

АВТОМАТИЗАЦІЯ ПРОЦЕСУ ПЕРЕКЛАДУ З ВИКОРИСТАННЯМ ТЕХНОЛОГІЙ ШТУЧНОГО ІНТЕЛЕКТУ

Безуглий В.О., Філімончук Т.В.

Харківський національний університет радіоелектроніки, Харків, Україна

Сучасні великомовні моделі (LLM) відкривають можливості для автоматизації перекладу різномірних документів у середовищах типу iPaaS. Завдяки розумінню контексту, семантики та структури тексту [1,2], вони забезпечують точність перекладу, збереження форматування та термінологічну узгодженість між різними форматами даних. Це створює основу для інтеграції інтелектуальних мовних сервісів у корпоративні екосистеми, підвищуючи ефективність багатомовної взаємодії.

Ключовою умовою масштабованості таких систем є уніфіковане подання даних і збереження структури під час трансформацій між форматами (DOCX, PDF, HTML). Для цього застосовують двомовний стандарт XLIFF, який сегментує текст і паралельно зберігає вихідний (source) та перекладений (target) контент, забезпечуючи узгодженість і стабільність під час багаторівневих перетворень.

Метою доповіді є представлення та експериментальна оцінка архітектури системи автоматизованого перекладу, яка побудована на основі файлового формату XLIFF із використанням великомовних моделей (OpenAI, Anthropic, Gemini) та інтеграцією в iPaaS-процеси.

У доповіді розглядається повний технологічний конвеєр обробки даних, що включає етапи: екстракції контенту, семантичної сегментації, формування «tag-aware» підказок для LLM, пост-редагування, отриманих результатів та зворотного складання перекладеного матеріалу у вихідний формат документа [3], а також наводяться результати вимірювань якості перекладу (BLEU/COMET) та цілісності розмітки (збереження порядку/вкладеності XLIFF-тегів) на вибірці технічних, наукових та вебтекстів для трьох режимів: базовий LLM, LLM з «tag-aware» інструкціями, гібридний (LLM + пост-редагування) [1, 3]. Наведені дані показують, що XLIFF як інтерфейс обміну та спеціальні інструкції до LLM зменшують втрати розмітки й обсяг пост-редагування, а семантична сегментація знижує контекстні помилки у довгих документах. Інтеграція в iPaaS спрощує впровадження й повторне використання робочих потоків.

Список літератури

1. Kleidermacher H.C., Zou J. Science Across Languages: Assessing LLM Multilingual Translation of Scientific Papers. arXiv, 2025. DOI: <https://doi.org/10.48550/arXiv.2502.17882>
2. [Amano T. та ін. AI-mediated translation presents two possible futures for global knowledge exchange. PMC, 2025. DOI: https://doi.org/10.1371/journal.pbio.3003215](https://doi.org/10.1371/journal.pbio.3003215)
3. Brach W., Kostal K., Ries M. The Effectiveness of Large Language Models in Transforming Unstructured Text to Standardized Formats. IEEE Access, 2025, PP(99):1–1. DOI: <https://doi.org/10.1109/ACCESS.2025.3573030>