

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

**ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МІСЬКОГО ГОСПОДАРСТВА імені О. М. БЕКЕТОВА**

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ

до самостійної роботи із навчальної дисципліни

**«ЕЛЕКТРОННИЙ ДОКУМЕНТООБІГ У ПРОФЕСІЙНІЙ
ДІЯЛЬНОСТІ»**

*(для студентів усіх форм навчання освітнього рівня «бакалавр»
спеціальностей 126 – Інформаційні системи та технології,
151 – Автоматизація та комп'ютерно-інтегровані технології)*

**Харків
ХНУМГ ім. О. М. Бекетова
2020**

Методичні рекомендації до самостійної роботи із навчальної дисципліни «Електронний документообіг у професійній діяльності» (для студентів усіх форм навчання освітнього рівня «бакалавр» спеціальностей 126 – Інформаційні системи та технології, 151 – Автоматизація та комп'ютерно-інтегровані технології) / Харків. нац. ун-т міськ. госп-ва ім. О. М. Бекетова ; уклад. М. Ю. Карпенко. – Харків: ХНУМГ ім. О. М. Бекетова, 2020. – 27 с.

Укладач канд. техн. наук, доц. М. Ю. Карпенко

Рецензент

О. В. Грицунов, доктор фізико-математичних наук, професор кафедри мікроелектроніки, електронних приборів і пристроїв (Харківський національний університет радіоелектроніки);

Рекомендовано кафедрою прикладної математики і інформаційних технологій, протокол № 1 від 29.08.2019.

ЗМІСТ

ВСТУП.....	4
1 Що таке сканування й розпізнавання	4
2 Які зображення можна розпізнати.....	5
3 Як розпізнати документ у програмі FineReader	6
3.1. Інтерфейс програми FineReader	7
3.2 Послідовність дій щодо розпізнавання документу	8
3.3 Завантаження зображень	8
3.4 Автокоригування зображень	9
3.5 Розмітка документу.....	11
3.6 Розпізнаємо зображення.....	13
3.7 Зберігаємо результат.....	14
3.8 Експортуємо результат	15
3.9 Розпізнавання таблиці складної структури	15
4 Коригування зображень перед їх розпізнаванням	17
4.1 Як перезняти зображення	17
4.2 Штучний спалах та його використання	17
4.3 Послідовність дій з коригування зображень	18
4.3.1. Завантажуємо файл	19
4.3.2. Коригуємо розмір зображення	20
4.3.3 Прибираємо зайві деталі	21
4.3.4. Коригуємо геометрію.....	22
4.3.5. Коригуємо яскравість	22
5 Завдання до виконання роботи	25
Список рекомендованих джерел.....	26

ВСТУП

Мета роботи – опанувати роботу з програмою FineReader. Навчитися розпізнавати документ за допомогою програми FineReader. Одержати навички щодо коригування зображень засобами графічного редактора PhotoShop для їх подальшого розпізнавання.

1 ЩО ТАКЕ СКАНУВАННЯ Й РОЗПІЗНАВАННЯ

Сканування – це процес одержання графічної копії оригінального зображення. Сканувати документ можна по-різному. Наприклад, – за допомогою сканера або фотоапарата. І в тому й в іншому випадках результатом буде растровий графічний файл. Цей файл може бути кольоровим або чорно-білим, мати різну роздільність, може бути представлений у різних форматах (найчастіше «*.jpg» або «*.tif»). Важливо розуміти головне: растровий файл – це набір точок. Те, що на екрані або друкованому аркуші він виглядає як текст, ще ні про що не свідчить. Обробити цей «текст» програмою Word або Excel неможливо. Усе, що ви зможете з ним зробити – це вставити у вигляді готового об’єкта. Відкоригувати растровий файл можна лише у спеціальній програмі – графічному редакторі. При цьому працювати з ним можна як з графічним зображенням, а не як з текстовим документом. Пояснимо цю різницю між текстом й сканованим документом. Уявіть, що ви вводите документ до програми Word, набираючи його на клавіатурі. Щоразу, натискаючи клавішу, до комп’ютера потрапляє її код. Фактично цей код і зберігається у «вордовському» документі. У цьому випадку для кожної букви документа відомо її походження. Наприклад, перший символ цього абзацу має код «0194», це буква «В», другий – код «0224», це символ «а» і т. д. При друкуванні або виведенні тексту на екран виконується **візуалізація** документа. Тобто виходячи з кодів, що є в тексті Word формує їх графічне зображення, іншими словами, «промальовує» букви. Коли ви коригуєте документ у програмі

Word, ви змінюєте коди символів у документі. Саме тому одне натискання клавіші може змінити лише один символ. У графічному зображенні цього немає. Сканер нічого не друкує. Він вводить піксели (крапки) зображення. Будучі розташовані у певному порядку, ці піксели виглядають як текст. Той факт, що якась частина цього зображення – це символ «В», а якась – символ «а» комп'ютеру невідома. Для нього усі ці «символи» – просто набір пікселів. Тому й відредагувати ці піксели програмою Word не можна.

Для того, щоб перетворити графічний файл до текстового формату, його потрібно розпізнати. Цим займаються спеціальні програми оптичного розпізнавання образів (OCR). Однією з найпоширеніших програм цього класу є FineReader. Мета у цих програм одна. Вони аналізують зображення, намагаючись знайти фрагменти із якимось змістом. Це можуть бути символи тексту, таблиці, малюнки. Програма розпізнавання намагається з'ясувати структуру документа: знайти букви, слова, речення, орієнтуючись на специфіку їх зовнішнього вигляду. Потім вона формує текстовий документ, перевіряє його за системою словників, граматичних правил й пропонує користувачеві для остаточного контролю. Коли перевірку завершено, результат розпізнавання можна зберегти у форматі процесора Word або таблиці Excel. Після цього ви отримаєте «нормальний» документ, з яким можете робити все, що завгодно.

Звичайно, на практиці не все так просто. Алгоритм розпізнавання – процес складний, у випадку неякісних зображень програма розпізнавання припуститься чисельних помилок. Щоб мінімізувати їх кількість, треба чітко уявляти, які мають бути зображення для доброго розпізнавання, як їх можна виправити у графічному редакторі, яким чином треба організувати сам процес розпізнавання.

2 ЯКІ ЗОБРАЖЕННЯ МОЖНА РОЗПІЗНАТИ

Зображення для вдалого розпізнавання мають відповідати певним вимогам, а саме:

– графічний файл повинен мати роздільність приблизно 300 dpi;

– його лінійний розмір при цій роздільності має бути близький до оригіналу. Навряд чи вам вдасться розпізнати сторінку документа з роздільністю 300 dpi розміром з поштову марку;

– зображення має бути чорно-білим або у відтінках сірого.

