

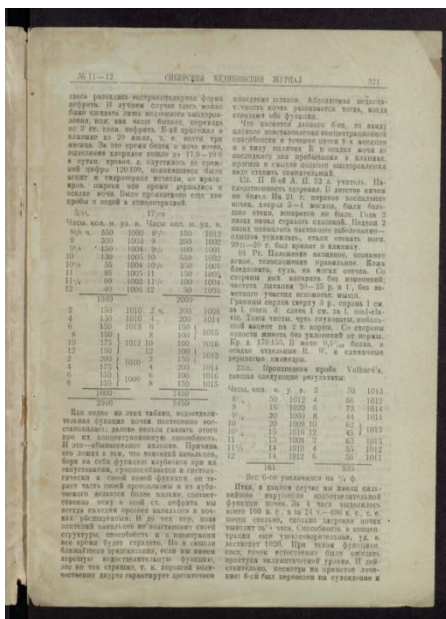
Методические рекомендации и технические требования к переводу в цифровую форму библиотечных фондов

Часть 2.

ТРЕБОВАНИЯ К ОБРАБОТКЕ СКАНКОПИЙ

В процессе оцифровки формируется цифровая копия документа, состоящая из трех копий:

- **Сканкопия** – файлы, полученные после сканирования (в формате TIFF). Каждая страница – отдельным файлом. Многостраничные TIFF не используются.



- **Пользовательская копия** – по сути это **Сканкопия**, только приведенная к аккуратному виду:
 - исправляем отклонения по вертикали/горизонтали, приобретенные при сканировании изображений страниц;
 - осуществляем кадрирование с целью приведения к единым размерам image-файлов документа;
 - сохраняем в формате TIFF LZW.
- **Пользовательская копия, оптимизированная для WEB** – пользовательская копия, предназначенная непосредственно для использования в web-среде:
 - уменьшаем физические габариты изображения до 70%;
 - сохраняем в формате JPEG сжатием до 70%.

Для долговременного архивного хранения цифровых копий документов допустимо оставлять:

- **пользовательскую копию** (без изменения физических габаритов изображения в формате TIFF LZW);
- **пользовательскую копию, оптимизированную для WEB** (в формате JPEG).

СТРУКТУРА ЭЛЕКТРОННЫХ КОПИЙ

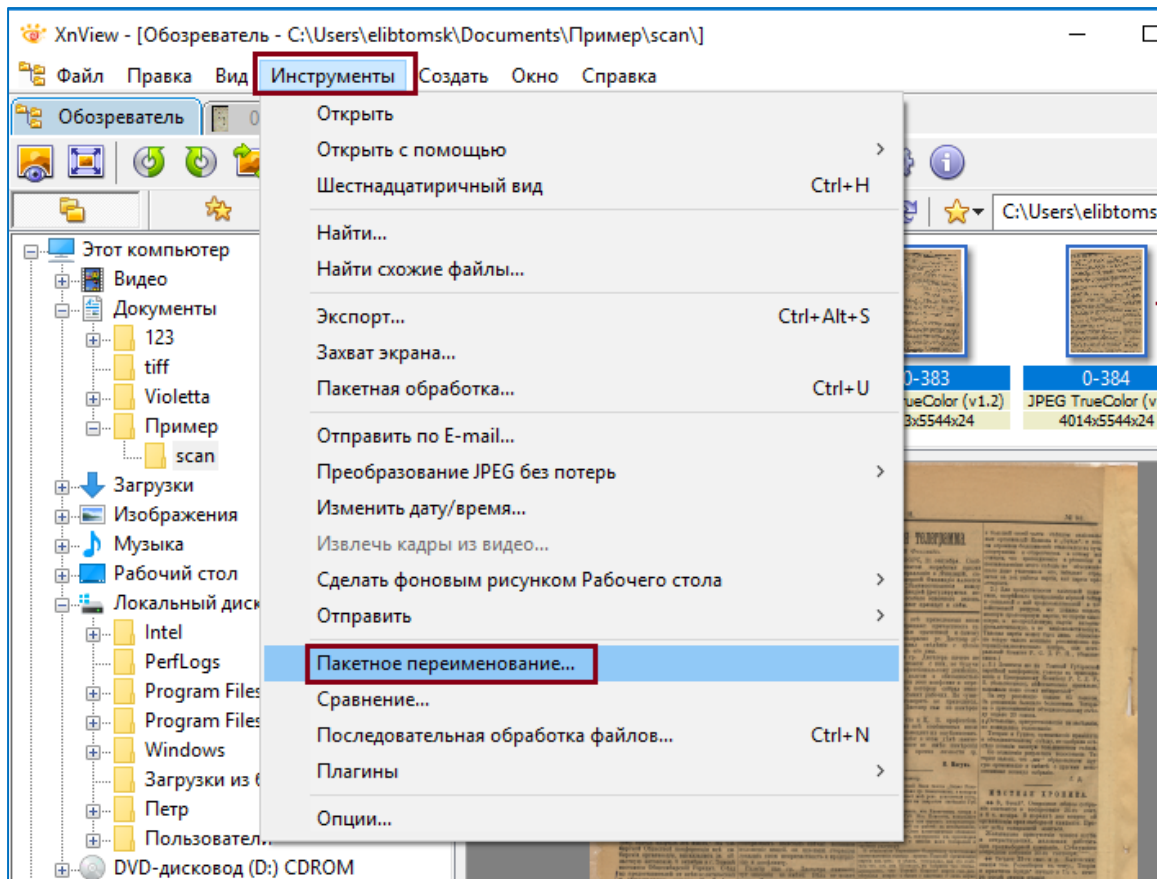
Электронные копии издания, полученные после сканирования, перед обработкой рекомендуется структурировать.

- 1.1. Структура электронных копий издания должна обязательно соответствовать оригиналу (сохранение последовательности страниц, внешний вид) и не нарушать содержательную часть документа.
- 1.2. Для электронных копий, созданных в виде набора графических файлов, устанавливаются следующие требования:
 - *каждую страницу издания сохраняют в отдельный файл;*
 - *имена файлов соответствуют номерам страниц печатного оригинала;*
 - *«*****.расширение» – рекомендуемый формат для именования файлов, это пять знаков: первые два знака отводятся под нумерацию разделов (используют в случае, если каждый раздел издания имеет свою независимую нумерацию), последние три знака отвечают за нумерацию страниц в пределах одного раздела. Общая структура электронной копии должна соответствовать изданию в печатном виде. Количество знаков в имени файлов может быть увеличено, если объем издания превышает 999 страниц.*



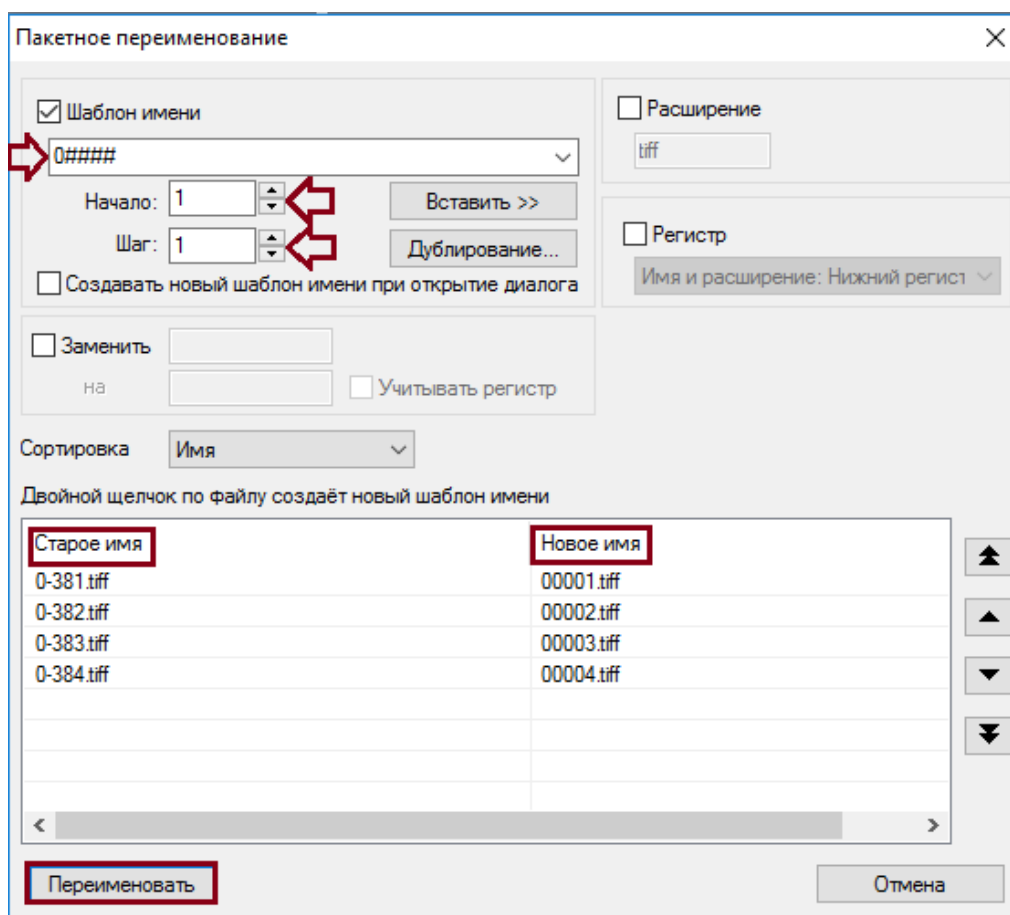
Если при сканировании файлы были пронумерованы каким-либо другим способом, то нумерацию необходимо привести в соответствие с перечисленными правилами. Для этого в **XnView** можно воспользоваться функцией **Пакетное переименование**:

- 1) Открыть папку с изображениями, полученными после сканирования;
- 2) Выделить эскизы страниц, которые необходимо переименовать (если издание имеет несколько разделов, каждый со своей независимой нумерацией, переименование осуществляем поэтапно: сначала выделяем файлы первого раздела – переименовываем, затем второго – переименовываем и т.д.);
- 3) В меню **Инструменты** выбрать **Пакетное переименование...**;



4) В окне **Пакетное переименование** настроить шаблон:

0#####, где 0 – номер раздела, ##### – количество символов в имени файла (без номера раздела), **Начало** – номер, который следует присвоить первой странице в выбранной группе, **Шаг** – величина, на которую отличаются номера двух соседних страниц в выбранной группе.

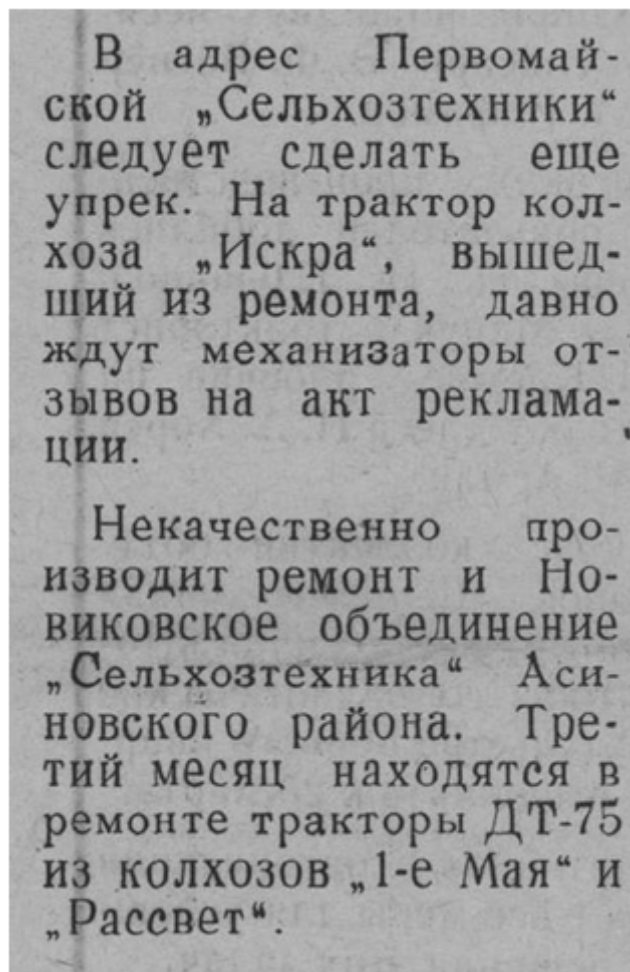


- 5) В нижней части окна **Пакетное переименование** можно увидеть предварительный результат переименования. Если он устраивает – нажать кнопку **Переименовать**.

ОБРАБОТКА ЭЛЕКТРОННЫХ КОПИЙ ДОКУМЕНТОВ

Обработка отсканированных изображений документа не должна приводить к существенным изменениям получаемых образов по сравнению с оригиналами. Потери информации или включение отсутствующей в оригинале информации не допускается. Для проведения обработки цифровых образов используют различные графические редакторы, программы для пакетной обработки и распознавания символов. Чаще всего это **Photoshop** (платный) или **Gimp** (бесплатный аналог). Для начинающих может быть рекомендована программа **XnView** (бесплатная для некоммерческого использования), сочетающая простоту, бесплатность и достаточное количество функций.

Однако следует отметить, что *иногда* в результате обработки программой **XnView** получаемые образы довольно сильно искажаются (текст неравномерно частично сжимается, а частично – растягивается, т.е. «идёт волнами»):



Если такое произошло, то можно попробовать поменять версию **XnView** или перейти на работу в другой программе. Так как этот эффект проявляется не всегда и не у всех, а программа действительно сочетает в себе простоту и достаточное количество функций, то поэтапная подготовка **Пользовательской копии** будет продемонстрирована в программе **XnView**.

Поэтапная подготовка Пользовательской копии в программе XnView.

Далее поэтапная подготовка **Пользовательской копии** будет продемонстрирована в программе XnView.



