Extractive Summarization

Project Group: Rahul Reddy Baddam , Sri Charan Vemuri , Shri Datta Madhira

Intrsuctions : The code is well commented and documented. Considering the dataset files are in the same directory, we can run the code sequentially from top , make observations and see the results , thank you

IMPORTS

In [1]:
```python
!pip install sumy
!pip install sacrebleu
!pip install rouge

import re
import nltk
import json
import spacy
import numpy as np
import pandas as pd
import seaborn as sn
import matplotlib.pyplot as plt

from time import time
from rouge import Rouge
from bs4 import BeautifulSoup
from collections import Counter
from sacrebleu.metrics import BLEU
from sumy.utils import get_stop_words
from sumy.nlp.stemmers import Stemmer
from sumy.nlp.tokenizers import Tokenizer
from sumy.summarizers.lsa import LsaSummarizer
from sumy.summarizers.luhn import LuhnSummarizer
from sumy.parsers.plaintext import PlaintextParser
from sumy.summarizers.lex_rank import LexRankSummarizer
from sumy.summarizers.edmundson import EdmundsonSummarizer

nlp = spacy.load('en_core_web_sm', disable=['ner', 'parser'])
print('The nltk version is {}.'.format(nltk.__version__)) # Verify version >3.5

nltk.download('punkt')
```

```
Requirement already satisfied: sumy in c:\users\rahul\anaconda3\envs\dmt\lib\site-pac
kages (0.11.0)
Requirement already satisfied: breadability>=0.1.20 in c:\users\rahul\anaconda3\envs
\dmt\lib\site-packages (from sumy) (0.1.20)
Requirement already satisfied: pycountry>=18.2.23 in c:\users\rahul\anaconda3\envs\dm
t\lib\site-packages (from sumy) (22.3.5)
Requirement already satisfied: nltk>=3.0.2 in c:\users\rahul\anaconda3\envs\dmt\lib\s
ite-packages (from sumy) (3.7)
Requirement already satisfied: docopt<0.7,>=0.6.1 in c:\users\rahul\anaconda3\envs\dm
t\lib\site-packages (from sumy) (0.6.2)
Requirement already satisfied: requests>=2.7.0 in c:\users\rahul\anaconda3\envs\dmt\l
ib\site-packages (from sumy) (2.28.2)
Requirement already satisfied: lxml>=2.0 in c:\users\rahul\anaconda3\envs\dmt\lib\sit
e-packages (from breadability>=0.1.20->sumy) (4.9.1)
Requirement already satisfied: chardet in c:\users\rahul\anaconda3\envs\dmt\lib\site-
packages (from breadability>=0.1.20->sumy) (4.0.0)
Requirement already satisfied: tqdm in c:\users\rahul\anaconda3\envs\dmt\lib\site-pac
kages (from nltk>=3.0.2->sumy) (4.64.1)
Requirement already satisfied: regex>=2021.8.3 in c:\users\rahul\anaconda3\envs\dmt\l
ib\site-packages (from nltk>=3.0.2->sumy) (2022.7.9)
Requirement already satisfied: click in c:\users\rahul\anaconda3\envs\dmt\lib\site-pa
ckages (from nltk>=3.0.2->sumy) (8.0.4)
Requirement already satisfied: joblib in c:\users\rahul\anaconda3\envs\dmt\lib\site-p
ackages (from nltk>=3.0.2->sumy) (1.1.1)
Requirement already satisfied: setuptools in c:\users\rahul\anaconda3\envs\dmt\lib\si
te-packages (from pycountry>=18.2.23->sumy) (65.6.3)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\rahul\anaconda3\envs
\dmt\lib\site-packages (from requests>=2.7.0->sumy) (1.26.14)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\rahul\anaconda3\envs\dm
t\lib\site-packages (from requests>=2.7.0->sumy) (2022.12.7)
Requirement already satisfied: idna<4,>=2.5 in c:\users\rahul\anaconda3\envs\dmt\lib
\site-packages (from requests>=2.7.0->sumy) (3.4)
Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\rahul\anaconda3\e
nvs\dmt\lib\site-packages (from requests>=2.7.0->sumy) (2.1.1)
Requirement already satisfied: colorama in c:\users\rahul\anaconda3\envs\dmt\lib\site
-packages (from click->nltk>=3.0.2->sumy) (0.4.6)
Requirement already satisfied: sacrebleu in c:\users\rahul\anaconda3\envs\dmt\lib\sit
e-packages (2.3.1)
Requirement already satisfied: portalocker in c:\users\rahul\anaconda3\envs\dmt\lib\s
ite-packages (from sacrebleu) (2.7.0)
Requirement already satisfied: lxml in c:\users\rahul\anaconda3\envs\dmt\lib\site-pac
kages (from sacrebleu) (4.9.1)
Requirement already satisfied: numpy>=1.17 in c:\users\rahul\anaconda3\envs\dmt\lib\s
ite-packages (from sacrebleu) (1.23.5)
Requirement already satisfied: colorama in c:\users\rahul\anaconda3\envs\dmt\lib\site
-packages (from sacrebleu) (0.4.6)
Requirement already satisfied: regex in c:\users\rahul\anaconda3\envs\dmt\lib\site-pa
ckages (from sacrebleu) (2022.7.9)
Requirement already satisfied: tabulate>=0.8.9 in c:\users\rahul\anaconda3\envs\dmt\l
ib\site-packages (from sacrebleu) (0.9.0)
Requirement already satisfied: pywin32>=226 in c:\users\rahul\anaconda3\envs\dmt\lib
\site-packages (from portalocker->sacrebleu) (305.1)
Requirement already satisfied: rouge in c:\users\rahul\anaconda3\envs\dmt\lib\site-pa
ckages (1.0.1)
Requirement already satisfied: six in c:\users\rahul\anaconda3\envs\dmt\lib\site-pack
ages (from rouge) (1.16.0)
The nltk version is 3.7.
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\rahul\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
```