Це, так би мовити, технічні вимоги. Зрозуміло, є інші, не менш важливі, умови для виконання цієї роботи. Наприклад, зображення має бути досить контрастним, букви – з чітко окресленими краями. Будь-яке «сміття» на сканованому зображенні у вигляді позначок олівцем, плям, дефектів паперу програма розпізнавання спробує ідентифікувати як символи. Ці обставини треба враховувати при підготовці зображень для подальшого розпізнавання.

3 ЯК РОЗПІЗНАТИ ДОКУМЕНТ У ПРОГРАМІ FINEREADER

Спочатку навчимося розпізнавати якісні зображення. Тобто такі, що не потребують обробки. Вони можуть бути отримані зі сканера, або внаслідок копіювання фрагментів екрану. Для прикладу ми візьмемо для розпізнавання документ з таблицею, як на рисунку 1. Для розпізнавання будемо використати програму «FineReader» компанії АBBYY, скачавши демо-версію з <http://www.abbyy.ru/Download/>.

Какие изображения можно распознать		
Чтобы процесс распознавания был успешным, графическое изображение должно соответствовать определенным требованиям, а именно.		
Параметр	Значение	Комментарий
Разрешение	300 dpi	
Формат	«Jpg», «Tif», «Bmp»	Практически любой растровый формат
Тип	Цветное, оттенки серого, черно-белое	

Рисунок 1 – Приклад документа для розпізнавання

3.1. Інтерфейс програми FineReader

Вікно програми зображено на рисунку 2. У ньому представлено меню, під ним – панелі інструментів.

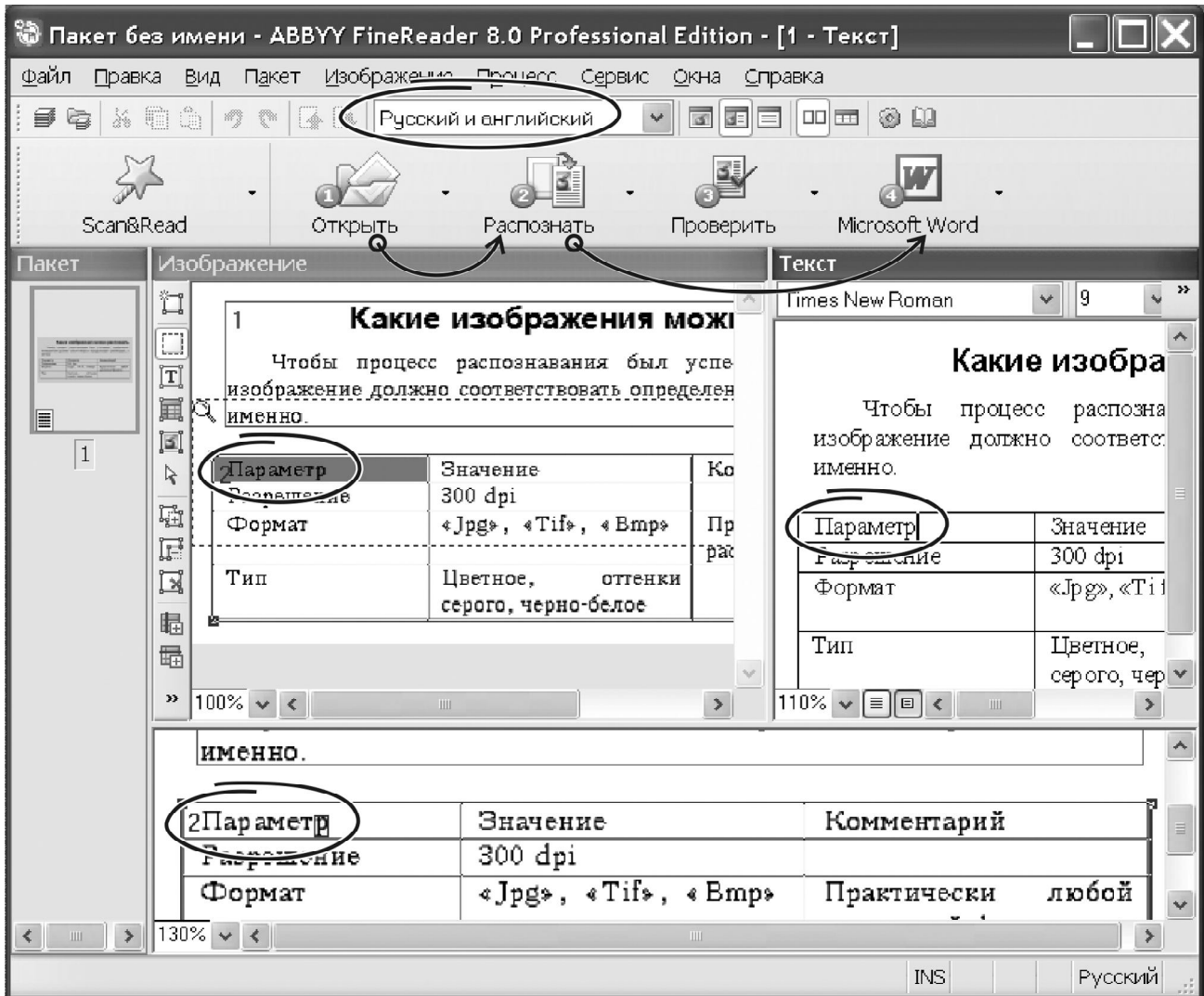


Рисунок 2 – Головне вікно програми FineReader

Нижче панелі інструментів знаходяться п'ять великих іконок для керування програмою, а саме: «Scan&Read», «Открыть», «Распознать», «Проверить», «Microsoft Word».

Основна частина вікна розбита на чотири області. Перелічимо їх у порядку зліва-направо й зверху-вниз. У лівій частині вікна міститься область сторінок документа. Кожна сторінка – це окреме зображення. Праворуч від

області сторінок розташоване віконце «Изображение», де є фрагмент відсканованого документа та його розмітка (про неї поговоримо трохи пізніше).

У правій частині вікна показано результат розпізнавання. Це – текстовий документ, його можна правити прямо у програмі FineReader. У нижній частині вікна розташована область зі збільшеною копією зображення.

Між областю сторінок та вікном «Изображение» знаходиться робоче меню для розмітки зображення.

