DataScience HW-10

• Perform NLP analysis on the following questions. Make sure to include interpretation of each result including text/visualizations/tables to support your answer

1. Select any article on the website of your interest and perform Name Entity Relation (NER) analysis on it. Show all the necessary output and visualization and explain your findings. Also mention if you see any shortcoming or error in the algorithm output.

Named Entity Recognition (NER) Analysis

In [2]:
```
import spacy
from spacy import displacy
import requests
from bs4 import BeautifulSoup
import pandas as pd


def url_to_text(url):
    response = requests.get(url)
    soup = BeautifulSoup(response.content, 'html.parser')
    paragraphs = soup.find_all('p')
    text = ' '.join([para.get_text() for para in paragraphs])
    return text

nlp = spacy.load("en_core_web_sm")

url = "https://www.automotive-technology.com/articles/regenerative-bra
text = url_to_text(url)

doc = nlp(text)

print([(X.text, X.label_) for X in doc.ents])
```

```
[('the past 10 years', 'DATE'), ('1', 'CARDINAL'), ('the past years',
'DATE'), ('2', 'CARDINAL'), ('The Basics of Regenerative Braking\nRegen
erative', 'ORG'), ('EV', 'ORG'), ('the electric motor', 'ORG'), ('seven
ty percent', 'PERCENT'), ('3', 'CARDINAL'), ('Regenerative Braking Bene
fits EVs', 'ORG'), ('one', 'CARDINAL'), ('10-30%', 'PERCENT'), ('Improv
ing Energy Efficiency\nDue', 'ORG'), ('EV', 'ORG'), ('4', 'CARDINAL'),
('The Types of Regenerative Braking Systems\nRegenerative', 'ORG'), ('t
wo', 'CARDINAL'), ('the electric motor', 'ORG'), ('Parallel Regenerativ
e Braking', 'WORK_OF_ART'), ('second', 'ORDINAL'), ('5', 'CARDINAL'),
('Regenerative Braking', 'NORP'), ('Limited Braking Power\nRegenerativ
e', 'WORK_OF_ART'), ('Battery Health and Capacity', 'ORG'), ('EV', 'OR
G'), ('6', 'CARDINAL'), ('The Future of Regenerative Braking', 'ORG'),
('7.', 'CARDINAL'), ('Regenerative Braking Impacts Driving Experience
\n', 'ORG'), ('EV', 'ORG'), ('one', 'CARDINAL'), ('one', 'CARDINAL'),
('8', 'CARDINAL'), ('today', 'DATE'), ('the past 10 years', 'DATE'),
('1.', 'CARDINAL'), ('the past years', 'DATE'), ('2.', 'CARDINAL'), ('T
he Basics of Regenerative Braking Regenerative', 'ORG'), ('EV', 'ORG'),
('the electric motor', 'ORG'), ('seventy percent', 'PERCENT'), ('3.',
'CARDINAL'), ('Regenerative Braking Benefits EVs', 'ORG'), ('one', 'CAR
DINAL'), ('10-30%', 'PERCENT'), ('Improving Energy Efficiency Due', 'OR
G'), ('EV', 'ORG'), ('4.', 'CARDINAL'), ('The Types of Regenerative Bra
king Systems Regenerative', 'ORG'), ('two', 'CARDINAL'), ('the electric
motor', 'ORG'), ('Parallel Regenerative Braking', 'ORG'), ('second', 'O
RDINAL'), ('5.', 'CARDINAL'), ('Battery Health and Capacity', 'ORG'),
('EV', 'ORG'), ('6.', 'CARDINAL'), ('The Future of Regenerative Braking
The', 'ORG'), ('7.', 'CARDINAL'), ('Regenerative Braking Impacts Drivin
g Experience Traditional', 'ORG'), ('EV', 'ORG'), ('one', 'CARDINAL'),
('one', 'CARDINAL'), ('today', 'DATE'), ('Automotive Technology', 'OR
G'), ('CRM', 'PRODUCT'), ('Ochre Media Pvt Ltd.', 'ORG'), ('2024', 'DAT
E')]
```

In [3]:
```
displacy.render(doc, style='ent', page=True)
```

Please fill the all required fields....!! Over the past 10 years DATE , electric vehicles have revolutionized mobility through sustainable energy solutions advancing more significantly in the regenerative brake system of automobiles. This is different from normal braking systems in which energy used by the vehicle or one tapped from the battery is used to create kinetic energy during braking that is never apprehended back so as to serve a later purpose. For information on how regenerative braking is implemented, where and how it can be applied, as well as its effects on now and future vehicles – read on.

1 CARDINAL . Understanding Conventional Braking

Before going into more detail about what is called regenerative braking, let us consider what traditional types of brakes are. In virtually, all gasoline or diesel automobiles, there is a type of braking commonly called friction braking. Whenever the driver applies pedal pressure, brake pallets apply pressure on brake rotors where they come into contact and thus slow down the wheels. This friction converts kinetic energy that is energy in motion, directly to heat energy which simply dissipates into space. It does a good job of returning the car to idle for a fairly long period of time and of course, all for naught.

Car manufactures in the past years DATE endeavored to enhance the efficiency of the friction brakes and yet its operating principle entails conversion of energy into heat energy. Regenerative braking work on premise of using this lost energy and harnessing it in some form or the other so as to use it to our advantage.