XnView для Windows

Версия 1.97.4 (Apr 30 2010)

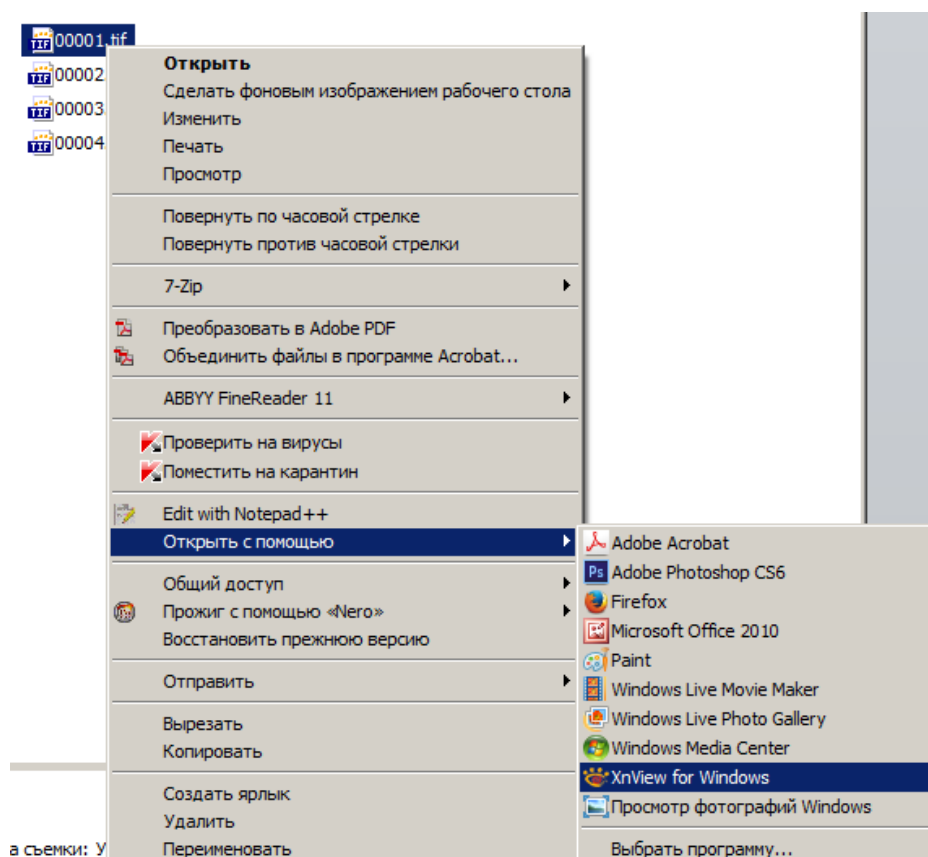
Версия библиотеки 5.62

Copyright 1991-2010 Pierre-e Gougelet

E-mail: webmaster@xnview.com - contact@xnview.com

URL: <http://www.xnview.com> - <http://www.xnview.org>

В **Проводнике** выделяем файл **00001.tiff**, удерживая указатель мыши на синем поле, щелкаем правой кнопкой мыши и выбираем **Открыть с помощью – XnView**.





Файл 0001.tiff открыт в программе XnView, обратите внимание на вкладку, соответствующую открытому файлу.

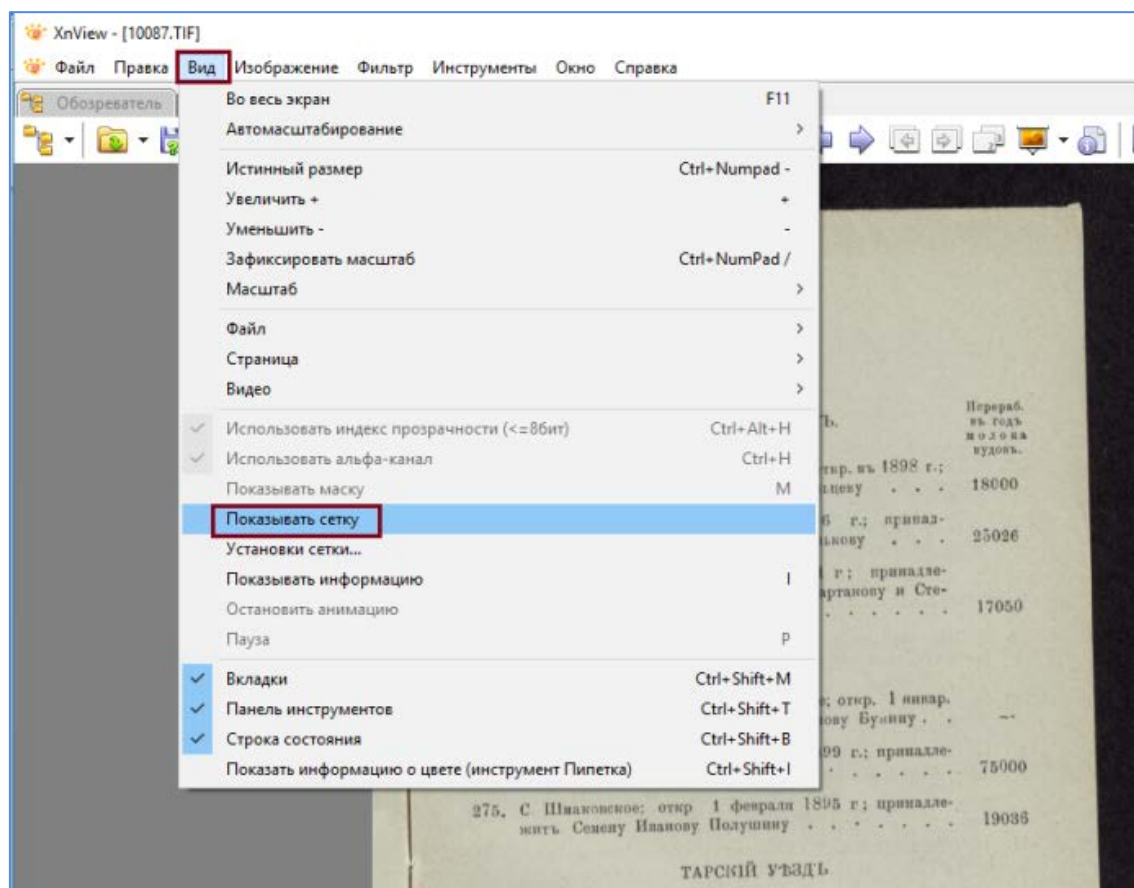
Обработка сканкопии проводится в несколько этапов.

- I. Исправление отклонений по вертикали/горизонтали, приобретенные при сканировании изображений страниц;
- II. Кадрирование с целью приведения к единым размерам image-файлов документа;
- III. Сохранение в формате TIFF LZW;
- IV. Контроль качества готовой пользовательской копии.

Действие I. Исправление отклонений по вертикали/горизонтали, приобретенные при сканировании изображений страниц

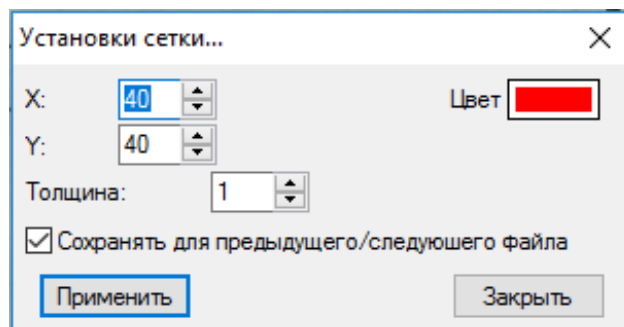
Для того чтобы провести выравнивание отсканированного изображения страницы в программе **XnView**, необходимо:

1. Открыть изображение в **XnView**.
2. Выбрать меню **Вид – Показать сетку**. Сетка помогает увидеть те изображения, которые нуждаются в выравнивании.

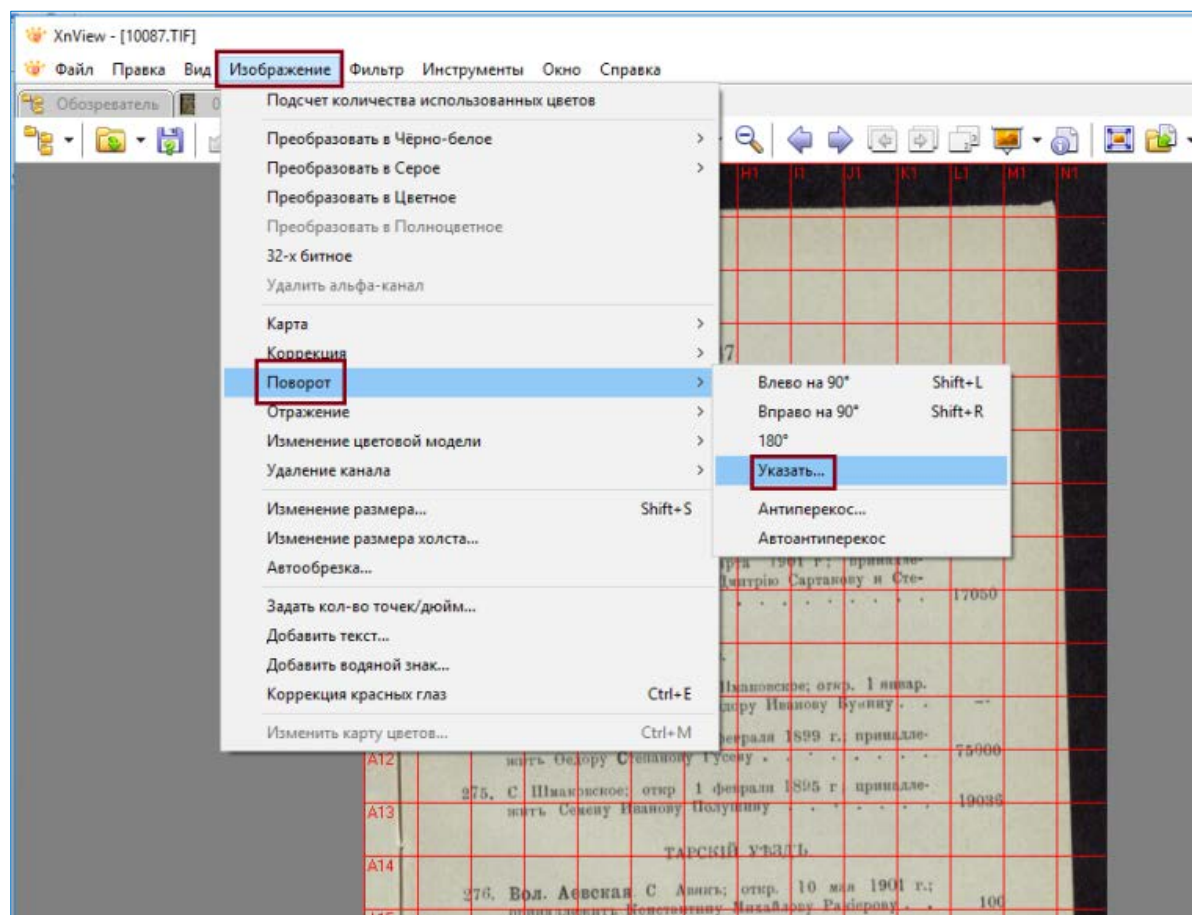


Сетку можно настраивать: задавать цвет, размер, толщину линий.

Для этого необходимо открыть меню **Вид – Установки сетки...** и выбрать те установки, которые подходят именно вам.



Далее можно переходить непосредственно к повороту изображения. Выбрать меню **Изображение – Поворот – Указать...**

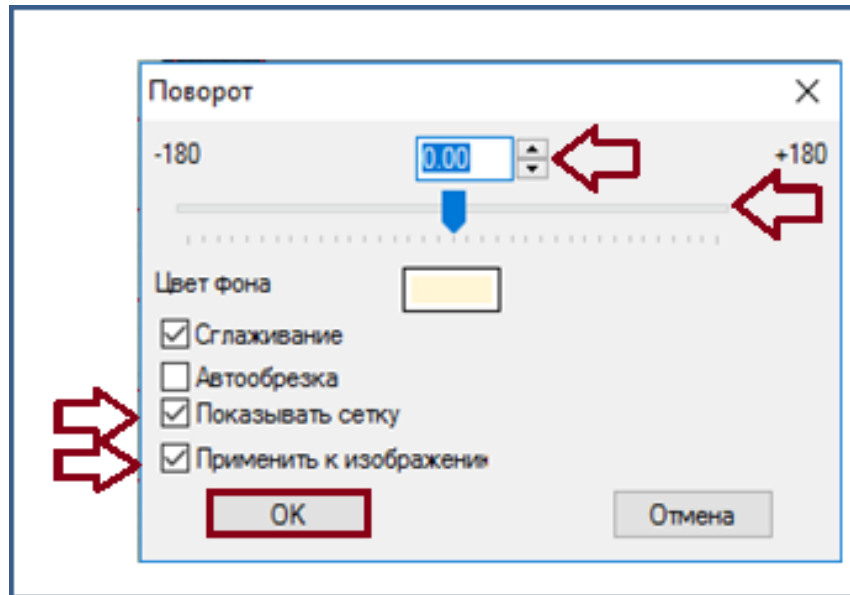


Появляется окно **Поворот**.

В этом окне, перемещая бегунок или выбирая значения угла наклона, можно развернуть изображение под нужным углом.

Галочки напротив опций **Показать сетку** и **Применить к изображению** помогают выбрать необходимый угол поворота.

3. Далее



нажимаем **ОК**.

Действие II. Кадрирование с целью приведения к единым размерам image-файлов документа

Кадрирование проводят без интерполяции (без изменения физических габаритов изображения), все страницы под единый размер по ширине и по высоте. Ширина и высота цифровых копий измеряется в пикселах (px). При кадрировании допускается обрезка краев страницы не более 1-5 мм.

