

30.03.2022

Тема: ИНСТРУМЕНТЫ РАСПОЗНАВАНИЯ ТЕКСТОВ И СИСТЕМЫ КОМПЬЮТЕРНОГО ПЕРЕВОДА. ОЦЕНКА КОЛИЧЕСТВЕННЫХ ПАРАМЕТРОВ ТЕКСТОВЫХ ДОКУМЕНТОВ.

Цель урока: ИЗУЧИТЬ ИНСТРУМЕНТЫ РАСПОЗНАВАНИЯ ТЕКСТОВ И СИСТЕМЫ КОМПЬЮТЕРНОГО ПЕРЕВОДА. НАУЧИТЬСЯ ОЦЕНИВАТЬ КОЛИЧЕСТВЕННЫЕ ПАРАМЕТРЫ ТЕКСТОВОГО ДОКУМЕНТА.

Ход урока

Изучение нового материала

Вопросы занятия:

- распознавание документов;
- компьютерные словари;
- программы-переводчики.

Современный мир меняется и улучшается каждую минуту. Ещё недавно для того чтобы почитать мы были вынуждены покупать книги или ходить в библиотеку и брать книги там. А сейчас достаточно зайти в интернет, найти интересную книгу и читать её с компьютера, телефона или специального устройства – электронной книги.

А задумывались вы, как в интернет попали все эти книги. Можно предположить, что какой-то ответственный человек сидел и набирал тексты книг на компьютере.

Как вы думаете, сколько он потратил на это времени? Давайте попробуем подсчитать. Предположим, что наш человек очень опытный наборщик текстов и его скорость набора 180 знаков в минуту. Ему нужно ввести в компьютер текст романа «Война и мир».



Посчитаем время набора текста романа «Война и мир».

1. $v = 180 \text{ зн./мин}$

2. $N = 478\,458 \text{ слов}, N_p = 437\,699 \text{ слов},$
 $N_\phi = 15\,040 \text{ слов}, N_n = 652 \text{ слова.}$

3. $N_s = 2\,966\,547.$

$$T = \frac{N_s}{v} = \frac{2\,966\,547}{180} \approx 16\,481 \text{ (мин)} \approx 275 \text{ (ч)} \approx 11 \text{ (сут.)}$$

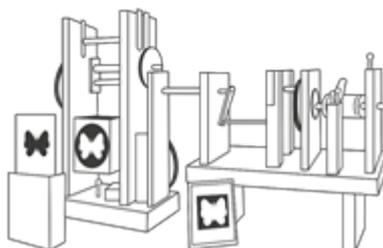
Как вы уже догадались, данный способ перевода печатного текста на компьютер не эффективен и не практичен.

Гораздо быстрее и удобнее использовать специальный инструмент, такой как например сканер, который за считанные секунды переведёт любую бумажную информацию в цифровую.



А когда же появился сканер? Давайте рассмотрим историю появления этого инструмента.

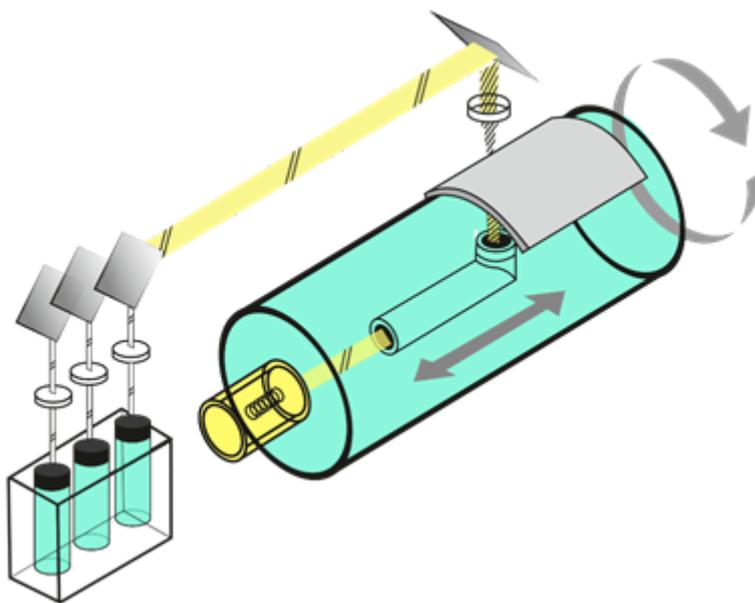
Можно сказать, что создание сканера началось со времён изобретения всем известного телеграфа. Был изобретён прибор, который передавал изображение на расстояния.