2 CARDINAL . The Basics of Regenerative Braking Regenerative ORG braking systems work on a simple principle: rather than energizing molecules to become heat, they store the kinetic energy and

turn it to electricity. When the driver pulls the brake pedal of an EV ORG or hybrid vehicle, the car's electric motor becomes a generator in reverse. This generator mode enables the motor to again capture the motion energy of a car and convert it into electrical energy. This energy is then sent back to the battery to provide a composite charge and this give the vehicle a longer charge.

In this system, the electric motor ORG plays a dual role: It is responsible for powering the vehicle during acceleration and charge generating during deceleration. This is made feasible through the reversible characteristic of electric motors. Regenerative braking system can capture a large percentage of energy lost during normal braking, something in the vicinity of seventy percent PERCENT . 3 CARDINAL . How Regenerative Braking Benefits EVs ORG

a) Extending Driving Range

Among the various problems associated with an electric vehicle, there is something called 'range anxiety'. There is also one CARDINAL other disadvantage of using these cars, which is how these cars drain the battery to produce power for the use in the vehicle; there is however regenerative braking by which a small amperage is charged to the battery with each stop. Regenerative breaks do not replace the need to charge but the additional distance that is gained can vary from 10-30% PERCENT depending on the driving cycle and efficacy of the system.

b) Improving Energy Efficiency Due ORG to the fact that regenerative braking handles the energy which would otherwise get wasted in the process, it automatically enhance on the energy usage of

the car. Traditional automobiles cannot retain the energy that was used to accelerate, the energy being dissipated during a braking process. Regenerative braking, though, captures some of that energy and brings efficiency in the drive cycle of the EV ORG higher. This is especially helpful in urban driving conditions where traffic is often congested and frequent slowing down creates the prospect of major energy efficiency gains.

c) Reducing Brake Wear and Maintenance Costs

This means that because regenerative braking reduces the use of friction brakes this also has the advantage of putting less wear and tear on brake pads and rotors. It can also result in reduced maintenance expenses for owners of the electric vehicles, hence you find that electric vehicles are cheaper expensive to maintain. Due to the less usage of friction brakes the life of the friction brakes is higher and therefore less number of brake wears during the life cycle of the vehicle. This is a positive factor that point to cost saving about the durability of regenerative braking.

4 CARDINAL . The Types of Regenerative Braking Systems Regenerative ORG braking system can be different depending on a specific make and model of electric or hybrid vehicle. There are two CARDINAL main types:

a) Series Regenerative Braking

In series regenerative braking, the electric motor ORG gets to work by slowing down the vehicle whenever the brake is applied. This system is most efficient when a gradual slowdown is required for the car. However, if there is an impendent standstill needed, or very fast deceleration, then the friction brakes will be added to the regenerative

system's action. In this system, the control of regenerative and friction braking is done via a control system of the vehicle.

b) Parallel Regenerative Braking WORK_OF_ART

The second ORDINAL type of regenerative braking is called parallel regeneration braking where regenerative and friction brakes are acting in synergy and slow down the car. This system is intended to provide the maximum use of regenerative braking but falls back on the friction brakes when more braking force is required. Parallel systems are most useful in high performance electric vehicles due to the possibly increased demand for fast stops or better braking.

5 CARDINAL . Challenges in Regenerative Braking NORP

Despite its benefits, regenerative braking does face a few challenges:

a) Limited Braking Power Regenerative WORK_OF_ART braking may not suffice to give cars enough stopping power especially in an emergency. In such instances, the friction brake must come to the aid of the situation. Manufacturers integrate both systems so they complement one another; however, using regenerative braking to stop suddenly is impossible.

b) Battery Health and Capacity ORG

The amount of energy that regenerative braking can collect hence varies with the state of charge of the battery. Unfortunately, near the end of a battery charge, the device cannot take much more current, and consequently, the regenerative system cannot recover much energy then either. In such situations, the role of friction braking will increase. Further, excessive charging and discharging cycles may be detrimental to battery life, however, this issue is essentially inapt in contemporary

models of EV ORG batteries.

c) Energy Recovery Efficiency

Some of the kinetic energy cannot be captured by the regenerative braking systems for instance. That is why a factor of speed, temperature or the state of the road has an impact on how much energy the system is capable of capturing and transforming.

6 CARDINAL . The Future of Regenerative Braking ORG

The regenerative braking is also simple to enhance as they improve the technology in electric vehicles. Future systems are expected to construct more complex energy management algorithms and improved battery control systems. Efforts are also being made in the development of ultra capacitor, which is known to have ability to store and discharge power more quickly than battery. Such innovations could greatly help improve operations of regenerative braking systems, especially in demanding driving applications.

Further, with advancement in technology especially on the area of self-driving mode, the regenerative braking could contribute more on the auto energy control. For instance, self-driven EVs could potentially bring better regenerative braking depending of course on sophisticated machine learning algorithms. For example, a car on autonomous mode might see stops a long time in advance and could have a much more efficient manner of slowing down.

7. CARDINAL How Regenerative Braking Impacts Driving Experience ORG Traditional brakes depend on friction, while regenerative brakes use an electric motor to slow the car down: new EV ORG drivers', have a relatively different kind of experience. Nearly all present-day EVs have a function referred to as " one

CARDINAL -pedal driving." The depressing of accelerator during one CARDINAL -pedal mode also enables the use of regenerative braking system that slows the vehicle without actually applying the brake pedal. The above feature can make the driving more natural and less boring, especially when driving in the interior city regions. But it will take some times to adapt, since it makes a completely new type of deceleration, which is not inherent to every car.

