МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ
Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Пензенский государственный университет» (ПГУ)

Обеспечение сохранности архивных документов с использованием информационных технологий

Методические указания
к выполнению лабораторных работ

С о с т а в и т е л и :
Н. Н. Надеева, М. А. Катышева

Пенза
Издательство ПГУ
2018

УДК 681.3
О-14

Р е ц е н з е н т

кандидат технических наук, доцент кафедры «Прикладная информатика»
Пензенского государственного технологического университета

Ю. Е. Усачев

О-14 **Обеспечение сохранности архивных документов с использованием информационных технологий** : метод. указания к выполнению лаб. работ / сост: Н. Н. Надеева, М. А. Катышева. – Пенза : Изд-во ПГУ, 2018. – 36 с.

Работа нацелена на освоение компетенций обучающихся в области сохранности архивного фонда и реставрации архивных документов с использованием информационных технологий.

Издание подготовлено на кафедре «Информационное обеспечение управления и производства» ПГУ и предназначено для обучающихся по направлению подготовки 46.03.02 «Документоведение и архивоведение».

УДК 681.3

© Пензенский государственный университет, 2018

СОДЕРЖАНИЕ

Лабораторная работа № 1. Освоение элементарной знаниевой компетенции «Знание нормативно-правового и методического обеспечения сохранности архивного фонда и реставрации архивных документов» 4
Лабораторная работа № 2. Освоение элементарной навыковой компетенции «Способность преобразовать бумажный документ в электронную форму при помощи сканера» 11
Лабораторная работа № 3. Освоение элементарной деятельностной компетенции «Способность распознать и откорректировать текстовый документ с помощью программы Fine Reader» 18
Библиографический список 30
Приложение А 31
Приложение Б 34

Лабораторная работа № 1
Освоение элементарной знаниевой компетенции «Знание нормативно-правового и методического обеспечения сохранности архивного фонда и реставрации архивных документов»

Цель работы – закрепить полученный на лекционных занятиях необходимый минимум теоретических знаний в области сохранности архивного фонда и реставрации архивных документов. Предполагается закрепление знаниевых и навыковых компетенций студентов, сформированных в рамках дисциплин «Информационные технологии в профессиональной деятельности», «Информационные технологии» и «Информационные технологии в ДОУ и архивном деле», а также освоение элементарной знаниевой компетенции (ЭЗК) «Знание нормативно-правового и методического обеспечения сохранности архивного фонда и реставрации архивных документов» (таблица 1.1).

Таблица 1.1 – Паспорт осваиваемой ЭЗК

Содержание ЭЗК	**Знание нормативно-правового и методического обеспечения сохранности архивного фонда и реставрации архивных документов**
1	2
Виды деятельности, в рамках которых данная ЭЗК востребована	Организационно-управленческая и научно-исследовательская
Задачи, для решения которых владение данной ЭЗК необходимо	**Задача 1** (*связана с организационно-управленческой деятельностью*): правовое обеспечение сохранности архивного фонда и реставрации архивных документов. **Задача 2** (*связана с научно-исследовательской деятельностью*): составление рекомендаций по правовому обеспечению сохранности архивного фонда и реставрации архивных документов
ЗК, на базе которых формируется данная ЭЗК	**ПК-8** – способность анализировать ценность документов с целью их хранения

Продолжение таблицы 1

1	2
ЗК, которые являются необходимыми для овладения данной ЭЗК	**ПК-22** – способность принимать участие в работе по проведению экспертизы ценности документов
Нормативные акты, необходимые для овладения (освоения) данной ЭЗК	– Федеральный закон «Об информации, информационных технологиях и о защите информации» от 27.07.2006 № 149-ФЗ; – Федеральный закон «Об архивном деле в Российской Федерации» от 22.10.2004 № 125-ФЗ; – Правила организации хранения, комплектования, учета и использования документов Архивного фонда Российской Федерации и других архивных документов в органах государственной власти, органах местного самоуправления и организациях, утв. Приказом Министерства культуры РФ от 31 марта 2015 г. № 526; – Правила организации хранения, комплектования и использования документов Архивного фонда Российской Федерации и других архивных документов в государственных и муниципальных архивах, музеях и библиотеках, организациях Российской академии наук, утв. Приказ Министерства культуры и массовых коммуникаций РФ от 18.01.2007 № 19; – ГОСТ Р 7.0.97–2016. СИБИД. ОРД. Требования к оформлению документов; – ГОСТ Р 7.0.8–2013. СИБИД. Делопроизводство и архивное дело. Термины и определения; – ГОСТ 7.50–2002. Межгосударственный стандарт. Система стандартов по информации, библиотечному и издательскому делу. Консервация документов. Общие требования; – Разработка методов восстановления угасающих хроматических (цветных) текстов архивных документов с использованием цифровых компьютерных технологий (Метод. рекомендации. М. : ВНИИДАД, 2014)
Название предмета (дисциплины), для которого эта ЭЗК является исходной компетенцией	«Архив организации», «Технологии оцифровки документов», «Технологии создания электронных архивов»
Название предмета (дисциплины), для которого эта ЭЗК является целевой компетенцией	«Обеспечение сохранности и реставрации документов информационными технологиями»

Продолжение таблицы 1

1	2
Критерии эффективности	Определен набор основных инструментов СПС «Консультант Плюс» для поиска нормативных документов. Набор основных инструментов СПС «Консультант Плюс» обоснован в зависимости от вида нормативного документа. Составлен список нормативно-правовых и методических документов, регламентирующих обеспечение сохранности архивного фонда и реставрацию архивных документов

Сохранность документов является необходимым условием обеспечения доступности информации для пользователей архивов.

Под **сохранностью документа** понимают его состояние, которое характеризуется степенью удержания эксплуатационных свойств (прочности, эластичности, износостойкости составляющих документ материалов).

Сохранность – это показатель пригодности документа для использования и постоянного хранения. Чаще всего термин «сохранность» употребляется для обозначения специальных мер по обеспечению необходимого уровня безопасности, а именно: поддержание стабильного режима хранения, нормативное размещение фондов, соблюдение правил использования документов и предупреждение чрезвычайных ситуаций.

Обеспечение сохранности документов – комплексная проблема, реализующаяся на разных структурных уровнях: от государственного до локального. Она предусматривает решение многих и разнообразных вопросов: правовых, финансовых, научно-методических, материально-технических, кадровых, организационно-управленческих.

Обеспечить сохранность архивных фондов как части культурного наследия и информационного ресурса страны возможно лишь при развитии следующих направлений деятельности:

– поддержание нормативного физического и санитарно-гигиенического состояния зданий, инженерных коммуникаций, помещений хранилищ;

– расширение площадей хранилищ в соответствии с ростом библиотечных фондов, строительство и реконструкция библиотечных зданий;

– оборудование архивов современными комплексами технических средств безопасности, их постоянное совершенствование и обеспечение бесперебойной работы;

– постоянное совершенствование материально-технической базы библиотек для обеспечения нормативного режима хранения документов, их стабилизация и реставрация;

– обеспечение процессов защиты, хранения и использования библиотечных фондов достаточным количеством квалифицированных кадров;

– создание единого страхового фонда документов архивов;

– четкая организация работы хранителей, служб консервации фондов и безопасности, контроль за безукоризненным выполнением ими своих должностных обязанностей в процессе сохранения и использования фондов.

На государственном уровне нормативно-правовое обеспечение осуществляется через подготовку и утверждение федеральными органами исполнительной власти нормативно-правовых документов, таких как:

1. **Федеральный закон «Об архивном деле в Российской Федерации»** от 22 октября 2004 г. № 125-ФЗ (изм. от 23 мая 2016 г. № 149-ФЗ, от 18 июня 2017 г. № 127-ФЗ). Обеспечению сохранности архивных документов посвящена **глава 4 «Хранение и учет архивных документов»**, а в **частности *статья 17 «Обязанности государственных органов, органов местного самоуправления, организаций и граждан, занимающихся предпринимательской деятельностью без образования юридического лица, по обеспечению сохранности архивных документов»***, в которой сказано:

1) государственные органы, органы местного самоуправления, организации и граждане, занимающиеся предпринимательской деятельностью без образования юридического лица, обязаны обеспечивать сохранность архивных документов, в том числе документов по личному составу, в течение сроков их хранения, установленных федеральными законами, иными нормативными правовыми актами Российской Федерации, а также перечнями документов, предусмотренными ч. 3 ст. 6 и ч. 1 и ч. 1.1 ст. 23 настоящего Федерального закона;

2) уничтожение документов Архивного фонда Российской Федерации запрещается;

3) в отношении особо ценных документов, в том числе уникальных документов, устанавливается **особый режим учета, хранения и использования. Создаются страховые копии этих документов**;

4) порядок отнесения документов Архивного фонда Российской Федерации к особо ценным документам, в том числе уникальным документам, порядок учета таких документов, создания и хранения их страховых копий определяются уполномоченным федеральным органом исполнительной власти в сфере архивного дела и делопроизводства.

