

```
In [4]: from docx import Document
```

```
In [6]: def read_word_files(file_paths):
    all_content = []
    for file_path in file_paths:
        try:
            doc = Document(file_path)
            text = []
            for paragraph in doc.paragraphs:
                text.append(paragraph.text)
            all_content.append('\n'.join(text))
        except Exception as e:
            print(f"Error reading the Word file '{file_path}': {e}")
    return all_content

word_file_paths = [r"D:\code\NLP\doc\Doc1.docx", r"D:\code\NLP\doc\Doc 2.docx", r"D:\code\NLP\doc\Doc 3.docx"]
contents = read_word_files(word_file_paths)
```

```
In [8]: if contents:
        for i, content in enumerate(contents, start=1):
            print(f"Content of file {i}:")
            print(content)
```

Content of file 1:

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. A Formula One season consists of a series of races, known as Grands Prix. 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.

Content of file 2:

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. 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'). 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.

Content of file 3:

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.

Content of file 4:

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.

Content of file 5:

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). All constructors are scored individually, even if they share either chassis or engine with another constructor (e.g., Williams-Ford, Williams-Honda in 1983).

Content of file 6:

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. It took about 70 years to complete. 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.' It would be necessary to recruit a team – moreover, a huge one – probably comprising hundreds and hundreds of unpaid amateurs, all of them working as volunteers.

```
In [9]: import nltk
        from nltk.tokenize import word_tokenize, sent_tokenize
```

```
In [13]: def tokenize_documents(contents):
        tokenized_docs = []
        for doc_content in contents:
            words = word_tokenize(doc_content)
            sentences = sent_tokenize(doc_content)
            tokenized_docs.append((words, sentences))
        return tokenized_docs

        tokenized_docs = tokenize_documents(contents)
```

```
In [14]: for i, (words, sentences) in enumerate(tokenized_docs, start=1):  
         print(f"\nTokens for Document {i}:")  
         print("Words:", words)  
         print("Sentences:", sentences)
```

Tokens for Document 1:

Words: ['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', '.', 'A', 'Formula', 'One', 'season', 'consists', 'of', 'a', 'series', 'of', 'races', ',', 'known', 'as', 'Grands', 'Prix', '.', 'Grands', 'Prix', 'take', 'place', 'in', 'multiple', 'countries', 'and', 'continents', 'around', 'the', 'world', 'on', 'either', 'purpose-built', 'circuit', 's', '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', '.']

Sentences: ["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.", "A Formula One season consists of a series of races, known as Grands Prix.", "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."]

Tokens for Document 2:

Words: ['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', '.', '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', '"', ')', '.', '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', '.']

Sentences: ["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.", "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').", "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."]

m 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.']

Tokens for Document 3:

Words: ['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', '.']

Sentences: ['A major rule shake-up in 2014 saw the 2.4-litre naturally aspirated V8 engine 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.']

Tokens for Document 4:

Words: ['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', '.']

Sentences: ['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.']

Tokens for Document 5:

Words: ['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', ')', '.', 'All', 'constructors', 'are', 'scored', 'individually', ',', 'even', 'if', 'they', 'share', 'either', 'chassis', 'or', 'engine', 'with', 'another', 'constructor', '(', 'e.g.', ',', 'Williams-Ford', ',', 'Williams-Honda', 'in', '1983', ')', '.']

Sentences: ['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).', 'All constructors are scored individually, even if they share either chassis or engine with another constructor (e.g., Williams-Ford, Williams-Honda in 1983).']

Tokens for Document 6:

Words: ['The', 'use', 'of', 'volunteers', 'was', 'integral', 'in', 'making', 'and', 'maintaining', 'Wikipedia', '.', 'However', ',', 'even', 'without', 'the', 'internet', ',', 'hug

e', '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', '.', 'It', 'took', 'about', '70', 'years', 'to', 'complete', '.', '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', '.', '"', 'It', 'would', 'be', 'necessary', 'to', 'recruit', 'a', 'team', '-', 'moreover', ',', 'a', 'huge', 'one', '-', 'probably', 'comprising', 'hundreds', 'and', 'hundreds', 'of', 'unpaid', 'amateurs', ',', 'all', 'of', 'them', 'working', 'as', 'volunteers', '.']