8 CARDINAL . Regenerative Braking Beyond EVs

Even though regenerative braking is today DATE considered as the characteristic feature of electric and hybrid vehicles, it may have the application in other types of transport. Originally introduced to locomotives, regenerative braking can be applied to any vehicle – from trains and buses to bicycles. Modern trains already use regenerative braking which leads to the feeding of power back into the electrical supply system whenever the train slows down. Since the drive towards the use of sustainable energy continues to be adopted, it is expected that regenerative braking will be adopted in more forms of vehicles and equipment.

Conclusion

It is, therefore, with great strides that regenerative braking can be noted as having been adopted in automobiles especially electric ones. Integrating this braking system into electric vehicles allows the conversion of the braking force into electricity, which at the same time increases efficiency, range and decreases the amount of wear on the brakes. While regenerative braking may not completely substitute conventional braking means, it becomes an important supplementary

system that increases energy efficiency and environmental friendliness of car operation. This paper has explored the subject of regenerative braking systems in much detail and given evidence of their effectiveness and reliability, it is expected that because of its importance in reducing undesirable energy outcomes, regenerative braking systems it will continue to evolve with advances in technology and become an even more important tool as we shift our alignment towards a more sustainable environment.

In an environment where people pay much attention to energy conservation and environmental conservation regenerative braking helps to make people aware that in this world, there are ways that make even acts such as slowing down productive. Since most of the drivers are signaling their intention to shift to EVs, regenerative braking is set to advance further, which remains a positive development in braking technology as the world goes through changes to live a sustainably conscious lifestyle. Over the past 10 years DATE , electric vehicles have revolutionized mobility through sustainable energy solutions advancing more significantly in the regenerative brake system of automobiles. This is different from normal braking systems in which energy used by the vehicle or one tapped from the battery is used to create kinetic energy during braking that is never apprehended back so as to serve a later purpose. For information on how regenerative braking is implemented, where and how it can be applied, as well as its effects on now and future vehicles – read on. 1. CARDINAL Understanding Conventional Braking Before going into more detail about what is called regenerative braking, let us consider what traditional types of brakes are. In virtually, all gasoline or diesel automobiles, there is a type of

braking commonly called friction braking. Whenever the driver applies pedal pressure, brake pallets apply pressure on brake rotors where they come into contact and thus slow down the wheels. This friction converts kinetic energy that is energy in motion, directly to heat energy which simply dissipates into space. It does a good job of returning the car to idle for a fairly long period of time and of course, all for naught. Car manufactures in the past years DATE endeavored to enhance the efficiency of the friction brakes and yet its operating principle entails conversion of energy into heat energy. Regenerative braking work on premise of using this lost energy and harnessing it in some form or the other so as to use it to our advantage. 2. CARDINAL The Basics of Regenerative Braking Regenerative ORG braking systems work on a simple principle: rather than energizing molecules to become heat, they store the kinetic energy and turn it to electricity. When the driver pulls the brake pedal of an EV ORG or hybrid vehicle, the car's electric motor becomes a generator in reverse. This generator mode enables the motor to again capture the motion energy of a car and convert it into electrical energy. This energy is then sent back to the battery to provide a composite charge and this give the vehicle a longer charge. In this system, the electric motor ORG plays a dual role: It is responsible for powering the vehicle during acceleration and charge generating during deceleration. This is made feasible through the reversible characteristic of electric motors. Regenerative braking system can capture a large percentage of energy lost during normal braking, something in the vicinity of seventy percent PERCENT . 3. CARDINAL How Regenerative Braking Benefits EVs ORG a) Extending Driving Range Among the various problems associated with

an electric vehicle, there is something called ‘range anxiety’. There is also one CARDINAL other disadvantage of using these cars, which is how these cars drain the battery to produce power for the use in the vehicle; there is however regenerative braking by which a small amperage is charged to the battery with each stop. Regenerative breaks do not replace the need to charge but the additional distance that is gained can vary from 10-30% PERCENT depending on the driving cycle and efficacy of the system. b) Improving Energy Efficiency Due ORG to the fact that regenerative braking handles the energy which would otherwise get wasted in the process, it automatically enhance on the energy usage of the car. Traditional automobiles cannot retain the energy that was used to accelerate, the energy being dissipated during a braking process. Regenerative braking, though, captures some of that energy and brings efficiency in the drive cycle of the EV ORG higher. This is especially helpful in urban driving conditions where traffic is often congested and frequent slowing down creates the prospect of major energy efficiency gains. c) Reducing Brake Wear and Maintenance Costs This means that because regenerative braking reduces the use of friction brakes this also has the advantage of putting less wear and tear on brake pads and rotors. It can also result in reduced maintenance expenses for owners of the electric vehicles, hence you find that electric vehicles are cheaper expensive to maintain. Due to the less usage of friction brakes the life of the friction brakes is higher and therefore less number of brake wears during the life cycle of the vehicle. This is a positive factor that point to cost saving about the durability of regenerative braking. 4. CARDINAL The Types of Regenerative Braking Systems Regenerative ORG braking system can

