

Введение .....	1
1. Сканирование .....	1
2. Обработка .....	3
3. Кодировка .....	8
4. Создание текстового слоя .....	10
5. Рюшечки .....	11
5. Где взять куда залить .....	11

## Введение

Это мини руководство предназначено для описания в картинках полного цикла создания электронной версии любимой научно-технической книги (не забудьте поделиться ее, после создания). Существует несколько апробированных методов создания достаточно качественных е-книг, все они характеризуются тем, что на выходе книга имеет разрешения 600 dpi (все книги в 300 dpi, несмотря на все старания создателей, явно проигрывают).

Рассматриваемая здесь метода, основана на сканировании в **300 dpi в градациях серого** (600 dpi ч/б будет после обработки). По этому поводу следует заметить, что уменьшение геометрического размера скана в 4 раза, по сравнению со сканированием в 600 dpi, практически компенсируется увеличением глубины цвета в 8 раз (зато скорость сканирования возрастает в 2 раза **J**), а также уменьшается количество паразитного мусора.

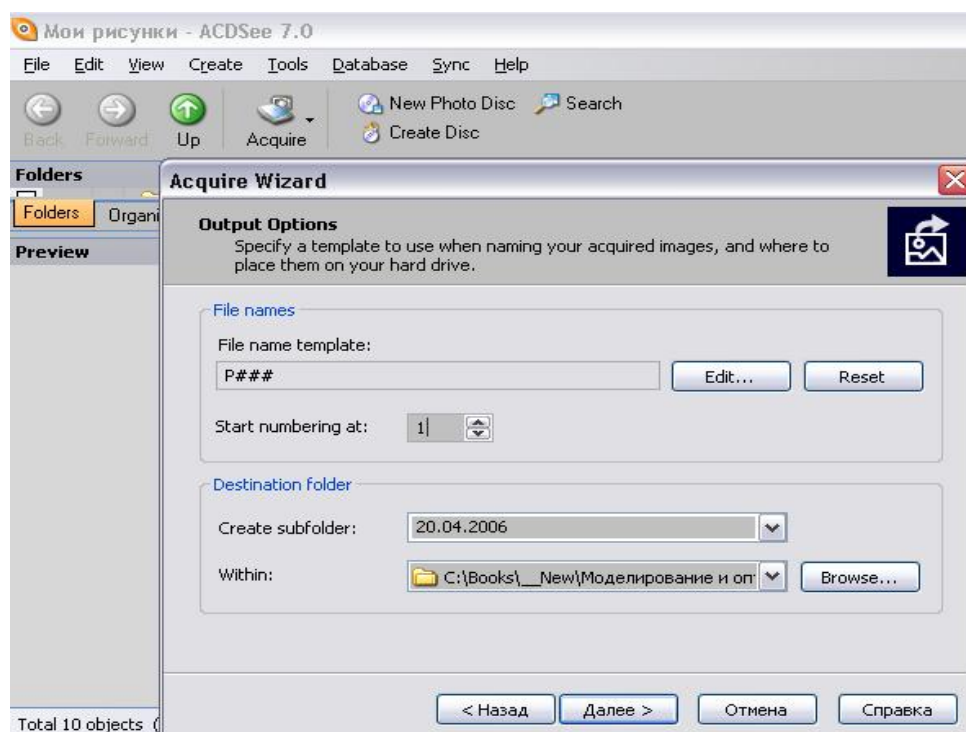
## 1. Сканирование

Беря в руки увесистую книгу, многие думают, что отсканировать ее может только маньяк. Совершенно верно, именно так. Без применения научно-организованного подхода, любая работа превращается в мучение, но, сделав работу незаметной, большого удовольствия не получишь, но дело сделаешь.

Для сканирования сгодится любая программа, способная взаимодействовать с TWAIN драйвером сканера и сохранять отсканированные изображения на диск, нумеруя их удобным способом. Сойдет любой просмотрщик графических файлов: **ACDsee**, **IrfanView**, **XnView**... Если ваш сканер поддерживается программой сканирования **VueScan**, можете использовать и ее.

Например, в **ACDsee 7.0** это выглядит примерно так:

- устанавливаем способ нумерации,
- выбираем каталог для сканов
- жмем на кнопку далее.



Сама техника сканирования незатейлива:

- берется книга, кладется разворотом (т.е. двумя страницами) на стекло, прижимается если надо сверху рукой (это быстрее, чем использовать груз).
- делается предварительное сканирование
- картинка разворачивается на 90 градусов (в нормальное положение)
- выбирается область сканирования с некоторым запасом, как правило по горизонтали (по вертикали трудно промахнуться)
- мышкой жмется кнопка основного сканирования
- после того, как данный разворот отсканирован, во время обратного движения каретки сканера, переворачиваем страницу книги, кладем на то же место и жмем опять на левую кнопку мыши (курсор ведь остался на кнопке сканирования), и так пока книга не кончится.

Т.е. идея проста, сканируем развороты в слепую. Этим достигаем максимальной скорости сканирования, которая ограничена только техническими характеристиками сканера, и полной свободы головы. Таким образом, во время сканирования Вы можете заниматься многими другими вещами, да хоть кино посмотреть.

