

УДК 004.056.5

Ю.І. Лехан, О.А. Пастух докт. техн. наук, проф.

Тернопільський національний технічний університет імені Івана Пулюя, Україна

ОБРОБКА ТА АНАЛІЗ ТЕКСТОВОЇ ІНФОРМАЦІЇ МЕТОДАМИ МАШИННОГО НАВЧАННЯ

Y.I. Lekhan, O. A. Pastukh Dr, Prof.

PROCESSING AND ANALYSIS OF TEXTUAL INFORMATION USING METHODS OF MACHINE LEARNING

Лінгвістична обробка природномовних текстів є однією з центральних проблем інтелектуалізації інформаційних технологій. Цій проблемі приділяється значна увага в розвинутих країнах Європи та США, свідченням цього є виділення величезних коштів на розробку лінгвістичного програмного забезпечення. Велику кількість науково-дослідних програм спрямовано на розвиток лінгвістичних інформаційних систем. У зв'язку з бурхливим розвитком Інтернету, інших комп'ютерних технологій ця проблема набуває ще більшої значущості.

Ще з середини 50-х років минулого століття значні зусилля науковців були спрямовані на розробку математичних алгоритмів та комп'ютерних програм обробки текстів природною мовою. Для автоматизації аналізу та синтезу текстів створювалися різноманітні моделі процесів обробки тексту, а також відповідні алгоритми та структури представлення даних. Традиційно аналіз природномовних текстів представлявся як послідовність процесів – морфологічний аналіз, синтаксичний аналіз, семантичний аналіз. Людська мова стала ньюанізованою формою спілкування, яка містить величезну кількість інформації, яка часто виходить за межі тих самих слів. NLP стане важливою технологією для подолання розриву між людськими комунікаціями та цифровими даними.

Обробка природної мови (Natural Language Processing – NLP) – це міждисциплінарна галузь, яка стоїть на перетині комп'ютерних наук, штучного інтелекту та обчислювальної лінгвістики, основним проблемним полем якої є забезпечення взаємодії між комп'ютерами та людськими (природними) мовами. Інтелектуальний аналіз тексту (Text mining) – напрям інтелектуального аналізу даних та штучного інтелекту, метою якого є отримання високоякісної інформації з колекцій текстових документів за допомогою застосування методів машинного навчання та обробки природної мови. Основна задача Text mining полягає в тому, щоб виявити інформацію, яка, можливо, невідома і прихована в контексті іншої інформації. Це досягається за допомогою різних методологій аналізу; обробка природної мови – одна з них, вона виконує лінгвістичний аналіз, що допомагає машині «читати» текст.

За допомогою NLP можна розробляти програмне забезпечення, яке може витягувати інформацію з вихідного тексту, знаходити зв'язок між словами, робити синтаксичний розбір, визначати емоційне ставлення автора, досліджувати тексти на схожість, а також розпізнавати людську мову.

Література

1. Introduction to NLP - Режим доступу:

<https://towardsdatascience.com/introduction-to-nlp-5bff2b2a7170>

2. Getting started with NLP - Режим доступу

<https://medium.com/@gon.esbuyo/get-started-with-nlp-part-ii-overview-of-an-nlp-workflow-7ba1f5948b24>