

Федеральное государственное автономное
образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Институт космических и информационных технологий
Кафедра «Вычислительной техники»

УТВЕРЖДАЮ
Заведующий кафедрой

_____ А.И. Легалов
подпись инициалы, фамилия
« ____ » _____ 20 ____ г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

**ГЕНЕРАЦИЯ УЧЕБНЫХ МАТЕРИАЛОВ ПО МЕТОДУ ПАЛМЕРА С
РАСПОЗНАВАНИЕМ ОМОНИМОВ МЕТОДОМ КОМПОНЕНТНОГО
АНАЛИЗА**

09.04.01 Информатика и вычислительная техника
09.04.01.04 Технология разработки программного обеспечения

Научный руководитель	_____	<u>К.Т.Н.</u>	<u>Д.В. Личаргин</u>
	подпись, дата		
Выпускник	_____		<u>В.В. Сотникова</u>
	подпись, дата		
Рецензент	_____	<u>д.ф.-м.н., доцент</u>	<u>К.В.Сафонов</u>
	подпись, дата		
Нормоконтролер	_____	<u>К.Т.Н., доцент</u>	<u>В.И.Иванов</u>
	подпись, дата		

Красноярск 2016

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	4
1 Анализ и сравнение разработок в области автоматической генерации осмысленного текста.....	7
1.1 Обзор предметной области	7
1.2 Основные определения.....	8
1.3 Сравнение функционала аналогичных разработок	9
1.3.1 Программа SeoGenerator	9
1.3.2 Программа GeneratingtheWeb	11
1.3.3 Программа Delirium 1.8.....	13
1.3.4 Программа ArticleCloneEasy.....	13
1.3.5 Программа СинонимайзерM1pluss	14
1.3.6 Программа SyMonum	16
1.3.7 Программа MonkeyWrite.....	17
1.3.8 Вывод по рассмотренным аналогам	18
2 Разработка программы генерации учебных материалов	20
2.1 Методы и средства разработки программы автоматической генерации осмысленного текста для учебных материалов	20
2.1.1 Метод Палмера и необходимость генерации учебных заданий по этому методу.....	21
2.1.2 Метод компонентного анализа и его применение для распознавания омонимических групп	23
2.1.3 Генерация подстановочных таблиц с фонетическими пояснениями ...	25
2.1.4 Принцип использования реляционных баз данных в работе алгоритма генерации подстановочных таблиц.....	30
2.1.5 Частотный и грамматический анализ как этапы работы алгоритма генерации подстановочных таблиц.....	30
2.1.6 Семантический анализ как этап работы алгоритма генерации подстановочных таблиц	32
2.1.7 Общая схема разрабатываемой программы «Генератор подстановочных таблиц»	33
2.2 Интерфейс программы«Генератор подстановочных таблиц»	34
3 Интеграция программы «Генератор подстановочных таблиц» в ряд программных продуктов.....	40
3.1 Описание программы Automatic Mouse and Keyboard	40
3.2 Генерация учебных материалов	41
3.3 Описание основных используемых программ	43
3.4 О системе электронного обучения Сибирского федерального университета	52
3.5 Наполнение электронного курса заданиями на основе сгенерированных материалов для системы дистанционного обучения СФУ	55
ЗАКЛЮЧЕНИЕ	57
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	58
ПРИЛОЖЕНИЕ А	61

ПРИЛОЖЕНИЕ Б.....	64
ПРИЛОЖЕНИЕ В.....	67

ВВЕДЕНИЕ

Проблемной областью является разработка системы автоматической генерации учебных материалов, обеспечения образовательного процесса учебными материалами, а так же повышения эффективности языковой подготовки в техническом вузе в целях индивидуализации траекторий обучения.

Актуальность работы. С развитием вычислительных технологий стали актуальны такие проблемы как построение экспертных систем, синтез и генерация осмысленной речи компьютером на различных языках. Для решения выше представленных проблем применяются различные средства и методы: реляционные таблицы, графы, декларативное программирование, порождающие грамматики и многие другие. По этим методам семантику можно раскладывать, решать конкретную задачу или делать конкретный программный продукт. Например: синонимизатор, «технический справочник», системы диалога с пользователем, экспертные системы, программы электронного перевода.

Также информационные технологии широко используются в сфере образования. Применение программного и технического обеспечения предоставляет большие возможности в традиционном обучении. Электронные образовательные ресурсы – важные компоненты, образующие информационно-образовательную среду. Именно образовательные ресурсы представляют собой базу для обеспечения высокого качества прохождения учащимися учебного процесса [1].

В настоящее время практически в каждой научной дисциплине применяют электронные ресурсы или издания. В традиционном учебном процессе активно используется интеграция электронных образовательных ресурсов на базе сетевых технологий.

В данной работе рассматривается использование генератора подстановочных таблиц, для обучения студентов по дисциплине английский язык, с функцией выбора омонимов и лексико-семантических вариантов на основе вектора смысловых признаков.

Компьютерная лингвистика, информатика, математика и математическая логика – такие аспекты предметных областей затрагивают данное исследование.

Поэтому актуальна задача разработки системы генерации учебных материалов по методу Палмера с распознаванием омонимов методом компонентного анализа. Она поможет решить такие проблемы как:

- 1) Снижение количества рутинной работы преподавателей за счет ее автоматизации, генерация учебных материалов;
- 2) Снижение не продуктивных затрат на монотонные действия пользователя при создании электронных ресурсов;
- 3) Решение проблемы разработки электронных курсов в кратчайшие сроки с учетом регулярного обновления ФОС;

4) Улучшение усвоения материала учащимися.

Новизна работы заключается в алгоритмизации методов компонентного анализа, при выборе омонимических групп слов, для генерации подстановочных таблиц на основе текста на естественном языке.

Цель диссертации: разработать алгоритмическую реализацию системы построения учебных подстановочных таблиц на основе произвольных предложений естественного языка на входе системы.

Задачи диссертации состоят в:

- 1) распознавании омонимов;
- 2) разработке алгоритмов построения подстановочных таблиц;
- 3) разработке алгоритмов оценки предпочтения омонимических групп;
- 4) разработка программной реализации построения учебных подстановочных таблиц на выходе, на основе произвольных предложений естественного языка на входе системы;
- 5) разработка программной системы «Генератор подстановочных таблиц»;
- 6) пополнение базы данных дефинимов английского языка;
- 7) переводе данных программ в единую систему с использованием общего формата данных с программами по компьютерной лингвистике магистратуры ИКИТ СФУ.

Необходимо описать метод компонентного анализа, который может быть использован для выявления соответствий смысловых признаков в омонимических группах, позволяющий ранжировать слова в порядке большего совпадения в каждом столбце для всех вариантов в омонимических группах генерируемой таблицы по методу Палмера.

Основная идея работы состоит в создании генератора подстановочных таблиц с функцией выбора омонимов и лексико-семантических вариантов на основе вектора смысловых признаков, также именуемый как «Генератор подстановочных таблиц» и в использовании полученных подстановочных таблиц для системы сопутствующих программных продуктов, учитывая их совместимость. Например, при использовании материалов полученных из программы «Генератор большого объема текстов по стандартным шаблонам» программный продукт «Text-BasedTaskGenerator» составит разнообразные варианты текстовых заданий на основе входного текста.

Назначение программы – необходимость преодоления недостаточно высокого уровня навыков аудирования у студентов высших учебных заведений, что не дает возможность осуществления беглой коммуникации в англоязычной среде с минимальным количеством фонетических ошибок.

Разработанная программа генератор подстановочных таблиц с функцией выбора омонимов и лексико-семантических вариантов на основе вектора смысловых признаков поможет лицам изучающим английский язык быстрое

усваивать материал. Решит проблему затруднения изучения, понимания языка и преодоление трудности перевода. По ним учащийся может составить свое первое предложение с возможными ошибками, выбирая из каждой колонки предложенные варианты. Также пополнить свой словарный запас и изучить различные взаимозаменяемые слова в конкретном контексте.

Данные факторы обеспечивают возможность выбора учащимся индивидуальной траектории обучения в зависимости от поставленных задач, что дает возможность достижения более качественного восприятия материала учащимися.

Подстановочные таблицы можно сохранить в текстовом документе и использовать при необходимости для повторения и закрепления пройденного материала.

Объектом исследования является системы генерации учебных материалов для обучения иностранным языкам.

Предметом исследования является средства генерации подстановочных таблиц с учетом преодоления сопутствующий проблем компьютерной лингвистике.

1 Анализ и сравнение разработок в области автоматической генерации осмысленного текста

1.1 Обзор предметной области

Компьютерная лингвистика возникла в результате развития компьютерных технологий. Разбирая язык на компоненты, семантику, грамматику можно увидеть конкретные языковые конструкции, по которым строится язык. С помощью компьютерных методов, используя математические модели, языковые конструкции, можно описать и смоделировать естественные языки. Во главе компьютерной лингвистики стоит понятие языка, формальный язык рассматривается с разных сторон [2].

В зависимости от конкретной задачи, в компьютерной лингвистике используются свои методы анализа языка:

- лексико-грамматический анализ;
- синтаксический анализ;
- морфологический анализ;
- семантический анализ;
- частотный т.д.

Электронное обучение.

Метод, для которого характерно активное взаимодействие между всеми участниками учебного процесса. Значение этого метода и интенсивность его использования существенно возрастает с развитием обучающих телекоммуникационных технологий. Именно этот метод, который ориентирован на групповую работу представляет наибольший интерес для дистанционного обучения. Именно этот метод предусматривает широкое использование исследовательских и проблемных способов обучения. В едином процессе такого метода объединяются три идеи: обучение в коллективе, взаимооценка, обучение в малых группах. Роль преподавателя при таком обучении сводится к тому, что он задает тему, а далее он должен создать и поддерживать такую благоприятную среду общения и психологический климат, при которых обучаемые могли бы работать в сотрудничестве. Обучение в сотрудничестве может иметь различные формы.

Широкое использование информационных технологий в сфере образования, обусловлено современным уровнем развития методики обучения и компьютерной техники. Применение программного и технического обеспечения предоставляет большие возможности в традиционном обучении. Электронные ресурсы являются важным компонентом в сфере образования, которые в совокупности с компьютерными технологиями образуют информационно-образовательную среду с едиными программными средствами, обеспечивающими высококачественный содержательный контент для предоставления информации лицам использующий данный ресурс (студентов, школьников, педагогов и т.д.) [2-5].

Преимущества использования электронных обучающих программ и ресурсов состоят в том, что:

- Учащиеся имеют доступ к процессу обучения в подходящее им время и в удобном месте;
- Имеет место нужное соотношение самостоятельной работы и контакта с преподавателем;
- Независимости от места проживания и отсутствия временных и финансовых расходов на транспорт;
- Доступ к актуальным материалам в электронном виде круглосуточно;
- Доступность получения высшего образования для лиц с ограниченными возможностями;
- Возможности обеспечения индивидуального графика обучения [27].

Формы и методы дистанционного обучения.

Методы и средства, используемые в учебном процессе, являются важным признаком при систематизации видов дистанционного обучения, но особую, ключевую роль в дистанционном обучении играет способ взаимодействия преподавателей с учащимися. Ниже приведены некоторые из возможных вариантов:

1. Участие преподавателя сводится к минимуму. Учащийся, посредством использования ПК, электронных курсов, баз данных, электронных справочников и других источников, доступных в web-режиме, изучает необходимый материал.
2. Индивидуальное дистанционное обучение происходит посредством связи с преподавателем по электронной почте, сотового телефона, Skype и др. Удаленное взаимодействие с преподавателем-экспертом для получения дистанционной консультации, оценки.
3. Преподаватель в online-режиме, либо по записи на видео, аудио файлы излагает материал (электронные лекции) учащимся, которые участвуют в пассивном режиме, только как слушатели.
4. Метод, при котором каждый обучающийся взаимодействует с другими участниками курса. Групповая деятельность используется при исследовательском направлении курса, помогает работать в коллективе, развить групповое мышление. Преподаватель задает тему для исследования и следит за качеством ее исполнения.

1.2 Основные определения

Подстановочная таблица (ПТ) представляет собой массив двумерных данных, обеспечивающая автоматическую генерацию фраз естественного языка, с учетом специфического синтаксиса подстановочных таблиц.

Существуют сложные подстановочные таблицы, состоящие из иерархически расположенных простых ПТ с возможностью выбора одной ПТ из списка.

При этом создастся более длинная и сложная конструкция, чем при использовании простой ПТ.

Генератор осмысленного текста представляет собой программу, которая используется для генерации текста, правильного с позиции большинства языковых норм, учитывающая признаки языка (синтаксические, морфологические, семантические, фонетические).

Комбинонимы – слова, замещающие друг друга при изменении смысла, но сохранении осмысленности в некотором контексте одного слова или группы слов. Например, слова «банкнот – чек – кредитная карточка – мелочь и другие» могут замещать друг друга во фразах типа «я расплатился банкнотами», «я сэкономил мелочь». Комбинонимами часто являются эквонимы [6, 7].

Эквонимы – Слова, находящиеся на одном уровне по степени общности и имеющие общий гипероним, например, слова «блюдо – чашка – стакан» относительно гиперонима «посуда» или слова «книга – газета – журнал» относительно слова «литература» (но не относительно слова «периодика», поскольку книга не является «периодикой»). В общем виде любые два слова являются эквонимами относительно слова «нечто» [6, 7].

Омонимы – разные по значению, но одинаковые по звучанию и написанию слова, морфемы и другие единицы языка [3].

В разных языках существуют свои омонимы, что приводит к затруднению изучения, понимания языка и трудности перевода.

1.3 Сравнение функционала аналогичных разработок

При использовании подстановочных таблиц возможно генерирование предложений естественного языка, выбирая из каждой колонки этой таблицы по одному слову в результате получая осмысленные предложения.

Поиск генераторов осмысленного текста в сети интернет не дает положительного результата. Были рассмотрены как бесплатные, так и платные аналоги генераторов осмысленного текста, но данные реализации не отвечают заявленным требованиям:

- прозрачности;
- доступности;
- простоты;
- универсальности.

1.3.1 Программа SeoGenerator

SeoGenerator (SEO AnchorGenerator) под win32, представляет собой инструмент генерации текстов, анкоров, названий, с сохранением результата в формате TXT [8].

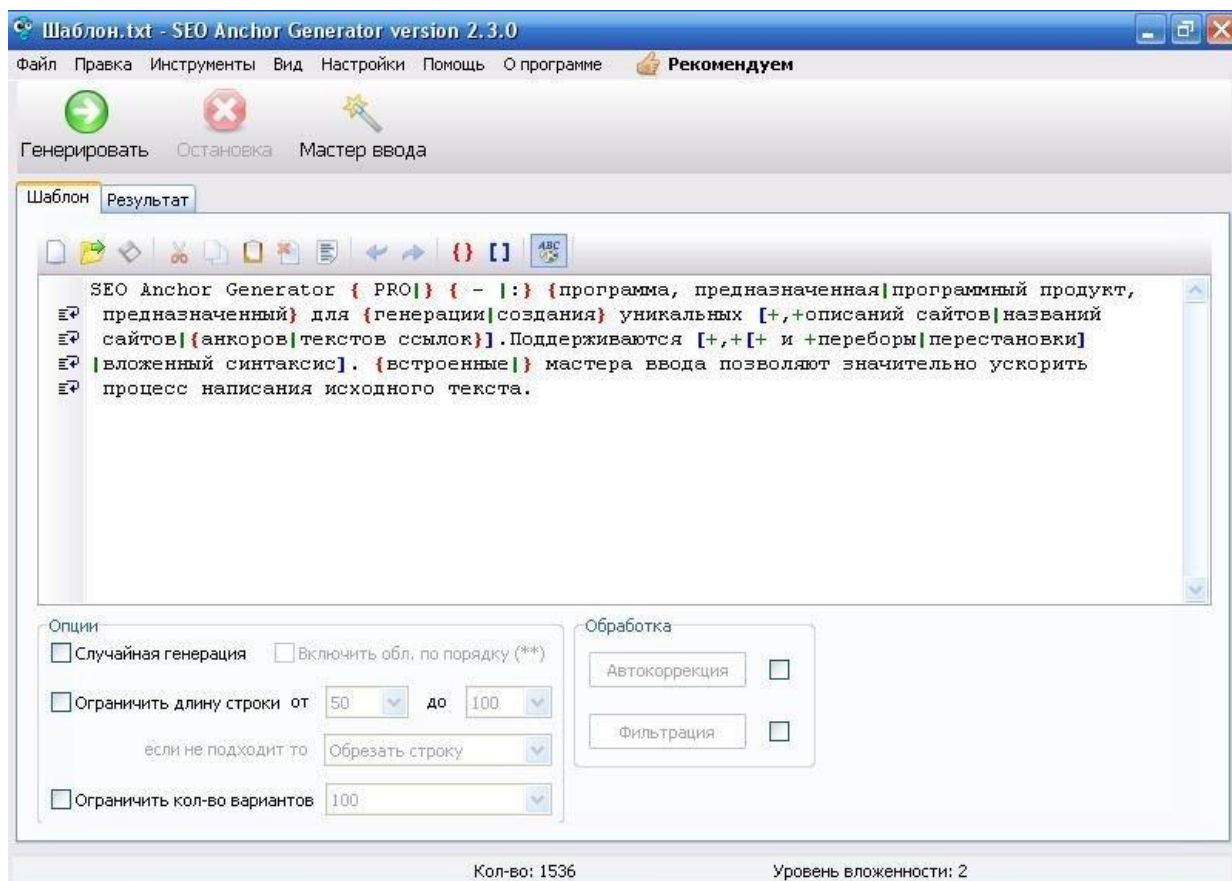


Рисунок 1 – Программа «SeoGenerator»

Есть два вида программы:

- SEO AnchorGenerator – для работы необходим компьютер с Windows. Позволяет быстро и удобно генерировать множество текста по заранее определенному шаблону, может работать с большими текстами. Также доступны дополнительные опции обработки текста, такие как поиск и удаление похожих текстов (рисунок 1);
- On-line SeoGenerator – это веб версия программы SEO Anchor Generator. Ею можно пользоваться с любого компьютера и любой платформы. Возможно удалять точные дубликаты и производить случайное перемешивание результата с выборкой определенного числа значений (рисунок 2) [8].

Данные программы поддерживают следующие виды шаблонов:

- {a|b|c} – выбирает одно из значений, например, «b»;
- [a|b|c] – выбирает комбинацию значений, например, «bac»;
- [+ и +a|b|c] – выбирает комбинацию значений, используя разделитель, например, «b и a и c»;
- Вложенные шаблоны – например, можно задать шаблон {a|b| [+ .+1|2] {x|y|z} 3}, глубина вложенности неограниченна;

- Область генерации по порядку (***) – если выделить с помощью скобок «(» и «)» текст содержащий один или несколько шаблонов, то при случайной генерации в этой области будут только уникальные сочетания[8].

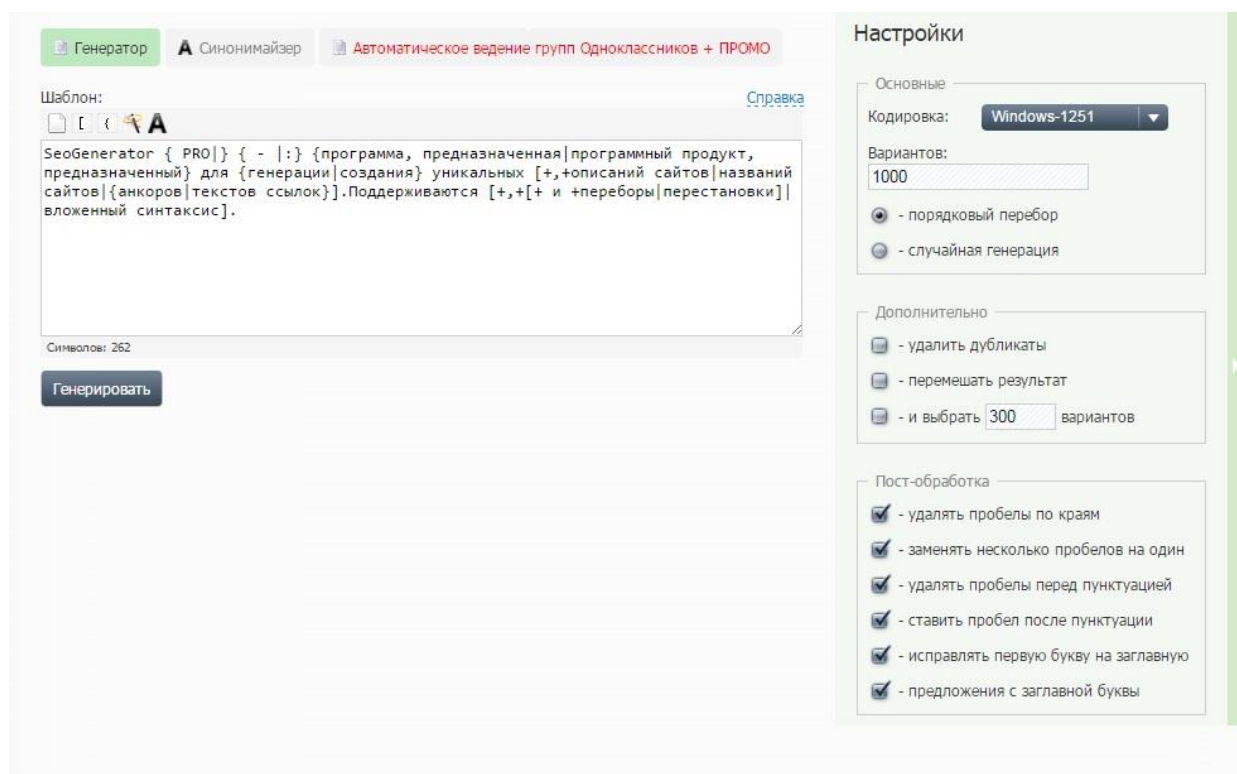


Рисунок 2 – Программа «On-line SeoGenerator»

Среди достоинств данной программы можно отметить, что On-lineSeoGenerator можно пользоваться с любого компьютера и любой платформы. Можно удалять точные дубликаты и производить случайное перемешивание результата с выборкой определенного числа значений.

Недостатки – ручной ввод шаблонов, использование только синонимов, ограниченная база синонимов.

Вывод – не отвечает заявленным требованиям.

1.3.2 Программа GeneratingtheWeb

GeneratingtheWeb – программа для генерации методом перебор, перестановок. Основные функции программы заключаются в генерации текстов, простановке ссылок с нужным анкором и синонимайзе [9].

