

УДК 621.3.082

**РЕАЛИЗАЦИЯ АЛГОРИТМА СТЕГАНОГРАФИЧЕСКОГО
ВСТРАИВАНИЯ ИНФОРМАЦИИ В ТЕКСТОВЫЕ ДАННЫЕ
С ИСПОЛЬЗОВАНИЕМ НЕЙРОСЕТЕВЫХ МЕТОДОВ
АНАЛИЗА ЕСТЕСТВЕННЫХ ЯЗЫКОВ**

Р. Р. Ахметзянов¹

*Самарский национальный исследовательский университет
имени академика С.П. Королева, г. Самара, Российская Федерация*

Научный руководитель: В. А. Митекин, к.т.н., доцент
*Самарский национальный исследовательский университет
имени академика С.П. Королева, г. Самара, Российская Федерация*

Ключевые слова: стеганография, естественный языка, нейронные сети

Целью работы является разработка метода скрытной передачи информации в текстовых данных на естественном языке. Таким образом, было необходимо:

1. Реализовать алгоритм встраивания информации в текстовый фрагмент путем правдоподобной замены отдельных его частей (слов, словосочетаний).

2. Реализовать алгоритм «слепого» извлечения информации.

3. Провести исследование эффективности алгоритмов с помощью метрики bit error rate.

К примеру, предложение «Собака шла по улице» должно заменяться на «Кошка шла по улице». Словами, которые могут быть заменены, являются подлежащее, либо же, если подлежащего в предложении нет, сказуемое. Определив слово, возможное для замены, мы оцениваем его длину. Если слово с чётной длиной, то считаем, что встроен бит со значением «0», в противном случае, когда слово нечётной длины, «1», то есть значением встроенного бита мы принимаем остаток от деления длины слова на 2: $bit = len(word) \bmod 2$. Если длина найденного слова не соответствует биту, подлежащему встраиванию, заменяем его на похожее по смыслу подходящей длины слово. Результирующий алгоритм встраивания информации выглядит следующим образом:

1. Определяем длину предложения в словах. Встраивания в короткие предложения может быть наиболее визуально заметно, поэтому они должны быть отброшены.

¹ Ахметзянов Риназ Рафикович, студент группы 6512-100503D,
email: rinaz.akhmetzyanov@yandex.ru

2. Находим в предложении подлежащее или, если подлежащего нет, сказуемое. Если сказуемого в предложении нет, то пропускаем предложение и переходим к следующему.

3. Считаем длину найденного слова, если оно эквивалентно значению бита, который должен быть встроен, то считаем информацию уже встроенной и переходим к следующему предложению.

4. Получаем список слов, похожих на данное, и находим среди них слово необходимой длины. Оцениваем схожесть слов, если она выше некоторого порогового значения, заданного ранее, то игнорируем их.

5. Из оставшихся в списке слов случайным образом выбираем одно и заменяем им слово в предложении.

Порядок предложений, в которые будет встроена информация, определяется генератором псевдослучайных чисел, что позволяет даже в случае обнаружения встроенной информации защитить ее от корректного извлечения.

Из полученного алгоритма встраивания получаем алгоритм извлечения информации из предложения:

1. Определяем длину предложения в словах. Если предложение слишком короткое, то отбрасываем его.

2. Находим в предложении подлежащее или, если подлежащего нет, сказуемое. Если сказуемого в предложении нет, то пропускаем предложение и переходим к следующему.

3. Считаем остаток от деления на два длины полученного слова и извлекаем встроенный бит.

Нахождение в предложении слов, подлежащих встраиванию, реализуется с помощью библиотеки нейросетевого анализа естественного языка spaCy, позволяющей разбить предложения на токены и назначить для каждого токена его часть речи. Слова, с помощью которых будет происходить замена, а также их схожесть с изначальным словом, предоставляет нам библиотека нейросетевого анализа естественного языка word2vec.

Алгоритмы оценивались метрикой bit error rate. BER представляет собой отношение неверно извлечённых битов к их общему числу. Соответственно, чем меньше значение данной метрики, тем более эффективны алгоритмы. Замеры проводились для двух видов текстов, художественного и технического, а также для двух языков: русского и английского. По итогам исследований встраивание и извлечение было наиболее эффективно в текстах на английском языке. В среднем технические тексты показывали лучшие результаты как для русского языка, так и для английского.

Разработанный стеганографический алгоритм наиболее эффективен для встраивания информации в технический текст на английском языке, где его средний BER составлял 0,08, что, учитывая возможность встраивания избыточной информации, позволяет нам извлекать информацию практически без потерь.

УДК 930

ПОЯВЛЕНИЕ ТЕРМИНА «БЕЛОЧЕХИ»

А. В. Бабикина¹

*Самарский национальный исследовательский университет
имени академика С.П. Королева, г. Самара, Российская Федерация*

Научный руководитель: М. М. Леонов, д.и.н., профессор
*Самарский национальный исследовательский университет
имени академика С.П. Королева, г. Самара, Российская Федерация*

Ключевые слова: термин «белочехи», гражданская война, исторический конструкт, чехословацкий легион

Целью статьи является попытка установить появление термина «белочехи» и контекст его употребления в период событий, соотносимых хронологически с этим термином, и в более позднее время в историографии.

Зачастую под понятием «белочехи» подразумевается вся масса антибольшевистского движения, что не вполне верно с исторической точки зрения. Сам термин имеет право на существование в том случае, если под ним подразумевается вполне конкретная социальная база, а именно: представители чешского этноса, выступающие на стороне белого движения, то есть разделявшие его программу.

Эвакуировавшиеся через Дальний Восток чехословаки придерживались позиции вооруженного нейтралитета, что не позволяет применительно ко всем представителям чешского этноса, действовавшим тогда на территории России, применять лексему «белочехи».

Впервые лексема «белочехи» встречается в резолюциях крестьянских восстаний села Чемодановка, где она имела политический контекст, но в дальнейшем в таком же виде не закрепились. Употреблялся этот термин и в приказах и воззваниях Л. Д. Троцкого, снова в политическом контексте.

В дальнейшем лексема «белочехи» будет употребляться в советской печати, но в большей степени с целью сформировать общий образ врага, нежели подразумевая под собой конкретную социальную базу.

¹ Бабикина Анастасия Владимировна, студент группы 5101-460301D,
email: anastaciababikova@yandex.ru