

Введение в обработку естественного языка

Урок 12. Модель Transformer-2

Практическое задание

Домашнее задание к уроку 11

Задание

Взять тот же датасет, который был на вебинаре и предобученную модель для задачи суммаризации

1. Проверить насколько хорошо она суммаризирует
2. (дополнительно) Сделать генерацию заголовков для статьи (обучить модель для генерации заголовков)

Выполнил **Соковнин ИЛ**

B []: !pip install transformers

```
Looking in indexes: https://pypi.org/simple, (https://pypi.org/simple,) https://us-python.pkg.dev/colab-wheels/public/simple/ (https://us-python.pkg.dev/colab-wheels/public/simple/)
Collecting transformers
  Downloading transformers-4.20.1-py3-none-any.whl (4.4 MB)
    |████████████████████| 4.4 MB 4.2 MB/s
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.7/dist-packages (from transformers) (1.21.6)
Requirement already satisfied: importlib-metadata in /usr/local/lib/python3.7/dist-packages (from transformers) (4.11.4)
Collecting huggingface-hub<1.0,>=0.1.0
  Downloading huggingface_hub-0.8.1-py3-none-any.whl (101 kB)
    |██████████████████| 101 kB 11.4 MB/s
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.7/dist-packages (from transformers) (4.64.0)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.7/dist-packages (from transformers) (21.3)
Requirement already satisfied: filelock in /usr/local/lib/python3.7/dist-packages (from transformers) (3.7.1)
Collecting pyyaml>=5.1
  Downloading PyYAML-6.0-cp37-cp37m-manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_12_x86_64.manylinux2010_x86_64.whl (596 kB)
    |██████████████████| 596 kB 45.3 MB/s
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.7/dist-packages (from transformers) (2022.6.2)
Collecting tokenizers!=0.11.3,<0.13,>=0.11.1
  Downloading tokenizers-0.12.1-cp37-cp37m-manylinux_2_12_x86_64.manylinux2010_x86_64.whl (6.6 MB)
    |██████████████████| 6.6 MB 39.4 MB/s
Requirement already satisfied: requests in /usr/local/lib/python3.7/dist-packages (from transformers) (2.23.0)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.7/dist-packages (from huggingface-hub<1.0,>=0.1.0->transformers) (4.1.1)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /usr/local/lib/python3.7/dist-packages (from packaging>=20.0->transformers) (3.0.9)
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (from importlib-metadata->transformers) (3.8.0)
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages (from requests->transformers) (2.10)
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packages (from requests->transformers) (3.0.4)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packages (from requests->transformers) (2022.6.15)
Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in /usr/local/lib/python3.7/dist-packages (from requests->transformers) (1.24.3)
Installing collected packages: pyyaml, tokenizers, huggingface-hub, transformers
  Attempting uninstall: pyyaml
    Found existing installation: PyYAML 3.13
    Uninstalling PyYAML-3.13:
      Successfully uninstalled PyYAML-3.13
Successfully installed huggingface-hub-0.8.1 pyyaml-6.0 tokenizers-0.12.1 transformers-4.20.1
```

B [2]: !wget -q https://www.dropbox.com/s/431702z5a5i2w8j/gazeta_train.txt
!wget -q https://www.dropbox.com/s/k2egt3sug0hb185/gazeta_val.txt
!wget -q https://www.dropbox.com/s/3gki5n5djs9w0v6/gazeta_test.txt

B [3]: !ls

gazeta_test.txt gazeta_train.txt gazeta_val.txt sample_data

B [4]: !pip install razdel networkx pymorphy2[fast] nltk rouge==0.3.1 summa lextank sumy
!pip install --upgrade datasets razdel spacy networkx seaborn scipy tqdm tensorflow-text
!python -m spacy download ru_core_news_md

```
Looking in indexes: https://pypi.org/simple, (https://pypi.org/simple,) https://us-python.pkg.dev/colab-wheels/public/simple/ (https://us-python.pkg.dev/colab-wheels/public/simple/)
Collecting razdel
  Downloading razdel-0.5.0-py3-none-any.whl (21 kB)
Requirement already satisfied: networkx in /usr/local/lib/python3.7/dist-packages (2.6.3)
Collecting pymorphy2[fast]
  Downloading pymorphy2-0.9.1-py3-none-any.whl (55 kB)
    |██████████████████| 55 kB 2.2 MB/s
Requirement already satisfied: nltk in /usr/local/lib/python3.7/dist-packages (3.7)
Collecting rouge==0.3.1
  Downloading rouge-0.3.1-py3-none-any.whl (7.4 kB)
Collecting summa
  Downloading summa-1.2.0.tar.gz (54 kB)
    |██████████████████| 54 kB 3.4 MB/s
Collecting lextank
  Downloading lextank-0.1.0-py3-none-any.whl (69 kB)
    |██████████████████| 69 kB 7.2 MB/s
Collecting sumy
  Downloading sumy-0.10.0-py2.py3-none-any.whl (94 kB)
    |██████████████████| 94 kB 3.4 MB/s
```

Загрузим данные с помощью библиотеки datasets

```
B [5]: # по сути данные те же можно пользоваться любым способом загрузки
from datasets import load_dataset

dataset = load_dataset('IlyaGusev/gazeta', revision="v1.0")["test"]

Downloading builder script: 0%|          | 0.00/2.98k [00:00<?, ?B/s]

Downloading metadata: 0%|          | 0.00/3.87k [00:00<?, ?B/s]

No config specified, defaulting to: gazeta/default

Downloading and preparing dataset gazeta/default (download: 545.11 MiB, generated: 542.44 MiB, post-processed: Unknown size, total: 1.06 GiB) to /root/.cache/huggingface/datasets/IlyaGusev___gazeta/default/1.0.0/ef9349c3c0f3112ca4036520d76c4bc1b8a79d30bc29643c6cae5a094d44e457...

Downloading data files: 0%|          | 0/3 [00:00<?, ?it/s]

Downloading data: 0%|          | 0.00/471M [00:00<?, ?B/s]

Downloading data: 0%|          | 0.00/48.6M [00:00<?, ?B/s]

Downloading data: 0%|          | 0.00/52.1M [00:00<?, ?B/s]

Extracting data files: 0%|          | 0/3 [00:00<?, ?it/s]

Generating train split: 0%|          | 0/52400 [00:00<?, ? examples/s]

Generating test split: 0%|          | 0/5770 [00:00<?, ? examples/s]

Generating validation split: 0%|          | 0/5265 [00:00<?, ? examples/s]

Dataset gazeta downloaded and prepared to /root/.cache/huggingface/datasets/IlyaGusev___gazeta/default/1.0.0/ef9349c3c0f3112ca4036520d76c4bc1b8a79d30bc29643c6cae5a094d44e457. Subsequent calls will reuse this data.

0%|          | 0/3 [00:00<?, ?it/s]
```

EDA (Exploratory data analysis)