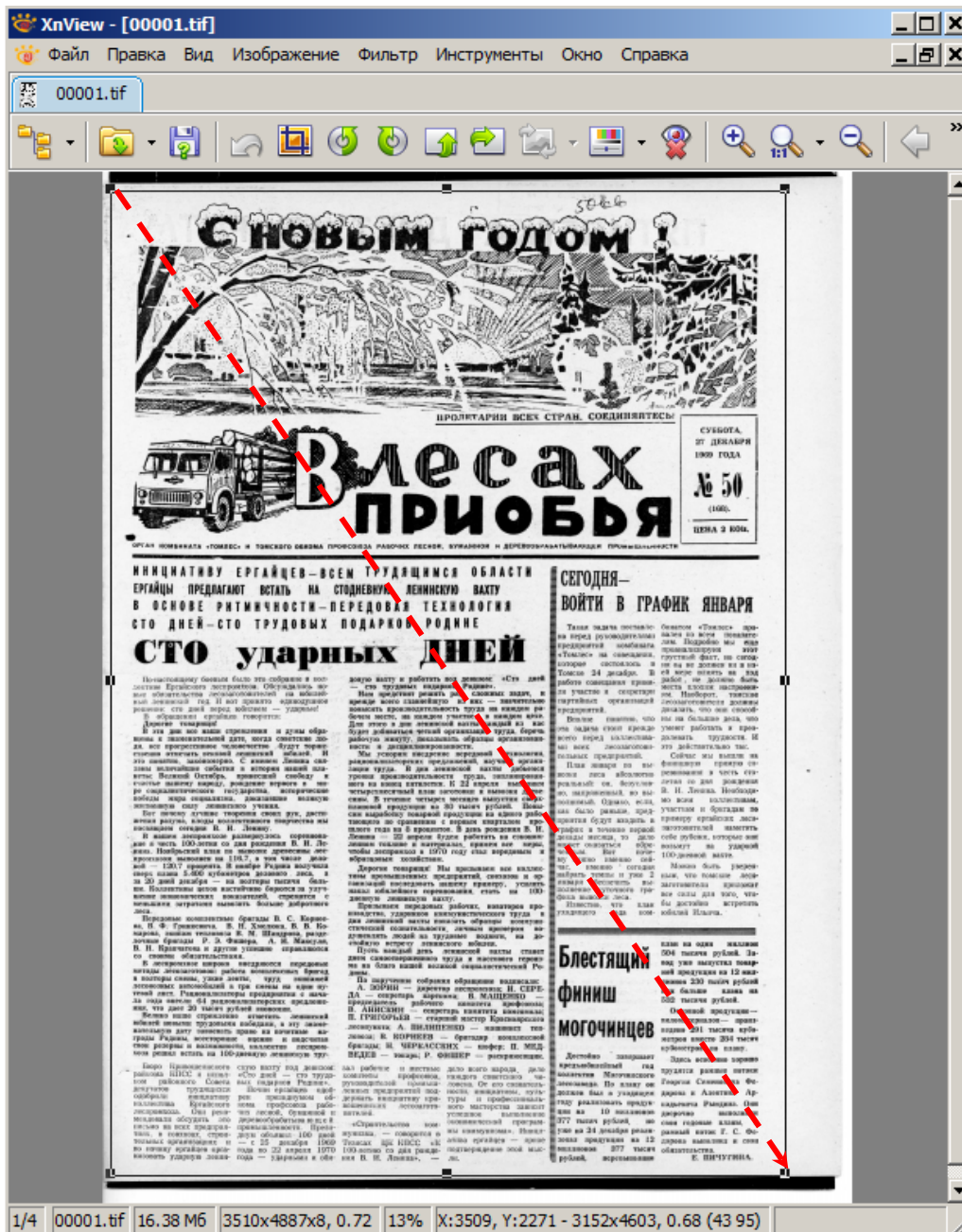


Рабочая область – это содержимое страницы без полей

Выделение для последующего кадрирования. От границ рабочей области оставляем одинаковые поля.

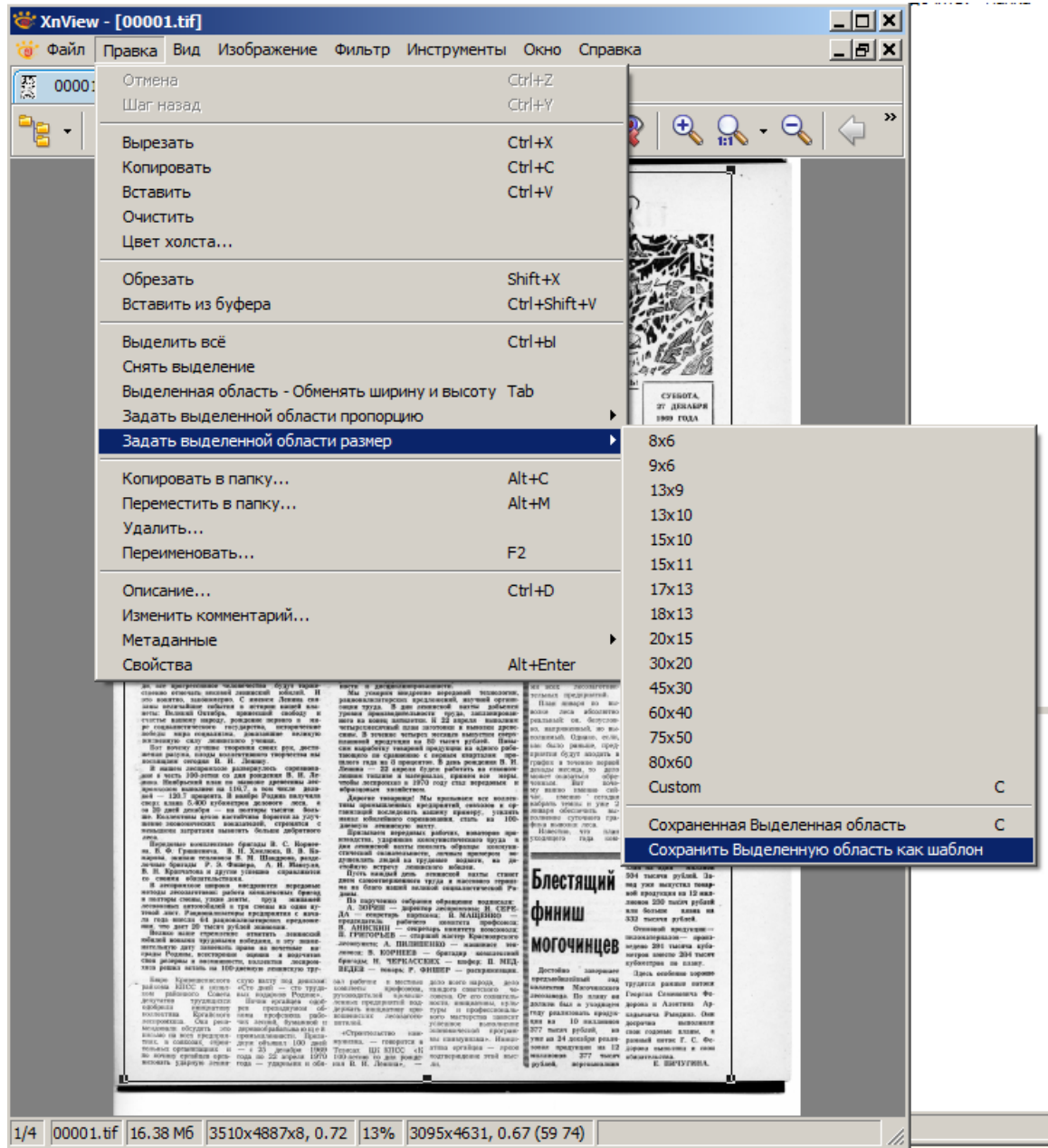
Если изображение отсканировано ровно и не нуждается в повороте или поворот уже осуществлен, то изображение можно кадрировать. Для этого необходимо:

1. Выделить область с помощью левой кнопки мыши;



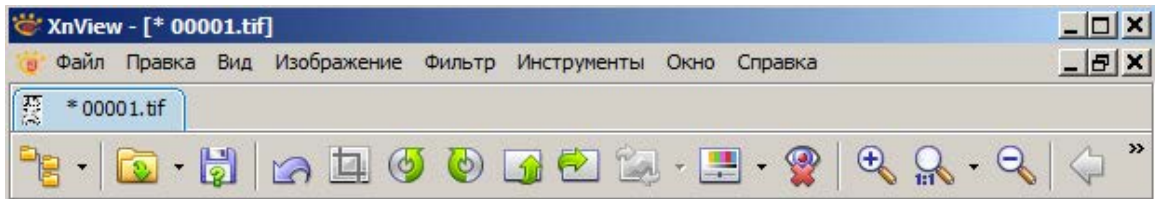
2. Откорректировать положение шаблона относительно кадрируемой страницы можно с помощью мыши (устанавливаем указатель мыши внутри шаблона и, удерживая нажатой левую кнопку мыши, выбираем область для кадрирования);

3. Выбрать меню **Правка – Задать выделенной области размер – Сохранить Выделенную область как шаблон** (для кадрирования других страниц документа далее по единому шаблону под единый размер);



4. Выбрать меню **Правка – Обрезать**

или нажать кнопку на панели инструментов



ИНИЦИАТИВУ ЕРГАЙЦЕВ – ВСЕМ ТРУДЯЩИМСЯ ОБЛАСТИ ЕРГАЙЦЫ ПРЕДАЮТ ВСТАТЬ НА СТОДНЕВНУЮ ЛЕНИНСКУЮ ВАХТУ В ОСНОВЕ РИТМИЧНОСТИ – ПЕРЕДОВАЯ ТЕХНОЛОГИЯ СТО ДНЕЙ – СТО ТРУДОВЫХ ПОДАРОКОВ РОДИНЕ **СТО ударных ДНЕЙ**

По инициативе лесоводов была эта собраны в колхозе Кудавского лесничества. Обсуждалась вся область лесничества...
Инициативу ергайцев – всем трудящимся области Ергайцы передают встать на стодневную Ленинскую вахту в основе ритмичности – передовая технология сто дней – сто трудовых подарков родине.

«Сто дней – это праздник природы Ергайцы»...
План предельно ярок, сложный, амбициозный...
Многие лесоводы считают, что это будет самая большая работа в жизни человека.

Впервые в истории области...
Впервые в истории области...
Впервые в истории области...

Впервые в истории области...
Впервые в истории области...
Впервые в истории области...

Впервые в истории области...
Впервые в истории области...
Впервые в истории области...

Впервые в истории области...
Впервые в истории области...
Впервые в истории области...

СЕГОДНЯ – ВОЙТИ В ГРАФИК ЯНВАРЯ

Тысячи людей поставили на своем рабочем месте...
Сегодня – войти в график января...
Сегодня – войти в график января...

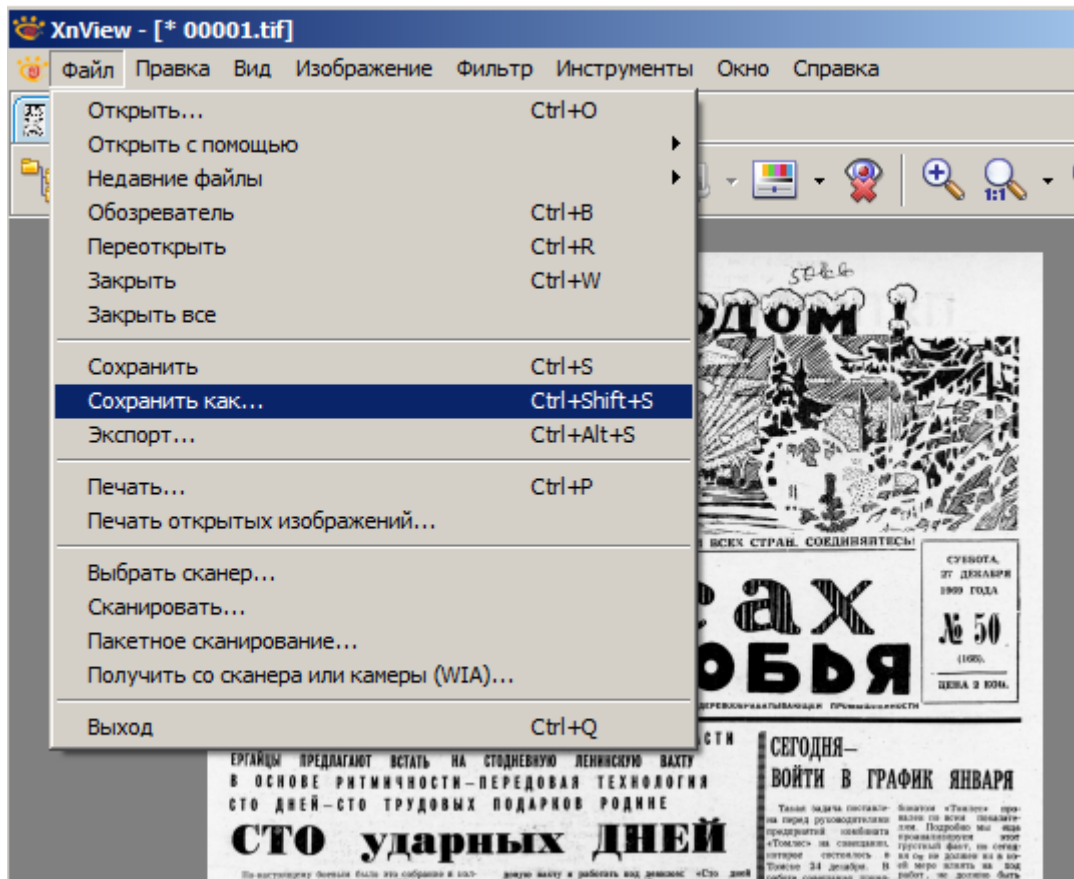
Блестящий Финиш МОГОЩИНЦЕВ

Двадцать человек...
Блестящий финиш Могощинцев...
Блестящий финиш Могощинцев...

