

Создание программного обеспечения для решения задач компьютерной лингвистики

Кобилев С., Раббимов И.

Самаркандский Государственный Университет

Одной из интересных задач компьютерной лингвистики является задача проектирования программных систем, используемых для изучения иностранных языков. До настоящего времени разработано большое количество таких систем, например, TomWord, Anki, Vocabilis, Learnwords и другие[1]. Анализ структур, выполняемых функций и принципов применения этих и других систем показывает некоторые их недостатки и неудобства, такие как: системы созданы в основном для изучения одного иностранного языка, чаще всего, английского. Их адаптация для других языков требует сложных программных изменений; большинство из них не имеет подсистемы, применяемые простыми пользователями для расширения и модификации базы данных словаря, изменения лексики общения, подготовки и вывода статистики по изучению языка; база систем содержит данные только об одной предметной области. Нет простых и удобных средств расширения информационного фонда данными различных предметных областей.

Всё это усиливает актуальность задачи разработки и программной реализации новых методов проектирования систем компьютерной лингвистики. Поэтому в данной работе раскрывается структура и описывается процесс создания простого программного комплекса с адаптируемым интерфейсом[2], применяемого для изучения иностранной лексики.

Структура базы данных. База содержит лексику, т.е. множество слов (словарь) изучаемых языков. Словарь можно организовать в виде двуязычного или многоязычного. Например, для хранения данных о трёх языках структура базы данных, описанная конструкциями паскалеобразного языка программирования может иметь вид:

```
type stroka = string [30];
```

slovo=recordnomersl: **integer**;sleng, slrus, sluzb: stroka**end**;

varbazaslov:file of slovo;

Комплекс программ.Для проектирования программного обеспечения системы была использована технология модульного программирования. Программный комплекс состоит из двух основных программ, каждая из которых применяет несколько вспомогательных модулей. Модули разработаны в виде пользовательских библиотек.

Первая программа – LexiHelp используется для создания и корректировки базы данных системы. Вспомогательные модули программы выполняют конкретные функции:

Inpword– создание словаря; Addword – добавление новых слов; Corword – корректировка словаря; Strword- сортировка словаря.

Вторая программа – Lexicon предназначена непосредственно пользователю, изучающему иностранную лексику. Основная идея, заложенная в алгоритме данной программы, заключается в следующем.

- а) Выводится сообщение об изучаемых языках.
- б) Определяется изучаемый язык.
- в) Случайно выбирается слово родного языка.
- г) Перевод этого слова отгадывается пользователем.
- д) При отгадывании выводятся слово родного языка, первая и последняя буквы слова иностранного языка.
- е) Допускается не более трёх, ошибочных попыток отгадывания.
- ж) В конце сеанса изучения даётся итоговая таблица (статистика) работы пользователя.

Литература

1. www.danilidi.ru/progi/Anki
2. КобиловС. Локализация программного обеспечения для образовательной информатики // Вестник КазНУ. Серия математика, механика, информатика. – 2010. -№4 (67). – С.74-77.