

Сканирование бумажной книги

Введение

Предположим, что Вы хотите отсканировать бумажную книгу, и уже [купили сканер](#) для этого. Каковы Ваши дальнейшие действия?

Ответ на этот вопрос зависит от того, в какой операционной системе Вы будете сканировать - Windows, Linux или Mac OS?

Под Linux сканирование производится в программе [SANE](#).

В этой статье мы рассмотрим наиболее популярный вариант - сканирование под Windows.

Теоретические основы

Перед тем, как начинать сканирование, Вам необходимо освоить следующие теоретические понятия:

1. Растровая графика

Любой скан - это растровый графический файл - или, иными словами, прямоугольник, наполненный рядами мельчайших квадратных цветных точек - *пикселями*.

Подробнее: [Растровая графика в Википедии](#)

2. Форматы графических файлов.

Файлы растровых изображений (которые мы будем использовать для сохранения сканов) могут быть разных форматов. Наиболее популярные форматы - это TIF, BMP, JPG, GIF, PNG. Для нас (в данном случае) важно лишь то, что только форматы TIF и BMP обеспечивают беспотерное сохранение графической информации - поэтому мы будем использовать только их (другие форматы используют "сохранение с потерями" - поэтому они нам не подходят).

Подробнее: [Графические форматы в Википедии](#)

3. Разрешение (DPI)

Чем мельче пиксели изображения - тем детальней будет наш скан. Разрешение - это и есть степень мелкости пикселей изображения, измеряемая в "DPI" (dots-per-inch - точек на дюйм) - то есть, берём один дюйм длины (или ширины) скана, и смотрим, сколько там уместилось пикселей - получаем величину "DPI".

Подробнее: [Что такое DPI и ресемплинг](#)

Подробнее: [DPI в Википедии](#)

4. Глубина цвета

Глубина цвета показывает, сколько битов в графическом файле занимает каждый пиксель изображения (скана). Для хранения чёрно-белого пикселя достаточно 1 бита, для хранения серого пикселя (256 возможных оттенков серого) нужно 8 битов, для хранения цветного пикселя (16,7 млн. возможных оттенков) нужно 24 бита.

Подробнее: [Глубина цвета в Википедии](#)

Варианты глубины цвета, встречающиеся в данной статье: "Цветной 24 бита", "Серый 8 бит", "Чёрно-белый".

Выбор программы сканирования

Для работы со сканером под Windows Вам, скорее всего, понадобится установить "программу для сканирования" (одну или две - см. далее).

"Скорее всего" - потому, что возможен вариант, что никаких программ устанавливать не потребуется - если Ваш сканер поддерживается WIA (см. далее), и Вас устраивает такой вариант.

В любом случае, установка программ сканирования обеспечит удобство работы со сканером - поэтому ставить такие программы рекомендуется всегда.

Чтобы разобраться с этими программами, рассмотрим схему работы любого сканера. Как сканер взаимодействует с компьютером, при помощи каких программ?

В любом случае, под Windows компьютер использует **две** (а не одну) программы для работы со сканером:

1. Драйвер сканера.
2. Программа сканирования.

Драйвер сканера

Драйвер сканера - это техническая программа низкого уровня, которая позволяет компьютеру "понять" конкретный сканер. Все сканеры немного отличаются друг от друга по внутреннему "железному" устройству - поэтому и драйвера у них разные. Драйвер сканера - это то, что "превращает" сканер из куска мёртвого металла в понятный данному компьютеру логический объект, с которым можно общаться при помощи простейших программных команд. Драйвер - это своего рода "программная прокладка" между любым сканером и компьютером, это "программная прошивка" для сканера.

Подробнее: [Драйвер в Википедии](#)

Драйвера для сканеров под Windows бывают двух типов:

- [TWAIN](#). Это старейший стандарт, он же и наиболее популярный, привычный и наиболее широко распространённый. Подавляющее большинство сканеров имеют TWAIN-драйвер на компакт-диске в комплекте поставки.

- [WIA](#). Это новомодный стандарт, он появился начиная с Windows XP. Сканеры более старых моделей его не поддерживают. WIA встроена в Windows XP (и старше), поэтому, если Ваш сканер поддерживается им, то ставить WIA-драйвер с компакт-диска, прилагаемого в комплекте сканера, не придётся - сканер определится автоматически при его подключению к компьютеру.

