

ОСОБЛИВОСТІ АРХІТЕКТУРИ RULE-BASED І ЛОКАЛЬНИХ МОВНИХ МОДЕЛЕЙ ДЛЯ ОЦІНЮВАННЯ ЯКОСТІ ТЕКСТУ РЕЗЮМЕ

Матвєєв М.С., Сердюк Н.М.

Кафедра Комп'ютерних інтелектуальних технологій та систем
Харківський національний університет радіоелектроніки,
Україна

E-mail: mykhailo.matvieiev@nure.ua,
nataliya.serdyuk@nure.ua

Abstract

The paper presents a conceptual design of an intelligent system for automated resume text analysis and improvement. The proposed approach combines rule-based algorithms with lightweight local large language models (LLMs) such as Mistral and Phi-3 Mini to evaluate the structure and quality of written content. The system performs syntactic, semantic, and stylistic analysis, detects typical weaknesses, and provides feedback in the form of adaptive recommendations aimed at enhancing precision, completeness, and linguistic clarity. The integration of rule-based and LLM approaches allows scalable implementation in HR software with minimal resource consumption and high data privacy.

Сучасний етап розвитку інформаційних технологій характеризується активним упровадженням інтелектуальних систем, здатних автоматизувати аналіз і покращення текстових матеріалів. Одним із найважливіших прикладних напрямів є створення програмних рішень, які допомагають користувачам формувати якісні професійні документи, зокрема резюме. Від структурованості та змістовності такого тексту значною мірою залежить сприйняття кандидата роботодавцем. Попри наявність великої кількості онлайн-редакторів і конструкторів, більшість із них виконують лише технічні функції, не враховуючи змістовну логіку й мовні особливості опису досвіду. Це зумовлює потребу у створенні інтелектуальної системи, здатної здійснювати не лише поверхневу перевірку, а й комплексний семантичний і стилістичний аналіз тексту [1].

Метою дослідження є розроблення інтелектуальної системи автоматизованого аналізу та покращення текстів резюме, що поєднує методи лінгвістичного аналізу, машинного навчання та генеративних моделей для підвищення якості представлення професійних даних. Запропонована архітектура складається з чотирьох основних модулів: блоку попередньої обробки тексту, rule-based аналізатора, генеративного помічника на основі локальної мовної моделі та модуля формування рекомендацій. Після введення тексту користувачем система проводить синтаксичний і семантичний аналіз, розпізнає відсутні або слабко виражені елементи структури (наприклад, ключові слова, показники результативності, зв'язки між навичками та досвідом), а потім передає результати до генеративного модуля, який формує зрозумілі й доречні поради з покращення.

Для реалізації підходу використано інструментарій Python та бібліотеки spaCy, NLTK і transformers, що забезпечують багаторівневу обробку тексту. Rule-based перевірка виконує лексичний аналіз, пошук стандартних помилок, надлишкових слів і логічних розривів, тоді як легка локальна LLM, розгорнута через платформу Ollama, оцінює семантичну цілісність тексту та пропонує способи підвищення його інформативності. Застосування локальних моделей, на відміну від систем типу Rezi чи Zety, запропоноване рішення не передає персональні дані у хмару та може функціонувати локально. Rule-based аналіз дає пояснюваність результатів, а LLM-компонент — адаптивність і гнучкість [2].

Попередні дослідження у сфері автоматизованого аналізу резюме свідчать, що використання комбінованих rule-based і мовних моделей забезпечує високу точність виявлення структурних та

семантичних недоліків у текстах. Такі системи демонструють здатність виявляти понад 80 % помилкових або надто узагальнених фраз, а також формувати рекомендації, які підвищують рівень конкретизації опису навичок у середньому на 25–30 %. Згідно з результатами наведених у літературі прикладів, поєднання чітко визначених правил із навченою мовною моделлю дозволяє досягти більшої гнучкості й адаптивності під час роботи з різними стилями текстів. Розроблені у відомих публікаціях інтерфейси відображають результати аналізу у вигляді інтерактивних підказок, які допомагають користувачу коригувати текст у режимі реального часу, наближаючи його до професійного стандарту [3].