True

READING THE FOOD REVIEW DATASET

In [2]:
```python
fr_data = pd.read_csv("Reviews.csv",nrows=100000)

fr_data.head()
```

Out[2]:

	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	1
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	0
2	3	B000LQOCH0	ABXLMWJIXXAIN	Natalia Corres "Natalia Corres"	1	1
3	4	B000UA0QIQ	A395BORC6FGVXV	Karl	3	3
4	5	B006K2ZZ7K	A1UQRSCLF8GW1T	Michael D. Bigham "M. Wassir"	0	0

In [3]:
```python
#Reformating the data (dropping duplicates and missing values, and selecting only the
fr_data.rename(columns={'Text': 'text', 'Summary': 'summary'}, inplace=True)

fr_data.drop_duplicates(subset=['text'], inplace=True)
fr_data.dropna(axis=0, inplace=True)

fr_data = fr_data[['summary', 'text']]

fr_data
```

Out[3]:

	summary	text
0	Good Quality Dog Food	I have bought several of the Vitality canned d...
1	Not as Advertised	Product arrived labeled as Jumbo Salted Peanut...
2	"Delight" says it all	This is a confection that has been around a fe...
3	Cough Medicine	If you are looking for the secret ingredient i...
4	Great taffy	Great taffy at a great price. There was a wid...
...	...	...
99995	yummy!	I just love it and will buy another box when I...
99996	Tastes like More!	My late father in law used to have a rating sy...
99997	Great ramen	This is my favorite brand of Korean ramen. It ...
99998	Spicy!!	I do like these noodles although, to say they ...
99999	This spicy noodle cures my cold, upset stomach...	I love this noodle and have it once or twice a...

88421 rows × 2 columns

PROCESSING THE DATASET

In [4]:
```python
# Cleans text data for extractive models by removing HTML tags and converting to lower
def preprocess_text_for_extractive_models(data):
    output = []
    for line in data:
        line = BeautifulSoup(line, "lxml").text
        line = line.lower()
        output.append(line)
    return output


def preprocess(df):
    # Preprocess the 'text' and 'summary' columns using the preprocess_text_for_extrac
    text_clean = preprocess_text_for_extractive_models(df['text'].apply(str))
    summary_clean = preprocess_text_for_extractive_models(df['summary'].apply(str))
    final_text = []
    final_summary = []

    for i in range(len(text_clean)):
        final_text.append(text_clean[i])
        final_summary.append(summary_clean[i])
    # Return a new DataFrame with the preprocessed text and summary columns
    return pd.DataFrame({'text': final_text, 'summary': final_summary})
```

In [5]:
```python
reviews_df = preprocess(fr_data)
# Preprocessed Data
reviews_df
```

Out[5]:

	text	summary
0	i have bought several of the vitality canned d…	good quality dog food
1	product arrived labeled as jumbo salted peanut…	not as advertised
2	this is a confection that has been around a fe…	"delight" says it all
3	if you are looking for the secret ingredient i…	cough medicine
4	great taffy at a great price. there was a wid…	great taffy
...	...	...
88416	i just love it and will buy another box when i…	yummy!
88417	my late father in law used to have a rating sy…	tastes like more!
88418	this is my favorite brand of korean ramen. it …	great ramen
88419	i do like these noodles although, to say they …	spicy!!
88420	i love this noodle and have it once or twice a…	this spicy noodle cures my cold, upset stomach…

88421 rows × 2 columns

In [6]:
```python
# Creates two maps: one maps review IDs to complete reviews, and the other mapsreview
def get_id_reviews_map(df):
    id_complete_review_map = {}
    id_summary_review_map = {}
    # Loop through each review in the DataFrame and create the mappings
    for id, review in enumerate(list(zip(df['text'], df['summary']))):
        complete_review, summary_review = review
        id_complete_review_map[id] = complete_review
        id_summary_review_map[id] = summary_review
    # Return the ID-review mappings
    return id_complete_review_map, id_summary_review_map

id_complete_review_map, id_summary_review_map = get_id_reviews_map(reviews_df)
```

