

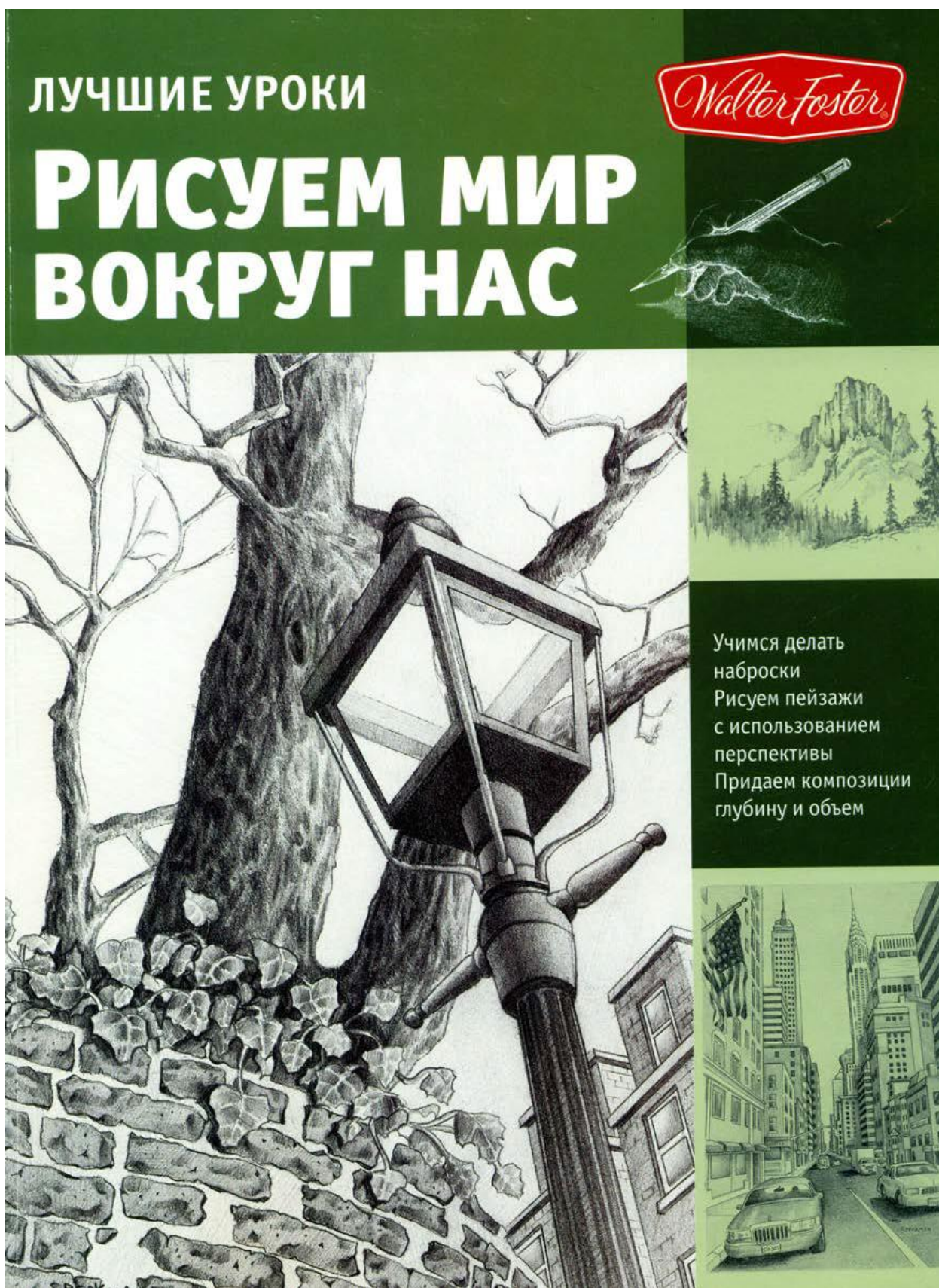
Лабораторная работа №6

«Конструкция, подключение и инсталляция сканера»

Цель работы: знакомство с программами оптического распознавания текстов. Сравнительный анализ бесплатных интернет OCR сервисов

Сервис	Адрес	Регистрация	Количество файлов и их размер	Входные форматы	Выходные форматы	Поддерживаемые языки	Качество распознавания
Online OCR2	http://www.onlineocr.net	Пользователь без регистрации ограничен в количестве страниц	15 страниц, до 15 МБ	5: PDF, JPG, BMP, TIFF, GIF	3: DOCX, XLSX, TXT	46 языков	Хорошее
Yandex OCR	https://translate.yandex.ru/ocr	Не нужна	Количество не ограничено, размер до 5 МБ	7: DOC, DOCX, PDF, XLS, XLSX, PPT, PPTX	1: TXT	97 языков	Плохое, текст не распознан
Convertio	https://convertio.co/ru/pdf-doc/	Пользователь без регистрации ограничен в количестве страниц	до 10 страниц, до 5 МБ	14: PDF, JPG, BMP, GIF, JP2, JPEG, PBM, PCX, PGM, PNG, PPM, TGA, TIFF, WBMP	11: DOCX, PDF, XLSX и TXT	76 языков	Отличное

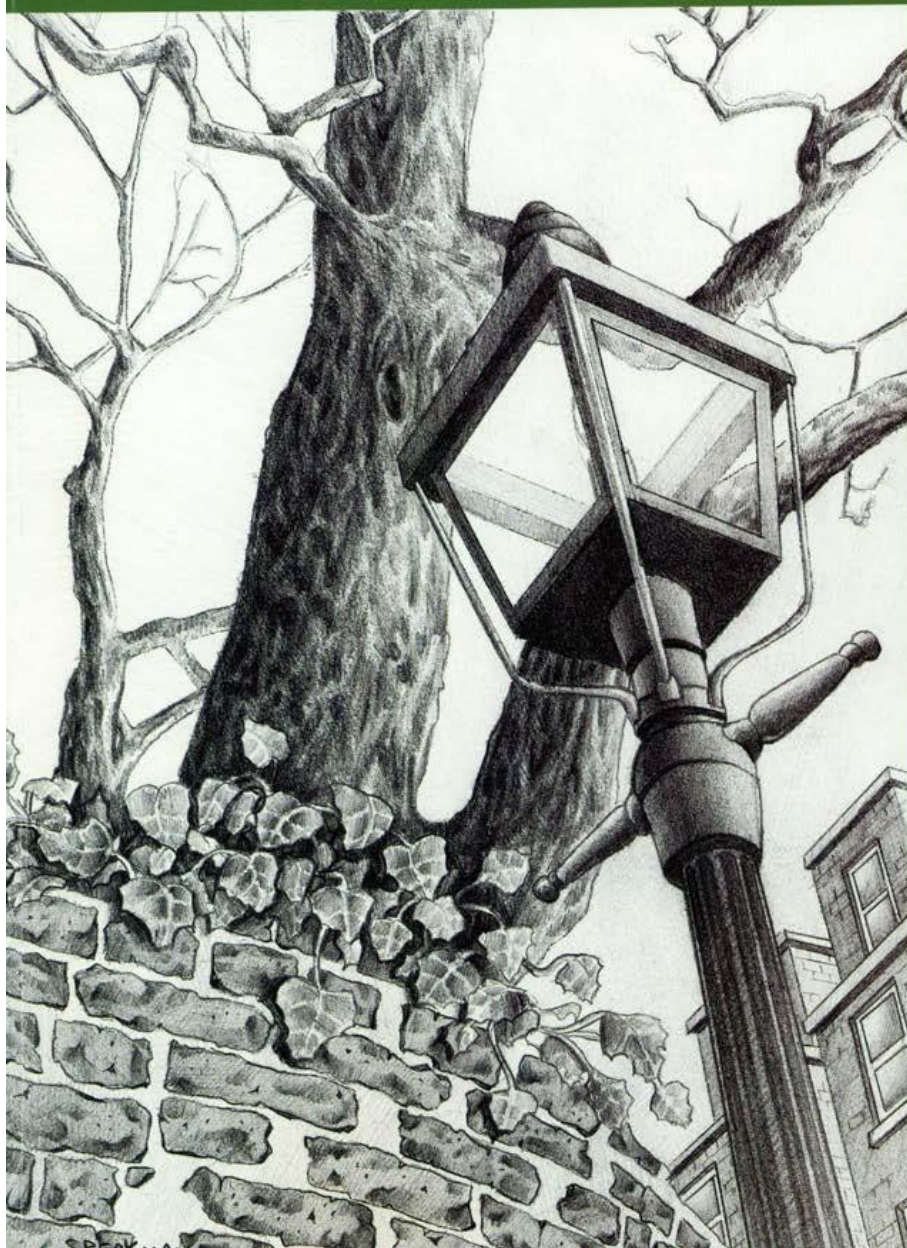
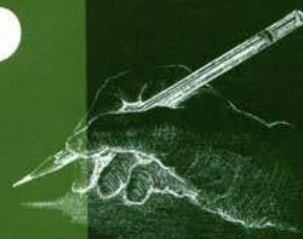
Результат конвертации, при помощи сайта Yandex OCR (из PDF в DOCX),
файла Рис6.pdf



ЛУЧШИЕ УРОКИ

РИСУЕМ МИР ВОКРУГ НАС

Walter Foster®



Учимся делать
наброски
Рисуем пейзажи
с использованием
перспективы
Придаем композиции
глубину и объем



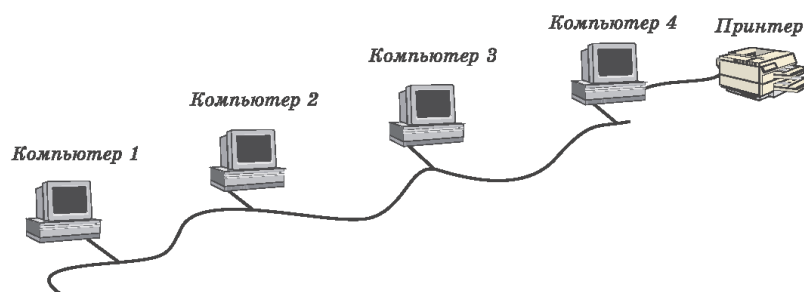


Рис. 1.1. Простейшая сеть: несколько компьютеров и общий принтер

Классификация компьютерных сетей

Возможно множество различных способов классификации компьютерных сетей. Здесь мы рассмотрим только основные из них.

- В зависимости от расстояния между связываемыми узлами сети можно разделить на три основных класса: локальные, региональные и глобальные (рис. 1.2).

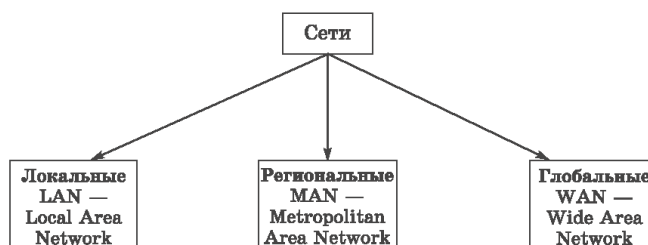


Рис. 1.2. Классификация сетей по расстоянию между узлами

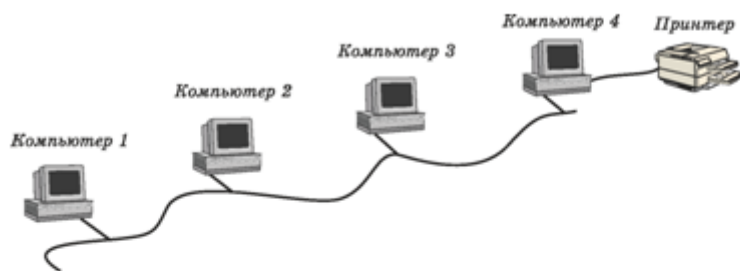


Рис. 1.1. Простейшая сеть: несколько компьютеров и общий принтер

Классификация компьютерных сетей

Возможно множество различных способов классификации компьютерных сетей, здесь мы рассмотрим только основные из них.

В зависимости от расстояния между связываемыми узлами сети можно разделить на три основных класса: локальные, региональные и глобальные (рис. 1.2).

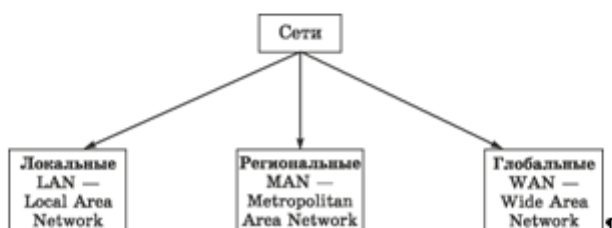


Рис. 1.2. Классификация сетей по расстоянию между узлами

Результат конвертации, при помощи сайта Convertio (из PDF в DOCX), файла Скан5.pdf

Прежде всего выясним, с какими объектами мы будем работать. Как и в других приложениях Microsoft Office, можно выделить две основные группы объектов: документы и инструменты.

Все документы MS Word создаются на основе выбранного шаблона (образца).

Шаблон — служебный файл, который содержит всевозможную информацию о структуре и оформлении документов конкретного типа.

Использование шаблонов позволяет создавать документы таким образом, чтобы все элементы оформления органично сочетались между собой.

Шаблоны являются специализированными программами и сохраняются как файлы с расширением .dot. В стандартный комплект пакета MS Office включены шаблоны для наиболее распространенных типов документов. При запуске Word без указания имени файла процессор по умолчанию предлагает начать создание нового документа под именем «Документ 1» на основе стандартного шаблона «Обычный». На рис. 5.2 показано, из каких объектов состоит объект *Документ*, а в табл. 5.2 приведены примеры свойств некоторых объектов.

Поясним некоторые термины.

Кегль (размер шрифта) — вертикальный размер, измеряемый в пунктах (1 пункт равняется 0,376 мм). Для большинства документов используются 10—12-пунктовые шрифты, в то время как газетная полоса может иметь только 8-пунктовые шрифты. Размер шрифта более 14 пунктов обычно используется лишь для заголовков и выделений.

Гарнитура (тип шрифта) определяет особенности написания целого набора символов, включающего в себя заглавные и строчные буквы, цифры, знаки пунктуации и специальные символы. Каждый тип шрифта имеет несколько стилей начертания символов (например, полужирный, курсив, полужирный курсив, обычный). Кроме того, можно ввести подчеркивание символов и фрагментов текста.

С помощью инструментов мы производим различные операции над документами. Так, оперируя различными символами, можно ввести текст, придать ему необходимый вид, расположить на странице.

Выполняя операции с фигурным текстом, можно оформить красивый заголовок на титульном листе. Операции с рисунками позволяют украсить внешний вид вашего документа и сделать его содержание более понятным.

Символ — это минимальный элемент текста. Он обладает следующими свойствами:

- размер;
- начертание (обычное, жирное, курсивное, подчеркивание);

Рис. 5.2. Классификация объектов, составляющих объект *Документ*

Объект	Свойства
Текст	1. Вид шрифта (кегель, гарнитура, цвет). 2. Стил абзаца. 3. Параметры страницы
Фигурный текст	1. Вид шрифта (кегель, гарнитура, цвет). 2. Формат текста. 3. Специальные эффекты (тень, наклон, поворот). 4. Размещение на странице
Рисунок	1. Формат графического файла. 2. Размер, масштаб. 3. Цветовое решение. 4. Размещение на странице
Таблица	1. Стил. 2. Число строк, столбцов. 3. Формат ячейки

Прежде всего выясним, с какими объектами мы будем работать. Как и в других приложениях Microsoft Office, можно выделить две основные группы объектов: документы и инструменты.

Все документы MS Word создаются на основе выбранного шаблона (образца).

Шаблон — служебный файл, который содержит всевозможную информацию о структуре и оформлении документов конкретного типа.

Использование шаблонов позволяет создавать документы таким образом, чтобы все элементы оформления органично сочетались между собой.

Шаблоны являются специализированными программами и сохраняются как файлы с расширением .dot. В стандартный комплект пакета MS Office включены шаблоны для наиболее распространенных типов документов. При запуске Word без указания имени файла процессор по умолчанию предлагает начать создание нового документа под именем «Документ 1» на основе стандартного шаблона «Обычный». На рис. 5.2 показано, из каких объектов состоит объект *Документ*, а в табл. 5.2 приведены примеры свойств некоторых объектов.

Поясним некоторые термины.

Кегль (размер шрифта) — вертикальный размер, измеряемый в пунктах (1 пункт равняется 0,376 мм). Для большинства документов используются 10—12-пунктовые шрифты, в то время как газетная полоса может иметь только 8-пунктовые шрифты. Размер шрифта более 14 пунктов обычно используется лишь для заголовков и выделений.

Гарнитура (тип шрифта) определяет особенности написания целого набора символов, включающего в себя заглавные и строчные буквы, цифры, знаки пунктуации и специальные символы. Каждый тип шрифта имеет несколько стилей начертания символов (например, полужирный, курсив, полужирный курсив, обычный). Кроме того, можно ввести подчеркивание символов и фрагментов текста.