Для генерации текста в GeneratingTheWeb используются следующие конструкции:

- 1) Генерация перебор – { 1 | 2 | 3 | 4 }
- 2) Генерация перестановок:
 - a) Простые перестановки – [1 | 2 | 3 | 4]
 - b) Перестановки с одинаковым разделителем – [<, > 1 | 2 | 3 | 4]

с) Перестановки с разными разделителями – [\langle, \rangle 1 | 2 | 3 \langle и \rangle 4]

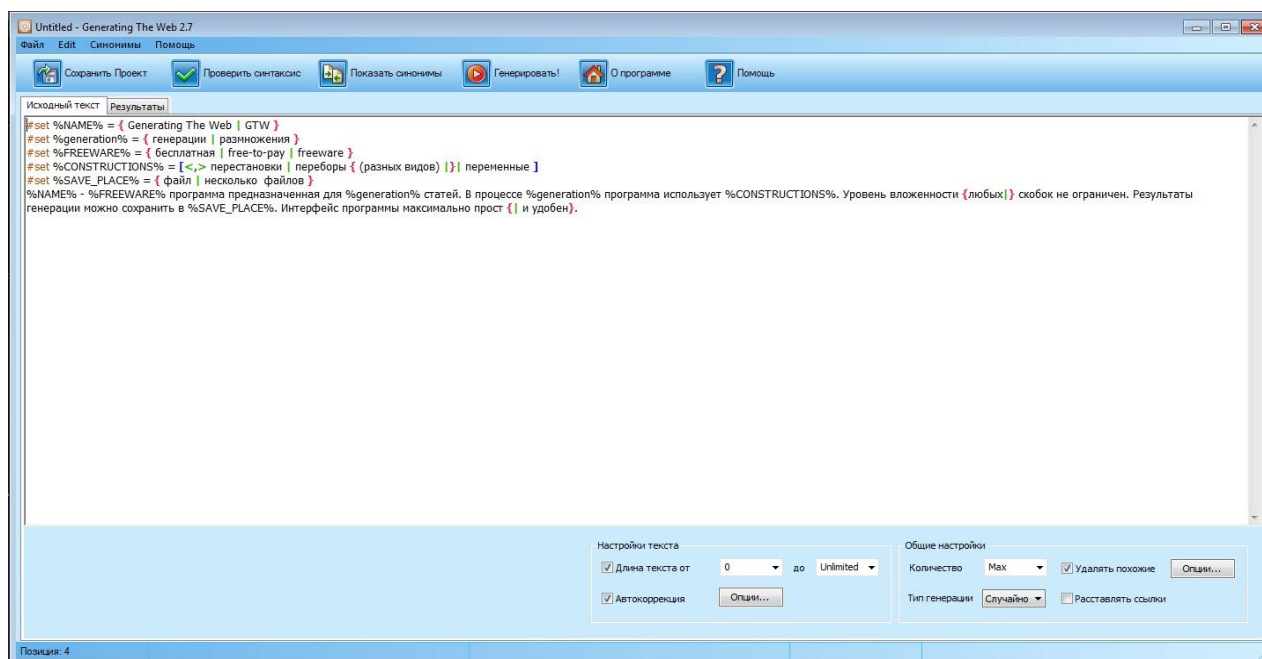


Рисунок 3 – Программа GeneratingtheWeb

В настройках можно выставить необходимую длину текста, отметить нужные позиции автокоррекции текста, максимальное количество вариантов текста на выходе, удаление похожих текстов, выставить тип генерации (по порядку или случайно), расставить ссылки, отфильтровать полученные варианты сгенерированных текстов по уникальности. Это можно сделать либо отфильтровав результаты с помощью опции "Удаление точных совпадений", либо с помощью метода шинглов. В случае использования метода шинглов, указываем длину шингла и максимально допустимый процент схожести текста. По умолчанию в программе стоит длина шингла 7 и процент схожести 20.

Чем меньший процент схожести - тем меньше вариантов текста на выходе, но тем более уникальными по отношению друг к другу они будут. От применения метода шинглов также зависит скорость генерации - чем меньше длина шингла, тем медленнее будет происходить процесс генерации [9].

Достоинства – возможность выставить необходимую длину текста, обозначить нужные позиции автокоррекции, удаление похожих текстов, расставление ссылок.

Недостатки – составление предложений методом переборов, перестановок, ручной ввод шаблонов, использует синонимы – получаются не достаточно уникальные тексты, ограниченная база синонимов.

Вывод – не отвечает заявленным требованиям.

1.3.3 Программа Delirium 1.8

Delirium 1.8 – эта программа генерирует тексты на основе существующих. Она использует алгоритм Марковских цепей, позволяющий создавать довольно длинные "осмысленные" фразы. Программа способна на генерацию текстов, без участия человека [10].

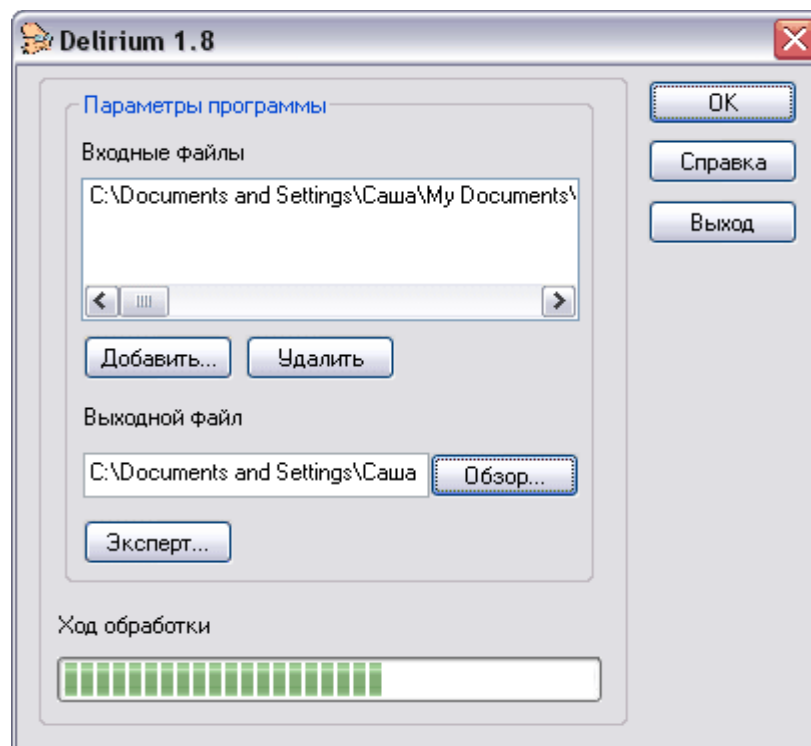


Рисунок 4 – Программа Delirium 1.8

Достоинства – генерация текстов, без участия человека.

Недостатки – осмысленный текст получается только из больших объемов текста, равномерного по содержанию.

Вывод – не отвечает заявленным требованиям.

1.3.4 Программа ArticleCloneEasy

ArticleCloneEasy – бесплатная программа для размножения статей при помощи генерации текстов по принципу формул (перебора и замены). В ее состав входят: сама программа, словари русских и английских синонимов, которые взяты из открытых источников, инструкция по применению на английском и русском языках, демонстрационный файл [11].

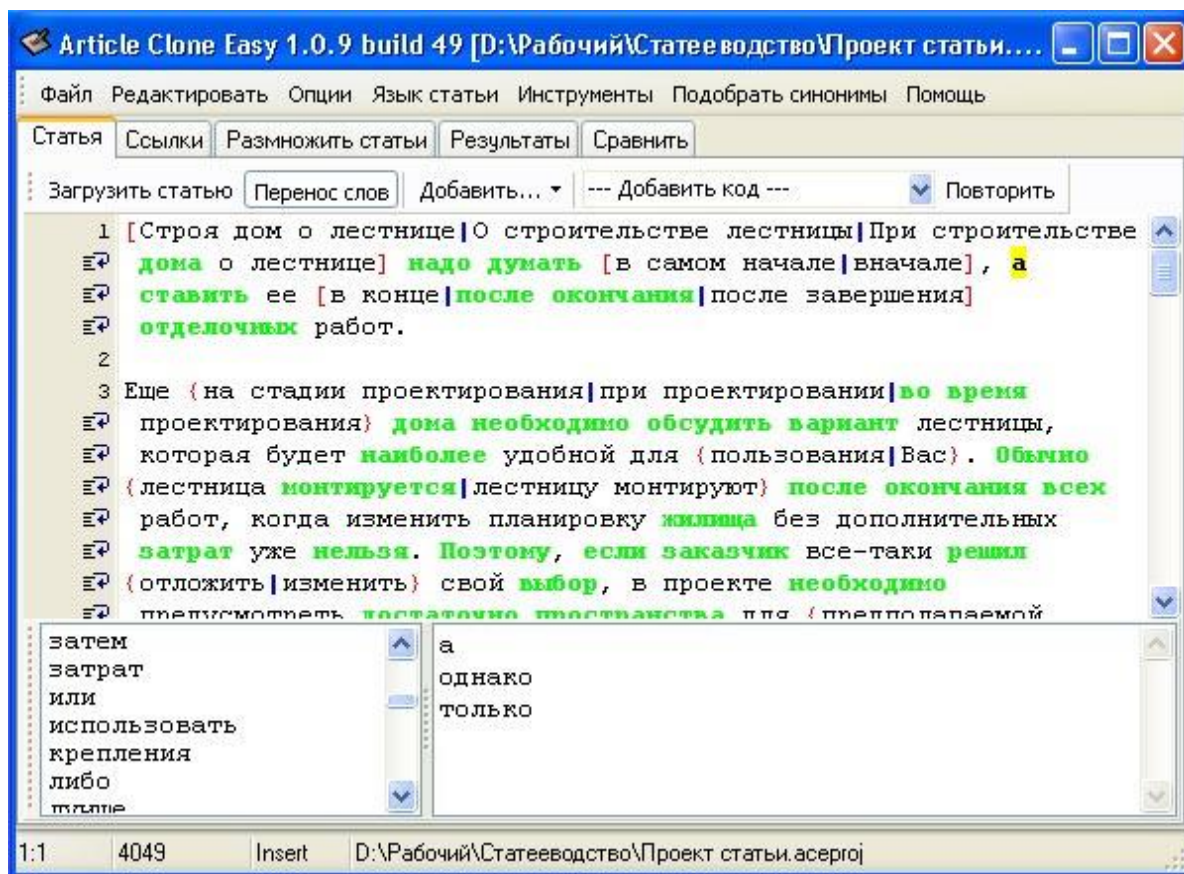


Рисунок 5 – Программа ArticleCloneEasy

Доступны такие функции:

- подбор синонимов;
- несколько вариантов перебора слов и выражений;
- неограниченная вложенность выражений;
- расстановка ссылок;
- штучное и групповое сравнение;
- многочисленные опции формирования и сохранения статей [11].

Достоинства – расстановка ссылок, многочисленные опции формирования и сохранения статей в различных форматах.

Недостатки – составление предложений по принципу формул (перебора и замены).

Вывод – не отвечает заявленным требованиям.

1.3.5 Программа СинонимайзерM1pluss

Синонимайзер M1pluss – большая база русских синонимов – 638 000 слов и словосочетаний, украинские и другие синонимы на латинице. Отбор избранных синонимов в отдельную базу, есть выбор случайных синонимов, подготовка текста для размножения. Продукт имеет постоянную поддержку. Есть демоверсия, но ограничение синонимов до 3000. Стоимость – 25\$/700 руб [12].

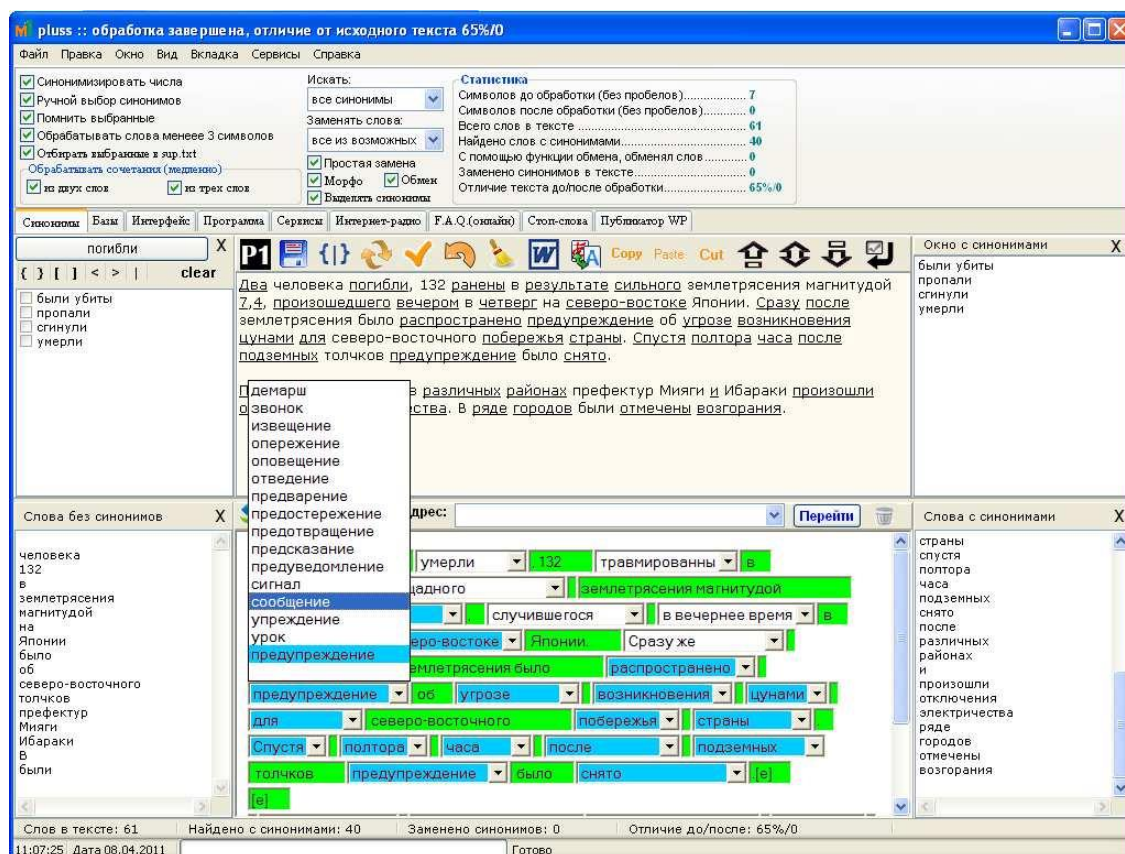


Рисунок 6 – Программа Синонимайзер M1plus

Доступны такие функции:

- Google переводчик в простом и пакетном режиме;
- пакетная обработка файлов типа txt, html, htm, php, sql;
- программа поддерживает кириллические и латинские синонимы;
- идет выбор случайных синонимов, если для слова их несколько;
- в комплекте идет 5 русских баз, ru1 и ru2 для ручного рерайта, ru3, ru4 и ru5 для автоматической обработки, и одна с английскими синонимами;
- стоп-слова и стоп-теги;
- функция обмен;
- функция конвертирования других баз;
- анализатор текста;
- оптимизатор текста;
- три режима обработки, морфология, простая замена и смешанная;
- генератор фраз и анкоров.

Требования к системе:

- ОС: Windows (2000, XP, Vista, Windows 7);
- Браузер: InternetExplorer;
- Проигрыватель WindowsMedia 9 или выше (для проигрывания радио) [12].

Достоинства – большая база русских синонимов, украинские и другие синонимы на латинице, пакетная обработка файлов типа txt, html, htm, php, sql, три режима обработки: морфология, простая замена и смешанная.

Недостатки – использование только синонимов, коммерческое программное обеспечение.

Вывод – не отвечает заявленным требованиям.

1.3.6 Программа SyMonum

SyMonum – программа для уникализации текстов, содержит закрытую базу синонимов из 68272 словоформы, которая позволяет производить синонимизацию текста и статей с высоким процентом читаемости [13].

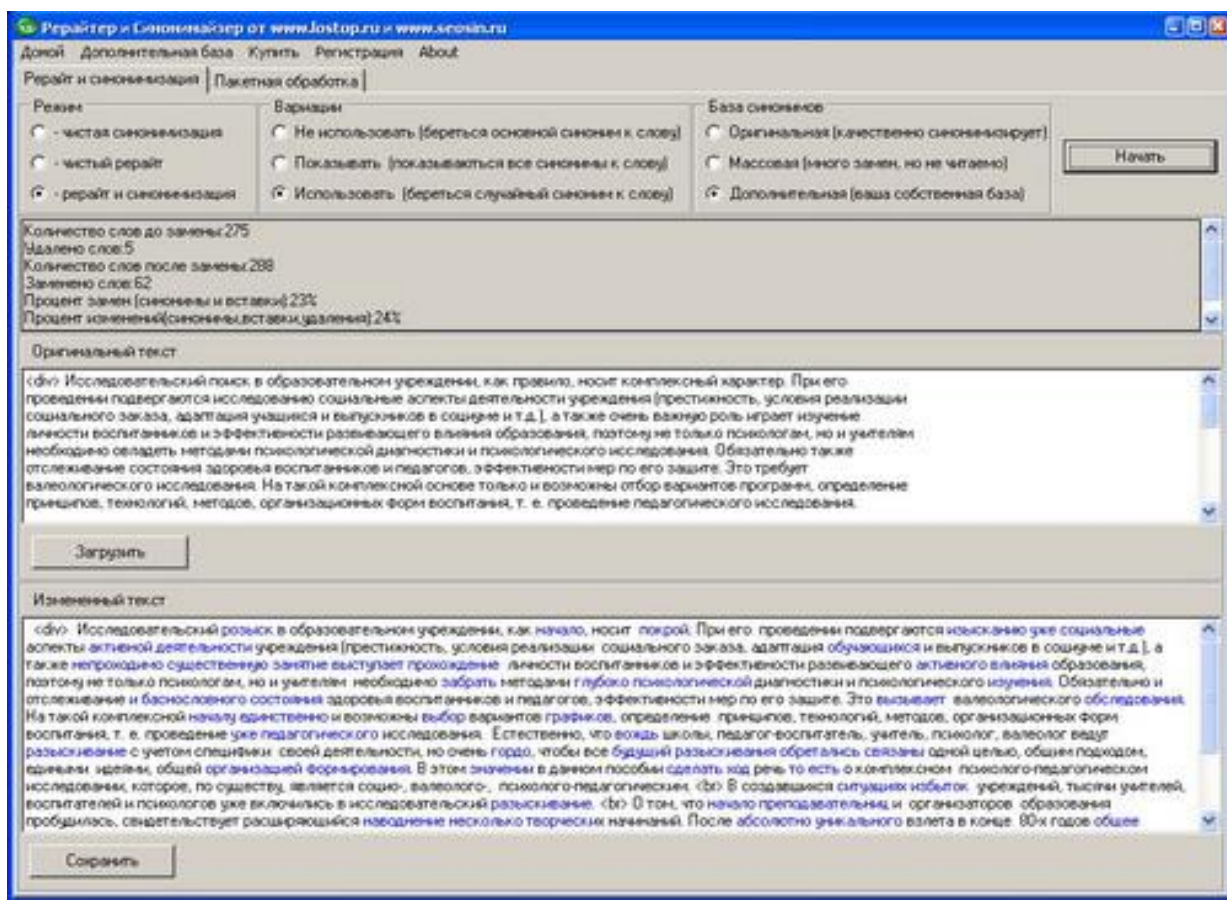


Рисунок 7 – Программа SyMonum

Так же имеется дополнительный словарь синонимов более 1 000 000 просклоненных слов и словоформ, высокая степень уникализации, но не высокий процент читаемости. Стоимость – 15\$. Есть демонстрационная версия [13].

Достоинства – большая база синонимов, имеется возможность использовать свои базы данных синонимов, программа может по пред настройкам изменять целые папки с текстами в формате ТХТ.

Недостатки – коммерческое программное обеспечение, использование только синонимов.

Вывод – не отвечает заявленным требованиям.

1.3.7 Программа MonkeyWrite

MonkeyWrite – это генератор текстов с мощным функционалом, где каждое предложение текста представляется отдельной строкой. Генератор позволяет размножать тексты, не только заменяя синонимами слова и выражения, но и перестраивая и перефразируя предложения. Имеется возможность автоматически в результирующих статьях менять местами абзацы или их группы, автоматизировано расставлять ссылки по заранее созданному анкор-листу [14].

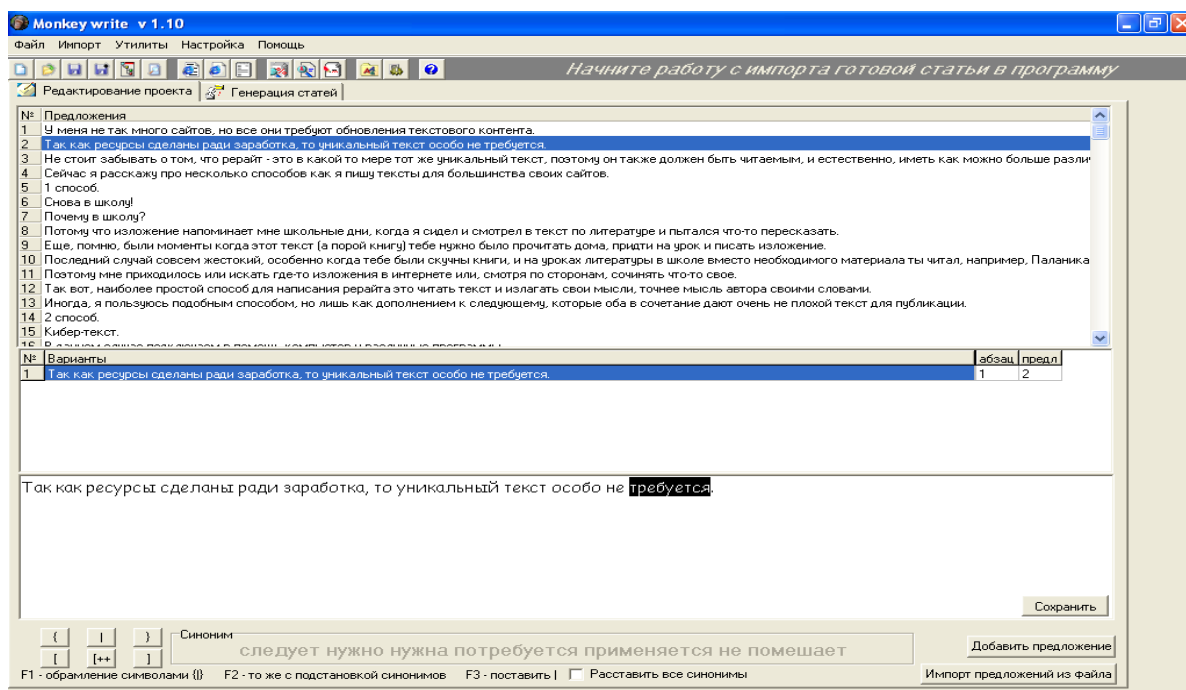


Рисунок 8 – Программа MonkeyWrite

Доступны такие функции:

- базу синонимов возможно автоматически дополнять;
- проверка грамматики средствами MicrosoftWord (с условием наличия MicrosoftWord на Вашем компьютере);
- сравнение статей на схожесть методом шинглов. Этот модуль может сравнивать статьи пакетно. Имеется возможность отсеивать самые неуникальные варианты;
- пакетный анализатор ссылок. Покажет, сколько ссылок, с какими анкерами и адресами содержится в пакете статей, есть импорт полученных данных в Word в виде таблицы;

- описание статей (при генерации можно создать отдельный файл с описаниями, программа запишет туда несколько первых предложений каждой статьи);
- система макросов. Чтобы сделать код более легким и читаемым, громоздкие выражения можно перенести в макросы.

Требования к системе: Windows XP [14].

Достоинства – не только заменяет синонимами слова и выражения, но и перестраивает и перефразирует предложения, расставление ссылок, проверка грамматики средствами MicrosoftWord.

