

УДК 004.934

## ДОСЛІДЖЕННЯ МЕТОДІВ МОДЕЛЮВАННЯ ІНФОРМАЦІЙНОГО ПОШУКУ ЗА ДОПОМОГОЮ ЛІНГВІСТИЧНИХ АВТОМАТІВ

Белінський Г. А.

Науковий керівник – проф. Шубін І. Ю.

Харківський національний університет радіоелектроніки, каф. ПІ  
м. Харків, Україна

email: [heorhii.bielinskyi@nure.ua](mailto:heorhii.bielinskyi@nure.ua)

The goal of research is to develop algorithms and software based on linguistic automata for the optimization of information retrieval. Additionally, the adaptability of these methods for various search tasks is explored, along with determining the most effective strategies for utilizing linguistic automata in this context.

To achieve the set goal, methods of mathematical modeling and analysis of linguistic automata are employed. Specifically, the research involves analyzing their potential in the context of optimizing information retrieval and determining optimal operating parameters.

Лінгвістичні автомати (далі по тексту – ЛА) є ефективним інструментом для моделювання процесів інформаційного пошуку. У сучасному інформаційному суспільстві, де об'єм даних постійно зростає, інструменти для швидкого та точного пошуку інформації стають все важливішими.

ЛА можуть знаходити застосування у різних галузях, включаючи науку, техніку, медицину, соціологію та інші. Особливо актуальним є їх використання у веб-пошуку, де ефективність алгоритмів визначає зручність користувачів та результативність досліджень.

Під час аналізу методів моделювання інформаційного пошуку з використанням ЛА, виявлено такі ключові напрями досліджень:

– Моделювання лінгвістичних структур.

Використання ЛА дозволяє побудувати моделі для аналізу та розпізнавання лінгвістичних одиниць, що полегшує розвиток нових алгоритмів для покращення якості пошуку.

– Оптимізація інформаційного пошуку.

ЛА дозволяють оптимізувати процес пошуку шляхом розробки алгоритмів, які ураховують семантичні зв'язки та контекстні особливості запитань користувачів.

– Аналіз текстової інформації.

ЛА використовуються для аналізу текстів, визначення ключових слів, витягування суттєвої інформації та покращення точності результатів пошуку.

Статистичні методи, такі як частотний аналіз, можуть визначати ключові слова та патерни у тексті на підставі їхньої частоти входження.

Використання цих методів дозволяє лінгвістичним автоматам ефективно розпізнавати та аналізувати мовний контент, що робить їх корисними для багатьох завдань, таких як покращення пошукових систем, аналіз текстів або розпізнавання мовлення.

Реалізація лінгвістичних автоматів для моделювання інформаційного пошуку включає в себе ряд технологій програмної інженерії. У процесі створення лінгвістичних автоматів для моделювання інформаційного пошуку широко використовуються різноманітні високорівневі мови програмування. Python, Java, C++, C# – це лише деякі з мов, що надаються у розробці завдяки своїй потужній функціональності та підтримці різних аспектів розробки [2].

Програмісти можуть обирати мову відповідно до конкретних потреб проекту. Наприклад, Python часто використовується в області обробки природної мови (NLP) завдяки своїй простоті та багатому екосистемі бібліотек, таких як NLTK та SpaCy.

Важливою складовою створення лінгвістичних автоматів є використання бібліотек для обробки природної мови (NLP). NLTK, SpaCy, TextBlob – це інструменти, що надають реалізації різноманітних алгоритмів обробки тексту, включаючи токенізацію, аналіз частотності та синтаксичний аналіз.

Використання цих бібліотек спрощує завдання розпізнавання та розуміння мови, роблячи легшим створення лінгвістичних моделей для інформаційного пошуку.

Застосування методів машинного навчання та глибокого навчання є важливою частиною розробки лінгвістичних автоматів. Використання фреймворків, таких як TensorFlow, PyTorch і scikit-learn, дозволяє тренувати моделі, які можуть виконувати завдання, пов'язані з інформаційним пошуком, такі як класифікація текстів або генерація контенту [3].

Моделі можуть навчатися визначати семантичні зв'язки між словами, розуміти концепції та класифікувати інформацію з великих обсягів даних.

Узагальнена схема лінгвістичного автомату [4] може включати наступні етапи:

Вхідні дані: Лінгвістичний автомат отримує вхідні дані у вигляді тексту або послідовності символів.

Токенізація: Вхідний текст розбивається на окремі токени або слова.

Побудова структури: Токени або слова аналізуються для визначення структури або синтаксичних відносин між ними, які можуть бути представлені у вигляді графа або дерева.

Аналіз: На основі побудованої структури проводиться аналіз тексту з використанням внутрішніх правил або алгоритмів.

Генерація вихідних даних: Лінгвістичний автомат може генерувати вихідні дані у формі відповіді на запит або інші види обробленої інформації.

Вихідні дані: Отримані вихідні дані можуть бути подані у вигляді тексту, графіків або інших структур для подальшого використання або відображення.

Висновок: лінгвістичні автомати виявляються ефективним інструментом для моделювання процесів інформаційного пошуку. Вони знаходять застосування у різних галузях, включаючи науку, техніку, медицину, соціологію та веб-пошук. Під час аналізу методів моделювання виявлено ключові напрями досліджень: моделювання лінгвістичних структур, оптимізація процесу пошуку та аналіз текстової інформації. Створення лінгвістичних автоматів включає в себе використання різних технологій програмної інженерії, таких як високорівневі мови програмування, бібліотеки для обробки природної мови та методи машинного навчання. Ці інструменти сприяють розробці ефективних систем інформаційного пошуку, що відповідають вимогам сучасного інформаційного середовища.

Список використаних джерел:

1. Д.В. Ланде, І.Ю. Субач, Ю.Є. Бояринова : «Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки» dwl.kiev, 2018. <http://dwl.kiev.ua/art/oiad/oiad.pdf>.

2. Manning, Christopher D., and Hinrich Schütze. "Foundations of Statistical Natural Language Processing." MIT Press, 1999. <https://www.amazon.com/Foundations-Statistical-Natural-Language-Processing/dp/0262133601>.

3. Bird, Steven, Ewan Klein, and Edward Loper. "Natural Language Processing with Python." O'Reilly Media, 2009. <https://www.amazon.com/Natural-Language-Processing-Python-Analyzing/dp/0596516495>.

4. Долуда В.П, «Дослідження методів синтаксичного аналізу для системи автоматичної обробки текстової інформації» openarchive.nure.ua, 2019. <https://openarchive.nure.ua/server/api/core/bitstreams/ecdf458-7717-4093-86a5-4ec0fc58d282/content>.