бирать данные по энергопотреблению. Опрос всех объектов осуществляется одновременно и занимает не более 30 сек. Реальная скорость передачи данных составляет 12 кбит/с. За счет своевременного отключения нагрузок и постоянного контроля потребляемого тока разработанная система позволяет снизить расходы на электроэнергию на 30%.

СПИСОК ЛИТЕРАТУРЫ

1. Калабанов, С.А. Система защиты и управления для бытовой электросети на основе технологии SMART GRID / С.А. Калабанов, А.В. Карпов, Р.С. Кириллов, Р.И. Шагиев – Ползуновский вестник, № 3-1, 2011. – С. 203-207.
2. Sanders, G. GPRS Networks // G. Sanders, L. Thorens, M. Reisky, O. Rulik, S. Deylitz. – England: Wiiley, 2003 – 294 p.
3. SAM3S Series Datashet [Electronic resource] / Atmel Corporation – Mode of Access: http://www.atmel.com/Images/Atmel_6500_32-bit-Cortex-M3-Microcontroller_SAM3S_Datasheet.pdf
4. FreeRTOS Reference Manual - API Functions and Configuration Options [Electronic resource] / Mode of Access: http://www.freertos.org/
5. MG2639 AT Command Manual [Electronic resource] / ZTE Corporation – Mode of Access: http://www.wless.ru/files/GSM/ZTE/2G/MG2639_AT_Command_Manual.pdf

*Аспирант **Р.И. Шагиев** – r3ntil@gmail.com; д.ф.м.н., проф. **А.В. Карпов** – Arkadi.Karpov@ksu.ru; к.ф.м.н. **С.А. Калабанов** – kazansergei@mail.ru; магистрант **Р.Р. Фатыхов** – ruslancomb@gmail.com - Казанский федеральный университет, Институт Физики, кафедра радиофизики.*

УДК: 004.651.519.765.008

РАЗРАБОТКА, НАПОЛНЕНИЕ И ОБСЛУЖИВАНИЕ БАЗЫ ДАННЫХ ДЛЯ КУЛЬТУРОМЕТРИЧЕСКОГО ИССЛЕДОВАНИЯ ХУДОЖЕСТВЕННЫХ ТЕКСТОВ

О. В. Головань, А.В. Ишков

В статье описана специализированная частотная база данных (БД), предназначенная для исследования статистически значимых частотных зависимостей в художественных текстах и их корпусах. По гипотезе, предложенной авторами, слова, имеющие самую высокую частоту в БД, имеют и особую, культурную значимость. Отыскание таких высокочастотных, культурнозначимых слов (лексем) и составляет задачу культурометрии, осуществляемой на основе материала, зафиксированного в языке и его текстах.

Ключевые слова: база данных, частота и ранг слов, информатика, художественный текст, культурология, культурометрия.

В настоящее время количественные методы все чаще используются даже в таких «прикладных» и далеких от математики областях человеческой деятельности, как лингвистика, культурология, искусствоведение и даже музыка [1, 2].Одним из таких новых, перспективных направлений междисциплинарных исследований и является *культурометрия* - раздел науки на стыке культуры, прикладной математики и статистики, который оценивает уровень культурного достояния в художественном творчестве, науке и искусстве на основе статистически значимых рейтингов [3].

Мы полагаем, что одним из наиболее статистически значимых показателей развития культуры является зафиксированная в ее художественном языке закономерность использования наиболее смысло-, культурно-, исторически- и культурнозначимых простейших элементов, имеющих единую, общечеловеческую ценность и значение [4].

Интеллектуальные системы исследования текста позволяют, анализируя смысловое и словарное содержание текста, представленность в нем отдельных слов и словосочетаний, частоту их употребления и пр., устанавливать основную идею или скрытый смысл, вложенный автором в произведение. Эксперименты такого рода часто применяются в прикладной социологии, психологии, педагогике, культурологии, криптографии [5].

