**Улучшение качества кластеризации навыков ИТ-вакансий с помощью тонкой настройки модели sentence transformer**

Рамазанова Валия Сапкиреевна

(Евразийский национальный университет им. Л.Н. Гумилева, г. Астна, РК)

В данной статье рассматривается применение моделей эмбеддингов предложений для кластеризации навыков, собранных с сайта интернет-рекрутмента hh.kz. Кластеризация навыков может быть использована для создания групп близких по смыслу навыков в графе знаний вакансий. Основная цель исследования – улучшение качества кластеризации навыков с помощью дообучения модели трансформеров предложений на доменных данных. Проведенные эксперименты показали значительное улучшение показателей качества модели на русском языке, включая метрики Cosine-Similarity, Manhattan-Distance, Euclidean-Distance и Dot-Product-Similarity. Применение агломеративной кластеризации к эмбеддингам навыков до и после дообучения модели продемонстрировало улучшение качества кластеров, что подтверждается увеличением среднего коэффициента силуэта и снижением процента ошибочных кластеров. В статье также обсуждаются ограничения проведенного исследования и предлагаются направления для дальнейшего развития.