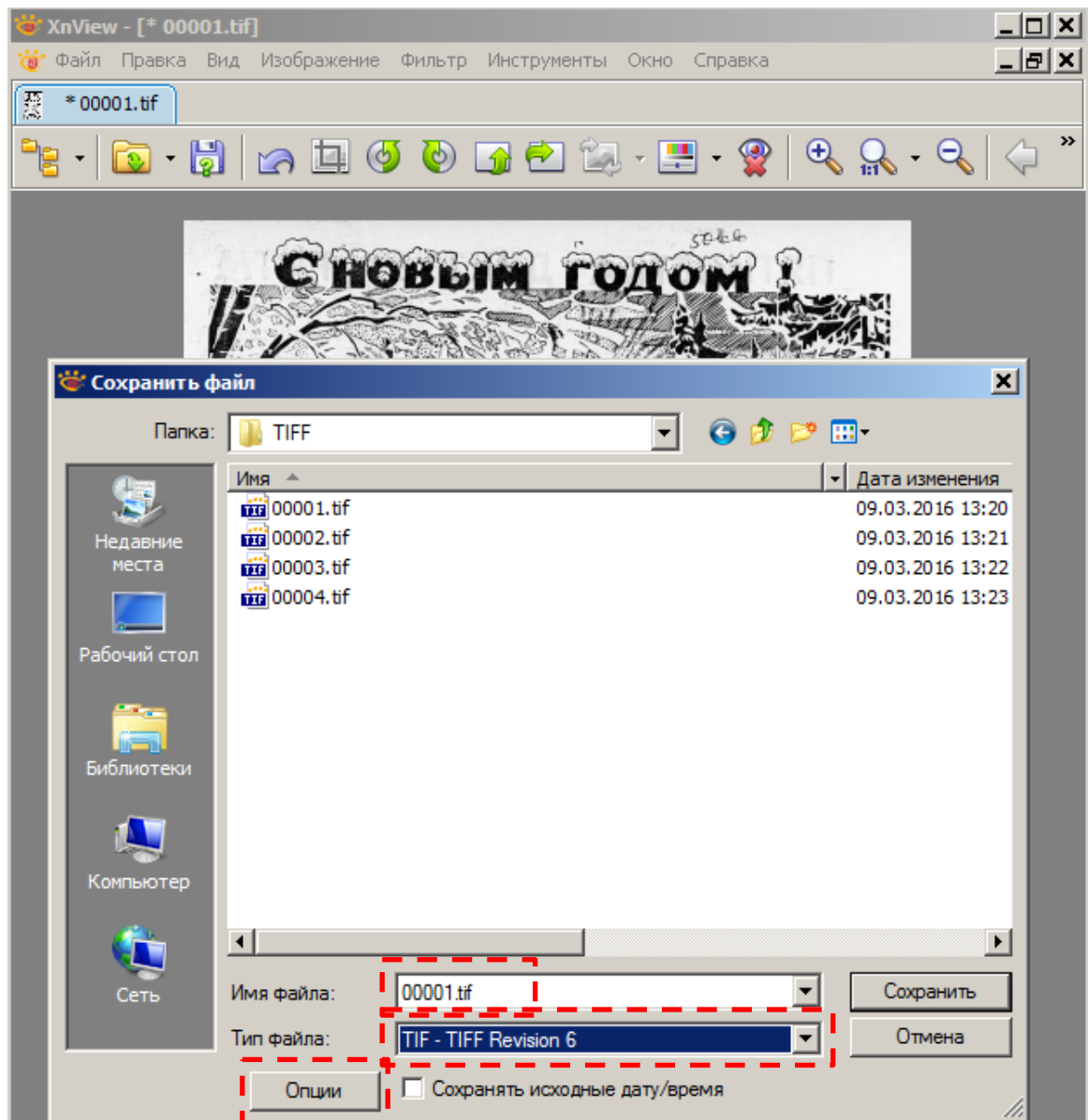
Действие III. Сохранение в формате TIFF LZW

Откадрированные изображения сохраняем в формате TIFF (LZW-сжатие), как готовую пользовательскую копию.

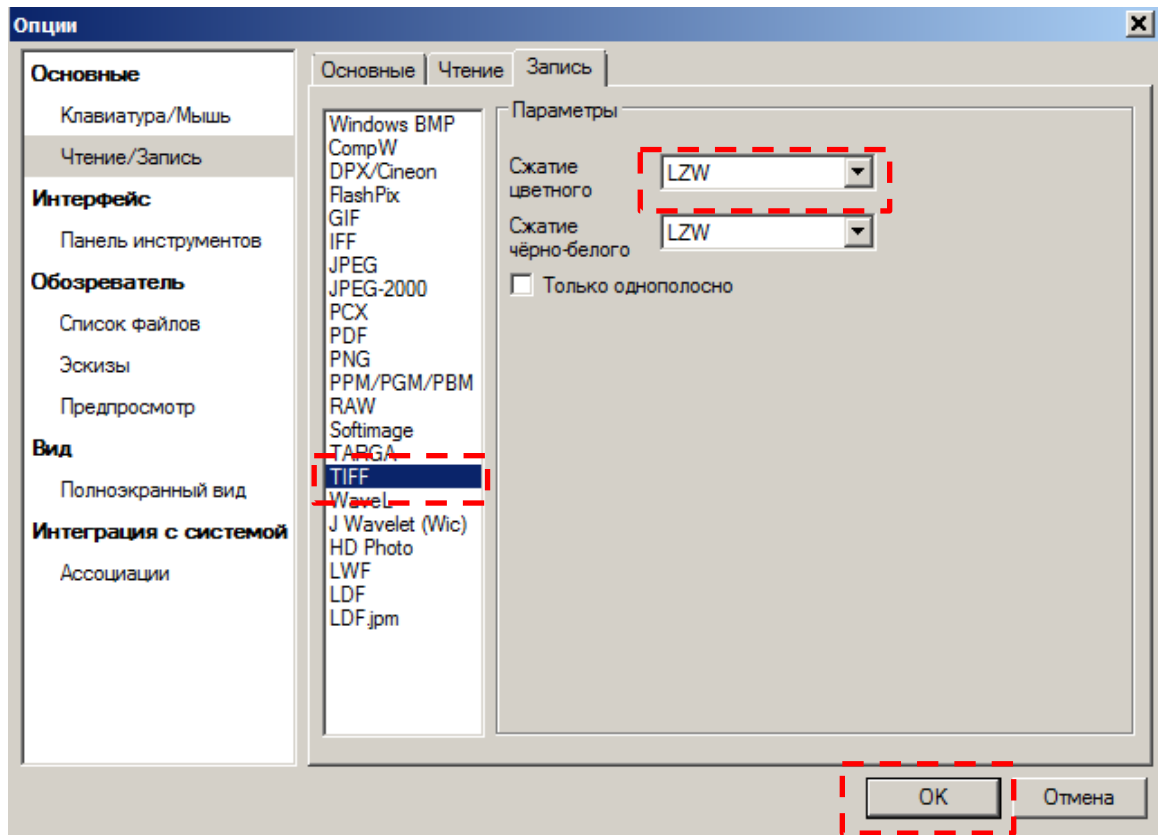
1. Обрезанное изображение сохранить в папку для обработанных изображений: меню **Файл – Сохранить как...** или сочетание клавиш **Ctrl+Shift+S**;



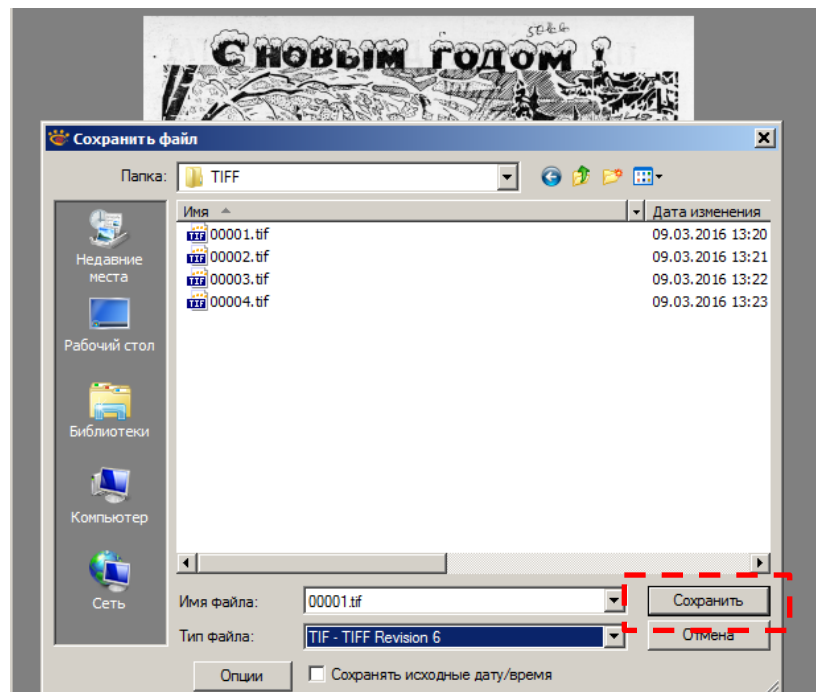
2. В диалоговом окне необходимо указать:
- Папку, куда сохранять,
 - имя файла (соответствует исходному скану),
 - тип файла tiff.



- открыть **Опции** и выбрать **TIFF – Сжатие цветного LZW**. Не забудьте подтвердить изменение настроек программы: нажать **ОК**;



- Далее в окне **Сохранить файл** необходимо нажать кнопку **Сохранить**;



3. Последний шаг: закрыть файл меню **Файл – Закреть**.
4. Перейти к следующему изображению.

Для обработки последующих страниц (при обработке страниц того же документа) повторяем действия, описанные в разделе **Подготовка Пользовательских копий**, кроме **Действия II**.

Если изображение отсканировано ровно и не нуждается в повороте или поворот уже осуществлен (согласно Действию I), то изображение можно кадрировать.



1. Выбрать меню **Правка – Задать выделенной области размер – Сохраненная выделенная область** или нажать клавишу **C** (на клавиатуре). Поверх изображения появится рамка для кадрирования заданных размеров (сохраненный шаблон);
2. Необходимо откорректировать расположение шаблона по поверхности страницы – устанавливаем указатель мыши внутри шаблона и, удерживая нажатой левую кнопку мыши, выбираем область для кадрирования;
3. Далее меню **Правка – Обрезать** или нажимаем кнопку на панели инструментов



Обрезанное изображение сохранить в папку для обработанных изображений в формате **LZW** (согласно Действию III).

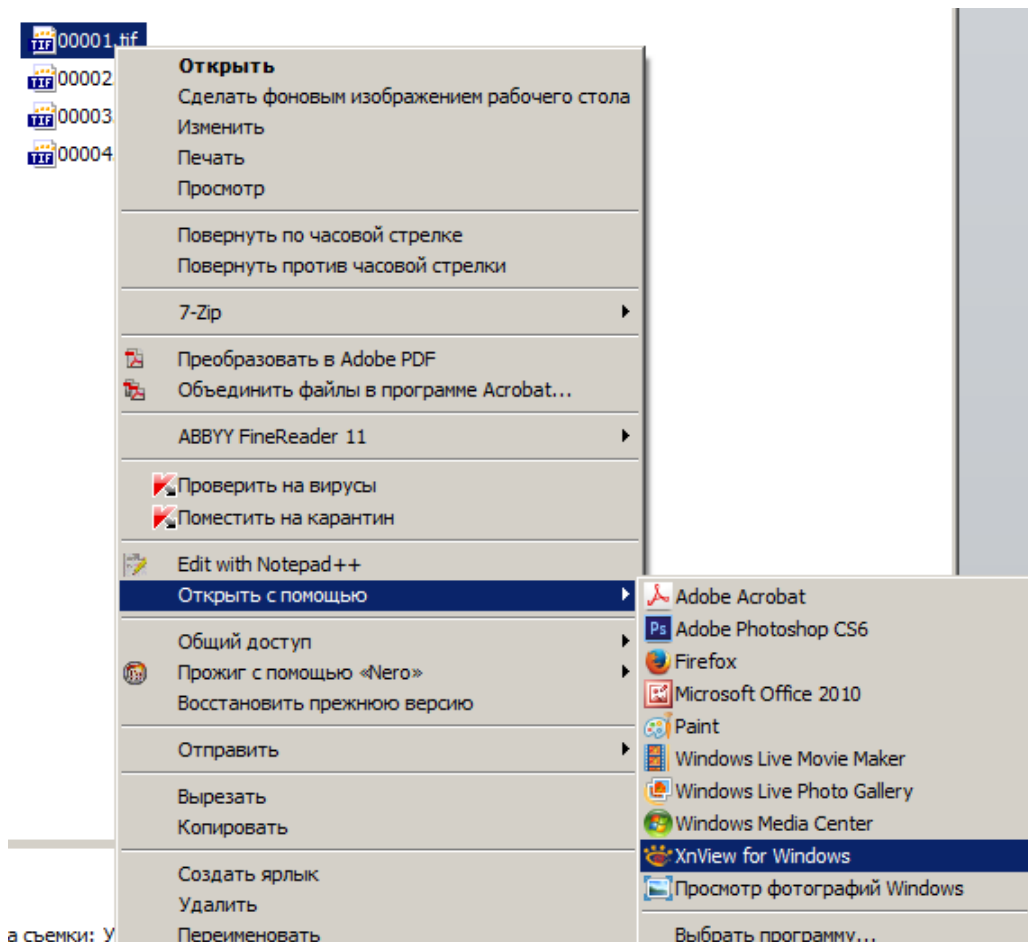
По окончании обработки не забывайте меню: **Файл – Заккрыть файл**.

Можно перейти к обработке следующего изображения.