2. **ГОСТ 7.50–2002 Межгосударственный стандарт. Система стандартов по информации, библиотечному и издательскому делу.**

Консервация документов. Общие требования. Устанавливает общие требования к консервации документов, в том числе к режиму хранения, технологическим процессам стабилизации, реставрации, изготовлению копий и используемым при этом материалам. Распространяется на документы, выполненные на бумаге, коже, пергаменте.

Инструктивно-методическое обеспечение осуществляется посредством таких документов, как:

1. **Правила организации хранения, комплектования, учета и использования документов Архивного фонда Российской Федерации и других архивных документов в органах государственной власти, органах местного самоуправления и организациях,** утв. Приказом Министерства культуры РФ от 31 марта 2015 г. № 526, в которых:

1) определен порядок хранения, комплектования, учета и использования документов Архивного фонда РФ и других подобных материалов в органах госвласти, местного самоуправления и организациях.

2) урегулированы вопросы создания архивных фондов.

3) предусмотрены требования к хранилищам документов.

4) закреплены правила использования документов Архивного фонда РФ и других архивных материалов организации.

2. **Разработка методов восстановления угасающих хроматических (цветных) текстов архивных документов с использованием цифровых компьютерных технологий** (Метод. рекомендации. М. : ВНИИДАД, 2014).

Предложены цифровые технологии компьютерной растровой графики на основе программы Adobe Photoshop при восстановлении выцветающих хроматических (цветных) текстов архивных документов.

В пособии на основе сведений по характеристике, анализу и синтезу цвета в моделях графического редактора рассматриваются собственно методики и процедуры технологических процессов. На основе принципов трехкомпонентной теории цвета обосновывается выбор цветовой модели, цветового канала и цвета в системах RGB и СМУК.

3. **Методические рекомендации по электронному копированию архивных документов и управлению полученным информационным массивом** (М. : ВНИИДАД, 2012).

В работе на основе проведенных исследований, практического опыта и анализа разнообразных источников информации представлены методические рекомендации по созданию электронных копий подлинников архивных документов на бумажной основе, их учету, хранению и управлению полученным цифровым контентом.

4. **Восстановление угасающих текстов и изображений архивных документов** (Метод. рекомендации. М. : ВНИИДАД, 2006).

В разделе 3 методических рекомендаций **«*Применение цифровых компьютерных технологий при восстановлении угасающей информации*»** описываются:

– физические и программные принципы сканирования и обработки изображений;

– цифровые технологии репродуцирования;

– требования к компьютерным компонентам;

– предварительное сканирование и корректировка параметров изображения по тону и цвету;

– особенности цифрового репродуцирования в различных диапазонах спектра;

– редактирование импортируемых изображений документов;

– возможности корректирования изображений программами компьютерной графики;

– вывод откорректированных документов на бумажный носитель.

Задания и контрольные вопросы

Задания

1. Запустить СПС «Консультант Плюс».
2. Создать в СПС «Консультант Плюс» последовательно группу папок, указав в их наименовании номер группы, название дисциплины, номер бригады, номер лабораторной работы.
3. Произвести поиск нормативных актов, регламентирующих сохранность архивного фонда и реставрацию архивных документов.
4. Произвести экспорт найденных документов в MS Word.
5. Сохранить результаты работы в созданную папку в СПС «Консультант Плюс».
6. Перечислить письменно способы реставрации документов согласно ГОСТ 7.50–2002. **Межгосударственный стандарт. Система стандартов по информации, библиотечному и издательскому делу. Консервация документов. Общие требования.**
7. Оформить отчет о проделанной работе (приложение Б).

Контрольные вопросы

1. Что понимается под сохранностью документа?
2. Как обеспечивается сохранность документов на государственном уровне?
3. Перечислите основные направления деятельности по обеспечению сохранности архивных фондов.
4. Какие нормативные акты регламентируют сохранность архивного фонда Российской Федерации?
5. Какие методические документы регламентируют реставрацию архивных документов при помощи информационных технологий?

Лабораторная работа № 2
Освоение элементарной навыковой компетенции «Способность преобразовать бумажный документ в электронную форму при помощи сканера»

Цель работы – закрепить полученный на лекционных занятиях необходимый минимум теоретических знаний в области преобразования бумажного документа в электронную форму при помощи сканера. Предполагается закрепление знаниевых и навыковых компетенций студентов, сформированных в рамках дисциплин «Информационные технологии в профессиональной деятельности», «Информационные технологии» и «Информационные технологии в ДОУ и архивном деле», а также освоить элементарную навыковую компетенцию «Способность преобразовать бумажный документ в электронную форму».

Начнем с изучения паспорта этой элементарной навыковой компетенции (ЭНК) (таблица 2.1). Она является элементарной по отношению к компетенции ОПК-2 (из ФГОС-3+): «Владение базовыми знаниями в области информационных технологий (программные продукты, используемые в управлении документами, системы электронного документооборота, технологии сканирования документов)».

Таблица 2.1 – Паспорт осваиваемой ЭНК

Содержание ЭНК	**Способность преобразовать бумажный документ в электронную форму при помощи сканера**
1	2
Виды деятельности, в рамках которых данная ЭНК востребована	Организационно-управленческая и научно-исследовательская
Задачи, для решения которых владение данной ЭНК необходимо	**Задача 1** (*связана с организационно-управленческой деятельностью*): получение навыков сканирования архивных документов. **Задача 2** (*связана с научно-исследовательской деятельностью*): составление рекомендаций по преобразованию бумажного документа в электронную форму
ЗК, на базе которых формируется данная ЭНК	**ПК-8** – способность анализировать ценность документов с целью их хранения
НК, которые являются необходимыми для овладения данной ЭНК	**ПК-22** – способность принимать участие в работе по проведению экспертизы ценности документов

Продолжение таблицы 2.1

1	2
Нормативные акты, необходимые для овладения (освоения) данной ЭНК	- Федеральный закон «Об информации, информационных технологиях и о защите информации» от 27.07.2006 № 149-ФЗ; - Федеральный закон «Об архивном деле в Российской Федерации» от 22.10.2004 № 125-ФЗ; - Правила организации хранения, комплектования и использования документов Архивного фонда Российской Федерации и других архивных документов в государственных и муниципальных архивах, музеях и библиотеках, организациях Российской академии наук, утв. Приказом Министерства культуры и массовых коммуникаций РФ от 18.01.2007 № 19; - ГОСТ Р 7.0.97–2016. СИБИД. ОРД. Требования к оформлению документов; - ГОСТ Р 7.0.8–2013. СИБИД. Делопроизводство и архивное дело. Термины и определения; - ГОСТ 7.50–2002. Межгосударственный стандарт. Система стандартов по информации, библиотечному и издательскому делу. Консервация документов. Общие требования; - Разработка методов восстановления угасающих хроматических (цветных) текстов архивных документов с использованием цифровых компьютерных технологий (Метод. рекомендации. М. : ВНИИДАД, 2014)
Название предмета (дисциплины), для которого эта ЭНК является исходной компетенцией	«Архив организации», «Технологии оцифровки документов», «Технологии создания электронных архивов»
Название предмета (дисциплины), для которого эта ЭНК является целевой компетенцией	«Обеспечение сохранности и реставрации документов информационными технологиями»
Критерии эффективности	Определен набор основных правил и методов сканирования документа. Набор правил и методов сканирования обоснован в зависимости от вида документа, который подлежит оцифровке

К аппаратным средствам получения цифровых растровых оригиналов в основном относятся **сканеры** и **цифровые фотокамеры**.