Посмотрим на то, как устроен датасет

```
B [11]: !head -n 1 gazeta_train.txt
!cat gazeta_train.txt | wc -l
!cat gazeta_val.txt | wc -l
!cat gazeta_test.txt | wc -l

{"url": "https://www.gazeta.ru/financial/2011/11/30/3852658.shtml", "text": "«По итогам 2011 года чистый отток может составить примерно $80 млрд, в следующем году – около $20 млрд. При этом мы ожидаем, что со второго полугодия 2012 года начнется приток капитала», – заявил «Интерфаксу» замминистра экономического развития Андрей Клепач. Официальные прогнозы по выводу капитала из России становятся все пессимистичными: еще летом власти полагали, что из страны уйдет не более $35 млрд, в сентябре Минэкономразвития назвал цифру $50 млрд, в начале ноября Центробанк пересмотрел оценку до $70 млрд. Очередное изменение прогноза было ожидаемо: по расчетам Центр обанка , за январь – октябрь чистый отток капитала достиг $64 млрд, причем в последние месяцы он ускорился: в сентябре он составил $14 млрд, в октябре – $13 млрд пр отив среднего ежемесячного оттока в $6–8 млрд в первом полугодии. «После октябрьских данных Минэкономразвития вынуждено было изменить оценку, настаивать на $70 млрд означало ожидать серьезного замедления оттока капитала на непонятно каких причинах», – говорит главный экономист BNP Paribas Юлия Цепляева. «В последние два месяца отток капитала ускорится, на декабрь приходится значительная часть выплат по внешним долгам, что приводит к усилению оттока, особенно если они не рефинансируются но выми кредитами», – соглашается главный экономист ФК «Открытие» Владимир Тихомиров. Прогнозируемый Минэкономразвития отток капитала – один из самых высоких за послед ние 20 лет. Больше ушло лишь в 2008 году на фоне разрастания финансового кризиса и российско-грузинской войны – $133,7 млрд. В кризисный 2009 год из России утекло $56,1 млрд. Главный фактор ускорения оттока капитала в 2011 году – нестабильность на внешних финансовых рынках и рост опасений относительно второй волны рецессии. «Это реакция на неуверенность, которую генерирует Европа с долговыми проблемами. В случае новой волны глобальной турбулентности Россия – одна из самых уязвимых стра н», – говорит Цепляева. Еще одна причина – ослабление рубля. «Привлекательность вложений снижается на фоне того, что рубль перестал укрепляться, а ставки по депозит ам достаточно низкие. В результате экспортеры не полностью возвращают экспортную выручку», – говорит Тихомиров. Внутри страны эксперты не видят особых причин для бе гства капитала. «Ситуация выглядит достаточно позитивно, очень хорошие макроэкономические результаты за год, особенно на фоне других стран. С политической точки зре ния все достаточно понятно и предсказуемо, итог выборов очевиден», – говорит экономист ИК «Тройка Диалог» Антон Струченевский. Тем не менее политический фактор игра ет роль. «Бизнесу важно не только, кто будет президентом, он ждет ясности с перестановками в правительстве. В наших условиях административный ресурс важнее всего дл я успешности бизнеса», – говорит Цепляева, добавляя, что отток капитала продолжится до завершения президентских выборов.", "title": "Прогноз не успевает за оттоко м", "summary": "В 2011 году из России уйдет $80 млрд, считают в Минэкономразвития. Менее месяца назад Центробанк давал оценку $70 млрд, повысив первоначальный прогн оз вдвое. Отток капитала из страны усиливается из-за кризиса в Европе, а в декабре российским компаниям выплачивать внешние долги. На движение капитала повлияли и в ыборы: несмотря на их предсказуемость, бизнес хочет ясности с перестановками в правительстве.", "date": "2011-11-30 18:33:39"}
52400
5265
5770

B [7]: import json
import random

def read_gazeta_records(file_name, shuffle=True, sort_by_date=False):
    assert shuffle != sort_by_date
    records = []
    with open(file_name, "r") as r:
        for line in r:
            records.append(json.loads(line))
    if sort_by_date:
        records.sort(key=lambda x: x["date"])
    if shuffle:
        random.shuffle
    return records

B [8]: train_records = read_gazeta_records("gazeta_train.txt")
val_records = read_gazeta_records("gazeta_val.txt")
test_records = read_gazeta_records("gazeta_test.txt")
```

```
B [14]: test_records[0]
```

```
Out[14]: {'date': '2020-02-14 16:39:11',
'summary': 'В NASA назвали четыре миссии в дальний космос, которые в этом десятилетии могут быть запущены американцами. Среди них – две миссии по изучению Венеры, полет к спутнику Юпитера и экспедиция к Тритону, спутнику Нептуна.',
'text': 'Американское аэрокосмическое агентство NASA огласило названия четырех космических миссий, которые в скором времени могут быть выбраны для реализации и запуск которых может состояться уже в конце этого десятилетия. Эти четыре проекта стали полуфиналистами конкурса, объявленного среди американских научных команд, в котором участвовало более десяти миссий. Все они были отобраны по критериям потенциальной пользы для науки и технической осуществимости проекта. В рамках программы Discovery NASA занимается планированием миссий, которые призваны дать ответы на фундаментальные вопросы о происхождении тел Солнечной системы и возможному наличию жизни на них. «Эти выбранные миссии могут трансформировать наше восприятие некоторых из наиболее активных и сложных миров в Солнечной системе, – заявил Томас Зурбучен, помощник директора NASA по науке. – Исследование каждого из этих небесных тел поможет раскрыть секреты о том, как они и им подобные объекты образовались в космосе». Каждый проект из выбранных полуфиналистов получит $3 млн на проработку концепции. Из четырех выбранных проектов NASA планирует отобрать две в 2021 году, чтобы профинансировать работу по их реализации в рамках программы Discovery. Два из четырех проектов посвящены исследованию Венеры, куда аппараты NASA не отправлялись уже свыше тридцати лет – с 1989 года, когда к этой планете был запущен зонд Magellan. Миссия VERITAS (Venus Emissivity, Radio Science, InSAR, Topography, and Spectroscopy) призвана картировать поверхность Венеры для лучшего понимания геологической истории планеты и получения ответа на вопрос о причинах ее кардинальных отличий от Земли. Карта будет построена при помощи радара, установленного на орбитальном аппарате. В ходе миссии планируется узнать, продолжают ли в наши дни на Венере тектонические и вулканические процессы. Кроме того, планируется картировать поверхность планеты в инфракрасном диапазоне. Миссия DAVINCI+ (Deep Atmosphere Venus Investigation of Noble gases, Chemistry and Imaging Plus) задумана для того, чтобы исследовать состав атмосферы Венеры, понять, как она формировалась и эволюционировала в прошлом, и ответить на вопрос о существовании когда-либо на планете жидкого океана. Сделать это планируется при помощи спускаемого аппарата, который будет проводить измерения от верхних слоев атмосферы до самой поверхности. Научные инструменты аппарата будут размещены внутри герметичного сферического корпуса, который должен защитить их от разрушительного воздействия высокой температуры атмосферы. + в названии миссии говорит о наличии в ее составе камер, которые будут присутствовать на борту орбитального и спускаемого аппаратов. «Проект DAVINCI+ пересекается по целям с российской миссией «Венера-Д», обе нацелены на посадку, поэтому в будущем возможно их совмещение в одном совместном проекте», – сказал «Газете.Ru» основатель проекта «Открытый космос» Виталий Егоров. Миссия TRIDENT должна исследовать Тритон – ледяной спутник Нептуна, известный ученым своей активностью. Данные, полученные американским зондом Voyager 2, показали, что поверхность Тритона активно обновляется. Ученые считают, что Тритон, на котором выпадает «снег» из органических веществ, может оказаться отличной целью для понимания вопросов происхождения аналогичных тел в Солнечной системе. Пролетев мимо Тритона, аппарат попытается картировать его поверхность, изучить активные процессы на его поверхности и определить, есть ли под его поверхностью жидкий океан. Миссия IVO (Io Volcano Observer) должна направиться к одному из самых интересных спутников Юпитера – Ио. Это тело с наиболее мощной вулканической активностью в Солнечной системе, которая поддерживается за счет мощного приливного воздействия со стороны Юпитера. Во время нескольких близких пролетов аппарат выяснит, как формируется лава на поверхности спутника.',
'title': 'Венера, Ио или Тритон: куда полетит NASA',
'url': 'https://www.gazeta.ru/science/2020/02/14_a_12960289.shtml'}
```

```
B [15]: train_records[0].keys()
```

```
Out[15]: dict_keys(['url', 'text', 'title', 'summary', 'date'])
```

```
B [16]: print(min([record["date"] for record in train_records]), end=' - ')
print(max([record["date"] for record in train_records]))
print(min([record["date"] for record in val_records]), end=' - ')
print(max([record["date"] for record in val_records]))
print(min([record["date"] for record in test_records]), end=' - ')
print(max([record["date"] for record in test_records]))
```

