

ДОДАТОК А

Звіт з результатами перевірки на унікальність тексту

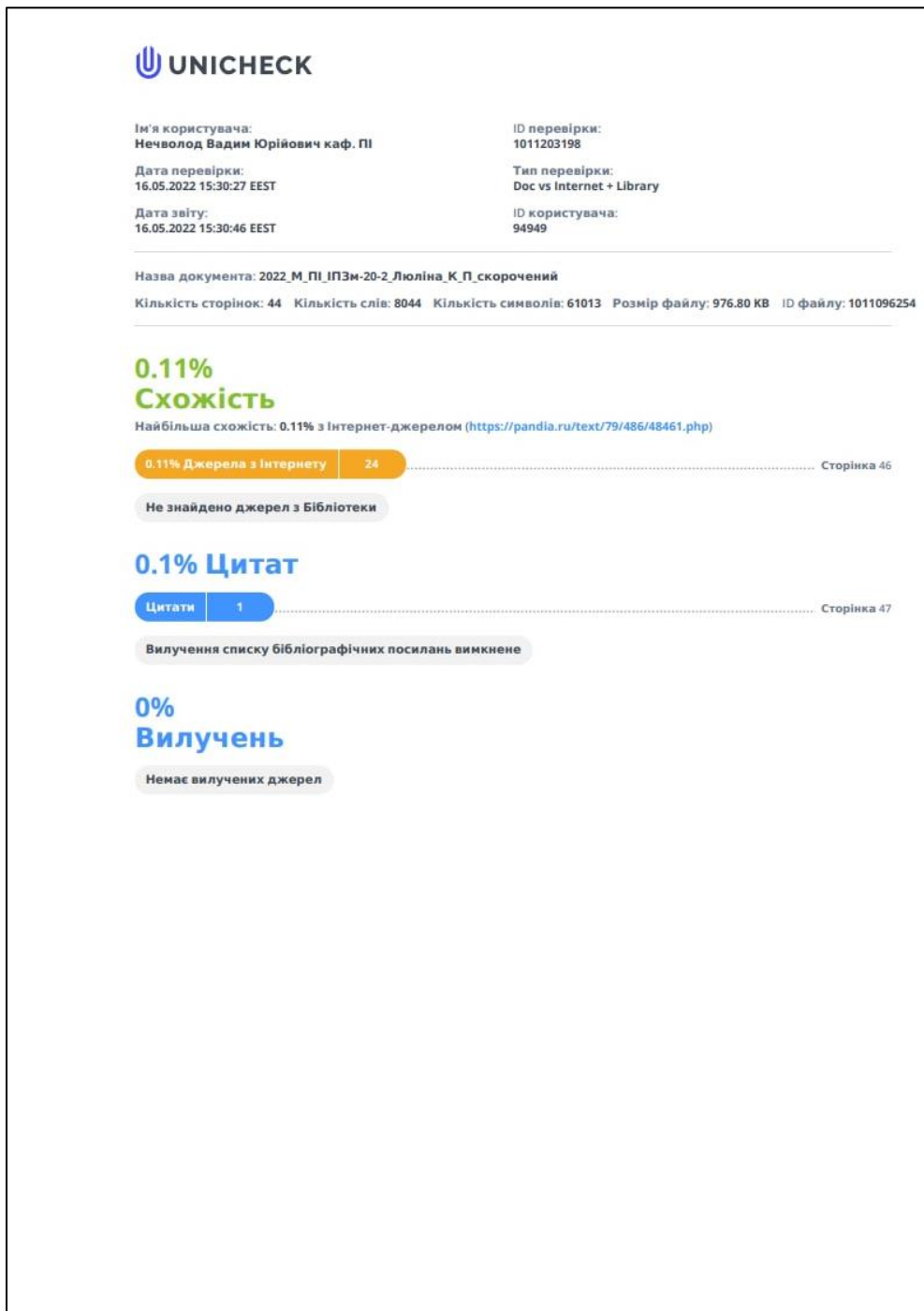


Рисунок А.1 – Результат перевірки тексту на унікальність

ДОДАТОК Б
Слайди презентації

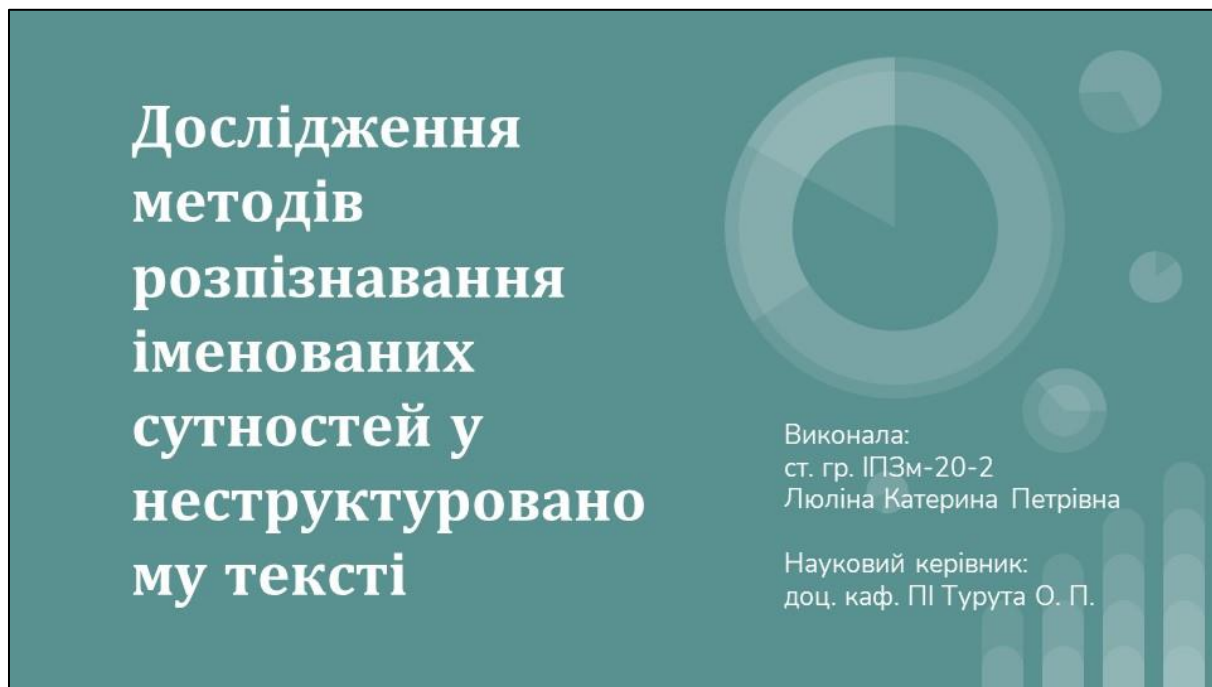


Рисунок Б.1 – Титульний слайд

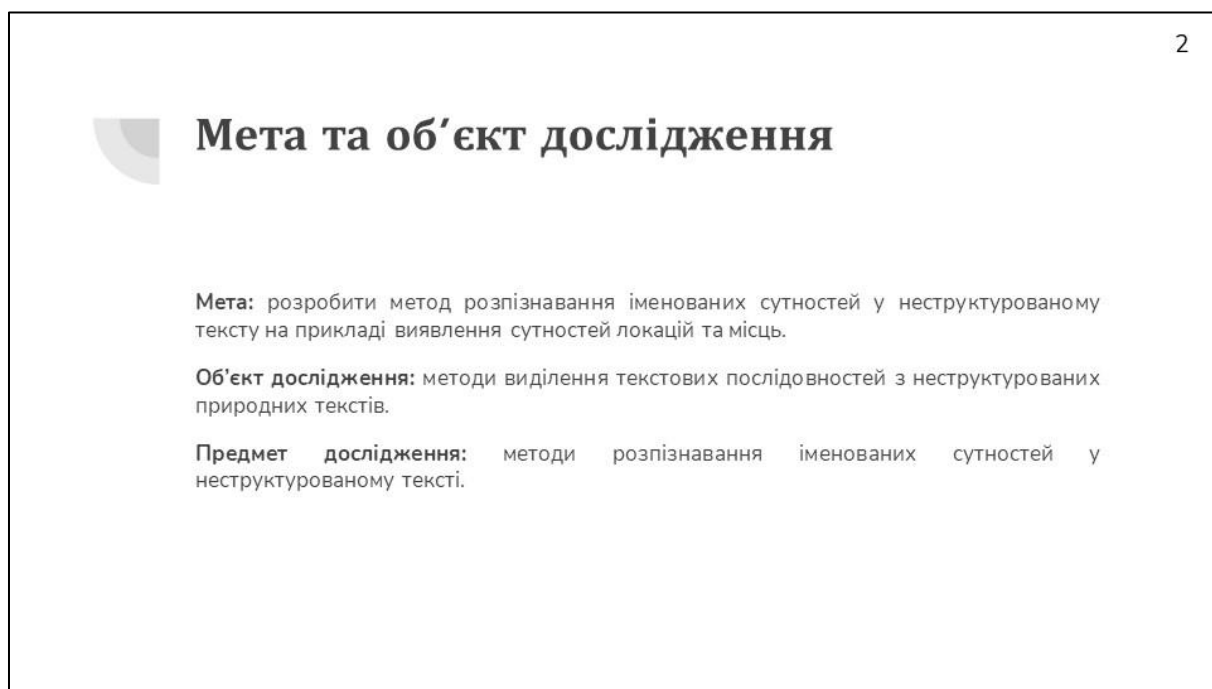


Рисунок Б.2 – Слайд «Мета та об'єкт дослідження»

3

Актуальність

1. Зростання обсягів інформації у соціальних мережах.
2. Зростання популярності української мови.