Недостатки – только под Windows XP, использование синонимов.

Вывод – не отвечает заявленным требованиям.

1.3.8 Вывод по рассмотренным аналогам

Ни одна из упомянутых программ не позволяет быстро и эффективно создать учебные материалы для работы с текстом в качестве предмета задания по иностранному языку для школьников и учащихся первых курсов в институтах.

Существует множество средств, дающих возможность выполнять некоторые приближения по генерации осмысленного текста, например, в форме всплывающих вариантов выбора синонимов.

Среди преимуществ разработанной программы «Генератор подстановочных таблиц» можно выделить:

- возможность составления учебного материала с привлечением подстановочных таблиц, создание на их основе тестовых заданий: выбор правильного варианта, множественный выбор, для проверки знаний учащихся, что должно повысить качество обучения;
- простоту работы по составлению материала;
- снижение количества рутинной работы преподавателей за счет ее автоматизации;
- использование комбинимов;
- распознавание омонимов;
- интерфейс позволяющий осуществлять выбор вариантов омонимов;
- функция выбора омонимов на основе предложенных;
- генерация подстановочных таблиц с функцией выбора омонимов и лексико-семантических вариантов на основе вектора смысловых признаков.
- генерация подстановочных таблиц с функцией выбора омонимических групп слов на основе вектора смысловых признаков.

Существующие системы генерации осмысленного текста не имеют следующих функций:

- интерфейса позволяющего осуществлять выбор вариантов омонимов;
- составление более уникальных и разнообразных вариантов осмысленного текста путем использования комбинонимов, а не путем замены слов на синонимы;
- функции редактирования результата путем ручного ввода омонимов либо их выбор на основе предложенных.

2 Разработка программы генерации учебных материалов

Модель данных реляционных таблиц данных со строками слов и сопоставленными ими векторами семантических признаков дает возможность реализовать функционал разрабатываемой программной системы на основе алгоритмов, в основе которых лежат представления о системе естественного языка.

Данная модель основывается на таком разделе естественного языка как грамматика и семантика.

Большинство авторов анализирует естественный язык с точки зрения синтаксиса, что не дает обычному пользователю возможность составлять осмысленные фразы языка в необходимом для процесса обучения количестве.

Модель данных реляционных таблиц данных со строками слов и сопоставленными ими векторами семантических признаков является отражением естественного языка, что обеспечивает возможность выделения множества осмысленных фраз из множества бессмысленных на основании критериев осмысленности. В частности в работах такого известного семантиста как О.Н. Селиверстова описываются понятия взаимозависимости семантически связанных групп слов.

Так, например, слово компьютер можно имплицировать как следующее слово – мощный, современный, дорогостоящий. При наличии нескольких таких групп слов происходит наложение ограничений на множество осмысленных, частотных и естественно звучащих слов генерируемых на основе таких соответствий.

Алгоритм учитывает частотные принципы организации структуры фраз естественного языка. В словаре комбинированных слов с наиболее большей частотностью расположены в начале словаря и их выбор программной системой предпочтителен.

В результате мы решаем проблему выделения не только осмысленных фраз из множества любых фраз с учетом бессмысленных, но и выделение из множества осмысленных фраз наиболее частотных.

Генерируемые подстановочные таблицы основываются на фразах отличающиеся частотность и естественностью с точки зрения пользователя, который осуществляет ввод входных предложений. В результате фразы генерируемы или составляемые вручную на основе полученных учебных материалов, на основе таблиц Палмера также во многом сохраняют признаки частотности и естественности, что сводит к минимуму процесс постредактирования подстановочных таблиц экспертом, осуществляющим проверку их качества.

2.1 Методы и средства разработки программы автоматической генерации осмысленного текста для учебных материалов

Большой трудность при создании подобной системы является многофакторность выбора той или иной (синонимической) группы омонимов и

лексико-семантических вариантов, выбор которой должен осуществляться заявленной системой автоматически. В результате было принято решение сделать вложенную систему алгоритмических реализаций. Первоначально множество отобранных групп слов анализируется на основе частотности. Далее заявленные варианты отсеиваются и проходят отбраковку грамматическим алгоритмом, и рассматривается ключевой момент в работе программной системы – это отбраковка определенных комбинаций омонимических групп с учетом векторов семантических признаков включаемых в них слов. В результате последующие эксперименты показали, что для фраз достаточного простого вида, распределение ошибок выбора омонимических групп данной системы снижаются кардинально.

2.1.1 Метод Палмера и необходимость генерации учебных заданий по этому методу

Метод Палмера появился в начале 20 века, сразу обрел всеобщее признание и повлиял на формирование таких методов как: аудиолингвальный и аудиовизуальный [15, 16].

В основу концепции изучения иностранного языка Палмер ставит подражание и заучивание (слов, предложений), но изучение необходимо начинать с устной речи (восприятие речи на слух, произношение звуков, использование материала в речи на повседневные темы с учетом определенного словаря).

Отбор словаря происходит не по словам, а по эргонам (лексические единицы, служебные слова, словосочетания) отвечающие определенным принципам:

- 1) Распространенность, должны встречаться общеупотребительные слова;
- 2) сочетаемость с другими словами;
- 3) конкретность (слова, имеющие конкретное значение);
- 4) учет частей речи;
- 5) целесообразность (применение слова в ПТ).

Также важным аспектом является выбор текстов.

Требования к содержанию:

- 1) Содержание должно соответствовать теме урока и быть информативным.
- 2) Темы текстов быть знакомы, сюжетные.

Требования к языку:

- 1) текст состоит из слов отобранного словаря;
- 2) учитывать значение слов;
- 3) используется незнакомые слова (значение слов понятно по контексту).

Необходим рациональный подход к выбору языкового материала, его систематизация и классификация. Организация в различных формах

(практические занятия, лекции, самостоятельная работа) всего процесса обучения на основе отобранного материала.

В методе Палмера грамматические конструкции, для лучшего усвоения, представлены в виде подстановочных таблиц (все элементы которой взаимосвязаны, выбрав по слову из каждой колонки можно составить предложение). Также используются разнообразные упражнения с повышением сложности, вопросы, подстановочные таблицы.

Самое обширное распространение для обучения языка приобрела идея подстановочных таблиц.

Код	Поле1	Поле2
25	American_English	американский_английский
26	British_English	британский_английский
27	Russian	русский
28	Ukrainian	украинский
29	German	немецкий
30	French	французский
31	Italian	итальянский
32	Spanish	испанский
33	Greek	греческий
34	Latin	латинский
35	- Combinonims -	- Combinonims -
36	cool	крутой
37	great	великолепный
38	super	супер
39	superb	супер-пупер
40	cool	крутой
41	great	великолепный
42	cool	крутой
43	great	великолепный
44	- Combinonims -	- Combinonims -
45	handsome	красивый
46	cute	крутой
47	cool	крутой
48	unattractive	непривлекательный
49	handsome	красивый

Рисунок 9 – База данных комбинонимов английского языка

На основе метода Палмера был составлен словарь комбинонимов, использующийся для выбора омонимической подгруппы в подстановочной таблице.

По позиции пары слов в словаре, можно судить являются ли они:

- 1) гиперонимами – овощ и огурец, говорить и лгать;
- 2) синонимами – очевидно, естественно, конечно;

- 3) антонимами – жить и умереть, жить и родиться, умереть и родиться;
- 4) различными лексико-семантическими вариантами – дерево (предмет/вещество из предмета/растения) и дерево (предмет/растение);
- 5) омонимами – коса (предмет/часть тела/волосы) и коса (место/суша/у воды);
- 6) синтагматическими парами – есть и торт, одевать и шляпа;
- 7) ассоциациями – ехать и телега, биология и животное, биология и зоопарк;
- 8) словами со смещенным ядром значения – учить и ученик, повар и готовить, готовить и блюдо;
- 9) многими другими элементарными и составными отношениями между словами[16].

Например, если в предложении написать слово «родственник», то в подстановочной таблице пользователю будет предложены такие слова как: «родитель», «предок», «теща», «свекровь», «невестка», «золовка», «свояченица», «мать», «мама», «бабушка», «тетя», «тесть», «свекор», «зять», «шурин», «отец», «дедушка», «дядя», «ребенок», «малыш», «кузина», «дочь», «сноха», «сын», «брат», «сестра» и другие.

Разработанная программа должна позволить определить не только грамматическую корректность, но и семантическую корректность фраз, и логическую связанность текстов.

2.1.2 Метод компонентного анализа и его применение для распознавания омонимических групп

В семантической классификации, каждая точка, это – отдельный омоним или ЛСВ, а в действительности словам соответствуют кластеры точек, таким образом, нам надо из кластера выбрать только одно понятие – точку смыслового пространства на основе функциональных отношений к другим точкам – словам из того же шаблона как подмножества многомерного пространства слов. Эту «точку» можно выбрать на основе таблиц, что может применяться как иллюстрация при изучении языка [17].

Примеры омонимов:

- ручка (письменная принадлежность) – ручка (дверная);
- конь (животное) – конь (шахматная фигура);
- дробь (число) – дробь (свинцовые шарики для стрельбы);
- резинка (для волос) – резинка (для удаления надписей с бумаги) – резинка (жевательная);
- бор (5 элемент таблицы Менделеева) – бор (хвойный лес);
- кисть (часть тела) – кисть (малярный инструмент);
- ключ (инструмент) – ключ (для шифрования) – ключ (от замка) – ключ (музыкальный знак).

Для выбора омонимической подгруппы подстановочной таблицы был взят компонентный анализа, входящий в структурную семасиологию, которая появилась вследствие использования метода структурного изучения языковых значений слов.

Для компонентного анализа характерно формирование упорядоченных последовательностей слов с единым значением, которые относятся к одной части речи – лексико-семантические группы. Данная группа слов отражает смысловое значение относительно друг друга, вследствие чего выявляются их общие и индивидуальные признаки [18-20, 21-23].

Этот метод подходит для установления соответствий или различий общих смысловых признаков, в частности он поможет выявить необходимость использования конкретного слова в омонимических группах.

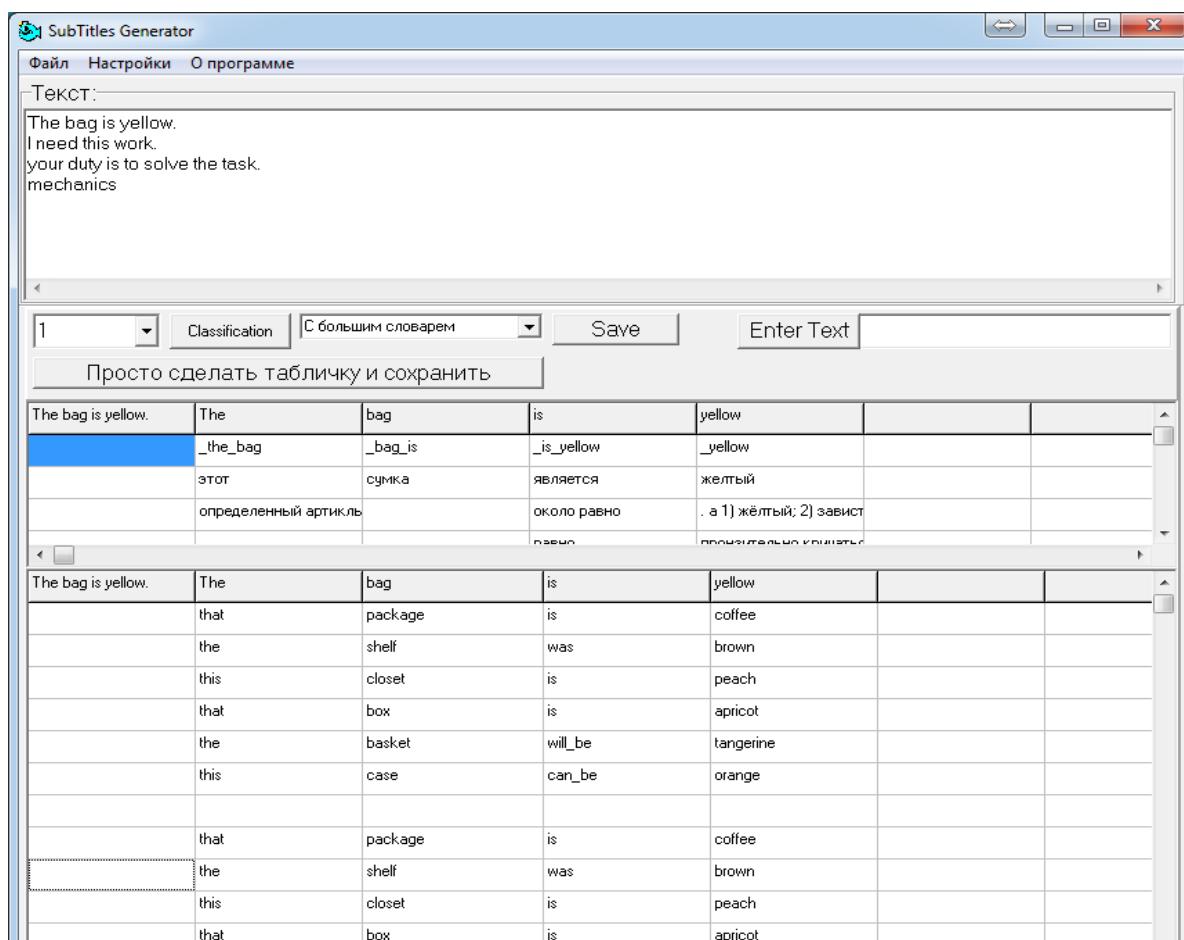


Рисунок 10 – Подстановочная таблица с вариантами омонимов и комбинимов, созданная с помощью программы «Генератор подстановочных таблиц»

Чем больше общих сем в омонимических подгруппах со словами в предложении, тем больше вероятность использования данной подгруппы в предложении. По такому же принципу осуществляется ранжирование слов в каждой омонимической подгруппе, во всплывающем списке слова

выстраиваются от большего совпадения к меньшему. По такому принципу работает алгоритм выбора омонимической группы слов.

Разработанная формула, описывающая данный метод:

$$\max \sum (p_{i'}^{k'} * s_{j_n}^{v'} * p_{i''}^{k''} * s_{j''}^{v''}) * k_3, \quad (1)$$

где p_i – положение слова в предложении;

s_j – сема внутри структуры слова;

i' – номер первого слова в предложении;

i'' – номер второго слова в предложении;

k' – номер омонимической группы слов определенной позиции для первого слова;

k'' – номер омонимической группы слов определенной позиции для второго слова;

j' – номер семы в словарном определении (векторе признаков) для первого слова;

j'' – номер семы в словарном определении (векторе признаков) для второго слова;

v' – вариант семы в позиции словарного определения для первого слова;

v'' – вариант семы в позиции словарного определения для второго слова;

Для того чтобы определить какое именно слово необходимо вставить в предложение нужно использовать компонентный анализ для каждого члена предложения, учитывая его значимость.

Формула для определения значимости:

$$k_3 = (i' - i'') / (k'_{\max} * k''_{\max}), \text{ где}$$

k_3 – коэффициент значимости

k'_{\max} – общее количество омонимических групп для i' -ой позиции для первого слова;

k''_{\max} – общее количество омонимических групп для i'' -ой позиции для второго слова.

2.1.3 Генерация подстановочных таблиц с фонетическими пояснениями

Векторные порождающие грамматики.

Классическим способом определения языков через порождающую процедуру является определение их с помощью порождающих грамматик, или грамматик Хомского.

«Порождающие грамматики были введены Н. Хомским в конце 50-х годов как средство описания формальных языков (точнее, как средство описания формальных моделей естественных языков)».

Порождающая грамматика нужна для вывода (порождения) цепочки языка из некоторой первоначальной цепочки с помощью обусловленных правил подстановки. Порождение представляет собой пошаговый процесс

получения новой цепочки из составленной на предыдущем шаге, с использованием правил замены.

Векторная транскрипция.

Порождающая грамматика задается четверкой: $G\langle S, T, N, R\rangle$, где S – начальный символ порождающей грамматики G , T – множество терминальных символов (совпадает с алфавитом языка, задаваемого грамматикой), N – множество нетерминальных символов (нетерминальные символы используются в промежуточных порождениях цепочек.), R – множество правил трансформации одной строки в другую.

Рассмотрим использование векторного представления для порождающих грамматик.

Векторное представление слова W : $\text{words} \rightarrow R_n$ — параметризованная функция, отображающая слова из некоторого естественного языка в векторы большой размерности.

Количество эмоциональных синонимов в языке огромно, и они захватывают очень многие пласты лексики литературного языка. Вариации эмоциональных значений часто сочетаются с небольшим видоизменением логического значения слов.

На основе таких параметров как напряженность, резкость, временной компонент (время до и после произнесения звука) и слитность (с плавными переходами между звуками) и др. составляется вектор признаков классификации. Признаки одного ряда могут иметь тождественные элементы с предыдущими, что обозначается знаком «<», что делает текст векторного фонетического описания более читабельным.

Из данных критериев составляется вектор признаков классификации. Признаки одного ряда имеют тождественные элементы, определяющие связи и логику классификации, выраженную в виде семантической формулы определяемых понятий.

Значения каждого элемента вектора являются понятиями другой классификации. Единицы слова разного уровня – буквы с различными признаками. Каждая классификация генерирует понятия для создания ряда признаков последующих классификаций. Каждую из классификаций задает свой вектор семантических признаков.

$T = t_1[\text{напряженность, резкость, временной компонент, слитность}], t_2[\text{напряженность, резкость, временной компонент, слитность}], \dots, t_n[\text{напряженность, резкость, временной компонент, слитность}]$, где T – слово языка, $t_1 \dots t_n$ – буква с рядом критериев.

Признаки классификации. Вектор признаков классификации задается перечислением множества значений фонетических элементов и их оттенков для каждого уровня классификации, то есть для каждого элемента вектора отдельно.

Приведем некоторые примеры использования векторных порождающих грамматик. Грамматики Хомского традиционно имевшие вид $AB..C \rightarrow DE..F$, могут быть записаны в виде: $A_i B_i .. C_i \rightarrow D_i E_i .. F_i$. Значения элементов векторов

могут быть как константы, так и переменные. Например: [напряженность, х...у, z...w] → [резкость, х...у, q...r] + [временной компонент, х...у, z...w].

По такому принципу генерируются строки с линейно представленной матричной структурой и учитываются фонетические особенности речевого строя английского языка.

При использовании векторных порождающих грамматик учитываются фонетические аспекты языка, которые позволяют эффективно и максимально быстро изучить язык, а также понять его эмоциональную сторону.

Рассмотрим различные признаки звуковых элементов при их произношении. Так, например, при произношении фразы: «It's time to have a bite», возможна различная раскладка векторов описания фонетических оттенков звуковых элементов языка [24].

Таблица 1 – Описание оттенков звуков при произношении в различных ситуациях

Звук	Длина	Форма при произношении	Тенденции при произношении	До	В начале	В середине	В конце	После
i	Как у взрывного звука	Готовясь к t, более высоко	Не вибрирует, грубовато, суховато	Резкое, слышное начало	Четко, громко	Менее громко и четко, гладко	Едва слышно, язык с разворотом вверх	Пауза с долей взрывного звука
t	Как у взрывного звука	Взрыв тонкой и узкой расслабленной перемычки	Очень маленькая амплитуда открытия щели	-	Щелчок	Сильный выдох в узкую щель	-	-
s	Как протяженность предыдущего звука	Тонкая и широкая щель	Вибрация при звучании	-	-	Узкая тонкая щель	Окончание выдоха, уход с места произношения	Пауза в 2-3 звука
t	Больше, чем у взрывного звука	Резкий взрыв напряженной перемычки	Полное открытие щели с переходом в открытую позицию	Напряженная пауза	Резкий приступ	Сильный выдох в расширяющуюся щель, придыхание	Усиливающееся придыхание	-
ai	Дольше чем у среднего дифтонга	На одном скольжении	С эмоцией изящества и игры	Язык еще загнут вверх, гласный уже слышен	Язык скользит вверх, челюсть вниз	Язык скользит вверх, челюсть останавливается	Язык вертикально «поглаживает» позицию i	Челюсть еще более смыкается
m	Как у гласной с затуханием	Легко сомкнутые губы	Монотонная, но яркая назальная вибрация	Губы смыкаются, мягкое небо начинает открываться	Губы сомкнуты, мягкое небо открыто		Пауза	

Как видно из таблицы приведенной выше, каждому звуковому элементу соответствуют определенные характеристики, как общие, так и соответствующие особым этапам звучания. В таблице приводятся лишь

некоторые элементы вектора признаков для элементов звучания фраз на естественном языке.

Характеристики каждого звука определяются соотношением всей фразы к различным классам и иерархии подклассов речевого высказывания по таким признакам классификации как: пол, комплекция, тип говорящего, его настроение, выражаемая система отношений и эмоциональной оценки [24].

Рекурсивные порождающие грамматики.

Рекурсивные порождающие грамматики могут рассматриваться как класс порождающих грамматик, который обеспечивает более высокое качество генерации строк и внутреннюю упорядоченность правил. Благодаря этому качество обучения языка для обучающегося возрастает, фонетический строй, особенности языка изучаются более глубоко.

Методы реализации данной функции состоят в гибридизации порождающих грамматик и подстрок команд в форме особых нетерминальных символов языка порождающих грамматик. В результате исследования сформулирован принцип организации множества деревьев, состояний порождаемой строки высказывания на естественном языке. Разработанный подход позволяет обеспечить выполнение алгоритмов генерации строк естественного языка с учетом использования командных нетерминальных символов порождающих грамматик. Такой инструментарий компьютерной лингвистики может найти свое применение в качестве средства моделирования и анализа естественных языков, в частности, в целях генерации осмысленной речи и осуществления языковых трансформаций. Модель системы рекурсивных порождающих грамматик позволит обучающимся произносить предложения на незнакомом или малознакомом языке с правильной интонационной окраской [25].

На основе порождающих грамматик были разработаны правила для генерации транскрипции иностранных языков, а также функции, реализующие генерацию верной и неверной транскрипций, что решает проблему составления вариантов ответов к тестам электронных курсов.

Создание таких правил решало актуальную проблему, появившуюся в связи с широким распространением подходов по проверке знаний учеников в электронной форме. В перспективе на основе использования рекурсивных порождающих грамматик с привлечением семантических моделей языка можно более эффективно решать задачи порождения грамматически и семантически осмысленной речи. С другой стороны на основе расширения традиционных порождающих грамматик можно повышать эффективность различных аспектов для более узких классов задач от генерации фонетической транскрипции до работы трансляторов языков программирования высокого уровня [9, 21].

Предшествующие типы порождающих грамматик не всегда справляются с задачей учета более широкого контекста и нуждаются в увеличении качества генерируемых строк, а также в устранении недостаточного упорядочения внутренней структуры множеств правил. Необходимо учесть традиционное отсутствие их деления на классы или группы. Предлагается создание данного

типа порождающих грамматик «рекурсивные порождающие грамматики», который позволит решать данный класс задач. Общеизвестно, что рекурсия заключается в принципе определения, описания или вычисления функции или набора правил, содержащего аналогичные структуры внутри самой этой функции, то есть когда объект вызывает сам себя, точнее аналогичный объект на другом этапе вычисления [27-33].