Небольших перекосов бояться не стоит, это будет исправлено при последующей обработке, но все же надо соблюдать аккуратность.

Не забываем, что сканируем с разрешением **300 dpi и в градациях серого** (gray scale), если будете сканировать в черно-белом режиме при 300 dpi, то просто потеряете время (хорошая книжка уже не получится).

На выходе этого этапа получаем так называемый сырой материал – файлы в формате **tiff с разрешением 300 dpi в градациях серого**, обычно размер каждого файла составляет примерно 8 мегабайт.

Скорость сканирования может достигать до 200 и даже более разворотов (400 страниц) в час, т.е. сканирование среднестатистической книги, займет не более 2 часов времени.

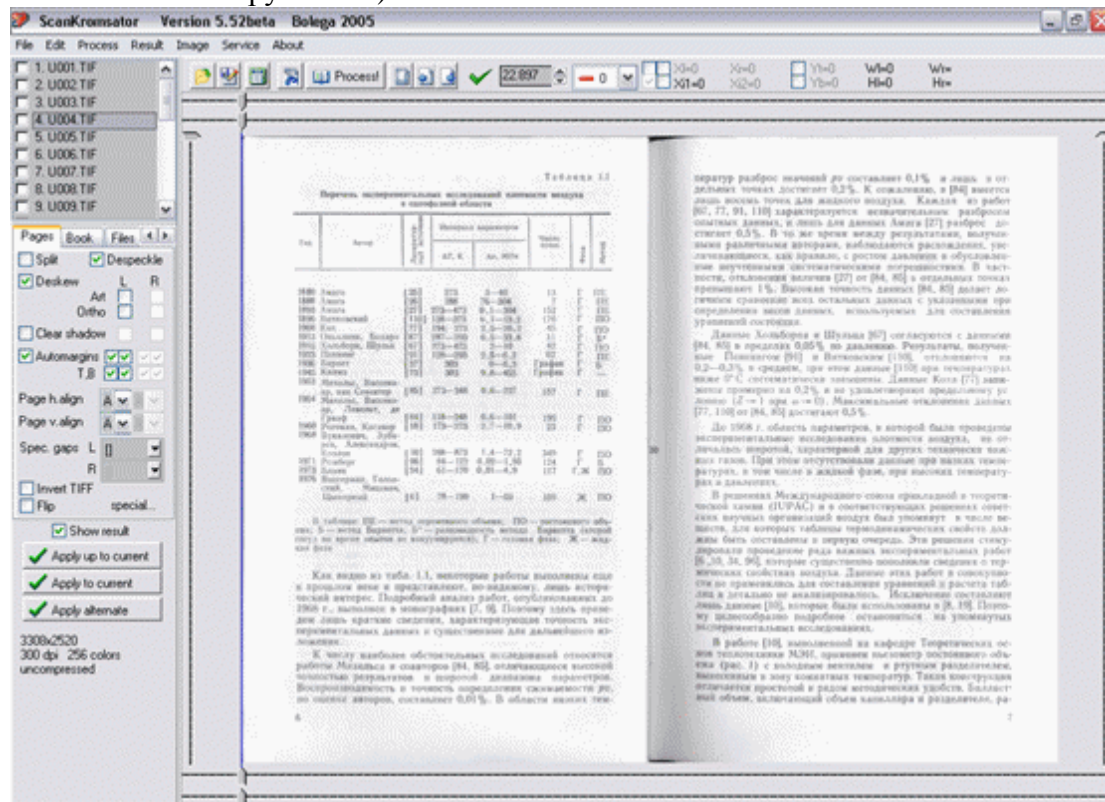
Обычно первый и последний разворот книги содержат по одной странице. Ну, так и сканируем их по одной, т.е. все-таки придется сделать 3 предварительных сканирования на книгу J.

## 2. Обработка

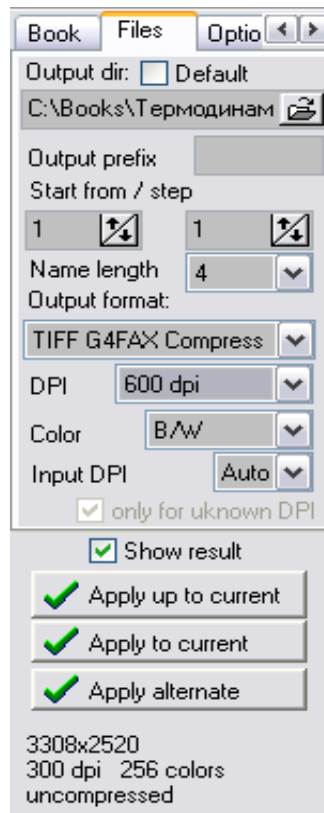
За обработку сырого материала отвечает замечательная, притом совершенно бесплатная, программа **ScanKromsator** от **bolega** (тут убедительная просьба, не надо сразу же бросаться и писать ему письма о том, как улучшить, углубить, да и просто спасибо, наверное, то же не надо посылать, просто сделайте **хорошо** несколько книг и поделитесь ими).

ScanKromsator это мощный инструмент, предназначенный для обработки сканированного материала, с целью создания качественных е-книг, со многими полезными и не очевидными для новичка функциями. Поэтому, просто следуйте пошаговой инструкции и все получится.