3.2 Послідовність дій щодо розпізнавання документу

Щоб розпізнати сканований документ, потрібно зробити так:

1. Завантажити одне або декілька зображень до програми FineReader.
2. Відредагувати зображення, щоб очистити його від «сміття», усунути перекручування рядків й т. і. Цей етап потрібен не завжди.
3. Розмітити зображення, тобто вказати, Як FineReader повинен інтерпретувати окремі ділянки малюнка. Звичайно така операція потрібна при складному форматуванні документа. Для простих випадком операцію розмітки FineReader прекрасно виконує в автоматичному режимі.
4. Розпізнати документ.
5. Переглянути результат роботи. Якщо FineReader помилився, можна відредагувати документ або поправити розмітку й повторити розпізнавання.
6. Імпортувати результат розпізнавання до зовнішнього додатку, наприклад, до програми Word або Excel.

3.3 Завантаження зображень

У нашому випадку дані для розпізнавання перебувають на жорсткому диску у вигляді графічних файлів. Натискаємо кнопку «Открыть» (на рисунку 2 вона зображена з номером «1»). З'явиться вікно Провідника. У ньому знаходимо папку, виділяємо потрібні файли й клацаємо на кнопці «Открыть»

або натискаємо «Enter». Завантажені зображення з'являться в області сторінок.

Для завантаження зображень натисніть комбінацію «Ctrl+O» (англійська буква «O»). Також можна скористатись меню «Файл → Открыть».

Кнопка «Scan&Read» служить для початку сканування. Якщо ви користуєтесь сканером, включіть його, покладіть документ й натисніть «Scan&Read». Подальша робота залежить від моделі сканера. Через кнопку «Scan&Read» можна також завантажити документ. Клацніть на маленькому трикутнику в лівій частині іконки «Scan&Read». Розкриється список можливих дій. Серед них буде операція «Открыть и распознать». Розкривши список дій кнопки «Открыть», ви побачите рядок «Сканировать изображение... Ctrl+K»... Таким чином, кнопки «Scan&Read» й «Открыть» багато у чому взаємозаміняють одна одну.

3.4 Автокоригування зображень

Операції коригування зображення зосереджені в меню «Изображение». Розберемо ті можливості, які доводиться використати найчастіше.

ПОВОРОТ ЗОБРАЖЕННЯ. Текст при розпізнаванні повинен бути зорієнтований «зверху вниз». Якщо сторінка документа сканована з поворотом, її бажано переорієнтувати. Для цього увійдіть до меню «Изображение → Повернуть/Отразить изображение». З'явиться меню з п'яти пунктів: «Повернуть по часовой стрелке», «Повернуть против часовой стрелки», «Повернуть на 180°», «Зеркально отразить относительно вертикали», «Зеркально отразить относительно горизонтали». На практиці звичайно вистачає перших двох можливостей.

КОРИГУВАННЯ РОЗДІЛЬНОЇ ЗДАТНОСТІ. Якщо зображення скановано з «неправильним» розділенням, зверніться до меню «Изображение → Исправить разрешение» або «Ctrl+Shift+T». У вікні

встановіть потрібні параметри. Значення за замовчуванням становить «300dpi», можна лишити його без змін. Якщо всі сторінки від скановані з низьким розділенням, переустановіть перемикач «Исправить разрешение» у положення «Всех изображений пакета».

ОЧИЩЕННЯ ЗОБРАЖЕНЬ ВІД ДЕФЕКТІВ. Тепер спробуємо поліпшити якість картинки. Відразу зауважимо: серйозні дефекти зображення FineReader не виправить. Він може прибрати лише незначні дефекти, такі як смітинки на папері, невеличкі плями, тощо. Причому цю операцію він зробить інтелектуально: текст залишиться у первісному стані, його пошкоджено не буде. Щоб поліпшити зображення і прибрати з нього незначні дефекти, робимо так:

– виділяємо одну або декілька сторінок документа;

– викликаємо меню «Изображение → Обработать изображение → Очистить изображение от мусора».

«ВИРІВНЮВАННЯ» РЯДКІВ. Один з типових дефектів сканування – це деформація рядків тексту. Це часто трапляється, наприклад, при скануванні книг, або інших переплетених видань, коли сторінки мають вигнуту форму. У цьому випадку рядки тексту на сканованому зображенні будуть виглядати не прямими, а деформованими. Виправити такий дефект може операція усунення деформації рядків. Правда, упоратися із цим завданням вона зможе не завжди!

Щоб вирівняти рядки у відсканованому зображенні зробіть так.

– виділіть одну або кілька сторінок документа;

– зверніться до меню «Изображение → Обработать изображение → Устранить искажение строк».

Користуйтеся інструментом вирівнювання рядків обережно. У певних випадках він може відпрацювати неправильно. Тому не застосовуйте вирівнювання відразу до всіх сторінок документа. Обробляйте їх по черзі, щораз переглядаючи отриманий результат. Цими можливостями ми поки що й обмежимося.

3.5 Розмітка документу

Інформацію в електронному документі можна умовно розділити на три категорії: текст, таблиця й графічне зображення (рисунок). У процесі розпізнавання програмі FineReader бажано «знати», з чим вона має справу. У принципі, аналізуючи сторінку, з'ясувати її структуру FineReader може й самостійно. Але при складному форматуванні він може вдатися до помилок. Наприклад, якщо в документі є зображення вікна Windows (або якоїсь програми), FineReader може спробувати розпізнати текст меню та інших елементів цього вікна, хоча за змістом такого робити не потрібно. Тому у простих випадках опирайтеся на автоматичну розмітку тексту. Якщо форматування складне, має декілька колонок, знану кількість ілюстрацій, специфічних таблиць, – робіть розмітку самі. виправляти помилки FineReader-а після невдалого розпізнавання буде довше. Тому для впевненої роботи з програмою FineReader треба освоїти «ручну» розмітку документа, тим більше, що зробити це зовсім не складно.

Подивіться на панель інструментів розмітки ліворуч від області сторінок. Почнемо з іконки з буквою «Т»: це інструмент для визначення текстового блоку. Щоб виділити текстовий блок у документі, зробіть так:

- клацніть на іконці лівою кнопкою миші;
- встановіть покажчик на зображенні документа там, де має починатися фрагмент тексту;
- утримуючи ліву кнопку миші, обведіть текстовий блок;
- відпустіть ліву кнопку миші.

На документі з'явилася прямокутна область зеленого кольору. Вона повинна охоплювати фрагмент зображення, який FineReader має розглядати як текст. На рисунку 2 текстовий блок охоплює всю частину документа, за винятком таблиці. Він позначений номером 1.

Щоб видалити блок розмітки потрібно виділити його щигликом миші й нажати клавішу «Del». Щоб змінити розмір блоку, виділіть його й розтягніть за кут або межу так, як ви міняєте розмір вікна Windows.

