

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 кл
вычислительные ресурсы

Source language

English

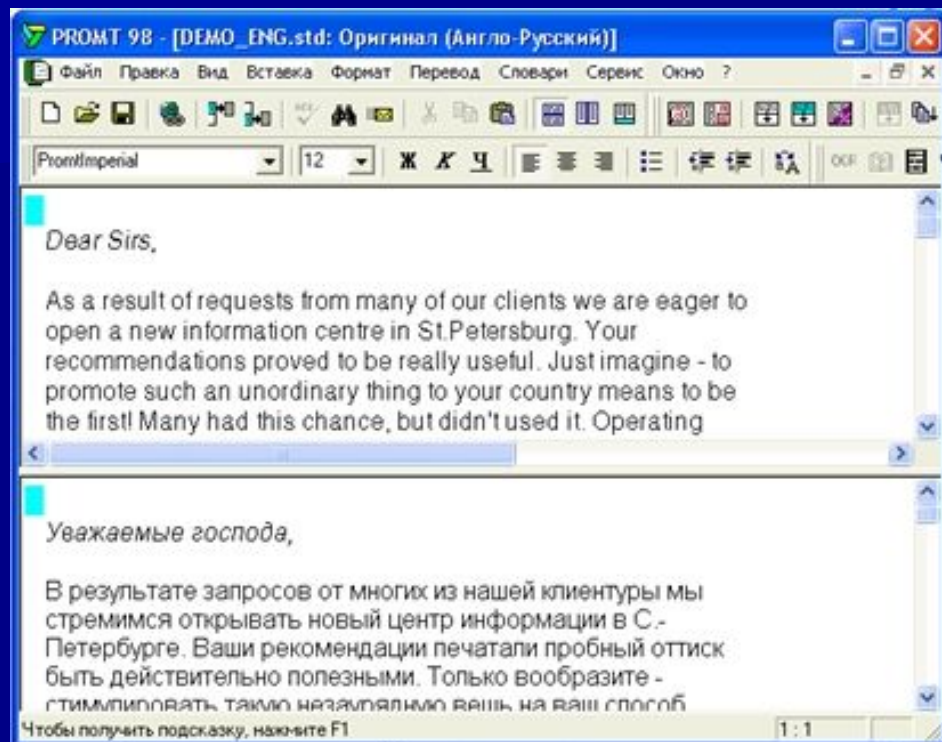
Target language

Russian

< Back

Next >

Cancel



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

Информатика — это наука об информации.

Результат перевода:

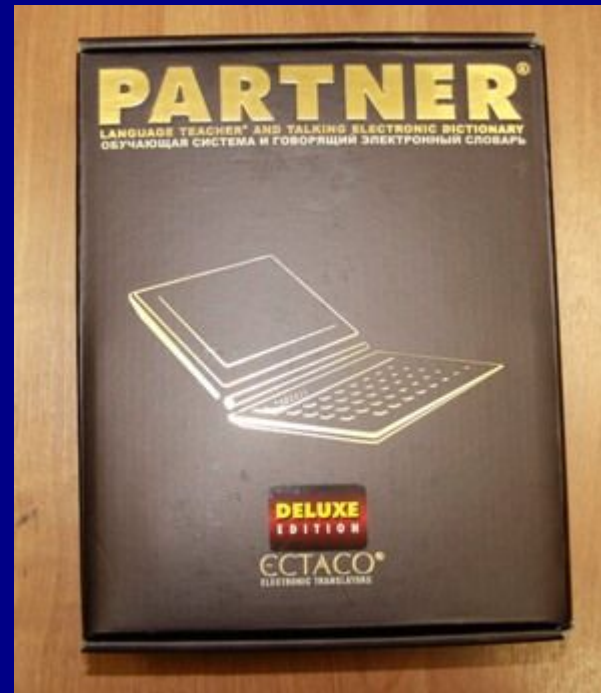
The computer science is an information science.

