Національний технічний університет України

«Київський політехнічний інститут

імені Ігоря Сікорського»

**Щ О Д Е Н Н И К ПРАКТИКИ**

                                Переддипломна практика

(вид і назва практики)

Студента Лисого Павла Олексійовича

Факультет прикладної математики

освітньо-кваліфікаційний рівень              бакалавр

напрям підготовки     113 «Прикладна математика»

спеціальність:                     «Прикладна математика»

курс       4      Група       КМ-71

Студент Лисий Павло Олексійович

(прізвище, ім’я, по батькові)

Прибув ФПМ, КПІ ім.Ігоря Сикорського

Печатка

підприємства, організації, установи «  12  »            квітня               2021 року

(підпис) (посада, прізвище та ініціали відповідальної особи)

Вибув із підприємства, організації, установи

Печатка

Підприємства, організації, установи «  16 »            травня               2021 року

(підпис) (посада, прізвище та ініціали відповідальної особи)

**Індивідуальне завдання з практики**

Тема     «Математичне та програмне забезпечення для

обробки й класифікації архіву новинних відеорепортажів»

Зміст

Протягом даної роботи розглядається процес обробки новинних сюжетів з використанням методів computer-vision та natural language processing. Під обробкою новинних сюжетів мається на увазі: алгоритмічне визначення start-end time кожного окремо сюжету, визначення дати сюжету, хронометражу сюжету, виділення назви сюжету, виділення тексту сюжету, виділення типу (новини, анонс та реклама), визначення журналістів та гостей, визначення ведучих, визначення тегів та класифікація сюжету(позитивний/негативний).

Керівник практики від університету \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Керівник практики від підприємства \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Календарний графік проходження практики**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| № з/п | Назви робіт | Тижні проходження практики | | | | | Відмітки про виконання |
| 1 | 2 | 3 | 4 | 5 |  |
|  | Огляд та порівняння існуючих методів | + |  |  |  |  |  |
|  | Проектування алгоритму |  | + | + | + |  |  |
|  | Тестування алгоритму |  |  |  | + |  |  |
|  | Підготовка розділу:  ПОСТАНОВКА ЗАДАЧІ |  |  |  |  | + |  |
|  | Підготовка розділу:  АНАЛІЗ ІСНУЮЧИХ МЕТОДІВ РОЗВʼЯЗАННЯ ЗАДАЧІ |  |  |  |  | + |  |
|  | Підготовка розділу:  МАТЕМАТИЧНЕ ЗАБЕЗПЕЧЕННЯ |  |  |  |  | + |  |
|  | Підготовка розділу:  ОПИС ПРОГРАМНИХ ЗАСОБІВ |  |  |  |  | + |  |
|  | Підготовка розділу: ЕКСПЕРЕМЕНТАЛЬНІ РЕЗУЛЬТАТИ |  |  |  |  | + |  |
|  | Підготовка розділу:  ВИСНОВКИ |  |  |  |  | + |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

Керівники практики:

від вищого навчального закладу                 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

(підпис) (прізвище та ініціали)

від підприємства, організації, установи \_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_

(підпис) (прізвище та ініціали)

**Р о б о ч і з а п и с и**

\_\_\_\_1\_\_\_ тиждень практики

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Дати 16.04.2021 Записи про виконання завдання

Протягом даного тижня було визначено мету та ціль роботи. Проведено ознайомлення з існуючими методами cv та nlp. Було вивчено декілька ключових технік, які дозволяють досягти поставленої цілі. Відбулося проектування архітектури алгоритму та взаємодії різних частин коду.

Також було визначено основні етапи роботи:

Перший етап передбачає розділення відеосюжету на окремі сюжети, визначення start-end time сюжету, визначення довжини кожного сюжету.

Другий підхід передбачає використання cv алгоритму, для парсингу відеопотоку. В результаті визначається назва сюжету та текст сюжету.

Третій підхід використовує nlp алгоритми для оброблення виділеного тексту для визначення журналістів та гостей, визначення ведучих та визначення тегів. Також завдяки методам nlp класифікуємо сюжет(позитивний/негативний).

**Р о б о ч і з а п и с и**

2 тиждень практики

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Дати 23.04.2021 Записи про виконання завдання

Протягом тижня була вивчена документація наступних бібліотек (moviepy, av, pytesseract, PIL). Розібрані попердньонавчені моделі для досягнення поставлених задач. Відбулося проектування 1 частини алгоритму – cv. Відбулося проектування 1 частини алгоритму – cv. Протягом тижня була розроблена cv частина задачі, а саме: парсинг тексту з використанням Tesseract ORC. Детальніший звіт представлений у додатку [диплом.docx, c.15-16].