Таким же чином ви можете виділити блок з таблицею або рисунком. Ці іконки розташовані нижче інструмента виділення текстового блоку. З областями рисунків все зрозуміло: ці ділянки зображення FineReader розпізнавати не буде. Він перенесе їх до результуючого документу «як є». А от з таблицями має сенс познайомитись докладніше. Зазвичай розпізнавання таблиці робиться так.

Вибираємо інструмент визначення блоку з таблицею.

Акуратно обводимо фрагмент зображення, де міститься таблиця. Зайвий простір при цьому краще не чіпати, але й зовнішня границя таблиці повинна повністю перебувати всередині блоку.

Клацаємо усередині створеного блоку правою кнопкою миші. З контекстного меню (рис. 3) вибираємо «Анализ структуры таблицы».

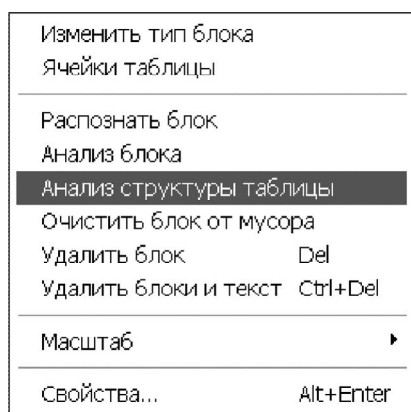


Рисунок 3 – Контекстне меню аналізу структури таблиці

Переглядаємо результат розмітки. Особливу увагу звертаємо на те, як розташувались окремі осередки таблиці. Для цього є сенс послідовно клацати лівою кнопкою миші по осередках таблиці в області розмітки й аналізувати правильність їх розташування.

Якщо FineReader усе зробив правильно, переходимо до наступного етапу роботи.

А що робити в тому випадку, якщо при розмітці таблиці є певні помилки? Наприклад, FineReader не розпізнав бліду лінію границі й об'єднав декілька осередків в один? Для цього на панелі розмітки блоків є інструменти: «Добавить вертикальную линию», «Добавить горизонтальную линию» й «Удалить линии». Користуватися ними дуже просто. Клацніть на іконці «Добавить вертикальную линию». Переведіть покажчик миші до віконця «Изображение» на таблиці. Разом з покажчиком по зображенню буде переміщатися вертикальна лінія розмітки. Поставте її на потрібне місце й клацніть лівою кнопкою миші. Таким же чином додайте всі відсутні вертикальні лінії в розмітці таблиці. Щоб закінчити роботу в режимі додавання ліній виберіть інший інструмент розмітки. Таким же чином розберіться із горизонтальними лініями. Інструмент «Удалить линии» допоможе стерти зайву розмітку таблиці.

При роботі з таблицями часто доводиться поєднувати або розділяти окремі осередки. Найзручніше зробити це через праву кнопку миші:

- встановіть покажчик на осередок таблиці;
- клацніть правою кнопкою миші, з контекстного меню (рис. 3) виберіть пункт «Ячейки таблицы», розкриються три розділи: «Разбить ячейки», «Объединить ячейки», «Объединить строки»;
- скористайтесь підходящим пунктом для вирішення проблеми; так, якщо FineReader помилково злив кілька осередків в одну, виберіть «Разбить ячейки»; якщо потрібно об'єднання осередків в одне ціле, – виділіть їх і зверніться до пункту «Объединить ячейки» контекстного меню.

3.6 Розпізнаємо зображення

Коли розмітку сторінок завершено, натисніть «Ctrl+Shift+R», – розпізнати всі сторінки. Цю дію можна виконати через меню

«Процесс → Распознать → Распознать все» або розкривши список доступних дій кнопки «Распознать».

Обов'язково перевірте, яка мова встановлена для розпізнавання. Вибрати мову можна прямо на панелі інструментів FineReader. Наприклад, на рисунку 2 встановлена пара «Русский и английский». Щигликом миші розкрийте список і виберіть варіант, що підходить для вашого випадку.

Результат розпізнавання можна подивитися у вікні «Текст» (рис. 2). Уважно прогляньте його. Спочатку переконаєтеся, що FineReader правильно обробив структуру документа. Якщо це не так, відкоригуйте розмітку й заново розпізнайте сторінку.

При повторному розпізнаванні не обов'язково заново обробляти весь документ. Ви можете обмежитися поточною або декількома виділеними сторінками. Для цього виділіть сторінку й клацніть на кнопці «Распознать» або натисніть «Ctrl+R».

3.7 Зберігаємо результат

Робота над великими документами може забрати багато часу й проходити в декілька сеансів. Щоб зберегти результат роботи в програмі FineReader використовується поняття **пакет**. Пакет поєднує вихідні зображення, їх розмітку, результати обробки в тім стані, що було на момент створення пакету. Зберегти пакет можна через меню «Файл → Сохранить пакет как...». З'явиться вікно, ідентичне вікну Провідника. Вибираємо папку, вказуємо ім'я пакета й результат буде записано на диск.

Пакет включає велику кількість файлів. Щоб не вносити сум'яття, не зберігайте пакет до папки з якимись даними. Краще зберігайте пакет в окремому каталозі.

3.8 Експортуємо результат

Експорт потрібний для збереження результатів у зовнішньому додатку або буфері обміну. Щоб зберегти поточну сторінку в Word клацніть на останній іконці з номером 4 і написом «Microsoft Word» (рис. 2). Відкриється вікно Word з розпізнаним документом. Це – не єдиний спосіб експортування даних. Клацніть на значку розкриття списку кнопки «Microsoft Word». З'являться такі варіанти: «Мастер сохранения результатов», «Сохранить страницы...», «Передать страницы в...», «Передать все страницы в...», «Опции».

Через «Сохранить страницы...» можна записати результат розпізнавання на диск. При записі є можливість вибрати всі або тільки поточну сторінку, а також визначити формат файлу для збереження.

Зазвичай на практиці користуються пунктами «Передать страницы в...» і «Передать все страницы в...». Ці дві можливості відрізняються тільки тим, що в першому випадку екпортується поточна або кілька виділених сторінок, у другому випадку FineReader передасть всі сторінки документа. Обидва меню містять кілька розділів. Їх кількість залежить від установлених програм на вашому комп'ютері. Можна обмежитися трьома: «Microsoft Word», «Microsoft Excel» й «Буфер обмена». Виберіть один із вказаних пунктів. FineReader викличе додаток і передасть до нього результат розпізнавання.

3.9 Розпізнавання таблиці складної структури

Таблиці неоднорідної структури – звичайна річ у практиці менеджера, бухгалтера, економіста. Тому, завершуючи тему розпізнавання документів, дамо пораду, як з ними працювати.

Перш ніж розпізнавати складну таблицю, визначите її «базову структуру». Тобто знайдіть той «оптимальний» набір рядків і колонок, що згодом мінімізує витрати на розподіл й об'єднання осередків. Коли структура зрозуміла, зробіть ручну розмітку таблиці.