При этом сами правила порождающих грамматик, имеющих вид «подстрока1» «подстрока2» «подстрокаn» «подстрока1'» «подстрока2'» «подстрокаn'», будут иметь следующие добавления в синтаксисе внутреннего описания правил, а именно после любой строки может находиться команда вида: «строкаi» «строкаi+1» {название_базы_правил, глубина_генерации, количество_строк}.

Пример использования предлагаемого вида правил рекурсивной порождающей грамматики:

«The» «nice» «girl» «like»+ «-(e)s»{endings,4,2} «to» «dance» «обстоятельство места».

Результатом применения вложенного набора правил порождающей грамматики на втором уровне рекурсии будет строка вида:

«The» «nice» «girl» «likes» «to» «dance» «обстоятельство места».

Далее при применении правила базы правил «endings» получаем строку:

«The» «nice» «girl» «starts» «dancing» «in» «the» «атрибут здания/организации» «club» и далее получаем строку вида: «The» «nice» «girl» «starts» «dancing» «in» «the» «youth» «club» или «The» «nice» «girl» «starts» «dancing» «in» «the» «sport» «club», ит.п.

Различные типы функций, реализующие различные принципы обработки строк, реализуются на основе обобщающих, что упрощает работу по трансформации строк.

В рамках данного исследования система генерации леса состояний генерируемой строки описывается моделью последовательного вызова правил определенного типа для конкретных уровней генерации с возможностью семантизации этих уровней. Данный подход является частью работы по формализации естественного языка с акцентом на его семантику и решение проблемы генерации осмысленной речи [35-38].

Таким образом, предложена модель системы рекурсивных порождающих грамматик для учебных целей. Предложены принципы синтаксического оформления, позволяющие реализовать базовый функционал данного типа порождающих грамматик, предложен алгоритм применения правил порождающих грамматик для построения интонационной разметки в отдельном слове, словосочетании, выражении, проведено обоснование алгоритмической реализации данного класса порождающих грамматик. Приведены примеры правил для системы генерации строк естественного языка с рекурсией, вызываемой командными нетерминальными символами порождающей грамматики.

Данные теоретические принципы помогут при решении таких проблем как выявление фразовых единств и учета контекста. Например, если программа

видит два слова подряд is to, то она интерпретирует его как фразовое единство, чтобы заменить фразу my duty is to solve the task, она генерирует фразу my purpose is to analyze the principle, для чего необходимо по рекурсии вызвать базу грамматических эквивалентов (со сдвигом по частям речи, например: solution, solved).

2.1.4 Принцип использования реляционных баз данных в работе алгоритма генерации подстановочных таблиц

Для создания подстановочных таблиц используется реляционная база данных. Структура базы данных состоит из множества взаимосвязанных таблиц. Информация об объектах одного вида хранится в отдельной таблице. В строках таблицы содержится информация о каждом объекте, а в столбцах их характеристики. В реляционных базах каждая таблица связана с другой.

Строки называются структурированными записями, состоящими из полей с атрибутами. Каждому полю таблицы соответствует одна характеристика объекта с определенным типом данных (число, текст, дата). У каждой записи есть первичный ключ, отвечающий за их идентификацию. Ключ включает в себя наборы полей из таблицы, скомбинировав которые можно точно определить каждую запись.

Для таблиц в реляционной базе данных характерны следующие свойства:

Отсутствие двух одинаковых строк;

Каждый столбец расположен в том порядке, как он был создан;

Строки могут отсутствовать, на наличие хотя бы одного столбца обязательно;

Каждому столбцу присвоено уникальное имя в рамках данной таблицы, значения каждого столбца принадлежат к одному типу.

Если ячейка, находящаяся на пересечении строки и столбца, имеет атомарное значение, то таблица с такими ячейками считается нормализованной.

2.1.5 Частотный и грамматический анализ как этапы работы алгоритма генерации подстановочных таблиц

Частотный анализ.

Частотный анализ необходим для выбора более вероятного значения слова из множества вариантов:

- 1) have– иметь, получать, быть должным;
- 2) can– мочь, уметь;
- 3) from–от, из, с;
- 4) look–взгляд, смотреть;
- 5) like – подобный, как бы, похожий, любить, нравиться и др.

Пример:

He can ride a horse.

Слово «can» в большинстве случаев является глаголом, поэтому этот вариант рассматривается в первую очередь, а такое значение этого слова как консервная банка учитывается в последнюю.

Также более распространенные слова в употреблении в речи ранжируются программой в начало списка, а остальные по убыванию.

Грамматический анализ.

Распознавание текста с помощью грамматического анализа с применением компьютерных методов и средств является сложной задачей. Необходимо автоматическим способом, без человеческого участия, определить каждый член предложения и их роли относительно друг друга (главные, подчиненные члены предложения), а так же автоматическое распознавание частей речи для каждого слова в предложении «наречие», «личное местоимение», «инфинитив или глагол», «опр. артикль», «существительное», «модальный глагол», «союз», «глагол be», «герундий».

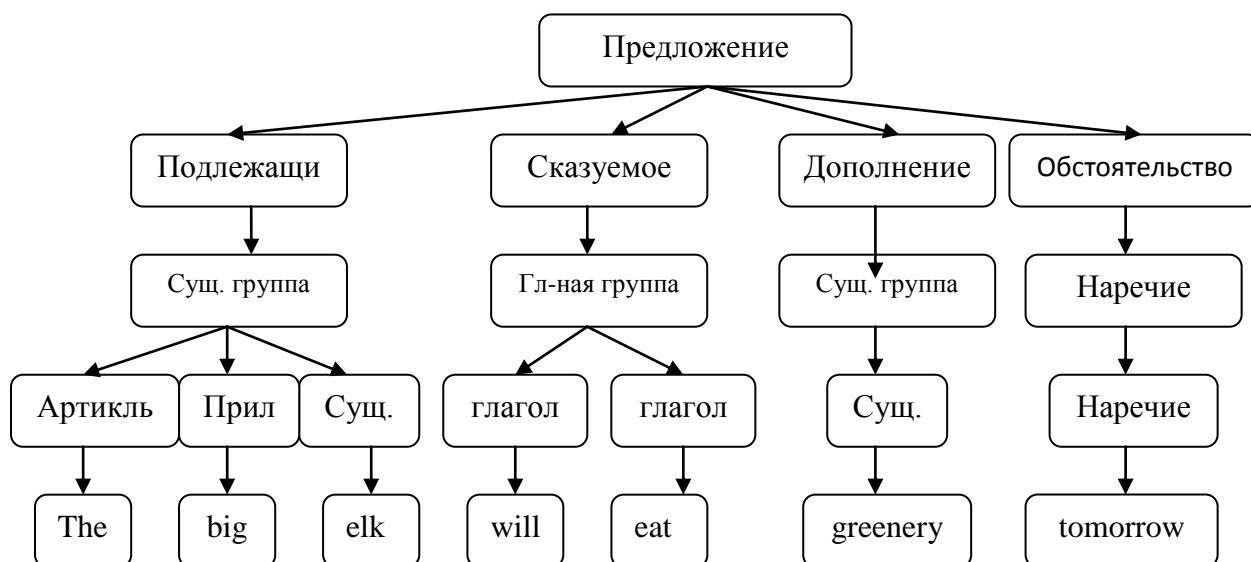


Рисунок 11 – Предложение, представленное в виде дерева

В лингвистике структура предложений естественного языка представлена в виде дерева. Каждое предложение воспринимается в виде цепочки символов. Необходимо правильно проинтерпретировать каждое слово в предложении, связать слова между собой и в результате вывести синтаксическое дерево. Пример предложения представленного в виде дерева строк с указанием частей речи представлен на рисунке 11.

В зависимости от морфологии языка, а так же от многозначности слов распознавание частей речи для каждого языка имеет разную точность распознавания. Например, для английского языка точность составляет примерно 90 процентов.

Для повышения точности используется словарь словоформ. Для каждого слова в словаре есть множество вариантов форм принадлежащих разным частям речи.

Для того чтобы определить к какой части речи относится слово, глагол или существительное, необходимо рассмотреть слово стоящее впереди. Если впереди стоит артикль, то это существительное. Если перед словом стоит существительное, а после глагол, то слово является глаголом.

Перевод слова «can» может быть как консервная банка, так и как глагол может.

Если перед «can» есть артикль, то это консервная банка. The can is big.

Если перед «can» стоит существительное, а после глагол, то «can» – в данном случае является модальным глаголом. You can used.

Для определения союз это или предлог необходимо определить его положение в предложении. Если слово стоит вначале предложения, а после существительное и глагол, тогда это слово является союзом, в ином случае предлог.

Пример:

For I knew her, I didn't be live her. She did it for me.

В результате проблема в том, что для генерации фраз осмысленного языка на основе огромного правил рассматривающего все более специфические случаи словоупотребления, необходимо большее количество человеческих ресурсов, часто такое решение задач упирается в комбинаторику. Рост количества правил необходим для генерации такого подмножества языка.

2.1.6 Семантический анализ как этап работы алгоритма генерации подстановочных таблиц

Семантика – раздел лингвистики, изучающий смысловое значение единиц языка. В качестве инструмента изучения применяют семантический анализ.

Сема – минимальная, предельная, далее не делимая составная часть лексического значения.

В базе сем используются следующие определения каждого слова: «слово», «транскрипция», «перевод слова», «слово на немецком языке», «слово на французском языке», «слово на испанском языке», «сущность» (отношение субъект-субъект, субъект-сущность-субъект и др.), «тип классов слов по абстрактности» (место, информация, идея, сознание, действие и др.), «частота использования» и другие.

Семантический анализ слов – это компонентная оценка количества одинаковых сем слов, определяющих основной его смысл со словом, на которое можно его заменить без потери осмысленности предложения.

Например слово sight – !!!!! 01[sight]01 02[сАйт]02 03[вид]03 04[die Art(Gestalt)]04 05[l'aspect]05 06[la vista]06 07[x]07 08[i]08 09[k]09 10[_]10 11[e0]11 12[02]12.

На основе данной базы, алгоритм семантического анализа сравнивает каждое слово входного предложения, ищет более подходящий вариант для использования слов в предложении и ранжирует в порядке убывания. Также запоминает самый подходящий по смыслу для автоматического сохранения при нажатии кнопки «Просто сделать табличку и сохранить».

2.1.7 Общая схема разрабатываемой программы «Генератор подстановочных таблиц»

На рисунке 12 показан пошаговый процесс создания подстановочной таблицы. Выбранная строка обрабатывается пошагово. На первом этапе разработки в результате были введены подгруппы слов.

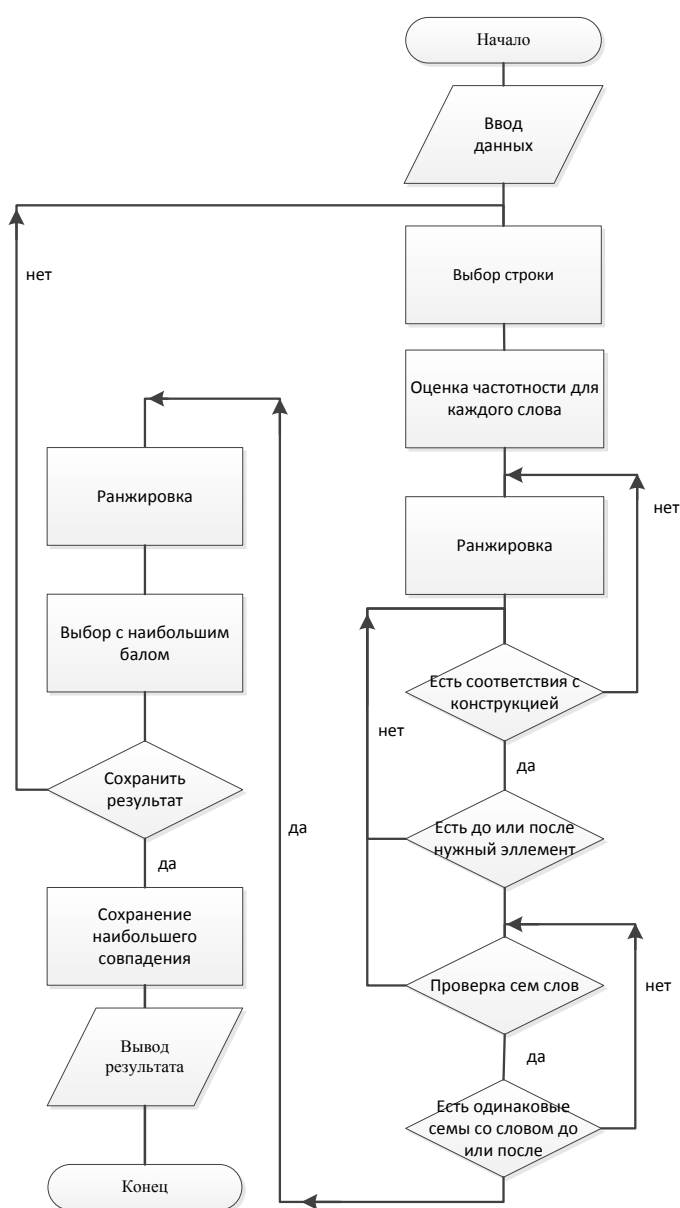


Рисунок 12 – Общая блок-схема программы

Процесс генерации разбит на этапы. Первый блок – частотный анализ. Второй – семантический анализ. Третий – грамматический анализ. Тем самым алгоритм был усложнен.

2.2 Интерфейс программы «Генератор подстановочных таблиц»

Программы «Генератор подстановочных таблиц» представляет собой программный продукт, предназначенный для автоматизации процесса обучения и генерации подстановочных таблиц на основе введенного текста для кафедры РИЯ ИКИТ СФУ.

Программа запускается локально на компьютере пользователя, к которому заявлены минимальные требования. Необходимое количество свободного пространства на диске составляет всего 25 Мб.

Скорость обработки введенной информации зависит от мощности компьютера и размера предложения.

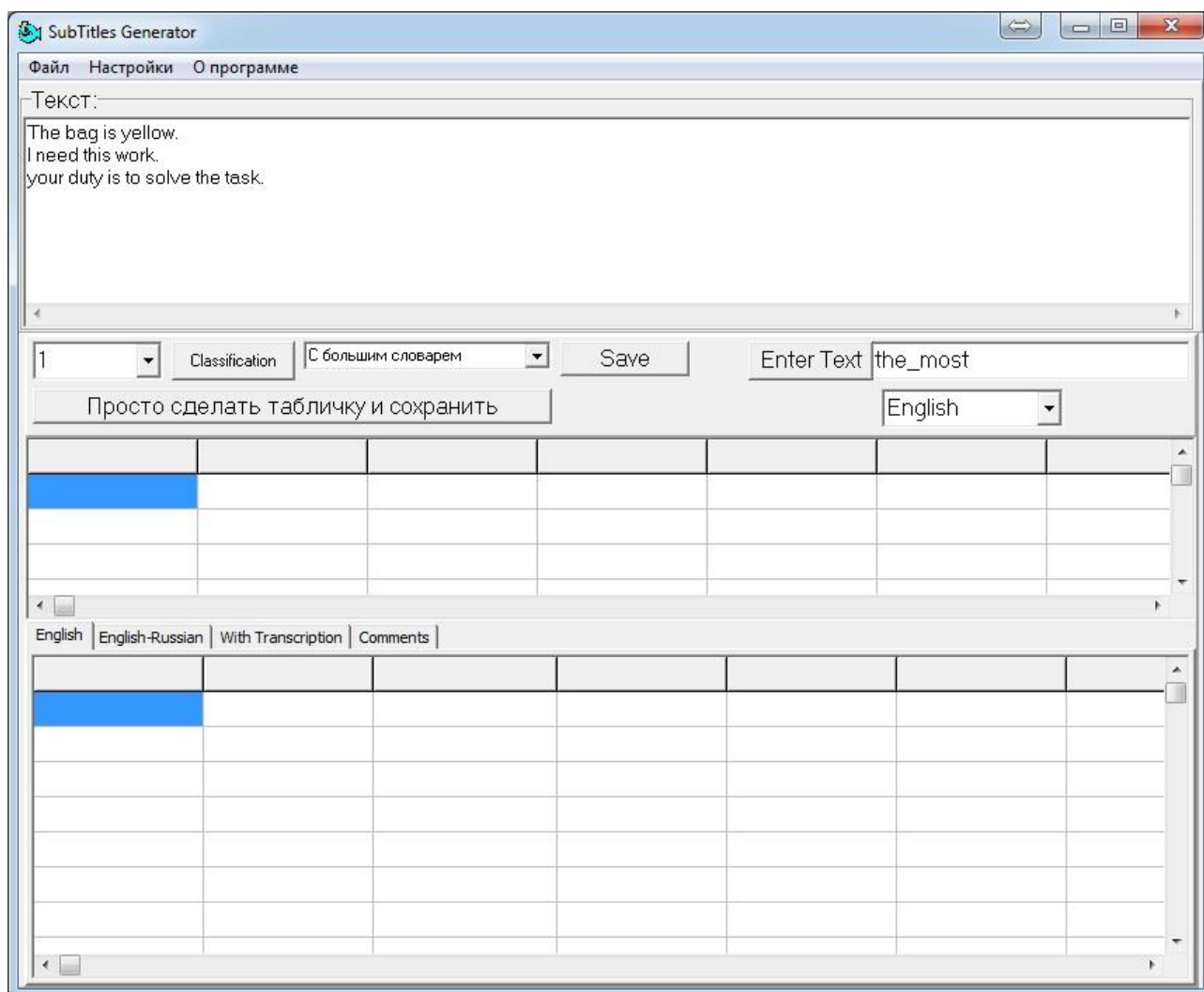


Рисунок 13 – Основное окно программы

Интерфейс программы «Генератор подстановочных таблиц» представленный в виде структурированной таблицы, включает в себя упорядоченную систему правил языка, что позволяет с наибольшей точностью представить структуру языка для конечного пользователя и максимально адаптировать ее для алгоритмов порождения речи.

Ввод и корректировка информации проводится до выбора строк для обработки данных.

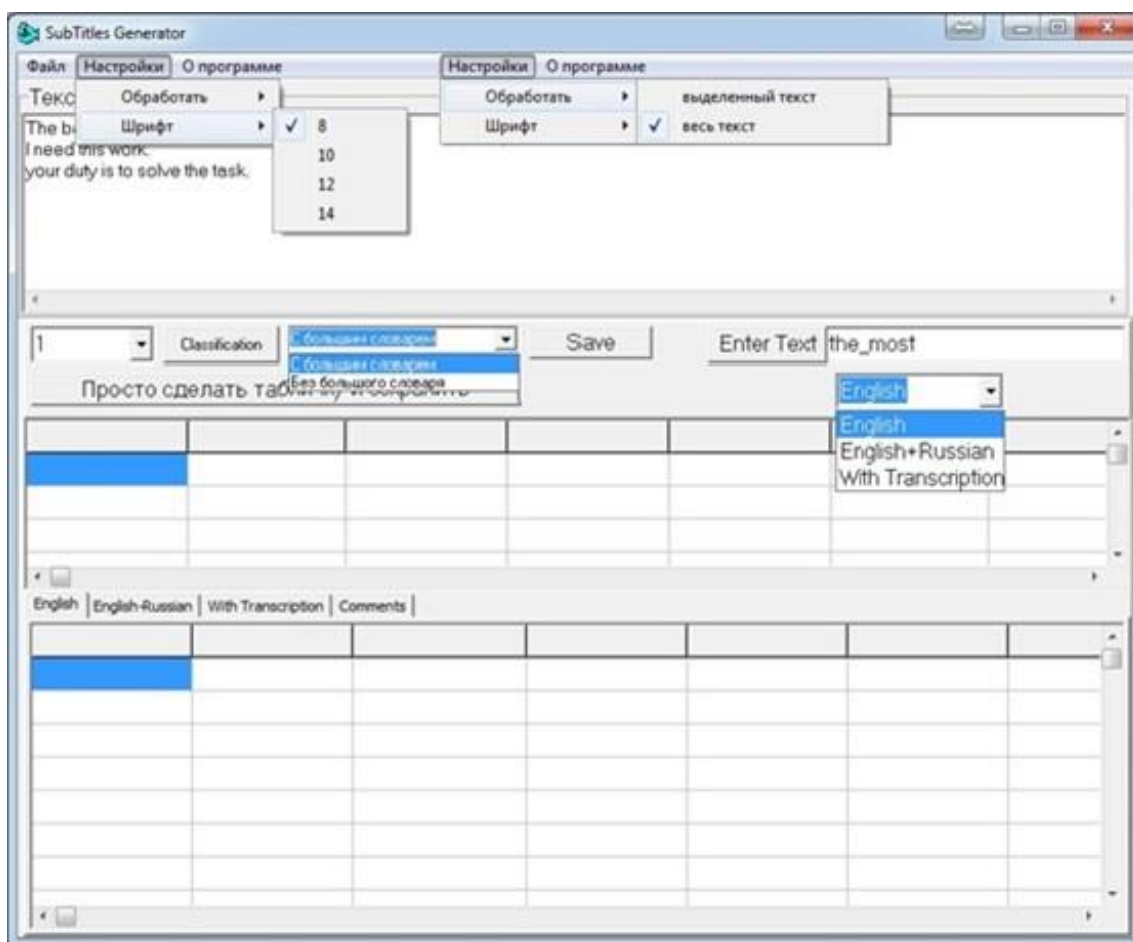


Рисунок 14 – Общие настройки программы

В программном продукте используется основные настройки для поля «Текст»:

Обработать.

Обработка выделенного текста

Обработка всего текста.

Шрифт.

С возможными вариантами 8, 10, 12, 14.

Настройки для возможных вариантов выводимого результата, подстановочны таблиц:

Выбор вариантов используемого словаря.

С большим словарем.

Без большого словаря.

Выбор вариантов вывода подстановочных таблиц.

English – только английские варианты слов

English+Russian – вывод английских вариантов слов + перевод на русском языке.

WithTranscription – вывод английских вариантов слов с транскрипцией.

