

## **РАЗРАБОТКА СИСТЕМЫ АВТОМАТИЗИРОВАННОЙ ПРОВЕРКИ ТЕКСТОВ НА ПРЕДМЕТ СОДЕРЖАНИЯ ИНФОРМАЦИИ ОГРАНИЧЕННОГО ДОСТУПА**

**Меренков Д.Н.** (федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

**Научный руководитель – кандидат технических наук, доцент ФБИТ Коржук В.М.**  
(федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

В данной работе рассматривается система автоматизированной проверки текстов на предмет содержания информации ограниченного доступа. Предложенная система направлена на снижение финансовых издержек при проведении идентификационной экспертизы для открытого опубликования информации, а также на повышение точности распознавания наличия информации ограниченного доступа в текстовых массивах данных.

### **Введение.**

Для открытого опубликования информации, организация, от имени которой будет опубликована информация, должна совершить ряд проверок, чтобы убедиться, что в материалах отсутствует информация ограниченного доступа: конфиденциальная информация, сведения, составляющие государственную тайну или информация о технологиях двойного назначения, в целях соблюдения требований законодательства Российской Федерации и избежания утечки такой информации при ее открытом опубликовании.

Актуальность работы определяется существенным количеством генерируемого научным сообществом и организациями-участниками внешнеэкономической деятельности научно-технического материала и необходимостью соблюдения требований законодательства Российской Федерации при открытом опубликовании результатов научно-технической деятельности, работ и результатов интеллектуальной деятельности.

В настоящее время проверка текстов на предмет содержания информации ограниченного доступа осуществляется сущностным или перечневым методом. Сущностный метод заключается в законодательном определении признаков сведений, составляющих информацию ограниченного доступа. Перечневый метод подразумевает помимо законодательного определения, использование перечней сведений, отнесенных к информации ограниченного доступа, как правило, подзаконных или локальных нормативных актов. В результате подобный способ определения наличия в тексте информации ограниченного доступа, ведет к низкой эффективности работы, ввиду ручной проверки, а также к увеличению уровня финансовых издержек, ввиду необходимости содержания в штате большого количества экспертов.

Целью работы является повышение эффективности проведения идентификационной экспертизы, проводимой для открытого опубликования информации за счет:

- 1) определения тематики входного массива данных
- 2) выявления набора двусмысленно интерпретируемых слов/предложений/словосочетаний
- 3) выбора наиболее компетентного в заданной сфере эксперта, на основании предварительно обученной или периодически обучаемой системы

## **Основная часть.**

Для решения поставленной проблемы предлагается использовать разработанную автором систему. Система может являться частью системы помощи принятия решения.

Работу системы можно разделить на 3 этапа. Во время первого этапа на вход системе подается текстовая информации, наличие сведений ограниченного доступа в которой необходимо проверить. Система, в свою очередь, при помощи методов машинного обучения, определяет тематику входного текста из предварительно заданного набора, а также находит и выделяет в тексте слова, предложения или словосочетания, которые могут являться признаками наличия в тексте информации ограниченного доступа.

Во время второго этапа, система, на основании найденных ранее закономерностей, выбирает наиболее компетентного эксперта в заданной области.

В заключение система уведомляет выбранного эксперта о результатах автоматизированного поиска во входном тексте признаков наличия информации ограниченного доступа. Эксперт, в свою очередь, убеждается в степени корректности предоставленных системой данных и принимает решение о заключении экспертизы.

## **Выводы.**

Организации-участники внешнеэкономической деятельности должны проверять любую открыто-публикуемую информацию на предмет содержания информации ограниченного доступа. Несоблюдение закона об экспортном контроле может повлечь за собой как административную, так и уголовную ответственность, а также запрет заниматься внешнеэкономической деятельностью в течение определенного периода времени, что может негативно сказаться на репутации или существовании организации, а разглашение государственной тайны может не только преследоваться уголовно, но и нанести ощутимый урон национальным интересам Российской Федерации. Для соблюдения требований законодательства Российской Федерации, а также в целях автоматизации процесса проведения идентификационной экспертизы, была предложена разработанная автором система автоматической проверки текстов на предмет содержания информации ограниченного доступа, которая может помочь в принятии решения о заключении экспертизы, а также снизить финансовые издержки и увеличить точность распознавания наличия в тексте информации ограниченного доступа.

Меренков Д.Н. (автор)

Коржук В.М. (научный руководитель)