

Міністерство освіти і науки України
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «КИЄВО-МОГИЛЯНСЬКА АКАДЕМІЯ»
Кафедра інформатики факультету інформатики

**Синтаксичний аналіз природного тексту на прикладі простих
геометричних задач**

**Текстова частина до курсової роботи
за спеціальністю „Інженерія програмного забезпечення” 121**

Керівник курсової роботи

Ст.викл. _____
(прізвище та ініціали)

(підпис)

“ _____ ” _____ 2019 р.

Виконав студент

(прізвище та ініціали)

“ _____ ” _____ 2019 р.

Анотація

Темою моєї курсової роботи було, розуміння природнього тексту. З моїм керівником ми вирішили розробляти систему, для розв'язування простих геометричних задач. Для цього я використав славнозвісну бібліотеку <https://www.pullenti.ru/>, і думаю результати виконання програми говорять самі за себе.

```
1 package snippet;
2
3 import java.io.IOException;
4
5
6
7 public class Kurs {
8     public static void main(String[] args) throws IOException, Exception {
9
10        int flag = 0;
11        com.pullenti.ner.Sdk.initialize(com.pullenti.morph.MorphLang.UA, com.pullenti.morph.MorphLang.UA);
12
13        String txt = "У прямокутника одна сторона дорівнює 5 сантиметрів, а друга сторона шість. Знайдіть площу";
14        System.out.println("Text: " + txt);
15
16        com.pullenti.ner.AnalysisResult are = com.pullenti.ner.ProcessorService.getEmptyProcessor().process(
17            new com.pullenti.ner.SourceOfAnalysis(txt), null, null);
18        Rect rect = new Rect();
19
20        for (com.pullenti.ner.Token t = are.firstToken; t != null; t = t.getNext()) {
21
22            NumberToken npt1 = com.pullenti.ner.core.NumberHelper.tryParseNumber(t);
23
24            if (npt1 == null) {
25                com.pullenti.ner.core.NounPhraseToken npt =
26                    com.pullenti.ner.core.NounPhraseHelper.tryParse(t, com.pullenti.ner.core.NounPhraseParseAttr.W);
27
28                if (npt == null) continue;
29
30            }
31        }
32    }
33 }
```

Problems | Javadoc | Declaration | Console

```
<terminated> Kurs [Java Application] C:\Program Files\Java\jdk-13.0.1\bin\javaw.exe (19 apr. 2020 r., 22:51:21)
Text: У прямокутника одна сторона дорівнює 5 сантиметрів, а друга сторона шість. Знайдіть площу
First side: 2.0
Second side: 6.0
Diagonal: 6.324555320336759
square: 12.0
Perimetr: 16.0
```

```
1 package snippet;
2
3 import java.io.IOException;
4
5
6
7 public class Kurs {
8     public static void main(String[] args) throws IOException, Exception {
9
10        int flag = 0;
11        com.pullenti.ner.Sdk.initialize(com.pullenti.morph.MorphLang.UA, com.pullenti.morph.MorphLang.UA);
12
13        String txt = "У прямокутника ширина дорівнює 6 метрів, а ширина 8 метрів. Знайдіть площу.";
14        System.out.println("Text: " + txt);
15
16        com.pullenti.ner.AnalysisResult are = com.pullenti.ner.ProcessorService.getEmptyProcessor().process(
17            new com.pullenti.ner.SourceOfAnalysis(txt), null, null);
18        Rect rect = new Rect();
19
20        for (com.pullenti.ner.Token t = are.firstToken; t != null; t = t.getNext()) {
21
22            NumberToken npt1 = com.pullenti.ner.core.NumberHelper.tryParseNumber(t);
23
24            if (npt1 == null) {
25                com.pullenti.ner.core.NounPhraseToken npt =
26                    com.pullenti.ner.core.NounPhraseHelper.tryParse(t, com.pullenti.ner.core.NounPhraseParseAttr.W);
27
28                if (npt == null) continue;
29
30            }
31        }
32    }
33 }
```

Problems | Javadoc | Declaration | Console

```
<terminated> Kurs [Java Application] C:\Program Files\Java\jdk-13.0.1\bin\javaw.exe (19 apr. 2020 r., 23:29:48)
Text: У прямокутника ширина дорівнює 6 метрів, а ширина 8 метрів. Знайдіть площу.
First side: 6.0
Second side: 8.0
Diagonal: 10.0
square: 48.0
Perimetr: 28.0
```

Як бачимо ми маємо вирішину задачу.