Сканер – это устройство ввода данных, получаемых путем анализа (разбиения на составляющие) образов изображений. Сканеры подразделяются на двухуровневые, полутоновые и цветные. Двухуровневые (бинарные) работают лишь с двумя уровнями яркости – черным и белым и могут быть использованы для сканирования штриховых рисунков, например чертежей.

При полутоновом способе воспроизведения каждый пиксель имеет несколько возможных уровней яркости (16, 64, 256 градаций серого цвета). Можно использовать для сканирования черно-белых фотографий. Цветные сканеры на сегодняшний день получили массовое применение.

Важным показателем сканера является **разрядность**, которая показывает количество одновременно сканируемых цветов и измеряется в битах: 24 бит – 16,7 млн цветов, 30 бит – 1 млрд.

Основным показателем сканера является **разрешающаяся способность (РС)** и измеряется в точках на дюйм (dpi). РС сканера подразделяется на оптическое (реальное) и программное. Оптическое разрешение показывает предельные возможности светочувствительного элемента сканера: впоследствии программными методами сканер может повысить качество изображения и, соответственно, разрешение.

Разрешение сканера имеет два показателя – по горизонтали и вертикали. Например, 600×300, 600×600, 800×800 dpi. Однако чаще всего употребляется первое значение – 100, 300, 400, 1200 dpi.

Выбор РС зависит от размеров штрихов текста. В таблице 2.2 приведены размеры штрихов рукописных и машинописных текстов.

Таблица 2.2 – Размеры штрихов текста, которые наиболее часто встречаются в документах

Вид текста	Перо	Шариковая ручка	Карандаш	Машинописный текст
Толщина штрихов, мкм	40–400	100–500	300–700	200–500

Из таблицы 2.2 следует, что самые тонкие детали штрихов принадлежат тексту, нанесенному пером. Сканер может предлагать различную оптическую РС (100; 150; 200; 300; 600; 2400; 3600 пикселей на дюйм). В таблице 2.3 приводятся размеры пикселей в зависимости от РС.

Таблица 2.3 – Размеры пикселей в зависимости от РС

РС, dpi	100	150	200	300	600	1200	2400
Пиксель, мкм	250	170	125	85	42	21	10

Из табл. 2.2 и 2.3 следует, что для полного воспроизведения штрихов текста в зависимости от их вида можно сканировать с РС, равной от 200 до 600 dpi. При этом надо иметь в виду, что с увеличением РС заметно растет размер файла. Это наглядно видно из таблицы 2.4.

Таблица 2.4 – Разрешение и размер файла при сканировании документа размером А4

Цветовая Модель	**Разрешение сканирования, dpi**						
	100	150	200	300	600	1200	2400
Размер файла, Мб							
RGB	2,9	6,5	11,6	26	104	416,1	1700
Серая шкала	0,963	2,2	3,9	8,7	34,7	138,7	554,7

Сканирование – это перевод графической (рисунки, фотографии) или текстовой информации в цифровой вид. Чтобы отсканировать текст, картинку или фотографию, потребуется программное обеспечение (ПО) – графический редактор и сканер, который нужно включить до запуска программы.

Процесс перевода бумажного документа в электронный состоит из следующих шагов:

– сканирование документа;
– сегментация документа;
– распознавание документа;
– редактирование и проверка результата;
– сохранение документа.

Первый этап работы – *сканирование.* На этом этапе используют сканер. Чтобы начать сканирование, нужно включить сканер, положить оригинал документа и щелкнуть кнопку **Сканировать**  или в раскрывающемся списке подкоманд данной кнопки использовать пункт **Сканировать** (рисунок 2.1).

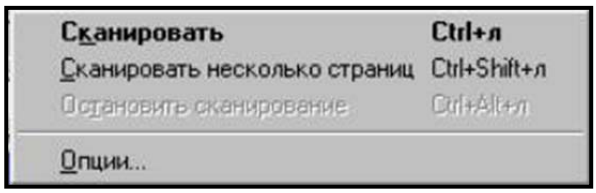


Рисунок 2.1 – Вид кнопки «Сканировать»

Программа способна работать со сканером как непосредственно, так и через драйвер ***TWAIN***. При непосредственном взаимодействии со сканером возможность сканирования цветных изображений не используется, так как предполагается, что текст в любом случае является одноцветным.

Программа использует для сканирования устройство, которое задано по умолчанию. Для того, чтобы выбрать такое устройство или изменить его настройку, нужно выбрать пункт **Опции**… в раскрывающемся списке кнопки **Сканировать** – откроется диалоговое окно **Опции** (рисунок 2.2).

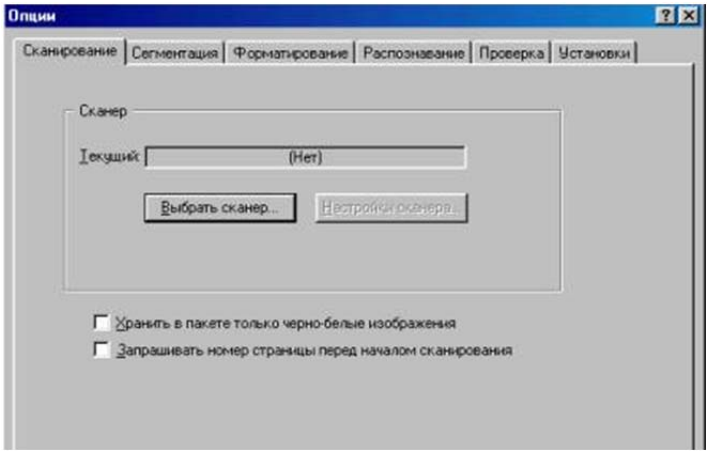


Рисунок 2.2 – Вид диалогового окна «Опции» на вкладке «Сканирование»

Если к компьютеру подключено несколько сканеров, или сканер был подключен после установки программы **Fine Reader**, следует щелкнуть на кнопке **Выбрать сканер**. В этом случае будет проведен поиск подключенных к компьютеру сканеров, это позволит выбрать нужный.

Для изменения настроек сканера используют кнопку **Настройка сканера**.

Когда сканер выбран, активизируются два флажка в нижней части диалогового окна. Если установить флажок **Показывать диалог TWAIN-драйвера**, то сканирование производится через протокол **TWAIN** с отображением специального диалогового окна (рисунок 2.3).

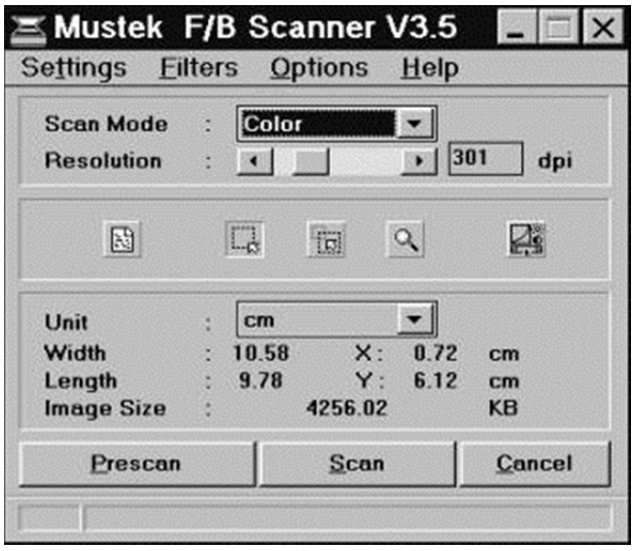


Рисунок 2.3 – Диалоговое окно драйвера TWAIN для сканера

Диалоговое окно интерфейса TWAIN может выглядеть по-разному – его вид определяется драйвером сканера. В нем с помощью пункта **Scan Mode** можно задать различные режимы сканирования:

– **Black&White** (черно-белый) – предназначен для сканирования текстов и контурных рисунков;

– **Gray (серый)** – 256 оттенков серого цвета, используют данный режим для сканирования черно-белых фотографий или для черно-белых принтеров;

– **Color (цветной)** – для получения цветных картинок и фотографий.