**а)** Запускаем программу и загружаем в нее файлы (список файлов слева сверху, под этим списком панель инструментов):

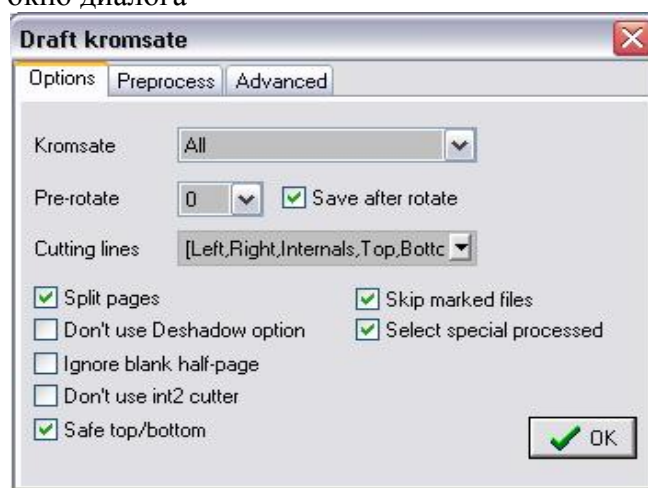


**б)** Выбираем путь для вывода результатов (закладка Files), тут же можно назначить способ нумерации выходных файлов, и что **очень важно, назначить выходное разрешение 600 dpi**.



с) Приступаем к черновому «кромсанию»:

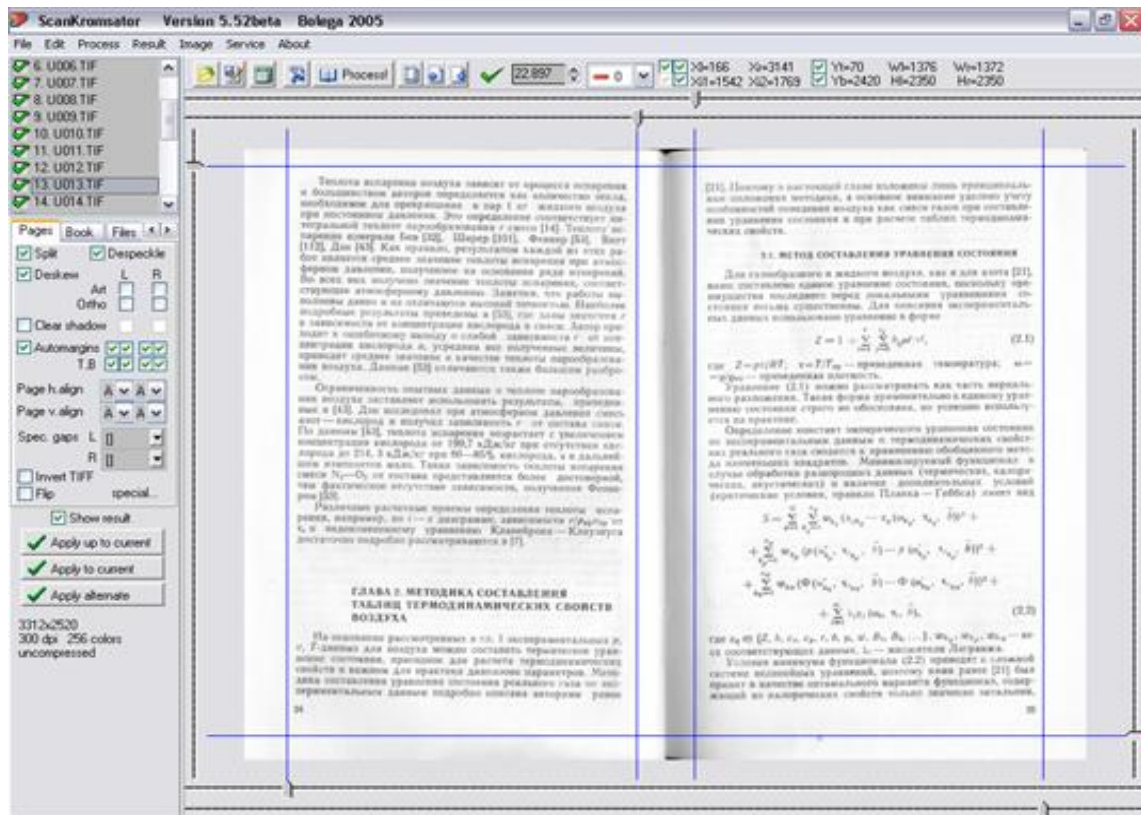
Находим левее кнопки с надписью **Process**, кнопку с ножницами (**Draft kromsate**), нажимаем, появляется окно диалога



ставим галочки на Split pages и safe top/bottom и ждем кнопку ОК.