Sentences: ['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.', 'It took about 70 years to complete.', '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.'", 'It would be necessary to recruit a team – moreover, a huge one – probably comprising hundreds and hundreds of unpaid amateurs, all of them working as volunteers.']

```
In [11]: def unique_words_in_documents(tokenized_docs):
    unique_words_per_doc = []
    for words, _ in tokenized_docs:
        unique_words = set(words)
        unique_words_per_doc.append(unique_words)
    return unique_words_per_doc

unique_words_per_doc = unique_words_in_documents(tokenized_docs)
```



```
In [12]: for i, unique_words in enumerate(unique_words_per_doc, start=1):  
         print(f"\nUnique words in Document {i}:")  
         print(unique_words)
```

Unique words in Document 1:

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

Unique words in Document 2:

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

Unique words in Document 3:

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

Unique words in Document 4:

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

Unique words in Document 5:

{'[', 'before', 'are', 'e.g.', 'entity', 'different', 'scored', 'is', 'or', 'if', 'either', 'Formula', 'in', 'the', 'credit', 'by', 'If', 'receives', 'of', 'even', 'both', 'with', 'credited', 'chassis', 'that', 'designing', 'Ferrari', 'Williams-Ford', 'and', 'All', 'One', 'for', 'companies', 'A', 'engine', 'as', 'sole', '1983', 'name', 'Williams-Honda', 'company', 'they', ']', 'same', 'constructor', 'individually', 'another', 'share', 'constructors', 'McLaren-Mercedes', 'designer', 'designed', '97', 'placed'}

Unique words in Document 6:

{'magazines', 'Oxford', 'Trench', 'them', 'years', 'the', 'action', 'ability', 'making', 'envisioned', 'To', 'word', 'beyond', 'many', 'amateurs', 'grand', 'Fawkes', 'Day', 'every', 'moreover', '5', 'without', 'comprising', 'he', 'Library', 'complete', 'in', 'unpaid', 'maintaining', 'even', 'similar', 'Dictionary', 'speech', 'literature', 'York', 'used', 'nature', 'must', 'and', 'According', 'freely', 'on', 'scheme', 'creation', 'any', 'be', 'November', 'combined', '70', 'Specifically', 'all', 'would', 'volunteers', 'newspapers', 'projects', 'new', 'took', 'Wikipedia', 'working', 'Simon', 'team', 'about', 'democratically', 'internet', 'by', 'had', 'with', 'a', 'London', 'man', 'complex', 'was', 'ne

```
cessary', 'peruse', 'author', '1857', 'one', "'the", 'huge', 'at', 'Winchester', 'dictionary', 'The', 'It', 'undertaking', 'New', 'Chenevix', 'Dr.', 'Guy', 'journals', 'However', 'probably', 'English', ',', 'language', 'most', 'hundreds', 'Richard', '-', 'instead', '.,', 'literate', 'made', 'as', 'integral', 'conceived', 'to', 'comb', 'use', 'recruit', 'of', 'said'}
```

```
In [15]: def combine_unique_words(unique_words_per_doc):
        combined_unique_words = set()
        for unique_words in unique_words_per_doc:
            combined_unique_words.update(unique_words)
        return combined_unique_words

combined_unique_words = combine_unique_words(unique_words_per_doc)
```

```
In [16]: print("\nCombined unique words from all documents:")
        print(combined_unique_words)
```

```
Combined unique words from all documents:
{'1990s', 'conform', 'Mercedes-Benz', 'formula', 'Max', 'force', 'before', 'often', 'magazines', 'Oxford', 'start', '31', 'pit', 'them', 'racing', 'teams', 'valid', '2000', 'years', 'if', 'receives', 'action', 'it', 'part', 'latter', 'ability', 'both', 'Williams-Ford', 'until', 'forms', 'nine', 'order', 'sanctioned', 'traction', 'Lewis', 'car', 'wall', 'envisioned', 'premier', 'track', 'word', 'power', 'One', 'many', 'circuit', 'followed', 'former', 'sole', 'teammate', 'name', 'crashing', 'amateurs', 'grand', 'Fawkes', 'tyres', 'Day', 'again', 'qualifying', 'since', 'Each', 'its', 'l'Automobile', '1950', 'designer', '97', 'place', 'he', 'shake-up', 'later', 'are', 'allows', 'complete', 'or', 'in', 'Alain', 'return', 'similar', 'Dictionary', '2.4-litre', 'between', 'multiple', 'open-wheel', 'that', 'York', 'must', 'used', 'winning', 'freely', 'out', '1.6-litre', 'claiming', 'may', 'collided', 'any', 'for', 'record', 'engine', '2001', 'combined', 'Specifically', 'from', 'all', 'months', 'lost', 'also', 'series', 'seasons', 'volunteers', 'Ford', 'no', 'period', 'wheel', 'de', 'Grand', 'closely', 'competitive', 'year', 'driver', 'youngest-ever', 'Championship', 'either', 'Ayrton', 'team', 'closed', 'an', 'accident', 'about', '2013', 'democratically', 'internet', 'Renault', 'aspirated', 'grip', 'starting', 'Grands', 'a', 'which', 'two', 'man', 'seven', '1994', 'titles', 'dominated', 'mistake', 'necessary', 'peruse', 'Senna', '1857', 'one', 'hold', '1988', 'they', ']', 'race', 'huge', 'lap', 'indivdually', 'share', 'were', 'dictionary', 'standards', 'FIA', 'It', 'championship', 'around', 'taking', 'New', 'Italian', 'Marino', 'Dr.', 'commonly', 'Formula', 'Prost', 'If', 'Canadian', ')', 'However', 'weekend', 'rivalry', 'aquaplaning', 'their', 'probably', 'off', 'prompted', 'exit', ',', 