be different depending on a specific make and model of electric or hybrid vehicle. There are two CARDINAL main types: a) Series Regenerative Braking In series regenerative braking, the electric motor ORG gets to work by slowing down the vehicle whenever the brake is applied. This system is most efficient when a gradual slowdown is required for the car. However, if there is an impendent standstill needed, or very fast deceleration, then the friction brakes will be added to the regenerative system's action. In this system, the control of regenerative and friction braking is done via a control system of the vehicle. b) Parallel Regenerative Braking ORG The second ORDINAL type of regenerative braking is called parallel regeneration braking where regenerative and friction brakes are acting in synergy and slow down the car. This system is intended to provide the maximum use of regenerative braking but falls back on the friction brakes when more braking force is required. Parallel systems are most useful in high performance electric vehicles due to the possibly increased demand for fast stops or better braking. 5. CARDINAL Challenges in Regenerative Braking Despite its benefits, regenerative braking does face a few challenges: a) Limited Braking Power Regenerative braking may not suffice to give cars enough stopping power especially in an emergency. In such instances, the friction brake must come to the aid of the situation. Manufacturers integrate both systems so they complement one another; however, using regenerative braking to stop suddenly is impossible. b) Battery Health and Capacity ORG The amount of energy that regenerative braking can collect hence varies with the state of charge of the battery. Unfortunately, near the end of a battery charge, the device cannot take much more current, and

consequently, the regenerative system cannot recover much energy then either. In such situations, the role of friction braking will increase. Further, excessive charging and discharging cycles may be detrimental to battery life, however, this issue is essentially inapt in contemporary models of EV ORG batteries. c) Energy Recovery Efficiency Some of the kinetic energy cannot be captured by the regenerative braking systems for instance. That is why a factor of speed, temperature or the state of the road has an impact on how much energy the system is capable of capturing and transforming. 6. CARDINAL The Future of Regenerative Braking The ORG regenerative braking is also simple to enhance as they improve the technology in electric vehicles. Future systems are expected to construct more complex energy management algorithms and improved battery control systems. Efforts are also being made in the development of ultra capacitor, which is known to have ability to store and discharge power more quickly than battery. Such innovations could greatly help improve operations of regenerative braking systems, especially in demanding driving applications. Further, with advancement in technology especially on the area of self-driving mode, the regenerative braking could contribute more on the auto energy control. For instance, self-driven EVs could potentially bring better regenerative braking depending of course on sophisticated machine learning algorithms. For example, a car on autonomous mode might see stops a long time in advance and could have a much more efficient manner of slowing down. 7. CARDINAL How Regenerative Braking Impacts Driving Experience Traditional ORG brakes depend on friction, while regenerative brakes use an electric motor to slow the car down: new EV ORG drivers', have a relatively different kind of

experience. Nearly all present-day EVs have a function referred to as “ one CARDINAL -pedal driving.” The depressing of accelerator during one CARDINAL -pedal mode also enables the use of regenerative braking system that slows the vehicle without actually applying the brake pedal. The above feature can make the driving more natural and less boring, especially when driving in the interior city regions. But it will take some times to adapt, since it makes a completely new type of deceleration, which is not inherent to every car. 8. Regenerative Braking Beyond EVs Even though regenerative braking is today DATE considered as the characteristic feature of electric and hybrid vehicles, it may have the application in other types of transport. Originally introduced to locomotives, regenerative braking can be applied to any vehicle – from trains and buses to bicycles. Modern trains already use regenerative braking which leads to the feeding of power back into the electrical supply system whenever the train slows down. Since the drive towards the use of sustainable energy continues to be adopted, it is expected that regenerative braking will be adopted in more forms of vehicles and equipment. Conclusion It is, therefore, with great strides that regenerative braking can be noted as having been adopted in automobiles especially electric ones. Integrating this braking system into electric vehicles allows the conversion of the braking force into electricity, which at the same time increases efficiency, range and decreases the amount of wear on the brakes. While regenerative braking may not completely substitute conventional braking means, it becomes an important supplementary system that increases energy efficiency and environmental friendliness of car operation. This paper has explored the subject of regenerative braking systems in much detail

and given evidence of their effectiveness and reliability, it is expected that because of its importance in reducing undesirable energy outcomes, regenerative braking systems it will continue to evolve with advances in technology and become an even more important tool as we shift our alignment towards a more sustainable environment. In an environment where people pay much attention to energy conservation and environmental conservation regenerative braking helps to make people aware that in this world, there are ways that make even acts such as slowing down productive. Since most of the drivers are signaling their intention to shift to EVs, regenerative braking is set to advance further, which remains a positive development in braking technology as the world goes through changes to live a sustainably conscious lifestyle. Thank you for your interest in publishing article with Automotive Technology ORG . Our client success team member will get in touch with you shortly to take this ahead. you're here, check out our informative and insightful article. Happy Surfing! Regards, Client Success Team (CRM PRODUCT), © Ochre Media Pvt Ltd. ORG , 2024 DATE . All rights reserved.

Shortcomings:

It seems to be working good, but there are a few miss classifciations. For example, whenever the word 'EV' appears, it assings this word to 'ORG', which is incorrect, since EV (Electric Vehicles) are a type of vehicles and not an organization or a company. Apart from these few mistakes, its identifying most of the entities correctly.