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

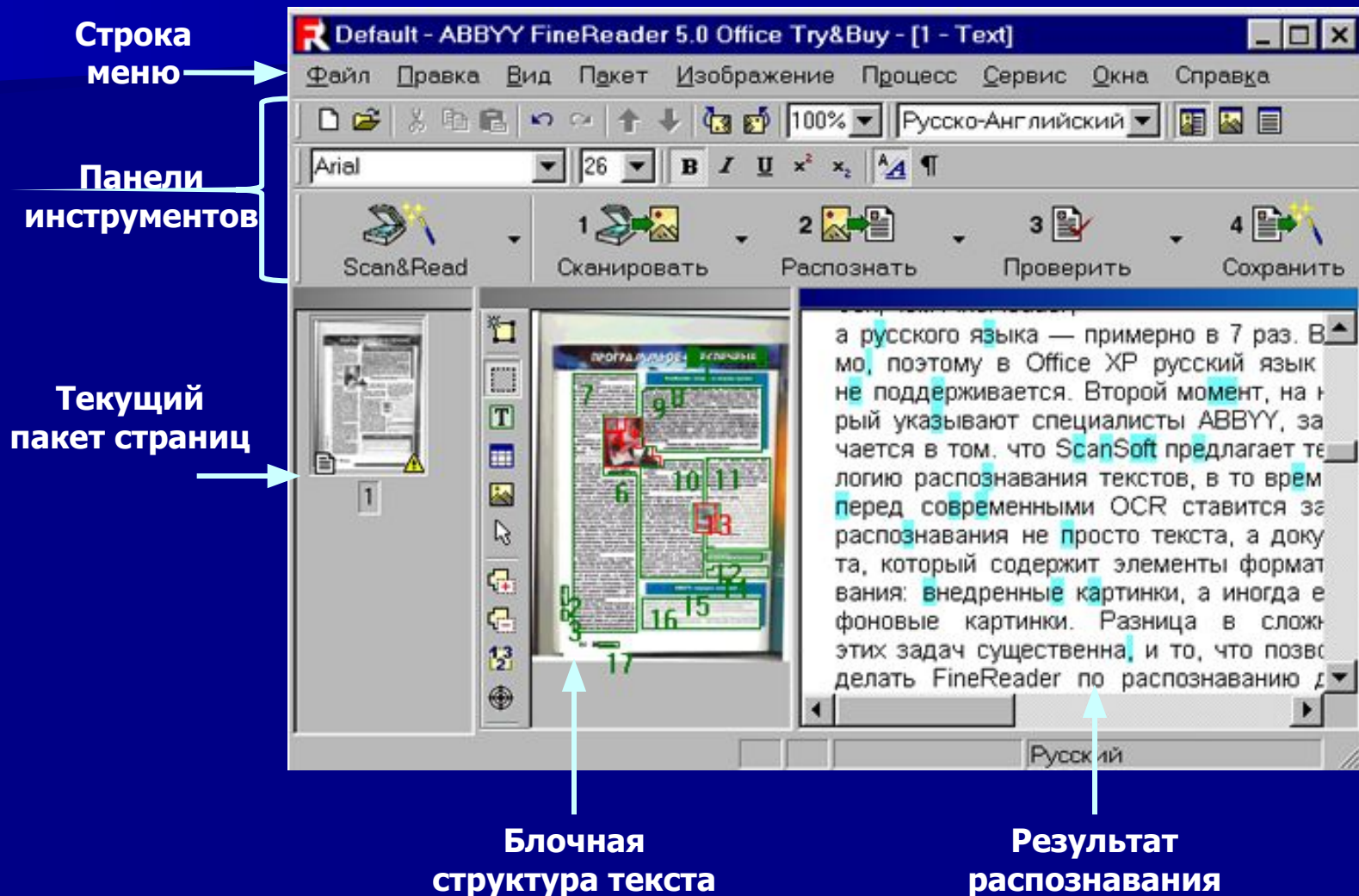


Программы оптического распознавания текста

Например ABBY Fine Reader



Окно программы FineReader

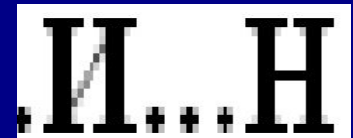


Процесс обработки FineReader

- 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.

форм

FineReader Forms

- **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