'English', 'most', 'hundreds', 'set', 'sport', 'Richard', 'saw', 'point-system', 'participants', 'No', 'retired', 'main', '(', 'win', 'Haas', 'Saturday', 'end', 'as', 'qualified', 'season', 'Fédération', 'Williams-Honda', 'assemble', 'became', 'comb', 'same', 'dying', 'San', 'V8', 'central', 'five', 'focus', 'recruit', '20', '2016', 'injuries', 'of', 'world', 'Licence', 'Rosberg', 'instead', 'engines', 'being', 'Trench', 'ground', 'sustained', 'Honda', 'chance', 'roads', 'F1', 'up', 'the', 'credit', 'Bianchi', '18', 'running', 'consists', 'annual', 'his', 'have', 'Ferrari', 'during', 'making', 'To', 'Internationale', 'held', 'although', 'wins', '1993', 'emerged', 'been', 'beyond', 'while', '1983', 'unit', 'Nico', 'Ratzenberger', 'every', 'warm-up', 'overtaking', 'another', 'age', 'constructors', 'points', 'moreover', '5', 'rule', 'without', 'safety', 'improve', 'comprising', 'continents', 'hybrid', 'Powered', 'Library', 'rules', 'worked', 'championships', 'entity', 'purpose-built', 'is', 'Tamburello', 'marshals', 'unpaid', 'maintaining', 'Verstappen', 'even', '16', 'speech', 'sixteen', 'regain', 'credited', 'chassis', 'literature', 'nature', 'and', '2015', 'after', 'According', 'Piquet', 'known', "'s", 'on', 'highest', 'turbocharged', 'following', 'formation', 'World', '``', 'Roland', 'scheme', 'creation', 'refers', 'be', 'recovery', 'November', 'third', '70', 'Japanese', 'winner', 'would', 'Super', 'warm', 'units', 'curve', 'newspapers', 'new', 'projects', 'Jules', 'continued', 'into', 'McLaren-Mercedes', 'determine', 'grid', 'took', 'Wikipedia', 'grade', 'working', 'international', 'Simon', 'who', 'time', 'early', 'makes', 'e.g.', '""', '1', 'different', 'licence', 'by', 'notorious', ':', 'with', 'had', 'vehicle', 'take', 'Brabham', 'naturally', 'title', 'designing', 'London', 'issued', 'fourth', 'public', 'Since', 'join', 'complex', 'All', 'has', '2014', 'single-seater', 'gives', 'themselves', 'won', '19', 'was', 'companies', 'A', 'Williams', 'equipment', 'form', 'author', 'Prix', 'tracks', "'the", 'circuits', 'Mercedes', 'constructor', 'check', 'at', 'three', 'replaced', 'Winchester', 'The', 'On', 'crews', 'designed', 'dominant', 'class', '[', 'referred', 'undertaking', 'begins', 'Chenevix', 'died', 'scored', 'races', '1980s', 'major', 'Guy', 'grade-rating', 'manufacturer', 'Nelson', 'journals', 'condition', 'inaugural', 'cars', 'Porsche', 'language', 'drivers', 'racers', 'where', '-', 'Championships', 'increase', '.,', 'McLaren', 'literate', 'made', 'Hamilton', 'clear', 'rival', 'company', 'Australian', 'integral', 'This', 'to', 'conceived', 'use', 'Drivers', 'Spain', 'this', 'countries', 'second', 'said', 'placed'}
```

```
In [22]: import string

def preprocess_text(text):
    text = text.translate(str.maketrans('', '', string.punctuation))
    text = text.replace('[', '').replace(']', '').replace('-', '')
    return text

preprocessed_contents = [preprocess_text(content) for content in contents]
preprocessed_tokenized_docs = [word_tokenize(content) for content in preprocessed_contents]

def get_unique_words(tokenized_docs):
    unique_words = set()
    for words in tokenized_docs:
        unique_words.update(words)
    return unique_words

unique_words_after_preprocessing = get_unique_words(preprocessed_tokenized_docs)

print("Unique words after preprocessing:")
print(unique_words_after_preprocessing)
```

