

Системы перевода и распознавания текстов

На дом: §17

Основные вопросы:

- как работают программы-переводчики;
- распознавание текста;
- ввод в компьютер печатного и рукописного текста.

Как работают программы-переводчики

Чтобы найти перевод неизвестного иностранного слова, пользователю электронного словаря достаточно ввести это слово в строке поиска, и уже через несколько мгновений будет получен перевод. Современные текстовые процессоры имеют в своем составе словари, позволяющие производить орфографическую проверку правильности написания слов (на разных языках).

File Alignment Wizard

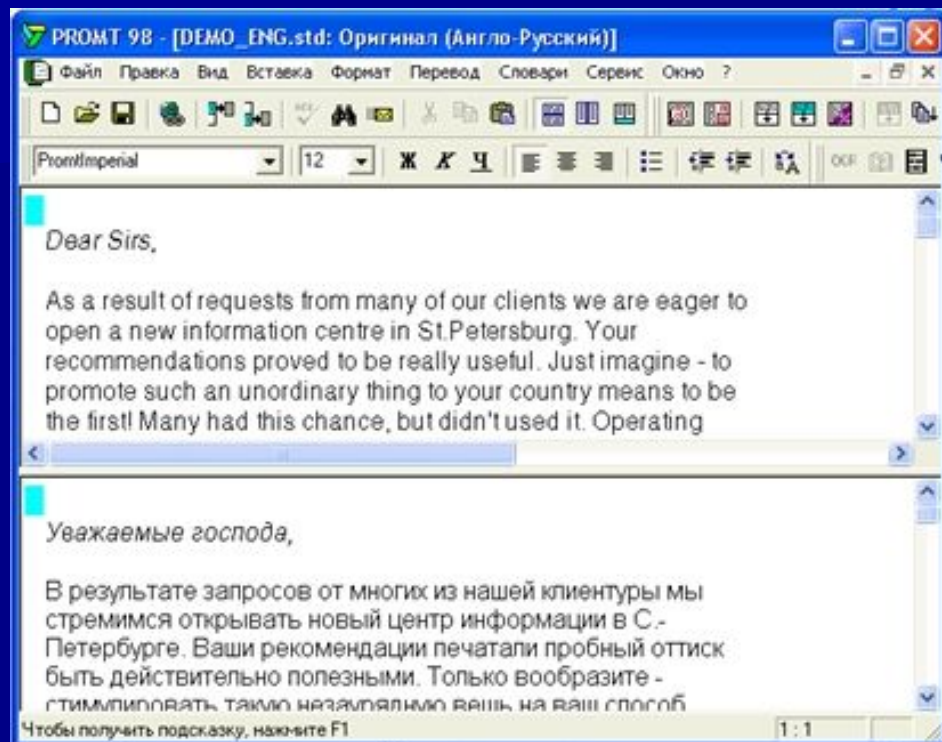
Excerpts from the source and target languages are shown below. Please select the appropriate language for each file from the dropdown list underneath it.

attract attention receive invitation client-server mid-range servers market computing resources	привлекать внимание получить приглашение клиент – серверный рынок серверов среднего к вычислительные ресурсы
---	--

Source language
English

Target language
Russian

< Back **Next >** Cancel



Рассмотрим простой пример. Переведем с помощью системы перевода на английский язык фразу:

Информатика — это наука об информации.

Результат перевода:

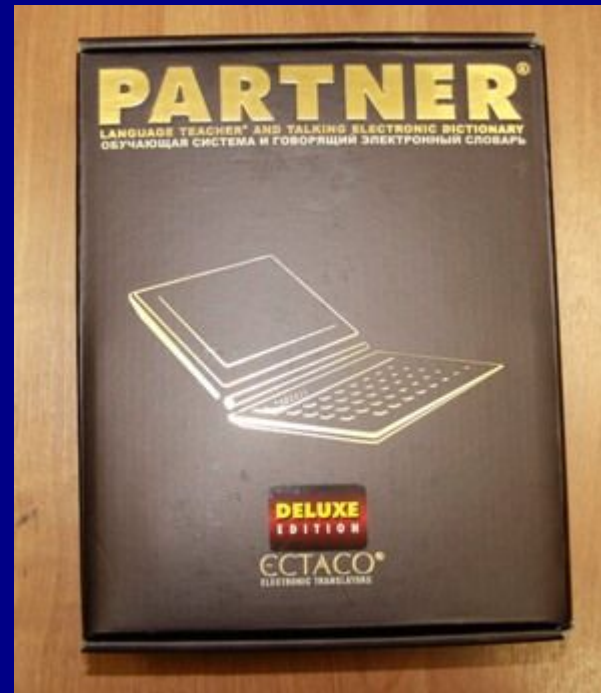
The computer science is an information science.

А теперь с помощью той же программы переведем эту фразу на русский язык. Получим:

Информатика — информатика.

Электронные устройства: Ectaco partner er 900 delux

может не только переводить слова и выражения с русского на английский и с английского на русский языки, но и распознавать вашу речь, проигрывать музыку и даже обучать английскому языку! Переводчик оснащён цветным сенсорным дисплеем, может воспроизводить музыку с карты памяти.



Корпус и комплектующие:



Крупные кнопки, цветной сенсорный экран

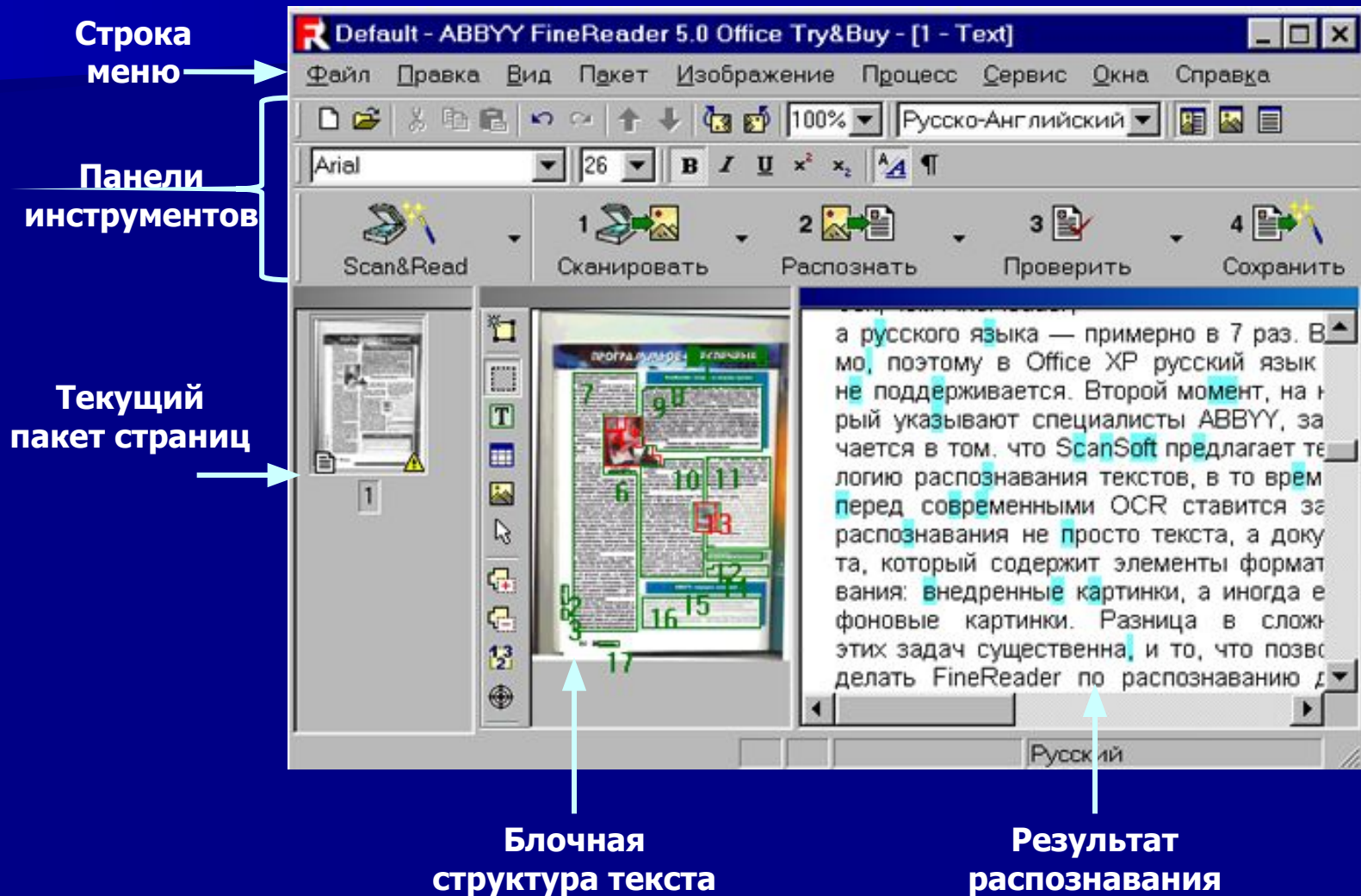


Программы оптического распознавания текста

Например ABBY Fine Reader



Окно программы FineReader

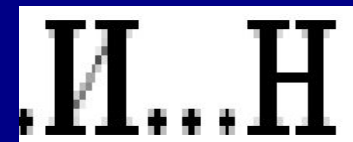


Процесс обработки FineReader

- Сканирование (сканер, цифровой фотоаппарат, цифровая видеокамера).
- Сегментация - выделение блоков на изображении.
- Распознавание – неоднозначно опознанные символы выделяются цветом.
- Проверка ошибок- можно провести проверку грамматики.
- Сохранение результатов в виде отформатированного или неотформатированного документа, или прямой передачи в другое приложение - WORD, Excel в буфер обмена Windows.