2. Select any research paper of your interest and perform word frequency and keyword extraction analysis. Use at least 3 different techniques/libraries. Show all the necessary output and visualization and explain your findings. Compare different techniques and mention if you see any shortcoming or error in the algorithm output.

In [4]:
```
import nltk
nltk.download('punkt_tab')

!pip install rake-nltk
```

```
[nltk_data] Downloading package punkt_tab to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt_tab.zip.
Collecting rake-nltk
  Downloading rake_nltk-1.0.6-py3-none-any.whl.metadata (6.4 kB)
Requirement already satisfied: nltk<4.0.0,>=3.6.2 in /usr/local/lib/pyt
hon3.10/dist-packages (from rake-nltk) (3.9.1)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-
packages (from nltk<4.0.0,>=3.6.2->rake-nltk) (8.1.7)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist
-packages (from nltk<4.0.0,>=3.6.2->rake-nltk) (1.4.2)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python
3.10/dist-packages (from nltk<4.0.0,>=3.6.2->rake-nltk) (2024.9.11)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-p
ackages (from nltk<4.0.0,>=3.6.2->rake-nltk) (4.66.6)
Downloading rake_nltk-1.0.6-py3-none-any.whl (9.1 kB)
Installing collected packages: rake-nltk
Successfully installed rake-nltk-1.0.6
```

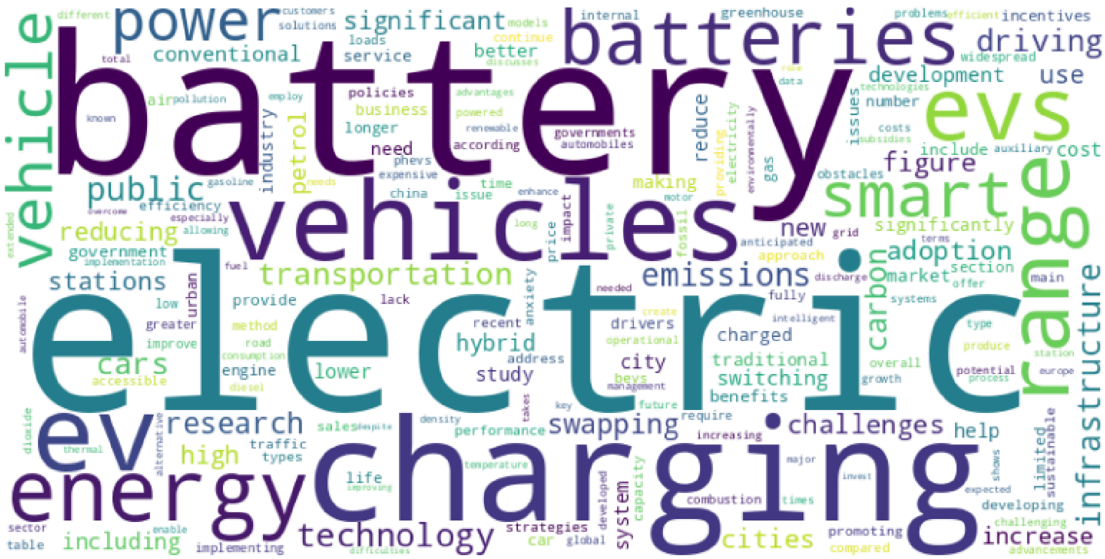
In [9]:
```
import chardet
from collections import Counter
from wordcloud import WordCloud
import matplotlib.pyplot as plt
from sklearn.feature_extraction.text import TfidfVectorizer
import spacy
from nltk.corpus import stopwords
import nltk
from rake_nltk import Rake

file_path = "Electric Vehicles.txt"

raw_data = open(file_path, "rb").read()
result = chardet.detect(raw_data)
encoding = result["encoding"]
print(f"Detected file encoding: {encoding}")

with open(file_path, "r", encoding=encoding, errors="ignore") as file:
    text = file.read()

nltk.download('punkt')
nltk.download('punkt_tab')
nltk.download('stopwords')

nltk_stopwords = set(stopwords.words("english"))
spacy_nlp = spacy.load("en_core_web_sm")

all_stopwords = nltk_stopwords.union(spacy_nlp.Defaults.stop_words)

words = [word.lower() for word in text.split() if word.isalpha() and w
word_counts = Counter(words)
```

```
wordcloud = WordCloud(width=800, height=400, background_color="white")
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis("off")
plt.title("Word Cloud of Research Paper (Excluding Stop Words)")
plt.show()

vectorizer = TfidfVectorizer(max_features=20, stop_words="english")
X = vectorizer.fit_transform([text])
tfidf_keywords = vectorizer.get_feature_names_out()

print("TF-IDF Keywords:")
print(tfidf_keywords)

rake = Rake()
rake.extract_keywords_from_text(text)
rake_keywords = rake.get_ranked_phrases()[:20]

print("\nRAKE Keywords:")
print(rake_keywords)

doc = spacy_nlp(text)
spacy_keywords = [chunk.text for chunk in doc.noun_chunks if chunk.tex

print("\nSpaCy Keywords:")
print(spacy_keywords)

print("\nComparison of Techniques:")
print(f"Word Cloud (Manual Frequency Count): Top {len(word_counts.most
print(f"TF-IDF: {tfidf_keywords}")
print(f"RAKE: {rake_keywords}")
print(f"SpaCy: {spacy_keywords}")
```

```
Detected file encoding: MacRoman
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package punkt_tab to /root/nltk_data...
[nltk_data]   Package punkt_tab is already up-to-date!
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
```