Использовать протокол ***TWAIN*** имеет смысл только в том случае, когда работа напрямую невозможна или дает некачественные результаты.

Кнопка **Scan** (сканировать, начать) запускает основной процесс сканирования, в результате которого мы видим на экране компьютера получившуюся фотографию. Остается сохранить ее на диске для дальнейшего использования.

Кнопка **Prescan** предназначена для пробного (предварительного) сканирования. Позволяет увидеть положение картинки и задать с помощью специального «пунктирного» прямоугольника (рамки) область сканирования.

Флажок **Показывать опции перед началом сканирования** из диалогового окна **Опции** (см. рисунок 2.2) применяют только в том случае, когда бумажные страницы документа существенно отличаются друг от друга по качеству (разные страницы печатались в разное время и разными средствами). В этом случае перед сканированием каждой страницы открывается диалоговое окно настройки сканера, чтобы пользователь мог отрегулировать качество процесса.

Сам процесс сканирования происходит в автоматическом режиме. Если требуется обработать много страниц, то лучше сначала все их отсканировать, а затем приступать к распознаванию. Это связано с тем, что сканирование требует присутствия пользователя из-за необходимости управления сканером (например, для смены страниц), а процесс распознавания может происходить в автоматическом режиме.

Когда процесс сканирования завершается, появляется окно с изображением текста.

Задания и контрольные вопросы

Задания

1. Включить сканер.
2. Запустить программу ***EPSON Scan***.

3. Отсканировать тексты, приведенные в приложении А:
– текст с рисунком;
– текст, разбитый на несколько колонок;
– сведения в табличном виде.
4. Сохранить каждый результат сканирования в отдельном пакете.
5. Оформить отчет о проделанной работе.

Контрольные вопросы

1. Из каких основных этапов состоит процесс сканирования?
2. Какой стороной нужно помещать «оригинал» в сканер?
3. Что такое «разрешение», в каких единицах оно измеряется?
4. Какие типы сканеров вы знаете?
5. Можно ли распознать фотографию текста, записанную в виде файла?
6. Какие режимы сканирования вы знаете?
7. Для чего предназначен «черно-белый» режим сканирования?
8. В каких случаях выбирается «цветной» режим сканирования?
9. В каких случаях используется «серый» режим сканирования?

Лабораторная работа № 3
Освоение элементарной деятельностной компетенции «Способность распознать и откорректировать текстовый документ с помощью программы Fine Reader»

Цель работы – закрепить полученный на лекционных занятиях необходимый минимум теоретических знаний в области распознавания отсканированных документов и отработать практические навыки, необходимые для работы в программе ***Fine Reader***. Предполагается закрепление знаниевых компетенций студентов, сформированных в рамках дисциплин «Информационные технологии в профессиональной деятельности», «Информационные технологии» и «Информационные технологии в ДОУ и архивном деле» и освоить элементарную деятельностную компетенцию «Способность распознать и откорректировать текстовый документ с помощью программы ***Fine Reader***».

Начнем с изучения паспорта этой элементарной деятельностной компетенции (ЭДК) (таблица 3.1). Она является элементарной по отношению к компетенции ОПК-2 (из ФГОС-3+): «Владение базовыми знаниями в области информационных технологий (программные продукты, используемые в управлении документами, системы электронного документооборота, технологии сканирования документов)».

Таблица 3.1 – Паспорт осваиваемой ЭДК

Содержание ЭДК	**Способность распознать и откорректировать текстовый документ с помощью программы Fine Reader**
1	2
Виды деятельности, в рамках которых данная ЭДК востребована	Организационно-управленческая и научно-исследовательская
Задачи, для решения которых владение данной ЭДК необходимо	**Задача 1** (*связана с организационно-управленческой деятельностью*): распознавание и корректировка текстовых документов с помощью программы ***Fine Reader***. **Задача 2** (*связана с научно-исследовательской деятельностью*): составление рекомендаций по распознаванию и корректировке текстовых документов с помощью программы ***Fine Reader***
ЗК, на базе которых формируется данная ЭДК	**ПК-8** – способность анализировать ценность документов с целью их хранения

Продолжение таблицы 3.1

1	2
НК, которые являются необходимыми для овладения данной ЭДК	**ПК-22** – способность принимать участие в работе по проведению экспертизы ценности документов
Нормативные акты, необходимые для овладения (освоения) данной ЭДК	- Федеральный закон «Об информации, информационных технологиях и о защите информации» от 27.07.2006 № 149-ФЗ; - Федеральный закон «Об архивном деле в Российской Федерации» от 22.10.2004 № 125-ФЗ; - Правила организации хранения, комплектования и использования документов Архивного фонда Российской Федерации и других архивных документов в государственных и муниципальных архивах, музеях и библиотеках, организациях Российской академии наук, утв. Приказом Министерства культуры и массовых коммуникаций РФ от 18.01.2007 № 19; - ГОСТ Р 7.0.97–2016. СИБИД. ОРД. Требования к оформлению документов; - ГОСТ Р 7.0.8–2013. СИБИД. Делопроизводство и архивное дело. Термины и определения; - ГОСТ 7.50–2002. Межгосударственный стандарт. Система стандартов по информации, библиотечному и издательскому делу. Консервация документов. Общие требования; - Разработка методов восстановления угасающих хроматических (цветных) текстов архивных документов с использованием цифровых компьютерных технологий (Метод. рекомендации. М. : ВНИИДАД, 2014)
Название предмета (дисциплины), для которого эта ЭДК является исходной компетенцией	«Архив организации», «Технологии оцифровки документов», «Технологии создания электронных архивов»
Название предмета (дисциплины), для которого эта ЭДК является целевой компетенцией	«Обеспечение сохранности и реставрации документов информационными технологиями»
Критерии эффективности	Определены критерии распознавания и корректировки документов, содержащих информацию различного вида. Набор критериев распознавания и корректировки документов обоснован в зависимости от вида документа, который подлежит оцифровке

Все операции, необходимые в ходе преобразования бумажного документа в электронную форму могут быть выполнены с помощью программы ***Fine Reader***.

Fine Reader – это система оптического распознавания текстов. Она преобразует полученное с помощью сканера графическое изображение (картинку) в текст (т.е. в коды букв, «понятные» компьютеру). Принцип действия этой программы следующий: сканируется любой текст, затем картинка текста преобразуется в «обычный электронный текст», такой, как если бы вы его напечатали с клавиатуры. Пользователю остается только сохранить текст на диске или скопировать его через буфер обмена в любой текстовый редактор.

Вся работа в программе ***Fine Reader*** ведется в пакетном режиме.

Пакет – это папка, в которой хранятся изображения и рабочие файлы программы. Каждое отсканированное изображение записывается как отдельная страница пакета.

Экран программы ***Fine Reader*** (рисунок 3.1) включает в себя строку меню, панели инструментов, окна **Пакет, Текст, Изображение**.