3. Необхідність аналізу користувацьких повідомлень у системах реагування на екстрені ситуації.

Рисунок Б.3 – Слайд «Актуальність»

4

Задачі дослідження

1. Вивчення стану наукових надбань у питаннях обробки природних мов.
2. Дослідження існуючих методів та наукових робіт щодо питання NER для локацій в українській та іноземних мовах.
3. Розгляд та опис наявних технологій, розробка програмного забезпечення та опис його архітектури.
4. Визначення метрик для вимірювання ефективності розробленого методу для вирішення питання розпізнавання іменованих сутностей локацій та місць.
5. Визначення метрик для вимірювання ефективності розробленого методу для вирішення питання розпізнавання іменованих сутностей локацій та місць

Рисунок Б.4 – Слайд «Задачі дослідження»

Аналіз наукової та патентної літератури

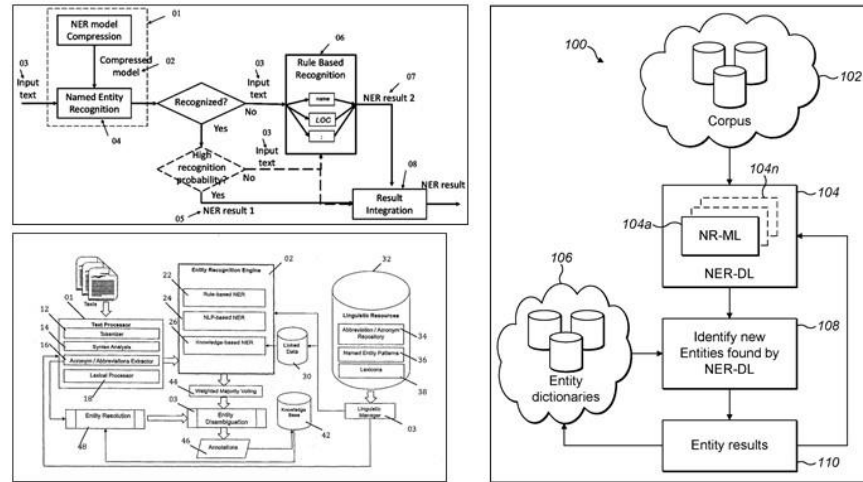


Рисунок Б.5 – Слайд «Аналіз наукової та патентної літератури»

Запропонований механізм

RE UD types:

1. Nominal: 'nmod', 'nummod', 'amod', 'det', 'flat', 'compound', 'reparandum'
2. Complex: 'appos', 'conj', 'acl', 'case', 'conj'.