Word Cloud of Research Paper (Excluding Stop Words)



```
TF-IDF Keywords:
['batteries' 'battery' 'cars' 'charging' 'cities' 'electric' 'emission
s'
 'energy' 'ev' 'evs' 'infrastructure' 'power' 'public' 'range' 'smart'
 'stations' 'technology' 'transportation' 'vehicle' 'vehicles']

RAKE Keywords:
['rbs ), electric vehicle producers may achieve low energy consumptio
n', 'suitable energy management strategy may regulate energy use instea
d', 'dynamically building ev recharging infrastructure helps alleviate
range anxiety', 'swapping station operator must continually modify char
ging', 'quick charges without suffering considerable power losses', 'en
hancing ev charging procedures – battery switching stations', 'download
keyboard_arrow_down browse figures versions notes abstract', 'forecast
“ real road ” driving energy consumption', 'balancing auxiliary loads a
uxiliary loads greatly impact', 'gmc hummer ev pickup edition 1', 'char
ging service providers would make significant investments', 'ev manufac
turers even provide complimentary rental automobiles', 'evs would consi
derably reduce greenhouse gas emissions', 'settled region employing mul
tiple technology devices', 'increasingly pushing electric vehicle shari
ng schemes', 'let electric vehicles replace individually constructed ve
hicles', 'significantly negatively influenced urban traffic ’', 'fcevs
): fcevs react hydrogen gas', '900 horsepower gta spano car', 'fayez al
anazi civil engineering department']

SpaCy Keywords:
['Electric Vehicles', 'Benefits', 'Challenges', 'Potential Solutions',
'Widespread Adaptation', 'Fayez Alanazi\n\nCivil Engineering Departmen
t', 'College', 'Engineering', 'Jouf University', 'Sakaka', 'Saudi Arabi
a\nAppl', 'https://doi.org/10.3390/app13106016\nSubmission', '10 May',
'/\xa0Accepted', '11 May', 'This article', 'the Special Issue', 'a Sust
ainable Future', 'The Role', 'Electric Vehicles']

Comparison of Techniques:
Word Cloud (Manual Frequency Count): Top 10 Words: ['electric', 'batter
y', 'charging', 'vehicles', 'ev', 'evs', 'energy', 'range', 'batterie
s', 'smart']
TF-IDF: ['batteries' 'battery' 'cars' 'charging' 'cities' 'electric' 'e
missions'
 'energy' 'ev' 'evs' 'infrastructure' 'power' 'public' 'range' 'smart'
 'stations' 'technology' 'transportation' 'vehicle' 'vehicles']
RAKE: ['rbs ), electric vehicle producers may achieve low energy consum
ption', 'suitable energy management strategy may regulate energy use in
stead', 'dynamically building ev recharging infrastructure helps allevi
ate range anxiety', 'swapping station operator must continually modify
charging', 'quick charges without suffering considerable power losses',
'enhancing ev charging procedures – battery switching stations', 'downl
oadkeyboard_arrow_down browse figures versions notes abstract', 'foreca
st “ real road ” driving energy consumption', 'balancing auxiliary load
s auxiliary loads greatly impact', 'gmc hummer ev pickup edition 1', 'c
harging service providers would make significant investments', 'ev manu
facturers even provide complimentary rental automobiles', 'evs would co
nsiderably reduce greenhouse gas emissions', 'settled region employing
multiple technology devices', 'increasingly pushing electric vehicle sh
aring schemes', 'let electric vehicles replace individually constructed
vehicles', 'significantly negatively influenced urban traffic ’', 'fcev
```

```
s ): fcevs react hydrogen gas', '900 horsepower gta spano car', 'fayez
alanazi civil engineering department']
SpaCy: ['Electric Vehicles', 'Benefits', 'Challenges', 'Potential Solut
ions', 'Widespread Adaptation', 'Fayez Alanazi\n\nCivil Engineering Dep
artment', 'College', 'Engineering', 'Jouf University', 'Sakaka', 'Saudi
Arabia\nAppl', 'https://doi.org/10.3390/app13106016\nSubmission', '10 M
ay', '/\xa0Accepted', '11 May', 'This article', 'the Special Issue', 'a
Sustainable Future', 'The Role', 'Electric Vehicles']
```

In [10]:
```
from sklearn.feature_extraction.text import TfidfVectorizer

tfidf = TfidfVectorizer(stop_words='english')

tfidf_matrix = tfidf.fit_transform([text])
feature_names = tfidf.get_feature_names_out()

scores = tfidf_matrix.sum(axis=0).A1
word_score = [(word, score) for word, score in zip(feature_names, scor
sorted_word_score = sorted(word_score, key=lambda x: x[1], reverse=Tru

for word, score in sorted_word_score[:10]:
    print(f"{word}: {score}")
```

```
electric: 0.4453479184622438
battery: 0.39527921757003887
vehicles: 0.31358817927223087
charging: 0.2529786992448249
ev: 0.24770830967722438
evs: 0.23716753054202333
batteries: 0.1686524661632166
range: 0.1554764922442153
energy: 0.15020610267661477
infrastructure: 0.11594857048721141
```

Comparison of Techniques

TF-IDF effectively highlights key terms like "electric," "vehicles," and "charging" based on importance but struggles with multi-word phrases and variations like "vehicle" and "vehicles" due to the lack of stemming. RAKE excels at extracting descriptive phrases, such as "enhancing EV charging procedures," but often includes irrelevant or verbose outputs from formatting issues. SpaCy identifies meaningful concepts like "Electric Vehicles" and "Potential Solutions" but sometimes captures irrelevant or poorly parsed phrases, lacking a prioritization mechanism. Word Cloud provides a quick visual of frequent terms but does not consider contextual relevance or importance.