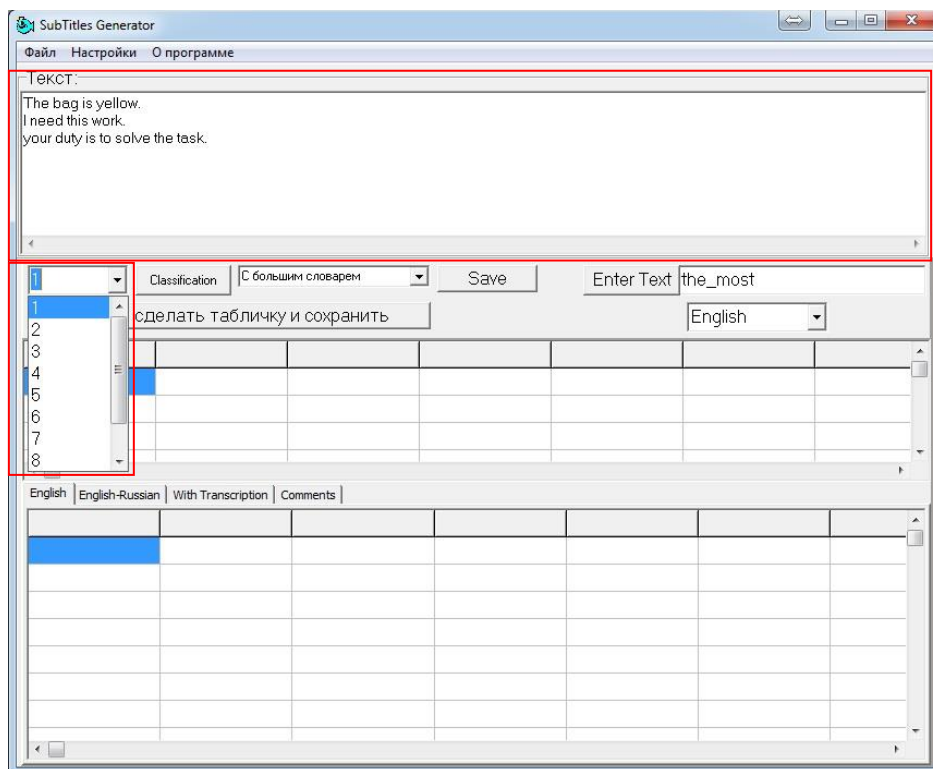


Рисунок 15 – Выбор строки обработки

Для начала работы программы необходимо ввести предложение в поле программы «Текст» либо воспользоваться исходным. Следующий этап – это выбор строки для ее обработки.

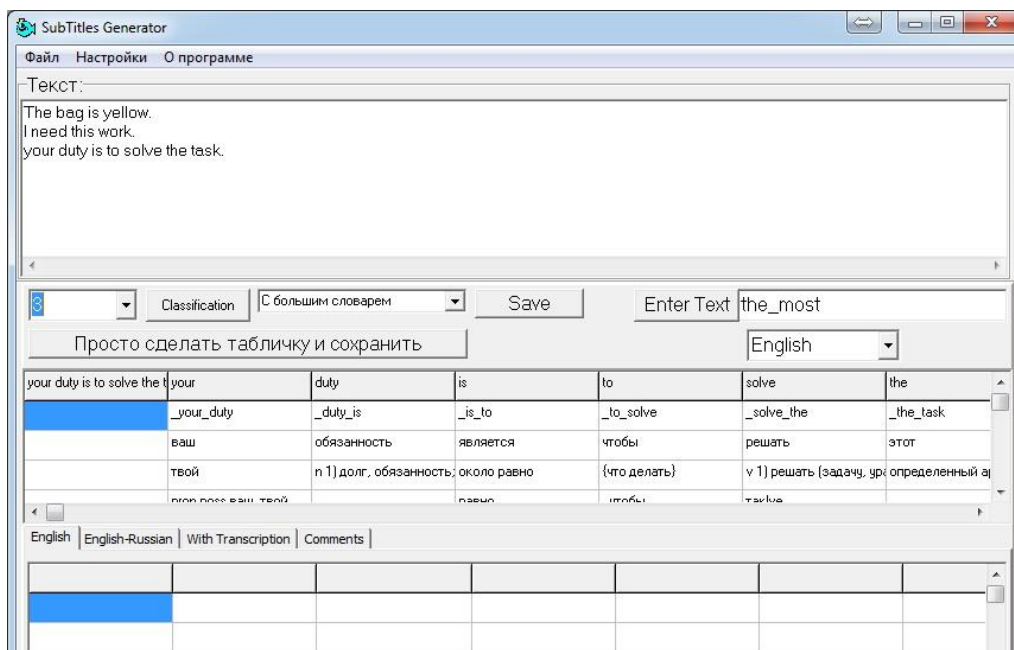


Рисунок 16 – Первичная обработка строки

После выбора строки происходит автоматическое разбитие предложения «Your duty is to solve the task» на слова.

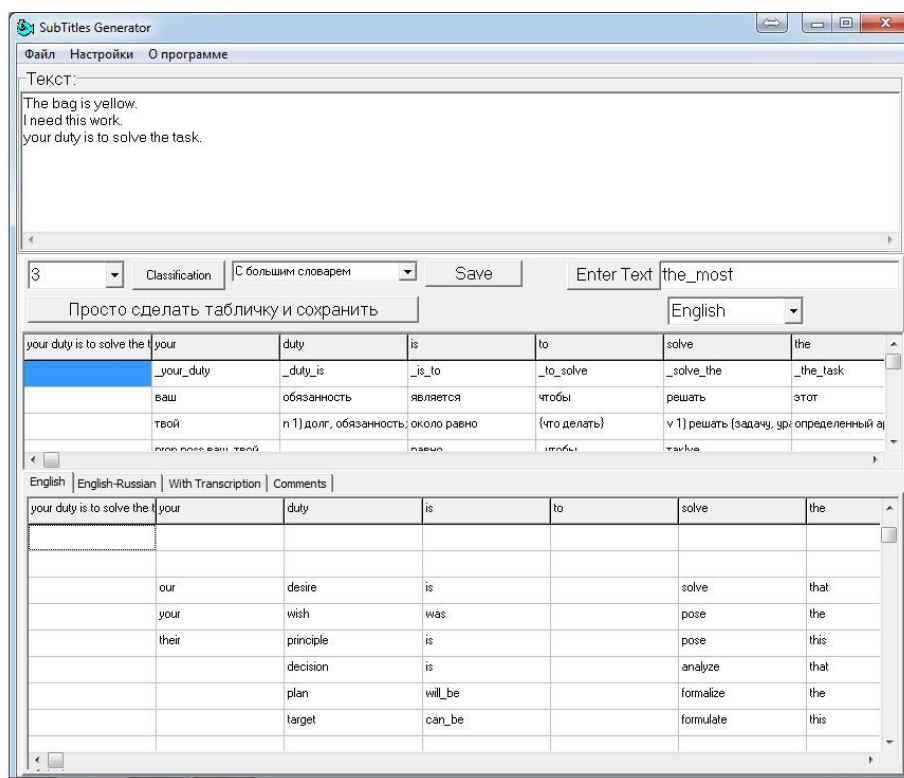


Рисунок 17 – Вывод вариантов

Для вывода вариантов групп слов необходимо нажать кнопку «Classification». После обработки выведется список с вариантами омонимических подгрупп слов для каждого слова в предложении.

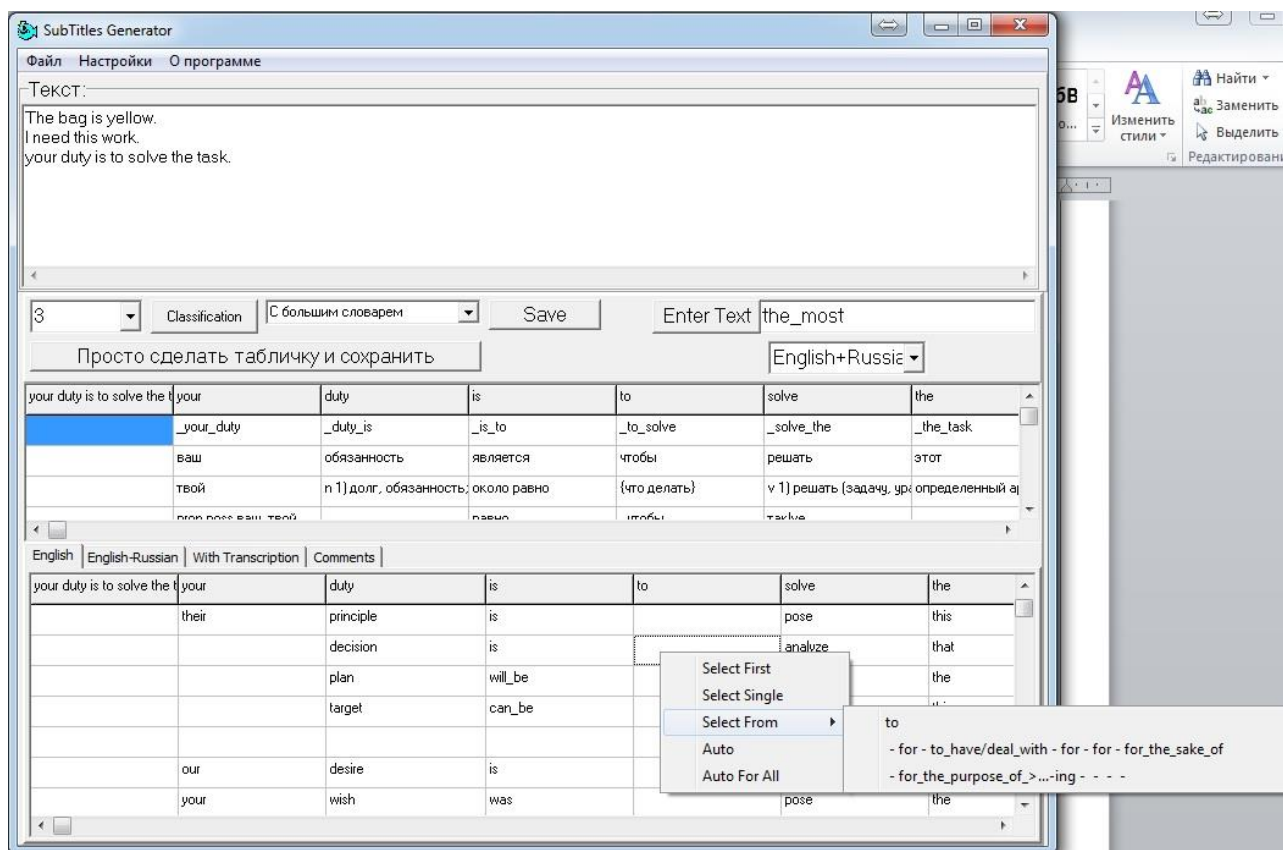


Рисунок 18 – Создание подстановочной таблицы и ее настройки

Пример выбора вариантов по каждому слову возможен в следующих вариациях:

Select First – выбор наиболее частотного варианта из представленных групп слов.

Select Single – выбор первого варианта из представленных групп слов.

Select From – выбор производится по каждой колонке пользователем самостоятельно, варианты указываются во всплывающем диалоговом окне.

Auto – автоматическое заполнение поля для данной колонки.

Auto For All – автоматическое (случайное) заполнение поля для всех групп.

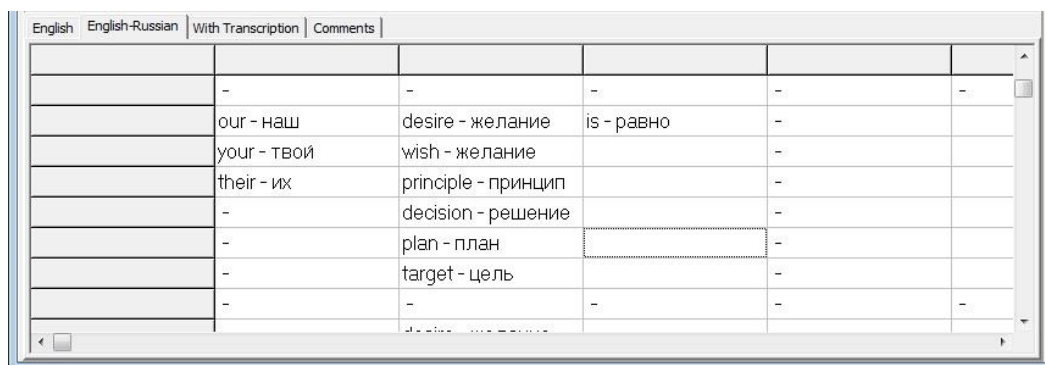


Рисунок 19 – Перевод слов

English	English-Russian	With Transcription	Comments			
	our - Ауэ - наш	desire - дизАйэ	is - Из - равно	--		that - ЗАЕ
	your - йО: - твой	wish - вИш - жел		--		the - Зэ - э
	their - ЗЕэ - их	principle - прИнт		--		this - ЗИс -
	--	decision - дисИж		--		that - ЗАЕ
	--	plan - плАЕн - п.		--		the - Зэ - э
	--	target - тАгэт - Л		--		this - ЗИс -
	--	--	--	--	--	--
	--	desire - дизАйэ		--	--	--

Рисунок 20 – Транскрипция слов

Для изучения выбранных слов используется вкладка с транскрипцией и переводом, которая поможет лицам, изучающим английский язык, правильно произносить и переводить сгенерированные слова.

Подстановочные таблицы можно сохранить в текстовом документе и использовать при необходимости для повторения и закрепления пройденного материала.

Для сохранения подстановочных таблиц используется файл с расширением ".doc".

Имя файла совпадает с текстом выбранного для обработки предложения и располагается в папке «Results».

Пример:

При обработке строки «Don't worry be happy».

Файл сохранится как «OpenAndSpeakOn - Don't worry be happy - topic.doc».

<u>your</u>	<u>artificial limb</u>	<u>is</u>	<u>wonderful</u>
<u>our</u>	<u>prosthetic device</u>	<u>was</u>	<u>excellent</u>
<u>her</u>	<u>tuft</u>	<u>is</u>	<u>great</u>
<u>their</u>	<u>forelock</u>	<u>is</u>	<u>good</u>
	<u>lock</u>	<u>will be</u>	<u>ideal</u>
	<u>curl</u>	<u>can be</u>	<u>perfect</u>

Рисунок 21 – Сохраненный результат работы программы

Сохраненные в файл подстановочные таблицы можно использовать для изучения либо в качестве входных данных для комплекса программных продуктов «GenerationOffice».

При нажатии на кнопку «Просто сделать табличку и сохранить» программа автоматически, без вывода результата на экран, сохраняет самый подходящий результат, отобранный по частотности, лексике и грамматике.

3 Интеграция программы «Генератор подстановочных таблиц» в ряд программных продуктов

Использование подстановочных таблиц, сгенерированных разработанной программой «Генератор подстановочных таблиц», возможно не только в качестве учебного материала, но и как входные данные для ряда программных продуктов через агенты. Реализация агентов происходит посредством использования программы Automatic Mouse and Keyboard, которая автоматизирует процесс наполнения электронных курсов учебными материалами. Программный продукт «Automatic Mouse and Keyboard» избавляет от монотонных действий пользователя, путем разовой записи нужных действий (существует возможность изменить нужный момент в записи) для дальнейшего выполнения.

3.1 Описание программы Automatic Mouse and Keyboard

Используя программный продукт «Automatic Mouse and Keyboard» можно при помощи клавиатуры и мыши имитировать самые разные действия. Принцип действия такой: программа записывает все передвижения и нажатия клавиш мышью компьютера в специальный макрос, который потом воспроизводится, для того чтобы автоматизировать рутинные операции. Программа позволяет выполнять самые разные задачи: например, вам необходимо сохранять фотографии из Интернета, вводить один и тот же текст, открывать некоторые файлы или нажимать на определенные кнопки. Чтобы постоянно не выполнять одинаковые действия, вы записываете порядок действий, и компьютер сам все сделает.

Автокликер в точности повторяет все записанные действия: как клавиатурные, так и движения мышки. «Automatic Mouse and Keyboard» позволяет записывать разные алгоритмы. Например, сначала вы выставляется курсор правильно на экране, потом делается нажатие, чтобы указатель переместился на поле, а дальше набирается текст. Это полезно, если часто рассылаются одинаковые письма или сообщения. Но это далеко не предел возможностей.

Программа полезна как для работы в офисе, так и для операций дома. Позволяет создавать часто повторяемые сценарии для игр или задач. При скачивании «Automatic Mouse and Keyboard» можно не только повторять действия, как это делают аналогичные программы, а и планировать их запуск, редактировать записанные сценарии, регулировать скорость их воспроизведения. Если сценариев много, вам доступна их настройка и определение, какие будут выполняться в первую очередь, а какие во вторую и так далее. Программа дает возможность оптимизировать и автоматизировать множество задач [39].

Отличительные черты Automatic Mouse and Keyboard:

- точная имитация действий клавиатуры и компьютерной мыши;
- автоматический ввод текста;

- возможность редактирования уже созданного сценария;
- функция планировщика запуска сценариев;
- запись действий мыши и клавиатуры;
- возможность скомпилировать скрипт в формат EXE.

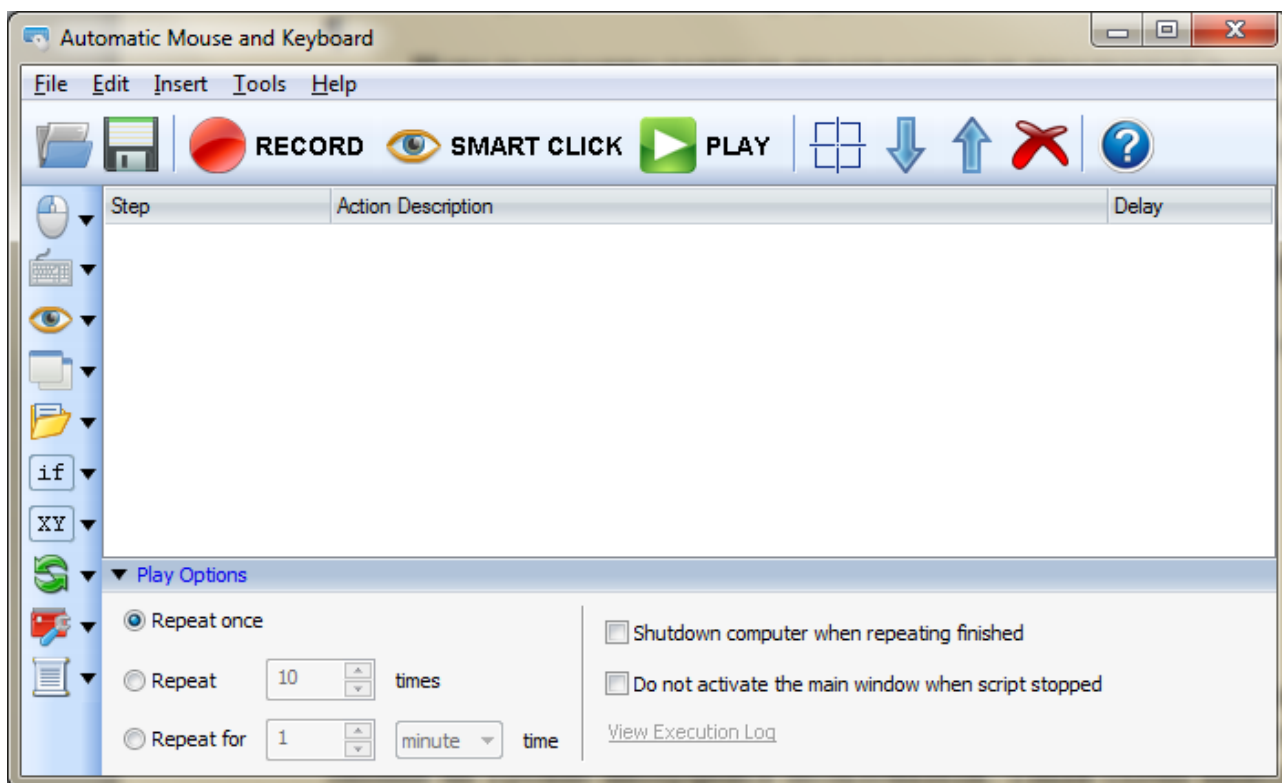


Рисунок 22–Основное окно программы

Требования к системе:

- Процессор: 800 MHz и выше;
- Разрядность: поддержка 32-разрядных и 64-разрядных систем;
- Оперативная память: не менее 512 Mb;
- Язык: Английский;
- Лицензия: Условно-бесплатно.

3.2 Генерация учебных материалов

Использование комплекса программ для генерации учебных материалов необходимо для обеспечения возможности повышения эффективности обучения иностранным языкам.

Подготовка новых учебных материалов производится в большом объеме и в разнообразных формах для электронного курса, используя более совершенные методы генерации учебных заданий, что должно повысить качество обучения.

Используются нижеперечисленные программы:

- 1) Open-and-ReadGenerator;

- 2) Text-Based Task Generator;
- 3) Semantic Electronic Dictionary;
- 4) Patterns Maker;
- 5) Генератор осмысленных фраз на основе векторов признаков;
- 6) Генератор текстовых заданий по стандартным шаблонам;
- 7) Генератор большого объема текстов по стандартным шаблонам;
- 8) Transcription Generated.

Данные программы предназначены соответственно для:

- 1) Генерация учебных материалов по методу Ильи Франка;
- 2) Составление большого объема вариантов текстовых заданий на основе входного текста;
- 3) Обеспечивает возможность выбора материалов из семантической классификации: в форме списков слов, подстановочных таблиц, с транскрипцией, без транскрипции, в алфавитном, в семантическом порядке и другими заданными параметрами;
- 4) Программа дающая возможность ранжирования и выборки учебных заданий к текстам, рисункам, аудио;
- 5) Генерация однотипных фраз, относящихся к разным темам на основе метода декларативного программирования;
- 6) Генерация предложений на основе шаблонов и составление учебных заданий на основе данных предложений. Заданий типа: выбор правильного варианта, множественный выбор и ранжирование слов в предложении;
- 7) Генерация большого объема ассоциативно связанных предложений в текст на основе различных шаблонов;
- 8) Генерация транскрипции на основе различных европейских языков, с генерацией ошибочных вариантов для их исправления в качестве учебных заданий [40].

Использование данных программных продуктов поможет:

- 1) обеспечить курсы разнообразными новыми материалами (текст, аудио, видео файлы, изображения по темам, тесты к пройденному материалу);
- 2) повысить качество обучения иностранного языка за счет всестороннего погружения в тему, благодаря информационной разнообразности наполняемого контента;
- 3) увеличить количество индивидуальных материалов для студентов СФУ ИКИТ.

3.3 Описание основных используемых программ

Программа «Electronic Dictionary».

Программа представляет собой визуализацию процесса работы порождающих грамматик в форме дерева вывода для некоторой части области знаний на уровне интерфейса пользователя. Такая схема вывода строится на базе правил, описывающих соответствующие релевантные классы объектов или их состояний и переходы между объектами состояниями, известными в этой области, заданными наборами правил порождающих грамматик[40].

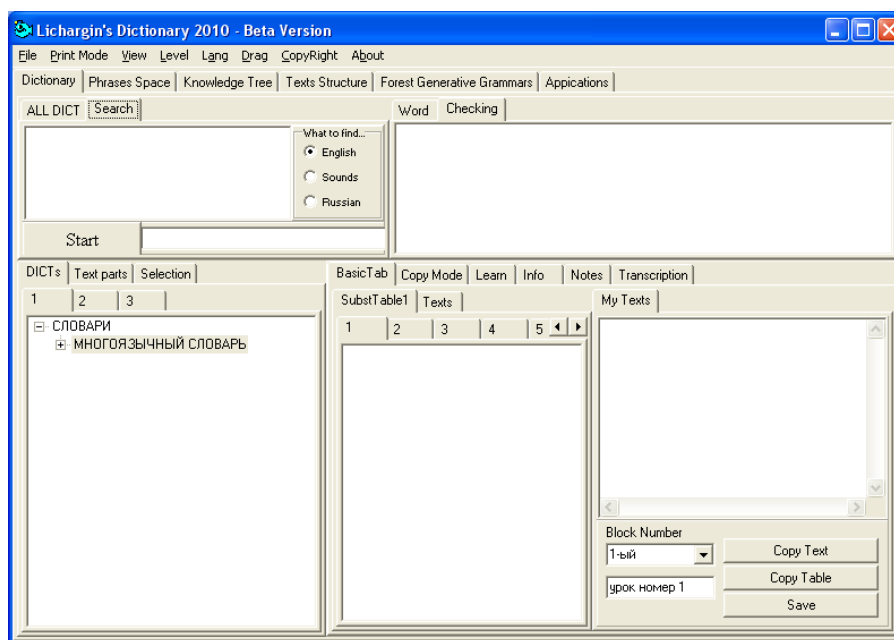


Рисунок 23—Основное окно программы

Из этой бесплатной программы делается импорт предложений или слов по указанной теме. Работа большинства программ строится на основе словарей.

Одной из ключевых программ для генерации таких словарей и подстановочных таблиц служит «Electronic Dictionary» выше представлено её рабочее окно.

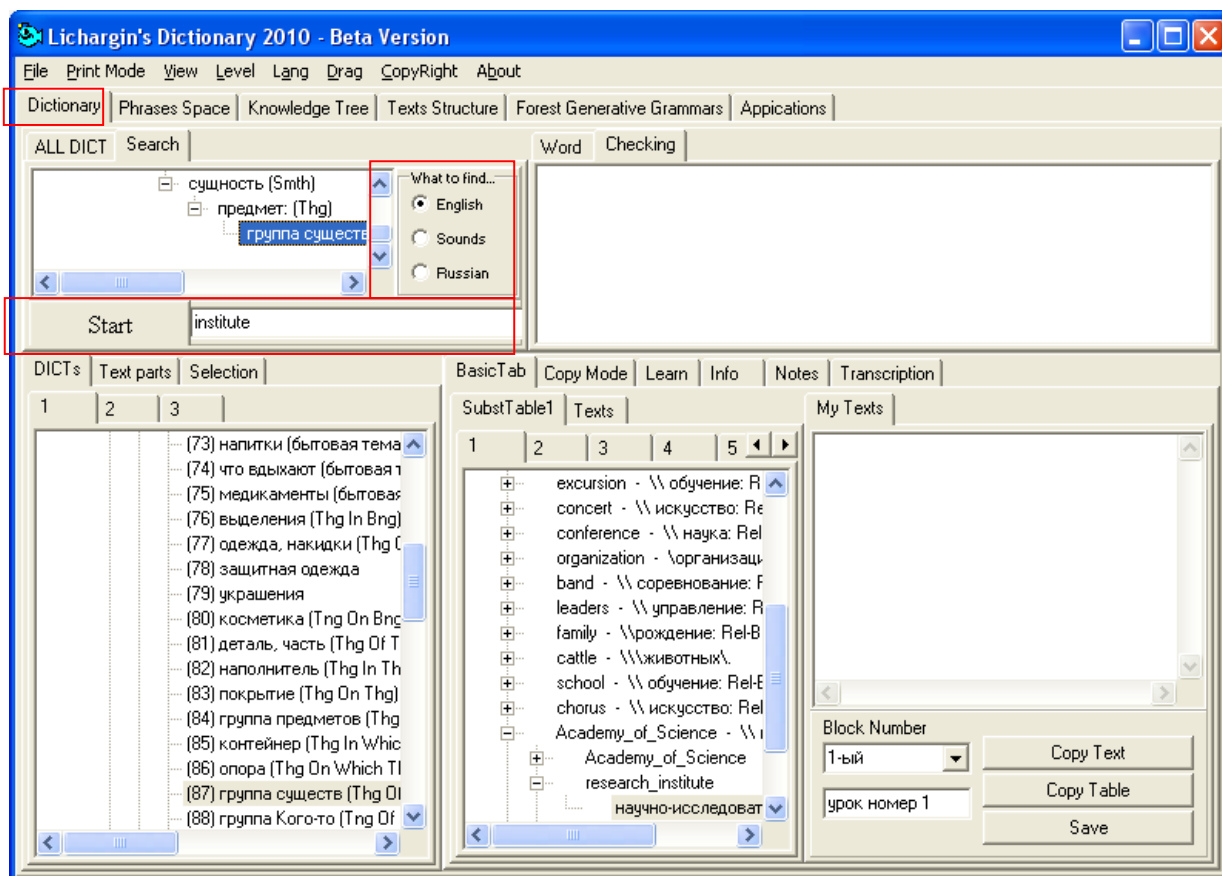


Рисунок 24– Подбор групп слов

Работа программы заключается в поиске слов по заданной теме. Для этого нужно выбрать параметры поиска (ищем по английскому слову, по транскрипции или по русскому слову).

В строке «поиск» пишем интересующее нас слово и нажимаем «Start». После этого программа из собственного общего словаря предложит несколько тематик его использования. Далее необходимо выбрать интересующую нас тематику [41].

Благодаря немалому объему словаря генерируется большое разнообразие слов, которые можно использовать как при изучении языка, так и для их использования в ряде бесплатных программных продуктов представленных ниже.

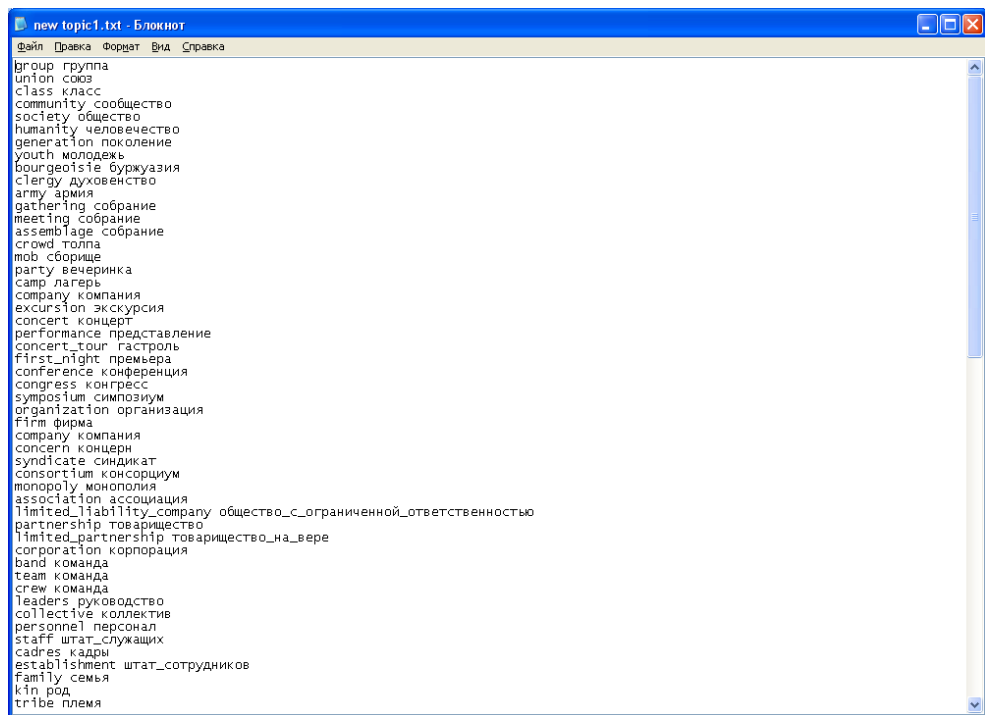


Рисунок 25 – Сохраненный файл с подобранными группами слов

Пример сохраненного файла по одной из тематик, где используется слово институт. Программа подобрала группу слов, в разных вариациях обозначающую группу людей, так и учебный процесс.

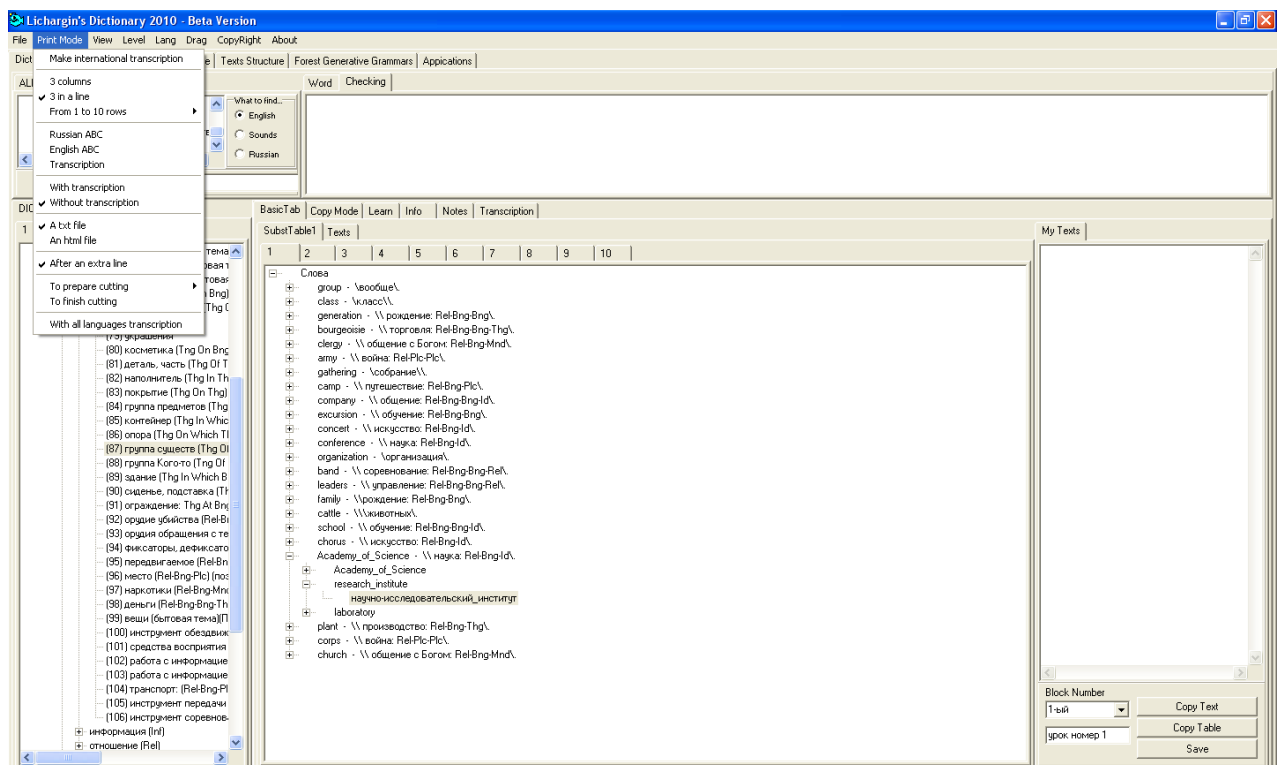


Рисунок 26 – Настройки вывода результата в программе «Electronic Dictionary»

В меню «Print Mode» можно выбрать настройки для отображения слов в

сохраняемом файле: 3 колонки, 3 в линию, от 1 до 10 строк (от 2 до 6 столбцов), с транскрипцией, без транскрипции. Также возможно сохранение в формате TXT либо в HTML. В словаре указывается как английское слово, так и его перевод, транскрипция, что упрощает восприятие и понимание результата работы программы как для пользователя с более высокими знаниями английского языка, так и для новичка.

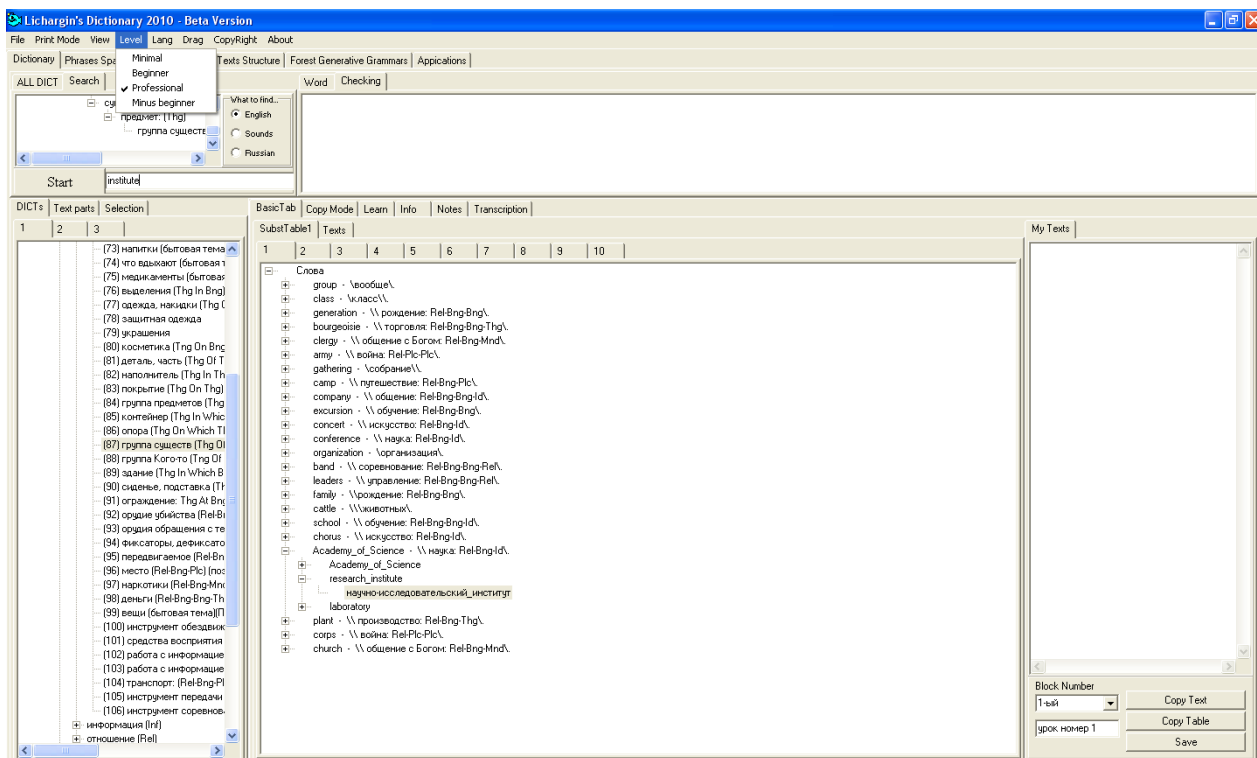


Рисунок 27 – Выбор сложности

Также возможен выбор уровня сложности подбора групп слов, для индивидуальной настройки.

Minimal – необходимы минимальные знания, небольшой запас слов.

Beginner – начальный уровень, есть минимальная база знаний английского языка.

Professional – хорошее владение языком, объемный словарный запас.

Minus beginner – при данной настройке отсеиваются простые слова, для пользователей с хорошим знанием языка.

Из данной программы возможна экспортация слов для базы омонимических групп слов, для ее расширения либо подбора базы по заданной тематике.

Возможно комбинирование нескольких вариантов тем в одну. Редактирование результата путем добавления и удаления слов в список результата.

Программа «Генератор текстовых заданий по стандартным шаблонам».

Сгенерированные подстановочные таблицы из программы «Генератор подстановочных таблиц» экспортируются в программу «Генератор текстовых заданий по стандартным шаблонам».

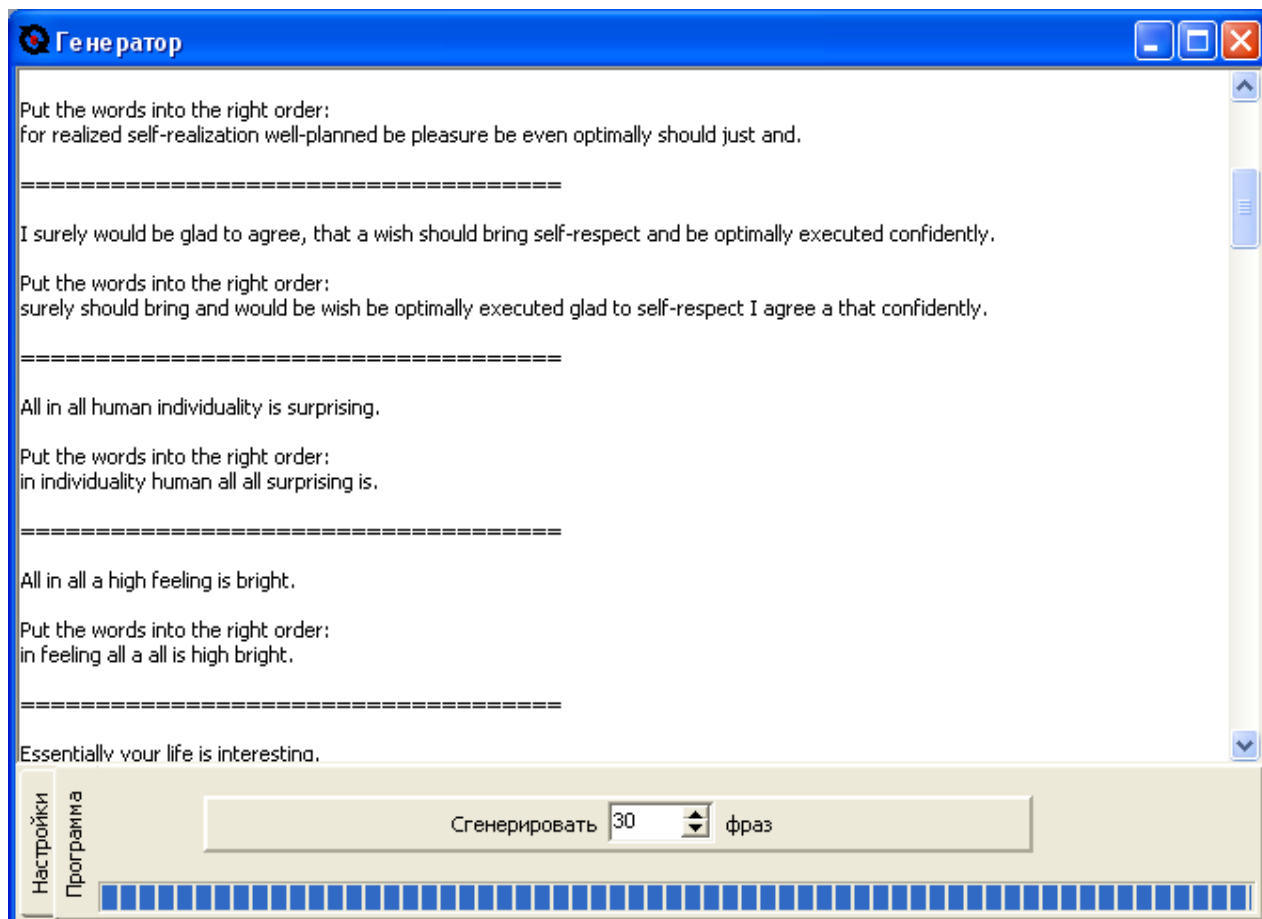


Рисунок 28 – Генератор текстовых заданий по стандартным шаблонам

На основе подстановочных таблиц строятся учебные задания, представленные на рисунке 19.

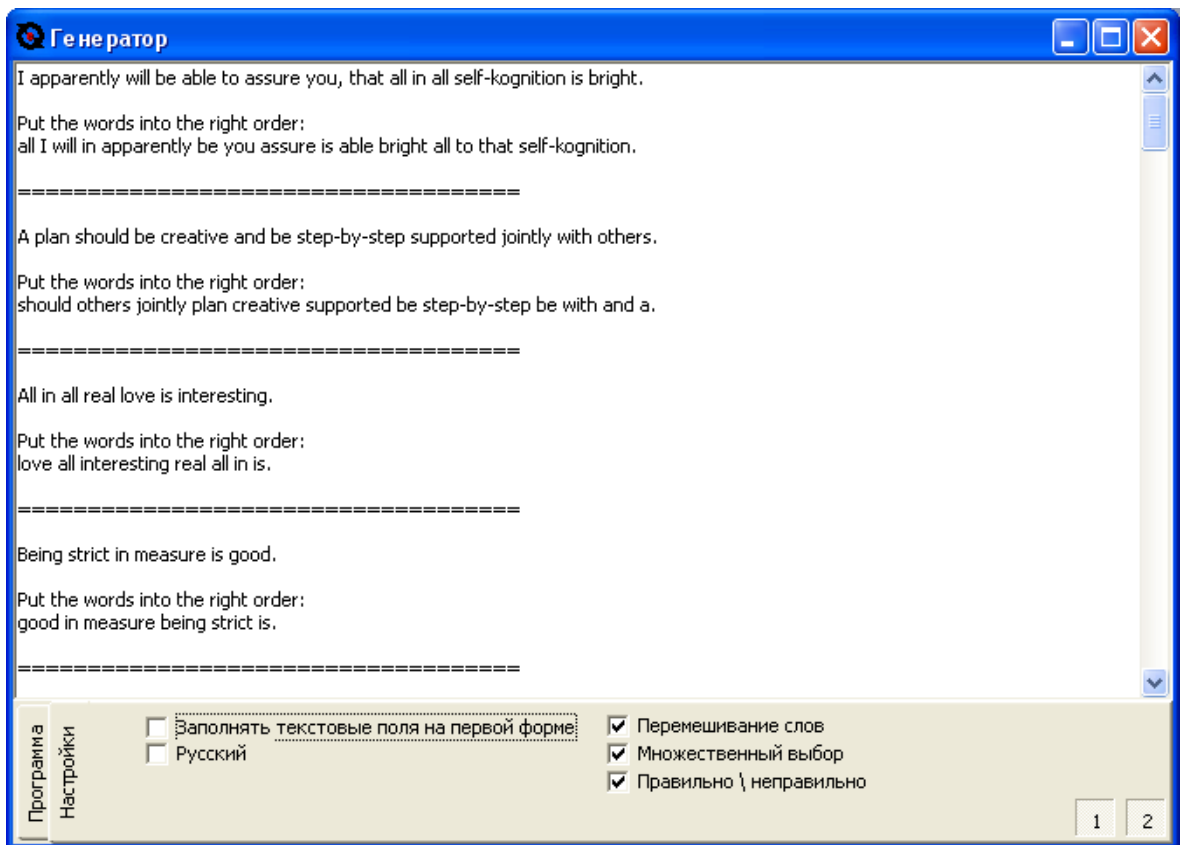


Рисунок 29 – Генерация фраз с учебными заданиями

В настройках можно выбрать тип учебных заданий для генерации:

Put the words into the right order (расположить слова в правильном порядке). Слова в исходных предложениях ранжируются в случайном порядке. Перед обучающимся ставится задача составить правильное с точки зрения грамматики и осмысленности предложение.

Multiple choice (множественный выбор). Выбор правильного варианта из предложенных пользователю.

Right/wrong (правильно / неправильно). Необходимо определить правильно указан перевод слова или нет.

По такому принципу программа «Генератор текстовых заданий по стандартным шаблонам» генерирует выбранное количество фраз с учебными заданиями.

С помощью программы «Patterns Maker» генерируются разнообразные задания к каждому уроку, для этого нужно выбрать количество заданий и нажать кнопку «Generate».

Задания составляются на русском языке. Результат выводится в текстовый файл.

