Министерство образования и науки Российской Федерации

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра теоретических основ компьютерной безопасности и криптографии

Поиск информации на персональном компьютере

АВТОРЕФЕРАТ

дипломной работы

студентки 6 курса 632 группы

специальности 10.05.01 Компьютерная безопасность

факультета компьютерных наук и информационных технологий

Мусугалиевой Альбины Геннадьевны

Научный руководитель

доцент, к.ф.-м.н. _____ А. В. Жаркова

18.01.2018 г.

Заведующий кафедрой

профессор, к.ф.-м.н. _____ В. Н. Салий

18.01.2018 г.

Саратов 2018

ВВЕДЕНИЕ

В наше время в связи с развитием информационных технологий и огромным ростом объема данных, доступных как отдельно взятому человеку, так и обществу, существует много проблем с обработкой информации. Жесткий диск компьютера может содержать сотни тысяч файлов. В этих условиях становится актуальной проблема поиска файлов.

Сейчас функции поиска файлов инкапсулированы во многие операционные системы, например, в операционную систему Windows встроен стандартный поиск файлов с заданными параметрами. Но если такого рода поиск является ключевой задачей вашей программы, требуется знать принципы организации этого поиска. При этом в готовых программах далеко не всегда все реализовано лучшим образом. Во-первых, в стандартных программах не всегда используются самые эффективные алгоритмы, а во-вторых, вполне возможно, что вам понадобится изменить стандартное поведение этих программ.

Тема защиты компьютерной информации стала очень популярной в последнее время. Связано это с широким распространением вычислительной техники, внедрением ее практически во все сферы человеческой деятельности. Любые нарушения в работе вычислительных систем с каждым годом становятся для человека все болезненнее и опаснее.

Одной из актуальнейших проблем, связанных со «здоровьем» компьютеров, является проблема их защиты от вредоносных объектов, в частности, их обнаружение на данной машине.

Целью данной работы является изучение различных алгоритмов поиска файлов с заданными параметрами и создание программного продукта для поиска информации на персональном компьютере.

Для достижения поставленной цели требуется решить следующие задачи:

- изучить теоретические сведения, касающиеся поиска информации

на компьютере;

- изучить различные алгоритмы поиска файлов с заданными параметрами;
- рассмотреть классификацию вредоносных объектов и их возможный поиск на компьютере.

Дипломная работа состоит из введения, 4 разделов, заключения, списка использованных источников и 1 приложения. Общий объем работы – 75 страниц, из них 54 страницы – основное содержание, включая 24 рисунка и 2 таблицы, список использованных источников из 23 наименований.

КРАТКОЕ СОДЕРЖАНИЕ

В первом разделе дипломной работы «Необходимые определения» приведены основные определения, которые взяты из [1–9] и в дальнейшем будут использованы в работе. Например, в данном разделе приводятся такие определения, как символ-джокер, вредоносная программа, компьютерный вирус, контрольная сумма.

Второй раздел «Алгоритмы поиска» состоит из трех подразделов: «Алгоритмы поиска подстроки в строке», «Точный поиск с символом-джокером» и «Поиск образца, заданного регулярным выражением». В первом подразделе приведены общие сведения об алгоритмах поиска подстроки в строке и подробно рассмотрены примитивный алгоритм [13], алгоритм Кнута – Морриса – Пратта [2], алгоритм Бойера – Мура [14] и алгоритм Рабина – Карпа [2]. Дана классификация согласно [10] и приведены алгоритмы, разделенные в соответствии с этой классификацией. Их описание можно найти в [1,11–13]. При выборе правильного алгоритма сравнения строк для конкретного приложения нужно учитывать длину образцов для поиска, дальнейшие действия с длинными образцами и текстами, вероятность нахождения совпадения с образцом, выполнение повторного поиска в одном и том же или разных текстах. [15] Во втором подразделе согласно [1] описан алгоритм точного поиска с символом-джокером. Этот алгоритм является усложнением задачи точного поиска для одного образца. Модификация заключается в введении специального символа-джокера δ, который совпадает с любым символом. В третьем подразделе приведены общие сведения о регулярных выражениях и рассмотрен алгоритм поиска образца, заданного регулярным выражением. Регулярное выражение используется для изменения строк и разбивки их на части разными способами. Это свойство реализовано в программе поиска файлов с заданными параметрами.