Shortcomings and Suggestions

RAKE and SpaCy occasionally extract irrelevant content due to formatting or parsing errors. TF-IDF and Word Cloud rely solely on frequency, missing nuanced phrases. Preprocessing steps like stemming, lemmatization, and noise removal

can improve all methods. Combining SpaCy's entity recognition with TF-IDF's scoring and RAKE's phrase extraction could yield more reliable and contextually rich results.

Conclusion

While each technique has its strengths, an integrated approach addressing their individual weaknesses would provide a more robust keyword extraction process.

3. Choose any keyword or twitter handler (page/account) of your choice and pull/scrap at least 20 (tweets) from Twitter. Manually put them in a text file. One tweet per line.

a. Read the text file line by line.

b. Perform Sentiment Analysis on the tweet and show total positive and negative sentiment counts.

c. Display one Positive and one Negative sentiment and discuss if the algorithm is accurate.

In [11]:
```
!pip install textblob
```

```
Requirement already satisfied: textblob in /usr/local/lib/python3.10/dist-packages (0.17.1)
Requirement already satisfied: nltk>=3.1 in /usr/local/lib/python3.10/dist-packages (from textblob) (3.9.1)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-packages (from nltk>=3.1->textblob) (8.1.7)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages (from nltk>=3.1->textblob) (1.4.2)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.10/dist-packages (from nltk>=3.1->textblob) (2024.9.11)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from nltk>=3.1->textblob) (4.66.6)
```

In [21]:
```
from textblob import TextBlob

file_path = "tweets.rtf"

with open(file_path, "r") as file:

    content = file.read()
    tweets = content.split("/n")


sentiment_scores = []

for tweet in tweets:
    tweet = tweet.strip()
```

```
    if not tweet:
        continue

    analysis = TextBlob(tweet)
    polarity = analysis.sentiment.polarity

    sentiment_scores.append((tweet, polarity))

sentiment_scores.sort(key=lambda x: x[1])

most_negative_tweet = sentiment_scores[0][0]
second_negative_tweet = sentiment_scores[1][0]

most_positive_tweet = sentiment_scores[-1][0]
second_positive_tweet = sentiment_scores[-2][0]

positive_count = sum(1 for _, polarity in sentiment_scores if polarity
negative_count = sum(1 for _, polarity in sentiment_scores if polarity
neutral_count = sum(1 for _, polarity in sentiment_scores if polarity

print(f"Positive Sentiment Count: {positive_count}")
print(f"Negative Sentiment Count: {negative_count}")
print(f"Neutral Sentiment Count: {neutral_count}")

print("\nMost Positive Tweet:")
print(most_positive_tweet)

print("\nSecond Most Positive Tweet:")
print(second_positive_tweet)

print("\nMost Negative Tweet:")
print(most_negative_tweet)

print("\nSecond Most Negative Tweet:")
print(second_negative_tweet)
```

```
Positive Sentiment Count: 9
Negative Sentiment Count: 3
Neutral Sentiment Count: 9

Most Positive Tweet:
\
His sense of humor is amazing

Second Most Positive Tweet:
\
Wow, this needs to be greatly simplified @DOGE!

Most Negative Tweet:
\
This is insane

Second Most Negative Tweet:
\
This is crazy
```

Discussion:

Although TextBlob's sentiment analysis offers a decent approximation of sentiment, it has drawbacks, particularly when it comes to deciphering context, sarcasm, and domain-specific terminology. Although the instances of positive sentiment are largely correct, tweets that contain terms like "insane" or "crazy" may be incorrectly labeled because they depend on the context, as TextBlob solely examines word polarity without taking into account larger context. The need for more sophisticated models that can comprehend subtle sentiment and irony, such as BERT or RoBERTa, is highlighted by this. Sentiment analysis can be helpful, but human review and model improvement would increase its precision, particularly for unclear or complex tweets.

4. Pick any dataset and build recommendation engines using

a. TF-IDF method

b. Using word2vec (doc2vec) model (deep learning)

c. Test both engines and compare the quality/accuracy of recommendations

Note:

The Dataset that I've used here is the IMDB 50k movie reviews dataset. Since 50,000 records were a bit too much for collab, I splitted the CSV into 5 differents CSVs, each containing 10,000 records. For this particular question, I have used the first one of those 5 CSVs.