Unique words after preprocessing:

```
{'1990s', 'conform', 'before', 'Max', 'often', 'Oxford', 'teams', '2000', 'if', 'it', 'lat  
ter', 'both', 'traction', 'envisioned', 'car', 'wall', 'premier', 'track', 'youngestever',  
'power', 'many', 'circuit', 'former', 'name', 'grand', 'Fawkes', 'tyres', 'qualifying', 's  
ince', '1950', 'designer', 'later', 'are', 'complete', 'in', 'Alain', 'return', 'between',  
'multiple', 'that', 'York', 'must', 'used', 'winning', 'freely', 'out', 'may', 'collided',  
'WilliamsHonda', 'any', 'openwheel', 'record', 'engine', '2001', 'from', 'months', 'lost',  
'also', 'seasons', 'Ford', 'no', 'de', 'Grand', 'closely', 'driver', 'Championship', 'tea  
m', 'an', 'about', '2013', 'democratically', 'internet', 'aspirated', 'shakeup', 'lAutomob  
ile', 'grip', 'starting', 'two', 'mistake', 'graderating', 'necessary', 'peruse', '1857',  
'they', 'lap', 'share', 'purposebuilt', 'It', 'championship', 'taking', '24litre', 'Formul  
a', 'weekend', 'rivalry', 'aquaplaning', 'their', 'probably', 'sport', 'saw', 'win', 'Haa  
s', 'Saturday', 'Fédération', 'assemble', 'comb', 'same', 'dying', 'focus', '2016', 'injur  
ies', 'world', 'Rosberg', 'Licence', 'instead', 'engines', 'ground', 'sustained', 'Honda',  
'chance', 'roads', 'F1', 'the', 'credit', 'Bianchi', '18', 'annual', 'his', 'Ferrari', 'du  
ring', 'held', 'wins', '1993', 'emerged', 'been', 'while', '1983', 'Ratzenberger', 'ever  
y', 'overtaking', 'age', 'moreover', '5', 'rule', 'without', 'improve', 'Powered', 'Librar  
y', 'championships', 'WilliamsFord', 'maintaining', 'Verstappen', 'Dr', 'sixteen', 'regai  
n', 'credited', 'and', '2015', 'after', 'Piquet', 'known', 'highest', 'turbocharged', 'fol  
lowing', 'World', 'scheme', 'creation', 'refers', 'recovery', 'November', 'third', '70',  
'winner', 'would', 'warm', 'units', 'curve', 'new', 'projects', 'Jules', 'continued', 'int  
o', 'determine', 'grid', 'Wikipedia', 'working', 'Simon', 'time', 'early', '1', 'licence',  
'notorious', 'with', 'vehicle', 'take', 'Brabham', 'naturally', 'title', 'designing', 'iss  
ued', 'fourth', 'Since', 'complex', 'All', '19', 'was', 'A', 'Williams', 'form', 'author',  
'tracks', 'at', 'three', 'replaced', 'Winchester', 'The', 'On', 'crews', 'engine97', 'desi  
gned', 'dominant', 'class', 'referred', 'undertaking', 'begins', 'died', 'races', 'major',  
'manufacturer', 'journals', 'cars', 'language', 'where', 'McLaren', 'Hamilton', 'clear',  
'rival', 'Australian', 'integral', 'use', 'Spain', 'countries', 'second', 'said', 'formul  
a', 'force', 'magazines', 'start', 'pit', 'them', 'racing', 'valid', 'years', 'receives',  
'action', 'part', 'ability', 'until', 'forms', 'nine', 'order', 'sanctioned', 'Lewis', 'wo  
rd', 'One', 'followed', 'sole', 'teammate', 'amateurs', 'crashing', 'Day', 'again', 'Eac  
h', 'its', 'place', 'he', 'allows', 'or', 'similar', 'Dictionary', 'claiming', 'for', 'com  
bined', 'sports', 'Specifically', 'all', 'series', 'volunteers', 'period', 'wheel', 'compe  
titive', 'year', 'either', 'Ayrton', 'closed', 'accident', 'Renault', 'Grands', 'a', 'whic  
h', 'man', 'McLarenMercedes', 'seven', '1994', 'titles', 'dominated', 'warmup', 'singlesea  
ter', 'Senna', 'one', 'hold', '1988', 'race', 'huge', 'individually', 'were', 'dictionar  
y', 'standards', 'FIA', 'around', 'New', 'Italian', 'Marino', 'commonly', 'Prost', 'If',  
'Canadian', 'However', 'off', 'prompted', 'exit', 'English', 'most', 'hundreds', 'set', 'R  
ichard', 'No', 'participants', 'retired', 'main', 'end', 'as', 'qualified', 'season', 'e  
g', 'became', 'San', 'V8', 'central', 'five', '16litre', 'recruit', '20', 'of', 'being',  
'Trench', 'up', 'F1s', 'running', 'consists', 'have', 'making', 'To', 'Internationale', 'a  
lthough', 'beyond', 'unit', 'Nico', 'another', 'constructors', 'points', 'safety', 'compri  
sing', 'continents', 'hybrid', 'rules', 'worked', 'entity', 'is', 'Tamburello', 'marshal  
s', 'unpaid', 'Prix31', 'even', '16', 'speech', 'literature', 'chassis', 'nature', 'Accord  
ing', 'on', 'MercedesBenz', 'formation', 'Roland', 'be', 'Japanese', 'Super', 'newspaper  
s', 'took', 'grade', 'international', 'who', 'makes', 'different', 'by', 'had', 'pointsyst  
em', 'London', 'public', 'join', 'has', '2014', 'gives', 'themselves', 'won', 'companies',  
'equipment', 'Prix', 'circuits', 'Mercedes', 'constructor', 'check', 'Chenevix', 'scored',  
'1980s', 'Guy', 'Nelson', 'condition', 'inaugural', 'Porsche', 'drivers', 'racers', 'Champ  
ionships', 'increase', 'literate', 'made', 'company', 'This', 'to', 'conceived', 'Driver  
s', 'this', 'placed'}
```

```
In [31]: unique_words_list = sorted(list(unique_words_after_preprocessing))
```

```
