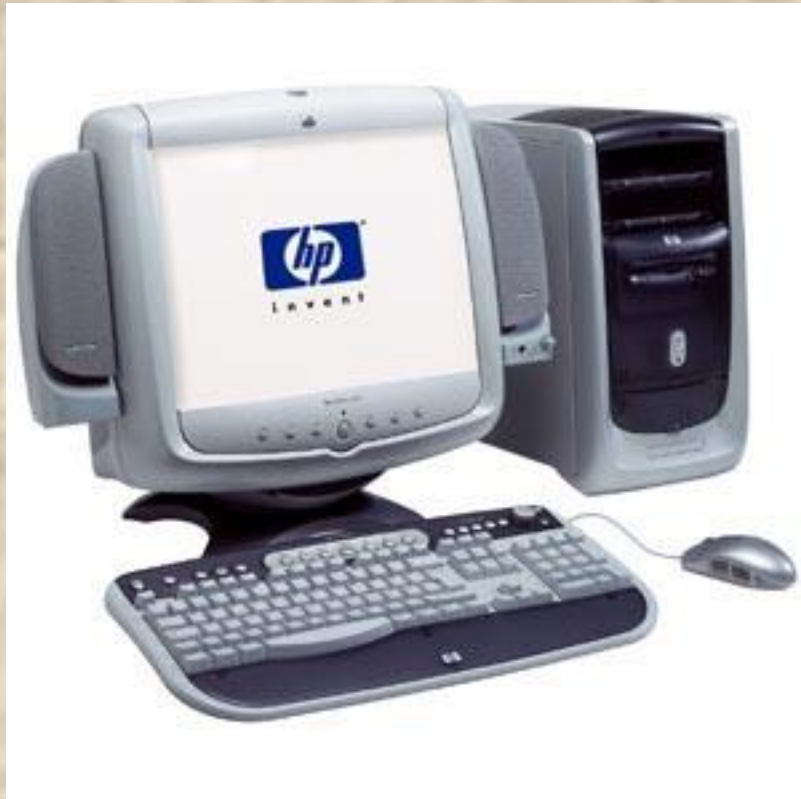
Компьютерные словари и системы машинного перевода текстов

Компьютерные словари

- Компьютерные словари могут содержать переводы на разные языки сотен тысяч слов и словосочетаний, а также предоставляют пользователю дополнительные возможности.



Возможности компьютерных словарей

- Компьютерные словари могут являться многоязычными - давать пользователю возможность выбрать языки и направление перевода (например, англо-русский, испанско-русский и т. д.);
- могут кроме основного словаря общеупотребительных слов содержать десятки специализированных словарей по областям знаний (техника, медицина, информатика и др.).

Возможности компьютерных словарей

- обеспечивают быстрый поиск словарных статей: "быстрый набор", когда в процессе набора слова возникает список похожих слов; доступ к часто используемым словам по закладкам; возможность ввода словосочетаний и др.;
- могут являться мультимедийными, т. е. предоставлять пользователю возможность прослушивания слов в исполнении дикторов, носителей языка.

Системы компьютерного перевода

- Способны переводить многостраничные документы с высокой скоростью (одна страница в секунду);
- переводить Web-страницы "на лету", в режиме реального времени;
- не применимы для перевода художественных произведений, так как не способны адекватно переводить метафоры, аллегории и другие элементы художественного творчества человека.

Системы оптического распознавания документов

- Используются при создании электронных библиотек и архивов путем перевода книг и документов в цифровой компьютерный формат.
- Сначала с помощью сканера необходимо получить изображение страницы текста в графическом формате. Далее для получения документа в текстовом формате необходимо провести распознавание текста, т. е. преобразовать элементы графического изображения в последовательность текстовых символов.

- Растровое изображение каждого символа последовательно накладывается на растровые шаблоны символов, хранящиеся в памяти системы оптического распознавания. Результатом распознавания является символ, шаблон которого в наибольшей степени совпадает с изображением

При распознавании документов с низким качеством печати (машинописный текст, факс и т. д.) используется векторный метод распознавания символов. В распознаваемом изображении символа выделяются геометрические примитивы (отрезки, окружности и др.) и сравниваются с векторными шаблонами символов.

Системы оптического распознавания символов являются "самообучающимися" (для каждого конкретного документа они создают соответствующий набор шаблонов символов), и поэтому скорость и качество распознавания многостраничного документа постепенно возрастают.

Системы оптического распознавания форм

- При заполнении документов большим количеством людей (например, при сдаче выпускником школы единого государственного экзамена (ЕГЭ)) используются бланки с пустыми полями. Данные вводятся в поля печатными буквами от руки. Затем эти данные распознаются с помощью систем оптического распознавания форм и вносятся в компьютерные базы данных.
- Сложность состоит в том, что необходимо распознавать символы, написанные от руки, которые довольно сильно различаются у разных людей. Кроме того, такие системы должны уметь определять, к какому полю относится распознаваемый текст.

Единый государственный экзамен

Бланк регистрации

Код региона | Код образовательного учреждения | Класс: Номер, Буква | Код пункта проведения ЕГЭ | Номер аудитории | Дата проведения ЕГЭ

Код предмета | Название предмета | Служебная отметка | Резерв 1

Заполнять гелевой или капиллярной ручкой ЧЕРНЫМИ чернилами ЗАГЛАВНЫМИ ПЕЧАТНЫМИ БУКВАМИ по следующим образцам:

А Б В Г Д Е Ё Ж З И Й К Л М Н О П Р С Т У Ф Х Ц Ч Ш Щ Ъ Ы Ь Э Ю Я 1 2 3 4 5 6 7 8 9 0 X V I L –

ВНИМАНИЕ! *Все бланки и листы с контрольными измерительными материалами рассматриваются в комплекте.*

Сведения об участнике единого государственного экзамена

Фамилия

Имя

Отчество (при наличии)