Примечание: также бывают встроенные драйвера - не относящиеся ни к одному из этих двух типов - см. ниже "Программа сканирования со встроенным драйвером".

Драйвер используется для сканера всегда - просто иногда он встроен либо в операционную систему (как WIA), либо в сканирующую программу (SANE в LINUX, VueScan под Windows).

Как правило, драйвер для сканера прилагается на лазерном диске, идущем в комплекте покупки сканера. Поэтому при подключении купленного сканера первое, что Вам нужно сделать - установить с диска драйвер сканера.

ВАЖНО: Рекомендуется всегда устанавливать драйвер сканера с прилагаемого компакт-диска.

Однако, драйвер для сканера бывает в наличии не всегда. Что делать в этом случае? Первое - нужно попробовать поискать драйвер на данный сканер в Интернете. Если же окажется, что для данного сканера по какой-то причине нет драйвера (обычно такого не бывает), то можно попробовать применить программу из раздела "Программа сканирования со встроенным драйвером" (см. далее).

В этой статье мы рассмотрим использование TWAIN-драйверов - как самый простой и популярный случай.

Программа сканирования

Программа сканирования - это та компьютерная программа, при помощи которой пользователь работает со сканером. Эта программа на самом деле взаимодействует с драйвером сканера - а не напрямую со сканером. То есть, если убрать драйвер сканера - то программа сканирования уже ничего не сможет сделать со сканером. Но и с одним только драйвером сканера (без программы сканирования) Вам поработать не удастся - слишком уж драйвер низкоуровневый.

Программа сканирования транслирует команды высокого уровня (типа "отсканировать страницу") в набор команд низкого уровня, непосредственно понятных драйверу сканера ("включить двигатель каретки", "протянуть каретку до конца стекла" и т.п.).

Программ сканирования существует великое множество. Все они отличаются друг от друга в основном степенью удобства и комфорта использования.

Программа сканирования обычно устанавливается после установки драйвера (и она должна "увидеть" установленный драйвер), хотя порядок их взаимной установки не принципиален.

Для тех, кто работает в операционной системе Windows - лучшим выбором для поточного сканирования является программа [IrfanView](#) (согласно коллективному мнению сообщества книгосканировщиков). Она свободно-бесплатна для коммерческого использования, и она удобна для массового сканирования страниц книг. Единственный её недостаток - она не умеет автоматически запускать очередной акт сканирования через заданные промежутки времени, т.е. лишена функции автосканирования.

Функция автосканирования есть в такой программе, как [VueScan](#). Это широко известная коммерческая программа для сканирования. Также функция автосканирования есть в некоторых других программах - ABBYY FineReader, [AutoScan](#) (673 КБ), [PaperCapture](#) (667 КБ). См. [подробнее](#).

Если Вы хотите использовать ABBYY FineReader как средство сканирования - то нужно помнить следующее:

Не используйте программу ABBYY FineReader версий 7, 8 или ниже для сканирования книг. Она корёжит получаемые скан-изображения страниц книги. Можно использовать программу ABBYY FineReader 9 (или старше) для этого - но перед сканированием сбросьте там в опциях галку "*Исправить перекос изображения*".

ВАЖНО: Не используйте программу ABBYY FineReader версий 7, 8 или ниже для сканирования книг.

Программа сканирования со встроенным драйвером

Некоторые программы сканирования уже содержат встроенные драйвера для целого набора сканеров. К ним относятся, например, [VueScan](#) и [SilverFast](#). Есть и иные аналогичные программы. Встроенные в них драйвера не относятся ни к типу TWAIN, ни к типу WIA. Все такие известные программы являются коммерческими. Их можно считать исключением из общей схемы сканирования под Windows.

Описание процесса сканирования

После того, как Вы установили драйвер сканера, и поставили сканирующую программу, можно приступать непосредственно к сканированию книги.

Перед началом сканирования обеспечьте на своём жёстком диске свободное пространство порядка 2 Гигабайт - для будущих сканов.

Процесс сканирования очень прост: берёте бумажную книгу, открываете её на первой странице, кладёте этой страницей на стекло сканера, нажимаете кнопку "Сканировать" и ждёте 20-30 секунд, пока сканер не отсканирует страницу. Затем снимаете книгу со стекла сканера, переворачиваете страницу на следующую, снова кладёте на книгу так же на стекло сканера (новой страницей), и снова нажимаете кнопку "Сканировать". И так до конца книги.