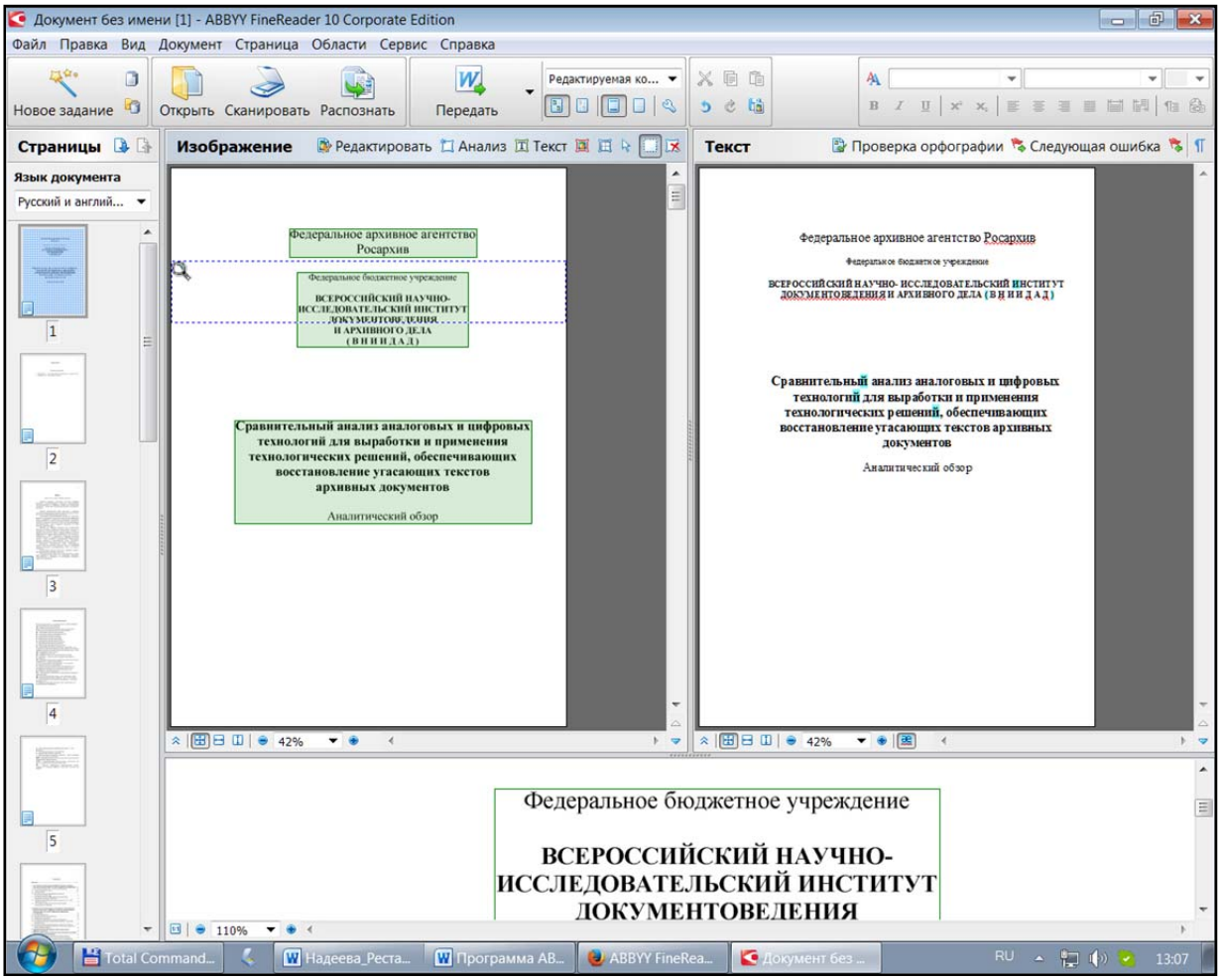


Рисунок 3.1 – Вид окна программы Fine *Reader 10*

Окно **Пакет** отражает страницы пакета.

Окно **Текст** – распознанный текст активной страницы окна **Пакет**.

Окно **Изображение** – графическое изображение (картинку) активной страницы окна **Пакет**.

В левой рабочей области располагается окно **Пакет**, содержащее список графических документов, которые должны быть преобразованы в текст. Эти графические файлы рассматриваются как части одного документа. Результаты их обработки в дальнейшем объединяются в текстовый файл. Форма значка, отмечающего исходные файлы, указывает, было ли произведено распознавание.

Окно **Крупный план** содержит фрагмент графического документа в увеличенном виде. С ее помощью можно оценить качество распознавания. Эту панель используют также при «обучении» программы в ходе распознавания текста. В окне **Крупный план** по умолчанию показывается черно-белое изображение независимо от того, какое именно (цветное, серое или черно-белое) изображение открывали.

Остальную часть рабочей области занимают окна документов (окно **Изображение**, окно **Текст**). Здесь располагается окно графического документа, подлежащего распознаванию, а также окно текстового документа, полученного после распознавания.

В верхней части окна приложения под строкой меню располагаются панели инструментов. На рисунке 3.2 включено отображение всех панелей, которые могут использоваться в программе ***Fine Reader 10***.

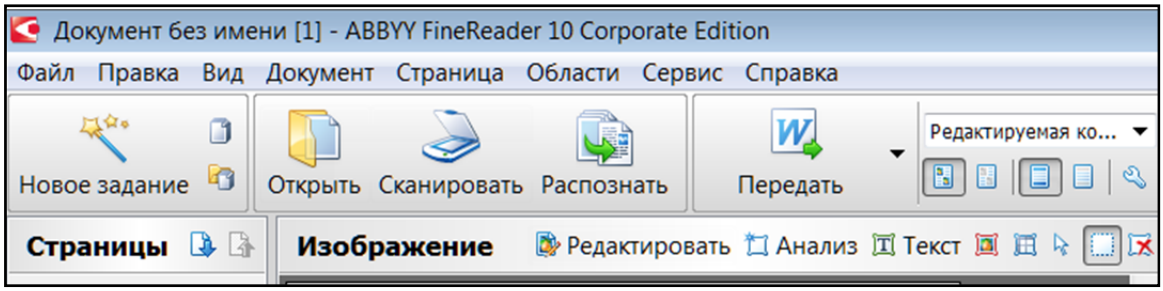


Рисунок 3.2 – Интерфейс программы *Fine Reader 10*

Панель инструментов **Стандартная** содержит кнопки для открытия документов и для операций с буфером обмена. Прочие кнопки этой панели служат для изменения представления документа. Панель **Инструменты** используют при работе с исходным изображением. В частности, она позволяет управлять сегментацией документа (разбиением страницы на блоки текста). С помощью элементов управления этой панели задают последовательность фрагментов текста в итоговом документе.

На панели инструментов программы находятся соответствующие кнопки: **Сканировать, Сегментировать, Распознать, Проверить, Сохранить**.

Прежде, чем приступить к распознаванию отсканированного документа, необходимо выполнить его ***сегментацию***. В бумажном документе, например на странице книги или журнала, текст не всегда располагается в фиксированном порядке. Он может размещаться в нескольких колонках (столбцах), содержать иллюстрации (и подписи к ним), другие элементы форматирования. Дополнительные врезки и данные, представленные в таблицах, также могут запутать естественный порядок текста. Поэтому, прежде чем включать текст в документ, его разбивают на *блоки*, содержащие цельные фрагменты. Блоки распознают последовательно. Полученный текст включается в документ в порядке нумерации блоков.

Сегментация – разбиение документа на блоки. Ручную установку блоков (рамка с помощью левой кнопки мыши) можно применить, если нужно распознать только часть текста.

Разные типы блоков обрабатываются программой по-разному. Программа ***Fine Reader 10*** поддерживает следующие типы блоков, которые можно выбрать из меню **Области** (рисунок 3.3).

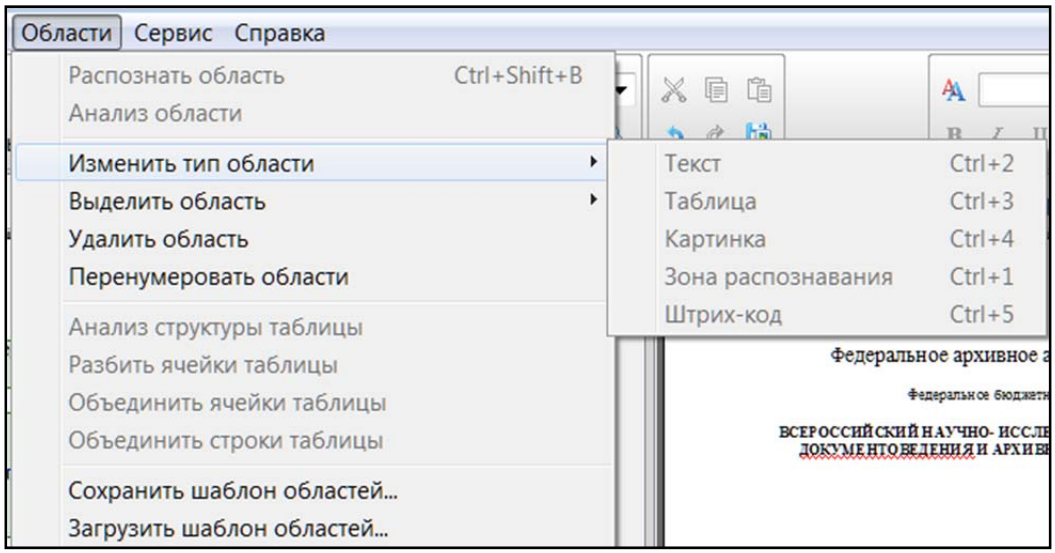


Рисунок 3.3 – Вид контекстно-зависимого меню «Области» при работе с блоками

– текстовый (**Текст**) – на этапе распознавания преобразуется в текст;

– табличный (**Таблица**) – представляет собой набор ячеек, каждая из которых преобразуется в текст по отдельности;

– изображение (**Картинка**) – включается в документ без изменений как графическая иллюстрация, если формат сохранения преобразованного документа допускает вставные объекты;

– лишний (**Зона распознавания**) – игнорируется;

– содержащий штрих-код (**Штрих-код**) – распознается как штрих-код.