В основе предлагаемого нами нового метода культурометрического исследования лежит систематический разбор обширных массивов художественных текстов, принадлежащих представителям (носителям) определенной культуры, на предмет поиска объективных закономерностей между частотой и рангом их отдельных слов по закону Ципфа [6]. Тогда, в соответствии с экспериментально обнаруженной ранее А. Вежбицкой закономерностью высокой частотности наиболее культурнозначи-

мых слов [7], по мере увеличения объема корпуса текстов, удастся зафиксировать такие слова. Но для этого потребуется создание и наполнение специально организованной базы данных (**БД**), которая позволит ставить в соответствие каждому отдельному слову его уникальный номер (ID), условно связанный с темой, словарем или конкретным автором, чтобы избежать повтора.

Целью настоящего исследования являлась разработка специализированной частотной базы данных для культурометрического исследования художественных текстов, а также ее наполнение, тестирование и оптимизация обслуживания.

Постановка задачи

Для оптимизации работы такой БД слова и словосочетания исходного художественного текста (или целых корпусов текстов) нужно будет разбивать на группы по частоте их употребления, признакам соответствия одному гнезду, однокоренности и др., выстраивая соответствующие зависимости. Таким образом, частотный анализ является первой ступенью интеллектуальной обработки и исследования текста. Проведение культурометрического исследования в этом случае производится путем составления частотного словаря текста, корпуса и(или)языка и определения ранга и частоты отдельных высокочастотных (культурнозначимых) слов.

Наиболее просто такая система может быть организована по принципу раздельного размещения алгоритма обработки текста, БД и обслуживающей программы с организацией гибкого взаимодействия между компонентами, например, посредством SQL-сервера.

Алгоритмическое и программное решение

На основе возможностей SQL-сервера FireBird v. 1.0 была разработана БД для частотного исследования художественных текстов и обслуживающая компьютерная программа, предназначенные для осуществления практических культурометрических исследований на языковом материале [8, 9].

Программа «Фрактальная размерность языка (LangFracDim)» позволяет проводить группировку слов в базе по определенным признакам (темам, алфавиту и т.п.) и устанавливать корреляционные связи между отдельными словами по частотным признакам [8]. Определение корреляций между квантитативными (частотными) характеристиками текстов основывается на установлении зависимости между частотой и рангом слова, параметры которой, после линеаризации, определяются методом наименьших квадратов. Работа основного алгоритма программы организована по блочному принципу путем взаимосвязанного выполнения четырех основных алгоритмов: загрузки текста из файла *.txt, разбора текста, и переноса слова или списка слов в БД (рисунок 1).

Рисунок 1 – Блок-схема основного алгоритма БД LangFracDimDB

Загрузка текста в буфер программы осуществляется стандартными средствами, предоставляемыми операционной системой Windows, для хранения данных в БД используется кодировка шрифта WIN1251, а исходный текст перед работой программы должен быть в текстовом (*.txt) формате.

После заполнения буфера запускаются алгоритмы разбора текста, позволяющие из общего набора символов кодировки WIN1251 выделить отдельные слова и сформировать на выходе список всех слов текста. Полученный список слов или отдельное слово затем переносится в БД с помощью соответствующих алгоритмов переноса и происходит инициализация переменных БД*i, l* и *w*, обозначающих слово, список слов и пустую строку, и далее запускается циклическая процедура проверки каждой ячейки массива текста на условия соответствия элемента букве, слову, пустой строке или пробелу.

Работа циклов завершается, как только все элементы текста, воспринимаемые машиной как отдельные слова (собственно слова как части речи, междометия, предлоги и пр.) не будут сформированы в новый список. На этом этапе работы программы «LangFracDim», путем сравнения кодировок элементов массива с кодами символов («,» - запятая), («.» - точка), («–» - дефис), других знаков препинания (!, «, ‘, “, ?, (, :, ..., и др.), происходит их исключение из списка слов. Пример блок-схемы алгоритма синтаксического разбора текста приведена на рисунке 2.