Вибір структури таблиці є дуже важливим. Не починайте обробку документа, поки не вирішите це питання. Подивіться на бланк податкової накладної, зображений на рисунку 4.

1	2	3	4	5	6	7	8	9	10	11
Оригінал (видається покупцю)				ЗАТВЕРДЖЕНО наказом ДПА України від 30.05.97 р. № 163 (у редакції наказу ДПА України від 08.10.98 р. № 469)						
Копія (записується у продавця)										
(Непотрібне виділити поміткою "X")										
ПОДАТКОВА НАКЛАДНА										
Дата виписки податкової накладної				05.05.07		Порядковий номер			505	
Продавець						Покупець				
Розділ	Дата відправлення (виконання, постачання) товарів (послуг) продавця	Номенклатура поставки товарів (послуг) продавця	Одиниця виміру товару	Кількість (об'єм, обсяг)	Ціна поставки одиниці продукції, ція без урахування ПДВ	Обсяги поставки (база оподаткування) без урахування ПДВ, що підлягають оподаткуванню за ставками				Загальна сума коштів, що підлягає оплаті
						20 %	0 % (реалізація на митній території України)	0 % (експорт)	Звільнення від ПДВ	
1	2	3	4	5	6	7	8	9	10	11
I	05.05.07	Маркер	шт.	10	1,5	15,00				
		Ручка	шт.	9	1,8	16,20				
		Ластик	шт.	4	0,65	2,60				
	Усього по розділу I						33,80			33,80
II	Товаротransпортні									1,80

Рисунок 4 – Приклад таблиці зі складною структурою

Спроба розпізнати такий документ в автоматичному режимі, скоріш за все, буде невдалою. В результаті автоматичного розпізнавання FineReader зробить набір неупорядкованих в текстових полів, фрагментів таблиць тощо. Навести лад у такій абракадабрі буде складніше, ніж надрукувати бланк «з нуля». Проблема вирішується просто. Більшість частин документа складається з 11 колонок. Вони пронумеровані й показані у верхній частині рисунку.

При розмітці виділили блок типу «Таблиця». Інструментами «Добавить вертикальную линию», «Добавить горизонтальную линию» зробить потрібну розмітку таблицю. Після цього в ручному режимі об'єднайте окремі осередки в групи, де це потрібно. Результат розпізнавання такого складного документа буде чудовий. Нагадаємо, що поєднувати й розділяти осередки зручно через праву кнопку миші (пункти «Разбить ячейки», «Объединить ячейки»). Якщо

ви вдало вибрали структуру для ручної розмітки, найбільш складні таблиці вам не страшні.

4 КОРИГУВАННЯ ЗОБРАЖЕНЬ ПЕРЕД ЇХ РОЗПІЗНАВАННЯМ

4.1 Як перезняти зображення

Якщо ви користуєтеся сканером, пропустіть цей розділ або прочитайте «по діагоналі». Цей матеріал потрібен тим, хто буде «сканувати» документи цифровим фотоапаратом або камерою мобільного телефону. Пропонуємо кілька порад щодо організації процесу.

1. Ахіллесова п'ята будь-якого фотоапарата – чутливість до освітлення. Чим більше світла, тим кращим буде результат. Знайдіть можливість добре освітити документ, перш ніж сфотографувати його.

2. Забезпечте рівномірне висвітлення сторінки документа з усіх боків. Іноді має сенс поставити лист білого паперу з протилежної сторони від джерела світла.

3. Стежте, щоб тінь від вашого тіла не падала на документ. Прибрати графічним редактором темну пляму на аркуші буде непросто і займе певний час. Навчіться критично оцінювати предмет, що фотографуєте.

4. Стежте за нахилом камери відносно сторінки. Типова помилка при перефотографуванні документів – зйомка під неправильним кутом. У цьому випадку зображення сторінки буде не прямокутним, а схожим на трапецію, виправлення геометрії забере певного часу. Тому краще тримати камеру перпендикулярно до площини документу.

4.2 Штучний спалах та його використання

Ось декілька простих рекомендацій щодо роботи зі спалахом.

1. Включіть спалах на камері, якщо він є. Це додасть світла, зображення буде чітким й контрастним. Але ж головне – всі зображення будуть приблизно однаково освітлені. Коригувати їх можна буде за однією схемою, а це спрощує завдання і скоротить час на обробку.

2. При роботі зі спалахом не знімайте з близької відстані. Швидше за все, спалах «пересвітить» кадр, документ вийде занадто світлим. Можна перезнімати з відстані 40-90 см. Точне значення вам підкаже досвід спілкування з вашою моделлю фотокамери. Щоб оригінал не вийшов дуже дрібним, виберіть ступінь збільшення (зуммування) у 1,5, максимум в 2 рази. Тоді і спалах кадр не «пересвітить», й погіршення геометрії будуть мінімальні.

3. Максимально використовуйте площину кадру. Стежте, щоб не було великих полів між краями аркуша й межами зображення. До речі, за цим критерієм можна оцінити ступінь зуммування.

4. Не встановлюйте граничне збільшення: швидше за все, виникнуть проблеми з різкістю зображення.

Запам'ятаєте, що від якості зйомки залежить обсяг роботи з подальшого коригування зображення. Чим краще ви перезняли оригінал, тим прийдеться його коригувати. Витративши пару хвилин на поліпшення умов освітлення, ви можете заощадити години роботи над коригуванням фотокопій.

І наостаннє, зупинимось на тих проблемах, що виникають незалежно від технічних питань. Це, в першу чергу, порушення геометрії, світла й тону.

4.3 Послідовність дій з коригування зображень

Отже документи зняті, файли переписані до жорсткого диску. Перш ніж приступити до розпізнавання, потрібно подивитися, що ми маємо у своєму розпорядженні. У нашому прикладі ми одержали зображення, як на рисунку 5.

Це зображення одержано за допомогою камери мобільного телефону з розділенням 2 Мп в приміщенні з поганим освітленням. Звичайно, високої якості очікувати тут не можна. На зображенні є всі типові вади для такого випадку. Центр і межі зображення засвічені нерівномірно. Папір виглядає сірим, майже чорним, у той час як реально він білого кольору. Є дефекти геометричні, текст на сторінці має трапецієподібну форму. Подивимось, що можна зробити. Для роботи нам знадобиться редактор растрової графіки. Зупинимо свій вибір на PhotoShop.

Мабуть, це кращий професійний редактор. PhotoShop – різноманітна й складна програма. Про її можливості ходять легенди. Ми не збираємося докладно їх вивчати. Нам потрібен тільки результат. Тому основні прийоми коригування ми будемо розглядати без докладних коментарів. Приймайте їх, як є, і дійте за принципом «роби як я».