2010-06-01 10:35:49 - 2019-05-31 23:56:26
2019-06-01 08:30:00 - 2019-09-30 23:11:23
2019-10-01 08:23:02 - 2020-03-23 22:16:23

Посчитаем статистику по словам

```
B [17]: from collections import Counter, namedtuple
import razdel
import pymorphy2

Stats = namedtuple("Stats", "vocabulary,lemma_vocabulary,words_counts,unique_words_counts")

def collect_stats(records, lower=True, text_max_words=3000, summary_max_words=100, nrows=1000):
    morph = pymorphy2.MorphAnalyzer()

    text_stats = Stats(Counter(), Counter(), list(), list())
    summary_stats = Stats(Counter(), Counter(), list(), list())

    def update_record_field_stats(field, stats, max_words):
        words = [word.text for word in razdel.tokenize(field)][0:max_words]
        lemmas = [morph.parse(word)[0].normal_form for word in words]
        stats.vocabulary.update(words)
        stats.lemma_vocabulary.update(lemmas)
        stats.words_counts.append(len(words))
        stats.unique_words_counts.append(len(set(words)))

    for i, record in enumerate(records):
        if i >= nrows:
            break
        text = record["text"]
        text = text if not lower else text.lower()
        update_record_field_stats(text, text_stats, text_max_words)

        summary = record["summary"]
        summary = summary if not lower else summary.lower()
        summary_words = [word.text for word in razdel.tokenize(summary)]
        update_record_field_stats(summary, summary_stats, summary_max_words)
    return text_stats, summary_stats
```

```
B [18]: train_text_stats, train_summary_stats = collect_stats(train_records)
print("Train texts vocabulary size: ", len(train_text_stats.vocabulary))
print("Train texts lemma vocabulary size: ", len(train_text_stats.lemma_vocabulary))
print("Train summaries vocabulary size: ", len(train_summary_stats.vocabulary))
print("Train summaries lemma vocabulary size: ", len(train_summary_stats.lemma_vocabulary))
print("Train common lemmas summary vs text: ", len(set(train_text_stats.lemma_vocabulary.keys()) & set(train_summary_stats.lemma_vocabulary.keys())))
```

Train texts vocabulary size: 82978
Train texts lemma vocabulary size: 37851
Train summaries vocabulary size: 15441
Train summaries lemma vocabulary size: 9030
Train common lemmas summary vs text: 8652

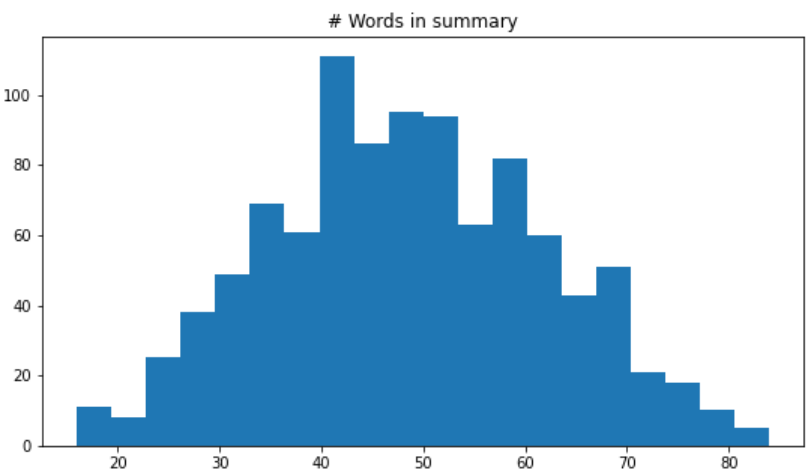
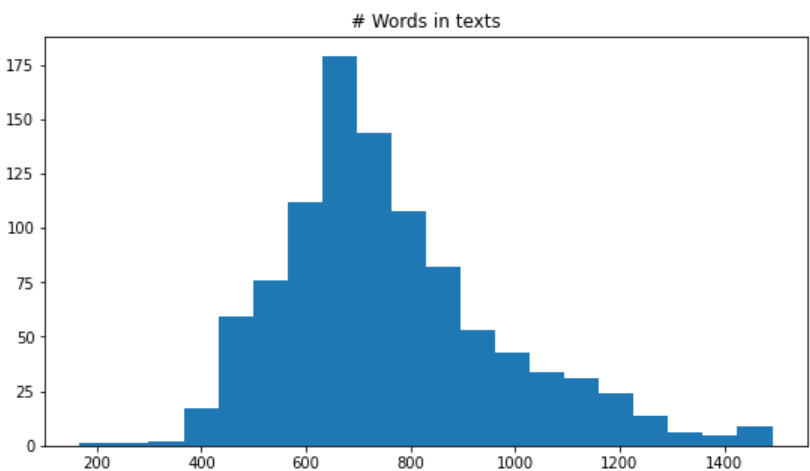
```
B [19]: train_text_stats.lemma_vocabulary.most_common(50)
```

```
Out[19]: [(',', 53334),
          ('.', 38444),
          ('в', 27978),
          ('и', 15538),
          ('«', 14801),
          ('»', 14714),
          ('на', 11761),
          ('не', 7995),
          ('с', 7964),
          ('–', 7920),
          ('что', 7498),
          ('быть', 6416),
          ('по', 5473),
          ('он', 5240),
          ('год', 4383),
          ('это', 4376),
          ('который', 3995),
          ('а', 3235),
          ('о', 2983),
          ('к', 2871),
          ('из', 2791),
          ('тот', 2791),
          ('они', 2720),
          ('за', 2715),
          ('как', 2542),
          (')', 2507),
          ('(', 2502),
          ('но', 2471),
          ('свой', 2417),
          (':', 2349),
          ('для', 2242),
          ('россия', 2082),
          ('от', 2027),
          ('она', 1999),
          ('этот', 1953),
          ('у', 1860),
          ('мы', 1801),
          ('до', 1761),
          ('мочь', 1722),
          ('я', 1665),
          ('один', 1595),
          ('также', 1587),
          ('после', 1540),
          ('-', 1490),
          ('весь', 1426),
          ('всё', 1414),
          ('первый', 1369),
          ('стать', 1369),
          ('%', 1290),
          ('уже', 1274)]
```

```
B [20]: import matplotlib.pyplot as plt

fig, axs = plt.subplots(1, 2, figsize=(20, 5))
axs[0].hist(train_text_stats.words_counts, 20)
axs[0].set_title('# Words in texts')

axs[1].hist(train_summary_stats.words_counts, 20)
axs[1].set_title('# Words in summary')
plt.show()
```



```
B [21]: test_text_stats, test_summary_stats = collect_stats(test_records)
print("Test texts vocabulary size: ", len(test_text_stats.vocabulary))
print("Test texts lemma vocabulary size: ", len(test_text_stats.lemma_vocabulary))
print("Test summaries vocabulary size: ", len(test_summary_stats.vocabulary))
print("Test summaries lemma vocabulary size: ", len(test_summary_stats.lemma_vocabulary))
print("Test common lemmas summary vs text: ", len(set(test_text_stats.lemma_vocabulary.keys()) & set(test_summary_stats.lemma_vocabulary.keys())))
```