(если первая и/или последняя страницы одинарные, т.е. не разворот, то можно предварительно покромсать их отдельно (поле Kromsate = Current), соответственно не надо для них ставить галочку Split pages)

д) Минут через 10-15:

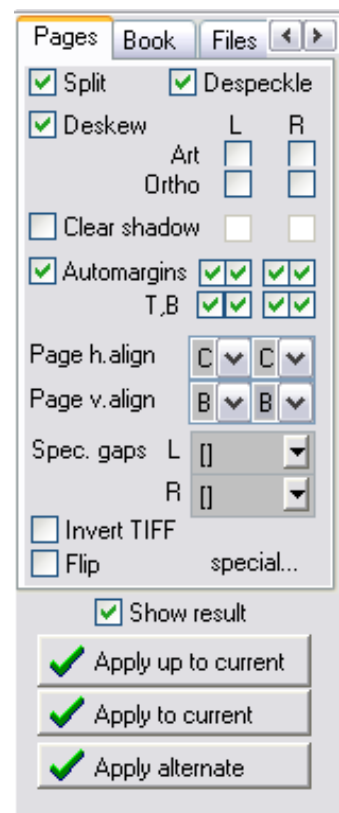


Обратите внимание на синенькие полосочки, это резак (по которым Вы безошибочно отличите эту программу от других **J**), за их пределами все будет безжалостно отрезано, а данная страница будет разделена на две (см. центральные резак). Посмотрите на то, что рядом с названиями страниц появились **зеленые галочки!**

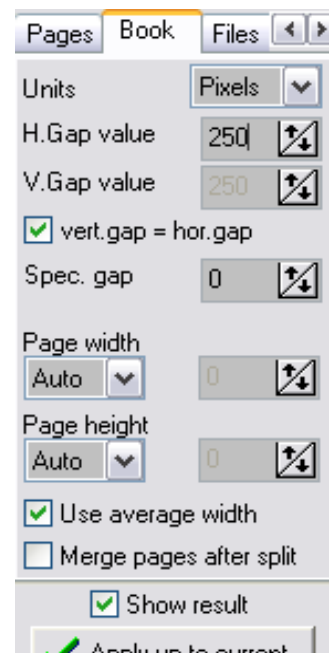
**е) Это короткий, но очень важный этап – расстановка опций.** Для этого пройдемся по закладочкам (слева в окне программы).

**Pages.** На ней выставляем способ центрирования. По умолчанию стоит **A** – автомат, это значит поместить изображение в верхний левый угол. Но, как правило (это у меня так) горизонтальное выравнивание ставится по центру (**Page h.align**) **C**, вертикальное в низ (**Page v.align**) **B** или вверх **T** это зависит от форматирования книги. **Despeckle** это убиение мелкого мусора. **Deskew** – выравнивание наклона страницы, если в результате страница получится криво выровненной, то ее можно переделать с помощью метода **Art** (включение этого метода для всех страниц замедляет процесс).

**Чтобы опция была применена ко всем страницам, при выборе ее удерживаем Ctrl.** Аналогично действуйте при выборе остальных опций, которые применяются ко всем страницам сразу.

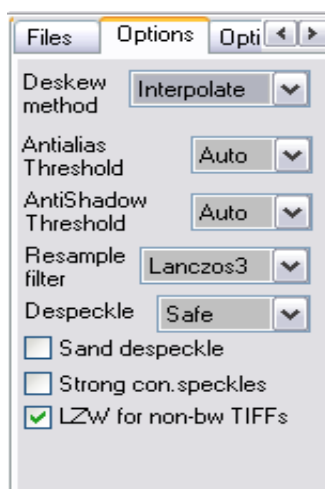


На закладке **Book** выставляем размеры выходных страниц, оставляем **Page width** и **height** в **Auto**. В поле **H.Gap value** ставим 200 (или 250) pixels.



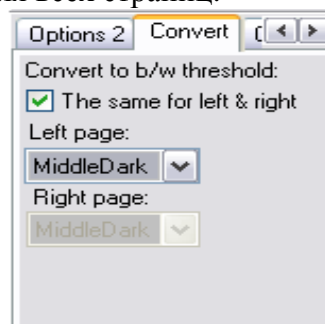
В закладке **Files**, как было сказано выше, ставим выходное **дпи 600** (иначе ничего хорошего не получится), это архи важно. От этого зависит весь окончательный результат.

Во вкладке **Options**, ставим **Deskew method = Interpolate**, для **Despeckle** метод **Safe** или **Fine+Normal** это интеллектуальный метод очистки. Например, он не вычищает точки над *i* и *j*.



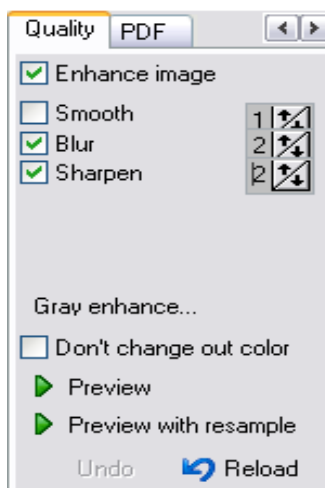
**Options 2** пропускаем.

Вкладка **Convert** – выставляем порог для преобразования из градаций серого в черно-белый. Для **Convert to b/w threshold** выбираем **MiddleDark**. Не забываем удерживать **Ctrl** при выборе опции, предназначенной для всех страниц.