Архітектура запропонованої системи включає послідовні етапи: введення тексту користувачем, попередня обробка даних, rule-based аналіз, генеративна обробка та формування рекомендацій. Взаємодія між модулями реалізована через локальний API, що дозволяє масштабувати рішення для інтеграції у більші HR-платформи або системи управління персоналом.

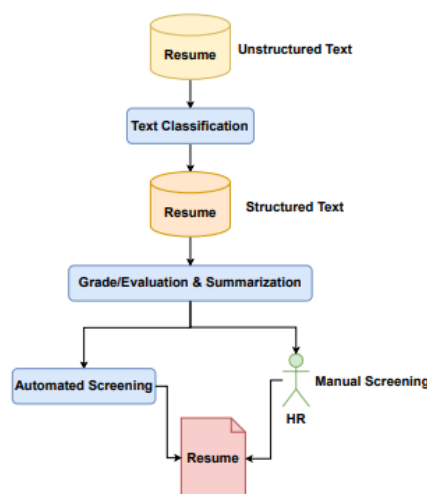


Рис. 1. Архітектура інтелектуальної системи автоматизованого аналізу та покращення тексту резюме

Останні публікації з тематики обробки природної мови у сфері рекрутингу підтверджують ефективність використання великих мовних моделей для аналізу та вдосконалення текстів резюме. У низці досліджень доведено, що поєднання NLP-технологій із алгоритмами класифікації дозволяє не лише автоматично вилучати навички та ключові слова, а й оцінювати рівень релевантності кандидатів і якість структурування інформації. Моделі великого масштабу, навчені на великих корпусах даних, демонструють високу точність під час визначення змістових елементів і логічних зв'язків у професійних описах. Окрему увагу приділяють етичним аспектам функціонування таких систем — питанням зниження ризику упередженості щодо статі, віку чи культурного походження під час автоматичного відбору [4]. Оглядові праці також підкреслюють перспективність комбінованих архітектур, де rule-based аналіз доповнюється семантичними методами розпізнавання, що створює основу для розвитку гнучких і пояснюваних інтелектуальних систем у сфері обробки текстових документів.

Розроблена система може демонструвати ефективність інтеграції rule-based і генеративних методів для покращення структурної та семантичної якості текстів резюме. Використання локальних мовних моделей гарантує конфіденційність і високу швидкодію, що дозволяє впроваджувати рішення у HR-середовищах з обмеженими ресурсами. Використання локальних мовних моделей забезпечує високу швидкість роботи та конфіденційність даних, що особливо важливо при обробці персональних документів. Подальша робота передбачає автоматичне формування профілю компетентностей на основі класифікації навичок і розширення функціоналу до оцінювання релевантності кандидата вакансії.

Література

1. Gan C., Zhang Q., Mori T. *Application of LLM Agents in Recruitment: A Novel Framework for Resume Screening* [Electronic resource] // arXiv preprint. – 2024. – arXiv:2401.08315. – Access mode: <https://arxiv.org/abs/2401.08315>
2. Sajid A., et al. *Resume Screening with Natural Language Processing (NLP)* // *Alphanumeric Journal*. – 2024. – Vol. 12(2).
3. IRJET Team. *NLP-Based Resume Analysis, Skill Enhancement and Automation of Screening* // *International Research Journal of Engineering and Technology (IRJET)*. – 2024. – Vol. 12(4). – Access mode: <https://www.irjet.net/archives/V12/i4/IRJET-V12I4268.pdf>
4. *Resume Screening with Natural Language Processing (NLP): A Case Study on Automated Candidate Evaluation* [Electronic resource] // *DergiPark Journal of Computer Science*. – 2024. – Access mode: <https://dergipark.org.tr/en/download/article-file/4159614>