С помощью инструментов мы производим различные операции над документами. Так, оперируя различными символами, можно ввести текст, придать ему необходимый вид, расположить на странице.

Выполняя операции с фигурным текстом, можно оформить красивый заголовок на титульном листе. Операции с рисунками позволяют украсить внешний вид вашего документа и сделать его содержание более понятным.

Символ — это минимальный элемент текста. Он обладает следующими свойствами:

- размер;
- начертание (обычное, жирное, курсивное, подчеркивание);

Рис. 5.2. Классификация объектов, составляющих объект *Документ*

Объект	Свойства
Текст	1. Вид шрифта (кегель, гарнитура, цвет). 2. Стил абзаца. 3. Параметры страницы
Фигурный текст	1. Вид шрифта (кегель, гарнитура, цвет). 2. Формат текста. 3. Специальные эффекты (тень, наклон, поворот). 4. Размещение на странице
Рисунок	1. Формат графического файла. 2. Размер, масштаб. 3. Цветовое решение. 4. Размещение на странице
Таблица	1. Стил. 2. Число строк, столбцов. 3. Формат ячейки

Прежде всего выясним, с какими объектами мы будем работать. Как и в других приложениях Microsoft Office, можно выделить две основные группы объектов: документы и инструменты.

Все документы MS Word создаются на основе выбранного шаблона (образца).

Шаблон — служебный файл, который содержит всевозможную информацию о структуре и оформлении документов конкретного типа.

Использование шаблонов позволяет создавать документы таким образом, чтобы все элементы оформления органично сочетались между собой.

Шаблоны являются специализированными программами и сохраняются как файлы с расширением .dot. В стандартный комплект пакета MS Office включены шаблоны для наиболее распространенных типов документов. При запуске Word без указания имени файла процессор по умолчанию предлагает начать создание нового документа под именем «Документ 1» на основе стандартного шаблона «Обычный». На рис. 5.2 показано, из каких объектов состоит объект *Документ*, а в табл. 5.2 приведены примеры свойств некоторых объектов.

Поясним некоторые термины.

Кегль (размер шрифта) — вертикальный размер, измеряемый в пунктах (1 пункт равняется 0,376 мм). Для большинства документов используются 10—12-пунктовые шрифты, в то время как газетная полоса может иметь только 8-пунктовые шрифты. Размер шрифта более 14 пунктов обычно используется лишь для заголовков и выделений.

Гарнитура (тип шрифта) определяет особенности написания целого набора символов, включающего в себя заглавные и строчные буквы, цифры, знаки пунктуации и специальные символы. Каждый тип шрифта имеет несколько стилей начертания символов (например, полужирный, курсив, полужирный курсив, обычный). Кроме того, можно ввести подчеркивание символов и фрагментов текста.

С помощью инструментов мы производим различные операции над документами. Так, оперируя различными символами, можно ввести текст, придать ему необходимый вид, расположить на странице.

Выполняя операции с фигурным текстом, можно оформить красивый заголовок на титульном листе. Операции с рисунками позволяют украсить внешний вид вашего документа и сделать его содержание более понятным.

Символ — это минимальный элемент текста. Он обладает следующими свойствами:

- размер;
- начертание (обычное, жирное, курсивное, подчеркивание);

Рис. 5.2. Классификация объектов, составляющих объект *Документ*

Объект	Свойства
Текст	1. Вид шрифта (кегель, гарнитура, цвет). 2. Стил абзаца. 3. Параметры страницы
Фигурный текст	1. Вид шрифта (кегель, гарнитура, цвет). 2. Формат текста. 3. Специальные эффекты (тень, наклон, поворот). 4. Размещение на странице
Рисунок	1. Формат графического файла. 2. Размер, масштаб. 3. Цветовое решение. 4. Размещение на странице
Таблица	1. Стил. 2. Число строк, столбцов. 3. Формат ячейки

Вывод: всего в ходе работы было протестировано 3 сайта: Online OCR2, Yandex OCR и Convertio. Лучше всего с задачей справился сайт Convertio, таблицы были верно конвертированы, текст совпадал с исходным (слова на английском языке были распознаны), даже колонтитулы были верно сделаны. Online OCR2 также вполне приемлемо распознал текст, однако его дополнительно нужно откорректировать. Yandex OCR не смог распознать текст на картинке и работал некорректно