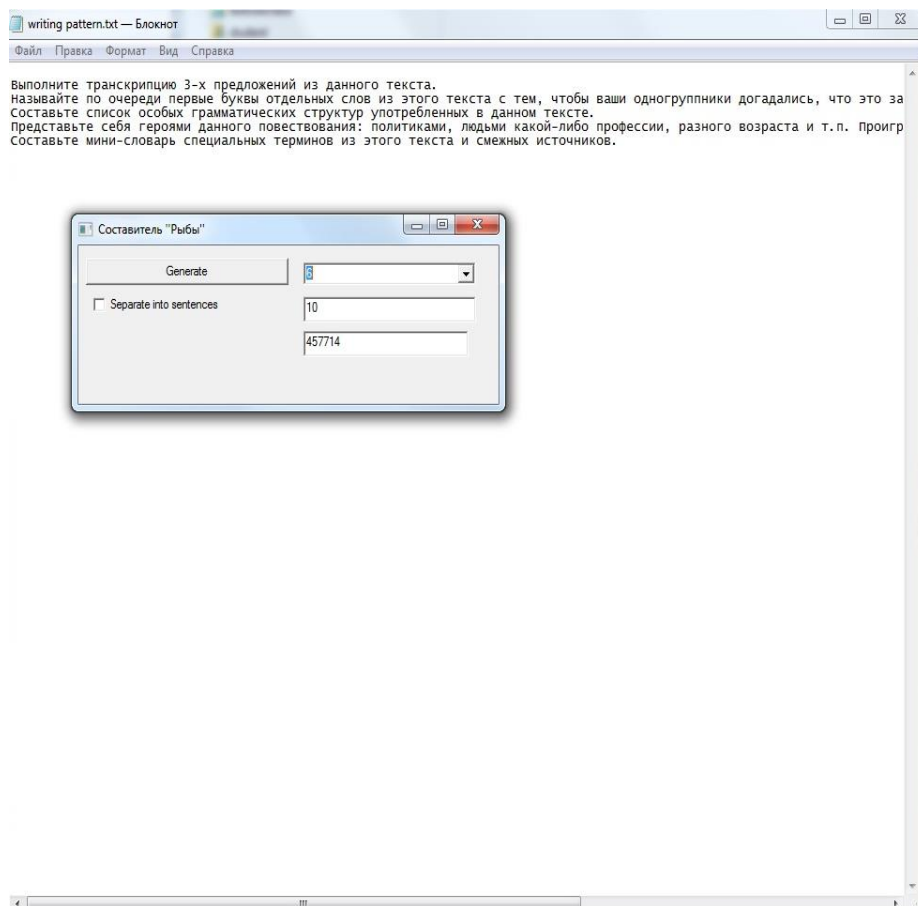


Рисунок 30– Генерация заданий к урокам

Задания типа:

- 1) Называйте по очереди первые буквы отдельных слов из этого текста, чтобы ваши одноклассники догадались, что это за предложение.
- 2) Составьте список особых грамматических структур употребленных в данном тексте.
- 3) Представьте себя героем данного повествования, проиграйте роль выбранного персонажа.
- 4) Составьте словарь специальных терминов из этого текста и смежных источников и др.

Получаемые фразы можно озвучить и выложить в электронный курс, с помощью программы для озвучивания текста.

Для создания аудио материалов была выбрана бесплатная программа «acapela», режим доступа: www.acapela-group.com.

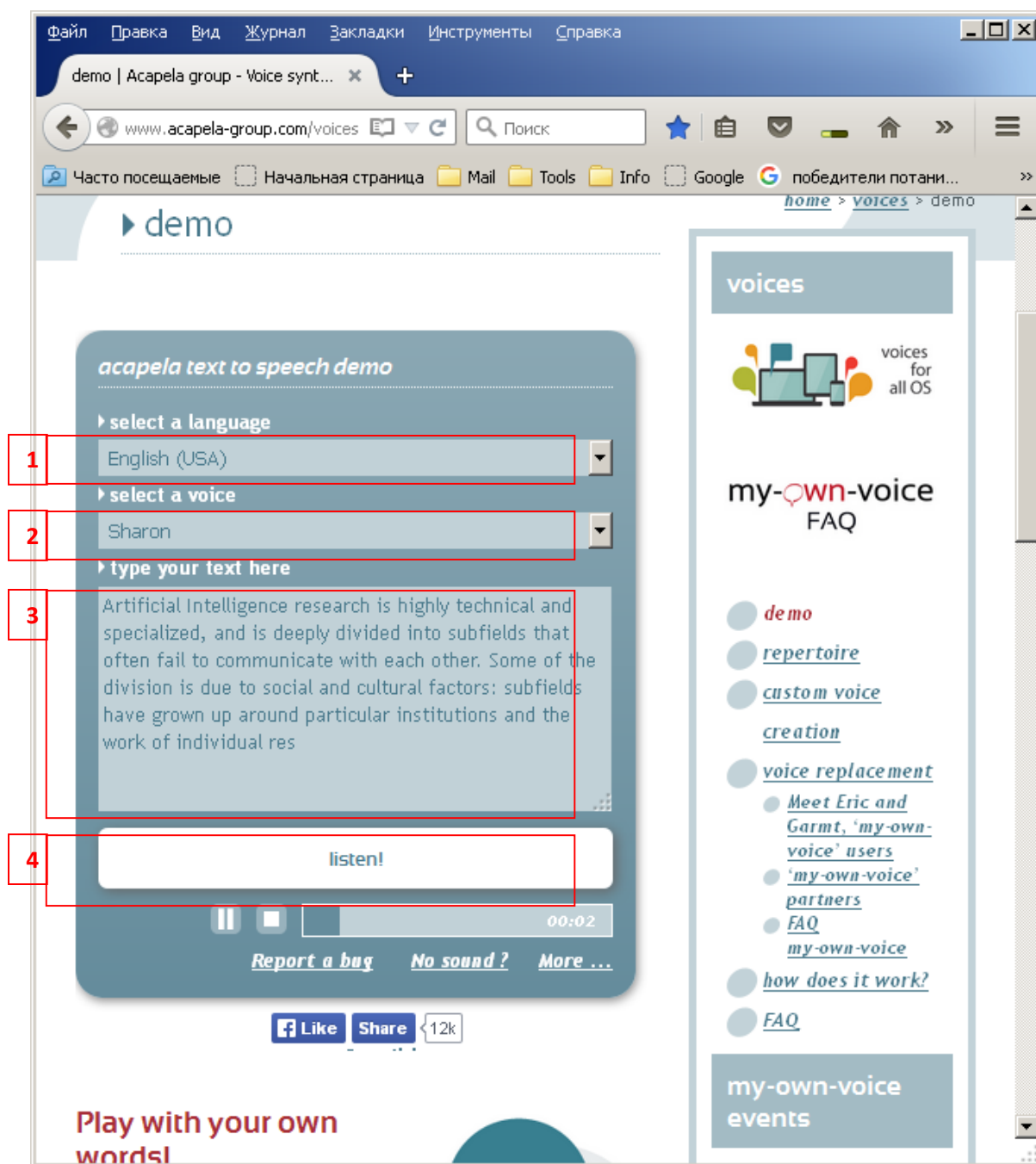


Рисунок 31– Озвучивание текста

Перед началом работы необходимо:

- 1) выбрать язык озвучиваемого текста;
- 2) выбрать голос;
- 3) ввести в 3 поле текст;
- 4) нажать кнопку «listen».



Рисунок 32– Расширения для Firefox

В браузере Firefox устанавливаем свободное расширение для Firefox, для обработки одной и нескольких ("все" и "Выбор") загрузок с нескольких внешних менеджеров FlashGot!, находящееся по адресу <https://flashgot.net/>. Для его применения нужно перейти по вкладке **Инструменты - FlashGot – Мультимедиа**.

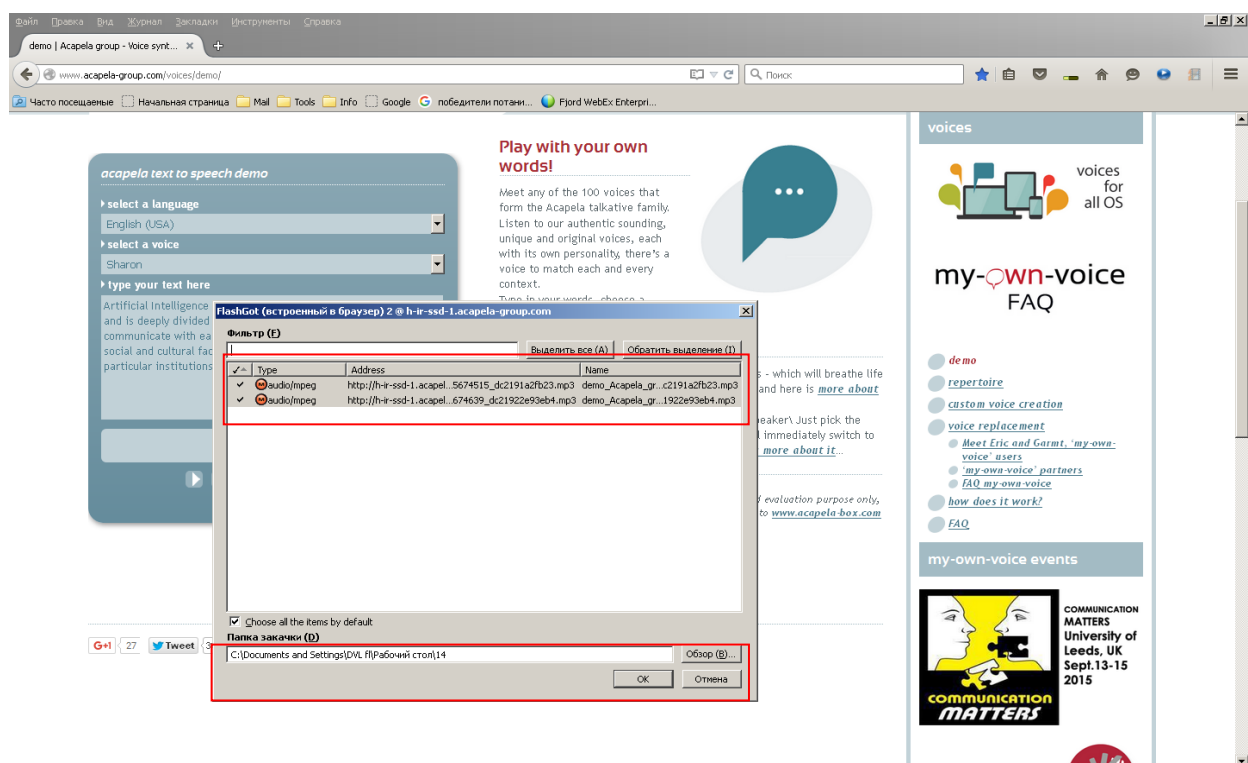


Рисунок 33– Сохранение аудио файлов

Для сохранения файлов в появившемся окне нужно отметить звуковые дорожки, выбрать папку для закивания аудио файлов.

3.4 О системе электронного обучения Сибирского федерального университета

В соответствии с новым законом об образовании в Сибирском федеральном университете широко практикуется использование электронного обучения, дистанционных образовательных технологий, что уже отражается в документации к основным образовательным программам и в учебных программах дисциплин.

Преподаватель, при организации аудиторных занятий, может планировать работу с учетом того, что студенты часть нагрузки выполняют с помощью ЭОР. В свою очередь, ЭОР могут избавить преподавателя и студента от рутинных действий (выдача заданий, проверка работ, комментирование и рецензирование, доступ к справочным и дополнительным материалам) [42].

Кроме того, с помощью ЭОР можно повысить уровень организации самостоятельной работы студентов.

По приказу ректора № 748 от 17 июня 2014 года ЭОР размещаются, режим доступа: e.sfu-kras.ru.

Новая система представляет собой ресурс, разработанный на платформе Moodle, и имеет следующий вид:

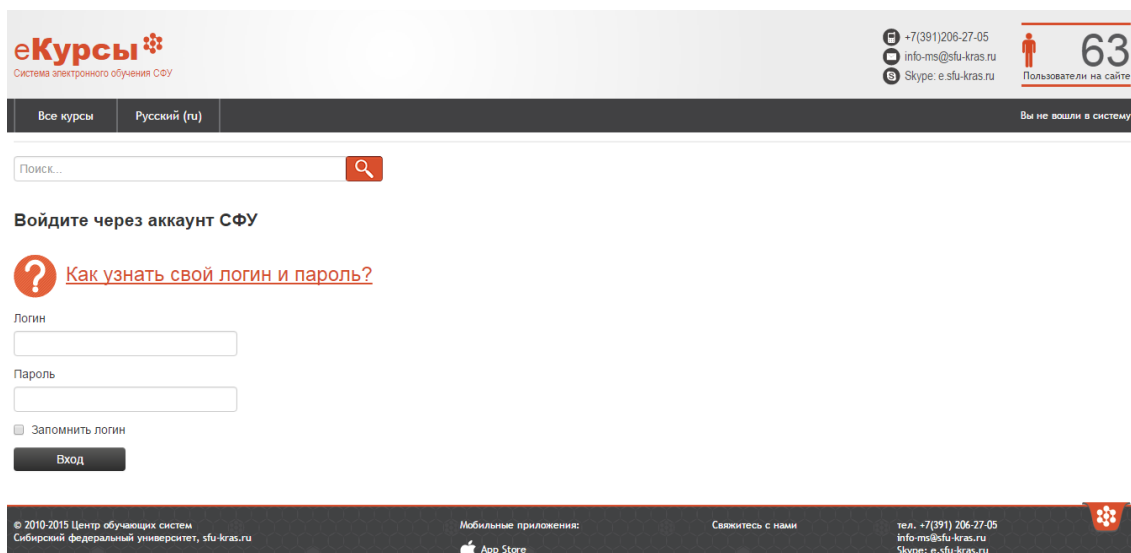


Рисунок 34 – Система электронного обучения СФУ на платформе Moodle

Moodle – модульная объектно-ориентированная динамическая учебная среда. Moodle – это пакет, который обычно определяют как CMS или LMS. Эти аббревиатуры можно расшифровать следующим образом:

- CMS – course management system – система управления курсами;
- LMS – learning management system – система управления обучением.

Moodle написана на языке программирования PHP с использованием SQL-базы данных профессором из Австралии Мартином Дунгиамосом и переведена на несколько десятков языков и используется для обучения более чем в ста пятидесяти странах мира.

Широкую популярность ей обеспечили простота использования и открытый исходный код. Сама система распространяется по лицензии GNU GPL и вам не придется производить лицензионные отчисления в пользу разработчика [43].

Основной учебной единицей Moodle являются учебные курсы. В рамках такого курса можно организовать:

1. Взаимодействие учеников между собой и с учителем. Для этого могут использоваться такие элементы как: форумы, чаты.
2. Передачу знаний в электронном виде с помощью файлов, архивов, веб-страниц, лекций.
3. Проверку знаний и обучение с помощью тестов и заданий. Результаты работы ученики могут отправлять в текстовом виде или в виде файлов.
4. Совместную учебную и исследовательскую работу учеников по определенной теме, с помощью встроенных механизмов wiki, семинаров, форумов и пр [43].

Moodle в современном образовании.

Современный мир стремительно меняется. Развитие компьютерной техники и средств связи кардинально меняют образ жизни человека. Не осталось в стороне от этих изменений и образование. Прошли те времена, когда для осуществления процесса обучения был необходим личный контакт учителя и ученика. Существуют способы интенсифицировать процесс обучения используя методы и средства электронного обучения. Использование возможностей электронного обучения, или как его еще называют e-learning, на основе системы управления электронными курсами Moodle обеспечит учащимся дополнительный источник учебных материалов [43].

Moodle можно использовать для организации:

1. Дистанционного обучения – при котором преподаватель и ученик большую часть времени не встречаются лично друг с другом.
2. Дистанционной поддержки очного образования – используя средства электронного обучения учащийся может получать задания и отправлять их на проверку используя систему Moodle.
3. Поддержки очного образования – выполнение отдельных практических заданий, тестов проходит вовремя учебных занятий в системе электронного обучения Moodle.

Система Moodle может обеспечить:

1. Выбор удобного времени и места для обучения как для преподавателя, так и для ученика.
2. Прочное усвоение знаний.
3. Контакт преподавателя с учеником по мере необходимости. Если ученик работает, он контактирует с преподавателем.
4. Индивидуализацию обучения.
5. Экономия времени и денег – отпадает необходимость тратить время и деньги на учебные занятия [43].

Преподаватели обладают необходимыми правами на редактирование размещенных материалов. Поэтому удалить, заменить, обновить учебные материалы можно самостоятельно в удобное время.

3.5 Наполнение электронного курса заданиями на основе сгенерированных материалов для системы дистанционного обучения СФУ

Заполнение электронных курсов происходит по каждой теме отдельно. Созданные текстовые материалы, тесты, аудио материалы дополняют каждую тему, помогают более углубленно погрузиться в изучаемый материал, узнать особенности культуры и речевого строя английского языка.

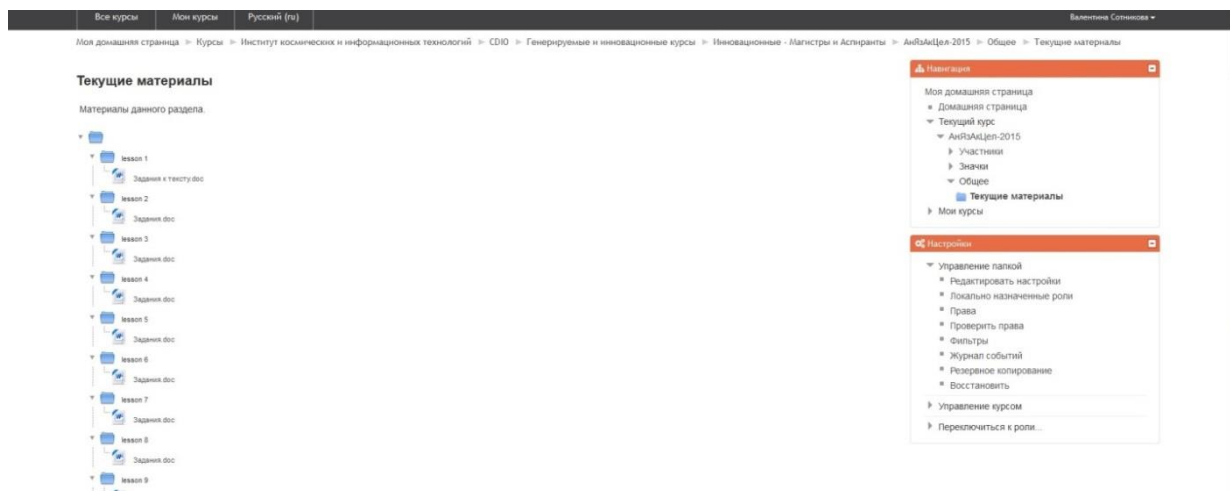


Рисунок 35 – Материалы к урокам

На рисунке выше представлен заполненный раздел с заданиями для каждого урока.

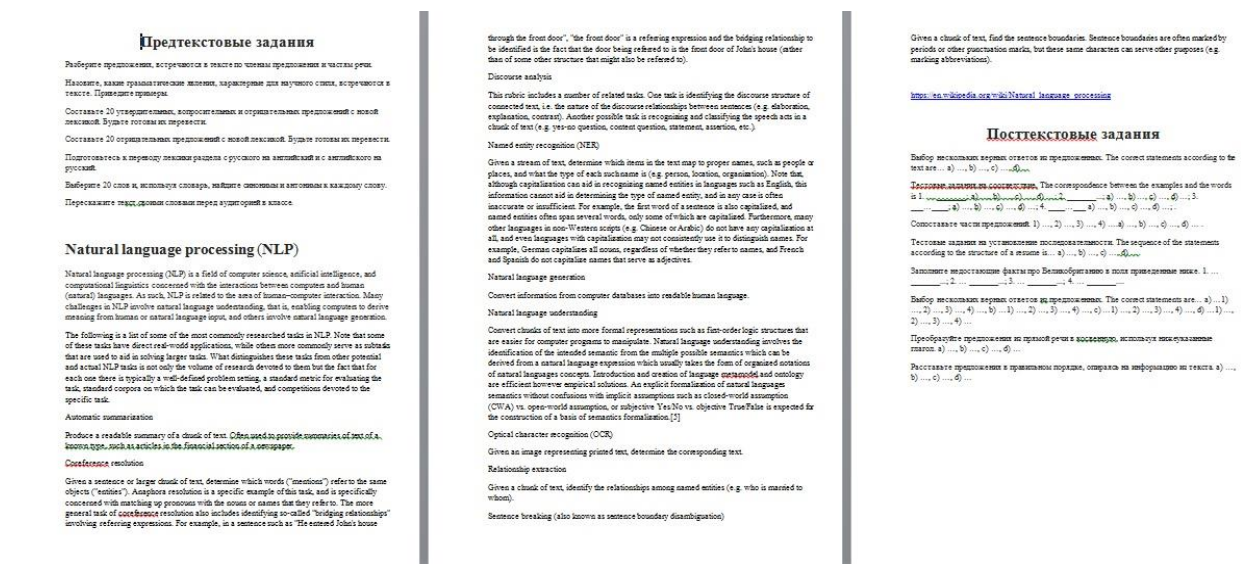


Рисунок 36 – Тексты с заданиями

Пример текста с пред текстовыми и пост текстовыми заданиями для электронных курсов, сгенерированный с помощью бесплатного комплекса программных продуктов, представлен на рисунке 27.

Данные задания ориентированы на развитие понимания лексики языка,

углубленного понимания материала, вынесенного для изучения на данном этапе.

Найдите источники и транскрипции текста:

Найдите источники и транскрипции текста: *[The following text is a dense, repetitive block of characters, likely representing a corrupted or placeholder image of a document page. It contains no legible text.]*

[The following text is a dense, repetitive block of characters, likely representing a corrupted or placeholder image of a document page. It contains no legible text.]

[The following text is a dense, repetitive block of characters, likely representing a corrupted or placeholder image of a document page. It contains no legible text.]

[The following text is a dense, repetitive block of characters, likely representing a corrupted or placeholder image of a document page. It contains no legible text.]

Рисунок 37 – Фонетический материал

Пример текста с практической и международной транскрипцией для раздела «Фонетический материал».

Такого рода задания используются в каждом разделе для всех тем.

ЗАКЛЮЧЕНИЕ

Была разработана программа «Генератор подстановочных таблиц», а также алгоритм системы построения учебных подстановочных таблиц на основе произвольных предложений естественного языка на входе системы.

Для улучшения качества обработки строк используется частотный, семантический, грамматический анализы. Их применение и описание изложено во второй главе.

Был описан метод компонентного анализа, который используется для выявления соответствий смысловых признаков в омонимических подгруппах генерируемых таблиц по методу Палмера, используемый для распознавания омонимов и оценки их предпочтения.

Разработан алгоритм построения подстановочных таблиц и расширена база комбинимов и дефинимов английского языка. Разработана база семантического анализа слов.

Также в работе рассматривается принцип реляционных баз данных, лежащие в основе генерации подстановочных таблиц, на основе произвольных входных предложений.

Для разработки программы «Генератор подстановочных таблиц» было выбрано решение объединения всех методов и разработана формула, описывающая данное решение.

Реализован перевод выходных данных программ в единую систему с использованием общего формата, с программами по компьютерной лингвистике магистратуры ИКИТ СФУ по средствам использования агентов.

