**BC/NW 2024 № 1 (41):15.3**

**ИСПОЛЬЗОВАНИЕ ОБЕЗЛИЧЕННЫХ ДАННЫХ В ЗАДАЧАХ МАШИННОГО ОБУЧЕНИЯ**

Бабак Н.Г Оцоков Ш.А.

Для развития своих сервисов у компаний возникает желание обрабатывать персональные данные пользователей, строить прогнозы на их основе. Из-за правовых ограничений и рисков понести репутационные потери данные в открытом виде нельзя обрабатывать и хранить в неспециализированных системах. Также в случае обработки данных в компаниях-партнёрах их необходимо передать по сети, что требует определённых соглашений и соответствующего уровня защиты. Решить, описанные выше проблемы, позволяет обезличивание данных.

Существующие системы обезличивания [1-2] обладают следующими недостатками:

1. используются узконаправленные подходы, где распознавание персональных данных реализовано вручную или на основе правил, решения ограничены применением в структурированных данных;
2. чаще всего обезличивание реализуется разрушающим способом, после чего такие данные невозможно использовать, например, в задачах машинного обучения;
3. отсутствует взаимосвязь между видами персональных данных и применяемым алгоритмом обезличивания.

Таким образом, эти недостатки приводят к тому, что обезличенные данные трудно или практически невозможно использовать, так как они теряют свою ценность. В частности, наиболее чувствительными являются задачи, связанные с машинным обучением и привлечением третьих лиц, недопущенных к персональным данным, например, перевод подрядчиками.

Для использования обезличенных данных необходимо обеспечить сохранение стилистических, семантических, лексических и морфологических особенностей текста. Выполнить данную задачу можно, реализовав автоматизированную систему синонимического обезличивания [3], которая обеспечивает сохранение ценности данных.

**Литература**

1. **Саксонов, Е. А.** Процедура обезличивания персональных данных / Е. А. Саксонов, Р. В. Шередин // Наука и образование: научное издание МГТУ им. Н.Э. Баумана. – 2011. – № 3. – С. 1.
2. **Oleksy M., Ropiak N., Walkowiak T.** Automated anonymization of text documents in Polish. Procedia Computer Science. 2021; 192. P. 1323-1333.
3. **Патент на изобретение RU 2804747 C1**. Способ и система обезличивания конфиденциальных данных. Заявление № 2022132305, 09.12.2022; опубликовано 04.10.2023 / Бабак Н. Г., Белорыбкин Л. Ю., Теренин А. А., Шаброва А. И.; заявитель ПАО Сбербанк.