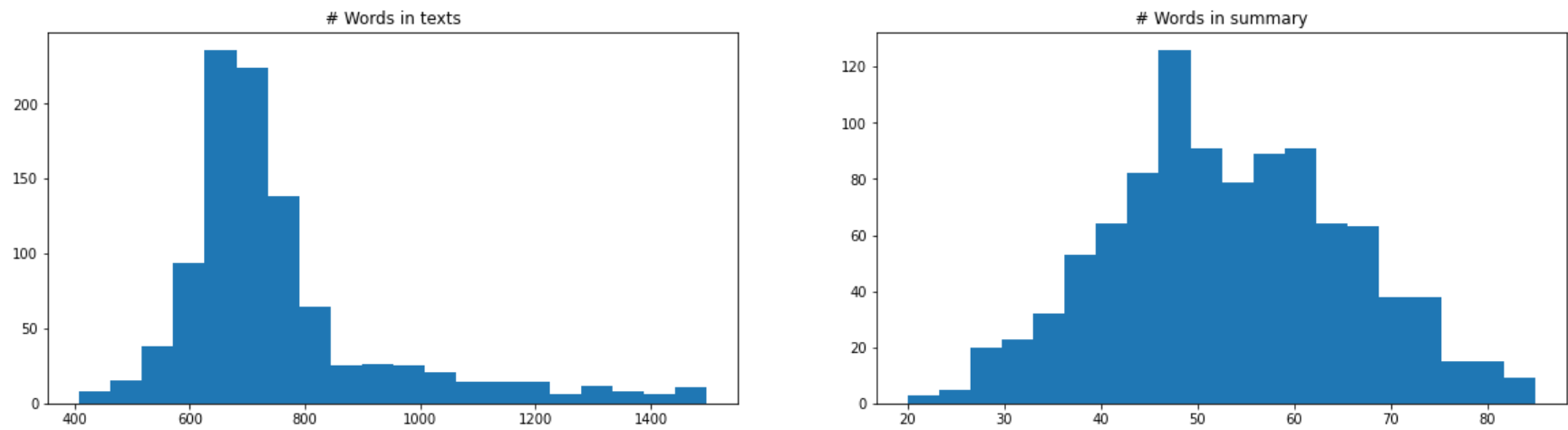
```
Test texts vocabulary size: 73282
Test texts lemma vocabulary size: 32583
Test summaries vocabulary size: 15196
Test summaries lemma vocabulary size: 8618
Test common lemmas summary vs text: 8353
```

```
B [22]: test_text_stats.lemma_vocabulary.most_common(50)
```

```
Out[22]: [(',', 54104),
          ('.', 38406),
          ('в', 28069),
          ('и', 14555),
          ('«', 12789),
          ('»', 12716),
          ('на', 11741),
          ('что', 9225),
          ('—', 8528),
          ('с', 8173),
          ('не', 7279),
          ('по', 6354),
          ('быть', 6240),
          ('он', 6065),
          ('это', 5560),
          ('год', 4990),
          ('который', 3920),
          ('о', 3888),
          ('тот', 3200),
          ('они', 3082),
          ('как', 2981),
          ('из', 2868),
          ('она', 2832),
          ('к', 2828),
          ('россия', 2713),
          ('а', 2597),
          ('свой', 2542),
          ('для', 2473),
          ('за', 2402),
          ('также', 2105),
          ('этот', 2089),
          ('от', 2061),
          ('я', 1865),
          ('мочь', 1855),
          ('мы', 1755),
          ('у', 1725),
          ('но', 1699),
          ('страна', 1657),
          ('русский', 1589),
          ('(', 1583),
          (')', 1583),
          ('человек', 1583),
          ('президент', 1582),
          ('сша', 1578),
          ('при', 1564),
          ('до', 1523),
          ('один', 1510),
          ('после', 1506),
          ('время', 1500),
          ('весь', 1476)]
```

```
B [23]: fig, axs = plt.subplots(1, 2, figsize=(20, 5))
axs[0].hist(test_text_stats.words_counts, 20)
axs[0].set_title('# Words in texts')

axs[1].hist(test_summary_stats.words_counts, 20)
axs[1].set_title('# Words in summary')
plt.show()
```



Используем предобученную модель: [csebuetnlp/mT5_multilingual_XLSum](https://huggingface.co/csebuetnlp/mT5_multilingual_XLSum) (https://huggingface.co/csebuetnlp/mT5_multilingual_XLSum)

B [10]:

```
# huggingface.co/models
# Знакомимся с сообществом huggingface, с его библиотекой transformers
!pip install transformers
!pip install transformers sentencepiece

Looking in indexes: https://pypi.org/simple, (https://pypi.org/simple,) https://us-python.pkg.dev/colab-wheels/public/simple/ (https://us-python.pkg.dev/colab-wheel
s/public/simple/)
Collecting transformers
  Downloading transformers-4.20.1-py3-none-any.whl (4.4 MB)
    |████████████████████| 4.4 MB 4.1 MB/s
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.7/dist-packages (from transformers) (4.64.0)
Requirement already satisfied: importlib-metadata in /usr/local/lib/python3.7/dist-packages (from transformers) (4.11.4)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.7/dist-packages (from transformers) (1.21.6)
Requirement already satisfied: filelock in /usr/local/lib/python3.7/dist-packages (from transformers) (3.7.1)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.7/dist-packages (from transformers) (6.0)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.7/dist-packages (from transformers) (2022.6.2)
Collecting tokenizers!=0.11.3,<0.13,>=0.11.1
  Downloading tokenizers-0.12.1-cp37-cp37m-manylinux_2_12_x86_64.manylinux2010_x86_64.whl (6.6 MB)
    |████████████████████| 6.6 MB 33.1 MB/s
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.7/dist-packages (from transformers) (21.3)
Requirement already satisfied: requests in /usr/local/lib/python3.7/dist-packages (from transformers) (2.23.0)
Requirement already satisfied: huggingface-hub<1.0,>=0.1.0 in /usr/local/lib/python3.7/dist-packages (from transformers) (0.8.1)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.7/dist-packages (from huggingface-hub<1.0,>=0.1.0->transformers) (4.1.1)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /usr/local/lib/python3.7/dist-packages (from packaging>=20.0->transformers) (3.0.9)
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (from importlib-metadata->transformers) (3.8.0)
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packages (from requests->transformers) (3.0.4)
Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in /usr/local/lib/python3.7/dist-packages (from requests->transformers) (1.25.11)
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages (from requests->transformers) (2.10)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packages (from requests->transformers) (2022.6.15)
Installing collected packages: tokenizers, transformers
Successfully installed tokenizers-0.12.1 transformers-4.20.1
Looking in indexes: https://pypi.org/simple, (https://pypi.org/simple,) https://us-python.pkg.dev/colab-wheels/public/simple/ (https://us-python.pkg.dev/colab-wheel
s/public/simple/)
Requirement already satisfied: transformers in /usr/local/lib/python3.7/dist-packages (4.20.1)
Collecting sentencepiece
  Downloading sentencepiece-0.1.96-cp37-cp37m-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (1.2 MB)
    |████████████████████| 1.2 MB 4.3 MB/s
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.7/dist-packages (from transformers) (1.21.6)
Requirement already satisfied: tokenizers!=0.11.3,<0.13,>=0.11.1 in /usr/local/lib/python3.7/dist-packages (from transformers) (0.12.1)
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.7/dist-packages (from transformers) (4.64.0)
Requirement already satisfied: huggingface-hub<1.0,>=0.1.0 in /usr/local/lib/python3.7/dist-packages (from transformers) (0.8.1)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.7/dist-packages (from transformers) (21.3)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.7/dist-packages (from transformers) (2022.6.2)
Requirement already satisfied: filelock in /usr/local/lib/python3.7/dist-packages (from transformers) (3.7.1)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.7/dist-packages (from transformers) (6.0)
Requirement already satisfied: importlib-metadata in /usr/local/lib/python3.7/dist-packages (from transformers) (4.11.4)
Requirement already satisfied: requests in /usr/local/lib/python3.7/dist-packages (from transformers) (2.23.0)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.7/dist-packages (from huggingface-hub<1.0,>=0.1.0->transformers) (4.1.1)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /usr/local/lib/python3.7/dist-packages (from packaging>=20.0->transformers) (3.0.9)
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (from importlib-metadata->transformers) (3.8.0)
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages (from requests->transformers) (2.10)
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packages (from requests->transformers) (3.0.4)
Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in /usr/local/lib/python3.7/dist-packages (from requests->transformers) (1.25.11)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packages (from requests->transformers) (2022.6.15)
Installing collected packages: sentencepiece
Successfully installed sentencepiece-0.1.96
```