Вот и всё - это и есть "сканирование книги".

К сожалению, сканирование книги - это весьма трудоёмкий и длительный процесс. Многим (поначалу) он даже кажется неподъёмным по количеству труда. Зато это очень простая деятельность - не требующая никакой квалификации. Проще говоря - это совершенно "обезьянья" бездумная работа, выполнять которую сможет любой человек.

В результате сканирования Вы получите сканы книги.

Вот как они примерно выглядят:

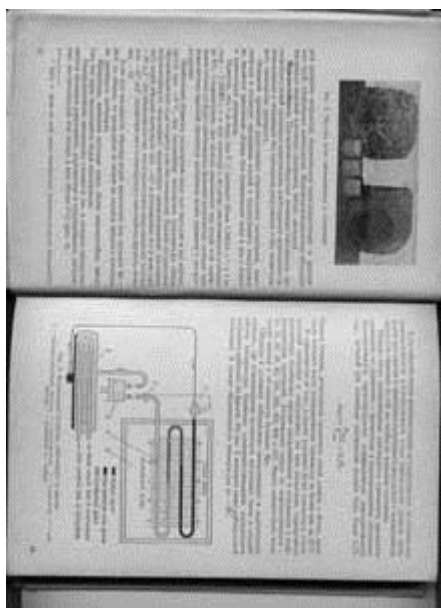


Рис. 1. Сдвоенный разворот - в исходном повернутом на бок виде

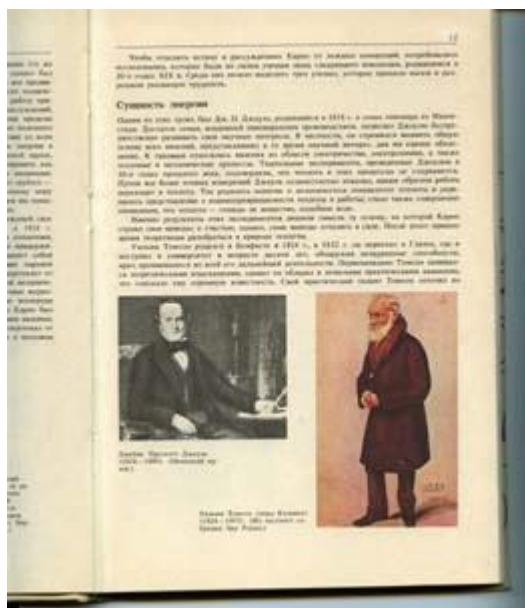


Рис. 2. Одиночный скан - с кусочком соседней страницы (крупноформатная книга)

Такие же на вид сканы должны получиться и у Вас. Сканы часто называются "сырыми" - потому что в дальнейшем они проходят облагораживающую обработку.

ВАЖНО: Запомните термин "сырые сканы".

Сканы - это "сырьё" для создания будущей электронной книги (в формате DjVu).

Сканирование при помощи Irfan View

Рассмотрим в деталях процесс сканирования на примере использования свободно-бесплатной программы Irfan View.

Если в комплекте со сканером имеется компакт-диск (скорее всего) - то первое, что Вам нужно сделать - установить с этого диска драйвер Вашего сканера.

Теперь нам нужно настроить Irfan View на работу со сканером (взаимодействие с драйвером сканера).

Открываем Irfan View. В меню **Файл** выбираем пункт **Выбрать TWAIN-источник...** . Появляется окно "Выбор источника":

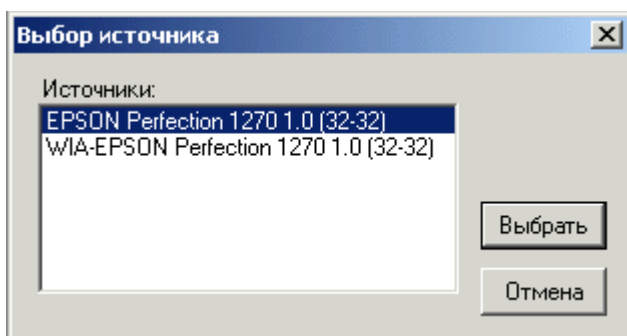


Рис. 5. Окно выбора драйвера сканера

Строки в этом окне у всех будут разными - в зависимости от сканера. В данном случае сканер даёт выбор из 2-х своих драйверов - TWAIN-драйвер и WIA-драйвер. Выбираем верхний (TWAIN-драйвер) - как наиболее привычный. Можно выбрать и 2-ой - это кому как нравится. Закрываем окно "Выбор источника".