При автоматической сегментации определение границ блока определяется автоматически. При этом учитываются поля документа, просветы между колонками, рамки.

Следующий этап работы программы – непосредственно ***распознавание***. Этот этап обычно не требует вмешательства пользователя, за исключением тех случаев, когда распознавание сопровождается «обучением».

Для большинства случаев сразу нужно нажать на кнопку инструмента **Распознать** 

и подтвердить автоматическое определение блоков. Процесс распознавания будет «иллюстрироваться» закраской участков текста.

Точность распознавания ***Fine Reader 10*** около 92–97 %, так как при печати книг и газет на бумаге иногда не прилипает или осыпается краска. Это будет хорошо видно в окне **Крупный план** с увеличенным масштабом. Результаты распознавания будут отражены в окне **Текст** (см. рисунок 3.1).

Последний этап – ***сохранение*** документа. Для этого необходимо из меню **Файл** выбрать соответствующую команду (рисунок 3.4). Можно сохранить распознанный текст или передать его в другую программу (например, MS Word) с помощью выбора соответствующего пункта из раскрывающегося списка подкоманд.

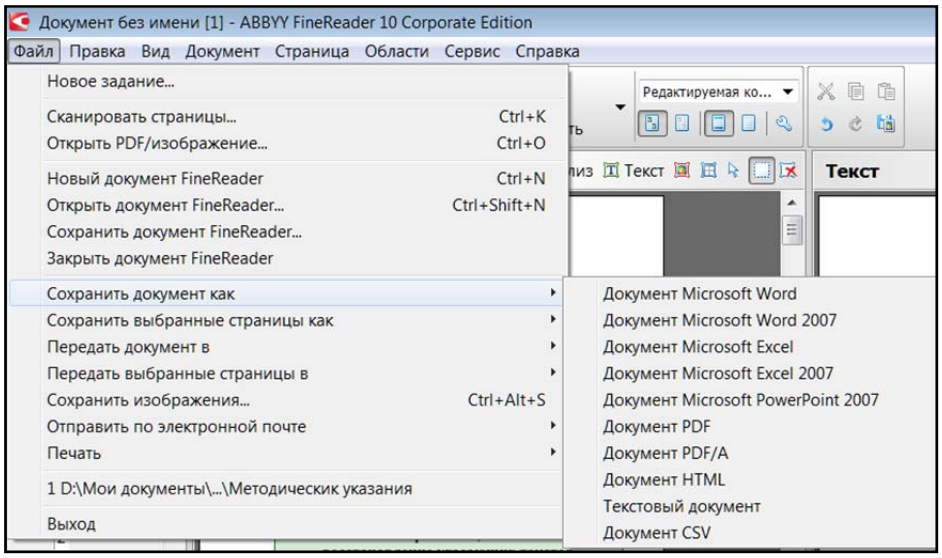


Рисунок 3.4 – Команды меню «Файл» для сохранения информации в различных форматах

Fine Reader сохраняет результаты распознавания в следующих форматах: ***.PDF, TXT, .DOC, .HTML, .CSV, .PDF/A***.

При сохранении в формате .doc Fine Reader 10 автоматически создает текстовые стили, которые называются FR1, FR2 и т.д.

Если сохраняется пакет для дальнейшей работы, результаты распознавания каждой страницы записываются в отдельный файл в расширением ***.FRF*** (***Fine Reader Format***).

Перед тем, как сохранять результаты распознавания в том или ином формате, нужно проверить, все ли нужные опции сохранения выставлены.

Корректировка текста документа в программе ***Fine Reader*** включает в себя следующее:

– проверка орфографии;

– выделение текста;

– копирование и перемещение выделенного текста;

– удаление текста;

– форматирование абзаца.

Для того, чтобы проверить орфографию, нужно нажать кнопку (рисунок 3.5). Чтобы вернуться к документу после того, как проверка закончена, нажать кнопку **Закрыть** в диалоге **Проверка**.

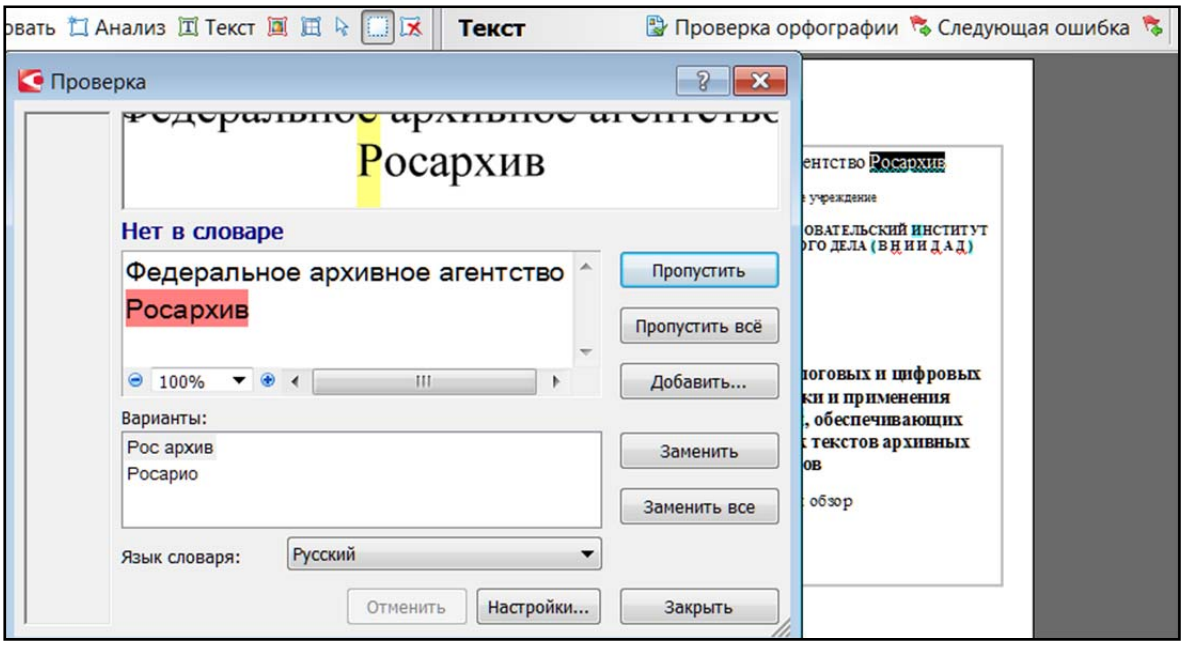


Рисунок 3.5 – Вид диалогового окна «Проверка»

Выделение текста в редакторе системы ***Fine Reader*** можно выполнить несколькими способами:

1. Чтобы выделить слово, нужно два раза нажать на нем мышью.

2. Чтобы выделить фрагмент текста, нужно нажать мышью в том месте, с которого начинается выделение, и потянуть ее в конец выбранного фрагмента, а затем отпустить кнопку мыши.

3. Чтобы выделить все содержимое окна **Текст**, в меню **Правка** выбрать пункт **Выделить все** (***Ctrl*** + ***A***).

Для того, чтобы скопировать и переместить выделенный текст, нужно нажать кнопку . Выделенный текст копируется в **Буфер обмена**. Установить курсор на то место в тексте, куда нужно перенести содержимое **Буфера обмена**. Нажать кнопку . Содержимое **Буфера обмена** копируется в текст.

Если не копировать, а переместить выделенный фрагмент, то нужно нажать кнопку .

