

*П. А. Ландаков¹**

Разработка программного инструмента редактирования метаданных файлов

¹ Национальный исследовательский ядерный университет «МИФИ», г. Москва,
Российская Федерация
* e-mail: p_landakov@mail.ru

Аннотация. В статье рассматривается разработка инструмента для редактирования метаданных файлов с целью обеспечения конфиденциальности данных и соблюдения законодательных норм в области информационной безопасности. Исследуется актуальность проблемы утечек конфиденциальной информации, особенно в контексте хранения метаданных в цифровых файлах. Целью работы является обеспечение конфиденциальности информации и ограничение неавторизованного первичного сбора данных об объекте исследования путем проектирования и программной реализации инструмента редактирования метаданных файлов. В ходе исследования проведен анализ существующих решений, разработаны UML-диаграммы компонент и классов, BPMN-диаграмма работы инструмента, а также блок-схема работы инструмента. Результатом работы стало создание инструмента редактирования метаданных файлов, способного обеспечить конфиденциальность информации и соответствие законодательству. Автоматизация процесса обработки файлов позволяет сократить риски утечек информации и повысить уровень защиты данных, делая разработанный инструмент важным компонентом в обеспечении информационной безопасности.

Ключевые слова: метаданные, Exif, конфиденциальность информации

*P. A. Landakov¹**

Development of a Software Tool for Editing Metadata Files

¹ National Research Nuclear University MEPhI, Moscow, Russian Federation
* e-mail: p_landakov@mail.ru

Abstract. The article discusses the development of a tool for editing metadata files aimed at ensuring data confidentiality and compliance with legislative norms in the field of information security. The relevance of the problem of leaks of confidential information is explored, especially in the context of storing metadata in digital files. The goal of the work is to ensure the confidentiality of information and restrict unauthorized primary data collection about the research object by designing and implementing a tool for editing metadata files. During the research, an analysis of existing solutions was conducted, UML diagrams of components and classes, BPMN diagram of the tool's operation, as well as a flowchart of the tool's operation were developed. The result of the work was the creation of a tool for editing metadata files capable of ensuring the confidentiality of information and compliance with legislation. Automating the file processing process allows reducing the risks of information leaks and increasing the level of data protection, making the developed tool an important component in ensuring information security.

Keywords: metadata, Exif, confidentiality of information

Введение

В эпоху цифровой информации важность эффективного управления данными все более возрастает. Файлы нередко имеют встроенные метаданные – дополнительные сведения о содержимом, такие как данные автора, дата создания, ключевые слова и другие атрибуты. Однако редактирование и управление этими метаданными часто остаются недооцененными аспектами процесса работы с файлами.

Из отчета команды Kaspersky «Значимые утечки данных в 2023 году» следует, что за последние несколько лет 71 % утечек пользовательских данных приходится на 2023 год, а значит, наблюдается высокий рост утечек [1]. Так как в метаданных файлов может храниться конфиденциальная информация, появляется актуальность создания инструмента, который поможет очищать файлы от метаданных, тем самым защищая данные от утечек. В контексте увеличивающейся угрозы нарушения конфиденциальности разработка такого инструмента становится важным шагом в обеспечении безопасности данных пользователей.

С точки зрения государства актуальность данной проблемы подчеркивается рядом законов, включая Федеральный закон 152-ФЗ «О персональных данных» и Федеральный закон «Об информации, информационных технологиях и о защите информации» [2, 3]. Эти законы устанавливают строгие требования к хранению и обработке персональных данных, что повышает значимость разработки специализированных инструментов для обеспечения соответствия законодательству. Таким образом, разработка инструмента редактирования метаданных файлов направлена не только на защиту данных от утечек, но и на обеспечение соблюдения законодательных норм и требований в области информационной безопасности.

Таким образом, целью данной работы является обеспечение конфиденциальности информации и ограничение неавторизованного первичного сбора данных об объекте исследования путем проектирования и программной реализации инструмента редактирования метаданных файлов.