Теперь приступаем к самому сканированию. В меню **Файл** выбираем пункт **Получить изображение/пакетное сканирование...** . Появляется окно "Настройка пакетного сканирования":

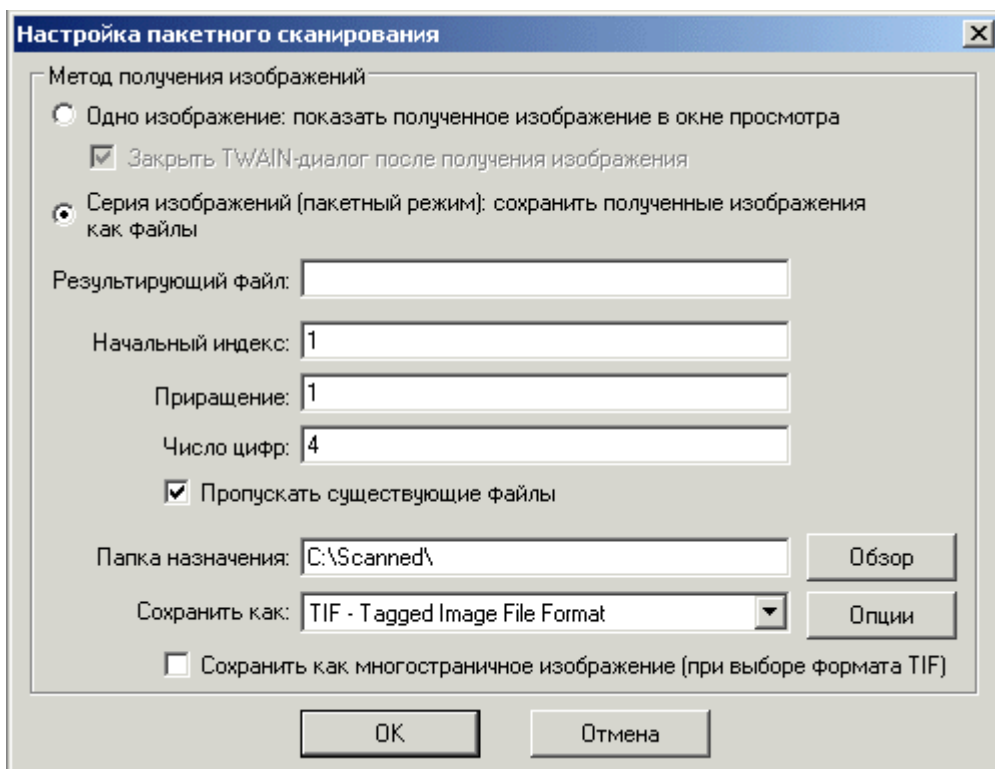


Рис. 6. Окно настройки пакетного сканирования

Выставляем в этом окне все настройки так, как показано на рисунке. Папку назначения можете выбрать свою. В поле "Сохранить как" можно указать формат **ВМР** вместо **TIF**. Никакие иные форматы сохранения сканов указывать **не допускается**.

ВАЖНО: Не сканируйте в формат **JPG**. Формат **JPG** "размывает" скан-изображения страниц книги.

Далее нажимаем на кнопку **Опции**. Появляется окно "Параметры сохранения TIFF":