В **Буфере обмена** содержится только та информация, которая была помещена туда в результате последней операции копирования или перемещения. Когда помещается в **Буфер обмена** новая информация, предыдущая исчезает.

Для копирования или перемещения выделенного фрагмента можно воспользоваться мышью.

Чтобы удалить текст необходимо выделить текст, затем нажать клавишу ***Del***. Выделенный текст исчезает.

Для отмены произведенного действия нужно нажать кнопку **Восстановить отмененное действие**.

Для установки языка для выделенного фрагмента нужно из списка **Язык документа** выбрать нужный язык.

Для того, чтобы установить шрифт для выделенного текста, из списка **Шрифт** на панели **Форматирование** выбрать нужный шрифт (рисунок 3.6).

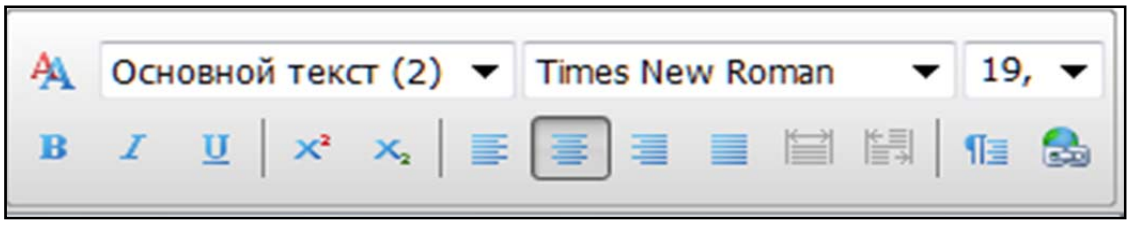


Рисунок 3.6 – Вид панели «Форматирование»

Можно изменить начертание шрифта и шрифтовые эффекты (Полужирный, *Наклонный*, Подчеркивание, Перечеркивание, Верхний индекс, Нижний индекс).

Для текущего абзаца можно установить следующие опции:

– выравнивание (по правому, левому краю, по центру, по обоим краям);

– отступы (справа, слева, сверху, снизу);

– красная строка;

– межстрочное расстояние.

Задания и контрольные вопросы

Задания

1. Выполнить настройку системы ***Fine Reader***. Для этого необходимо установить язык распознавания и тип текста. При распознавании текста на том или ином языке выбрать нужный язык из списка на панели **Страницы** из списка **Язык документа.**

2. Указать сканер, с которым будете работать:

– из меню **Сервис** выбрать пункт **Опции...**;

– в диалоге **Опции** выбрать закладку **Сканирование/Открыть**.

– в группе **Сканер** в строке **Драйвер** должен быть указан сканер.

3. Установить характер расположения текста на странице.

Программа ***Fine Reader*** автоматически определяет раскладку текста на странице. Для книг, газет, факсов, отчетов и т.п. подходит положение **Автоматическое определение**. В редких случаях, например, при распознавании оглавлений и листингов программ, нужно специально указывать программе, что текст напечатан в одну колонку. Для установки характера расположения текста на странице выполнить следующее:

4. Установить автоматический подбор ориентации страницы при сегментации и распознавании. Для этого необходимо:

– в меню **Сервис** выбрать пункт **Опции...**;

– в диалоге **Опции** выбрать закладку **Сканировать/Открыть**;

– отметить пункт **Автоматически распознавать полученные изображения**. Этот пункт следует отмечать, если в одном пакете распознаются страницы разной ориентации.

5. Установить уровень выделения ошибок.

В программе ***Fine Reader*** есть несколько уровней выделения ошибок:

Минимальный – когда цветом отмечаются нераспознанные символы и символы, распознанные с очень большой степенью неуверенности.

Стандартный – когда цветом отмечаются нераспознанные символы, символы, распознанные с очень большой степенью неуверенности и слова, которых нет в словаре языка распознавания.

Максимальный – когда помимо нераспознанных символов и не словарных слов цветом отмечаются символы, распознанные с любой (даже небольшой) степенью неуверенности.

Процедура установления уровня выделения ошибок следующая:

– из меню **Сервис** выбрать пункт **Опции**...;

– в диалоге **Опции** выбрать закладку **Дополнительно**;

– нажать кнопку **Проверка орфографии**;

– выполнить необходимые установки в окне **Настройка проверки орфографии** (рисунок 3.7).

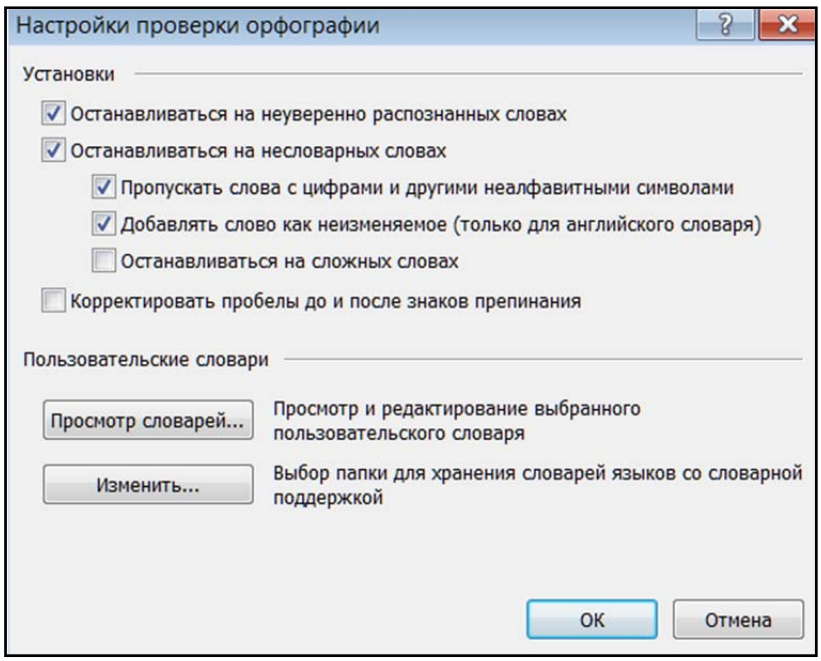


Рисунок 3.7 – Вид диалогового окна «Настройка проверки орфографии»

6. Выполнить распознавание пакета документов (приложение А).

Распознаваемое изображение должно иметь стандартную ориентацию: текст должен читаться сверху вниз и строки должны быть параллельны нижнему краю экрана. Поэтому при сканировании нужно указать программе, чтобы она автоматически подбирала ориентацию страницы.

7. Сохранить результаты распознавания.

Прежде чем сохранять результаты распознавания в том или ином формате, указать, сохранять или не сохранять форматирование, разделение на строки и т.п.

8. Передать распознанный текст в приложения ***MS Word*** и ***MS Excel***.

Для того, чтобы передать результаты распознавания в одно из следующих приложений: ***MS Word*** или ***MS Excel*** – нужно из меню **Файл** выбрать пункт **Сохранить как**, выбрать диск, папку и формат, задать имя сохраняемому файлу. Нажать кнопку **Сохранить.**

9. Выполнить проверку орфографии распознанного текста документа.

10. Изменить начертание и размер шрифта в распознанном тексте документа.

11. Для всех абзацев распознанного текста документа установить следующие опции:

– выравнивание текста в абзаце – по обоим краям;

– отступы: сверху – 3 пт; снизу – 6 пт;

– красная строка – 1,25 см;

– межстрочное расстояние – 1,5 строки;

12. Произвести вычисления для отсканированного документа, переданного в программу ***MS Excel.***

13. Показать результат работы преподавателю.

14. Оформить отчет о проделанной работе.

Контрольные вопросы

1. Рассказать о принципах работы программы ***Fine Reader***.

2. Что включает в себя интерфейс программы ***Fine Reader***?

3. Что включает в себя понятие «сегментация отсканированного документа»?

4. Как можно установить автоматический подбор ориентации страницы при сегментации и распознавании?

5. Какие типы блоков поддерживает программа ***Fine Reader***?

6. В каких форматах можно сохранять результаты распознавания в программе ***Fine Reader***?

7. Какие уровни выделения ошибок можно установить в программе ***Fine Reader***?

8. Описать процедуру установления уровня выделения ошибок.

9. В какие приложения можно передать результаты распознавания?