В результате получен качественный программный продукт, отвечающий поставленным целям и задачам.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Личаргин, Д. В. К проблеме генерации расширенной фонетической транскрипции на основе векторных порождающих грамматик. [Электронный ресурс] / Д. В. Личаргин, В. В. Сотникова, А. А. Усова, В. В. Бутовченко // Современные проблемы науки и образования. – 2015. № 2. . – Режим доступа: <http://www.science-education.ru/ru/article/view?id=20933>.
2. Усова, А. А. Разработка веб-приложения по генерации учебных заданий к тексту на естественном языке на основе порождаемых шаблонов. / А. А. Усова, В. В. Сотникова, Д. В. Личаргин // Красноярск: Международная конференция студентов, аспирантов и молодых ученых «Перспектив-2015», посвященная 70-летию Великой Победы. – 2015. – С. 7-10.
3. Borko, H. Abstracting Concepts and Methods / H. Borko, C.L. Bernier // – New York: Academic Press, 1975. – P. 140.
4. Carbonell, J. G. The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries / J. G. Carbonell, J. Goldstein // 21st International ACM SIGIR Conference Research and Development in Information Retrieval. – New York, ACM Press, 1998. – P. 335-336.
5. Личаргин, Д. В. Векторные порождающие грамматики и их применение для генерации учебных материалов по фонетике языка. / А. А. Усова, В. В. Сотникова, И. В. Колбасина, Е. Д. Старовойт // Новосибирск: X Международная научно-практическая конференция: «Научные перспективы XXI века. Достижения и перспективы нового столетия». – 2015. - С. 60-62.
6. Ando, R. K. Multidocument Summarization by Visualizing Topical Content / R. K. Ando // ANLP/NAACL 2000 Workshop on Automatic Summarization, 2000. – P. 79-88.
7. Higashinaka R., Miyazaki N., Nakano M., Aikawa K. Evaluating Discourse Understanding in Spoken Dialogue Systems // ACM Translation on Speech and Language Processing. 2004. Vol. 1. P. 1-20.
8. SeoGenerator [сайт]. – Режим доступа: <http://seogenerator.ru/>
9. Generating The Web [сайт]. – Режимдоступа: <http://www.generating-the-web.com/generating-the-web.shtml>.
10. Delirium 1.8 [сайт]. – Режим доступа: <http://softsearch.ru/programs/143-408-delirium-download.shtml>.
11. Мой SEO Гид [сайт]. – Режим доступа: <http://myseogid.ru/tools/196-article-clone-easy.html>.
12. TemplateBest [сайт]. – Режим доступа: <http://www.templatebest.ru/prog01.php>.
13. SyMonum [сайт]. – Режим доступа: <http://www.seosin.ru/page/6/>.
14. ProWebber [сайт]. – Режим доступа: <http://prowebber.ru/web/programs/11662-monkey-write-v123-unikalizator-teksta.html>.

15. Grosz-Ngate, M. Collaborative Design in a Networked Multimedia Environment: Emerging Communication Patterns / M. Grosz-Ngate // Journal of Research on Computing in Education, Vol.26, Iss.3, 1994. – P. 418-419.
16. Hahn, U. Knowledge-Based Text Summarization: Saliency and Generalization Operators for Knowledge-Based Abstraction / U. Hahn, U. Reimer // – USA, Massachusetts, Cambridge, MIT Press, 1999. – P. 215-232.
17. Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. / J. J. Hopfield // – National Academy Science, 1982. – P. 254-258.
18. DeVito, J. A. The communication handbook: A dictionary / J. A. DeVito // – New York: Harper&Row, 1986. – P. 4-10.
19. Everett, D. R. Computer-Mediated Communication as a Teaching Tool: a Case Study / D. R. Everett, T.C. Ahern // Journal of Research on Computing in Education, Vol.26, Iss.3, 1994. – P. 336-337.
20. Francis, W. N. Manual of Information to accompany A Standard Corpus of Present-Day Edited American English, for use with Digital Computers / W. N. Francis, H. Kucera // – Brown University Providence, Rhode Island Department of Linguistics Brown University, 1979 – P. 18-21.
21. Goldstein, J. External and Internal Noises [Электронный ресурс] / J. Goldstein // The institute for strategic clarity. – 2007. – Режим доступа: <http://www.instituteforstrategicclarity.org/epnoise.htm>.
22. Goodfellow, R. CALL Programs for Vocabulary Instruction / R. Goodfellow // Computer Assisted Language Learning Journal, Vol.8, №2, 1995. – P. 205-226.
23. Graziadei, W. D. Virtual Instructional Classroom Environment in Science (VICES) in Research, Education, Service & Teaching (REST) / W. D. Graziadei // – New York: SUNY Press, 1993. – P. 257- 258.
24. Личаргин, Д. В. Разработка приложения по генерации учебных заданий к тексту на естественном языке на основе порождаемых шаблонов. / А. А. Усова, В. В. Сотникова, С. А. Липман, В. В. Бутовченко. // Современные проблемы науки и образования. – 2015. № 6. – Режим доступа: <http://www.science-education.ru/ru/article/view?id=22636>.
25. Личаргин, Д. В. К вопросу об упорядочении многоуровневой семантической сети на дереве семантической классификации. / Д. В. Личаргин, К. В. Сафонов, О. И. Егорушкин, Е. П. Бачурина // Вестник СибГАУ. – 2014. №2. С. 44-50.
26. Hutchins, J. Summarization: Some Problems and Methods / J. Hutchins, K. Jones // Informatics: Meaning the Frontier of Informatics. – London, Aslib, 1987. – P. 151-173.
27. Chomsky, N. Syntactic Structures: Mouton de Gruyter. 2002/ P. 119
28. Comrie, B. Language Universals and linguistic typology / B. Comrie // University Of Chicago Press, 2nd edition, 1989. – P.230-239.
29. Cornelius, T. Leondes Intelligent systems: technology and applications. / T. Cornelius // – CRC Press, 2002. – P. 127-128.

30. Crystal, D., English Language / D. Crystal // Wikipedia. Free Encyclopedia, режимдоступа http://en.wikipedia.org/wiki/English_language.
31. Филатова, Н. Н. Центр «Компьютерные технологии образования» [Текст] / Н. Н. Филатова, О. Л. Ахремчик // EducationalTechnology&Society, №3, 2000, режим доступа: http://ifets.ieee.org/russian/depository/v3_i2/html/5.html.
32. Хомский, Н. Синтаксическиеструктуры. В кн.: Новое в зарубежной лингвистике. Вып. V. Москва, 1962.
33. Беляков, В. М. Разработка функциональной модели автоматизированной обучающей системы по русскому языку как иностранному: автореферат диссертации к.ф.н. / В. М. Беляков // – Москва, 1996. – С. 45-47.
34. Штрик, А. С. CASE: автоматизированное проектирование программного обеспечения [Текст] / А. С. Штрик // Монитор, №4, 1992. – С. 4-6.
35. Личаргин, Д. В. Статистические методы анализа естественного языка как способ повышения эффективности его генерации на основе семантических шаблонов / Д. В. Личаргин, А. Ю. Маглинец, М. В. Рыбков, Е. П. Бачурина. // Информатизация образования и науки. –2014. №4(24). – С. 92-103.
36. СТО 4.2–07–2014 Система менеджмента качества. Общие требования к построению, изложению и оформлению документов учебной деятельности. – введ. 09.01.2014. – Красноярск: ИПК СФУ, 2014. – 60 с.
37. Терехина, А. Ю. Анализ данных методами многомерного шкалирования [Текст] / А. Ю. Терехина // – Москва: Наука, 1986. – 168 с.
38. Терехов, С. А. Нейросетевые информационные модели сложных инженерных систем. Нейроинформатика [Текст] / С. А. Терехов // – Новосибирск. Наука. 1998. – С.101-136.
39. All-freeload.net [сайт]. – Режим доступа: <http://all-freeload.net/avtoklikery/1948-automatic-mouse-and-keyboard>.
40. Сотникова, В. В. Векторные порождающие грамматики и их применение для генерации учебных материалов по фонетике языка. / В. В. Сотникова, А. А. Усова, Д. В. Личаргин //Красноярск: Международная конференция студентов, аспирантов и молодых ученых «Перспектив Свободный-2015», посвященная 70-летию Великой Победы. – 2015. – С. 0-92.
41. Личаргин, Д. В. Разработка веб-приложения по генерации учебных заданий к тексту на естественном языке на основе порождаемых шаблонов. / В. В. Сотникова, А. А. Усова, И. В. Колбасина, Е. Д. Старовойт // Новосибирск: X Международная научно-практическая конференция: «Научные перспективы XXI века. Достижения и перспективы нового столетия». – 2015. - С. 57-60.
42. Домрачев, В. Г. О классификации компьютерных образовательных информационных технологий / В.Г. Домрачев, И.В. Ретинская // Информационные технологии, №2, 1996. – С. 10-14.
43. Moodle.ru [сайт]. – Режим доступа: <http://moodle.ru/course/view.php?id=18>

ПРИЛОЖЕНИЕ А

Пример базы комбинированных слов в таблице Access

Таблица1		
№	Слово	Перевод
1	physics	физика
2	mechanics	механика
3	chemistry	химия
4	physical_culture	физическая_культура
5	philosophy	философия
6	history	история
7	geometry	геометрия
8	computing	информатика
9	logic	логика
10	mathematics	математика
11	maths	математика
12	biology	биология
13	zoology	зоология
14	botany	ботаника
15	linguistics	лингвистика
16	rhetoric	риторика
17	semantics	семантика
18	grammar	грамматика
19	phonetics	фонетика
20	punctuation	пунктуация
21	orthography	орфография
22	spelling	орфография
23	language	язык
24	English	английский
25	American_English	американский_английский
26	British_English	британский_английский
27	Russian	русский
28	Ukrainian	украинский
29	German	немецкий
30	French	французский
31	Italian	итальянский
32	Spanish	испанский

Таблица 1		
№	Слово	Перевод
33	Greek	греческий
34	Latin	латинский
35	- Combinonims -	- Combinonims -
36	cool	крутой
37	great	великолепный
38	super	супер
39	superb	супер-пупер
40	cool	крутой
41	great	великолепный
42	cool	крутой
43	great	великолепный
44	- Combinonims-	- Combinonims -
45	handsome	красивый
46	cute	крутой
47	cool	крутой
48	unattractive	непривлекательный
49	handsome	красивый
50	cute	крутой
51	cool	крутой
52	unattractive	непривлекательный
53	- Combinonims	- Combinonims -
54	ugly	уродливый
55	hefty	здоровенный
56	muscular	мускулистый
57	brawny	мускулистый
58	strong	сильный
59	weak	слабый
60	healthy	здоровый
61	ill	больной
62	diseased	заболевший
63	sick	больной
64	lame	хромой
65	better	лучше
66	thick	толстый
67	fat	жирный
68	plump	полный

Таблица 1		
№	Слово	Перевод
69	thin	тонкий
70	- Combinonims	- Combinonims -
71	slim	стройный
72	tall	высокий
73	short	низкий
74	hairy	волосатый
75	thick-haired	густоволосый
76	bushy	растрепанный
77	shaggy	лохматый
78	- Combinonims -	- Combinonims -
79	sweet	очаровательный
80	beautiful	красивый
81	good-looking	симпатичный
82	lovely	миловидный
83	pretty	симпатичный
84	plain	не_отличающийся_красотой
85	fine	хорошо_себя_чувствующий
86	tall	высокий
87	nice	милый
88	attractive	привлекательный

ПРИЛОЖЕНИЕ Б

Пример базы сем

!!!! 01[START]01 02[START]02 03[СТАРТ]03 04[]04 05[]05 06[ART]06 07[_]07
08[_]08 09[_]09 10[_]10 11[_]11 12[03]12

!!!! 01[science]01 02[сАйэнс]02 03[наука]03 04[die Wissenschaft]04 05[la
science]05 06[la ciencia]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[00]12

!!!! 01[physics]01 02[фИзикс]02 03[физика]03 04[Physik]04 05[la physique]05
06[la fñsica]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[03]12

!!!! 01[mechanics]01 02[микАЕникс]02 03[механика]03 04[der Mechaniker]04
05[le m'ecanicien]05 06[месбñicos]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11
12[13]12

!!!! 01[chemistry]01 02[кЕмистри]02 03[химия]03 04[die Chemie]04 05[la
chimie]05 06[la qñmica]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[03]12

!!!! 01[physical..culture]01 02[фИзикэлкАлчэ]02 03[физическаякультура]03
04[der Sport]04 05[la culture physique]05 06[la cultura fñsica]06 07[x]07 08[i]08
09[y]09 10[y]10 11[e0]11 12[03]12

!!!! 01[philosophy]01 02[филОсэфи]02 03[философия]03 04[die Philosophie]04
05[la philosophie]05 06[la filosofна]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11
12[03]12

!!!! 01[history]01 02[хИстэри]02 03[история]03 04[die Geschichte]04
05[l'histoire]05 06[la historia]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[03]12

!!!! 01[geometry]01 02[джиОмэтри]02 03[геометрия]03 04[die Geometrie]04
05[la g'eom'etrie]05 06[la геометрна]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11
12[03]12

!!!! 01[computing]01 02[кэмпйУ:тиН]02 03[информатика]03 04[die Informatik]04
05[l'informatique]05 06[computar]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11
12[03]12

!!!! 01[logic]01 02[лОджик]02 03[логика]03 04[die Logik]04 05[la logique]05
06[la lygica]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[13]12

!!!! 01[mathematics]01 02[мАЕСэМАЕтикс]02 03[математика]03 04[die
Mathematik]04 05[le math'ematicien]05 06[las matemбticas]06 07[x]07 08[i]08
09[y]09 10[y]10 11[e0]11 12[03]12

!!!! 01[maths]01 02[мАЕСс]02 03[математика]03 04[]04 05[]05 06[]06 07[x]07
08[i]08 09[y]09 10[y]10 11[e0]11 12[13]12

!!!! 01[biology]01 02[байОлэджи]02 03[биология]03 04[die Biologie]04 05[la
biologie]05 06[la biologна]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[03]12

!!!! 01[zoology]01 02[зэОлэджи]02 03[зоология]03 04[die Zoologie]04 05[la
zoologie]05 06[la zoologна]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[03]12

!!!! 01[botany]01 02[бОтэни]02 03[ботаника]03 04[die Botanik]04 05[la
botanique]05 06[la botбnica]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[13]12

!!!! 01[linguistics]01 02[лингВИстикс]02 03[лингвистика]03 04[Linguistik]04
05[la linguistique]05 06[la lingбñstica]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11
12[03]12

!!!! 01[rhetoric]01 02[рЕСэрик]02 03[риторика]03 04[die Rhetorik]04 05[la rh'etorique]05 06[la retyrica]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[13]12

!!!! 01[semantics]01 02[семАЕнтикс]02 03[семантика]03 04[Semantik]04 05[la s'emantique]05 06[semбнтико]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[03]12

!!!! 01[grammar]01 02[грАЕмэ]02 03[грамматика]03 04[die Grammatik]04 05[la grammaire]05 06[la gramбtica]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[02]12

!!!! 01[phonetics]01 02[фоунЕтикс]02 03[фонетика]03 04[die Phonetik]04 05[la phon'etique]05 06[fonйtico]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[02]12

!!!! 01[punctuation]01 02[панктыЕйшн]02 03[пунктуация]03 04[die Interpunktion]04 05[la ponctuation]05 06[la puntuaсиун]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[03]12

!!!! 01[orthography]01 02[о:СОгрэфи]02 03[орфография]03 04[die Orthografie]04 05[l'orthographe]05 06[la ortografна]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[03]12

!!!! 01[spelling]01 02[спЕлиН]02 03[орфография]03 04[die Orthografie]04 05[l'orthographe]05 06[el deletreo]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[02]12

!!!! 01[language]01 02[лАЕнгВидж]02 03[язык]03 04[die Sprache(Zunge)]04 05[la langue]05 06[el idioma]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[02]12

!!!! 01[English]01 02[И:нглиш]02 03[английский]03 04[die englischeSprache(englische)]04 05[l'anglais]05 06[El inglйs]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[02]12

!!!! 01[American English]01 02[эмАЕрикэнИ:нглиш]02 03[американский английский]03 04[die amerikanischeenglischeSprache(englische)]04 05[l'anglaisam'ericain]05 06[El inglйs americano]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[03]12

!!!! 01[British English]01 02[брИ:тишИ:нглиш]02 03[британскийанглийский]03 04[die britischeenglischeSprache(englische)]04 05[l'anglaisbritannique]05 06[El inglйs inglйs]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[03]12

!!!! 01[Russian]01 02[рАшн]02 03[русский]03 04[russisch]04 05[russe]05 06[El ruso]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[02]12

!!!! 01[Ukrainian]01 02[йукрЕйниэн]02 03[украинский]03 04[ukrainisch]04 05[d'Ukraine]05 06[Ukrainian]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[14]12

!!!! 01[German]01 02[джЭ:мэн]02 03[немецкий]03 04[deutsch]04 05[allemand]05 06[El alemбn]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[02]12

!!!! 01[French]01 02[фрАЕнч]02 03[французский]03 04[franzosisch]04 05[franc,ais]05 06[El francйs]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[02]12

!!!! 01[Italian]01 02[итАЕлиэн]02 03[итальянский]03 04[italienisch]04 05[italien]05 06[El italiano]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[02]12

!!!! 01[Spanish]01 02[спАЕниш]02 03[испанский]03 04[spanisch]04 05[espagnol]05 06[El espacol]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[02]12

!!!! 01[Greek]01 02[грИ:к]02 03[греческий]03 04[griechisch]04 05[grec]05 06[El griego]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[03]12

!!!! 01[Latin]01 02[ЛАЕтин]02 03[латинский]03 04[lateinisch]04 05[latin]05 06[El latнn]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[03]12

!!!! 01[art]01 02[А:т]02 03[искусство]03 04[die Kunst]04 05[l'art]05 06[el arte]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[00]12

!!!! 01[music]01 02[мйУ:зик]02 03[музыка]03 04[die Musik]04 05[la musique]05 06[la тьsica]06 07[x]07 08[i]08 09[y]09 10[y]10 11[e0]11 12[00]12

!!!! 01[national anthem]01 02[наЕшэнэлАЕнСэм]02 03[национальныйгимн]03 04[die nationaleHymne]04 05[l'hymne national]05 06[himnonacional]06 07[x]07 08[i]08 09[t]09 10[_]10 11[e0]11 12[03]12

!!!! 01[medley]01 02[мЕдли]02 03[попурри]03 04[das Potpourri]04 05[le pot - pourri]05 06[medley]06 07[x]07 08[i]08 09[t]09 10[_]10 11[e0]11 12[13]12

!!!! 01[picture]01 02[пИкчэ]02 03[картинка]03 04[das Bild]04 05[картинка]05 06[el retrato]06 07[x]07 08[i]08 09[k]09 10[_]10 11[e0]11 12[00]12

!!!! 01[vision]01 02[вИжн]02 03[изображение]03 04[die Darstellung]04 05[la repr'esentation]05 06[la visiyn]06 07[x]07 08[i]08 09[k]09 10[_]10 11[e0]11 12[02]12

!!!! 01[view]01 02[вйУ:]02 03[вид]03 04[die Art(Gestalt)]04 05[l'aspect]05 06[el panorama]06 07[x]07 08[i]08 09[k]09 10[_]10 11[e0]11 12[02]12

!!!! 01[background]01 02[бАЕкграунд]02 03[заднийплан]03 04[der Hintergrund]04 05[le fond]05 06[el fondo]06 07[x]07 08[i]08 09[k]09 10[_]10 11[e0]11 12[02]12

!!!! 01[foreground]01 02[фО:граунд]02 03[переднийплан]03 04[der Vordergrund]04 05[le premier plan]05 06[foreground]06 07[x]07 08[i]08 09[k]09 10[_]10 11[e0]11 12[02]12

!!!! 01[image]01 02[Имидж]02 03[образ]03 04[die Weise(Gestalt)]04 05[l'image]05 06[la imagen]06 07[x]07 08[i]08 09[k]09 10[_]10 11[e0]11 12[02]12

!!!! 01[scene]01 02[сИ:н]02 03[сцена]03 04[die Szene(Buhne)]04 05[la sce`ne]05 06[la escena]06 07[x]07 08[i]08 09[k]09 10[_]10 11[e0]11 12[02]12

!!!! 01[sight]01 02[сАйт]02 03[вид]03 04[die Art(Gestalt)]04 05[l'aspect]05 06[la vista]06 07[x]07 08[i]08 09[k]09 10[_]10 11[e0]11 12[02]12

!!!! 01[speech]01 02[спИ:ч]02 03[речь]03 04[die Rede]04 05[les paroles]05 06[el habla]06 07[x]07 08[i]08 09[raai]09 10[_]10 11[e0]11 12[02]12

ПРИЛОЖЕНИЕ В

Примеры сгенерированных подстановочных таблиц

Таблица В.1 – Пример подстановочной таблицы OpenAndSpeak On – The bag is yellow – topic

Слово 1	Слово 2	Слово 3	Слово 4
that	package	is	coffee
the	shelf	was	brown
this	closet	is	peach
that	box	is	apricot
the	basket	will_be	tangerine
this	case	can_be	orange

Таблица В.2 – Пример подстановочной таблицы OpenAndSpeak On – I need this work – topic

Слово 1	Слово 2	Слово 3	Слово 4
i	love	that	time-work
i	love	the	job
we	like	this	profession
we	need	that	labour
	wish	the	drudgery
	want	this	sinecure

Таблица В.3 – Пример подстановочной таблицы OpenAndSpeak On – Her hair is splendid – topic

Слово 1	Слово 2	Слово 3	Слово 4
your	artificial_limb	is	wonderful
our	prosthetic_device	was	excellent
her	tuft	is	great
their	forelock	is	good
	lock	will_be	ideal
	curl	can_be	perfect

Таблица В.4 – Пример подстановочной таблицы OpenAndSpeak On – Don't worry be happy – topic

Слово 1	Слово 2	Слово 3	Слово 4
would	spit_on	appear_to_be	content
will	be_indifferent_to	become	satisfied
must	like	get	
should	love	happen_to_be	
shall	dislike		
might	hate		

Таблица В.5 – Пример подстановочной таблицы OpenAndSpeakOn – I need water – topic

Слово 1	Слово 2	Слово 3
i	love	the_Pacific
i	love	Atlantic_Ocean
we	like	Arctic_Ocean
we	need	Indian_Ocean
	wish	sea
	want	gulf

Таблица В.6 – Пример подстановочной таблицы OpenAndSpeak On – the user is in bed – topic

Слово 1	Слово 2	Слово 3	Слово 4	Слово 5
that	client	is	onthis	bedding
the	dummy	was		sheet
this	manufacturer	is		cushion
that	creator	is		pillow
the	maker	will_be		parquet
this	producer	can_be		parquetry

Таблица В.7 – Пример подстановочной таблицы OpenAndSpeak On – the user is strange – topic

Слово 1	Слово 2	Слово 3	Слово 4
that	client	is	worthless
the	dummy	was	useless
this	manufacturer	is	abominable
that	creator	is	disgusting
the	maker	will_be	hideous
this	producer	can_be	terrible

Таблица В.8 – Пример подстановочной таблицы OpenAndSpeak On – the god is mine – topic

Слово 1	Слово 2	Слово 3	Слово 4
that	spook	is	ours
the	spectre	was	yours
this	phantom	is	theirs
that	ghost	is	
the	soul	will_be	
this	spirit	can_be	