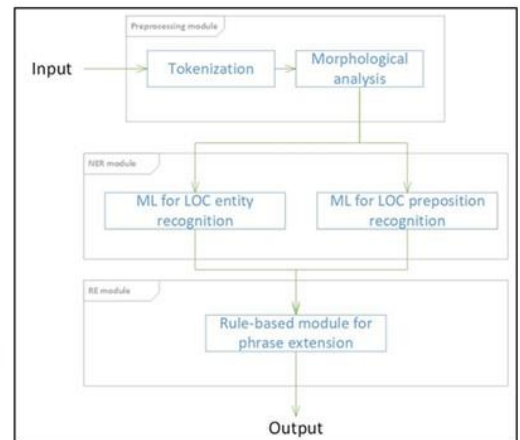


Рисунок Б.6 – Слайд «Запропонований механізм»



Підготовка даних для проведення експериментів

Мова: українська

Джерела:

1. Телеграм-канали та чати (160 записів).
2. Українська художня література (40 записів).

Підготовлений набір тестових даних наведено за [посиланням](#).

Він живе на третьому поверсі.
Він живе у другому під'їзді.
У вівторок на Плеканівській, 73 відновлює роботу відділ державної реєстрації актів цивільного стану.
У різних населених пунктах Харківської області набори відрізняються, іноді якимись нюансами, інколи ж зовсім інший асортимент.
В албанській Тирані автобус після ДТП застряг над річкою.
О 18:30 вони обіцяють розпочати на площі Франції безстроковий страйк.
У результаті пожежі на об'єкті у Білгородській області різні пошкодження отримали сім будинків.
Губернатор Білгородської області повідомив, що на кордоні трьох муніципалітетів – Борисівського та Білгородського районів та Ян
У мережі з'явилися фото мосту.

Рисунок Б.7 – Слайд «Підготовка даних для проведення експериментів»



Метрики для оцінки ефективності рішення

TP – кількість істинних позитивних рішень.

FP – кількість помилково позитивних рішень

FN – кількість помилково негативних рішень.

$Precision$ - відсоток вірно виявлених сутностей серед усіх розпізнаних сутностей.

$Recall$ - відсоток вірно розпізнаних іменованих сутностей серед тих, що мали бути розпізнані.

F - гармонійне середнє між $Precision$ та $Recall$.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F = 2 \frac{Precision * Recall}{Precision + Recall}$$

Рисунок Б.8 – Слайд «Метрики для оцінки ефективності рішення»