Ну, наконец, последняя, но очень важная вкладка **Quality**. В **Enhance image** ставим галочки для **Blur** и **Sharpen**, значения для них обычно 2 (набор этих опций и их значения не догма, можете поэкспериментировать, но для начала поставьте как на рисунке).



**И опять очень важно**, если у вас исходники это разворот книги, то жмем на **Gray enhance** и появляется диалог **Gray image enhance**, в нем ставим **cleaner passes** в 1, ставим галку на **Correct low contrast** и главное, что бы это заработало, ставим галку на **Enable**. По этой опции происходит выравнивание освещенности (особенно важно это для центра разворота), что убирает черные полосы и кучу мусора. Незаменимая штука.



Здесь приведен вид диалогового окна для **Gray image enhance** для версии программы 5.6A, в предыдущих версиях нет **Enable**.

Можно попробовать **Smooth**, по утверждению автора программы, применение этого метода приводит к некоторому уменьшению размера книги, за счет сглаживания контура букв.

А в целом я советую, сделать 3-4 разворота, так, как описано, посмотреть результат, только потом экспериментировать по улучшению, но повторяю, в 90% случаев должно получиться неплохо.

Кстати, часть опций можно сохранить на будущее в **File->Task options settings**.

**f)** Самый скучный, но к счастью не очень долгий этап. Надо пройтись по всем страницам, с целью проверки правильности расстановки резаков.

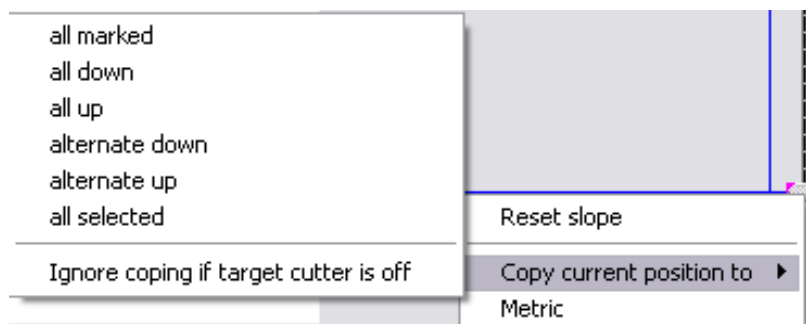


Да, если не стоит большая зеленая галка, то поставьте ее (для запоминания вносимых изменений).

Если Вы увидите, что для какой либо страницы резаки установлены не правильно, то их надо поправить. Передвигаем резаки, если надо меняем способ центрирования для данной

страницы (если текст на странице развернут на 90°, то для данной странице ставим **Deskew = Ortho** на закладке **Pages**).

Оптимально это делается так: левая рука отвечает за листание – кнопки **q** и **w**, правая за мышь, которой мы передвигаем, если надо резаки. Если Вы уверены, что для части страниц положение резака будет одинаково, то Вы можете скопировать их положение, нажав правую кнопку мыши на резаке, выберите нужную опцию (**Copy current position to**).



**g)** Кстати, знаете ли Вы, чтобы все не делать заново, задание можно сохранить (пункт основного меню **File->Save Task**)

**е)** Жмем большую кнопку **Process**. Тут появляется предупреждения, в здравом ли мы уме, что меняем разрешение, но нам уже все равно, мы все уже сделали.

Все, основное дело сделано, пусть компьютер теперь думает.

На компьютере Pentium M 1400 MHz, этот процесс занимает порядка одной минуты на разворот (это для версии 5.6A, в предыдущих версиях в 1.5÷2 раза медленнее).

Часа через 2-3, в указанной ранее папке, нас ждет результат, просматриваем его внимательно, иногда могут быть несколько неправильно выровненных страниц. Их переделываем отдельно.

Особо дотошные, могут почистить остатки вручную, так называемая тонкая очистка. Лично я это не делаю, за исключением убирания библиотечных штампов и записей на полях. Как правило, и так все замечательно.

(Если у Вас 256 и менее мегабайт памяти, то возможен глюк, он характеризуется постоянным свопом диска при смене dpi, тогда просто снимите задачу, выгрузите лишнее из памяти, запустите процесс кромсания заново)