In [43]: import pandas as pd
```

```
In [56]: df = pd.DataFrame(unique_words_list, columns=["Word"])  
  
for doc in range(1, 7):  
    df[f"Doc{doc}"] = df["Word"].apply(lambda word: 1 if word in preprocessed_tokenized_doc
```

```
In [57]: print(unique_words_list)
```

```
['1', '16', '16litre', '18', '1857', '19', '1950', '1980s', '1983', '1988', '1990s', '1993', '1994', '20', '2000', '2001', '2013', '2014', '2015', '2016', '24litre', '5', '70', 'A', 'According', 'Alain', 'All', 'Australian', 'Ayrton', 'Bianchi', 'Brabham', 'Canadian', 'Championship', 'Championships', 'Chenevix', 'Day', 'Dictionary', 'Dr', 'Drivers', 'Each', 'English', 'F1', 'F1s', 'FIA', 'Fawkes', 'Ferrari', 'Ford', 'Formula', 'Fédération', 'Grand', 'Grands', 'Guy', 'Haas', 'Hamilton', 'Honda', 'However', 'If', 'Internationale', 'It', 'Italian', 'Japanese', 'Jules', 'Lewis', 'Library', 'Licence', 'London', 'Marino', 'Max', 'McLaren', 'McLarenMercedes', 'Mercedes', 'MercedesBenz', 'Nelson', 'New', 'Nico', 'No', 'November', 'On', 'One', 'Oxford', 'Piquet', 'Porsche', 'Powered', 'Prix', 'Prix31', 'Prost', 'Ratzenberger', 'Renault', 'Richard', 'Roland', 'Rosberg', 'San', 'Saturday', 'Senna', 'Simon', 'Since', 'Spain', 'Specifically', 'Super', 'Tamburello', 'The', 'This', 'To', 'Trench', 'V8', 'Verstappen', 'Wikipedia', 'Williams', 'WilliamsFord', 'WilliamsHonda', 'Winchester', 'World', 'York', 'a', 'ability', 'about', 'accident', 'action', 'after', 'again', 'age', 'all', 'allows', 'also', 'although', 'amateurs', 'an', 'and', 'annual', 'another', 'any', 'aquaplaning', 'are', 'around', 'as', 'aspirated', 'assemble', 'at', 'author', 'be', 'became', 'been', 'before', 'begins', 'being', 'between', 'beyond', 'both', 'by', 'car', 'cars', 'central', 'championship', 'championships', 'chance', 'chassis', 'check', 'circuit', 'circuits', 'claiming', 'class', 'clear', 'closed', 'closely', 'collided', 'comb', 'combined', 'commonly', 'companies', 'company', 'competitive', 'complete', 'complex', 'comprising', 'conceived', 'condition', 'conform', 'consists', 'constructor', 'constructors', 'continents', 'continued', 'countries', 'crashing', 'creation', 'credit', 'credited', 'crews', 'curve', 'de', 'democratically', 'designed', 'designer', 'designing', 'determine', 'dictionary', 'died', 'different', 'dominant', 'dominated', 'driver', 'drivers', 'during', 'dying', 'early', 'eg', 'either', 'emerged', 'end', 'engine', 'engine97', 'engines', 'entity', 'envisioned', 'equipment', 'even', 'every', 'exit', 'five', 'focus', 'followed', 'following', 'for', 'force', 'form', 'formation', 'former', 'forms', 'formula', 'fourth', 'freely', 'from', 'gives', 'grade', 'graderating', 'grand', 'grid', 'grip', 'ground', 'had', 'has', 'have', 'he', 'held', 'highest', 'his', 'hold', 'huge', 'hundreds', 'hybrid', 'if', 'improve', 'in', 'inaugural', 'increase', 'individually', 'injuries', 'instead', 'integral', 'international', 'internet', 'into', 'is', 'issued', 'it', 'its', 'join', 'journals', 'known', 'lAutomobile', 'language', 'lap', 'later', 'latter', 'licence', 'literature', 'literature', 'lost', 'made', 'magazines', 'main', 'maintaining', 'major', 'makes', 'making', 'man', 'manufacturer', 'many', 'marshals', 'may', 'mistake', 'months', 'moreover', 'most', 'multiple', 'must', 'name', 'naturally', 'nature', 'necessary', 'new', 'newspapers', 'nine', 'no', 'notorious', 'of', 'off', 'often', 'on', 'one', 'openwheel', 'or', 'order', 'out', 'overtaking', 'part', 'participants', 'period', 'peruse', 'pit', 'place', 'placed', 'points', 'pointsystem', 'power', 'premier', 'probably', 'projects', 'prompted', 'public', 'purposebuilt', 'qualified', 'qualifying', 'race', 'racers', 'races', 'racing', 'receives', 'record', 'recovery', 'recruit', 'referred', 'refers', 'regain', 'replaced', 'retired', 'return', 'rival', 'rivalry', 'roads', 'rule', 'rules', 'running', 'safety', 'said', 'same', 'sanctioned', 'saw', 'scheme', 'scored', 'season', 'seasons', 'second', 'series', 'set', 'seven', 'shakeup', 'share', 'similar', 'since', 'singleseater', 'sixteen', 'sole', 'speech', 'sport', 'sports', 'standards', 'start', 'starting', 'sustained', 'take', 'taking', 'team', 'teammate', 'teams', 'that', 'the', 'their', 'them', 'themselves', 'they', 'third', 'this', 'three', 'time', 'title', 'titles', 'to', 'took', 'track', 'tracks', 'traction', 'turbocharged', 'two', 'tyres', 'undertaking', 'unit', 'units', 'unpaid', 'until', 'up', 'use', 'used', 'valid', 'vehicle', 'volunteers', 'wall', 'warm', 'warmup', 'was', 'weekend', 'were', 'wheel', 'where', 'which', 'while', 'who', 'win', 'winner', 'winning', 'wins', 'with', 'without', 'won', 'word', 'worked', 'working', 'world', 'would', 'year', 'years', 'youngestever']
```