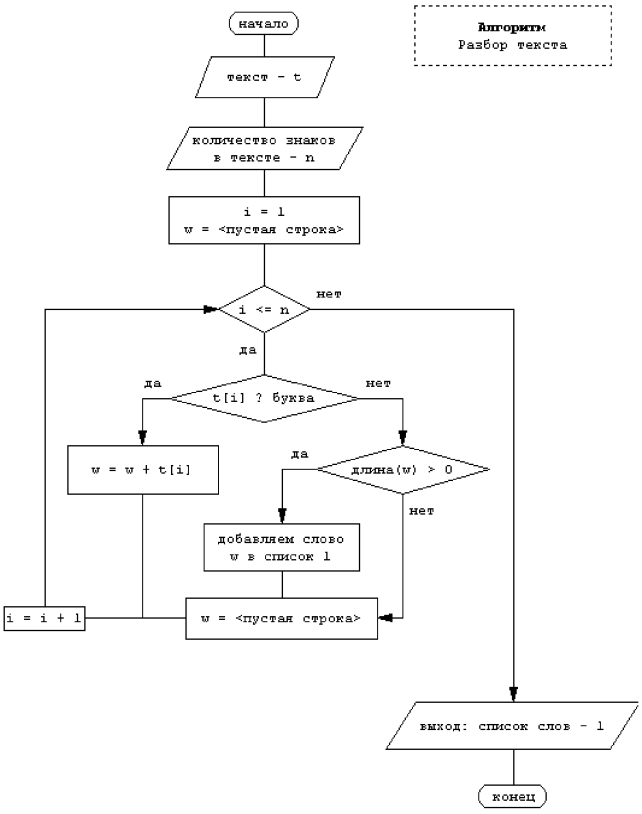

Рисунок 2 – Блок-схема алгоритма разбора текста в список слов программой LangFracDim

Алгоритм переноса слова в БД (рисунок 3), идентифицирует каждое новое слово по трем признакам: собственно самому слову как набору символов (*w*), соответствию слова определенной, выбранной оператором теме (*t*), и соответствию слова определенному словарю (*d*).

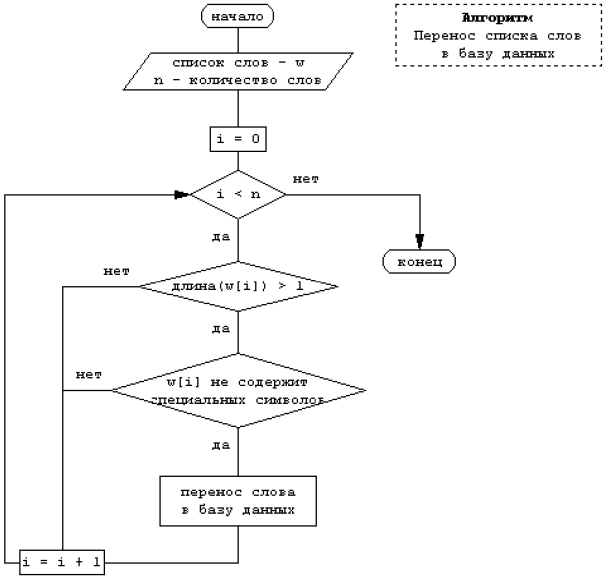

Рисунок 3 – Блок-схема алгоритма переноса списка слов в БД

Организация взаимодействия работы программы LangFracDimи с БД позволяет ставить в соответствие каждому слову его уникальный номер, связанный с темой и словарем, что позволяет избегать повтора одинаковых слов и ускорять извлечение слова из БД SQL-сервером. После разбора текста осуществляется окончательный перенос полученного списка слов в БД с помощью соответствующего алгоритма (рисунок 4).

Как видно из рисунка 4, одновременно с занесением слова в БД программой производится пересчет частоты встречаемости слова с учетом текущего содержимого всех словарей и тем БД.

Физически база данных «Фрактальная размерность языка» представляет собой один файл с оригинальным именем «WORDFRACDIM.GDB», который создан и функционирует под управление SQL-сервера Firebird 1.0. Отдельные таблицы БД связаны между в соответствии с рисунком 5. БД также содержит хранимые процедуры и функции, задаваемые пользователем.