Но очень бурное развитие сканера началось в начале XX века, в те времена, когда был изобретён фототелеграф, как мы привыкли его называть – телефакс. В 1902 году, немецкий физик Артур Корн запатентовал технологию фотоэлектрического сканирования, получившую впоследствии название телефакс.



Передаваемое изображение закреплялось на прозрачном вращающемся барабане, луч света от лампы, перемещающейся вдоль оси барабана, проходил сквозь оригинал и через расположенные на оси барабана призму и объектив попадал на селеновый фотоприёмник.



Эта технология до сих пор применяется в барабанных сканерах.



После этого учёными было сделано, новое прорывное в будущее изобретение – был изобретён новый способ сканирования, который гордо носит название – планшетный.

Рассмотрим принцип действия планшетных сканеров.

Сканируемый объект кладётся на стекло планшета сканируемой поверхностью вниз. Под стеклом располагается подвижная лампа, движение которой регулируется шаговым двигателем. Свет, отражённый от объекта, через систему зеркал попадает на чувствительную матрицу и передаётся в компьютер. За каждый шаг двигателя сканируется полоска объекта, все полоски потом объединяются программным обеспечением в общее изображение.

Данная конструкция имеет ряд преимуществ перед своими предками, основное из которых – это то, что сканер передаёт изображение в компьютер.

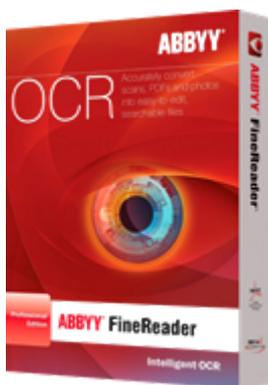
На сегодняшний день рынок компьютерной техники насчитывает огромное количество моделей сканеров.



Многие привыкли называть сканер – фотоаппарат, который фотографирует изображения.

Сканеры используются для ввода изображений на компьютер. С помощью сканера мы получаем на своём компьютере электронный снимок. Сканер иногда заменяет нам факс. Любой документ, любую фотографию мы можем отсканировать и передать по электронной почте.

Итак, отсканированный снимок попал в компьютер. Но дальше эту информацию нужно распознать, отредактировать и изменить. Для этого нам нужно воспользоваться специальной программой ABBYY FineReader.



Остановимся подробнее на работе такой программы:

- 1) В начале, вам необходимо отсканировать текст с помощью сканера. Для этого вы помещаете текст под крышку сканера, и в программе отдаёте команду **Сканировать и распознать**. Тем самым мы создаём цифровую копию исходного документа в формате графического изображения.
- 2) Затем программа анализирует структуру документа, выделяя на его страницах блоки текста, таблицы и картинки. Строки разбиваются на слова, а слова на буквы. Затем программа сравнивает найденные символы с шаблонами из своей памяти. Кроме того, в программу встроены словари, которые позволяют производить более точный анализ и распознавание и проверку распознанного текста. Проанализировав огромное количество вариантов, программа выдаёт окончательное решение пользователю распознанного текста.
- 3) И наконец, распознанный текст переносится в окно текстового редактора, например, Microsoft Word.

Однако, качество распознанного текста зависит от того как хорошо виден исходный текст. Отсканированный и распознанный текст обязательно нужно перечитать и отредактировать.

Современные технологии позволяют не использовать сканер, а вместо него воспользоваться фотоаппаратом или даже телефоном.

А теперь вспомните те далёкие годы, когда ещё не было печатных станков и нашим предкам приходилось вручную переписывать рукописи, чтобы они дошли до наших дней. Каким тяжёлым и важным был труд писарей, и как ценна была каждая книга.