In [58]: df

Out[58]:

	Word	Doc1	Doc2	Doc3	Doc4	Doc5	Doc6
0	1	1	0	0	0	0	0
1	16	0	0	1	0	0	0
2	16litre	0	0	1	0	0	0
3	18	0	0	1	0	0	0
4	1857	0	0	0	0	0	1
...	...	...	...	...	...	...	...
437	world	1	0	0	0	0	0
438	would	0	0	0	0	0	1
439	year	0	0	1	0	0	0
440	years	0	1	0	0	0	1
441	youngestever	0	0	1	0	0	0

442 rows × 7 columns

```
In [64]: def binary_search_query(tokenized_docs, query_word):
    matching_docs = []
    for i, (words, _) in enumerate(tokenized_docs, start=1):
        if query_word in words:
            matching_docs.append(i)
    return matching_docs

query_word = "race"
result = binary_search_query(tokenized_docs, query_word)
print(f"Documents containing the word '{query_word}': {result}")
```

Documents containing the word 'race': [3, 4]

```
In [67]: def binary_search_query(tokenized_docs, query_words):
    matching_docs = []
    for i, (words, _) in enumerate(tokenized_docs, start=1):
        if all(word in words for word in query_words):
            matching_docs.append(i)
    return matching_docs

query_words = ["forms", "formula"]
result = binary_search_query(tokenized_docs, query_words)
print(f"Documents containing all the words '{', '.join(query_words)}': {result}")
```

Documents containing all the words 'forms, formula': [1]

```
In [68]: def binary_search_query(tokenized_docs, query_words, not_word):
    matching_docs = []
    for i, (words, _) in enumerate(tokenized_docs, start=1):
        if all(word in words for word in query_words) and not_word not in words:
            matching_docs.append(i)
    return matching_docs

query_words = ["forms", "formula"]
not_word = "grand"
result = binary_search_query(tokenized_docs, query_words, not_word)
print(f"Documents containing all the words '{', '.join(query_words)}' but not '{not_word}':
```

Documents containing all the words 'forms, formula' but not 'grand': [1]

```
In [70]: def binary_search_query(tokenized_docs, query_words, not_words):  
    matching_docs = []  
    for i, (words, _) in enumerate(tokenized_docs, start=1):  
        if all(word in words for word in query_words) and not any(not_word in words for not  
            matching_docs.append(i)  
    return matching_docs  
  
    query_words = ["forms", "formula"]  
    not_words = ["track", "hundreds"]  
    result = binary_search_query(tokenized_docs, query_words, not_words)  
    print(f"Documents containing all the words '{', '.join(query_words)}' but not any of '{', '")
```

Documents containing all the words 'forms, formula' but not any of 'track, hundreds': [1]

In [ ]: