**ЗАТВЕРДЖУЮ:**

Старший майстер КЦПТО

\_\_\_\_\_\_\_\_\_\_\_\_\_ А.В.Сорочинський

“\_\_\_\_” грудня 2016 р.

**ПЛАН УРОКУ**

**Тема програми:** Автоматична обробка документів

**Тема уроку:** Встановлення програми сканування та розпізнавання.

**МЕТА УРОКУ**

**Навчальна:** Навчити учнів працювати програмами сканування та розпізнавання.

**Розвиваюча:** Розвивати зацікавленість учнів до обраної професії.

**Виховна:** Формувати самостійність у виконанні окремих операцій, виховувати бережливе відношення до обладнання та майна лабораторії.

**Вид уроку:** Урок одержання нових знань, вивчення та закріплення прийомів роботи.

**Метод проведення:** Комбінований

**Матеріально-технічне забезпечення:** комп’ютери, конспекти.

**ХІД УРОКУ**

**Організаційна частина – 5 хв.**

Привітання.

Перевірка наявності учнів.

**Вступний Інструктаж – 50 хв**.

Повідомлення учням теми та мети уроку.

Опитування учнів попереднього матеріалу раніше вивченої теми.

Вивчення нової теми за планом уроку.

Закріплення нової теми.

Інструктаж з охорони праці перед початком роботи.

Доведення денного завдання та критеріїв оцінювання.

**Поточний інструктаж – 205 хв.**

Перевірка організації робочих місць.

Робота учнів з програмами сканування та розпізнавання.

Надання допомоги невстигаючим учням.

Перевірка вірності виконаних вправ.

Надання консультацій учням, у яких виявлені помилки.

**Заключний інструктаж – 10 хв.**

підведення підсумків роботи за урок:

аналіз допущених помилок та недопущення їх в подальшій роботі;

відзначення успішних робіт учнів;

оголошення оцінок за виконане завдання.

домашнє завдання: вивчити з конспекту нові записи і терміни даної теми.

**Актуалізація опорних знань.**

1. Які програми для сканування тексту ви знаєте?
2. Які програми для розпізнання тексту ви знаєте?
3. Де можна взяти ПО для сканування та розпізнання тексту?
4. Алгоритм встановлення програм для сканування тексту?
5. Алгоритм встановлення програм для розпізнання тексту?
6. Настройка ПО для розпізнання і сканування тексту?

**Повідомлення нового матеріалу**

**Сканування і розпізнавання тексту**

Сканування і розпізнавання тексту є необхідними стадіями в процесі перетворення документа на твердому носії в цифровий формат. Це займає набагато менше часу, ніж ручний спосіб простого перенабор тексту. В результаті ви отримуєте готовий документ, який можна редагувати в будь-якому текстовому редакторі.

На цьому етапі задіюється спеціальне обладнання сканер, що дозволяє отримати оптичним способом електронну копію текстового документа. Однак отримані файл схоже цифрової фотографії, не дивлячись на те, що начебто на ній чітко зображений текст. Тобто, редагувати такий файл доведеться засобами графічних програм, що представляється абсолютно не раціональним.

Для того щоб працювати з текстом, звичним чином в текстовому редакторі, текст доведеться розпізнати. **Розпізнавання тексту** - це процес перекладу графічного зображення символів (букв) в комп'ютерні текстові символи. Зробити це можна маючи якісну цифрову копію оригінального тексту і набір сучасних комп'ютерних програм для розпізнавання тексту.

**Етапи сканування і розпізнавання тексту**

Опускаючи процес підготовки оригіналів, сам процес сканування і розпізнавання можна розділити на наступні етапи.

**1. Вибір режиму сканування**. Текстовий чорно-білий документ, який не містить картинок і ілюстрацій, або ж ці елементи так само чорно-білі, досить вибрати режим чорно-білого сканування або в градаціях сірого. Однак якщо текст не чорно-білий, присутні кольорові елементи, картинки, графіки, схеми, і нам потрібно отримати точну копію, режим сканування повинен бути відповідним, тобто кольоровим. Якщо кольоровість копії не принципова, то можна обмежитися режимом сканування у відтінках сірого.

**2. Далі приступаємо до розпізнавання тексту**. Для початку за допомогою спеціалізованих програм аналізується структура документів. На цьому етапі важливим є контроль фахівця, так як можливі помилки на цьому етапі зажадають серйозних зусиль по їх усуненню надалі.

Наступний етап - безпосереднє розпізнавання тексту. Цей процес так само довіряємо комп'ютерним програмам. Однак для успішного розпізнавання і мінімізації помилок, фахівець повинен, перш за все, правильно підготувати документ для розпізнавання, налаштувати обладнання для отримання максимально якісної цифрової копії оригіналу.

Свіжовіддруковані на принтері текст і стара газета з вицвілим шрифтом і пожовклим папері вимагають до себе різного підходу на всіх етапах сканування і розпізнавання. Однак професійні досвідчені фахівці за допомогою новітніх програмно-апаратних засобів відмінно справляються з будь-якими завданнями такого роду.

**3. Перевірка результату і усунення помилок**. Після закінчення процесу розпізнавання, не дивлячись на постійно вдосконалюється програмне забезпечення, що вийшов результат необхідно перевірити і відкоригувати вручну. Чим гірше стан оригіналу, тим більше виникає помилок і тільки людина може провести корекцію і виправити всі недоліки.

Далі відбувається перевірка на предмет синтаксичних і орфографічних помилок, розставляються, в разі потреби, розділові знаки і спеціальні символи.

**4. І останній етап це форматування документа**. Перевіряється і якщо потрібно коригується розмір шрифту, стилі заголовків і тексту, розбивається на абзаци, глави, перевіряється нумерація сторінок і зміст документа. Також перевіряється загальна структура документа та верстка. Графічні елементи, картинки, графіки, схеми та інші ілюстрації так само проходять перевірку на предмет відповідності.

Після проведення всіх вищеописаних процедур, ми отримуємо точну і найголовніше, що редагується копію оригіналу. Тепер ми з легкістю можемо вносити зміни, копіювати і ділиться цим документом, отримати необхідну кількість твердих копій. На підставі цього документа можна створювати свої власні документи. При цьому документ буде зберігатися, без ризику бути зіпсованим, втраченим або вкраденим.

Відсканувавши і розпізнавши всі свої документи, перевівши в цифровий вигляд свою бібліотеку, можна створити електронний архів, зручність використання якого важко переоцінити.

**Що потрібно для сканування і розпізнавання?**

**1) Сканер**

Для перекладу друкованих документів в текстовий вигляд, вам для початку потрібен сканер і відповідно, "рідні" програми і драйвери, які з ним йшли. За допомогою них можна буде сканувати документ і зберегти його для подальшої обробки.

Можна скористатися і іншими аналогами, але софт, який йшов зі сканером в комплекті, звичайно працює швидше і має більше опцій.

Залежно від того, який у вас сканер - швидкість роботи може істотно відрізнятися. Є сканери, які можуть отримати картинку з листа за 10 сек., Тобто які будуть отримувати за 30 сек. Якщо бажаєте сканувати книгу на 200-300 аркушів - думаю, не важко підрахувати у скільки разів буде різниця в часі?

**2) Програма для розпізнавання**

ABBYY FineReader 11

Офіційний сайт: http://www.abbyy.ru/

Одна з кращих програм в своєму роді. Вона призначена для того, щоб розпізнати текст на зображенні. Вбудовано безліч опцій і функцій. Може розібрати купу шрифтів, підтримує навіть рукописні варіанти.

Як правило, різні версії ABBYY FineReader не сильно відрізняються один від одного. Ви без праці зробите те ж саме і в інший. Головні відмінності можуть бути в зручності, швидкості роботи програми і її можливості. Наприклад, більш ранні версії відмовляються відкривати документ PDF і DJVU.

**Параметри сканування тексту**

У всіх сканерах є одні й ті ж настройки, які сильно можуть вплинути на швидкість і якість вашої роботи.

**1) Якість сканування - DPI**

По-перше, якість сканування поставте в опціях не нижче 300 DPI. Бажано навіть виставити побільше, якщо це можливо. Чим вище показник DPI - тим чіткіше вийти ваша картинка, ну і тим самим, швидше пройде подальша обробка. До того ж чим вище якість сканування - тим менше помилок вам надалі доведеться виправляти.

Оптимальний варіант забезпечує, зазвичай, 300-400 DPI.

**2) Кольори**

Цей параметр дуже сильно впливає на час сканування (до речі, DPI теж впливає, але ті так сильно, і тільки коли користувач ставить високі значення).

Зазвичай виділяють три режими:

- Чорно-білий (відмінно підійде для простого тексту);

- Сірий (підійде для тексту з таблицями і малюнками);

- Кольоровий (для кольорових журналів, книг, в загальному, документів, де важлива кольоровість).

Зазвичай від вибору кольоровості залежить час сканування. Адже якщо документ у вас великий, то навіть зайві 5-10 секунд на сторінці в цілому виллються в пристойний час ...

**3) Фотографії**

Документ ви можете отримати не тільки скануванням, але і сфотографувавши його. Як правило, в цьому випадку у вас будуть деякі інші проблеми: спотворення картинки, розмиття. Через це може знадобитися більш тривала подальша правка і обробка отриманого тексту.

Важливо відзначити, що не кожен такий документ вийде розпізнати, тому що якість сканування у нього може бути вкрай низьким.

**Розпізнавання тексту документа**

Найчастіше відскановані сторінки є формати: tif, bmb, jpg, png. В общем-то, для ABBYY FineReader - це не сильно важливо.

Після відкриття в ABBYY FineReader картинки, програма, як правило, на автоматі починає виділяти області і розпізнавати їх. Але іноді вона робить це не правильно. Для цього-то ми і розглянемо виділення потрібних областей вручну.

Важливо! Не всі відразу розуміють, що після відкриття документа в програмі, зліва у вікні відображається вихідний документ, в якому ви і виділяєте різні області. Після натискання на кнопку "розпізнавання" програма в вікні праворуч виведе вам готовий текст. Після розпізнавання, до речі, доцільно перевірити текст на помилки в тому ж самому FineReader.

**Текст**

Ця область використовується для виділення тексту. Картинки і таблиці потрібно виключати з неї. Рідкісні і незвичайний шрифти доведеться вводити вручну ...

Для виділення текстової області, зверніть увагу на панель у верхній частині FineReader. Там є кнопка "Т". Натискаючи на неї, потім на зображенні нижче виділяєте акуратно прямокутну область, в якій розташовується текст. До речі, в деяких випадках потрібно створювати текстових блоків по 2-3, а іноді по 10-12 на сторінку, тому що форматування тексту може бути різним і одним прямокутником всю область і не виділити.

Важливо відзначити, що в текстову область не повинні потрапляти картинки! Надалі це вам заощадить купу часу.

**Картинки**

Використовується для виділення картинок і тих областей, які важко розпізнати через погану якість, або незвичайності шрифту.

Зазвичай цю область використовують для виділення погано відсканованих таблиць, для виділення нестандартного тексту та шрифту, само-собою картинок.

**Таблиці**

На скріншоті нижче показана кнопка для виділення таблиць. Взагалі, особисто я її використовую вкрай рідко. Справа в тому, що вам доведеться досить рутинно малювати (фактично) кожну лінію на таблиці і показувати що і як програмою. Якщо таблиця невелика і в не дуже хорошій якості, я рекомендую для цих цілей використовувати область "картинка". Тим самим заощадите купу часу, а таблицю можна потім в Word зробити швиденько на основі картинки.

**Непотрібні елементи**

Важливо відмітити. Іноді на сторінці є непотрібні елементи, які заважають розпізнати текст, або взагалі не дають вам виділити потрібну область. Їх можна за допомогою "гумки" видалити зовсім.

Для цього переходимо в режим редагування зображення.

Вибираємо інструмент "ластик" і виділяємо непотрібну область. Вона зітреться і на її місці буде білий аркуш паперу.

**Перевірка помилок і збереження результатів роботи**

**По-перше, потрібна перевірка документа.**

Щоб її включити, після розпізнавання, у вікні праворуч, буде кнопка "перевірка". Після її натискання програма FineReader буде автоматично показувати вам ті області, де у програми виникли помилки і вона не змогла достовірно визначити той чи інший символ. Вам залишиться тільки вибирати, або ви згодні з думкою програми, або вводите свій символ.

До речі, в половині випадків, приблизно, програма буде вам пропонувати готове правильне слово - вам залишиться лише тоді мишкою вибрати потрібний варіант.

**Після перевірки потрібно вибрати формат, в який ви збережете результат своєї роботи**.

Тут FineReader дає вам розвернутися на повну котушку: можна просто передати інформацію в Word один в один, а можна зберегти її в одному з десятків форматів. Але хотілося б виділити інший важливий аспект. Який формат б не вибрали, більш важливо вибрати тип копії! Розглянемо найцікавіші варіанти.

**Точна копія**

Всі області, які ви виділяли на сторінці в розпізнаний документі будуть відповідати точнісінько вхідного документа. Дуже зручний варіант, коли вам важливо не втратити форматування тексту. До речі, шрифти так само будуть дуже схожі на оригінал. Рекомендую при такому варіанті передавати документ у Word, щоб уже там продовжити подальшу роботу.

**Копія, що редагується**

Цей варіант хороший тим, що ви отримаєте вже форматований варіант тексту. Тобто відступів з "кілометр", які можливо були в початковому документі - ви не зустрінете. Корисна опція, коли ви будете значно редагувати інформацію.

Правда, не варто вибирати, якщо вам важливо зберегти стилістику оформлення, шрифти, відступи. Іноді, якщо розпізнавання пройшло не дуже успішно - ваш документ може "перекосити" через зміненого форматування. У цьому випадку доцільно вибрати точну копію.

**Простий текст**

Варіант для тих, кому потрібен просто текст зі сторінки без всього іншого. Підійде для документів без картинок і таблиць.

**Питання до закріплення**

1. Розпізнавання тексту – це…?
2. Етапи сканування - …?
3. Вибір режиму сканування - …?
4. Розпізнавання тексту - …?
5. Перевірка результату і усунення помилок - …?
6. Форматування документа - …?
7. Режим точна копія – це…?
8. Режим копія, що редагується – це…?
9. Режим простий текст – це…?

**Майстер в/н А.О. Божко**