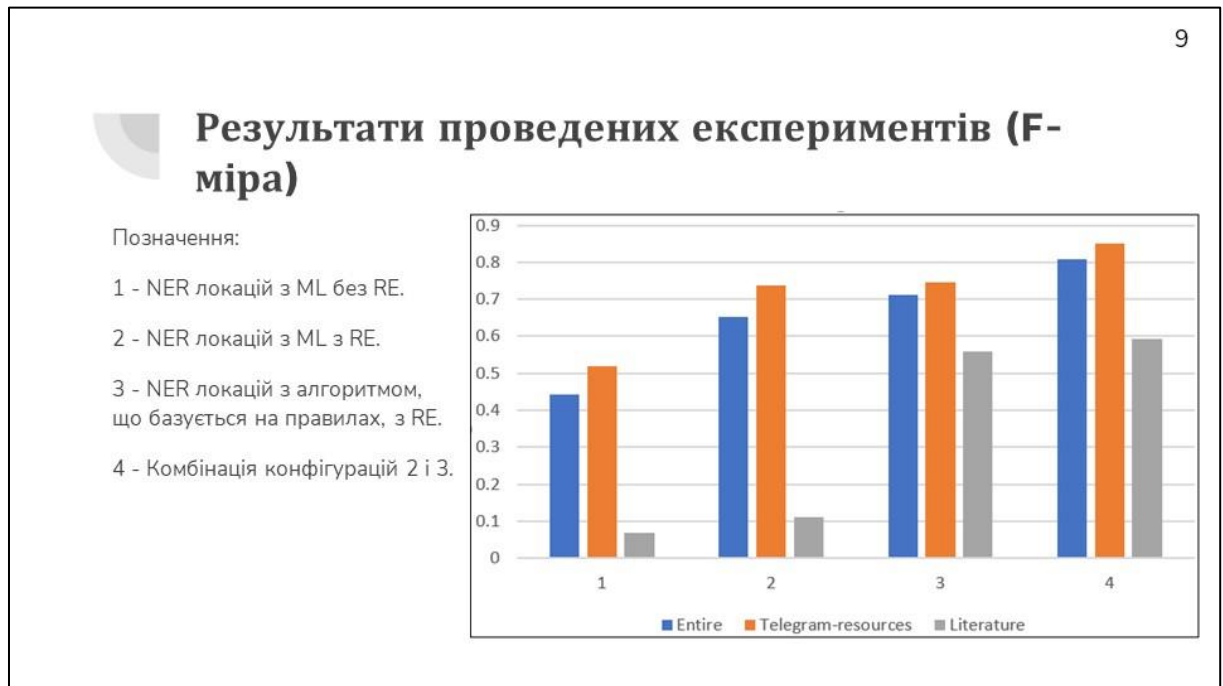


Рисунок Б.9 – Слайд «Результати проведених експериментів (F-міра)»

10

Технології

Мова програмування: Python.

Бібліотека для обробки природної мови: [Stanza](#).

Допоміжні бібліотеки:

- [NumPy](#) - для роботи з масивами даних;
- [Pandas](#) - для роботи з текстовими даними.

Середовище розробки: PyCharm Community Edition 2022.1.

Рисунок Б.10 – Слайд «Технології»



Впровадження

1. Опубліковано статтю «Дослідження методів розпізнавання іменованих сутностей у неструктурованому тексті».

Люліна К. П., Тірута О. П. Дослідження методів розпізнавання іменованих сутностей у неструктурованому тексті // Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління. Тези доповідей дванадцятій міжнародній науково-технічній конференції 27 – 28 квітня 2022 року. – Баку – Харків – Жиліна, 2022, т. 2. с. 143.

1. Потенціал практичного застосування рішення для систем, де потребується аналіз локацій: системи доставки послуг та товарів, системи реагування на екстренні ситуації.

Рисунок Б.11 – Слайд «Впровадження»



Демонстрація

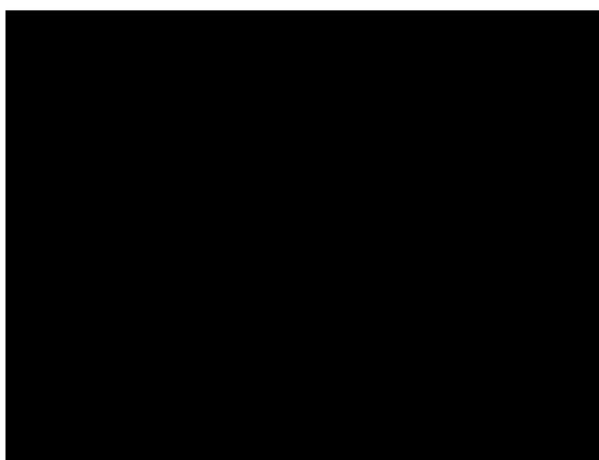


Рисунок Б.12 – Слайд «Демонстрація»

13

Висновки

- 1. Вивчено існуючі засоби та системи розпізнавання іменованих сутностей.
- 2. Розглянуто патенти та наукові роботи, присвячені розпізнаванню іменованих сутностей.
- 3. Досліджено ефективність використання алгоритмів машинного навчання, алгоритмів, що базуються на правилах, а також механізму RE для NER локацій у неструктурованому тексті українською мовою.

Рисунок Б.13 – Слайд «Висновки»

Дякую за увагу!

Рисунок Б.14 – Слайд «Дякую за увагу!»