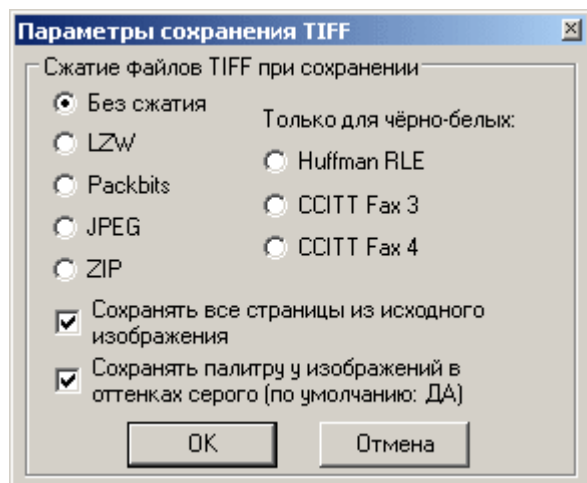


Рис. 7. Окно параметров сохранения TIFF

Выставляем в этом окне все настройки так, как показано на рисунке. Вместо опции "Без сжатия" допускается указать LZW - но это (предположительно) может увеличить время сканирования (зато вдвое уменьшит размер дискового пространства, необходимого для сохранения создаваемых сканов). Закрываем окно "Параметры сохранения TIFF".

Переходим непосредственно к процессу сканирования. В окне "Настройка пакетного сканирования" нажимаем кнопку ОК. Запускается драйвер сканера и загружается окно сканирования. В зависимости от модели сканера и версии его драйвера, это окно будет иметь различный вид. У меня оно выглядит так (сканер Epson Perfection 1270):

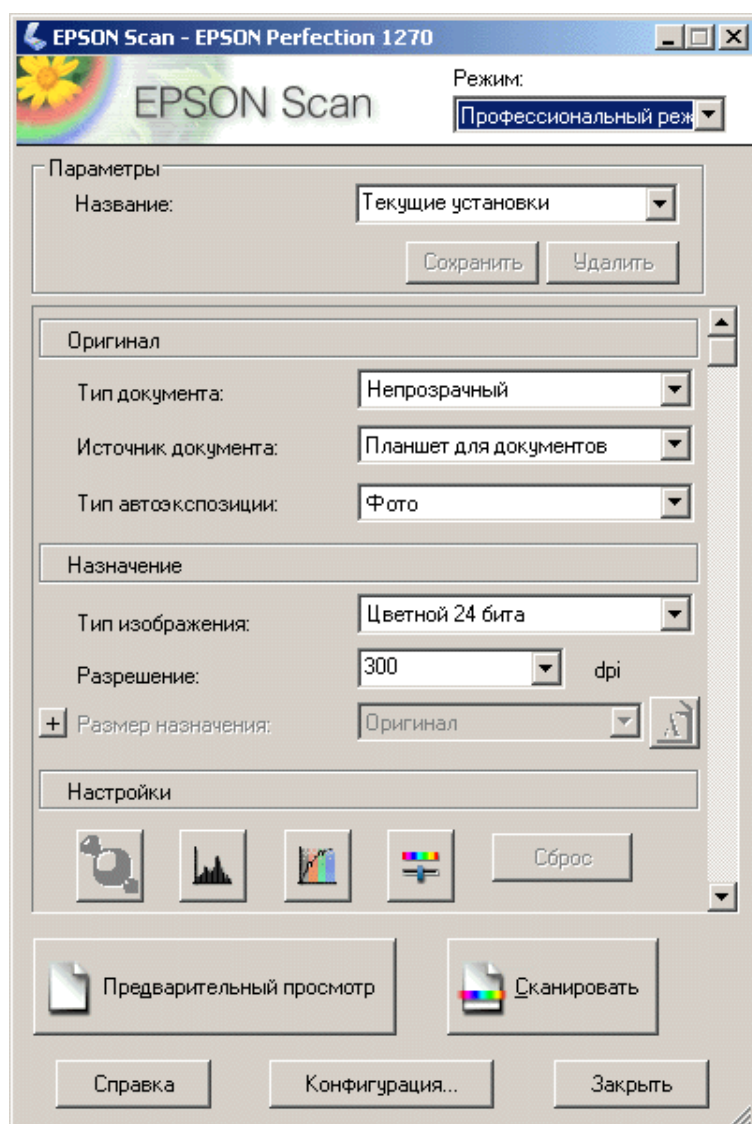


Рис. 8. Окно драйвера сканирования сканера Epson Perfection 1270

У Вас это окно будет выглядеть как-то по-другому - но аналогично.

Единственные параметры, которые нам потребуется выставить в этом окне - это **Тип изображения** и **Разрешение**. Тип изображения - это та глубина цвета, с которой сканер будет сканировать страницы нашей бумажной книги.