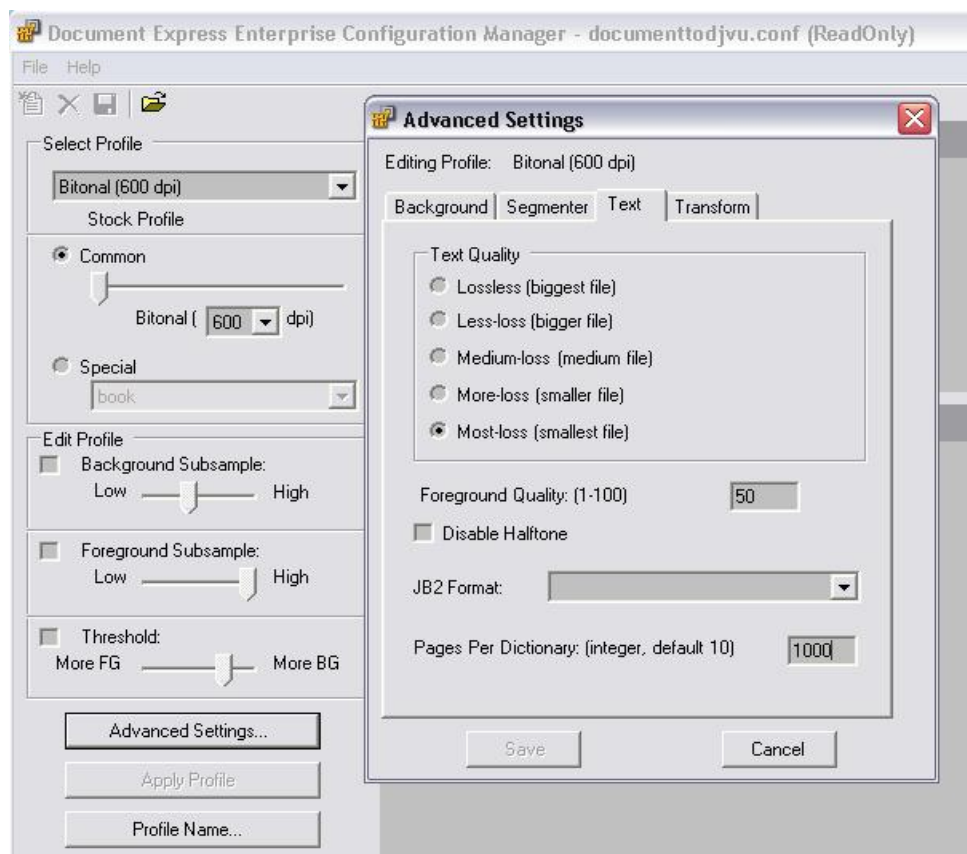
### 3. Кодировка

Кодировать в djvu можно двумя способами.

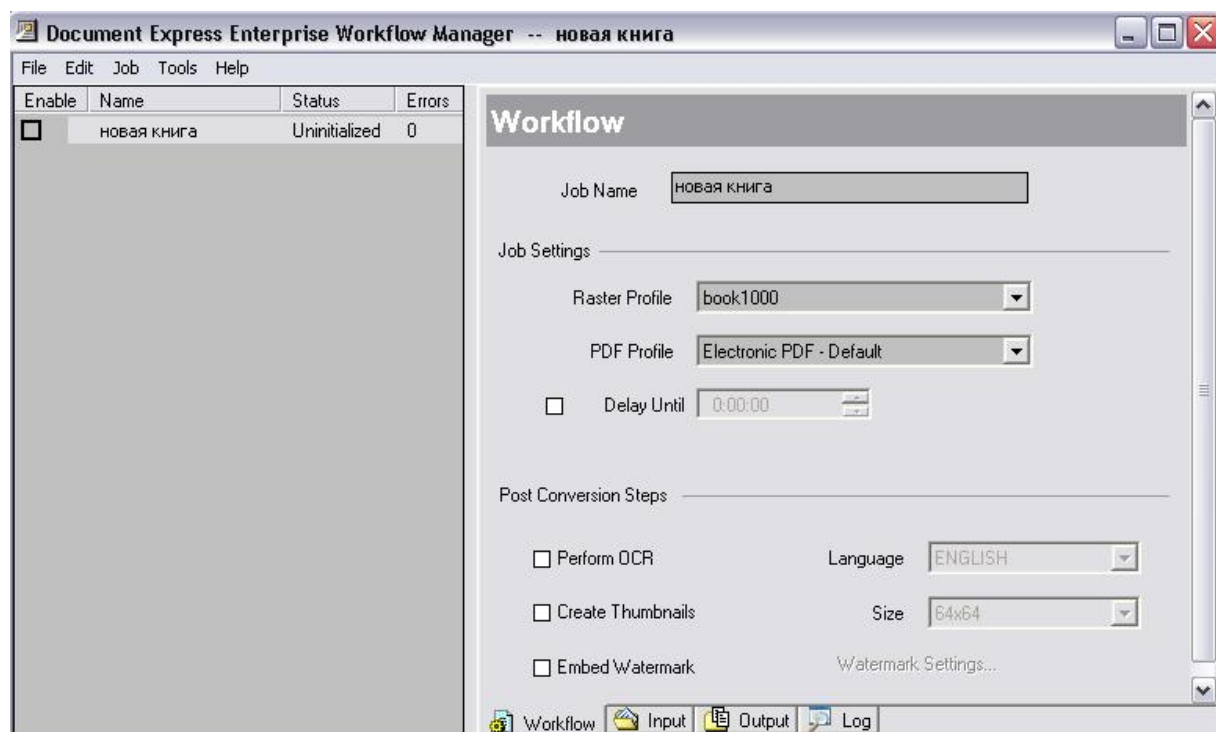
Первый, использовать или свободно распространяемую программу **Solo 3.1** (в этой программе используется старый алгоритм кодирования) или **Document Express Editor** версий от 4 до 6. Это делается просто, запускаем программу, загружаем первую страницу (только одну!), добавляем к первой странице остальные, но не более 500 (обычно 200÷300). Сохраняем с выбором профиля scanned и с разрешением 600 dpi.