In [7]:
```python
# This function calculates the BLEU score between a hypothesis and a reference using t

def get_bleu_score(hypothesis, reference):
    # Initialize the BLEU scorer
    bleu_scorer = BLEU(effective_order=True)

    # Calculate the BLEU score and return it
    score = 0
    try:
        score = bleu_scorer.sentence_score(hypothesis=hypothesis, references=[referenc
    except:
```

```python
            pass
        return score.score/100 if score else 0


    # This function calculates the ROUGE-L score between a hypothesis and a reference usir

    def get_rouge_score(hypothesis, reference):
        # Initialize the Rouge scorer
        rouge_scorer = Rouge()

        # Calculate the ROUGE-L score and return it
        score = 0
        try:
            score = rouge_scorer.get_scores(hyps=hypothesis, refs=reference)
        except:
            pass
        return score[0]["rouge-l"]["f"] if score else 0
```

In [8]:
```python
language = "english"
sentence_count = 1  # Denotes the number of sentences that should be in machine genera

# Summarize text using a specified summarizer, compute BLEU and ROUGE scores, and retu
def summarize(summarizer, summarizer_name, id_complete_review_map, id_summary_review_m

    summarizer_results = []

    # Iterate through each review in the id_complete_review_map
    for id in id_complete_review_map:
        # Retrieve the review text and summary reference for the current review id
        text = id_complete_review_map[id]
        reference = id_summary_review_map[id]
        parser = PlaintextParser(text, Tokenizer(language))
        # Generate the summary using the specified summarizer and sentence count
        summary = summarizer(parser.document, sentence_count)

        # Convert the summary from a list of sentences to a single string
        hypothesis = " ".join(str(s) for s in summary)

        # Calculate the BLEU and ROUGE scores for the summary
        bleu_score = get_bleu_score(hypothesis, reference)
        rouge_score = get_rouge_score(hypothesis, reference)

        summarizer_results.append([id, text, reference, hypothesis, bleu_score, rouge_

    # Create a DataFrame from the summarizer_results list and sort by BLEU and ROUGE s
    summarized_df = pd.DataFrame(summarizer_results, columns=["Document Id", "Original
    bleu_sorted = summarized_df.sort_values(by='BLEU Score', ascending=False)
    rouge_sorted = summarized_df.sort_values(by='ROUGE Score', ascending=False)

    # Return the sorted DataFrames
    return bleu_sorted, rouge_sorted
```

In [9]:
```python
# Function to print each sentence of a text in a new line
def print_sentences_in_new_line(text):
    s = ""
    for x in text.split("."):
        s += x + "\n"
    return s
# Function to pretty print the summary results
```

```
def pretty_print(df, summarizer_name):
    for index, row in df:
        print("\033[1m Original Text (Cleaned): \033[0m\n", print_sentences_in_new_lir
        , "\n\033[1m Reference Summary (Cleaned): \033[0m\n", print_sentences_in_new_line(
        , "\n\033[1m Candidate Summary: \033[0m\n", print_sentences_in_new_line(row["Candi
        , "\n\033[1m BLEU Score: \033[0m", str(row["BLEU Score"])
        , "\n\033[1m ROUGE Score: \033[0m", str(row["ROUGE Score"]))
        print("\n\033[1m ------------------------------------------------------------
```

EXPERIMENTS:

In [10]:
```
# Experiments for LSA Model
lsa_summarizer = LsaSummarizer(Stemmer(language))
lsa_summarizer.stop_words = get_stop_words(language)

lsa_bleu_sorted, lsa_rouge_sorted = summarize(lsa_summarizer, "LSA", id_complete_revie

# Print the top 5 summaries along with their BLEU and ROUGE scores
print("\033[1m LSA Top 5 NEWS AND SUMMARIES - BLEU SCORES \033[0m")
pretty_print(lsa_bleu_sorted.head(5).iterrows(), "LSA")
print("\033[1m LSA Top 5 NEWS AND SUMMARIES - ROUGE SCORES \033[0m")
pretty_print(lsa_rouge_sorted.head(5).iterrows(), "LSA")
```

```
C:\Users\rahul\anaconda3\envs\dmt\lib\site-packages\sumy\summarizers\lsa.py:76: UserW
arning: Number of words (1) is lower than number of sentences (2). LSA algorithm may
not work properly.
  warn(message % (words_count, sentences_count))
C:\Users\rahul\anaconda3\envs\dmt\lib\site-packages\sumy\summarizers\lsa.py:76: UserW
arning: Number of words (1) is lower than number of sentences (5). LSA algorithm may
not work properly.
  warn(message % (words_count, sentences_count))
C:\Users\rahul\anaconda3\envs\dmt\lib\site-packages\sumy\summarizers\lsa.py:76: UserW
arning: Number of words (5) is lower than number of sentences (7). LSA algorithm may
not work properly.
  warn(message % (words_count, sentences_count))
C:\Users\rahul\anaconda3\envs\dmt\lib\site-packages\sumy\summarizers\lsa.py:76: UserW
arning: Number of words (3) is lower than number of sentences (8). LSA algorithm may
not work properly.
  warn(message % (words_count, sentences_count))
C:\Users\rahul\anaconda3\envs\dmt\lib\site-packages\sumy\summarizers\lsa.py:76: UserW
arning: Number of words (3) is lower than number of sentences (4). LSA algorithm may
not work properly.
  warn(message % (words_count, sentences_count))
C:\Users\rahul\anaconda3\envs\dmt\lib\site-packages\sumy\summarizers\lsa.py:76: UserW
arning: Number of words (1) is lower than number of sentences (3). LSA algorithm may
not work properly.
  warn(message % (words_count, sentences_count))
C:\Users\rahul\anaconda3\envs\dmt\lib\site-packages\sumy\summarizers\lsa.py:76: UserW
arning: Number of words (4) is lower than number of sentences (6). LSA algorithm may
not work properly.
  warn(message % (words_count, sentences_count))
C:\Users\rahul\anaconda3\envs\dmt\lib\site-packages\sumy\summarizers\lsa.py:76: UserW
arning: Number of words (4) is lower than number of sentences (5). LSA algorithm may
not work properly.
  warn(message % (words_count, sentences_count))
C:\Users\rahul\anaconda3\envs\dmt\lib\site-packages\sumy\summarizers\lsa.py:76: UserW
arning: Number of words (3) is lower than number of sentences (6). LSA algorithm may
not work properly.
  warn(message % (words_count, sentences_count))
C:\Users\rahul\anaconda3\envs\dmt\lib\site-packages\sumy\summarizers\lsa.py:76: UserW
arning: Number of words (6) is lower than number of sentences (9). LSA algorithm may
not work properly.
  warn(message % (words_count, sentences_count))
C:\Users\rahul\anaconda3\envs\dmt\lib\site-packages\sumy\summarizers\lsa.py:76: UserW
arning: Number of words (5) is lower than number of sentences (6). LSA algorithm may
not work properly.
  warn(message % (words_count, sentences_count))
C:\Users\rahul\anaconda3\envs\dmt\lib\site-packages\sumy\summarizers\lsa.py:76: UserW
arning: Number of words (14) is lower than number of sentences (19). LSA algorithm ma
y not work properly.
  warn(message % (words_count, sentences_count))
```