Третий раздел «О вредоносных программах» состоит из двух подразделов «Классификация вредоносных программ» и «О поиске вредоносных программ». В первом подразделе рассматривается классификация вредоносных программ, которая предлагается лабораторией Касперского. Подробно дается описание каждой группы вредоносных программ, некоторые группы подразделяются на классы. Также для вредоносных программ описывается их действия и способы заражения компьютера. [16–19] Во втором подразделе рассматриваются методы обнаружения вредоносных программ согласно государственному стандарту РФ ГОСТ Р 51188-98. [8] Для каждого метода описаны сильные и слабые стороны его применения. При выборе методов обнаружения вредоносных программ следует руководствоваться сведениями о сущности каждого из них, а также дополнительными пояснениями об их функциях, возможностях, достоинствах и недостатках.

Четвертый раздел «Программная реализация» содержит описание двух подпрограмм: поиска файлов на персональном компьютере и специальной подпрограммы-сканера для поиска вредоносных программ на компьютере с возможностью пополнения базы. Поиск файлов осуществляется по различным параметрам (такие как имя файла, его расширение, размер и содержимое) с помощью следующих алгоритмов поиска образца в тексте: примитивный алгоритм, алгоритм Рабина – Карпа, алгоритм Бойера – Мура, алгоритм Кнута – Морриса – Пратта. Также в этом подразделе произведены и проанализированы замеры времени работы поиска с различными параметрами и алгоритмами. Можно заметить, что в общем случае при поиске по содержимому для образцов малой и средней длины и содержащих повторения фрагментов быстрее всех работает алгоритм Кнута – Морриса – Пратта, а хуже всех – примитивный алгоритм. На длинных образцах лучше всего использовать алгоритм Бойера – Мура, но при коротких образцах у него наихудшее время работы. Если образец не содержит повторения фрагментов, то на любых длинах образца лучше всех работает алгоритм Бойера – Мура, хуже всех –

примитивный алгоритм. При малой длине имени файла быстрее всех работает алгоритм Кнута – Морриса – Пратта, хуже всех – алгоритм Бойера – Мура, при больших длинах лучше всего использовать алгоритм Кнута – Морриса – Пратта или алгоритм Бойера – Мура; расширение почти не влияет на поиск; относительно размера файлов выводы аналогичны выводам о поиске по длине имен файлов. Каждый из алгоритмов имеет свои достоинства и недостатки, их использование зависит напрямую от входных данных, поэтому каждый пользователь сам решает какой из алгоритмов использовать в конкретной ситуации. Подпрограмма-сканер имеет в своей основе алгоритм Кнута – Морриса – Пратта, наилучший алгоритм из рассмотренных согласно нашим замерам, и осуществляет поиск вредоносных программ по их контрольным суммам, размеру файла (последовательности) и названию. [20–22] Также в этом разделе приводятся для сравнения аналогичные программы поиска файлов: стандартный поиск Windows 8.1 и программа Everything. [23] Преимущество разработанной программы от описанных программ состоит в следующем: возможность выбора алгоритма поиска, что влияет на быстродействие программы, более широкий выбор параметров поиска, что позволяет более эффективно использовать программу, а также наличие специальной подпрограммы-сканера для поиска вредоносных программ на компьютере с возможностью пополнения базы.