ДОДАТОК В

Експертний висновок результатів перевірки кваліфікаційної роботи

1

1 из 1

Експертний висновок результатів перевірки кваліфікаційної роботи

студент
(посада)

програмної інженерії
(кафедра)

ІПЗМ-20-2
(група)

Люліна К.П.
(прізвище, ім'я, по батькові)

Зауваження

Пункт ДСТУ 3008-2015	Зміст пункту	Сторінка кваліфікаційної роботи
1	2	3
	7.1 Загальні положення	
	7.5 Рисунки	
	7.6 Таблиці	
7.6.9	Якщо рядки або колонки таблиці виходять за межі формату сторінки, таблицю поділяють на частини, розміщуючи одну частину під іншою або поруч, чи переносять частину таблиці на наступну сторінку. У кожній частині таблиці повторюють її головку та боковик. У разі поділу таблиці на частини дозволено її головку чи боковик замінити відповідно номерами колонок або рядків, нумеруючи їх арабськими цифрами в першій частині таблиці. Слово «Таблиця» подають лише один раз над першою частиною таблиці. Над іншими частинами таблиці з абзацного відступу друкують «Продовження таблиці» або «Кінець таблиці ____» без повторення її назви.	39, далі за текстом.
	7.7 Переліки	
	7.10 Формули та рівняння	
	7.15 Додатки	
	Бал за оформлення роботи (мах.10)	9,9

Експерт

 (підпис)
 20.5.2022

Вадим НЕЧВОЛОД
 (прізвище, ім'я, по батькові)

Рисунок В.1 – Експертний висновок результатів перевірки кваліфікаційної роботи

ДОДАТОК Г

Публікація у збірнику «Сучасні напрями розвитку інформаційно комунікаційних технологій та засобів управління»

Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління

ДОСЛІДЖЕННЯ МЕТОДІВ РОЗПІЗНАВАННЯ ІМЕНОВАНИХ СУТНОСТЕЙ У НЕСТРУКТУРОВАНОМУ ТЕКСТІ

Люліна К. П., Турута О. П.

Харківський національний університет радіоелектроніки, Харків, Україна

Виникнення технології розпізнавання іменованих сутностей пов'язане з проблематикою автоматичної обробки текстів та їх розуміння у подальшому за допомогою програмних систем. Дане питання було вперше розглянуто в рамках шостої Конференції з розуміння повідомлень (Sixth Message Understanding Conference – MUC-6) у 1995 році [1]. На цій конференції вперше було сформульовано визначення іменованої сутності, що полягало у наступному: іменована сутність - це вставлений у текст тег стандартної узагальненої мови розмітки задля маркування слів.

Методи NER можуть базуватися окремо або одночасно на основі лексики (lexicon-based), правил (rule-based) та машинного навчання (machine learning based).

Метою доповіді є порівняльний аналіз існуючих підходів до проблеми розпізнавання іменованих сутностей, виокремлення суттєвих переваг та недоліків, а також побудова на основі позитивного та негативного досвіду структури системи розпізнавання іменованих сутностей.

В доповіді наводиться аналітичний огляд досліджуваної проблеми.

Було розглянуто застосування кожного типу методів NER на приладі трьох патентів.

Слід зазначити, що найбільш ефективними є методи, базовані на правилах, так як мають змогу визначати іменовані сутності за завчасно визначеними патернами [2]. Їх ефективність обумовлена тим, що природні мови мають достатньо завчасно визначених правил та форматів, за якими можна розпізнавати текст.

Проте методів, базованих на правилах недостатньо для аналізу текстів природною мовою, що у свою чергу обумовлено наявністю контексту в реченнях, багатозначністю слів, а також наявністю зворотів та словосполучень, що мають відхилення від загальноприйнятих правил правопису. Саме тому на практиці алгоритми NER є комбінацією різних методів розпізнавання іменованих сутностей, що працюють у комплексі.

Список літератури

1. Overview of results of the MUC-6 evaluation : веб-сайт. URL: <https://aclanthology.org/X96-1048.pdf>.
2. Computational Analysis and Understanding of Natural Languages: Principles, Methods and Applications: веб-сайт. URL: <https://www.sciencedirect.com/topics/computer-science/named-entity-recognition-system#:~:text=7.6%20Named%20Entity%20Recognition&text=There%20are%20three%20major%20approaches,based%2C%20and%20machine%20learning%20based.>

Рисунок Г.1 – Публікація на тему «Дослідження методів розпізнавання іменованих сутностей у неструктурованому тексті»