

НЕОБХІДНІСТЬ РОЗРОБКИ СИСТЕМИ АНАЛІЗУ ТА ОБРОБКИ ІНФОРМАЦІЇ ОРИГІНАЛЬНИХ ТЕКСТІВ

Сучасний період розвитку людського суспільства характеризується широким використанням цифрових технологій. Більшість населення не працює з друкованими варіантами даних надаючи перевагу зручному опрацюванню величезної кількості інформації в електронному вигляді. При роботі над текстовими документами користувачу необхідно дуже багато роботи виконувати вручну. Результатом такої роботи як правило повинен виступати повноцінний текст, що характеризуватиметься відсутністю будь-якого роду помилок. Для досягнення такої цілі автору необхідно багаторазово перечитувати свою роботу для знаходження всіх недоліків. Перед користувачем виникає необхідність зробити перевірку, знайти фрагменти тексту з помилками, оцінити характер помилок, передбачити можливі вирішення, прийняти певні рішення для покращення якості викладеного змісту. Ці причини змушують розробників шукати нові способи для представлення даних з можливістю ефективно керувати їх змістом, виділяти основні частини та проводити якісний первинний аналіз, застосовувати комплексний підхід для вибіркового розгляду окремих текстових частин.

За останні роки в області обробки текстів природною мовою отримали розвиток дослідження, пов'язані з моніторингом соціальних мереж і обробкою навмисно спотворених текстів. Розроблено різні методи вирішення задач інформаційного пошуку та машинного перекладу [1]. При цьому базові принципи обробки пов'язуються із комп'ютерною лінгвістикою. Вона перебуває в постійному розвитку і адаптує розроблені в цих науках методи та інструменти. Початки зародження комп'ютерної лінгвістики приводять до досліджень відомого американського лінгвіста Н. Хомського в області формалізації структури природної мови. Відомий вчений революціонував лінгвістику, розглядаючи мову – ключовий фактор в розумінні штучного інтелекту. Сюди можна також віднести перші експерименти з машинним перекладом, виконані програмістами та математиками, а також розроблені перші програми розуміння природної мови [2]. Негативні звіти про якість машинного перекладу змусили призупинити дослідження в галузі обробки природної мови, але з появою глибокого навчання були створені алгоритми для обробки природної мови, що показали вражаючі результати. Прорив відбувся в 2006 році, коли Джеффри Хінтон, Саймон Осіндеро та Йі-Уай Тех розробили алгоритми підготовки мережі з трьома прихованими шарами. Керівник компетенції AltexSoft з питань науки Олександр Кондуфоров прокоментував так роботу дослідників: «Після того, як вчені дізналися, як тренувати глибокі нейронні мережі, вони створили революцію в комп'ютерному зорі (computer vision) та обробці звуку (audio processing), а також в NLP (Natural Language Processing). Впровадження нових архітектур нейронних мереж, таких як Transformers, дозволило зробити величезний крок у галузі обробки текстів [3].»

До завдань, які вимагають застосування нових методів обробки текстів, можна віднести вилучення думок, визначення емоційного забарвлення текстів, аналіз реального впливу джерел інформації, обробку некоректних або навмисно спотворених текстів [4].

Важливими завданнями автоматичного аналізу електронних текстів виступає корекція граматичних і морфологічних помилок [5]. Сучасні пошукові системи і текстові редактори (Google, Word Office, META) частково вирішують це завдання. Вони містять орфографічні коректори, які зберігають всі форми слів і статистику помилок. Орфографічні коректори подібного типу добре працюють в хмарних обчисленнях, але показують невисоку швидкість на персональних комп'ютерах з обмеженими обчислювальними ресурсами.

Список використаної літератури

1. Lizunov P. Near-duplicate detection for tables on the base of the local-sensitive hashing and the nearest neighbor search methods / . Lizunov, a. Biloshchytskyi, a. Kuchansky, S. Biloshchytska, L. chala // – Eastern-European Journal of enterprise technologies. – 2016 – №6 – p. 4– 10.
2. Виноград Т. Программа, понимающая естественный язык — М.: Мир, 1976.
3. Eххact - [Електронний ресурс] – Режим доступу до статті: <https://blog.eххactcorp.com/deep-learning-in-natural-language-processing-history-and-achievements/>
4. Sutskever I. Sequence to Sequence Learning with Neural Networks [Text] / I. Sutskever, O. vinyals, Q. v.Le // In advances in Neural Information processing Systems. – 2014. – p. 3104–3112.
5. Luong T. addressing the rare word problem in Neural machine Translation / T. Luong, I. Sutskever, Q Le // – In proceedings of the acL-IJcNLP. – 2013. – p. 11 – 19.