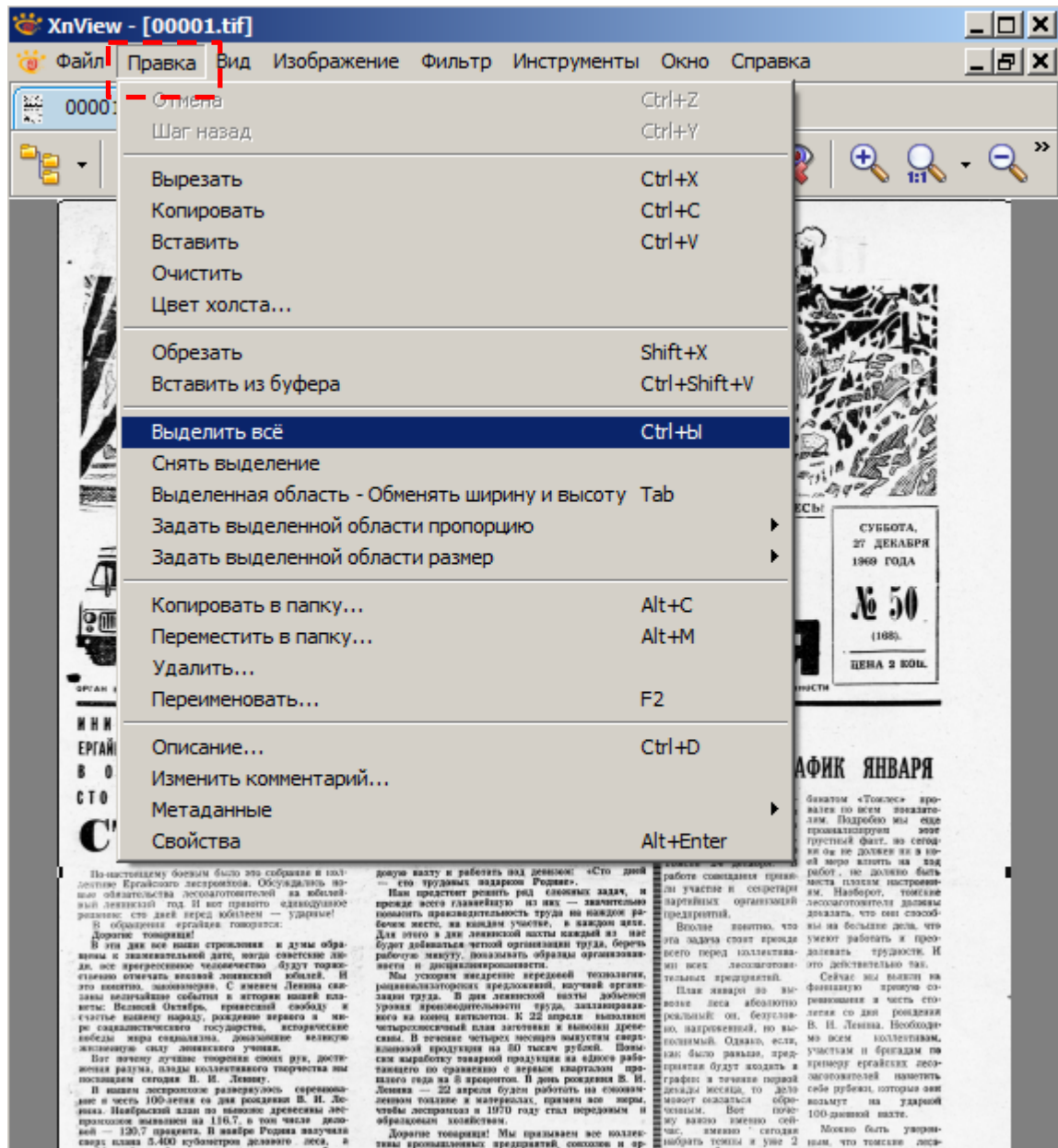
Внимание:

В случае, если шаблон области выделения был утрачен, можно воссоздать шаблон области выделения по имеющему кадрированному изображению.

В Проводнике выделяем ранее обработанный файл в формате TIFF LZW, устанавливаем указатель мыши синей области и нажимаем правую кнопку мыши, далее выбираем **Открыть с помощью – XnView**.



Далее в меню **Правка – Выделить всё**




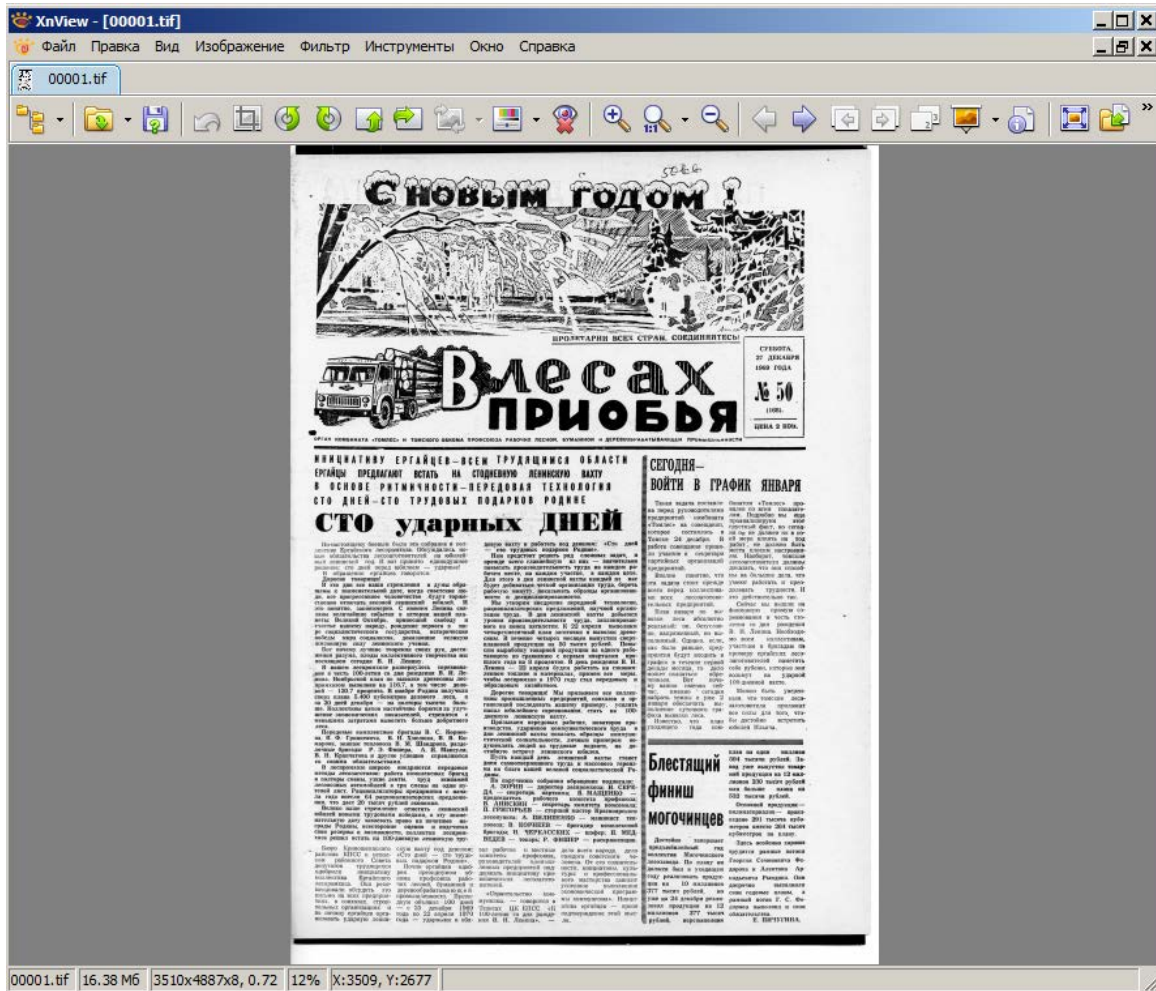
Появится рамка области выделения, далее в меню **Правка – Задать выделенной области размер – Сохранить Выделенную область как шаблон.**

Шаблон области выделения восстановлен – можно продолжить работу.

Действие IV. Контроль качества готовой пользовательской копии.

1. Открыть программу XnView.

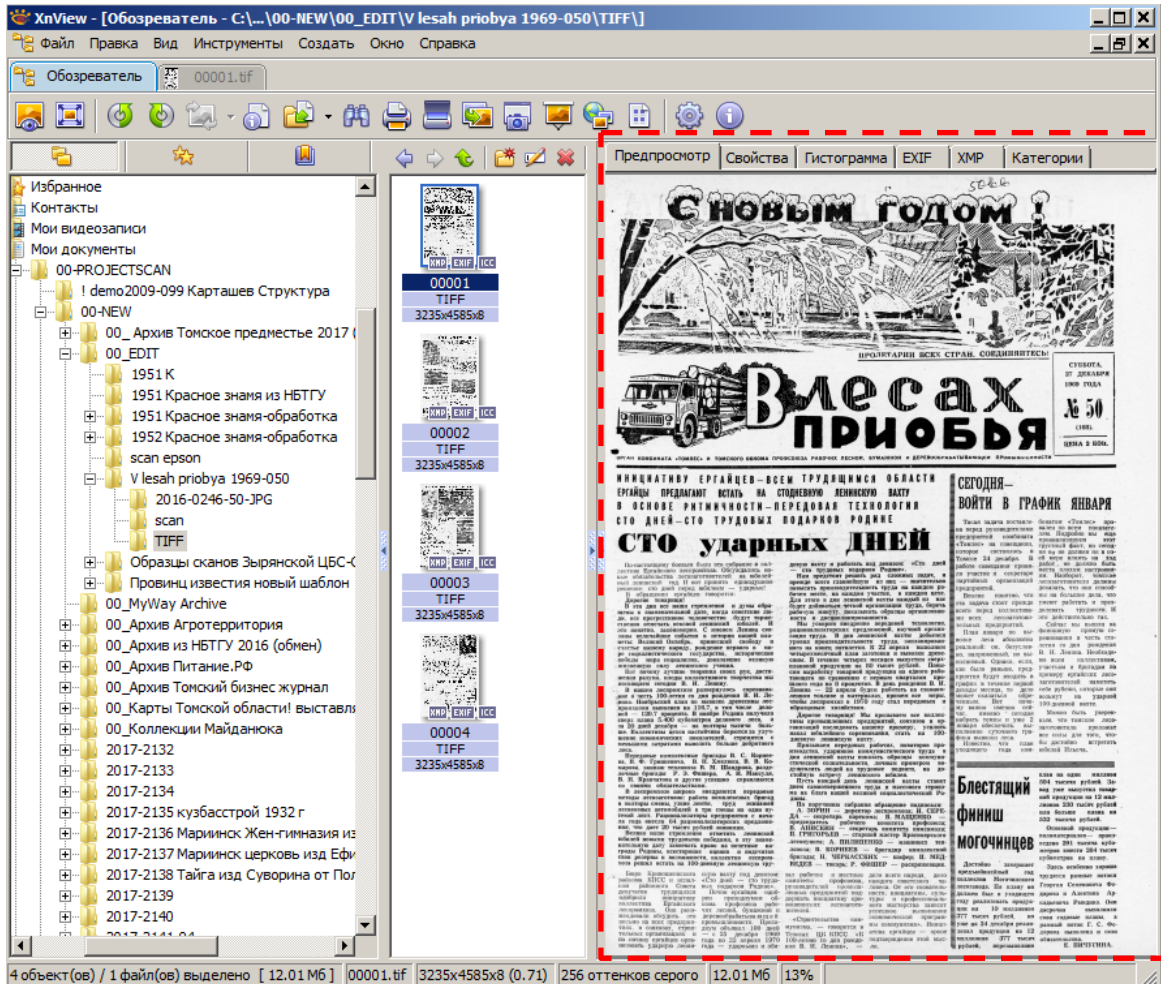
2. В окне программы с помощью кнопки **Обозреватель**  перейдем из режима Редактирования изображения в режим Обозревателя.



3. В режиме **Обозревателя** окно программы будет поделено на несколько областей. Область аналогично **Проводнику Windows**, позволяющая выбрать папку с файлами, содержащую подготовленную пользовательскую копию.

Блок, содержащий эскизы image-файлов. Здесь выбираем файл.

Третья область **Предпросмотра** выводит предпросмотр выбранного изображения, в ней же можно посмотреть его **Свойства**.



Свойства – размер, цветность, формат, сжатие, цветовая модель файла и т.д.

The screenshot displays the XnView application window. The title bar reads "XnView - [Обозреватель - C:\... \00-NEW\00_EDIT\lesah priobya 1969-050\TIFF\]". The menu bar includes "Файл", "Правка", "Вид", "Инструменты", "Создать", "Окно", and "Справка". The toolbar contains various icons for file operations. The left sidebar shows a folder tree with "00-PROJECTSCAN" expanded to "00_EDIT" and "lesah priobya 1969-050". The central area shows a list of image thumbnails, with "00001" selected. A red dashed box highlights the selected image and its corresponding properties in the right-hand panel.

The properties panel is divided into sections:

- Файл** (File):

Имя файла	00001.tif
Описание	Файл "TIF"
Размер файла	12 592 480
Создан	01.08.2016 16:34
Изменён	09.03.2016 14:20
Открыт	01.08.2016 16:34
- Изображение** (Image):

Формат	TIFF
Ширина	3235
Высота	4585
Кол-во бит	8
Цветовая модель	RGB
Точек/дюйм	300 x 300
Размер при печати	27.39 x 38.82 см, 10.78 x 15.28 дюймов
Сжатие	LZW
Кол-во изображений	1
Местоположение	Верх-Слева
Цветовой профиль	sGray
- Дополнительно** (Advanced):

Photometric Interpretation	Black=0
PhotometricInterpretation	1
PlanarConfiguration	1
SamplesPerPixel	1

The status bar at the bottom indicates: "4 объект(ов) / 1 файл(ов) выделено [12.01 Мб] 00001.tif 3235x4585x8 (0.71) 256 оттенков серого [12.01 Мб 13%".

Выделяя поочередно файлы Пользовательской копии, во вкладке **Свойства** проверяем на соответствие установленным требованиям:

- Имя файла на соответствие – порядковая нумерация страниц издания;
- Формат изображения (формат файла) – TIFF;
- Ширина и высота изображения в пикселах (для всех страниц документа единые размеры);
- Кол-во бит – 8 (для файлов, сканированных в режиме Серого/ Grayscale);
- Точек/дюйм 300x300 (параметр, задаваемый при сканировании 300 dpi);
- Размер при печати в см должен соответствовать размерам оригинала;
- Сжатие: LZW.