Original Text (Cleaned):
love this stuff, dn't need cream/mik or sugar for my coffee, saves 100+ calories per cup!

Reference Summary (Cleaned):
love this stuff, dn't need cream/mik or sugar for my coffee, saves 100+ calories per cup!

Candidate Summary:
love this stuff, dn't need cream/mik or sugar for my coffee, saves 100+ calories per cup!

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

Original Text (Cleaned):
i like this coffee
 that says it all! why should i go on when that is all to be said

Reference Summary (Cleaned):
i like this coffee

Candidate Summary:
i like this coffee

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

Original Text (Cleaned):
it's a shame to use so much corn syrup for babies! i'm very disappointed when i noticed that
the same kind sold in china is even better

Reference Summary (Cleaned):
it's a shame to use so much corn syrup for babies!

Candidate Summary:
it's a shame to use so much corn syrup for babies!

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

Original Text (Cleaned):
vya sweet vermouth elevates your mixed drink beyond expectations! i just discovered this brand's vermouth this past week off the internet blogs
 i'd been plowing thru other vermouth brands adjusting my manhattans
 this one is spectacular
i'm going out tomorrow to hunt down their dry and whisper dry for my martini's, and i think this one is going to be my favorite gift to give for the forseeable future! th

is will be fun to introduce to my friends
buy stock! via can only get more popular
  as i said

spectacular

 Reference Summary (Cleaned):
 vya sweet vermouth elevates your mixed drink beyond expectations!

 Candidate Summary:
 vya sweet vermouth elevates your mixed drink beyond expectations!

 BLEU Score: 1.0000000000000004
 ROUGE Score: 0.999999995

 Original Text (Cleaned):
 nice and simple, basic tadka daal! you cant go wrong with it

 Reference Summary (Cleaned):
 nice and simple, basic tadka daal!

 Candidate Summary:
 nice and simple, basic tadka daal!

 BLEU Score: 1.0000000000000004
 ROUGE Score: 0.999999995

 LSA Top 5 NEWS AND SUMMARIES - ROUGE SCORES
 Original Text (Cleaned):
 think thin products are great unless they are shipped from all star health and then
they are spoiled
 be very careful when buying think thin bars from amazon
  i bought some bars
 they made me sick
  amazon and all star health did nothing about it

 Reference Summary (Cleaned):
 think thin products are great unless they are shipped from all star health and then
they are spoiled

 Candidate Summary:
 think thin products are great unless they are shipped from all star health and then
they are spoiled

 BLEU Score: 0.942615147681512
 ROUGE Score: 0.999999995

 Original Text (Cleaned):
 open air cups - not a good idea

better off with real k cups
 open air cups - not a good idea
better off with real k cups

Reference Summary (Cleaned):
open air cups - not a good idea

Candidate Summary:
open air cups - not a good idea

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

Original Text (Cleaned):
 fantastic!you can not beat this taste and you can not resist to it only one if you a
re spice lover

Reference Summary (Cleaned):
 fantastic!you can not beat this taste and you can not resist to it only one if you a
re spice lover

Candidate Summary:
 fantastic!you can not beat this taste and you can not resist to it only one if you a
re spice lover