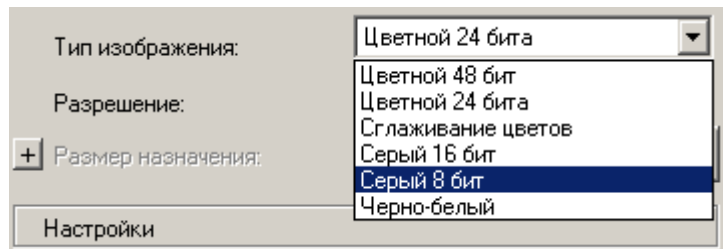


Рис. 9. Возможные значения глубины резкости

Для подавляющего большинства книг это значение нам следует установить в "**Серый 8 бит**". Такое значение выставляется для всех таких бумажных книг, в которых отсутствуют цветные элементы на страницах - а есть только лишь чёрно-белый текст и (может быть) чёрно-белые (серые) иллюстрации. Под цветными элементами понимается всё что угодно из цветного: цветной текст, цветные иллюстрации, цветной орнамент и т.п.

Если же в бумажной книге **есть** цветные объекты - тогда глубину цвета мы выставляем как "**Цветной 24 бита**" - чтобы не потерять при сканировании информацию о цвете.

Если Вы сомневаетесь, какой режим выбрать "**Цветной 24 бита**" или "**Серый 8 бит**" - можете *всегда*, для *любой* книги, выставлять режим "**Цветной 24 бита**". В дальнейшем, уже на этапе сканобработки, можно будет для каждого отдельно взятого скана понизить глубину цвета с 24 бит до любого иного значения. Скорость же сканирования одинакова - для обоих режимов - "**Цветной 24 бита**" и "**Серый 8 бит**". Разница лишь в том, что при режиме "**Цветной 24 бита**" полученные сканы будут занимать значительно **большее** место на Вашем жёстком диске.

Некоторые пользователи спрашивают: "Зачем нужно сканировать в режиме серого - всё равно ведь книга потом приводится к чёрно-белому режиму"?

Ответ прост: режим серого позволяет сохранить на получаемом скане ту *полезную избыточность* информации, которая при дальнейшей сканобработке даст простор для всевозможных видов улучшающей обработки данного скана.

Аналогичный пример из жизни: на важных мероприятиях любой фотограф старается сделать как можно больше снимков - чтобы потом отобрать из общей массы самые удачные.

ВАЖНО: Не сканируйте в чёрно-белом режиме! Только в сером или цветном! Это принципиально важно - для качества последующей сканобработки.

Параметр **Разрешение** для *подавляющего* большинства книг следует установить в 300 (dpi). В отдельных редчайших случаях можно поставить 600 dpi - либо когда требуется высокое качество, либо когда полиграфическое качество бумажной книги крайне низко.

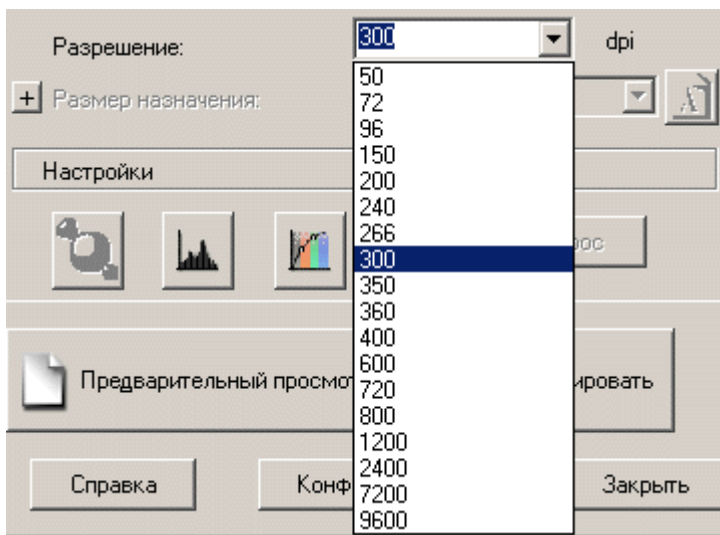


Рис. 10. Возможные значения разрешения

ВАЖНО: Не сканируйте в разрешении ниже 300 dpi! Слишком низкое разрешение при сканировании превратит Ваш труд в полную бессмыслицу - потому что читать такую электронную книгу будет крайне неудобно.