Используем предобученную модель: IlyaGusev/mbart_ru_sum_gazeta

B [11]:

```
from transformers import MBartTokenizer, MBartForConditionalGeneration

model_name = "IlyaGusev/mbart_ru_sum_gazeta"
tokenizer = MBartTokenizer.from_pretrained(model_name)

model_mbart_ru_sum_gazeta = MBartForConditionalGeneration.from_pretrained(model_name)

article_text = "Американское аэрокосмическое агентство NASA огласило названия четырех космических миссий, которые в скором времени могут быть выбраны для реализации и запуска которых может состояться уже в конце этого десятилетия. Все они были отобраны по критериям потенциальной пользы для науки и технической осуществимости."

input_ids = tokenizer(
    [article_text],
    max_length=600,
    padding="max_length",
    truncation=True,
    return_tensors="pt",
)["input_ids"]

output_ids = model_mbart_ru_sum_gazeta.generate(
    input_ids=input_ids,
    no_repeat_ngram_size=4
)[0]

summary = tokenizer.decode(output_ids, skip_special_tokens=True)
print(summary)
```

Downloaded: 0%| | 0.00/4.83M [00:00<?, ?B/s]

Downloaded: 0%| | 0.00/406 [00:00<?, ?B/s]

Downloaded: 0%| | 0.00/287 [00:00<?, ?B/s]

Downloaded: 0%| | 0.00/1.05k [00:00<?, ?B/s]

Downloaded: 0%| | 0.00/3.23G [00:00<?, ?B/s]

Американское аэрокосмическое агентство NASA огласило названия четырех космических миссий, которые в скором времени могут быть выбраны для реализации и запуска которых может состояться уже в конце этого десятилетия. Все они были отобраны по критериям потенциальной пользы для науки и технической осуществимости.

B [3]:

```
# print(input_ids)
```

Используем предобученную модель: csebuetnlp/mT5_multilingual_XLSum

```
B [12]: import re
from transformers import AutoTokenizer, AutoModelForSeq2SeqLM

WHITESPACE_HANDLER = lambda k: re.sub('\s+', ' ', re.sub('\n+', ' ', k.strip()))

# article_text = """Videos that say approved vaccines are dangerous and cause autism, cancer or infertility are among those that will be taken down, the company said."
# article_text = "Американское аэрокосмическое агентство NASA огласило названия четырех космических миссий, которые в скором времени могут быть выбраны для реализации"

test_records[0]["text"]

model_name = "csebuetnlp/mT5_multilingual_XLSum"

# tokenizer = AutoTokenizer.from_pretrained(model_name)
tokenizer = AutoTokenizer.from_pretrained("csebuetnlp/mT5_multilingual_XLSum")
model_mT5_multilingual_XLSum = AutoModelForSeq2SeqLM.from_pretrained(model_name)

input_ids = tokenizer(
    [WHITESPACE_HANDLER(article_text)],
    return_tensors="pt",
    padding="max_length",
    truncation=True,
    max_length=512
)["input_ids"]

# print(input_ids)

output_ids = model_mT5_multilingual_XLSum.generate(
    input_ids=input_ids,
    max_length=84,
    no_repeat_ngram_size=2,
    num_beams=4
)[0]

summary = tokenizer.decode(
    output_ids,
    skip_special_tokens=True,
    clean_up_tokenization_spaces=False
)

print(summary)
```

Downloading: 0%| | 0.00/375 [00:00<?, ?B/s]
Downloading: 0%| | 0.00/730 [00:00<?, ?B/s]
Downloading: 0%| | 0.00/4.11M [00:00<?, ?B/s]
Downloading: 0%| | 0.00/65.0 [00:00<?, ?B/s]

/usr/local/lib/python3.7/dist-packages/transformers/convert_slow_tokenizer.py:435: UserWarning: The sentencepiece tokenizer that you are converting to a fast tokenizer uses the byte fallback option which is not implemented in the fast tokenizers. In practice this means that the fast version of the tokenizer can produce unknown tokens whereas the sentencepiece version would have converted these unknown tokens into a sequence of byte tokens matching the original piece of text.
"The sentencepiece tokenizer that you are converting to a fast tokenizer uses the byte fallback option"

Downloading: 0%| | 0.00/2.17G [00:00<?, ?B/s]

Космические миссии, которые могут дать ответы на вопросы о происхождении тел Солнечной системы, будут запущены в 2021 году.

```
B [13]: def model_mT5_multilingual_XLSum_summary(article_text, model):
    """
        Самморизация.
        Используем предобученную модель: csebuetnlp/mT5_multilingual_XLSum
    """
    # model_mT5_multilingual_XLSum = AutoModelForSeq2SeqLM.from_pretrained(model_name)

    input_ids = tokenizer(
        [WHITESPACE_HANDLER(article_text)],
        return_tensors="pt",
        padding="max_length",
        truncation=True,
        max_length=512
    )["input_ids"]

    # print(input_ids)

    output_ids = model_mT5_multilingual_XLSum.generate(
        input_ids=input_ids,
        max_length=84,
        no_repeat_ngram_size=2,
        num_beams=4
    )[0]

    summary = tokenizer.decode(
        output_ids,
        skip_special_tokens=True,
        clean_up_tokenization_spaces=False
    )

    return summary
```

```
B [ ]: # test_records[0]["text"]
```

```
B [14]: article_text = test_records[0]["text"]
summary = model_mT5_multilingual_XLSum_summary(article_text, model_mT5_multilingual_XLSum)

print()
print("Текст: {}".format(test_records[0]["text"]))
print("Итоговый реферат: {}".format(summary))
print("Правильный реферат: {}".format(test_records[0]["summary"]))
```