Исследование предметной области

На этапе исследования предметной области был разработан виртуальный тестовый стенд, на котором были проанализированы существующие решения. Тестовый стенд развернут на виртуальной машине VirtualBox, так как она проста в конфигурировании и является хорошим подспорьем для развертывания виртуального тестового стенда на базе операционной системы семейства Debian Linux [4, 5]. В качестве рассматриваемых решений были выбраны следующие наиболее распространенные инструменты: Exiftool, Exiv2, Mat2 [6]. На стенд были также установлены тестовые наборы данных, которые включали в себя набор файлов разных форматов, однако преимущественно были выбраны медиафайлы, так как в их метаданных зачастую хранятся конфиденциальные данные [7]. В цифровом мире формат jpeg превалирует над другими форматами изображений по частоте использования, однако для более полного тестирования были добавлены и

файлы других форматов, включая видеофайлы: png, mp4 и avi [8]. Далее представлена структура разработанного виртуального тестового стенда (рис. 1).

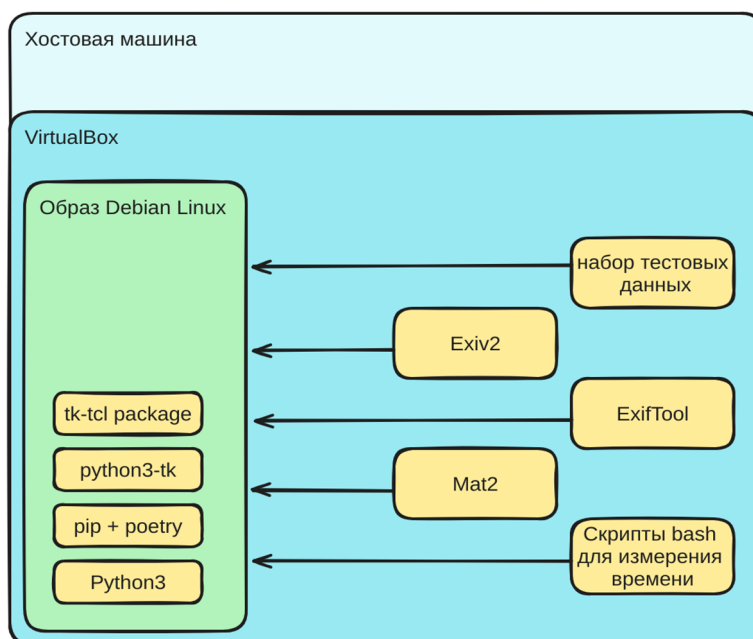


Рис. 1. Структура виртуального тестового стенда

На операционную систему были установлены все необходимые зависимости для корректной работы инструментов и тестов: tk-tcl package, python3-tk, pip, poetry, python3 [9, 10]. Для автоматизации были написаны bash-скрипты, которые замеряли время обработки тестовых данных каждым из инструментов.

После проведения тестирования были получены результаты сравнительного анализа инструментов, которые приведены в табл. 1.

Таблица 1

Сравнительный анализ существующих решений

	Время обработки 100 файлов	Поддержка форматов	Функционал	Поддержка стандартов	Возможность группового редактирования
Mat2	7400 мс	аудио, видео, фото, документы	только удаление метаданных	EXIF, IPTC, XMP	присутствует
Exif-Tool	720 мс	аудио, видео, фото, документы	удаление и редактирование	EXIF, IPTC, XMP	присутствует
Exiv2	690 мс	только изображения	удаление и редактирование	EXIF, IPTC, XMP	присутствует

Анализ полученных данных выступил в качестве отправной точки для определения ключевых критериев, которым должен удовлетворять разрабатываемый инструмент: необходимость поддержки нескольких стандартов метаданных, а именно EXIF и IPTC, так как они являются наиболее распространенными; должна быть возможность проводить не только очистку, но и добавление и редактирование метаданных, а также инструмент не должен уступать по скорости обработки файлов существующим решениям [11].

В компьютерных сетях используются сложные схемы соединения конечных компьютеров, в связи с этим используются разные механизмы коммутации и маршрутизации. Для повышения безопасности используются межсетевые экраны, DLP-системы и некоторые другие инструменты безопасности. Разрабатываемый инструмент может быть интегрирован в одну из таких систем, используя автоматизированное очищение конфиденциальных данных [12, 13].

