Практическая работа № 18.
Сканирование и распознавание текстов

Уровень A.

1) Выберите любую страницу с напечатанным на ней текстом.
2) Сфотографируйте эту страницу или сделайте ее скан (если есть сканер) с разрешением не меньше 300 dpi
3) С помощью программы или онлайн-сервисов (например, http://en.pdf24.org/onlineConverter.html) преобразуйте полученную картинку в формат PDF.
4) Найдите в Интернете информацию о формате DjVu.
5) Используя сервис https://convertonlinefree.com/djvutopdfru.aspx, преобразуйте картинку в формат DjVu. Сравните качество и размеры полученных PDF-файла и DjVu-файла.

Формат	JPG	PDF	DjVu
Качество (1-5)			
Размер			

6) Выполните распознавание текста с помощью программы (например, *CuneiForm*) или онлайн-сервисов, например:
   http://www.newocr.com/
   http://www.free-ocr.com/
   http://www.ocronline.com/ru/
   http://www.onlineocr.net/ru/
   http://finereader.abbyyonline.com/
7) Сохраните результат распознавания в виде документа в формате DOC и исправьте ошибки распознавания, если они были. Затем преобразуйте документ в формат PDF.

Уровень B.

8) Преобразуйте полученный PDF-документ в формат DjVu (с помощью сервиса https://convertonlinefree.com/djvutopdfru.aspx) и сравните размеры полученных файлов, заполнив таблицу:

Тип документа	**Размер**
Без распознавания	
Формат JPG	
Формат PDF	
Формат DjVu	
После распознавания	
Формат DOC	
Формат PDF	
Формат DjVu	

9) Отсканируйте ту же страницу с разрешениями 200 ppi, 150 ppi и 100 ppi, и проверьте, как изменяется качество распознавания. Сделайте выводы.

Уровень C.

10) Попробуйте отсканировать и распознать страницу документа, содержащего рисунки, таблицы и формулы. Сделайте выводы.