Текст: Американское аэрокосмическое агентство NASA огласило названия четырех космических миссий, которые в скором времени могут быть выбраны для реализации и запуск которых может состояться уже в конце этого десятилетия. Эти четыре проекта стали полуфиналистами конкурса, объявленного среди американских научных команд, в котором участвовало более десяти миссий. Все они были отобраны по критериям потенциальной пользы для науки и технической осуществимости проекта. В рамках программы Discover у NASA занимается планированием миссий, которые призваны дать ответы на фундаментальные вопросы о происхождении тел Солнечной системы и возможному наличию жизни на них. «Эти выбранные миссии могут трансформировать наше восприятие некоторых из наиболее активных и сложных миров в Солнечной системе, – заявил Томас Зурбучен, помощник директора NASA по науке. – Исследование каждого из этих небесных тел поможет раскрыть секреты о том, как они и им подобные объекты образовались в космосе». Каждый проект из выбранных полуфиналистов получит \$3 млн на проработку концепции. Из четырех выбранных проектов NASA планирует отобрать две в 2021 году, чтобы профинансировать работу по их реализации в рамках программы Discovery. Два из четырех проектов посвящены исследованию Венеры, куда аппараты NASA не отправлялись уже свыше тридцати лет – с 1989 года, когда к этой планете был запущен зонд Magellan. Миссия VERITAS (Venus Emissivity, Radio Science, InSAR, Topography, and Spectroscopy) призвана картировать поверхность Венеры для лучшего понимания геологической истории планеты и получения ответа на вопрос о причинах ее кардинальных отличий от Земли. Карта будет построена при помощи радара, установленного на орбитальном аппарате. В ходе миссии планируется узнать, продолжают ли в наши дни на Венере тектонические и вулканические процессы. Кроме того, планируется картировать поверхность планеты в инфракрасном диапазоне. Миссия DAVINCI+ (Deep Atmosphere Venus Investigation of Noble gases, Chemistry and Imaging Plus) задумана для того, чтобы исследовать состав атмосферы Венеры, понять, как она формировалась и эволюционировала в прошлом, и ответить на вопрос о существовании когда-либо на планете жидкого океана. Сделать это планируется при помощи спускаемого аппарата, который будет проводить измерения от верхних слоев атмосферы до самой поверхности. Научные инструменты аппарата будут размещены внутри герметичного сферического корпуса, который должен защитить их от разрушительного воздействия высокой температуры атмосферы. + в названии миссии говорит о наличии в ее составе камер, которые будут присутствовать на борту орбитального и спускаемого аппаратов. «Проект DAVINCI+ пересекается по целям с российской миссией «Венера-Д», обе нацелены на посадку, поэтому в будущем возможно их совмещение в одном совместном проекте», – сказал «Газете.Ру» основатель проекта «Открытый космос» Виталий Егоров. Миссия TRIDENT должна исследовать Тритон – ледяной спутник Нептуна, известный ученым своей активностью. Данные, полученные американским зондом Voyager 2, показали, что поверхность Тритона активно обновляется. Ученые считают, что Тритон, на котором выпадает «снег» из органических веществ, может оказаться отличной целью для понимания вопросов происхождения аналогичных тел в Солнечной системе. Пролетев мимо Тритона, аппарат попытается картировать его поверхность, изучить активные процессы на его поверхности и определить, есть ли под его поверхностью жидкий океан. Миссия IVO (Io Volcano Observer) должна направиться к одному из самых интересных спутников Юпитера – Ио. Это тело с наиболее мощной вулканической активностью в Солнечной системе, которая поддерживается за счет мощного приливного воздействия со стороны Юпитера. Во время нескольких близких пролетов аппарат выяснит, как формируется лава на поверхности спутника.

Итоговый реферат: Космические миссии, которые могут дать ответы на вопросы о происхождении тел Солнечной системы, будут запущены в 2021 году.

Правильный реферат: В NASA назвали четыре миссии в дальний космос, которые в этом десятилетии могут быть запущены американцами. Среди них – две миссии по изучению Венеры, полет к спутнику Юпитера и экспедиция к Тритону, спутнику Нептуна.

```
B [16]: article_text = test_records[1]["text"]
summary = model_mT5_multilingual_XLSum_summary(article_text, model_mT5_multilingual_XLSum)

print()
print("Текст: {}".format(test_records[1]["text"]))
print("Итоговый реферат: {}".format(summary))
print("Правильный реферат: {}".format(test_records[1]["summary"]))
```

Текст: Около 11 тысяч зрителей увидели все самое лучшее, что есть на сегодняшний день в культуре Бурятии. В Кремле выступил Бурятский государственный академический театр оперы и балета, Национальный цирк, Бурятский национальный театр песни и танца «Байкал», ставший победителем шоу «Танцуют все!» на телеканале «Россия», а также другие профессиональные и самодеятельные коллективы региона. Более 300 артистов из одного региона на главной сцене страны - похоже это рекорд России. Зрителям рассказали, что Республике Бурятия, чье население составляет 1 миллион человек, сохранилась и развивается культура десятков национальностей, включая русских, бурятов, староверов (семейских), эвенков. И все они были представлены в Москве. Как писали после шоу зрители в соцсетях: «А мы думали, что в Бурятии только буряты живут...». Для неподготовленного зрителя это вообще были вечера открытий. Например, когда еще в Кремлевском дворце выстраивалась очередь из желающих попасть на прием к врачам-пульмонологам и ламам-астрологам? А между тем буквально в паре метров можно было увидеть выставку Национального музея Бурятии о культуре и быте бурятского народа. И она не менее уникальна, чем китайская, или монгольская времен Чингисхана. И там же современное искусство, например, знаменитые работы скульптора Даши Намдакова. Не обошлось без политической составляющей. Перед началом праздничного шоу с главной сцены страны первый заместитель руководителя администрации президента России Сергей Кириенко зачитал поздравление президента Российской Федерации Владимира Путина, в котором говорилось: «Сердечно поздравляю вас с наступлением Нового года по лунному календарю и с наступлением Белого месяца! Этот светлый, особо почитаемый последователями буддизма праздник символизирует обновление природы, стремление человека к гармонии и чистоте помыслов. Он обращает верующих к духовным и нравственным истокам этой древней религии, ее непреходящим устоям и ценностям. Важно, что российские буддисты бережно хранят и передают из поколения в поколение традиции предков...». Заместитель председателя правительства России, полномочный представитель президента РФ в ДФО Юрий Трутнев, видимо, также был удивлен ажиотажем и аншлагом в Кремле и рассказал журналистам: «Я спросил у [главы Бурятии] Алексея Цыденова, сколько людей приехало из Бурятии на праздник, и он ответил: «Тысяча из шести». Это значит, что большая часть людей на празднике – из Москвы, значит, им это интересно!». Сам же Цыденов в своем выступлении посетовал на то, что в России все же мало что знают о Бурятии и ее традициях, но был оптимистичен: «Республика Бурятия богата своей историей, своими традициями. Наша цель – показать все возможности и огромный потенциал республики, чтобы не только жители Бурятии, но и вся Россия гордилась достижениями нашего народа. Мы хотим, чтобы жители всей России знали, что есть такая жемчужина – Бурятия». Небольшим, отдаленным и дотационным регионам всегда непросто быть замеченными властями. А их губернаторам и главам сложно быть ближе к высшему руководству страны по сравнению с их коллегами из более богатых субъектов. Но, похоже, главе Бурятии Алексею Цыденову удастся переломить эту не самую приятную для многих регионов традицию. Цыденов пришел в Бурятию с поста замминистра транспорта России без опыта управления регионом. Но за три года сумел стать опытным региональным политиком, способным на неожиданные ходы. Как пример, нынешние посвященные Восточному новому году концерты в Москве. Пока другие регионы со своими местными праздниками и юбилеями стоят в очереди, чтобы «затащить» к себе руководителей страны на торжества, Цыденов, наоборот, сумел привезти свой регион в Москву. Причем не куда-нибудь, а в самый Кремль. Таким образом, всего за два столичных дня Бурятия решила сразу несколько задач, в том числе подтвердила свой статус буддийского лидера России и познакомила жителей всей страны со своей богатой культурой. Зрители после концерта отвечали Бурятии взаимностью. Можно обратить внимание лишь на некоторые высказанные в соцсетях мнения об увиденном в кремлевском дворце: «Люди в Бурятии очень талантливые», «Вот это красота! Срочно летим в Буратию», «Буряты, вы крутые!», «Бурятия, мы вам завидуем!».