Теперь разберёмся для чего предназначены компьютерные словари и программы-переводчики

Основная цель современных компьютерных технологий упростить, улучшить жизнь человека, делать её комфортнее и динамичнее. Одним из таких удобств являются компьютерные словари и программы-переводчики, которые позволяют мгновенно найти нужное слово, перевести его и даже если нужно - прослушать произношение. Ещё совсем недавно это казалось невозможным. Люди, которые попадали за границу, должны были иметь при себе разговорник или карманный словарь, чтобы иметь возможность общаться с местными жителями.



Сегодня достаточно установить специальное приложение на телефон и общение проходит быстро и комфортно. Вы можете, как набрать на телефоне нужную фразу, так и произнести её, а приложение переведёт её на нужный язык.

Остановимся подробнее на программах переводчиках, установленных на компьютер.



Данные программы основаны на формальном знании языка – правил словообразования и правил построения предложения. Программа в начале анализирует текст на исходном языке, а затем конструирует этот текст на том языке, который необходим. Данные программы предназначены для перевода технической документации, деловых переписок и другие материалов, написанных на так называемым «сухим» языке. А вот перевод художественных текстов, эмоционально окрашенных, содержащих шутки, необычные сравнения может выполнить только человек.

Перейдём к практической части урока.

Давайте попробуем воспользоваться возможностями программы-переводчика Гугл и переведём английский язык следующий текст: «Информатика (от «информации» и «автоматика») — это наука о методах и процессах сбора, хранения, обработки, передачи, анализа и оценки информации с применением компьютерных технологий, обеспечивающих возможность её использования для принятия решений».

Затем полученный результат с помощью той же программы переведём на русский язык и сравним исходной текст и конечный результат.

Итак, откроем поисковую систему интернета Гугл. в поисковик запишем переводчик Гугл. В результатах поиска нажимаем на Открыть Google Переводчик.

Скопируем текст из текстового редактора и вставим в окно переводчика. В окошках исходный текст и перевод выберем необходимые языки.

Как видим наш текст мгновенно перевёлся на английский язык.

Теперь скопируем текст на английском языке и вставим его в окно Исходный текст.

Сравним исходный текст и переведённый.

«Информатика (от «информация» и «автоматика») — это наука о методах и процессах сбора, хранения, обработки, передачи, анализа и оценки информации с применением компьютерных технологий, обеспечивающих возможность её использования для принятия решений».

"Компьютерная наука (от «информация» и «автоматика») - наука о методах и процессах сбора, хранения, обработки, передачи, анализа и оценки информации с использованием компьютерных технологий, для использования (использования 2 раза) её (информации) для принятия решений (для тоже 2 раза)

Как видим в первом же слове отличия. Слово Информатика переводчик перевёл как Компьютерные науки.

Анализируя текст, мы видим, что переводчик переводит тексты не точно и может допускать ошибки.

То есть у программ есть свои достоинства и недостатки. В зависимости от задачи, одна особенность может и достоинством, и недостатком.

Достоинством программ является высокая скорость перевода. Всего несколько секунд, и Вы получаете перевод многостраничного текста.

Недостатком же можно назвать то, что программы-переводчики работают не точно. Они годятся для того, чтобы в общих чертах передать, о чём идёт речь, и допускают ошибки в переводе.

Пришло время подвести итоги урока.

Для ввода текстов в память компьютера с бумажных носителей используют **сканеры и программы распознавания символов**. Возможности современных компьютеров по хранению больших массивов информации и осуществлению в них быстрого поиска положены в основу разработки **компьютерных словарей и программ-переводчиков**. **Компьютерные словари** выполняют перевод отдельных слов и словосочетаний. Для перевода текстовых документов применяются **программы-переводчики**.

Домашнее задание

Проработать материал п 4.5. стр. 174-177, ответить на вопросы №4,5 с 177

Выполненную работу присылайте учителю на электронную почту

ekaterinaefremova160283@gmail.com