Вступ

Соціальні мережі є надзвичайно цікавим об'єктом дослідження, що відкриває можливості для аналізу соціальної взаємодії між людьми, прогнозування їх поведінки, класифікації та типології соціальної поведінки, моделювання інформаційних потоків у мережах, розвитку бізнес проектів. Соціальні мережі є ідеальним експериментальним майданчиком для вивчення соціальної поведінки людей, апробації технологій впливу та реклами, просування сучасних бізнес проектів.

Розвиток соціальних мереж поставив питання про аналіз даних, якими вони наповнені. Соціальні дані є невичерпним джерелом інформації про особисте життя та інтереси, уподобання та поведінку реальних людей. В соціальних мережах персональні дані стали публічно доступними. Доступними стали дані біографії, листи, щоденники, листування, повідомлення про подорожі тощо. Наявність таких даних відкриває надзвичайні можливості для вирішення дослідницьких завдань, з'ясування напрямів проектування бізнесу та створення програмних продуктів для користувачів соціальних мереж. Різні дослідницькі центри та бізнес компанії проявляють підвищений інтерес до збору та аналізу даних користувачів соціальних мереж.

Актуальність проблеми аналізу уподобань користувачів полягає в тому, що, вивчивши уподобання користувачів, на основі даних соціальних мереж можна здійснити моделювання соціальних, економічних та політичних процесів та створювати нові сервіси, бізнес рішення, програмні продукти.

Однак, на сьогоднішній день, не існує готових програмних рішень, що дозволять аналізувати висловлювання користувачів соціальних мереж. Немає чітких критеріїв та підходів до аналізу висловлювань в контексті соціальних мереж. За мету даної роботи було поставлені розробка та реалізація комплексного підходу до аналізу висловлювань користувачів соціальних мереж, а також реалізація програмного прототипу для надання ефективного управління процесом аналізу та отриманими результатами.

Робота складається з чотирьох розділів.

Перший розділ присвячено огляду теоретичних засад опрацювання даних в режимі реального часу. Розглянуто існуючі методи роботи моделі клієнт-сервер, а також оптимізації, опрацювання та збереження великих даних (BigData). Порівняно основні платформи для роботи з BigData(BD).

В другому розділі розглянуто основні методи машинного навчання та критерії для фільтрування висловлювань, а також наведено основні підходи до аналізу висловлювання.

Третій розділ присвячено розробці Web-прототипу для аналізу висловлювань користувачів соціальної мережі Twitter. Описано архітектуру прототипу, а також налаштування системи розподіленого опрацювання та збереження висловлювань. Спроектровано прототип для проведення і

збереження результатів статистичного аналізу висловлювань та описано етапи реалізації.

В четвертому розділі проведено базовий аналіз розробленого прототипу, а також визначено подальші напрямки розвитку.

Постановка задачі:

- 1) Розробити прототип, що дозволяє збирати та аналізувати висловлювання користувачів соціальних мереж в режимі реального часу;
- 2) Розробити критерії фільтрування висловлювань користувачів, що дозволить зменшити об'єм неінформативних висловлювань;
- 3) Застосувати наївний баєсівський класифікатор з машинним навчанням для первинного аналізу повідомлень;
- 4) Реалізувати можливість експертного навчання для покращення ефективності аналізатора;
- 5) Надати користувачам прототипу можливість переглядати результати аналізів висловлювань користувачів.

