

АНАЛІЗ МЕТОДІВ ОБРОБКИ ПРИРОДНОЇ МОВИ ДЛЯ СЕМАНТИЧНОГО ЗІСТАВЛЕННЯ ОСВІТНІХ ПРОГРАМ ТА СТАНДАРТІВ ПРОФЕСІЙНИХ КОМПЕТЕНЦІЙ

Д.В. Арсенюк¹

¹ *магістрант кафедри програмної інженерії та інтелектуальних технологій управління, НТУ «ХПІ», Харків, Україна*

На даний момент освітні установи, пропонують багато навчальних курсів, описи яких є неструктурованими, хоча ринок вимагає чітко верифікованих навичок від працівників. Це створює значний "розрив" між знаннями та реальними вимогами ринку праці. Проблема полягає в тому, що традиційні підходи до аудиту освітніх програм базуються на ручному зіставленні описів курсів із професійними стандартами. Такий процес є немасштабованим в умовах тисяч навчальних дисциплін. Застосування методів обробки природної мови (NLP) є перспективним рішенням, оскільки воно дозволяє автоматизувати аналіз великих обсягів тексту та розраховувати семантичну близькість між документами. Враховуючи це сфера інтелектуального аналізу освітніх текстів активно розвивається, пропонуючи інструменти для автоматизації подібних завдань [1]. Водночас важливо враховувати виклики, пов'язані з неоднозначністю природної мови та складністю самих моделей. Тому метою дослідження є аналіз та порівняння можливостей сучасних методів NLP для автоматизованого семантичного зіставлення описів навчальних курсів та стандартів професійних компетенцій.

Для вирішення задачі семантичного зіставлення можуть використовуватись різні класи методів обробки природної мови. Традиційні лексичні підходи, такі як TF-IDF або Bag-of-Words, дозволяють проводити швидке зіставлення за ключовими словами, але вони повністю ігнорують синонімію та семантичний контекст. Більш просунуті методи, що базуються на статичних векторних моделях (наприклад, Word2Vec або GloVe), здатні враховувати контекст окремих слів, проте все ще обмежені у розумінні повних речень. Найбільш ефективними на сьогодні є сучасні моделі-трансформери, такі як SBERT [2]. Ці моделі здатні створювати контекстні векторні представлення для цілих документів. Саме такий підхід дозволяє їм розраховувати семантичну близькість між описом курсу та стандартом компетенції з високою точністю, враховуючи окремі терміни та загальний зміст. Проте важливо пам'ятати, що практична інтеграція таких складних моделей у існуючі системи управління навчанням вимагає значних обчислювальних ресурсів, через що доцільно використовувати модульну архітектуру систем, при якій NLP-сервіси підключаються до основної платформи через API.

Таким чином, застосування сучасних NLP-моделей для аналізу освітніх програм відкриває нові можливості для прозорості та валідації академічних кваліфікацій. Використання методів семантичного аналізу дозволяє автоматизувати складний процес аудиту навчальних планів та, як наслідок, значно підвищити цінність і конкурентоспроможність дипломів на ринку праці.

Список літератури:

1. *Papamitsiou, Z. A review of text mining in education / Z. Papamitsiou, A. Economides // Journal of Educational Data Mining. – 2014. – Vol. 6, № 1. – P. 51–90.*
2. *Reimers, N. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks / N. Reimers, I. Gurevych // Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP). – 2019. – P. 3982–39*