Система оптического распознавания СИМВОЛОВ

- Если исходный документ имеет типографическое качество (достаточно крупный шрифт, отсутствие плохо напечатанных символов или исправлений), то задача распознавания решается методом сравнения с растровым шаблоном.



- При распознавании документов с низким качеством печати (машинописный текст, факс и т.д.) используется метод распознавания структурных элементов (отрезков, колец, дуг и др.) символов. В искаженном символьном изображении выделяются характерные детали и сравниваются со структурными шаблонами символов.

Система оптического распознавания форм

Единый государственный экзамен - 2005
Бланк регистрации

Заполнить гелевой или капиллярной ручкой ЧЕРНЫМИ чернилами ЗАГЛАВНЫМИ ПЕЧАТНЫМИ БУКВАМИ по следующему образцу:
А В С D E F G H I J K L M N O P Q R S T U V X Y Z Ч Ш Щ Ъ Ы Ь Э Ю Я І І 2 3 4 5 6 7 8 9 0 X V I L C

ВНИМАНИЕ! Данный бланк использовать только совместно с двумя другими бланками из данного пакета

Сведения об участнике единого государственного экзамена

Фамилия _____
Имя _____
Отчество _____

Документ Серия _____ Номер _____ Пол Ж М

Резерв - 1 _____ Резерв - 2 _____ Резерв - 3 _____ Факт выхода из аудитории во время экзамена

ЗАМЕЧАНИЯ участника ЕГЭ по процедуре проведения ЕГЭ.
Заполнение НЕ ОБЯЗАТЕЛЬНО.
Отметьте замечания по проведению экзамена:

<input type="checkbox"/> Организованная доставка участников в пункт проведения ЕГЭ при самостоятельном времени в пути более 1 часа	<input type="checkbox"/> Присутствие в аудитории преподавателей общеобразовательного предмета, по которому проводится ЕГЭ
<input type="checkbox"/> Вскрытие доставочного пакета осуществлялось НЕ в присутствии участника ЕГЭ	<input type="checkbox"/> Наличие нарушений дисциплины в аудитории

FineReader
Forms

Единый государственный экзамен - 2005
Бланк ответов № 1

Заполнить гелевой или капиллярной ручкой ЧЕРНЫМИ чернилами ЗАГЛАВНЫМИ ПЕЧАТНЫМИ БУКВАМИ по следующему образцу:
А В С D E F G H I J K L M N O P Q R S T U V X Y Z

Внимание! Данный бланк использовать только совместно с двумя другими бланками из данного пакета

Номера заданий типа А с вариантами ответов и предложенных вариантов ответов и правильных ответов

Номер задания	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
Тип А																					
Тип В																					

Результаты выполнения заданий типа В в отрывках в краткой форме

Задание	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
В1																					
В2																					
В3																					
В4																					
В5																					
В6																					

Задания с краткими ответами на заданиях типа В

- **Бланком** называется стандартный лист бумаги, на котором размещается постоянная информация и отведено место для переменной.
- Сложность состоит в том, что необходимо распознать написанные от руки символы, довольно сильно различающиеся у разных людей.
- Кроме того система должна определить, к какому полю относится распознаваемый текст.

Система оптического распознавания форм

- Для обработки бланков предназначено специальное приложение FineReader Forms.
- Для распознавания содержимого бланка необходимо предварительно создать шаблон формы.

Сервис/ Шаблоны

- Шаблон используют на этапе сегментации. Сегментация в данном случае состоит в наложении шаблона.
- Положение шаблона корректируется в соответствии с тем, насколько ровно был размещён бланк при сканировании.
- Заключительный этап состоит в распознавании содержимого бланка.

Система распознавания рукописного текста

- С появлением первого карманного компьютера Newton фирмы Apple в 1990 году начали создаваться системы распознавания рукописного текста. Такие системы преобразуют текст, написанный на экране карманного компьютера специальной ручкой, в текстовый компьютерный документ.

Распознавание рукописного текста