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

Original Text (Cleaned):
good coffee great price
doesn't look like the kips but works just as well
needs to be stored in an air tight container
i will purchase again

Reference Summary (Cleaned):
good coffee great price

Candidate Summary:
good coffee great price

BLEU Score: 0.6687403049764218
ROUGE Score: 0.999999995

Original Text (Cleaned):
good basic greek coffee
 can depend on this coffee to be the same from bag to bag

Reference Summary (Cleaned):
good basic greek coffee

Candidate Summary:
good basic greek coffee

BLEU Score: 0.6687403049764218
ROUGE Score: 0.999999995

In [11]:
```python
# Experiments for LexRank Model
from sumy.summarizers.lex_rank import LexRankSummarizer
lex_summarizer = LexRankSummarizer(Stemmer(language))
lex_summarizer.stop_words = get_stop_words(language)

lex_bleu_sorted, lex_rouge_sorted = summarize(lex_summarizer, "LexRank", id_complete_r

# Print the top 5 summaries along with their BLEU and ROUGE scores
print("\033[1m LexRank Top 5 NEWS AND SUMMARIES - BLEU SCORES \033[0m")
pretty_print(lex_bleu_sorted.head(5).iterrows(), "LexRank")
print("\033[1m LexRank Top 5 NEWS AND SUMMARIES - ROUGE SCORES \033[0m")
pretty_print(lex_rouge_sorted.head(5).iterrows(), "LexRank")
```

Original Text (Cleaned):
love this stuff, dn't need cream/mik or sugar for my coffee, saves 100+ calories per cup!

Reference Summary (Cleaned):
love this stuff, dn't need cream/mik or sugar for my coffee, saves 100+ calories per cup!

Candidate Summary:
love this stuff, dn't need cream/mik or sugar for my coffee, saves 100+ calories per cup!

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

Original Text (Cleaned):
loved this drink mix
 tasty, easy to use
 loved the lower calories
 plan to order this product, again
 thirst quenching

Reference Summary (Cleaned):
loved this drink mix

Candidate Summary:
loved this drink mix

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

Original Text (Cleaned):
 best ramen hands down!!! the noodles are bigger and have better flavor and the broth has such better flavoring than top ramen and maruchan
 they are pricier but if you're eating ramen by itself as a meal, this is the way to go
 if you're just boiling noodles for stir fry or whatever, you can get away with other brands
 the kimchi flavor is by far the best
 i tried this brand/flavor when it was on sale on a whim not knowing if it would be good or not and boy was i surprised
the noodles come in a brick that is circular! no more breaking up square bricks to fit in small pans! they come with a seasoning packet and a vegetable packet which is nice
 you definitely don't need the entire seasoning packet

in fact you can get by with half of it without it tasting too weak, but greatly reducing your sodium intake
 they are on the spicy side but so is kimchi(cabbage rubbed in chili powder and pickled)
if you just want to try a single bowl or foil package try your local asian grocer or

sometimes my local wally-world carries the styrofoam bowls in the kimchi flavor around a buck

Reference Summary (Cleaned):
best ramen hands down!!!

Candidate Summary:
best ramen hands down!!!

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

Original Text (Cleaned):
 great value & taste!!i love coconut water and we drink a lot in my household
it's great (actually the perfect thing) for rehydration after a night of drinking, long weekend, or after hard work
it has potassium, a few vitamins, and a weakly flavored but pleasant taste that really grows on you
i've tasted a lot of brands and zico has probably the best taste of them all
and i've shopped around on amazon, which is already much cheaper than your grocery store or gas station
and this item, with it's larger volumes (which stay fresh), and low pricehas proven to be the lowest priced (cheapest) price per ounce -- definitely the greatest value you can find!so i highly recommend this item, especially if you like coconut water and want the cheapest for the best product

Reference Summary (Cleaned):
 great value & taste!

Candidate Summary:
 great value & taste!

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

Original Text (Cleaned):
 want a pleasant way to start off your day? if your answer is yes, then start it off with a big cup of paramount's blueberry muffin coffee
  i find the paramount brand of coffee to be of very good quality and taste
  so far, blueberry muffin is my favorite but i am looking forward to trying their cherry flavored coffee next
  enjoy!

Reference Summary (Cleaned):
 want a pleasant way to start off your day?

Candidate Summary:
 want a pleasant way to start off your day?

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

LexRank Top 5 NEWS AND SUMMARIES - ROUGE SCORES
Original Text (Cleaned):
great product, nice sweet after dinner treat or trail pick me up
 healthy and delicious! everyone i share one with, loves them too
 i should be getting a commission

Reference Summary (Cleaned):
great product, nice sweet after dinner treat or trail pick me up

Candidate Summary:
great product, nice sweet after dinner treat or trail pick me up

BLEU Score: 0.9193227152249175
ROUGE Score: 0.999999995

Original Text (Cleaned):
my dog loves these
 i received a sample at the vet's office where i got my puppy and he loved them
 i like the fact that they help to clean his teeth as well as satisfy his need to ch
ew

Reference Summary (Cleaned):
my dog loves these

Candidate Summary:
my dog loves these

BLEU Score: 0.6687403049764218
ROUGE Score: 0.999999995

Original Text (Cleaned):
 very satisfied
 case arrived in 3 days and it was "fresh", meaning the expiration date was over 2 1/
2 years from now, may, 2015, which is better than the stuff that i was getting locall
y at pet valu
 the kitties like it

Reference Summary (Cleaned):
 very satisfied

Candidate Summary:
 very satisfied

BLEU Score: 0.5503212081491042
ROUGE Score: 0.999999995

Original Text (Cleaned):
my poodle loves greenies! we joke that they're like crack to her

she demands one every night at 8 pm
   i can't complain too much though - her breath is great!

