**ОУД.07 Информатика(13 группа)**

**-Шатерникова Ю.А .** **126\_ula@mail.ru**

**О**сновная литература

1.Михеева Е.В.,Титова О.И. Информатика: Учебник.- М.: Изд. Центр «Академия».

2.Макарова Н.В. Информатика и ИКТ: Учебник: СПб. «Лидер»

3.Колмыкова Е.А.,Кумскова И.А. Информатика 9-е издание,М.:Изд.центр «Академия»,2012г.

**Дополнительная литература**

1Интернет-ресурсы.

**ТЕМА:Программы переводчики.Возможности систем распознавания текстов.**

**Задание на дом:**

 1.Законспектировать тему.

ЛЕКЦИЯ:Программы переводчики.Возможности систем распознавания текстов.

***Компьютерные словари и системы машинного перевода текстов.***

В настоящее время существуют тысячи словарей для перевода между сотнями языков (англо-русский, немецко-французский и другие), причем каждый из них может содержать десятки тысяч слов.

В бумажном варианте словарь – это толстая книга с большим количеством страниц, поиск в нем довольно трудоемкий процесс.

Компьютерные словари (например, Lingvo, «Контекст») тоже содержат перевод слов, но они предоставляют дополнительные возможности.

* Компьютерные словари в основном являются многоязычными, то есть дают пользователю возможность выбрать языки и направление перевода (например, англо-русский, испано-русский и другие).
* Кроме основного словаря общеупотребительных слов, часто они содержат десятки специализированных словарей по областям знаний (техника, медицина, информатика и другие).
* Они 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.
* Некоторые компьютерные словари предоставляют пользователю возможность прослушивания слов в исполнении дикторов, носителей языка, то есть являются мультимедийными.

Кроме того, существуют системы машинного перевода, позволяющие переводить не только отдельные слова и словосочетания, но и целый документ (текст) с высокой скоростью, и Web-страницу в режиме реального времени.

Лучшими среди российских систем машинного перевода считаются PROMT и «Сократ».

Системы машинного перевода осуществляют перевод текстов, основываясь на формальном «знании» языка (синтаксиса языка – правил построения предложений, правил словообразования) и использовании словарей. Программа-переводчик сначала анализирует текст на одном языке, а затем конструирует этот текст на другом языке.

Современные системы машинного перевода используются для перевода технической документации, деловой переписки и других специализированных текстов,

но они неприменимы для перевода художественной литературы, так как им недоступны аллегории, метафоры и другие элементы художественного творчества человека.

***3. Системы оптического распознавания документов.***

Переход от бумажного документа к электронному состоит из двух этапов.

1. Сканирование. С помощью сканера получается изображение страницы текста в графическом файле.
2. Распознавание текста. Для преобразования элементов графического изображения в последовательности символов используются системы оптического распознавания символов.

Запустив такую систему, сначала надо распознать структуру размещения текста на странице: выделить колонки, таблицы, изображения и так далее. Далее текстовые фрагменты графического изображения страницы преобразовываются в текст.

**Существует два метода распознавания:**

***1. Метод сравнения с растровым шаблоном.***

Используется, если исходный документ имеет типографическое качество (достаточно крупный шрифт, отсутствие плохо напечатанных символов и исправлений).

Сначала растровое изображение страницы разделяется на изображения отдельных символов. Затем каждый из них последовательно накладывается на шаблоны символов (см. рис.), имеющихся в памяти системы, и выбирается шаблон с наименьшим количеством отличных от входного изображения точек.



***2. Метод распознавания символов*** по наличию в них определенных структурных элементов (отрезков, колец, дуг и других).

Используется при распознавании документов с низким качеством печати (машинописный текст, факс и так далее).

Любой символ можно описать через эти элементы и значения параметров их взаимного расположения.

Например, буквы «Н» и «И», состоят из трех отрезков, два из которых расположены параллельно друг другу, а третий соединяет эти отрезки. Различаются же эти буквы величиной углов, которые образуются третьим отрезком с двумя другими.

Современные системы оптического распознавания (FineReader, CuneiForm) используют оба метода и являются«самообучающимися» (то есть для каждого конкретного документа они создают соответствующий набор символов, поэтому скорость и качество распознавания постепенно возрастают).

Для распознавания бланков (форма), заполненных рукопечатным текстом (данные вводятся в поля печатными буквами от руки), используются системы оптического распознавания форм. Эта задача сложнее, так как печатные символы, написанные от руки разными людьми, сильно отличаются, к тому же необходимо определить, к какому полю относится распознаваемый текст.

В последнее время создаются системы распознавания рукописного текста, но они очень несовершенны.