ЗАКЛЮЧЕНИЕ

В процессе выполнения дипломной работы были изучены различные алгоритмы поиска в строках: примитивный алгоритм, алгоритм Рабина – Карпа, алгоритм Бойера – Мура, алгоритм Кнута – Морриса – Пратта, а также рассмотрены вопросы поиска вредоносных программ на компьютере. Каждый из рассмотренных алгоритмов эффективно работает в определенных классах задач, поэтому выбирать алгоритм поиска для реализации в конкретной задаче нужно только после определения точных целей и функциональности, которые она будет выполнять.

В результате работы на языке программирования C# была разработана и реализована программа, выполняющая следующие задачи: поиск файлов по различным параметрам (такие как имя файла, его расширение, размер и содержимое) с помощью следующих алгоритмов поиска образца в тексте: примитивный алгоритм, алгоритм Рабина – Карпа, алгоритм Бойера – Мура, алгоритм Кнута – Морриса – Пратта, а также поиск вредоносных программ в определенной директории.

Разработанное приложение рекомендуется использовать для поиска информации на персональном компьютере. В результате тестирования программы была сделана следующая оценка эффективности приложения: разработанная программа использует более широкий инструментарий для решения поставленных задач поиска информации, чем аналогичные программы, а также обладает специальной подпрограммой-сканером для поиска вредоносных программ на компьютере с возможностью пополнения базы.

Таким образом, все поставленные задачи были полностью решены.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1 Гасфилд, Д. Строки, деревья и последовательности в алгоритмах: Информатика и вычислительная биология [Электронный ресурс] / Д. Гасфилд. СПб. : Невский диалект, БХВ-Петербург, 2003. 654 с. Загл. с экрана. Яз. рус.

2 Кормен, Т. Алгоритмы: построение и анализ [Электронный ресурс] / Т. Кормен, Ч. Лейзерсон, Р. Риверст. М. : Изд-во «Вильямс», 2005. 1296 с. Загл. с экрана. Яз. рус.

3 Символы-джокеры [Электронный ресурс] // Википедия [Электронный ресурс] : свободная энциклопедия. URL: https://ru.wikipedia.org/wiki/Символы-джокеры (дата обращения: 30.09.2017). Загл. с экрана. Яз. рус.

4 Шаблон поиска [Электронный ресурс] // Википедия [Электронный ресурс] : свободная энциклопедия. URL: https://ru.wikipedia.org/wiki/Шаблон_поиска (дата обращения: 30.09.2017). Загл. с экрана. Яз. рус.

5 Богомолов, А. М. Алгебраические основы теории дискретных систем / А. М. Богомолов, В. Н. Салий. М. : Наука. Физматлит, 1997. 368 с.

6 ГОСТ Р 50922-2006. Защита информации. Основные термины и определения [Электронный ресурс] // Электронный фонд правовой информативно-технической документации [Электронный ресурс]. URL: https://www.niisva.su/wp-content/uploads/ 2014/09/ГОСТ_Р_50922-2006-Защита-информации.-Основные-термины-и-определения.pdf (дата обращения: 20.12.2017). Загл. с экрана. Яз. рус.

7 Климентьев, К. Е. Компьютерные вирусы и антивирусы: взгляд программиста [Электронный ресурс] / К. Е. Климентьев. М. : ДМК Пресс, 2013. 656 с.: ил. Загл. с экрана. Яз. рус.

8 ГОСТ Р 51188-98. Защита информации. Испытания программных средств на наличие компьютерных вирусов. Типовое руководство [Электронный ресурс] // docs.cntd.ru [Электронный ресурс] : электронный фонд правовой и нормативно-технической документации. URL:

http://docs.cntd.ru/document/gost-r-51188-98 (дата обращения: 23.12.2017). Загл. с экрана. Яз. рус.

9 Контрольная сумма [Электронный ресурс] // Википедия [Электронный ресурс] : свободная энциклопедия. URL: https://ru.wikipedia.org/wiki/Контрольная_сумма (дата обращения: 23.12.2017). Загл. с экрана. Яз. рус.