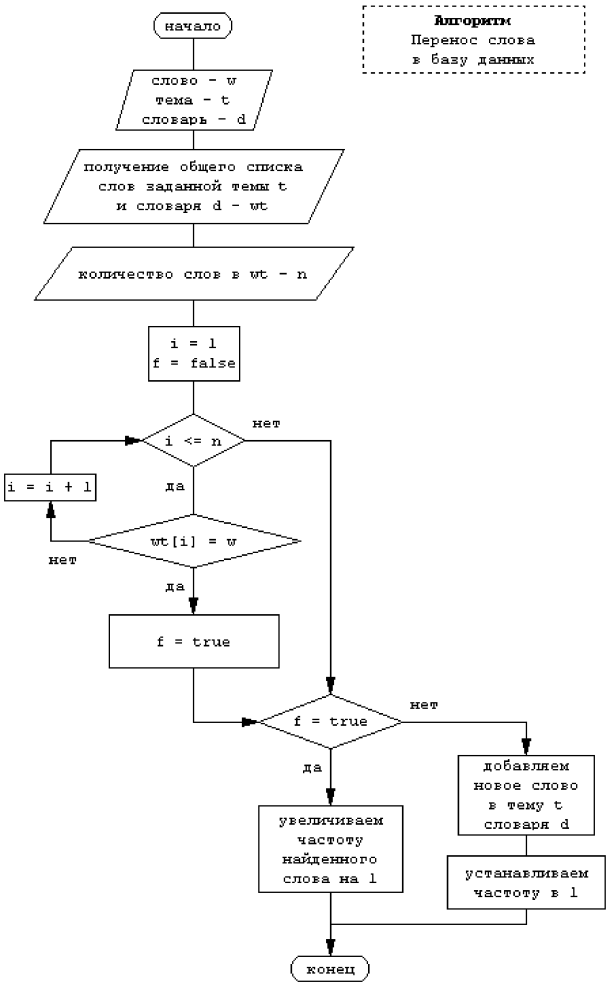

Рисунок 4 – Блок-схема окончательного алгоритма переноса слова в БД

В состав базы входят следующие таблицы: T_DICTIONARY – таблица словарей; T_WORD – таблица слов (в эту таблицу добавляются слова алгоритмом «Перенос слова в базу данных»); T_THEME – таблица тем; T_TEXT – таблица текстов; T_LINK – таблица частот слов в темах (эта таблица модифицируется алгоритмом «Перенос слова в базу данных»). Уникальность первичных ключей таблиц обеспечивается генераторами, входящими в структуру базы.

Для хранения данных БД использует следующие домены. D_RANG – для описания полей, хранящих ранги слов в частотном словаре; D_STRING – для описания коротких строк (до 50 символов), D_STRING_LONG – для описания длинных строк (до 250 символов), и D_TEXT – для полей, хранящих текстовую информацию большого переменного объема.

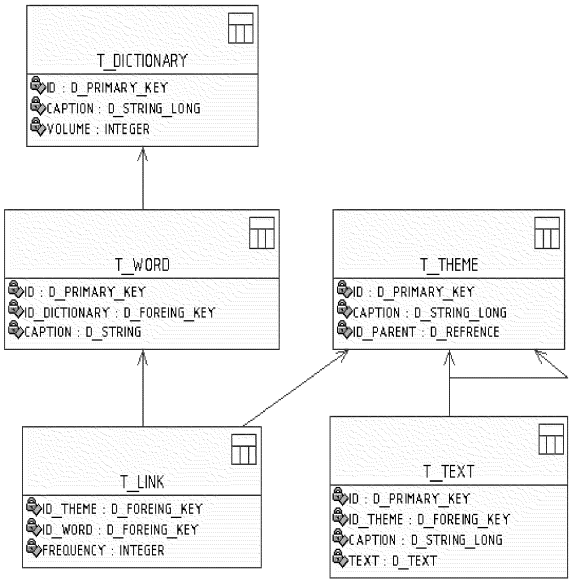

Рисунок 5. Структурная схема БД

В состав БД входят также программно-оформленные процедуры и функции, осуществляющие внутреннее функционирование базы. Примеры конкретного наполнения таблиц и интерфейса БД приведены на рисунках 6, 7.

Из сервисных возможностей в программе «LangFracDim» предусмотрены: экспорт базы данных в виде частотного словаря или списка слов, упорядоченных по другому признаку, в текстовый файл, вычисление частоты и ранга слов, представление зависимости между квантитативными характеристиками в виде графика в двойных логарифмических координатах, установление параметров искомой зависимости.

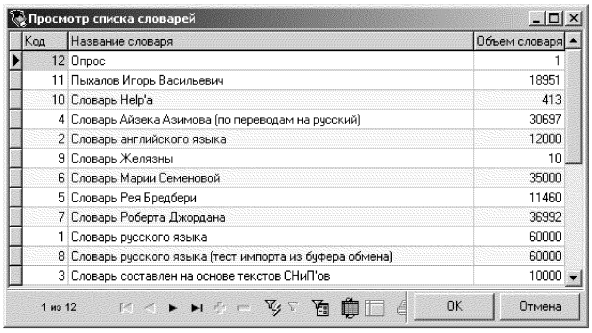

Рисунок 6 – Пример наполнения таблицы T_DICTIONARY

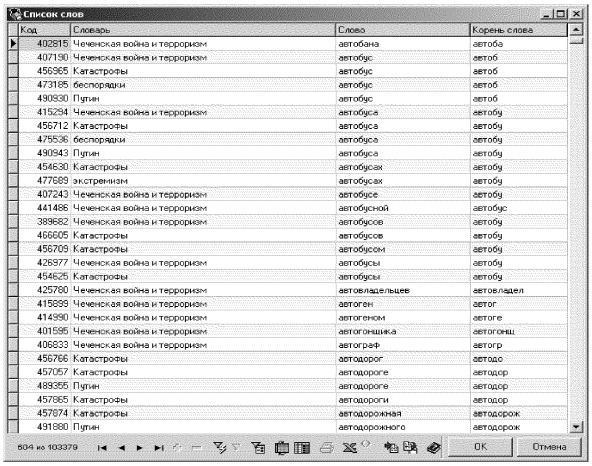

Рисунок 7 – Пример наполнения таблицы T_WORD

В качестве одновременно описательного и исследовательского критерия при лингвокультурологическом описании текста, корпусов и языка в целом, на основе которого может быть сделана выборка составляющих концепта «трагическое», нами будет использоваться не только частота слов, но и характеристик самой функциональной зависимости между частотой и рангом слов, выражаемой известным степенным законом Ципфа [6]:

$$Ч = k \times P^{\alpha}, \qquad (1)$$

где $Ч$ - частота встречаемости слова в тексте; k - коэффициент пропорциональности; P - ранг слова; α - степень развитости и наполняемости текста различными лексическими единицами (для современных языков близок к 1).

Показатель степени в (1) отражает, насколько сильно данная категория включена в тезаурус респондента, а, значит, и в его внутренний языковой, смысловой и культурный мир. Значения же коэффициента пропорциональности лишь показывают, как много слов из предложенных (прочитанных) текстов опрашиваемый связывает с конкретной темой.

Для удобства последующей работы интерфейс обслуживающей БД программы

«Фрактальная размерность языка» позволяет представлять данные как в различные табличные формы, связывающие статистические показатели языковых единиц (по принадлежности к одному или нескольким текстам, темам, словарям, однокоренности и пр.), так и в графическом виде (частотограммы) с их характеристиками, подписями слов, указанием на осях реперных частот, рангов (рисунок 8).