TF-IDF Recommendation

In [11]:
```
import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import cosine_similarity
import nltk
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
import string

nltk.download('stopwords')

file_path = 'IMDB Dataset-1.csv'
df = pd.read_csv(file_path)

reviews = df['review']

stop_words = set(stopwords.words('english'))
stemmer = PorterStemmer()

def preprocess_review(review):
```

```
    review = review.lower()

    review = ''.join([char for char in review if char not in string.pu

    words = review.split()
    filtered_words = [stemmer.stem(word) for word in words if word not

    return ' '.join(filtered_words)

processed_reviews = reviews.apply(preprocess_review)

tfidf_vectorizer = TfidfVectorizer(stop_words='english')
tfidf_matrix = tfidf_vectorizer.fit_transform(processed_reviews)

cosine_sim = cosine_similarity(tfidf_matrix[0:1], tfidf_matrix)

most_similar_review_idx = cosine_sim.argsort()[0][-2]
most_similar_review = reviews.iloc[most_similar_review_idx]

print(f"Most similar review using TF-IDF:\n{most_similar_review}")
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
Most similar review using TF-IDF:
Oz is by far the best show ever to grace television. Better than The So
pranos, yes, ER, yes, CSI, absolutely. Uncompromising, daring, and utte
rly disturbing yet profoundly moving. Oz took us past any image of pris
on that anyone had ever conjured up on television. Tom Fontana truly di
d a brilliant job with the writing. No topic is taboo. Rape, drugs, mur
der. Oz is evidence of just how good TV can be. It follows characters o
f all different backgrounds and all different races, but always comes b
ack to your everyman Tobias Beecher, in jail for vehicular manslaughte
r. We see what we don't want to see, pain, death, mayhem. Oz will distu
rb you, make you cringe, make you look away, but most of all it will ma
ke you think. To see Oz is to see a truly magnificent television produc
tion
```

Word2Vec (Doc2Vec) Recommendation

In [9]:

```
import pandas as pd
from gensim.models import Word2Vec
from nltk.tokenize import word_tokenize

file_path = 'IMDB Dataset-1.csv'
df = pd.read_csv(file_path)

reviews = df['review']

tokenized_reviews = [word_tokenize(review.lower()) for review in revie

model_w2v = Word2Vec(tokenized_reviews, vector_size=100, window=5, min

first_review_tokens = tokenized_reviews[0]

first_review_vector = sum(model_w2v.wv[word] for word in first_review_
```

```
similarity_scores = []
for review in tokenized_reviews:
    review_vector = sum(model_w2v.wv[word] for word in review if word
    similarity = model_w2v.wv.cosine_similarities(first_review_vector,
    similarity_scores.append(similarity)


most_similar_review_w2v = reviews.iloc[similarity_scores.index(max(sim

print(f"Most similar review using Word2Vec:\n{most_similar_review_w2v}
```

```
Most similar review using Word2Vec:
One of the other reviewers has mentioned that after watching just 1 Oz
episode you'll be hooked. They are right, as this is exactly what happe
ned with me.<br /><br />The first thing that struck me about Oz was its
brutality and unflinching scenes of violence, which set in right from t
he word GO. Trust me, this is not a show for the faint hearted or timi
d. This show pulls no punches with regards to drugs, sex or violence. I
ts is hardcore, in the classic use of the word.<br /><br />It is called
OZ as that is the nickname given to the Oswald Maximum Security State P
enitentary. It focuses mainly on Emerald City, an experimental section
of the prison where all the cells have glass fronts and face inwards, s
o privacy is not high on the agenda. Em City is home to many..Aryans, M
uslims, gangstas, Latinos, Christians, Italians, Irish and more....so s
cuffles, death stares, dodgy dealings and shady agreements are never fa
r away.<br /><br />I would say the main appeal of the show is due to th
e fact that it goes where other shows wouldn't dare. Forget pretty pict
ures painted for mainstream audiences, forget charm, forget romance...O
Z doesn't mess around. The first episode I ever saw struck me as so nas
ty it was surreal, I couldn't say I was ready for it, but as I watched
more, I developed a taste for Oz, and got accustomed to the high levels
of graphic violence. Not just violence, but injustice (crooked guards w
ho'll be sold out for a nickel, inmates who'll kill on order and get aw
ay with it, well mannered, middle class inmates being turned into priso
n bitches due to their lack of street skills or prison experience) Watc
hing Oz, you may become comfortable with what is uncomfortable viewin
g....thats if you can get in touch with your darker side.
```

Comparison of Quality & Accuracy of Recommendations:

The two recommendation methods, **TF-IDF** and **Word2Vec**, take different approaches to identifying similar reviews, each with its own advantages and drawbacks. TF-IDF operates by comparing word frequencies between the target review and others in the dataset. The most similar review produced by TF-IDF characterizes *Oz* as a show that is bold and emotionally resonant, highlighting its writing and the unsettling yet meaningful themes it explores. While this review is certainly pertinent, TF-IDF tends to focus on matching surface-level terms like "TV," "violence," and "Oz," without fully grasping the deeper context or thematic subtleties.

On the other hand, Word2Vec captures **contextual meaning** and the **semantic relationships** between words. The most similar review generated by Word2Vec highlights the harsh, unflinching portrayal of prison life and violence, which

closely aligns with the tone and content of the original review. Word2Vec's ability to understand the connections between words allows it to offer a more thematically accurate recommendation. It not only identifies commonly used words but also uncovers the deeper meanings behind them, making its recommendation more contextually relevant to the original review. Therefore, while both methods yield relevant suggestions, **Word2Vec** generally excels over TF-IDF in terms of capturing **contextual relevance** and **thematic accuracy**.