10 Поиск подстроки [Электронный ресурс] // Википедия [Электронный ресурс] : свободная энциклопедия. URL: https://ru.wikipedia.org/wiki/Поиск_подстроки (дата обращения: 30.10.2017). Загл. с экрана. Яз. рус.

11 Смит, Б. Методы и алгоритмы вычислений на строках [Электронный ресурс] / Б. Смит, М. : Вильямс, 2006. 496 с.

12 Aho, A. V. Efficient string matching: An aid to bibliographic search [Электронный ресурс] // A. V. Aho, M. J. Corasick. Communications of the ACM, 1975. V. 18, № 6. P. 333–340. URL: https://dl.acm.org/citation.cfm?doid=360825.360855&preflayout=flat# (дата обращения: 20.12.2017). Загл. с экрана. Яз. англ.

13 Алгоритмы поиска в строке [Электронный ресурс] // Хабрахабр [Электронный ресурс]. URL: https://habrahabr.ru/post/111449/ (дата обращения: 25.10.2017). Загл. с экрана. Яз. рус.

14 Макконнелл, Дж. Основы современных алгоритмов [Электронный ресурс] / Дж. Макконнелл. 2-е изд., доп. М. : Техносфера, 2004. 368 с. Загл. с экрана. Яз. рус.

15 Скиена, С. Алгоритмы. Руководство по разработке [Электронный ресурс] / С. Скиена. 2-е изд. СПб. : БХВ-Петербург, 2011. 720 с. (дата обращения: 23.11.2017). Загл. с экрана. Яз. рус.

16 Классификация вредоносных программ [Электронный ресурс] // Лаборатория Касперского [Электронный ресурс]. URL: https://www.kaspersky.ru/blog/klassifikaciya-vredonosnyx-programm/2200/ (дата обращения: 09.11.2017). Загл. с экрана. Яз. рус.

17 Что такое компьютерный вирус и компьютерный червь? [Электронный ресурс] // Лаборатория Касперского [Электронный ресурс]. URL: https://www.kaspersky.ru/resource-center/threats/viruses-worms (дата обращения: 09.11.2017). Загл. с экрана. Яз. рус.

18 Что такое троянская программа? [Электронный ресурс] // Лаборатория Касперского [Электронный ресурс]. URL: https://www.kaspersky.ru/resource-center/threats/trojans (дата обращения: 09.11.2017). Загл. с экрана. Яз. рус.

19 Что такое вредоносные утилиты? [Электронный ресурс] // Лаборатория Касперского [Электронный ресурс]. URL: https://www.kaspersky.ru/resource-center/threats/malicious-tools (дата обращения: 09.11.2017). Загл. с экрана. Яз. рус.

20 Баричев, С. Г. Основы современной криптографии [Электронный ресурс] / С. Г. Баричев, В. В. Гончаров, Р. Е. Серов // ИКТ [Электронный ресурс] : информационно-коммуникационные технологии в образовании. М. : Горячая линия – Телеком, 2006. 152 с. URL: http://www.ict.edu.ru/ft/002447/crypto1-3.pdf (дата обращения: 23.10.2017). Загл. с экрана. Яз. рус.

21 ClamAV Download [Электронный ресурс] // ClamAV [Электронный ресурс] . URL: http://www.clamav.net/downloads (дата обращения: 28.10.2017). Загл. с экрана. Яз. англ.

22 The Anti-Malware Testfile [Электронный ресурс] // EICAR [Электронный ресурс] . URL: http://www.eicar.org/86-0-Intended-use.html (дата обращения: 28.10.2017). Загл. с экрана. Яз. англ.

23 Everything [Электронный ресурс] // Voidtools [Электронный ресурс]. URL: https://www.voidtools.com/ru-ru/ (дата обращения: 20.12.2017). Загл. с экрана. Яз. рус.