Моделирование предметной области

В данном разделе рассматривается задача моделирования разрабатываемого инструмента, опираясь на тщательный анализ требований и сценариев использования, представленных в предыдущих разделах. Здесь внимание фокусируется на структуре и функциональности приложения, учитывая множество аспектов, начиная от пользовательского интерфейса до основных механизмов обработки метаданных. В рамках моделирования была реализована BPMN-диаграмма работы инструмента (рис. 2).

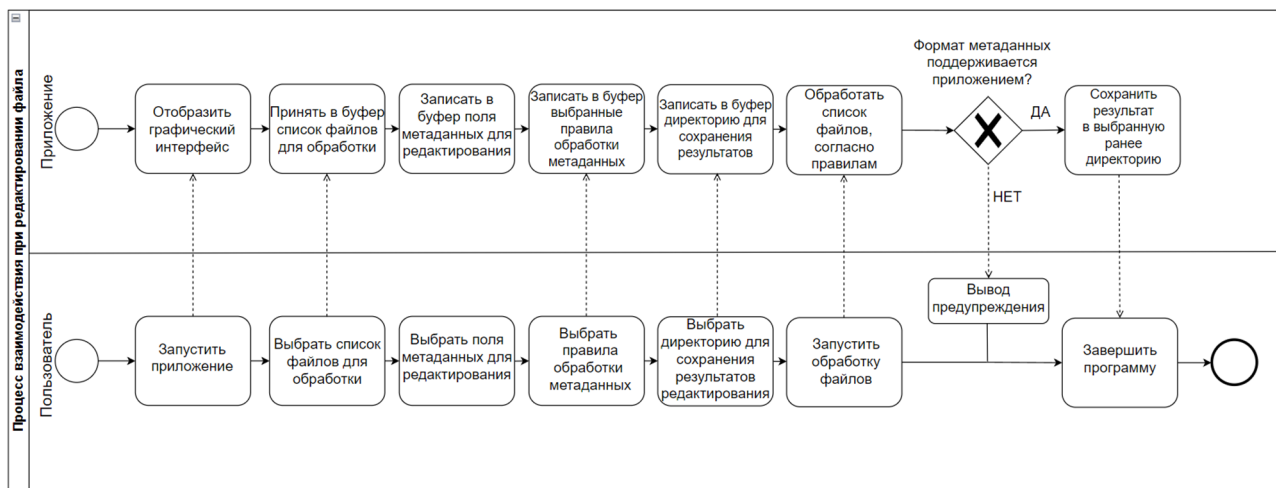


Рис. 2. BPMN-диаграмма работы инструмента редактирования метаданных файлов

Данная диаграмма представляет собой визуализацию основных этапов взаимодействия пользователя с приложением. Из диаграммы видно, что почти на каждом этапе приложение вносит полученные данные в буфер, а в конце обрабатывает записанные правила и сохраняет результат в заранее выбранную директорию, после чего работа программы завершается.

Это важное инструментальное средство позволяет детально изучить процессы работы, выявить потенциальные проблемные зоны и оптимизировать последовательность действий для повышения эффективности инструмента. Анализ данной диаграммы способствует выявлению узких мест и неэффективных этапов работы приложения, что помогает разработчикам предпринимать целенаправленные меры для улучшения его производительности и функциональности.

Проектирование предметной области

Данный этап представляет собой важную фазу, на которой определяются основные концепции, структуры данных и архитектурные решения, необходимые для создания эффективного и функционального продукта.

При проектировании стоит разобрать и учесть много составляющих работы и устройство самого инструмента. Для этой цели используем UML-диаграммы. На UML-диаграмме компонент (рис. 3) показано, какие компоненты должны быть реализованы в инструменте.

