

Как отсканировать отчет

Выбор сканера

Сейчас в продаже появилось очень много дешевых сканеров, однако не любой аппарат сгодится для сканирования отчетов. Все бытовые сканеры можно разделить по типу сканирующего сенсора на два типа. С сенсором CIS и сенсором CCD.

Сканеры с сенсором CIS появились в продаже не так давно. Они легкие, тонки и очень дешевые. Можно найти сканеры ценой меньше тысячи рублей. Однако с ними не так все хорошо. Несмотря на общую привлекательность сканирующая головка таких сканеров имеет очень маленькую глубину резкости. Т.е. текст со стороны переплета, где отчет плохо прижат к стеклу сканера будет нечитабелен. Такие сканеры покупать не следует.

Сканеры с сенсором CCD производятся уже довольно давно. Раньше они были с галогеновой подсветкой, сейчас - со светодиодной. Такие сканеры намного толще моделей с CIS сенсором, потому что в сканирующую головку встроена оптическая система. Однако Глубина резкости этих сенсоров позволяет сканировать даже очень толстые отчеты.

Еще следует обратить внимание на наличие на сканере кнопок и возможность запрограммировать сканер таким образом, чтобы при нажатии на кнопку он автоматически сканировал страницу и сохранял в указанную папку. Это сэкономит лишние телодвижения при больших объемах сканирования.

Крышка сканера должна откидываться на 180 градусов, либо вообще сниматься.



Некоторые модели сканеров специально позиционируются для сканирования книг. Их стекло вплотную подходит к краю поверхности, чтобы книгу можно было приложить как можно плотнее к стеклу. Например, сканеры серии optiscbook фирмы plustek. Однако такие сканеры дороже.

Я купил себе самый дешевый сканер Epson V33 за пару тысяч рублей и вполне им доволен. Следует отметить, что у этой модели через пару месяцев службы появляются тонкие цветные полосы на сканированном изображении, но сканированию текста и фотографий из отчетов это не сильно мешает.

Отдельно следует упомянуть сообщество diybookscanner.org Эти ребята из подручных средств собирают отличные сканеры для оцифровки книг, не уступающие фирменным. Очень интересно ознакомиться вот с этим видео

<http://www.youtube.com/watch?v=a4-qMc2QSOw&feature=g-like&context=G2eb41c9ALT2coMgAAAA>

Сканирование

При сканировании надо найти компромисс между качеством и скоростью. Люди, увлекающиеся качественно оцифровкой книг рекомендуют сканировать отдельно текст, отдельно рисунки и отдельно фотографии. При этом разрешения сканирования следующие: Текст (даже с примечаниями мелким шрифтом) - 400 dpi. Рисунки, в случае, когда сканируются отдельно от текста - минимум 300 dpi. Географические карты и фотографии - 300-600 dpi.

Это рекомендуют профессионалы, я же вполне довольствуюсь сканированием всего в разрешении 300 dpi в формат tiff. Это занимает сильно больше места чем JPG, но зато в этом формате не происходит потеря информации об изображении. Например, один лист A4 в tiff занимает 25 мегабайт. Естественно, такие изображения надо потом обработать и сконвертировать для уменьшения объема.

Обработка

Здесь можно пойти двумя путями. Первый - самый простой путь. Пережать все сканы в JPG, установив степень сжатия на уровне 60-80%, проверить, что текст и фотографии остались читабельны и на этом остановиться. Для массовой конвертации файлов хорошо подходят бесплатные программы FastStone image viewer или IrfanView.

Второй путь - обработать сканы, чтобы они выглядели более опрятно и занимали меньше места. Существует несколько программ для обработки сканов - это ScanTailor, ScanKromsator и ArtScan. ScanKromsator очень сложный, навороченный и тормозной. Я при обработке использую ScanTailor. Это очень простая и доступная программа для полуавтоматической обработки сканов. Обработка страниц состоит из нескольких шагов:

1. Разворот листов
2. Обрезка полей и корешков
3. Коррекция угла наклона
4. Выделение полезной области
5. Установка полей
6. Вывод

Подробнее следует рассказать про вывод. Программа умеет выводить только в формат tiff, однако тифф тиффу рознь. Существует три режима:

1. Черно-белое изображение. Картинка переводится всего в два цвета. Черный и белый. Этот режим очень хорошо подходит для страниц с текстом, черно белых схем, кроков. При этом разрешение вывода следует поставить в 600 dpi даже если вы сканировали текст в 300 dpi. Тем самым потерю информации в цвете вы компенсируете увеличением разрешения. Если при переводе в черно-белый режим мелкий текст становится нечитаемым, либо теряются детали рисунков и схем, то следует использовать режим «Цветной-серый», либо «смешанный». На выходе получится tiff файл всего с двумя цветами. Такой файл занимает меньше места чем любой другой формат будь то JPG или PNG. Получившийся tif следует еще сконвертировать одной из вышеупомянутых программ в формат tif с компрессией G4FAX. Это сжатие без потерь уменьшит его размер еще на 30 процентов. К примеру страница A4 в разрешении 600 dpi с ч-б текстом в tiff занимает около 150кб. В JPG эта же страница занимает больше мегабайта из-за того, что

JPG не умеет сжимать двухцветные изображения. Они их переводит в 24 битный цвет и потом сжимает. В PNG эта страница занимает в два раза больше места.

2. Цветная фотография, рисунок или фотография в оттенках серого, карта. Для вывода таких изображения следует выбрать режим «Цветной-серый». При этом для уменьшения размера выходного файла можно поставить 300 dpi вместо 600. На выходе все равно получится Tiff, но с 24-битным цветом внутри. Такие файлы следует после сконвертировать в JPG.
3. Когда на странице содержится текст и что-либо из пункта 2, то проще всего сделать также. Но в ScanTailor существует специальный режим «смешанный» для таких страниц. При этом становится доступна вкладка «зоны картинок». В ней можно обозначить зоны картинок, которые программа должна оставить в полном 24-битном цвете. Остальные участки страницы будут переведены в черно-белый цвет. Никаких выигрышей в сжатии изображений такой подход не дает. На выходе все равно будет 24-битный тифф, который лучше всего пережать в JPG.

Часто попадают очень старые или плохо отсканированные отчеты. В них при переводе страницы в черно-белый режим получается плохо-читаемый текст. Такие страницы лучше всего оставить в цвете и сжимать в JPG.

При проведении ряда тестов формат PNG оказался плохо пригоден для сжатия отчетов. При сжатии цветных изображений он проигрывает JPG, а при сжатии черно-белых Tiffy.

Сылки

- [Подборка материалов и инструментов для сканирования и обработки книг](#)
- [Руководство по программе ScanTailor](#)
- [Сайт программы ScanTailor](#)
- [Сайт программы FastStone Image Viewer](#)
- [Сайт программы IrfanView](#)

From: <http://pereval.g-utka.ru/> - Неофициальный перечень высокогорных перевалов

Permanent link:

http://pereval.g-utka.ru/doku.php/trash/%D0%BA%D0%B0%D0%BA_%D0%BE%D1%82%D1%81%D0%BA%D0%B0%D0%BD%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D1%82%D1%8C_%D0%BE%D1%82%D1%87%D0%B5%D1%82

Last update: 2017/07/10 10:00