10. Что включает в себя корректировка текста документа в программе ***Fine Reader***?

11. Какие действия необходимо выполнить для того, чтобы проверить орфографию в программе ***Fine Reader***?

12. Какими способами можно выполнить выделение текста в редакторе системы ***Fine Reader***?

13. Каким образом можно установить язык для выделенного фрагмента отсканированного документа?

14. Каким образом можно установить шрифт для выделенного текста в программе ***Fine Reader***?

15. Как можно изменить начертание шрифта и шрифтовые эффекты в программе ***Fine Reader***?

16. Какие опции можно установить для текущего абзаца в программе ***Fine Reader***?

Библиографический список

1. Об информации, информационных технологиях и о защите информации : федер. закон № 149-ФЗ от 27.07.2006 // Российская газета. – 2006. – 29 июля. – № 165.

2. Об архивном деле в Российской Федерации : федер. закон № 125-ФЗ от 22.10.2004 // Российская газета. – 2006. – 8 декабря. – № 277.

3. Правила организации хранения, комплектования, учета и использования документов Архивного фонда Российской Федерации и других архивных документов в органах государственной власти, органах местного самоуправления и организациях : утв. приказом Министерства культуры РФ от № 526 31 марта 2015 г. (зарег. в Минюсте РФ 07.09.2015).

4. ГОСТ Р 7.0.97–2016. Национальный стандарт Российской Федерации. Система стандартов по информации, библиотечному и издательскому делу. Организационно-распорядительная документация. Требования к оформлению документов : утв. приказом Росстандарта № 2004-ст от 08.12.2016.

5. ГОСТ Р 7.0.8–2013. Национальный стандарт Российской Федерации. Система стандартов по информации, библиотечному и издательскому делу. Делопроизводство и архивное дело. Термины и определения : утв. приказом Росстандарта № 1185-ст от 17.10.2013.

6. ГОСТ 7.50–2002. Межгосударственный стандарт. Система стандартов по информации, библиотечному и издательскому делу. Консервация документов. Общие требования.

7. Информатика. Базовый курс для бакалавров / под ред. С. В. Симоновича. – СПб. : Питер, 2012. – 640 с.

8. Мишенев, А. И. Adobe Photoshop CS 4. Первые шаги в Creative Suite 4 / А. И. Мишенев. – М. : ДМК Пресс, 2009. – 144 с.

9. Разработка методов восстановления угасающих хроматических (цветных) текстов архивных документов с использованием цифровых компьютерных технологий : метод. рекомендации. – М. : ВНИИДАД, 2014. – URL: .http://archives.ru/sites/default/files/2014-metod-rekomendacii-vosstanov

10. Сравнительный анализ аналоговых и цифровых технологий для выработки и применения технологических решений, обеспечивающих восстановление угасающих текстов архивных документов. Аналитический обзор. – М. : ВНИИДАД, 2012. – URL: http://archives.ru/documents/methodics/obzor_restore-text-archival-document.shtml

Приложение А

(*обязательное*)

Задания к лабораторной работе № 2

1. Текст с рисунком

УПРАВЛЕНИЕ ПО ДЕЛАМ АРХИВОВ
КИРОВСКОЙ ОБЛАСТИ

**ГОСУДАРСТВЕННОЕ
ОБЛАСТНОЕ УЧРЕЖДЕНИЕ
«ГОСУДАРСТВЕННЫЙ АРХИВ
КИРОВСКОЙ ОБЛАСТИ»**
(ГОУ «ГАКО»)

Карла Маркса ул., д. 142, г. Киров, 610027
Телефон (факс) 67-51-05
ОГРН 1034316518928
ИНН/КПП 4346018048/434501001

ДворянкинуС. А.
Аустрина ул.,
д. 143 а, кв. 31,
г. Пенза,
440015

АРХИВНАЯ СПРАВКА

31.12.2009 № 1-24/616

На № _____ от _____

В архивном фонде Вятской духовной консистории в метрической книге Преображенской церкви с. Зашижемье Орловского уезда за 1910 год имеется актовая запись под порядковым номером 70 о рождении 03 июня 1910 года у «временноотпускного солдата» починка Малого Коропова мальчика Александра, крещен 05 июня.

Родители: отец – Димитрий Спиридонов Целищев (имя и отчество – так в документе)

мать – Елизавета Иларионова Целищева (отчество – так в документе).

Основание: ф. 237, оп. 226, д. 1666, л. 510 об.

Директор архива Е. Г. Комарова

Главный специалист В. В. Косолапов

2. Текст, разбитый на несколько колонок

Этапы решения задач на электронно-вычислительных машинах (ЭВМ)

Решение задач на ЭВМ – сложный процесс, состоящий из следующих этапов:

1. Постановка задачи. Уяснение проблемы, определение целей решения и области применения результатов.
2. Математическое описание. Выбор математических методов и моделей, запись формул, обеспечивающих решение задачи. Составление плана решения.
3. Программирование задачи. Состоит из трех частей:
– графическое изображение метода решения (составление схемы алгоритма);
– написание программы на языке программирования (кодирование);
– ввод текста в ЭВМ.
4. Отладка программы. Выявление ошибок (кодирования и методов решения), проверка правильности результатов.
5. Обработка данных и получение результатов. Выполнение готовой программы на ЭВМ, выдача результатов.

Из этих этапов первые три не требуют ЭВМ, наиболее сложные первые два, а наиболее трудоемкий – четвертый. Все они сопровождаются соответствующей документацией, состав которой определяется стандартами.

3. Сведения в табличном виде

Наименование работы	Единица измерения	Годовой объем работы	Норма времени, ч	Нормативная трудоемкость, ч
1	2	3	4	5
Ведение контрольной картотеки	Одна РКК	10860	0,15	
Запись в книге учета отправляемых документов	Одна запись	6000	0,04	
Опись входящих документов	Одна опись	120	0,23	
Заполнение реестра	Один реестр	3733	0,19	
Обработка телеграмм	Одна телеграмма	650	0,11	
Рассылка документов	Один адресат	6000	0,023	
Регистрация получаемых документов коллегиального органа	2 л.	4800	0,03	
Регистрация внутренних приказов	2 л.	120	0,08	
Учет приказов коллегиального органа	Один документ	120	0,07	
Учет решений (постановлений) коллегиального органа	Один документ	120	0,1	
Учет телеграмм коллегиального органа	Одна телеграмма	150	0,05	
Учет сводок коллегиального органа	Одна сводка	60	0,48	
Учет описи документов коллегиального органа	Одна опись	420	0,12	
Учет писем-извещений коллегиального органа	Одно письмо	1260	0,05	
Учет постановлений коллегиального органа	Один документ	120	0,08	
Итого:				
Максимальное количество годового объема работ				
Минимальное количество годового объема работ				
Среднее количество годового объема работ				

Приложение Б

(*информационное*)

Требования к оформлению отчета

Отчет по лабораторной работе оформляется в соответствии с ГОСТ 2.105–95. ЕСКД. Общие требования к текстовым документам.

Отчет должен содержать:

1. Титульный лист с указанием названия номера лабораторной работы, номера группы ФИО студентов, должности и ФИО преподавателя.
2. Формулировку цели работы.
3. Порядок выполнения работы.
4. Выводы о проделанной работе.
5. Приложения, содержащие результаты выполнения работы (*при необходимости*).

Учебное издание

Обеспечение сохранности
архивных документов
с использованием
информационных технологий

С о с т а в и т е л и :
Надеева Наталья Николаевна,
Катышева Марина Александровна

Редактор *Л. Ю. Зимина*
Технический редактор *Р. Б. Бердникова*
Компьютерная верстка *Р. Б. Бердниковой*

Подписано в печать 09.10.2018. Формат 60×84$^{1}/_{16}$.
Усл. печ. л. 2,09. Тираж 100.
Заказ № 610.

Издательство ПГУ.
440026, Пенза, Красная, 40.
Тел./факс: (8412) 56-47-33; e-mail: iic@pnzgu.ru


ПензГУ

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ
Федеральное государственное бюджетное
образовательное учреждение высшего образования
«ПЕНЗЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Обеспечение сохранности архивных документов с использованием информационных технологий

Методические указания
к выполнению лабораторных работ

ПЕНЗА 2018