Reference Summary (Cleaned):
my poodle loves greenies!

Candidate Summary:
my poodle loves greenies!

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

Original Text (Cleaned):
always a good condiment
   if you are unfamiliar with it, it is a spicy pepper sauce in an oil base without th
e saltiness
   i like it
jim

Reference Summary (Cleaned):
always a good condiment

Candidate Summary:
always a good condiment

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

In [12]:
```python
# Experiments for LuhnSummarizer
from sumy.summarizers.luhn import LuhnSummarizer
luhn_summarizer = LuhnSummarizer(Stemmer(language))
luhn_summarizer.stop_words = get_stop_words(language)

luhn_bleu_sorted, luhn_rouge_sorted = summarize(luhn_summarizer, "Luhn", id_complete_r

# Print the top 5 summaries along with their BLEU and ROUGE scores
print("\033[1m Luhn Top 5 NEWS AND SUMMARIES - BLEU SCORES \033[0m")
pretty_print(luhn_bleu_sorted.head(5).iterrows(), "Luhn")
print("\033[1m Luhn Top 5 NEWS AND SUMMARIES - ROUGE SCORES \033[0m")
pretty_print(luhn_rouge_sorted.head(5).iterrows(), "Luhn")
```

Luhn Top 5 NEWS AND SUMMARIES - BLEU SCORES
Original Text (Cleaned):
yummy! i can't believe its 5lbs, i had to give some away there's so much! i personally wouldn't have purchased this at the retail price, but had to jump on it during amazon's 50% off chocolates sale!

Reference Summary (Cleaned):
yummy!

Candidate Summary:
yummy!

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

Original Text (Cleaned):
"i received a free sample of 5-hour energy® from smiley360
" when i first tried this i was apprehensive that it might make me jittery
  not the case
  i was a great pick me up just the right amount
  no jitters or odd feelings
  and when it was time for bed i had no problem going to sleep and didn't feel any crash sensation
  i would recommend this if you have things to do or need a little pick me up in the afternoon
  i tried it before 3 pm and by 9 pm i went to bed with no problem and slept like a baby
"i received a free sample of 5-hour energy® from smiley360
"

Reference Summary (Cleaned):
"i received a free sample of 5-hour energy® from smiley360
"

Candidate Summary:
"i received a free sample of 5-hour energy® from smiley360
"

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

Original Text (Cleaned):
the best chewable texture and mildly seasoned
different than jack's, bill's and pete's jerky
i had one bag once that seemed to be tougher but otherwise quality control is very consistant

Reference Summary (Cleaned):
the best chewable texture and mildly seasoned

Candidate Summary:
the best chewable texture and mildly seasoned

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

Original Text (Cleaned):
some kinda good !!! plain and simple food is sometimes the best and it is here !half
the calories of their other spam products and a better alternative to fast food
make your own lunch with this and enjoy a product made in america that is some kinda
good !!!

Reference Summary (Cleaned):
some kinda good !!!

Candidate Summary:
some kinda good !!!

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

Original Text (Cleaned):
very refreshing! ingredients are 3 different fruit juices, sparkling water, and no
sugar! about 150 calories per can, and they're smaller than regular soda cans
  so the ingredients are pretty natural, but 150 calories might be too much for some
people!

Reference Summary (Cleaned):
very refreshing!

Candidate Summary:
very refreshing!

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

Luhn Top 5 NEWS AND SUMMARIES - ROUGE SCORES
Original Text (Cleaned):
great product
 as described and delivered promptly
 the gel is easy to make and goes well with juice
 we use this daily and my kids think it is hysterical!

Reference Summary (Cleaned):
great product

Candidate Summary:
great product

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

Original Text (Cleaned):

see's candy is delicious
there were shops in california where i grew up
 now i live in new england and there is no other way to get it except mail order
i'll take it! ;)

Reference Summary (Cleaned):
see's candy is delicious

Candidate Summary:
see's candy is delicious

BLEU Score: 0.6687403049764218
ROUGE Score: 0.999999995

Original Text (Cleaned):
 who doesn't love terry's chocolate orange! and cos it is british chocolate it is "c
orn syrup free" so no rashes the next day! i love this brand, so rich and creamy yet
not as sweet as cadbury's which is too sickening for me! this is the perfect treat!

Reference Summary (Cleaned):
who doesn't love terry's chocolate orange!

Candidate Summary:
who doesn't love terry's chocolate orange!

