

О. С. Михоленко
(УО «БГТУ», Минск)

АНАЛИЗ ТЕКСТА ФОРМАТА DOCX НА ПРЕДМЕТ НАЛИЧИЯ ОСАЖДЕННОЙ ИНФОРМАЦИИ

Одним из самых распространенных форматов хранения текстовой информации является формат DOCX. Для тайной передачи или хранения информация в таких электронных документах используются стеганографические методы [1].

Алгоритм поиска и извлечения из текста осажженной информации в общем случае можно свести к следующим трем шагам:

1. Анализ текста на предмет наличия в нем несоответствий между показателями свойств одинаковых элементов (различное число пробелов между словами и т.д.). Результат – набор найденных нестандартных значений, который будет основой для выявления закономерности.

2. Нахождение закономерностей между выявленными отклонениями с их последующей группировкой.

3. Сопоставление выделенных групп с одним из алфавитов и дальнейшее декодирование полученной записи.

Был проведен анализ производительности описанного алгоритма (таб. 1). В качестве осажженной информации выбрана двоичная последовательность, каждый символ (набор символов) которой был

закодирован определенным числом пробелов. Столбец `textlength` содержит число символов, столбцы `time(ticks)` и `time(s)` – время выполнения алгоритма в тиках и секундах соответственно.

Таблица 1 – Результаты измерения производительности алгоритма

<code>textlength</code>	<code>time (ticks)</code>	<code>time (s)</code>
6907	2835720	1,056
18760	131084497	47,802
21989	172898949	69,702

Можно сделать вывод, что программная реализация алгоритма позволяет значительно ускорить стеганографический анализ документов.

Литература

1 Урбанович, П. П. Защита информации методами криптографии, стеганографии и обфускации: учеб.-метод. пособие / П. П. Урбанович. – Минск: БГТУ, 2016. – 220 с.