Файл	
Имя файла	00001.tif
Описание	Файл "TIFF"
Размер файла	12 592 480
Создан	01.08.2016 16:34
Изменён	09.03.2016 14:20
Открыт	01.08.2016 16:34
Изображение	
Формат	TIFF
Ширина	3235
Высота	4585
Кол-во бит	8
Цветовая модель	RGB
Точек/дюйм	300 x 300
Размер при печати	27.39 x 38.82 см, 10.78 x 15.28 дюймов
Сжатие	LZW
Кол-во изображений	1
Местоположение	Верх-Слева
Цветовой профиль	sGray
Дополнительно	
Photometric Interpretation	Black=0
PhotometricInterpretation	1
PlanarConfiguration	1
SamplesPerPixel	1

Оптимизация пользовательской копии для WEB

Пользовательская копия в формате TIFF LZW предназначена для архивного хранения и для локального использования. Из готовой пользовательской копии создаем цифровую копию для использования в web (оптимизированную для web):

- уменьшаем физические габариты изображений до 70%;
- файлы конвертируем из формата TIFF LZW в формат JPG сжатие до 70%.

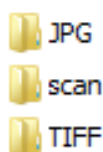
Оптимизированные цифровые материалы должны легко и комфортно читаться при просмотре на мониторе с разрешением 1920x1080 пикселей, при этом изображения не должны иметь эффекта растривания вокруг символов или появления муара.

Для примера сравним пользовательскую копию до и после оптимизации:

Пользовательская копия (TIFF LZW 8-бит) 300 dpi Исходный размер изображения 3510 x 5100 в пикселях	Пользовательская копия, оптимизированная для web (JPG 8-бит) 300 dpi Размер изображения уменьшен до 70 % 2457 x 3570 в пикселях
Размер файла 11 Мб	Размер файла 966 Кб

В качестве примера газетный номер «В лесах Приобья» 1969 года издания, №50, в количестве 4 страниц. Цифровая копия в процессе оцифровки готовится в 3 вариантах:

V_lesah_priobya_1969_050



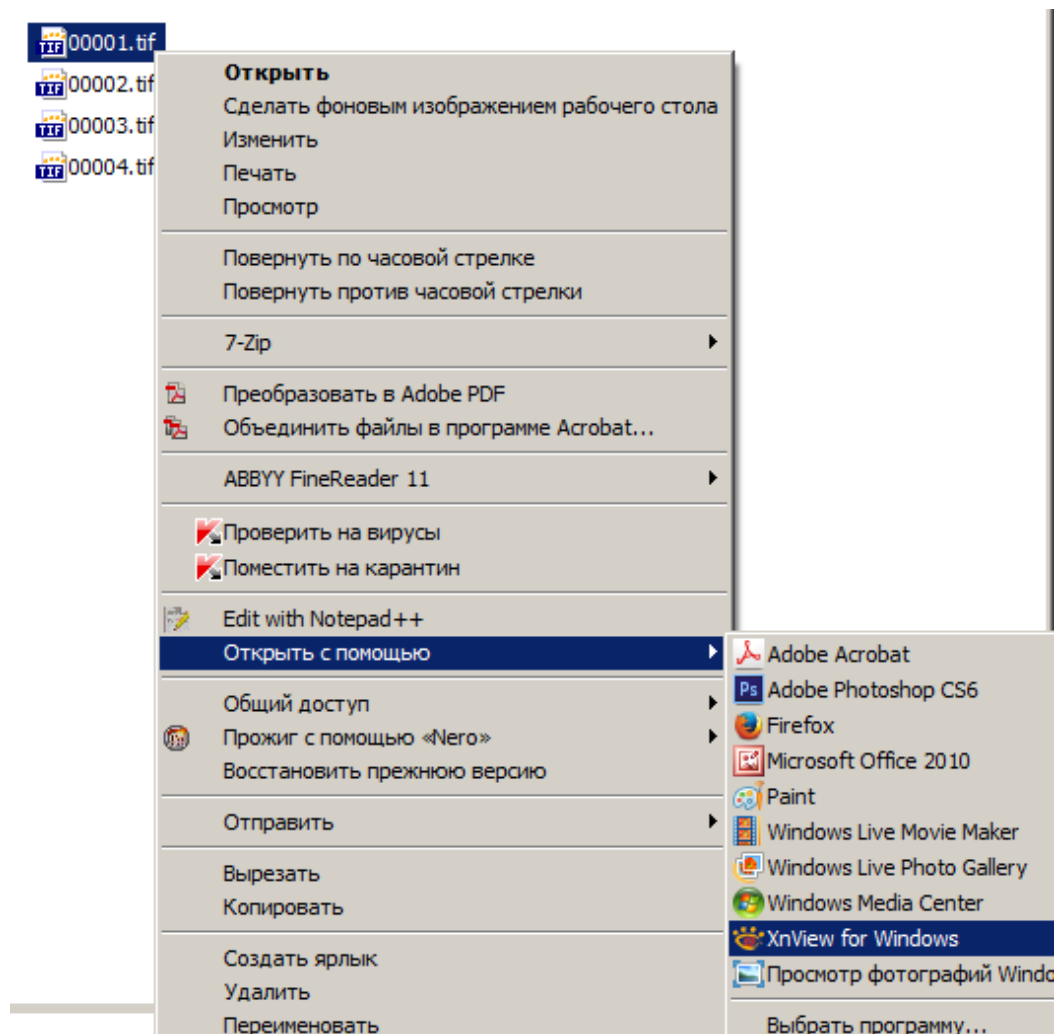
папка «scan» для файлов после сканирования,

папка «tiff» для обработанных сканов (пользовательских копий),

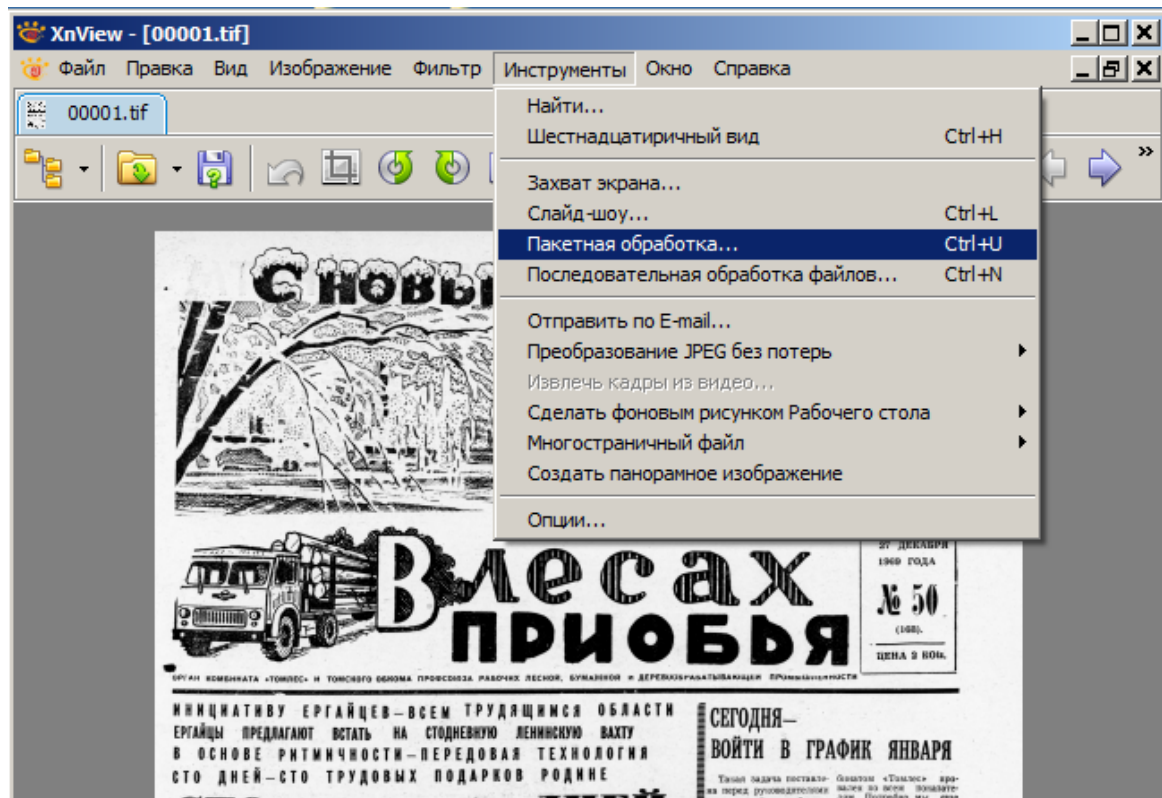
папка «jpg» для копии оптимизированной для web.

Конвертирование пользовательской копии в формат JPG в пакетном режиме (создание оптимизированной копии в автоматическом режиме документа/издания целиком) с помощью программы **XnView**:

- 1) Открыть с помощью программы **XnView** первый файл готовой пользовательской копии в формате TIFF LZW;

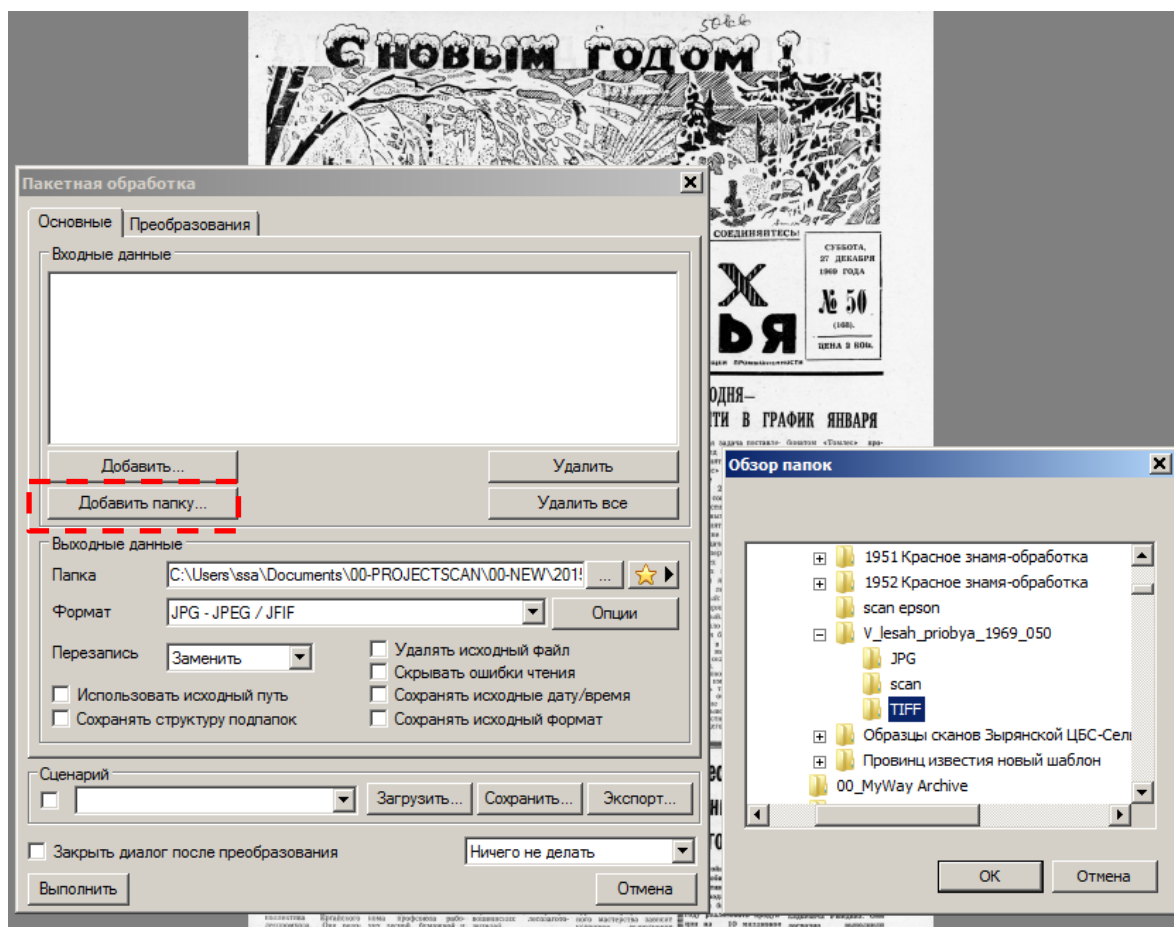


2) Выбрать меню **Инструменты – Пакетная обработка**;

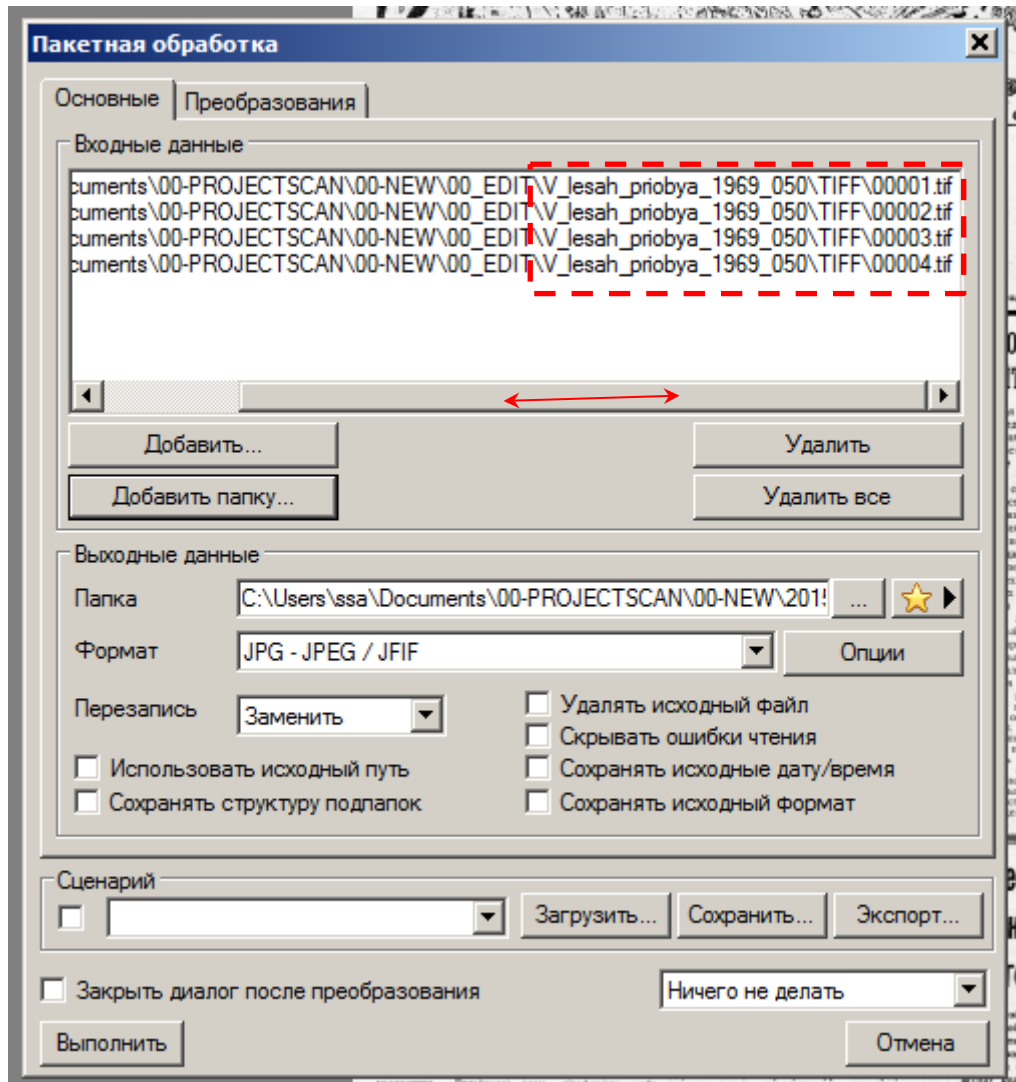


3) В появившемся окне **Пакетная обработка** во вкладке **Основные**:

Используя кнопку **Добавить папку...**, в появившемся окне **Обзор папок** указать папку с Пользовательской копией (TIFF LZW);

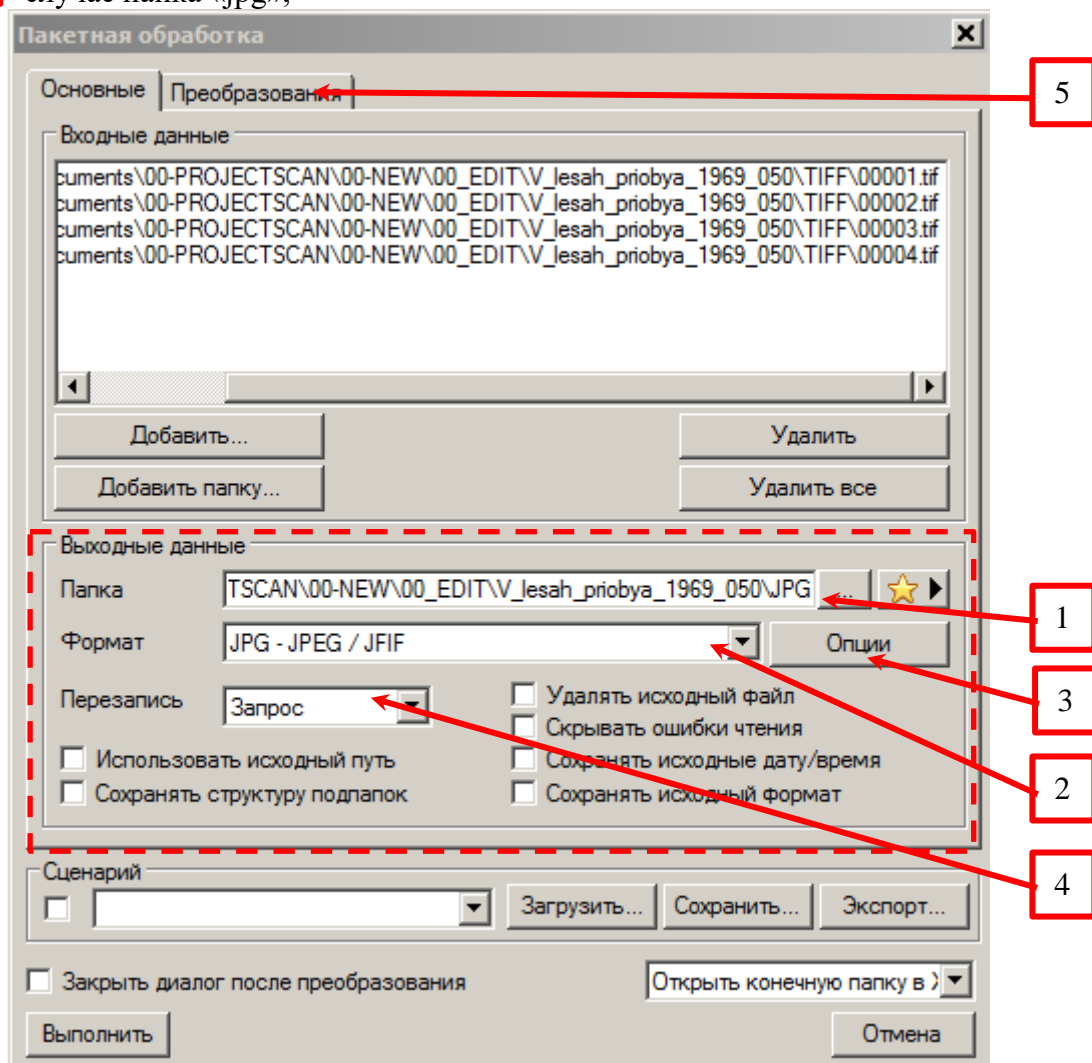


В окне **Входные данные** в виде списка будут перечислены файлы из выбранной папки с указанием места расположения их на вашем ПК (файлы Пользовательской копией в формате TIFF LZW);



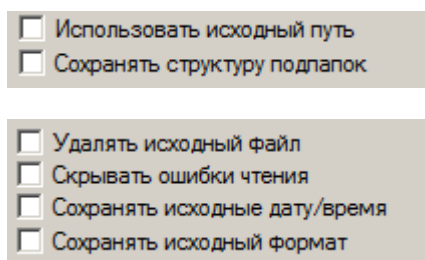
В области **Выходные данные**, указываем:

- 1 место для сохранения оптимизированной пользовательской копии, в вашем случае папка «jprg»;



- 2 устанавливаем формат **JPG – JPEG**;

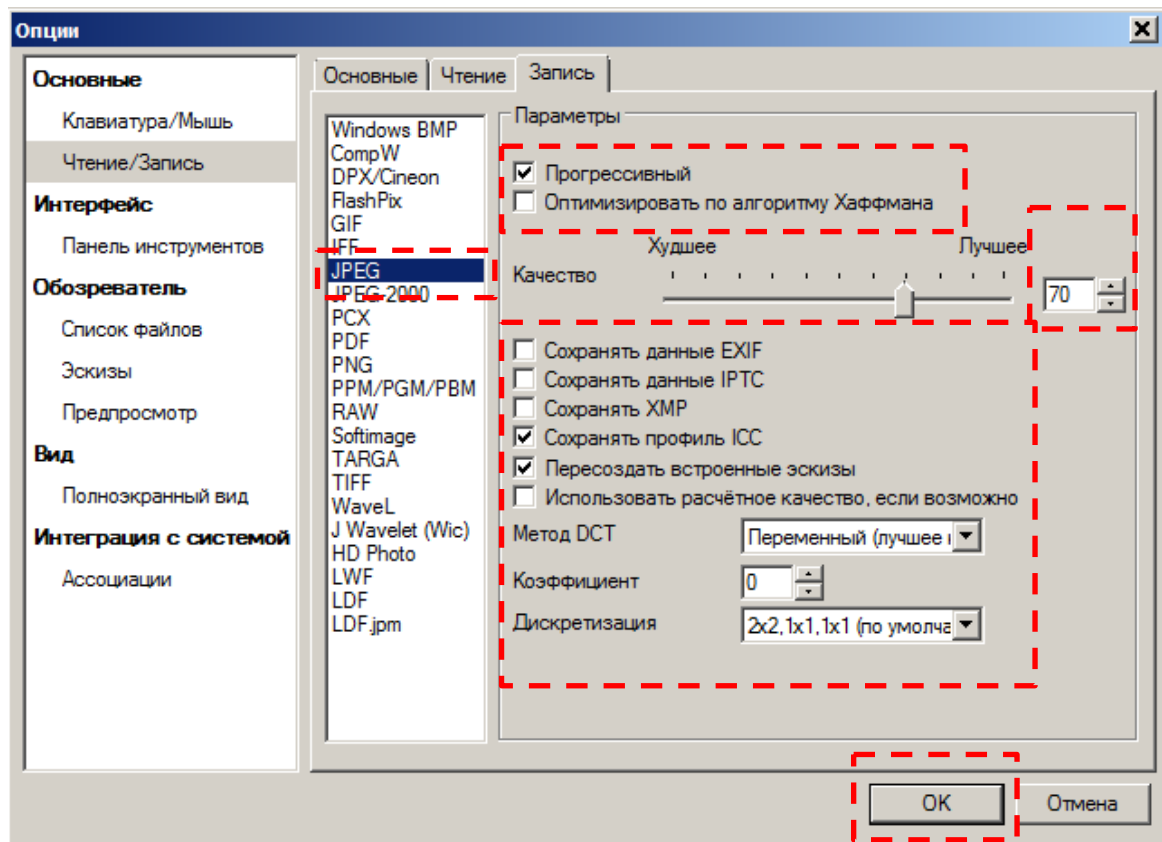
- 4 в поле **Перезапись** выбираем из списка **Запрос**;



все опции отключены, в белых клеточках «крюжички» отсутствуют;

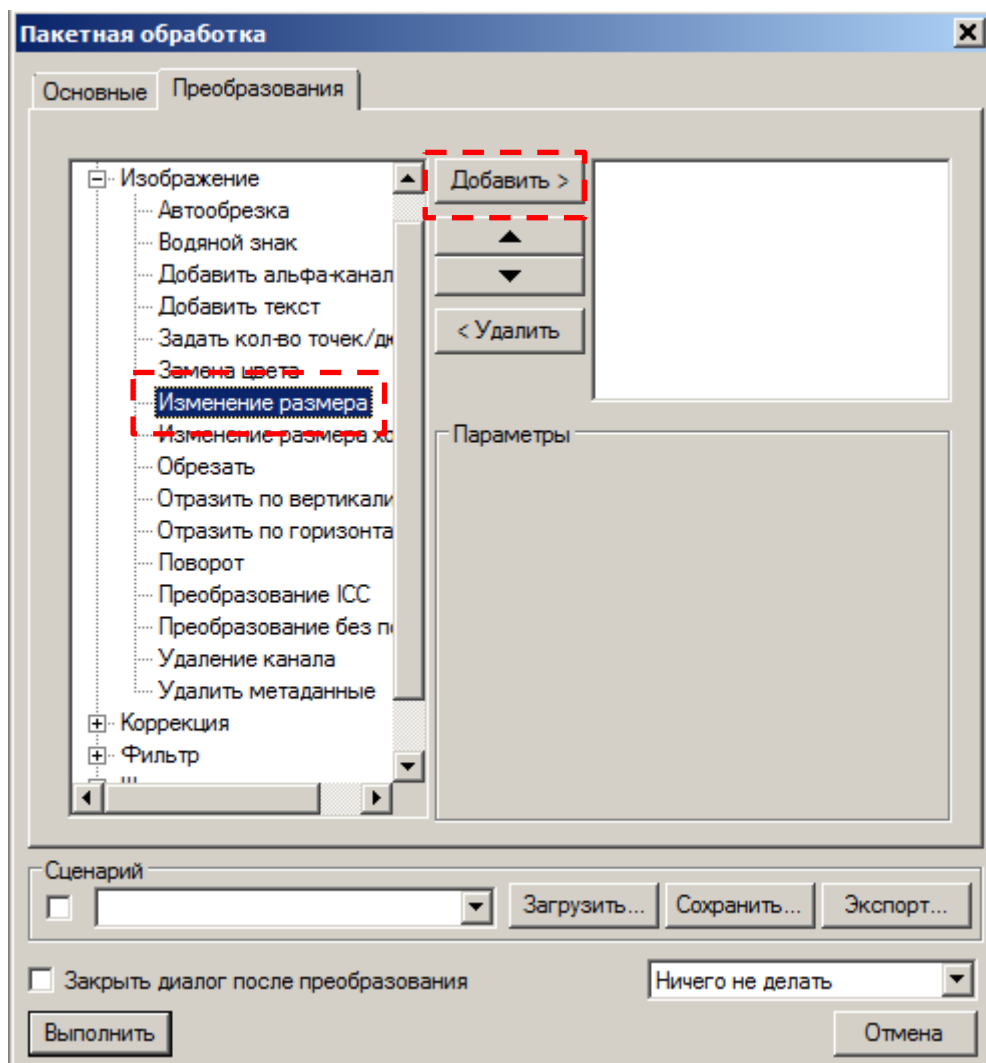
- 3

задаем **Опции** (как на рисунки ниже) и подтверждаем кнопкой **ОК**;