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

Original Text (Cleaned):
never showed up
don't know what happend to it
no one has tried to get in contact with me
not to happy about this

Reference Summary (Cleaned):
never showed up

Candidate Summary:
never showed up

BLEU Score: 0.5946035575013604
ROUGE Score: 0.999999995

Original Text (Cleaned):
loved this drink mix
 tasty, easy to use
 loved the lower calories
 plan to order this product, again
 thirst quenching

Reference Summary (Cleaned):
loved this drink mix

Candidate Summary:
loved this drink mix

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

In [13]:
```python
# Experiemnts for TextRankSummarizer
from sumy.summarizers.text_rank import TextRankSummarizer
text_rank_summarizer = TextRankSummarizer(Stemmer(language))
text_rank_summarizer.stop_words = get_stop_words(language)

text_rank_bleu_sorted, text_rank_rouge_sorted = summarize(text_rank_summarizer, "TextF

# Print the top 5 summaries along with their BLEU and ROUGE scores
print("\033[1m TextRank Top 5 NEWS AND SUMMARIES - BLEU SCORES \033[0m")
pretty_print(text_rank_bleu_sorted.head(5).iterrows(), "TextRank")
print("\033[1m TextRank Top 5 NEWS AND SUMMARIES - ROUGE SCORES \033[0m")
pretty_print(text_rank_rouge_sorted.head(5).iterrows(), "TextRank")
```

TextRank Top 5 NEWS AND SUMMARIES - BLEU SCORES
Original Text (Cleaned):
great product, love it! add cinnamon to it, and it will taste even better :)

Reference Summary (Cleaned):
great product, love it!

Candidate Summary:
great product, love it!

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

Original Text (Cleaned):
loved this drink mix
 tasty, easy to use
 loved the lower calories
 plan to order this product, again
 thirst quenching

Reference Summary (Cleaned):
loved this drink mix

Candidate Summary:
loved this drink mix

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

Original Text (Cleaned):
i like this coffee
 that says it all! why should i go on when that is all to be said

Reference Summary (Cleaned):
i like this coffee

Candidate Summary:
i like this coffee

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

Original Text (Cleaned):
expired chips!!!my husband and i went to london on our honeymoon and loved these chips
 i thought purchasing these chips would be a great addition to my anniversay gift to him
 big mistake

the chips expired over 7 months ago

Reference Summary (Cleaned):
expired chips!!

Candidate Summary:
expired chips!!

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

Original Text (Cleaned):
love this stuff, dn't need cream/mik or sugar for my coffee, saves 100+ calories per
cup!

Reference Summary (Cleaned):
love this stuff, dn't need cream/mik or sugar for my coffee, saves 100+ calories per
cup!

Candidate Summary:
love this stuff, dn't need cream/mik or sugar for my coffee, saves 100+ calories per
cup!

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

TextRank Top 5 NEWS AND SUMMARIES - ROUGE SCORES
Original Text (Cleaned):
order and delivery eas excellent
coffee i am still looking for a rich good taste coffee
did nott find in this flavor but better then the mountain

Reference Summary (Cleaned):
i am still looking for a rich good taste coffee

Candidate Summary:
coffee i am still looking for a rich good taste coffee

BLEU Score: 0.9036020036098446
ROUGE Score: 0.999999995

Original Text (Cleaned):
mccormick montreal chicken seasoning is excellent! it tastes really good on chicken
 it is not too salty
 it is just right
 it has a mixture of the perfect blends
 using this product takes a lot of the guess work out of seasoning your meat for a b
arbebecue

i hope mccormick does not ever discontinue this product
 two thumbs up on this one mccormick!!

Reference Summary (Cleaned):
mccormick montreal chicken seasoning is excellent!

Candidate Summary:
mccormick montreal chicken seasoning is excellent!

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

Original Text (Cleaned):
good product at a good price
the texture is excellent
cooks evenly
living in the country far from a source of asian noodles this service is great
now if they will start offering japanese spaghetti sauce

Reference Summary (Cleaned):
good product at a good price

Candidate Summary:
good product at a good price

BLEU Score: 1.0000000000000004
ROUGE Score: 0.999999995

 Original Text (Cleaned):
 great tea
  if anyone has not tried loose tea they need to, because they don't know what they a
re missing
  i was very glad to find this product again
  many thanks for keeping this wonderful beverage
  fast shipment and great service

 Reference Summary (Cleaned):
 great tea

 Candidate Summary:
 great tea

 BLEU Score: 0.5503212081491042
 ROUGE Score: 0.999999995

 Original Text (Cleaned):
 i love the product this is the fourth time i've re-ordered
 i like my coffee iced, even in winter!