**Р о б о ч і з а п и с и**

3 тиждень практики

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Дати 30.04.2021 Записи про виконання завдання

Протягом тижня була вивчена документація наступних бібліотек (spacy, nltk, google\_trans\_new, speech\_recognition). Розібрані попердньонавчені моделі для досягнення поставлених задач. Відбулося проектування 2 частини алгоритму – nlp. Протягом тижня була розроблена nlp частина задачі, а саме: розпізнавання мови завдяки бібліотеці speech\_recognition, переклад отриманого тексту на англійський за допомогою бібліотеки google\_trans\_new, визначення тегів завдяки бібліотеці spacy, nltk та метрики TF-IDF.

Детальніший звіт представлений у додатку [диплом.docx, c.16-19].

**Р о б о ч і з а п и с и**

4 тиждень практики

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Дати 7.05.2021 Записи про виконання завдання

Протягом тижня було завершено розробку алгоритму та написання скрипта на мові програмування python. Було проведено об’єднання 2 частин алгоритму (cv + nlp). Був створений модуль для створення таблиці з готовим результатом, а також запис готового результату до документу Excel. Приклад вихідного файлу Excel зображений на рисунку 1.

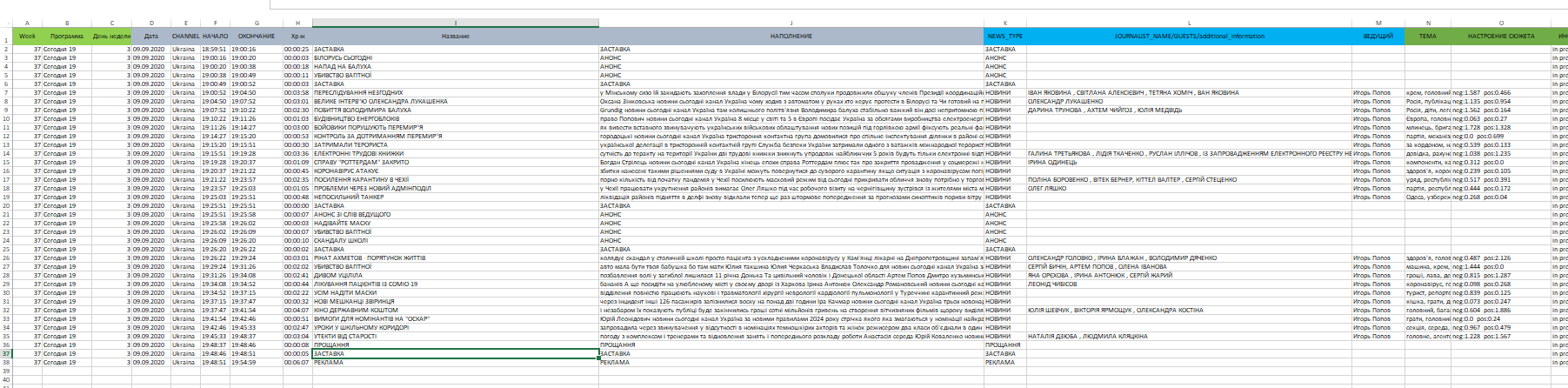


Рис.1 – вихідний файл Excel.

Також було протестовано алгоритм на роботі 84годин відеосюжетів. Точність розпізнавання start-end time частини складає 96,7%. Точність по іншим параметрам близька до 100%. Повний код наведений у додатку [main.ipynb].

**Р о б о ч і з а п и с и**

5 тиждень практики

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Дати 14.05.2021 Записи про виконання завдання

Протягом тижня були розроблені та оформлені основні розділи дипломної роботи, а саме: постановка задачі, аналіз існуючих методів розв’язання задачі, опис загальної схеми обробки медіасюжетів, опис програмних засобів, опис експериментальних результатів та висновки.

Детальніший звіт представлений у додатку [диплом.docx, c.16-19].

**Відгук та оцінка роботи студента на практиці**

ФПМ

(назва підприємства, організації, установи)

|  |
| --- |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |

Керівник практики від підприємства, організації, установи \_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(підпис) (прізвище та ініціали)

Печатка

«      »                                         2021 року

**Відгук осіб, які перевіряли проходження практики**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Висновок керівника практики від вищого навчального закладу про проходження практики**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Дата складання заліку «       »                                  2021 року

Оцінка:

за національною шкалою

(словами)

кількість балів

(цифрами і словами)

Керівник практики від вищого навчального закладу

\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(підпис) (прізвище та ініціали)