Если размер сканируемой книги заметно меньше, чем размер стекла сканера, то тогда для ускорения процесса сканирования можно применить такой приём, как уменьшение сканируемой площади:

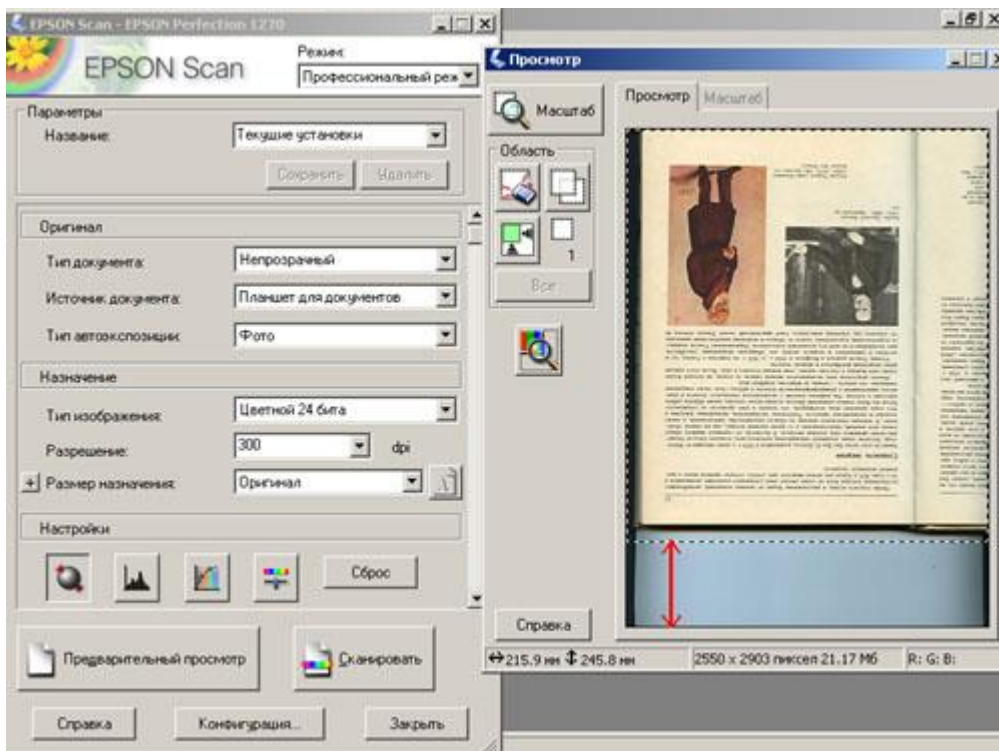


Рис. 11. Уменьшение сканируемой площади

Для этого нужно запустить процедуру "Предварительный просмотр" (нажатием одноимённой кнопки). Сканер выполнит предварительное сканирование, и в появившееся окно предварительного просмотра загрузится изображение текущего скана. В этом окне нужно мышкой изменить размер мерцающего пунктирного прямоугольника - так, чтобы он лишь немного превосходил по размеру размер скана. На рис. 11 красной стрелкой показано, на сколько удалось уменьшить сканируемую площадь для данного скана.

Последующие сканы при этом нужно класть на стекло сканера точно так же, как и контрольный скан (на котором мы сделали предварительный просмотр).

Этот приём даёт заметное сокращение времени сканирования.

Ход процедуры сканирования

Сама по себе техника сканирования исключительно проста. Кладём книгу разворотом на стекло сканера, прижимаем сверху грузом (чтобы разворот как можно плотнее прилегал к стеклу сканера - но так, чтобы не выдавить стекло сканера), нажимаем кнопку сканирования (в моём случае - это клавиша "Пробел" на клавиатуре или кнопка "Сканировать" в окне сканера), ждём, пока текущая страница отсканируется, и повторяем процесс заново - пока не отсканируем всю книгу.

Большинство книг помещаются на стекле сканера в виде разворота - т.е. 2 соседние страницы одновременно (см. Рис.1). При этом изображение страниц оказывается повёрнутым набок на 90 градусов - это не страшно, пусть это Вас не смущает. Это будет исправлено на этапе сканобработки.