Второй, рекомендованный способ, это использовать **Document Express Enterprise 5.1** (доступна облегченная версия этого пакета, объемом около 20 мегабайт). Вначале надо создать профиль для кодирования (делается это не часто, можно один раз), для этого запускаем **Document Express Enterprise Configuration Manager** из этого пакета, из списка выбираем профиль **Bitonal(600)**, нажимаем на кнопку **Advanced Settings...**, в диалоговом окне **Advanced Settings** выбираем закладку **Text** и ставим **Pages Per Dictionary** равным 1000. Сохраняем этот профиль под новым именем.





Запускаем **Document Express Enterprise Workflow Manager**, загружаем все страницы зараз, в поле **Job Name** пишем название книги, из списка **Raster Profile** выбираем, подготовленный ранее профиль, переключаемся на закладку **Output** и из списка **Separate Document(s) by** выбираем **One document only**. Ставим галочку (с самого левого края под **Enable**) и ждем конца кодирования, следим или пока эта галка исчезнет или по закладке **Log**.



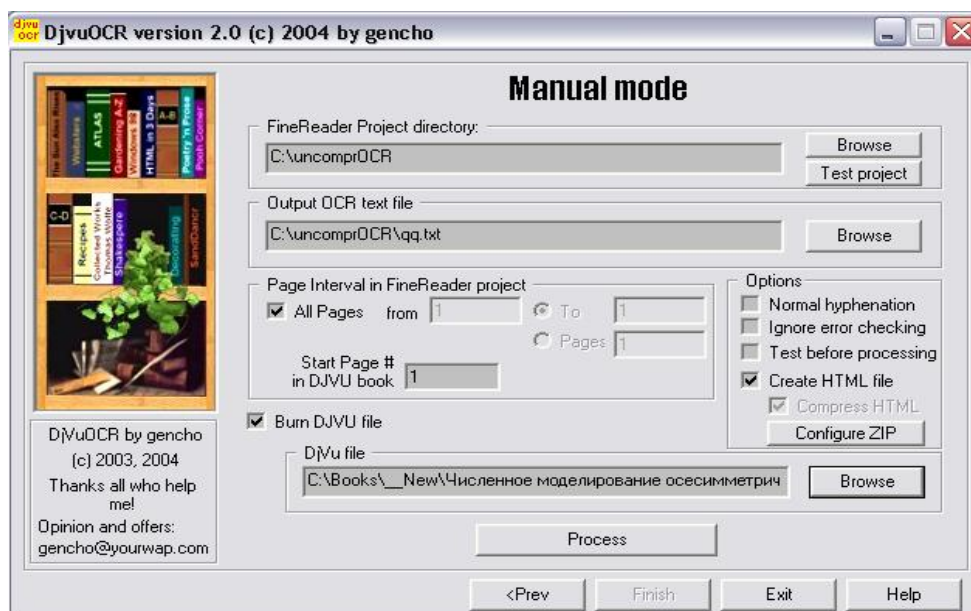
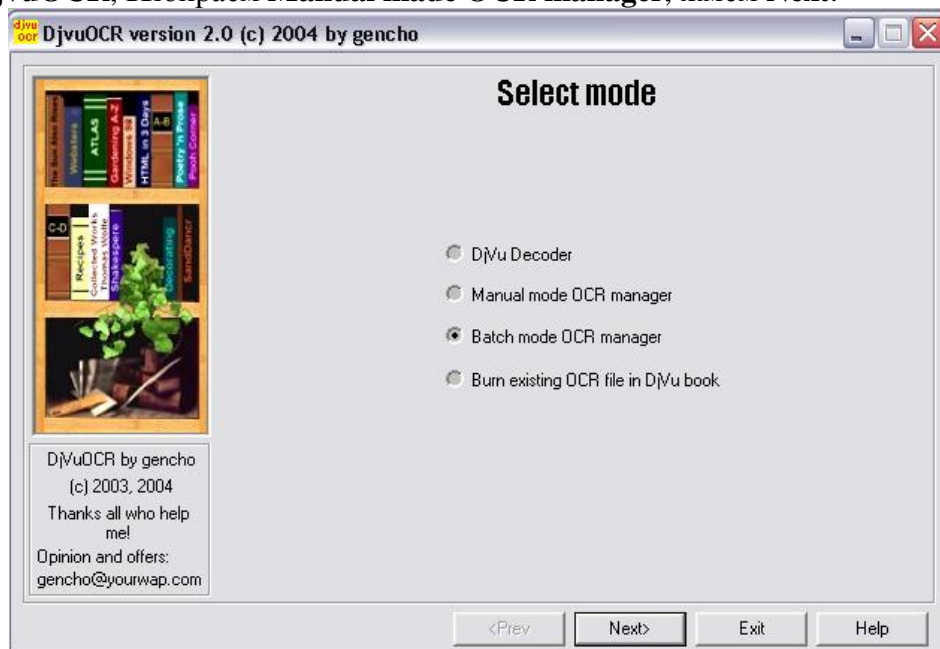
#### 4. Создание текстового слоя

После того, как все уже сделано, остановится на этом просто себя не уважать, не говоря уж об остальных.