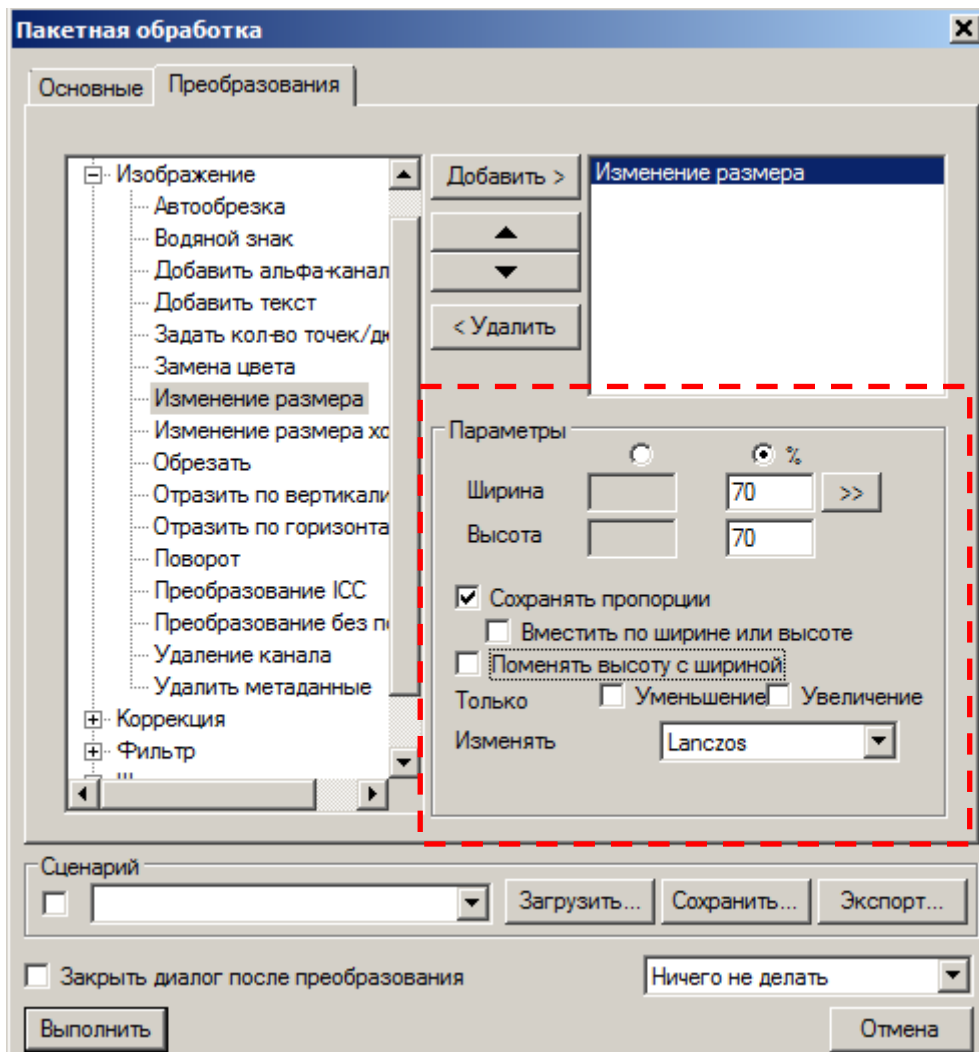


5

Перейти во вкладку **Преобразования** в левой области окна выбрать **Изменение размера** и нажать кнопку **Добавить**;




В области **Параметры** появятся настройки, которые устанавливаем вручную (как на рисунке ниже)

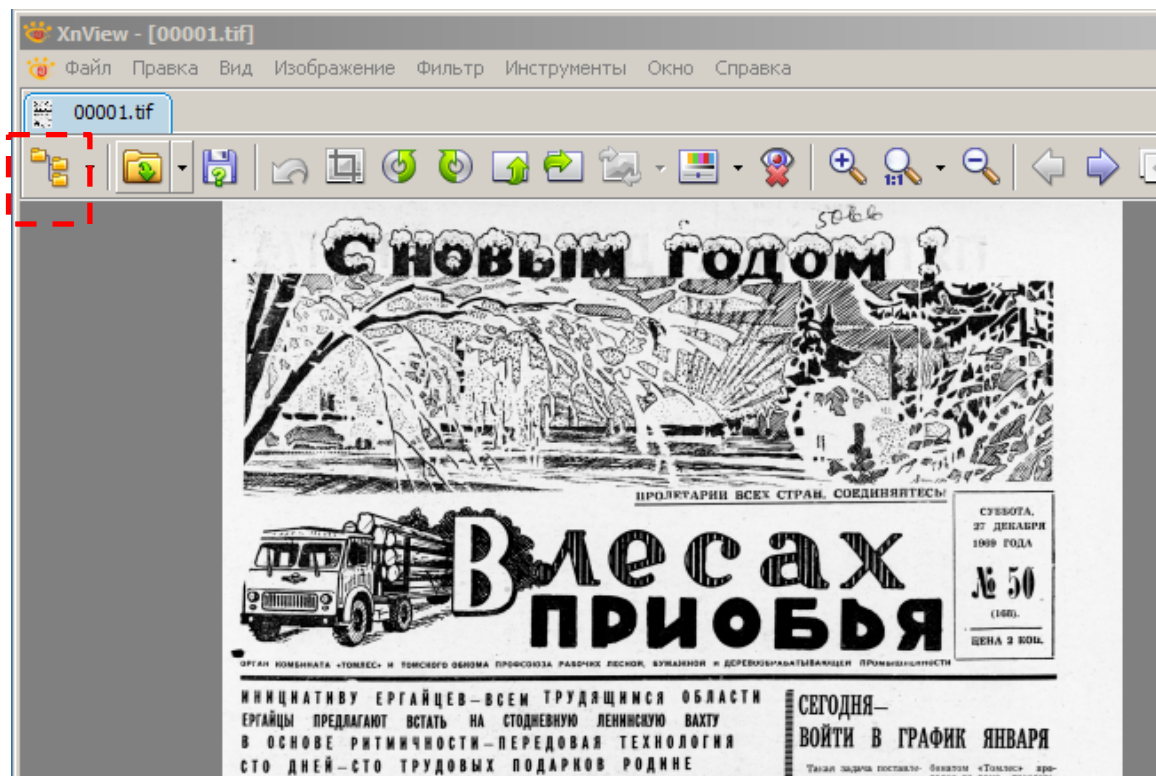


Последний шаг – нажать кнопку **Выполнить**.

Последний этап – контроль качества готовой пользовательской копии, оптимизированной для WEB.

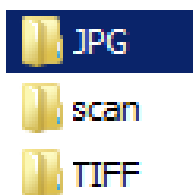
В окне программы XnView переходим в режим **Обозреватель**

(с помощью кнопки  или через меню **Файл – Обозреватель**)

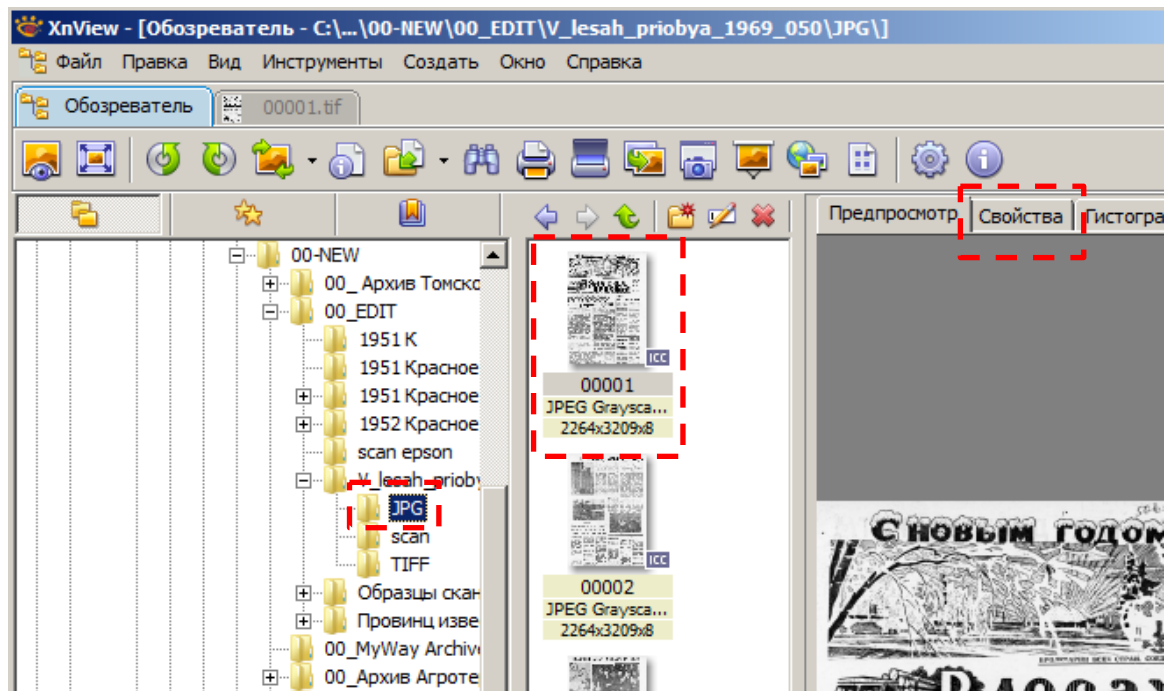


Готовая оптимизированная для web копия, в данном случае была сохранена при конвертации в папку «JPG» (кол-во файлов в папке должно соответствовать количеству страниц документа):

[V_lesah_priobya_1969_050](#)



Выбираем папку «JPG», выделяем файл в области эскизов, открываем вкладку **Свойства**

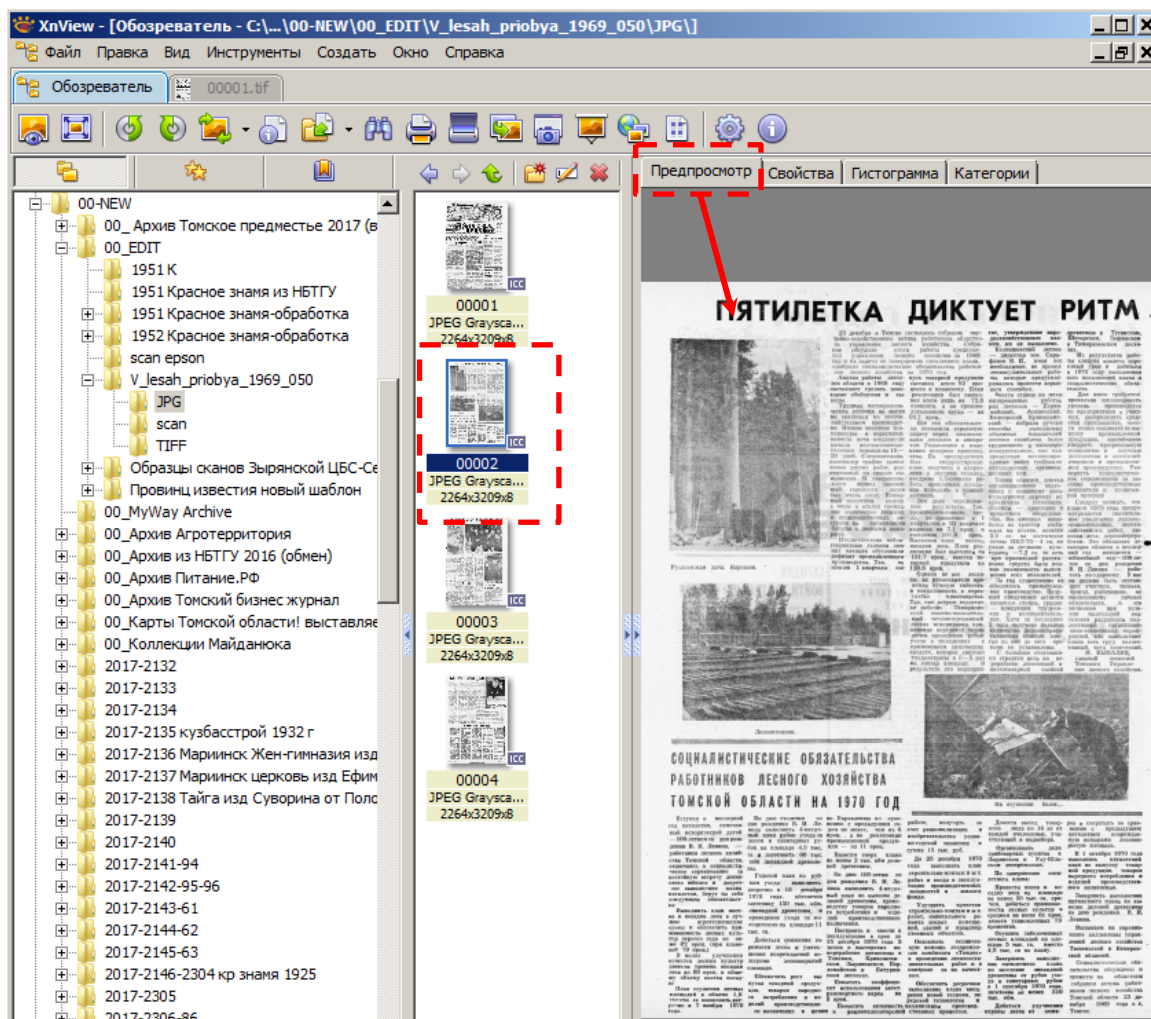


Во вкладке **Свойства** необходимо проконтролировать у каждого изображения:

- Формат изображения (формат файла) – **JPEG Grayscale** (если сканировали в режиме **Градации Серого**);
- Ширина и высота изображения в пикселах уменьшились до 70%;
- Кол-во бит – **8** (для файлов, сканированных в режиме Серого/ Grayscale);
- Точек/дюйм – **300x300** (300 dpi);
- Сжатие – **JPEG**.

Предпросмотр	Свойства	Гистограмма	Категории
<input type="checkbox"/>	Файл		
	Имя файла	00001.jpg	
	Описание	Файл "JPG"	
	Размер файла	1 483 443	
	Создан	01.08.2016 16:34	
	Изменён	01.09.2017 14:53	
	Открыт	01.09.2017 14:53	
<input type="checkbox"/>	Изображение		
	Формат	JPEG Grayscale (v 1. 1)	
	Ширина	2264	
	Высота	3209	
	Кол-во бит	8	
	Цветовая модель	RGB	
	Точек/дюйм	300 x 300	
	Размер при печати	19.17 x 27.17 см, 7.55 x 10.70 дюймов	
	Сжатие	JPEG	
	Кол-во изображений	1	
	Местоположение	Верх-Слева	
	Цветовой профиль	sGray	
<input type="checkbox"/>	Дополнительно		

Открыв **Предпросмотр**, можно визуально оценить качество изображения.



Составители:
Старкова Светлана Александровна,
Ретивых Виолетта Геннадьевна,
Отдел Электронной библиотеки Томской ОУНБ им. А.С. Пушкина,
тел.: 8 (3822) 51-38-06, e-mail: ssa@lib.tomsk.ru

Дата редакции 01.04.2020 г.