Однако, бывают книги большого формата - которые не помещаются разворотом на стекле сканера. Такие книги приходится сканировать по одной странице (см. Рис.2). При этом на скан попадает кусочек (полоска) соседней страницы - на это тоже не обращайте внимания. Это также будет исправлено на этапе сканобработки.

В процессе сканирования можно даже выключить монитор (и слушать музыку или смотреть фильм по телевизору) - запуская каждое новое сканирование клавишей "Пробел" на клавиатуре. Это возможно потому, что как только Вы хотя бы один раз нажали кнопку "Сканировать" в окне сканера, программа будет сохранять фокус выделения на этой кнопке - и она будет нажиматься даже от нажатия клавиши "Пробел" на клавиатуре.

Сканирование не требует большого старания. Всё, о чём нужно позаботиться при сканировании - это о том, чтобы книга была достаточно плотно прижата к стеклу сканера (во время сканирования). Остальное неважно. Для плотного прижатия рекомендуется класть на книгу (лежащую на сканере) какой-нибудь увесистый предмет - например, любую тяжёлую книгу (причём класть его всякий раз, когда Вы сканируете очередную страницу).

Отсканированные сканы будут автоматически сохраняться в назначенной папке - в нашем случае это "C:\Scanned". Будьте готовы к тому, что сканы займут на жёстком диске значительное пространство - от 2 до 4 Гигабайт. Это не страшно - при последующей сканобработке размер занимаемого места резко сократится, но на этапе сканирования Вам нужно обеспечить наличие достаточного места на жёстком диске.

Завершение сканирования

Вам не обязательно сразу же сканировать всю книгу. Это можно сделать порциями - за 2-3 вечера, к примеру. Сканировать удобно поздно вечером перед сном - когда усталость не позволяет заниматься каким-либо более интеллектуальным делом.

После того, как Вы отсканируете всю книгу, Вам нужно будет обязательно СРАЗУ ЖЕ просмотреть полученные сканы - и проверить их на предмет следующего:

1. Не было ли заломов уголков страниц при сканировании?
2. Не было ли недостаточно сильно прижатых к стеклу сканера разворотов - таких, у которых буквы в районе корешка книги сильно исказились?
3. Проверьте КОМПЛЕКТНОСТЬ отсканированных страниц. Очень часто многие люди случайно пропускают при сканировании (или дважды сканируют) некоторые страницы.

В случае обнаружения проблемных страниц их нужно сразу же пересканировать заново - и заменить дефектные сканы исправленными.

Напоследок отсканируйте цветную обложку книги (если есть) в режиме "**Цветной 24 бита**". Обложка - это лицо книги, поэтому не стоит забывать о ней.

Распространённые ошибки сканирования

1. Сканирование в чёрно-белом режиме. Это грубейшая ошибка, которая сводит на нет все возможности последующей сканобработки.
2. Сканируют на низком разрешении - менее 300 dpi - ради скорости и снижения размера (особенно, если делают PDF). Это также грубейшая (преступная) ошибка.
3. Сканирование в формат JPG. Это приводит к "размыванию" сканов. Сканировать надо в форматы TIF или BMP.
4. Сканирование в формат PDF. На некоторых МФУ сохранение в PDF реализовано криво - что потом приводит к проблемам при декодировании PDF -> TIF (для последующей сканобработки).
5. Сканирование посредством ABBYY FineReader 8 или ниже. Это "корёжит" строки текста на сканах.
6. Некоторые стараются при сканировании уложить книгу на стекло сканера максимально параллельно краям сканера. Это абсолютное излишество - угловой перекося скана устраняется автоматически при последующей сканобработке.
7. Забывают проверить комплектность сканов страниц после сканирования (пропуски страниц на сканах).
8. Пытаются сканировать книги на CIS-сканере (из-за малой глубины резкости CIS-сканеров изображение "расплывается" у корешка книги).
9. Игнорируют сканирование обложки книги.

Заключение

Вот и всё - по завершению сканирования бумажная книга сразу же становится совершенно не нужной (для последующего процесса "электронизации"). Поэтому, если Вы взяли бумажную книгу у друга или в библиотеке на 1-2 дня - можете её возвращать.

Я также рекомендую после сканирования отдохнуть 1-2 дня - и только после этого приступать к дальнейшей обработке сканов.