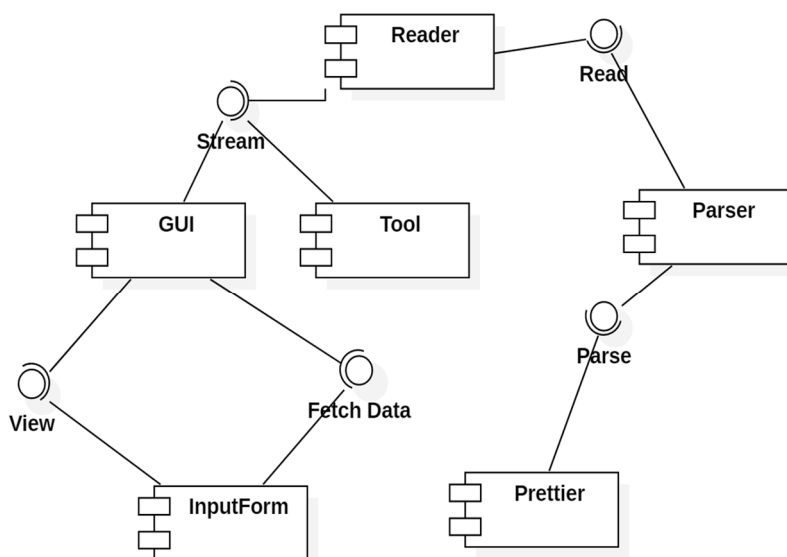


Рис. 3. UML-диаграмма компонент

Компонента reader, отвечающая за считывание данных из файла, предоставляет интерфейс компоненте parser. Компонента parser, принимая информацию от reader, проводит анализ и извлечение метаданных. Полученные данные выводятся пользователю в удобном формате с помощью компоненты prettier. В дополнение внедрены компоненты gui, tool и inputform. Взаимодействие между ними представлено на диаграмме: inputform обеспечивает интерфейсы view и fetch data для gui, а gui в свою очередь служит интерфейсом для tool и reader.

Не менее важной частью проектирования является разработка UML-диаграммы классов (рис. 4). Она помогает разработчикам и аналитикам лучше понять, как классы взаимодействуют друг с другом, какие атрибуты и методы они имеют, и как они связаны с другими классами.

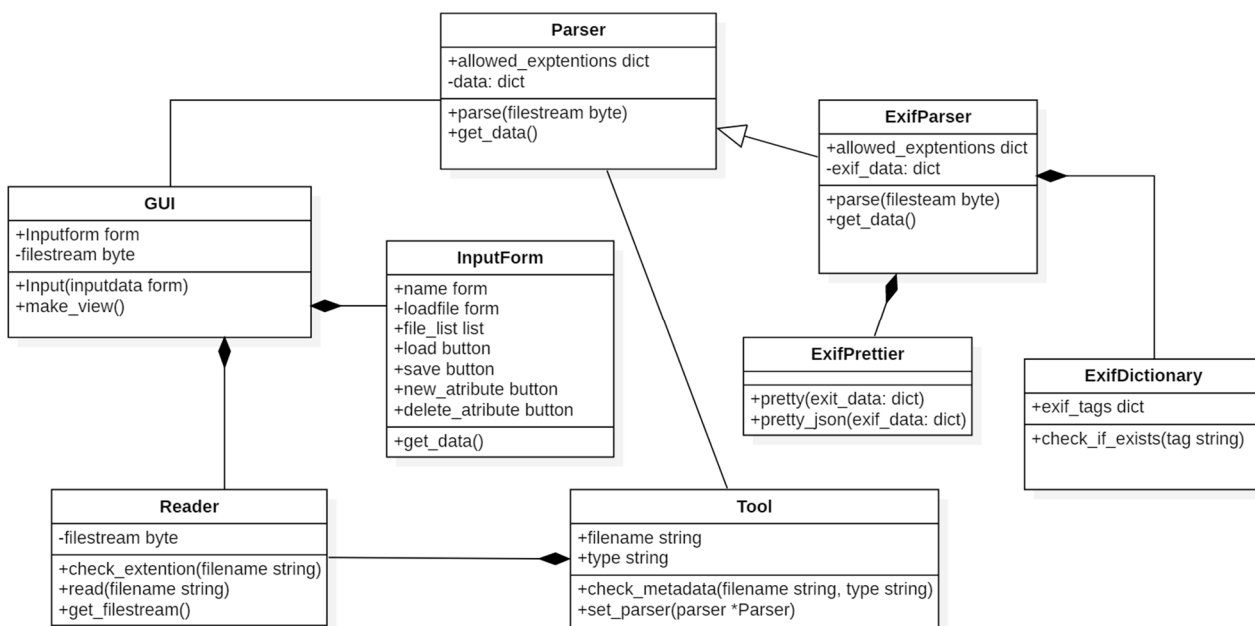


Рис. 4. UML-диаграмма классов

Созданные классы, такие как Reader, Tool, Parser, ExifParser и ExifPrettier, основаны на проектировании компонентной диаграммы и модели системы. Reader отвечает за считывание файла и передачу информации в Parser, описанный внутри класса Tool. Parser обрабатывает информацию и извлекает метаданные с помощью методов parse и get_data, а ExifParser предоставляет дополнительные возможности. Полученные данные обрабатываются и форматируются классом ExifPrettier для удобства использования.