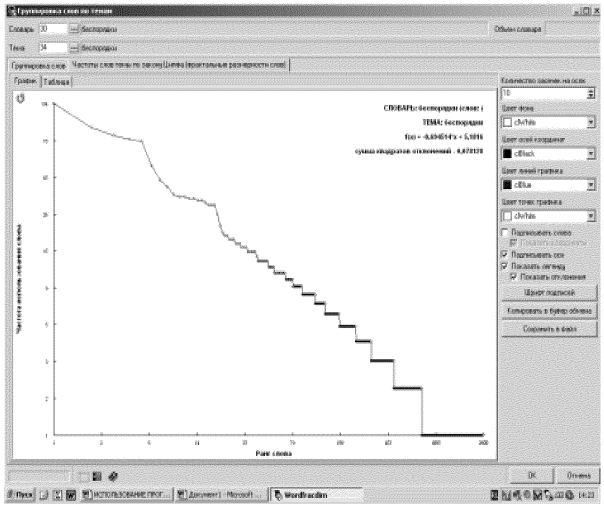

Рисунок 8 – Интерфейс программы LangFracDim с графической формой представления данных

Таким образом, использование комплекса программ «Фрактальная размерность языка», «Концепт-анализ» [10] и специализированной БД, формируемой средствами СУБД FireBird, позволяет не только увеличить верификацию и оперативность лингвокультурологических исследований, но и выявлять содержательную сторону смыслового компонента корпуса текстов, отдельного текста или концепта, численно оценивая уровень понимания читателем предложенного художественного текста

Выводы

1. На основе разработанных алгоритмов разбора текстов на списки слов, отдельные слова, их переноса в БД и пересчета частоты встречаемости удается обнаруживать в текстах высокочастотные, культурнозначимые слова.
2. Разработанная БД и программа LangFracDim позволяют определять частоту и ранг слов для отдельного текста или корпуса, сгруппированного по произвольному признаку.

СПИСОК ЛИТЕРАТУРЫ

1. Петров, В.М. Количественные методы в искусствознании / В.М. Петров -Вып. 1. -М.: Смысл, 2000.
2. Капица, С.П. Синергетика и прогнозы будущего / С.П. Капица, С.П. Курдюмов, Г.Г. Малинецкий -М.: Наука, 1997.
3. Быстров, М.В. «Культурометрия» или «квантитативная культурология»? / М.В. Быстров // Вопросы культурологии, 2009. №11. -С. 11-14.
4. Андреев, Н.Д. Статистико-комбинаторные методы в теоретическом и прикладном языковедении / Н.Д. Андреев -С-Пб., 1997.
5. Ахо А. Теория синтаксического анализа, перевода и компиляции,Т.1./ А. Ахо, Д. Ульман. -М.: Мир, 1978.
6. Zipf, G.K. The psycho-biology of language / G.K.Zipf -Boston, 1935.
7. Вежбицкая, А. Язык. Культура. Познание / А. Вежбицкая -М.: 1997.
8. Головань, О.В. Фрактальная размерность языка. Программа для ЭВМ / Св-во № 2005620308 (RU) от 28.11.2005 // Бюлл. № 4. 2005.
9. Головань, О.В. Фрактальная размерность языка БД. База данных/ Св-во рег. пр. для ЭВМ № 2005610982 (RU) //Бюлл. № 2. 2005.

к.ф.н. ***Головань О.В.,*** *доцент АлтГТУ; д.т.н., профессор* ***Ишков А.В.,*** *профессор, , olg168@rambler.ru - АГАУ*

УДК: 004.061

ИНФОРМАТИЗАЦИЯ ИСПЫТАТЕЛЬНОЙ ЛАБОРАТОРИИ ООО «БИЙСКОГО ЗАВОДА СТЕКЛОПЛАСТИКОВ»

А. Ю. Хорохордин, М. Ю. Локтев, В. А. Абанин

Рассмотрены вопросы построения информационной системы испытательной лаборатории завода с применением современных средств автоматизации хранения, обработки и предоставления данных.

Ключевые слова: информационная система, база данных, испытательная лаборатория, LabVIEW, MySQL.

Введение

Объективная количественная информация о состоянии параметров качества продукции, а также о параметрах производственного процесса на всех его стадиях служит основой для принятия эффективных управ-