

## ТЕОРЕТИЧНІ ОСНОВИ ОПРАЦЮВАННЯ ПРИРОДНОЇ МОВИ

### **Крошняк Петро Ярославович**

здобувач другого рівня вищої освіти за спеціальністю 014.09 Середня освіта (Інформатика)  
Тернопільський національний педагогічний університет імені Володимира Гнатюка,  
kroshnyak\_py@fizmat.tnpu.edu

### **Карабін Оксана Йосифівна**

кандидат педагогічних наук, доцент кафедри інформатики та методики її навчання,  
Тернопільський національний педагогічний університет імені Володимира Гнатюка,  
karabin@tnpu.edu.ua

Сучасний світ переживає бурхливий розвиток інформаційних технологій та штучного інтелекту. Завдяки стрімкому зростанню обчислювальної потужності, доступності великих обсягів даних і інновацій в галузі алгоритмів, технології розвиваються з неймовірною швидкістю, впливаючи на всі сфери життя. Однією з ключових галузей, яка відіграє важливу роль у цьому розвитку, є опрацювання природної мови (далі – NLP). NLP є відокремленою галуззю штучного інтелекту, яка спеціалізується на вивченні і розробці методів аналізу та обробки людської мови за допомогою комп'ютерів. Ця галузь є важливою складовою сучасних технологій та має значний вплив на багато сфер життя. Існує кілька ключових тенденцій, які визначають розвиток інформаційних технологій та штучного інтелекту, і NLP грає важливу роль у цьому контексті.

Перша тенденція – це зростання потужності обчислення. Сучасні комп'ютери і сервери стають все потужнішими і можуть ефективно обробляти великі обсяги даних. Це робить можливим використання складних NLP алгоритмів та моделей, що раніше були недосяжними.

Друга тенденція – це зростання доступності даних. Завдяки інтернету та сучасним засобам збору і зберігання даних, стає доступним велика кількість текстової інформації, яка може бути використана для навчання та покращення NLP моделей.

Третя тенденція – це розвиток глибокого навчання (deep learning). Моделі, засновані на глибокому навчанні, стали домінуючими в галузі NLP та дозволили досягти значних результатів у завданнях, таких як автодоповнення тексту, машинний переклад та аналіз тексту.

Опрацювання природної мови (NLP) – це галузь штучного інтелекту, яка займається вивченням і розробкою методів обробки і аналізу людської мови за допомогою комп'ютерних систем. NLP відіграє важливу роль в різних аспектах сучасного життя, включаючи машинний переклад, аналіз тексту, чат-боти, розпізнавання мови та багато інших застосувань. Теоретичні основи опрацювання природної мови є ключовими для розуміння і розвитку цієї галузі.

Морфологія та синтаксис: основні теоретичні концепції NLP включають в себе вивчення морфології (будови слів) та синтаксису (структури речень) мови. Морфологічний аналіз включає в себе розпізнавання словоформ, лематизацію (зведення словоформ до їх базової форми), і визначення граматичних характеристик слів. Синтаксичний аналіз досліджує відносини між словами в реченні і розуміння їхньої структури.

**Семантика:** семантика в NLP вивчає значення слів та речень. Це включає в себе семантичний аналіз, який допомагає розуміти смислові відношення між словами, і виразові семантичні моделі, такі як Word2Vec та GloVe, які допомагають векторизувати слова та вирази.

**Дискурс:** дискурсивний аналіз досліджує способи, якими речення пов'язані в тексті для створення змістовного контексту. Він включає в себе вивчення семантичних та прагматичних аспектів комунікації та визначення відношень між пропозиціями.

**Мовні моделі:** мовні моделі використовуються для прогнозування наступних слів або символів в тексті. Вони використовуються у багатьох NLP завданнях, таких як машинний переклад, автодоповнення тексту та генерація тексту. Такі моделі, як Recurrent Neural Networks (RNNs) та Transformer, стали ключовими для досягнень в цій галузі.

**Семантичні мережі та глибоке навчання:** застосування глибокого навчання до NLP дозволило досягти значних результатів у вирішенні різноманітних завдань. Моделі, які базуються на глибокому навчанні, можуть автоматично витягати корисну інформацію з тексту, розпізнавати об'єкти та відносини, і здійснювати багато інших NLP завдань.

**Датасети та оцінка результатів:** для навчання та оцінки NLP моделей важливо мати доступ до великих датасетів з різноманітними текстами. Також важливо використовувати метрики, які дозволяють виміряти якість роботи моделей, такі як точність, відновлення та F-міра.

Теоретичні основи опрацювання природної мови розвиваються швидко і стають все більше досконалими завдяки поєднанню мовознавчих теорій з інноваціями в галузі штучного інтелекту. Це робить NLP потужним інструментом для різних сфер життя, від бізнесу та науки до соціальних мереж та освіти. Розуміння та використання цих теоретичних основ допомагають досягти нових досягнень в галузі обробки природної мови.

Таким чином, загалом, сучасні тенденції розвитку інформаційних технологій та штучного інтелекту свідчать про важливість NLP як ключового компонента цього розвитку. Вивчення теоретичних основ опрацювання природної мови стає все більш важливим для досягнення нових досягнень у цій захоплюючій галузі. Важливість NLP полягає в тому, що ця галузь дозволяє комп'ютерам розуміти та генерувати людську мову. Це має велике значення для автоматизації багатьох завдань, від відповідей на запитання користувачів та аналізу великих обсягів тексту до створення інтелегентних чат-ботів та систем машинного перекладу. NLP також відкриває нові можливості в науці, дозволяючи дослідникам аналізувати інформацію з великих корпусів тексту для виявлення нових знань.

### Список використаних джерел:

1. What is Natural Language Processing (NLP)? URL: [https://aws.amazon.com/what-is/nlp/?nc1=h\\_ls](https://aws.amazon.com/what-is/nlp/?nc1=h_ls) (дата звернення: 02.11.2023).
2. Basics of Natural Language Processing for Beginners. URL: <https://medium.com/geekculture/basics-of-natural-language-processing-for-beginners-d86351df9d09> (дата звернення: 04.11.2023).
3. Artificial intelligence. URL: <https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence> (дата звернення: 05.11.2023).