Разработка

Используя все данные, полученные из этапов исследования, проектирования и моделирования, была реализована блок-схема работы инструмента (рис. 5), где описывается алгоритм обработки файлов.

После запуска происходит выбор файлов для обработки через графический интерфейс. Если ранее не загружались другие файлы, создается список медиафайлов для обработки и переход на следующий этап. Затем пользователь добавляет правила обработки. После этого выбирается место для сохранения результата обработки. Если все правила обработаны, файлы сохраняются в указанную директорию, и программа завершает работу. Если есть невыполненные правила, происходит их последовательная проверка и действия в зависимости от условий.

С целью оптимизации процесса разработки и развертывания кода, следует использовать методологию DevSecOps, которая призвана обеспечить безопасность на всех этапах жизненного цикла разработки ПО и является частью более глобальной методологии DevOps, а также вносить автоматизацию в процессы развертки кода [14, 15].

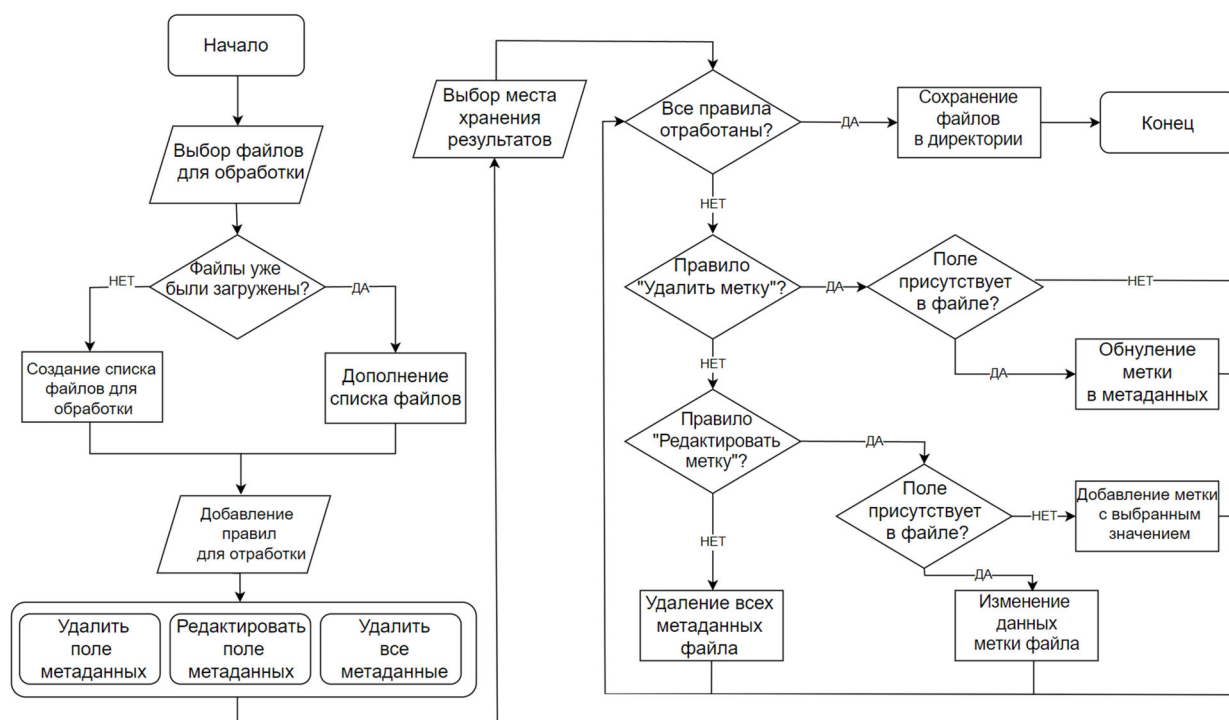


Рис. 5. Блок-схема работы инструмента редактирования метаданных файлов

Заключение

Роль метаданных в цифровой жизни человека высока, так как метаданные могут хранить в себе конфиденциальные данные. Данный факт обеспечивает актуальность и важность разработки специального инструмента, позволяющего обеспечивать конфиденциальность информации путем редактирования метаданных файлов.