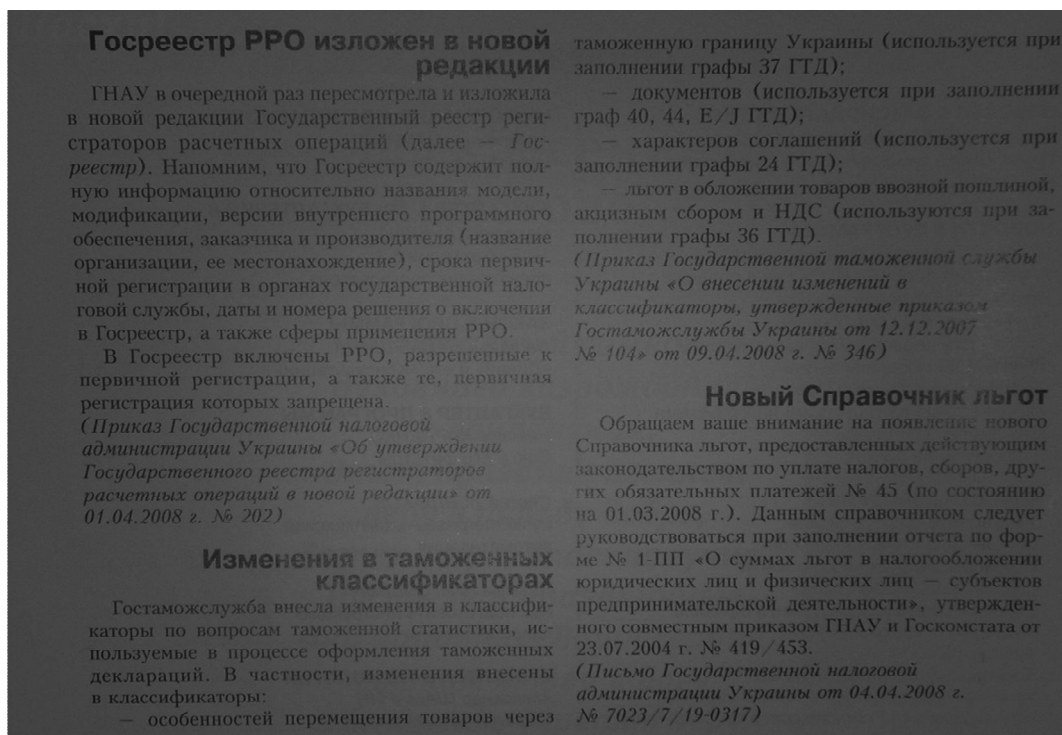


Рисунок 5 – Зображення для коригування

4.3.1. Завантажуємо файл

Викликаємо PhotoShop через «Пуск → Програми». Відкриваємо в ньому зображення. Для цього тиснемо «Ctrl+O», з'явиться вікно Провідника. У ньому знаходимо файл, подвійним щигликом відкриваємо його в програмі. Вікно PhotoShop із завантаженим зображенням показане на рисунку 6.

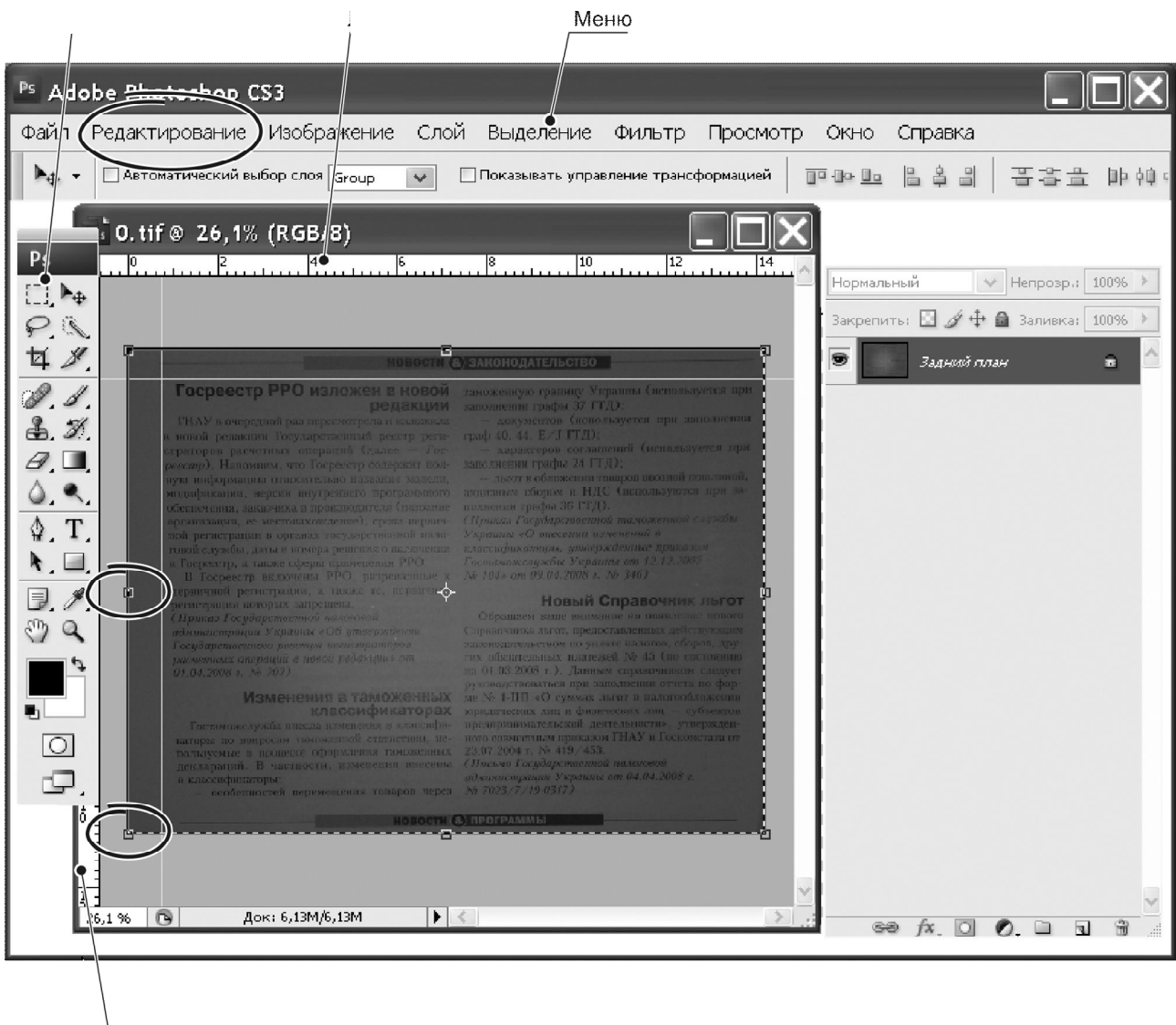


Рисунок 6 – Головне вікно програми PhotoShop

4.3.2. Коригуємо розмір зображення

Це перше, з чого ми маємо почати роботу. І ось чому. Матриця камери була 2 мегапікселі. Це означає, що оригінал, займаючи весь кадр, буде розміром приблизно 9×12 см у розділенні 300 dpi. Тепер має сенс повернути розміри до первісних: великі букви FineReader розпізнаватиме краще.