Итоговый реферат: В воскресенье в Москве состоялось праздничное шоу, на котором можно было увидеть все самое лучшее, что есть на сегодняшний день в культуре Бурятии.

Правильный реферат: 25 и 26 февраля в Кремлевском дворце съездов праздновали Сагаалган – Восточный Новый год. Бурятия - центр российского буддизма и один из немногих регионов страны, где новый год встречают официально дважды.


```
B [17]: article_text = test_records[2]["text"]
summary = model_mT5_multilingual_XLSum_summary(article_text, model_mT5_multilingual_XLSum)

print()
print("Текст: {}".format(test_records[2]["text"]))
print("Итоговый реферат: {}".format(summary))
print("Правильный реферат: {}".format(test_records[2]["summary"]))
```

Текст: 7 ноября в Белоруссии прошли выборы членов совета республики – верхней палаты парламента. В столице и шести областях республики были избраны по восемь сенаторов. Еще восемь членов совета назначит президент Белоруссии Александр Лукашенко. Голосование осуществлялось на заседаниях депутатов местных советов базового уровня и Мингорсовета. Выборы в верхнюю палату парламента прошли безальтернативно – президиумы районных и городских советов депутатов, а также исполнительных комитетов выдвинули восемь претендентов на восемь мест. Для победы каждому кандидату необходимо было набрать простое большинство голосов. Из предыдущего состава верхней палаты парламента планировали переизбраться всего 11 человек, то есть новый состав совета республики будет значительно обновлен. Все выдвинутые кандидатуры получили поддержку депутатов. Большинство членов совета республики беспартийные – это одна из главных особенностей белорусского парламента. Среди утвержденных сенаторов руководителей промышленных и сельскохозяйственных предприятий, общественных организаций, научные сотрудники университетов, руководители школ и медицинских заведений. Результаты голосований будут утверждены 12 ноября. А через пять дней Белоруссию ждут выборы по мажоритарной системе в палату представителей – нижнюю палату парламента. Полномочия нынешнего состава парламента истекают в следующем году – в один год с президентскими. Однако проводить в 2020-м выборы президента и парламента глава страны Александр Лукашенко не захотел. Таким образом, срок полномочий парламентариев был немного сокращен, а новый состав начнет свою работу до конца 2019 года. Причем выборы в совет республики впервые проходят 7 ноября. Александр Лукашенко предложил приурочить их ко Дню Октябрьской революции. А вот выборы в палату представителей пройдут в воскресенье, к ним приковано гораздо больше общественного внимания. Всего были зарегистрированы 523 кандидата. С учетом того, что в палату представителей проходят 110 депутатов, средний конкурс составляет пять человек на место. Глава Центральной избирательной комиссии Лидия Ермошина 6 ноября заявила, что свои предвыборные программы опубликовали только 330 кандидатов. Двухпалатный парламент в Белоруссии несет, скорее, символическую роль. Начиная с 1996 года палата представителей не отклонила ни одного законопроекта, представленного президентом или правительством. Вся реальная власть сосредоточена в руках главы государства, который часто любит рассуждать об этом, особенно когда речь заходит о выборах. В основном, кстати, белорусский лидер сам поднимает вопрос о несменяемости власти в стране. «Мы уже четко определились, что в текущем году проводим парламентские выборы, а в следующем году – президентские, несмотря на прогнозы некоторых «знатоков» о том, что Лукашенко в нынешнем году, поскольку тихо и спокойно, быстро пройдет президентские выборы, чтобы опять схватить кресло и сесть в него надолго. Это полное заблуждение. Как я и обещал, все будет по закону и в рамках конституции», – говорил Лукашенко в ходе своей встречи с председателем ЦИК Лидией Ермошиной. В отличие от сегодняшних голосований, на выборах 17 ноября будут присутствовать наблюдатели. Аккредитованы почти 17,4 тыс. национальных наблюдателей. Причем, как ожидает председатель ЦИК Лидия Ермошина, их количество увеличится более чем в два раза. Пока на выборы в палату представителей получил аккредитацию 551 международный наблюдатель от СНГ, ОБСЕ, ПАСЕ. Для Лукашенко присутствие зарубежных наблюдателей – тема важная. К выборам в нижнюю палату президент относится как экзамену. Для Лукашенко выборы – возможность изменить свою репутацию «последнего диктатора Европы» в глазах своих западных партнеров. При этом президент все же подчеркивает, что Белоруссия проводит выборы без оглядки на Запад. «Я не думаю, что нам надо перед кем-то прогибаться и кого-то тащить в страну, приглашать на одни и вторые выборы. Но надо четко дать сигнал: кто хочет – приезжайте, – говорил он. – Все должно быть по закону. Как определили по конституции и закону, так мы и проведем выборы. Они для нашего народа, они не для кого-то. Хотя хотелось бы сохранить лицо и внешне выглядеть нормально». В ходе прошлых парламентских выборов Лукашенко проверку западных партнеров прошел. Впервые за 20 лет в палату представителей прошли сразу два оппозиционных кандидата – член Объединенной гражданской партии (ОГП) Анна Канопацкая и лингвист Елена Анисим, критиковавшая правительство из-за игнорирования вопросов национальной культуры и языка. Прохождение двух оппозиционеров в палату Евросоюз оценил как положительный признак и в первый раз за 12 лет признал парламентские выборы в Белоруссии. Затем последовало награждение от Брюсселя – отношения с Минском сдвинулись с мертвой точки. Более того, ЕС выразил готовность к сотрудничеству, в частности, к запуску переговоров по таможенным вопросам, а также перспективам возобновления энергетического диалога. «В то же время мы полностью осознаем, что не было ни конкретных, ни структурных изменений в ситуации с демократией и правами человека в Белоруссии, – говорила в 2016 году еврокомиссар Корина Крещу. – Тем не менее, мы приветствуем попадание в парламент первых непростых депутатов впервые за 12 лет. Надеемся, что они внесут с собой воздух и новые идеи в орган, который сам за последнюю каденцию инициировал меньше пяти законопроектов». Вскоре Европейский союз снял большинство санкций с высокопоставленных белорусских чиновников, включая самого Лукашенко. Всем им, кроме разрешения вести дела с зарубежными структурами, вернули также право на въезд в ЕС. В 2016-м Лидия Ермошина назвала выборы в парламент свободными и «самыми демократичными с 2000 года». Чего ожидать 17 ноября – пока неизвестно. Однако на этот раз ни Канопацкую, ни Анисим не зарегистрировали кандидатами для участия в грядущих парламентских выборах. У ЦИК возникли претензии к сбору ими подписей. В регистрации отказано выдвинувшему от ОГП Геннадию Трубачу, оппозиционной журналистке Ирине Халип, члену общественной организации «Говори правду» Илье Борисову, Александру Абрамовичу, Наталье Кушнеровой и ряду других независимых кандидатов. Кроме того, была отменена регистрация кандидата от гражданской кампании «Европейская Беларусь» Полины Шарендо-Панасюк. В ходе своего телевизионного выступления она призвала Лукашенко покинуть пост президента. В «Европейской Беларуси» считают, что именно это заявление послужило причиной для снятия кандидата с гонки. Однако, как отмечает в разговоре с «Газетой.Ru» доцент кафедры политической истории МГИМО Кирилл Коктыш, у оппозиции и так низкие шансы пройти в парламент – и причина состоит в ее низкой популярности. «При действующих раскладах симпатия и так на стороне власти, там ей не требуются манипуляции, чтобы сохранить свои позиции. На самом деле, в любом случае, если говорить про белорусскую оппозицию, которая есть, там нет содержательной программы, внятной альтернативы, и в этом плане уровень поддержки всегда достаточно невысокий. Он никогда не выходил за 10%», – отмечает эксперт. Вместе с тем выборы 17 ноября все же могут принести определенные важные изменения в белорусский парламент – свое представительство могут увеличить политические партии, никогда не игравшие роль в политике постсоветской Белоруссии. Из зарегистрированных 560 кандидатов 367 (65,5%) являются членами партий. 19 апреля в своем ежегодном послании к народу Лукашенко, анонсируя изменения в конституции, заявил, что белорусский парламент должен «подтянуть политические партии, чтобы они генерировали нормальных руководителей». Изменения, к слову, коснутся перераспределения полномочий между ветвями власти. Об этом сказал сам президент в своем послании, отметив, что таким образом он готовит почву для своего будущего преемника. «Мне задают часто вопрос: хотел бы ты оставить эту конституцию своему преемнику? Я четко и прямо говорю: нет. Мы уже прошли тот этап, когда нам нужна была сильная власть, какая она сегодня есть, – сказал Лукашенко в апреле в ходе своего послания к белорусскому народу. – Надо нагружать властными полномочиями другие структуры и ветви власти. Чтобы, к примеру, за экономику отвечало правительство вместе с парламентом».