В рамках исследования предметной области был создан виртуальный тестовый стенд, где были протестированы существующие решения. На основании результатов тестирования была реализована таблица сравнительного анализа существующих решений. В ходе моделирования предметной области была реализована BPMN-диаграмма работы инструмента редактирования метаданных файлов. На этапе моделирования были реализованы UML-диаграмма компонент и UML-диаграмма классов. На основе полученных данных была разработана блок-схема работы инструмента редактирования метаданных файлов.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Аналитические отчеты Лаборатории Касперского «О значимых утечках данных в 2023 году» // Kaspersky-labs. – URL: <https://content.kaspersky-labs.com/se/media/ru/data-leaks-report-2023.pdf>
2. Борисенко О. В. Анализ федерального закона № 152-ФЗ «о персональных данных» // Электронное приложение к Российскому юридическому журналу. – 2012. – №. 2. – С. 26-30.
3. Инюшкин А. А. Особенности применения Федерального закона от 27. 07. 2006 № 149-ФЗ «Об информации, информационных технологиях и о защите информации» к правоотношениям, связанным с программами для ЭВМ, базами данных и сайтами в сети Интернет // Общество: политика, экономика, право. – 2017. – №. 12. – С. 136-139.

4. Oracle VM VirtualBox User Manual // VirtualBox. – URL: <https://download.virtualbox.org/virtualbox/UserManual.pdf>
5. Басыня Е.А. Системное администрирование и информационная безопасность: Учебное пособие. – Новосибирск: Изд-во НГТУ, 2018.
6. Toevs V. Processing of metadata on multimedia using exiftool: A programming approach in python // 2015 Annual Global Online Conference on Information and Computer Technology (GOCICT). – IEEE, 2015. – С. 26-30.
7. Басыня Е. А., Сафронов А. В. Разработка и исследование системы управления метаданными изображений // Актуальные проблемы электронного приборостроения АПЭП-2016. – 2016. – С. 155-157.
8. Fernandez J. J., Pandian N. Jpeg metadata: A complete study // 2018 international conference on recent trends in advance computing (ICRTAC). – IEEE, 2018. – С. 34-40.
9. Г. Россум, Ф.Л.Дж. Дрейк, Д.С. Откидач, М. Задка, М. Левис, С. Монтаро, Э.С. Реймонд, А.М. Кучлинг, М.-А. Лембург, К.-П. Йи, Д. Ксиллаг, Х.Г. Петрилли, Б.А. Варсав, Дж.К. Ахлстром, Дж. Роскинд, Н.Шеменор, С. Мулендер. Язык программирования Python. / 2001 – 454 с.
10. Pip documentation v24.0 - User Guide. – URL: https://pip.pypa.io/en/stable/user_guide/
11. Mana S. C., Sasiprabha T. A study on various semantic metadata standards to improve data usability // 2019 International Conference on Computational Intelligence in Data Science (ICCIDS). – IEEE, 2019. – С. 1-4.
12. Олифер Виктор, Олифер Наталья. Компьютерные сети. Принципы, технологии, протоколы: Юбилейное издание, 2020. – С. 1008.
13. Таненбаум Эндрю, Фимстер Ник, Уэзеролл Дэвид. Компьютерные сети. 6-е изд. – Санкт-Петербург : Питер, 2023. – 992 с
14. Басыня Е. А. Сетевая информационная безопасность: Учебник. – Москва : НИЯУ МИФИ, 2023. – 224 с.
15. Басыня, Е. А. Автоматизированная установка и конфигурирование серверных решений / Е. А. Басыня, М. С. Лукина // Современные материалы, техника и технологии. – 2016. – No 2(5). – С. 21-26.

© П. А. Ландаков, 2024