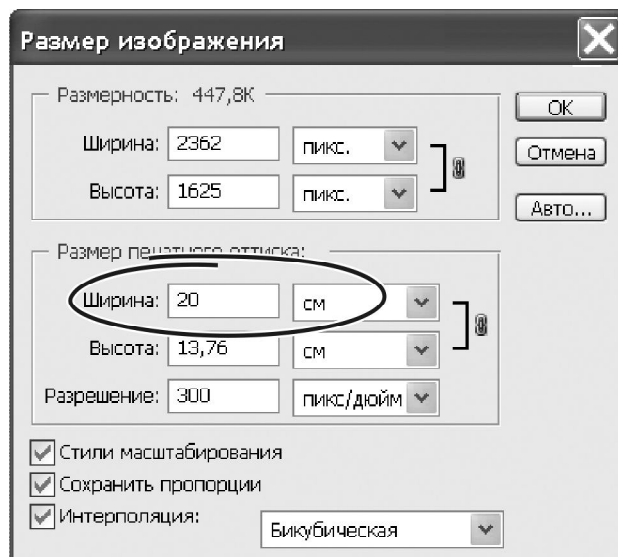


Рисунок 7 – Вікно налаштувань розміру зображень

Натискаємо «Alt+Ctrl+I» (або «Изображение → Размер изображения»). З'явиться вікно, як на рисунку 7. Вимірюємо (приблизно) розмір оригіналу. Досить визначити один вимір, наприклад – ширину. Вводимо отримане значення в поле «Ширина».

Прапорці «Інтерполяція» й «Зберегти пропорції» повинні бути включені. Значення параметра «Разрешение» має бути не менший за «300 dpi». У нашому випадку ширина зображення була близько 21 см. Вказуємо у полі «Ширина» значення «21» і натискаємо «ОК». Якщо все зроблено правильно, розмір зображення (у мегабайтах) має змінитися.

4.3.3 Прибираємо зайві деталі

Все зайве, що буде на зображенні, програма FineReader намагатиметься розпізнати та включити до готового документу. Тому перш ніж робити розпізнавання, із зображення потрібно прибрати усі зайві деталі. Робимо так.

1. Натискаємо «M», букву «M» натискайте на англійській клавіатурі. Ми вибрали інструмент «Прямоугольное выделение».

2. Утримуючи ліву кнопку миші, обводимо площу на зображенні.
3. Відпускаємо кнопку миші, навколо області з'явиться пунктирна рамка.
4. Викликаємо меню «Изображение → Кадрировать». Photoshop прибере фрагменти документу, що розташовані поза виділеною областю.

4.3.4 Коригуємо геометрію

Для впевненої роботи програми FineReader бажано зробити так щоб фото документа мало прямокутну форму, щоб воно не мало «діжкопдібного» дефекту. Для коригування геометрії робимо так.

1. Ставимо покажчик миші на ліву лінійку вікна із зображенням (рис. 6).
2. Утримуючи ліву кнопку миші, переміщаємо її на область документа. З'явиться вертикальна напрямна лінія.
3. Повторюємо дії, перетягнувши напрямну від верхньої лінійки на документ; у такий спосіб ми одержали орієнтири для контролю перпендикулярності.
4. Натискаємо «Ctrl+A» (символ «A» тиснемо на англійській клавіатурі); ми виділили весь документ, довкола нього з'явиться пунктирна рамка.
5. Натискаємо «Ctrl+T» («Редактирование → Свободная трансформация. Навколо зображення з'явиться область із прямокутними мітками на кожній стороні. Розтягуючи область, ми збільшуємо розмір зображення в тому або іншому напрямку. Але нам потрібно інше: трансформувати його геометрію, тобто зробити ширшою нижню частину зображення. Для цього ставимо покажчик миші на прямокутник у правому нижньому куті.
6. Утримуємо «Ctrl» і тягнемо його вправоруч, поки текст не вирівняється відносно сторінки або напрямних ліній.

4.3.5 Коригуємо яскравість

Щоб відкоригувати яскравість зображення (іноді цей процес називають коригуванням експозиції), робимо так.

1. Натискаємо «Shift+D», щоб зняти виділення (про всякий випадок).

2. Дивимось на значок виділення областей (рис. 6). Якщо він прямокутний, то натискаємо «M», потім «Shift+M». Значок виділення стане овальним. Якщо значок був овальний, – просто натискаємо «M».

3. Утримуючи ліву кнопку миші, обводимо найбільш яскравий фрагмент зображення (у центрі). Відпускаємо ліву кнопку миші.

4. Натискаємо «Q». Зображення покриється червоною «маскою». Залишиться тільки «отвір» у центрі із чіткими межами.

5. Натискаємо «Фільтр → Размытие → Размытие по Гауссу». З'явиться вікно з єдиним параметром «Радіус». Збільшуйте його, щоб зробити плавне розмиття меж «отвору». Зазвичай радіус становить 150-200 пт.ю

6. Натискаємо «OK», щоб закрити вікно.

7. Натискаємо «Q». Червона «маска» пропаде (ми перейшли до режиму роботи із зображенням).

8. Натискаємо «Ctrl+I» (інвертувати виділену область).

9. Натискаємо «Ctrl+L». З'явиться вікно, як на рисунку 8. Ставимо покажчик миші на центральний повзунок, утримуючи ліву кнопку, пересуваємо його вправо. Краї зображення мають стати світлішими.