Итоговый реферат: В воскресенье в Белоруссии пройдут выборы в нижнюю палату парламента, а через пять дней парламент начнет свою работу.

Правильный реферат: В Белоруссии в день годовщины Октябрьской революции выбрали членов верхней палаты парламента. Сегодняшнее голосование – пролог основной части электорального процесса: через десять дней в республике пройдут выборы в палату представителей, которые станут тестом для Александра Лукашенко, ведь за ними будут внимательно следить на Западе.

Вывод

Качество суммаризации достаточно высокое

```
B [ ]:
```

```
B [19]: article_text = test_records[0]["text"]
summary = model_mT5_multilingual_XLSum_summary(article_text, model_mbart_ru_sum_gazeta)

print()
print("Текст: {}".format(test_records[0]["text"]))
print("Итоговый реферат: {}".format(summary))
print("Правильный реферат: {}".format(test_records[0]["summary"]))
```

Текст: Американское аэрокосмическое агентство NASA огласило названия четырех космических миссий, которые в скором времени могут быть выбраны для реализации и запуск которых может состояться уже в конце этого десятилетия. Эти четыре проекта стали полуфиналистами конкурса, объявленного среди американских научных команд, в котором участвовало более десяти миссий. Все они были отобраны по критериям потенциальной пользы для науки и технической осуществимости проекта. В рамках программы Discovery у NASA занимается планированием миссий, которые призваны дать ответы на фундаментальные вопросы о происхождении тел Солнечной системы и возможному наличию жизни на них. «Эти выбранные миссии могут трансформировать наше восприятие некоторых из наиболее активных и сложных миров в Солнечной системе, – заявил Томас Зурбучен, помощник директора NASA по науке. – Исследование каждого из этих небесных тел поможет раскрыть секреты о том, как они и им подобные объекты образовались в космосе». Каждый проект из выбранных полуфиналистов получит \$3 млн на проработку концепции. Из четырех выбранных проектов NASA планирует отобрать две в 2021 году, чтобы профинансировать работу по их реализации в рамках программы Discovery. Два из четырех проектов посвящены исследованию Венеры, куда аппараты NASA не отправлялись уже свыше тридцати лет – с 1989 года, когда к этой планете был запущен зонд Magellan. Миссия VERITAS (Venus Emissivity, Radio Science, InSAR, Topography, and Spectroscopy) призвана картировать поверхность Венеры для лучшего понимания геологической истории планеты и получения ответа на вопрос о причинах ее кардинальных отличий от Земли. Карты будут построены при помощи радара, установленного на орбитальном аппарате. В ходе миссии планируется узнать, продолжают ли в наши дни на Венере тектонические и вулканические процессы. Кроме того, планируется картировать поверхность планеты в инфракрасном диапазоне. Миссия DAVINCI+ (Deep Atmosphere Venus Investigation of Noble gases, Chemistry and Imaging Plus) задумана для того, чтобы исследовать состав атмосферы Венеры, понять, как она формировалась и эволюционировала в прошлом, и ответить на вопрос о существовании когда-либо на планете жидкого океана. Сделать это планируется при помощи спускаемого аппарата, который будет проводить измерения от верхних слоев атмосферы до самой поверхности. Научные инструменты аппарата будут размещены внутри герметичного сферического корпуса, который должен защитить их от разрушительного воздействия высокой температуры атмосферы. + в названии миссии говорит о наличии в ее составе камер, которые будут присутствовать на борту орбитального и спускаемого аппаратов. «Проект DAVINCI+ пересекается по целям с российской миссией «Венера-Д», обе нацелены на посадку, поэтому в будущем возможно их совмещение в одном совместном проекте», – сказал «Газете.Ru» основатель проекта «Открытый космос» Виталий Егоров. Миссия TRIDENT должна исследовать Тритон – ледяной спутник Нептуна, известный ученым своей активностью. Данные, полученные американским зондом Voyager 2, показали, что поверхность Тритона активно обновляется. Ученые считают, что Тритон, на котором выпадает «снег» из органических веществ, может оказаться отличной целью для понимания вопросов происхождения аналогичных тел в Солнечной системе. Пролетев мимо Тритона, аппарат попытается картировать его поверхность, изучить активные процессы на его поверхности и определить, есть ли под его поверхностью жидкий океан. Миссия IVO (Io Volcano Observer) должна направиться к одному из самых интересных спутников Юпитера – Ио. Это тело с наиболее мощной вулканической активностью в Солнечной системе, которая поддерживается за счет мощного приливного воздействия со стороны Юпитера. Во время нескольких близких пролетов аппарат выяснит, как формируется лава на поверхности спутника.

Итоговый реферат: Космические миссии, которые могут дать ответы на вопросы о происхождении тел Солнечной системы, будут запущены в 2021 году.

Правильный реферат: В NASA назвали четыре миссии в дальний космос, которые в этом десятилетии могут быть запущены американцами. Среди них – две миссии по изучению Венеры, полет к спутнику Юпитера и экспедиция к Тритону, спутнику Нептуна.

```
B [ ]:
```