Reference Summary (Cleaned):
i love the product this is the fourth time i've re-ordered

Candidate Summary:
i love the product this is the fourth time i've re-ordered

BLEU Score: 0.9036020036098446
ROUGE Score: 0.999999995

RESULTS

In [14]:
```python
# Define the data
summarizers = ['LSA', 'LexRank', 'Luhn', 'TextRank']
BLEU_scores = [lsa_bleu_sorted['BLEU Score'].mean() * 100,
               lex_bleu_sorted['BLEU Score'].mean() * 100,
               luhn_bleu_sorted['BLEU Score'].mean() * 100,
               text_rank_bleu_sorted['BLEU Score'].mean() * 100]

# Define the colors for the bars
colors = ['#c7e9c0', '#41b6c4', '#225ea8', '#253494']

# Create the plot
fig, ax = plt.subplots()
ax.bar(summarizers, BLEU_scores, color=colors)

# Set the axis labels and title
ax.set_xlabel('Summarizer')
ax.set_ylabel('BLEU Score')
ax.set_title('Average BLEU Scores for Different Summarization Techniques')

# Add text over the bars
for i, v in enumerate(BLEU_scores):
    ax.text(i, v + 1, str(round(v, 2)), ha='center', fontweight='bold')

# Set the y-axis limits
ax.set_ylim([0, max(BLEU_scores) * 1.4])

# Show the plot
plt.show()
```
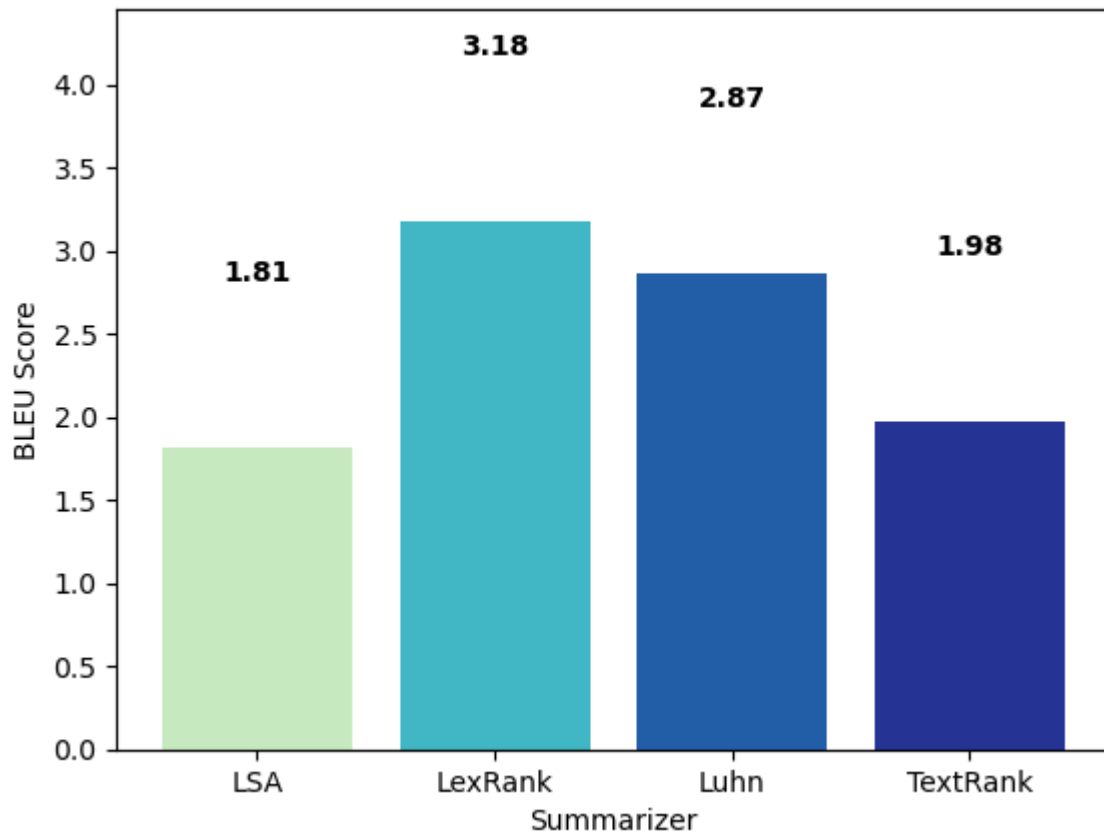
Average BLEU Scores for Different Summarization Techniques


```
In [15]:  # Define the data
          summarizers = ['LSA', 'LexRank', 'Luhn', 'TextRank']
          ROUGE_scores = [lsa_rouge_sorted['ROUGE Score'].mean() * 100,
                          lex_rouge_sorted['ROUGE Score'].mean() * 100,
                          luhn_rouge_sorted['ROUGE Score'].mean() * 100,
                          text_rank_rouge_sorted['ROUGE Score'].mean() * 100]
          colors = ['#FF5733', '#E74C3C', '#C0392B', '#922B21']

          # Create the plot
          fig, ax = plt.subplots()
          ax.bar(summarizers, ROUGE_scores, color=colors)

          # Set the axis labels and title
          ax.set_xlabel('Summarizer')
          ax.set_ylabel('ROUGE Score')
          ax.set_title('Average ROUGE Scores for Different Summarization Techniques')

          # Add text over the bars
          for i, v in enumerate(ROUGE_scores):
              ax.text(i, v + 1, str(round(v, 2)), ha='center', fontweight='bold')
          ax.set_ylim([0, max(ROUGE_scores) * 1.2])
          plt.show()
```

Average ROUGE Scores for Different Summarization Techniques