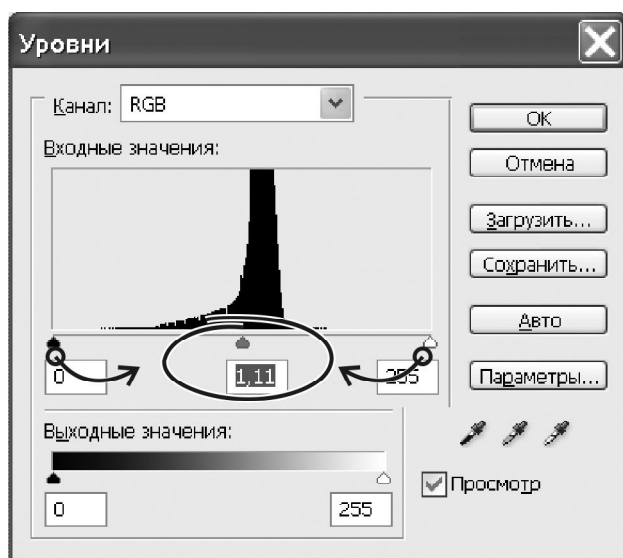


Рисунок 8 – Вікно роботи з рівнями для коригування тону

10. Натискаємо «ОК», – підтверджуємо зроблене коригування та закриваємо вікно «Уровни».

11. Скасовуємо виділену область , – комбінація «Ctrl+D».

12. Знову натискаємо комбінацію «Ctrl+L». Щоб висвітлити все зображення, переміщаємо правого повзунка ліворуч. Щоб підвищити щільність тексту, переміщаємо лівого повзунка праворуч. Не прагніть одержати текст на ідеально білому фоні. Текст на сіруватому фоні FineReader теж розпізнає.

13. Затискаємо «ОК», – закриваємо вікно «Уровни».

14. Зберігаємо документ комбінацією «Shift+Ctrl+S» (меню «Сохранить как...») або «Ctrl+S» (просто «Сохранить»). На рисунку 9 показано результат нашої роботи.

Зверніть увагу, – цей результат було отримано за допомогою мобільного телефону з пересічною камерою. Залишається тільки розпізнати файл.

Госреестр РРО изложен в новой редакции

ГНАУ в очередной раз пересмотрела и изложила в новой редакции Государственный реестр регистраторов расчетных операций (далее – *Госреестр*). Напомним, что Госреестр содержит полную информацию относительно названия модели, модификации, версии внутреннего программного обеспечения, заказчика и производителя (название организации, ее местонахождение), срока первичной регистрации в органах государственной налоговой службы, даты и номера решения о включении в Госреестр, а также сферы применения РРО.

В Госреестр включены РРО, разрешенные к первичной регистрации, а также те, первичная регистрация которых запрещена.

(Приказ Государственной налоговой администрации Украины «Об утверждении Государственного реестра регистраторов расчетных операций в новой редакции» от 01.04.2008 г. № 202)

Рисунок 9 – Відкоригований документ для розпізнавання

Описаний прийом вимагає певного досвіду. Від того, як вдало побудована виділена область, як плавно вона розмита залежить результат коригування рівнів. Ви можете повторити процес з іншим розміром виділеної області. Але в будь-якому разі описаних способів дозволяє отримати прийнятний результат.

5 ЗАВДАННЯ ДО ВИКОНАННЯ РОБОТИ

Вхідну інформацію для роботи студент обирає сам та погоджує з викладачем. Це має бути текстовий документ (або декілька різних документів) на паперовому носії. Метою роботи буде розпізнавання такого документу. У документі для розпізнавання має бути не менше 5 таблиць та 5 рисунків. Причому рисунки мають бути розташовані за текстом. Розпізнавання сторінок, де розташовано один рисунок, не розглядається. У звіті щодо виконання роботи студент має надати детальне описання кожного кроку розпізнавання документа. Кожний етап роботи треба проілюструвати відповідним рисунком (скріншотом).

Звіт опублікуйте на дистанційному курсі № 285.

СПИСОК РЕКОМЕНДОВАНИХ ДЖЕРЕЛ

1. Баронов, В. В. Информационные технологии и управление предприятием / В. В. Баронов. - М.: ДМК пресс 2005. - 328 с.
2. Ларьков, Н.С. Документоведение: учебное пособие / Н. С. Ларьков. - М.: Восток-Запад, 2006. - 427 с.
3. Кулешов, С. Г. Управлінське документознавство: навч. посібник / С. Г. Кулешов. - К.: ДАКККиМ, 2003. - 57с.
4. Долженков В.А. Самоучитель Microsoft Office 2010 / В. А. Долженков, Ю. В. Колеников. – СПб.: БХВ – Петербург, 2011. – 368 с.
5. Гандапас Р. Презентационный конструктор / Р. Гандапас. – М.: Вершина, 2006. – 192 с.
6. Карпенко М. Ю. Конспект лекцій з дисципліни «Електронний документообіг в проектувальній діяльності» (для студентів 1 курсу денної форми навчання за напрямами: 6.030504 – «Економіка і підприємництво», 6.030509 – «Облік і аудит») / М. Ю. Карпенко, В. Б. Уфимцева. – Харків : ХНУМГ, 2012. – 90 с.
7. Методичні вказівки «Оформлення презентацій у програмі Power Point» та дипломного проектування (для студентів 3 курсу денної форми навчання освітньо-кваліфікаційного рівня бакалавр, у галузі знань 0305 «Економіка та підприємництво», за напрямом підготовки 6.030504 «Економіка підприємства»)/ М. Ю. Карпенко, О. В. Димченко, В. М. Тюріна, В.Б. Уфимцева. – Х. : ХНУМГ, 2010. – 30 с.
8. Методичні вказівки до виконання лабораторної роботи «Оформлення презентацій у програмі Impress» / М. Ю. Карпенко, Т. В. Момот, В. Б. Уфимцева, Н. В. Макогон. – Харків : ХНУМГ, 2011. – 26 с.
9. Дистацнійний курс «Електронний документообіг в проектувальній діяльності».

Виробничо-практичне видання

Методичні рекомендації
до самостійної роботи із навчальної дисципліни

«ЕЛЕКТРОННИЙ ДОКУМЕНТООБІГ У ПРОФЕСІЙНІЙ ДІЯЛЬНОСТІ»

*(для студентів усіх форм навчання освітнього рівня «бакалавр»
спеціальностей 126 – Інформаційні системи та технології,
151 – Автоматизація та комп'ютерно-інтегровані технології)*

Укладач **КАРПЕНКО** Микола Юрійович

Відповідальний за випуск *М. Ю. Карпенко*
За авторською редакцією
Комп'ютерне верстання *М. Ю. Карпенко*

План 2020, поз. 288 М

Підп. до друку 24.06.18. Формат 60 × 84/16.

Друк на ризографі. Ум. друк. арк. 1,6.

Тираж 30 пр. Зам. № __.

Видавець і виготовлювач:

Харківський національний університет
міського господарства імені О. М. Бекетова,
вул. Маршала Бажанова, 17, Харків, 61002
Електронна адреса: rectorat@kname.edu.ua
Свідоцтво суб'єкта видавничої справи:
ДК № 5328 від 11.04.2017.