Для добавления распознанного текстового слоя в djvu книгу потребуется две программы. Попробуйте найти **FineReader 7.0** (в восьмерке был изменен формат вывода), благо для него есть генератор ключей. Вторая программа, это небольшая утилита **DjvuOCR 2.0 final** от болгарского камрада **Gencho** (ждем третью версию под 8<sup>ой</sup> файнридер J ).

Загружаем все тифы в **FineReader**, те из которых была сделана djvu книжка, и распознаем в пакетном режиме. После распознавания, ничего не правим, не вычитываем, просто сохраняем пакет на диск. С **FineReader** все.

Запускаем **DjvuOCR**, выбираем **Manual made OCR manager**, жмем **Next**.



Далее, тоже все просто:

- **FineReader Project directory** – выбираем каталог с проектом.
- **Output OCR text file** – это любой, пустой текстовый файл, помещенный в каталог с проектом.

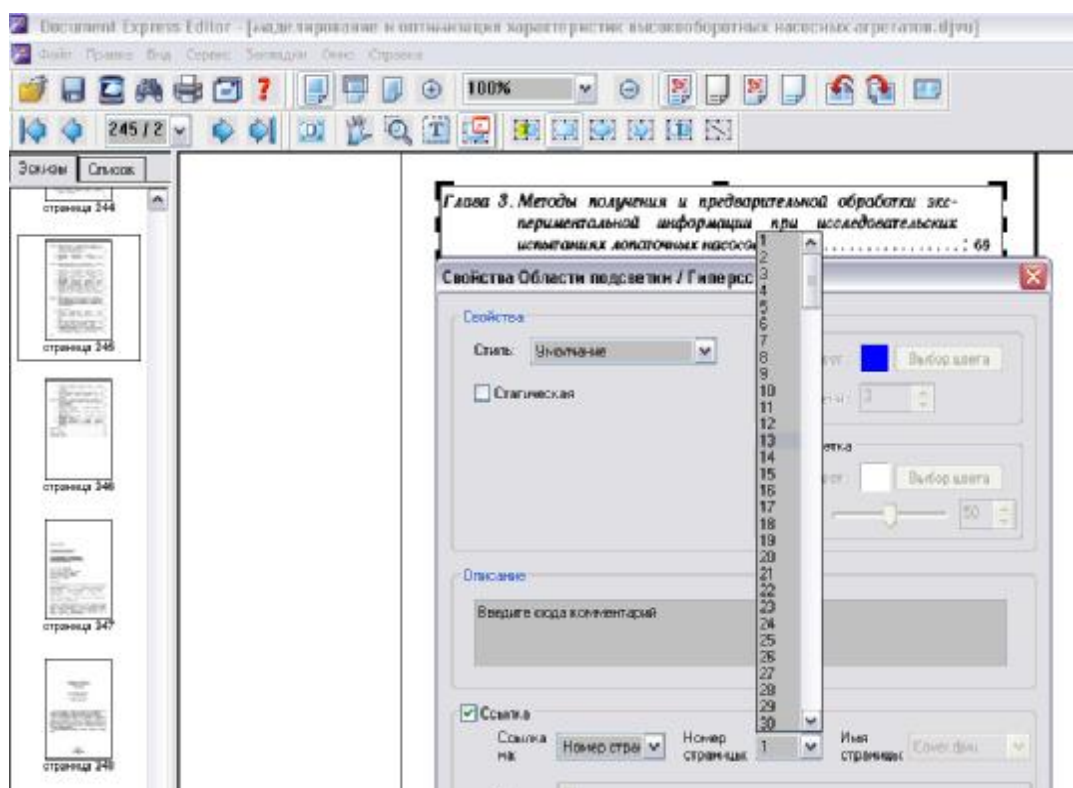
- Ставим галочку на **Burn DJVU file** и выбираем djvu книжку.
  - Нажимаем **Process**.
  - Ждем несколько минут.
- И всё.

## 5. Рюшечки

Почти все готово, осталось, но это по желанию, несколько штрихов.

Не поленитесь, добавьте обложку.

Можно потратить еще 5 минут и сделать оглавление (частичное, основные главы, полностью не обязательно, хотя есть и для этого инструменты). В 5 и 6 версии **Document Express Editor** это делается мышкой. Ждем на кнопку – прямоугольная гиперссылка, обводим пункт меню, выскакивает окно диалога – свойство гиперссылки, в котором выбираем линк на номер страницы и затем соответственно сам этот номер. Ну и так далее, пока рука не отсохнет. Только делаем это в самый последний момент, после добавления обложки, вкладок и пр., иначе ссылки сдвинутся.



## 5. Где взять, куда залить

Прежде чем делать книгу, проверьте, может она уже есть, посетите поисковый ресурс [www.poissknig.ru](http://www.poissknig.ru). Хотя если Вам встретится некачественный экземпляр, то никто не мешает его переделать (практически любая djvu книга, сделанная в 300 dpi ч/б и менее, может считаться браком).

Упомянутые программы можно скачать, поискав их на [forum.ru-board.com](http://forum.ru-board.com).

Поделиться книгой, можно выставить её на руборде, [www.avaxhome.ru](http://www.avaxhome.ru), [www.natahaus.ru](http://www.natahaus.ru), [www.kpnemo.ru](http://www.kpnemo.ru) и др.