sentence_tf(sentence, tf_scores)) for sentence in sentences]
        sorted_sentence_tfs = sorted(sentence_tfs, key=lambda x: x[1], reverse=True)[:num_sentences]
        return [sentence for sentence, _ in sorted_sentence_tfs]

top_sentences = get_top_sentences(document, tf_scores)
def print_top_sentences(top_sentences):
    print("Top Sentences:")
    print("-----")
    for i, sentence in enumerate(top_sentences, 1):
        print(f"Sentence {i}: {sentence}")
        print("-----")

print_top_sentences(top_sentences)
```

Top Sentences:

Sentence 1: No driver died of injuries sustained on the track at the wheel of a Formula One car for 20 years until the 2014 Japanese Grand Prix, where Jules Bianchi collided with a recovery vehicle after aquaplaning off the circuit, dying nine months later from his injuries.

Sentence 2: Powered by Porsche, Honda, and Mercedes-Benz, McLaren won sixteen championships (seven constructors' and nine drivers') in that period, while Williams used engines from Ford, Honda, and Renault to also win sixteen titles (nine constructors' and seven drivers').

Sentence 3: The FIA worked to improve the sport's safety standards since that weekend, during which Roland Ratzenberger also died in an accident during Saturday qualifying.

```
In [8]: top_sentences = get_top_sentences(document, tf_scores)
def print_top_sentences(top_sentences):
    print("Top Sentences:")
    print("-----")
    for i, sentence in enumerate(top_sentences, 1):
        print(f"Sentence {i}: {sentence}")
        print("-----")

print_top_sentences(top_sentences)
```

Top Sentences:

Sentence 1: No driver died of injuries sustained on the track at the wheel of a Formula One car for 20 years until the 2014 Japanese Grand Prix, where Jules Bianchi collided with a recovery vehicle after aquaplaning off the circuit, dying nine months later from his injuries.

Sentence 2: Powered by Porsche, Honda, and Mercedes-Benz, McLaren won sixteen championships (seven constructors' and nine drivers') in that period, while Williams used engines from Ford, Honda, and Renault to also win sixteen titles (nine constructors' and seven drivers').

Sentence 3: The FIA worked to improve the sport's safety standards since that weekend, during which Roland Ratzenberger also died in an accident during Saturday qualifying.

In []: