

Г.А. Черкасова, Анна Филиппович

ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ ПОДГОТОВКИ ТЕКСТА САР ДЛЯ СОЗДАНИЯ ЕГО ЛИНГВИСТИЧЕСКОЙ БАЗЫ ДАННЫХ

Основной компонентой Интегрированной инструментальной информационно-программной среды для автоматизации исследований САР является лингвистическая база данных. Технология ее создания в проекте рассматривается как последовательное решение четырех задач в течение двух лет: 1. Разработка лингвистической базы данных. 2. Разработка программного комплекса автоматизированного ввода текста словарных статей в базу данных. 3. Подготовка и ввод в базы данных текста 1, 2 и 3 томов САР. 4. Подготовка и ввод в базы данных текста 4, 5 и 6 томов САР.

Задача разработки базы данных традиционно сводится к разработке инфологической и даталогической моделей, которые обеспечивают эффективность хранения и использования данных. Разработка лингвистической базы данных САР связана с двумя его основными особенностями: первая состоит в том, что для организации лексического материала в нем используется гнездовой способ и вторая — в том, что исходным материалом является оригинал-макет его переиздания, который был сверстан с использованием программы Page Maker, т.е. исходными данными являются компьютерные файлы в формате *.р65.

Особенность гнездовой организации находит свое выражение в том, что, рассматривая какую-либо словарную статью (гнездо) с целью ее формального описания (это является условием представления ее в форме элемента базы данных), предварительно нельзя быть уверенным том, что ее структура эквивалентна хотя бы одной из уже известных (заранее заданных). Количество гнезд в САР конечно, и в итоге последовательного анализа материала Словаря для каждого из них можно установить конкретную структуру, и даже обнаружить некоторые типовые структуры, использованные его составителями. Такие типовые структуры описаны в исследованиях А.Н.Тихонова [1] и М.И.Чернышевой [2]. Основными формальными признаками, позволяющими выявить структуру конкретной словарной статьи, являются шрифтовые выделения, разделители (знаки препинания), пометы (сокращения), а также признаки пространственного расположения материала на страницах Словаря (колоннотитулы, отступы и др.).

Особенность компьютерного представления исходного текста САР

в формате Page Maker требует дополнительных операций их преобразования в иной текстовый формат для последующего структурного и формального преобразований в формат данных конкретной СУБД.

Приведенные особенности существующих материалов САР после завершения работ по его переизданию оказали влияние на содержание информационной технологии разработки базы данных. Технология была разделена на два этапа: *подготовительный* и *конструкторский*.

Целью подготовительного этапа является получение материалов Словаря в виде удобном для последующего автоматизированного ввода в сконструированную базу данных. На этом этапе решаются две крупные задачи: первая, формирование данных Словаря в виде набора (фрагментов) текстовых файлов в формате MS Word; вторая, подготовка этих фрагментов для последующего автоматизированного ввода в сконструированную базу данных.

На втором конструкторском этапе должна быть создана лингвистическая база САР. Здесь выделяются четыре задачи: первая, конструирование (разработка) структур баз данных для фрагментов САР; вторая, разработка технологического процесса заполнения базы данных и автоматизированный ввод словарных статей; третья, исследование эффективности технологического процесса заполнения базы данных; четвертая, интеграция технологий и баз данных фрагментов словарных материалов.

Существенное влияние на любую информационную технологию оказывают, помимо условий ее реализации и используемого инструментария, временных и ресурсных ограничений, субъектные особенности. К ним, прежде всего, относятся количественные ограничения на возможности привлечения к реализации технологических процедур и операций людей, их квалификация, мотивация качественного исполнения работ и др. В связи с этим информационная технология создания лингвистической базы данных САР изначально рассматривалась как технология обработки данных большим количеством исполнителей. В качестве таких исполнителей были выбраны студенты¹, одной из конкретных целей

¹ В качестве исполнителей работ в информационной технологии создания лингвистической базы данных САР были привлечены студенты 5 курса МГТУ им.Н.Э.Баумана, изучающие на кафедре «Систем обработки информации и управления (ИУ–5)» дисциплины «Лингвистическое обеспечение АСОИУ» и «Семиотика информационных технологий». Выполненные ими преобразования словарных данных и разработки являлись частью заданий курсовой работы 10 семестра 2005–2006 уч.года

которых являлось «приобретение навыков разработки лингвистических баз данных».

Первая задача подготовительного этапа

Формирования данных Словаря в виде набора текстовых файлов в MS Word это не простое конвертирование формата PageMaker в формат *.rtf, а преобразование множества шрифтов, используемых при верстке в один, так как СУБД может использовать только один шрифт, да и шрифтовые выделения (курсив, полужир, капитель и др.) допускаются только в полях Метод, по которым невозможен поиск данных.

При подготовке оригинал-макета САР использовалась специально разработанная гарнитура Andrew Daskova и ее модификация, содержащая буквы с ударением (прямым и обратным) — Andrew Daskova Ud. Эти гарнитуры используются в три начертаниях: обычное, курсивное и жирное; и на ряду с современным русским алфавитом включают старославянские буквы Ѣ, Ѣ, Ѣ, Ѣ и ряд специальных знаков. Для примеров в статьях Словаря, содержащих церковную лексику, использовалась также гарнитура шрифта «Ochrid».

При подготовке текстовых файлов в MS Word были приняты следующие правила:

- а) текст САР должен использовать одну гарнитуру;
- б) обратное (тупое) ударение заменялось прямым;
- в) ударные гласные отображаются *самой гласной и следующим за ней апострофом*, т.е. *á* заменялась на *a'*;
- г) шрифт капитель задается меньшим кеглем и разрядкой;
- д) для сохранения книжного представления (по строкам) знак «разрыв строки» заменен на «|» (вертикальная линия);
- е) словарные статьи САР, расположенные в двух и более колонках, собирались вместе и помещались в колонку с наименьшим номером.

Итак, для формирования данных Словаря в виде набора текстовых файлов выполнялись следующие операции:

1. Исходные данные, сформированные в издательском формате в программе PageMaker, процедурой экспорта сохранялись в формате *.rtf. Отдельные фрагменты содержали 100 колонок или 50 полос кни-

— «Создание базы данных Словаря Академии Российской 1789–1794 гг. (Вариант 5)». Курсовую работу выполняли следующие студенты: Н.С. Петров, (группа ИУ5-10), М.Р.Маркус (группа ИУ5-104), К.Е.Иовков (группа ИУ5-104), Е.О.Жмакина (группа ИУ5-101), А.В.Дон (группа ИУ5-103), О.М.Антонок (группа ИУ5-103).

ги. Имя файла задавалось следующим форматом: Т_NNN-KKK.rtf, где Т — номер тома, NNN — номер начальной колонки в тексте словаря, KKK — номер конечной колонки;

2. Ручная проверка и корректировка «разрыва» последней статьи в файле. В случае необходимости перенос концов словарных статей из начала следующего файла.

3. В экспортированных файлах задавались следующие параметры страницы: левое поле — 2 см., правое — 10 см., что позволяет отбрасывать текст в формате одной колонки.

4. В файлах межбуквенный интервал в режиме *шрифт* задавался — *обычный*. Капитель (слова 6-го и 9-го кегля) выделяется *разрядкой* в 3 п.

5. Средствами Microsoft Word проводились следующие замены:

5.1. Гласные с обратным (тупым) ударением заменялись буквами с обычным ударением. Операция выполнялась с учетом регистра, а также старославянских букв «ять», «фита» и др.

5.2. Все гласные ударного шрифта переведены на безударные и к ним добавлен знак «'» (после них).

5.3. Проверка и замена знака *табуляция* и *двойных пробелов* на *один пробел*.

5.4. Знак «разрыв строки», фиксирующий в тексте словаря точное повторение строк по оригиналу 1789-1794 гг., заменялся на «|».

6. Последние абзацы каждой полосы, включающие несколько букв и имеющие стиль <kon_str> удалялись, так как их содержимое полностью повторялось в начале следующей полосы.

7. Ручная проверка и корректировка «разрыва» словарных статей по колонкам, т.е. соединялся текст переходящий из одной колонки в следующую.

8. Проверка наличия других шрифтов, кроме Andrew Daskova и их замена.

Вторая задача подготовительного этапа

К решению этой задачи были привлечены студенты. Каждый из студентов получил файл с отрезком САР (пример 1). Они должны были выполнить следующие два задания: 1) подготовить словарные статьи конкретного фрагмента САР к вводу в базу и 2) провести декомпозицию словарных статей на заданные компоненты. Ниже приведены требования к выполнению заданий.

Задание 1.

Подготовка словарных статей к вводу в базу данных состоит в их записи в таблицы Microsoft Word (2003). При подготовке нужно выполнить следующие действия:

Установить шрифты;

Выделить цветом элементы словарных статей: **желтым** — заголовочные слова, **зеленым** — грамматические характеристики и пометы, **бирюзовым** — цитаты, **красным** — шифр источника (пример 2);

Заполнить таблицу 1 «Заголовочные слова»: <№ столбца>, <№ словарной статьи>, <заголовочное слово>, <словарная статья полностью> (пример 3).

Заполнить таблицу 2: <№ столбца>, <№ словарной статьи>, <заголовочное слово>, <№ значения>, <значение> (пример 4).

Задание 2.

Декомпозиция словарных статей состоит в переносе ее элементов в таблицы 3 и 4, имеющие следующие структуры (примеры 5 и 6):

Таблица 3. «Грамматические характеристики заголовочных слов»

— <№ столбца>, <№ словарной статьи>, <**заголовочное слово**>, <варианты>, <**Часть речи**>, <**Род**>, <**Число**>, <**Вид**>, <**Помета**>.

Таблица 4. «Характеристики значений» —

<№ столбца>, <№ словарной статьи>, <**заголовочное слово**>, <№ значения>

<Толкование>, <Примеры>, <**Цитата**>, <**Источник**>.

Пример 1.

Исходный текст отрезка САР 1789–1794 гг..

2

ЗАБОБОНЫ, бонѣ. с. ж. множ. Бред-|ни, сказки по большей части основан-|ныя на суевѣрїи, и не достойныя |вниманїя умнаго челоуѣка. Сей ли |лутѣ сласенїя, яко помратенѣ забоб-|бонами не знаешь, что ты долженѣ |еси Богу, государю, отечеству? Теоф. |Проккоп.

Ты жиѣ ло тѣмѣ законамѣ,|Которы лисаѣ,|Смѣялся забабонамѣ. М. Лом.

Забабонщина, щины. с. ж. проспонар. |Рѣчи и рассказы пустыя, основанныя |на суевѣрїи. Какой забабонщинѣ ты |вѣришь!

ЗАБОТА, пы. с. ж. Попеченіе о чемѢ,|щаніе, спараніе соединенное съ без-|покойствомъ и печалію. *Домаш- нія|заботы сокрушили его здоровье. Весь|вѣкъ живу въ заботахъ.*

Межъ тѣмъ на Горизонтѣ заря восходитъ,|И смертныхъ бдѣнія, и тѣмы заботъ возводитъ.

В. Петр. Ен.

3

Заботливый, вая, вое. *Заботливъ*,|ва, во. прил. Спарашельный, попечи-|пельный; который о чемѢ, или о комѢ|печется, забошится. *Заботливая хо-|зяйка. Заботливъ о подлинныхъ.*

Заботливо. нар. Спарашельно, попечи-|пельно. *Этотъ хозяинъ заботливо|за всемъ смотритъ.*

Заботливость, спи. с. ж. Свойство по-|печительнаго, спарашельнаго, пща-|пельнаго челоѡка; попечитель-|ность. *Съ его заботливостію всякой|трудъ пре- одолѣть можно.*

Заботный, ная, ное. *Заботенъ*, пна, пно. |прил. Съ хло-|пошами соединенный. |*Заботная должность. Заботной день.*

Заботно. нар. Хлопотно, многодѣльно,|суешливо. *Всяко- му крестьянину въ|лѣтнюю пору заботно.*

Заботу, пишь, забошилъ, пишь. гл. д. |просп. Причиняю кому заботу, хло-|попы. *Малрасно ты меня такою без-|дѣлицею заботишь.*

Заботусь, ся, пишся, забошися, забо-|пишся. гл. возв. Спараяся, пекуся о|комѢ или о чемѢ; заботу имѣю. *Забо-|тусь о воспитаніи дѣтей. Добрый те-|ловѣкъ о другихъ заботится.*

4

Беззаботливый, вая, вое. *Беззаботливъ*, |ва, во. прил. Безпечный, заботливости |неимѣющій; небрежущій о

помѢ, что |надобно. *Такѡ беззаботливѡ, что ни |за
темѡ не смотритѡ.*

Беззаботливо. нар. БезѢ хлопотѢ, без-|печно. *Беззабот-
ливо вѣкѡ провож-|даетѡ.*

Беззаботливость, спи. с. ж. просп. |Безпечность, нера-
дивость. *Ты своею |беззаботливостію проронилѡ это
|мѣсто.*

Беззаботный, ная, ное. *Беззаботенѡ,|пна, пно.* прил. 1)
Безпечный, нера-|дивый. *Беззаботный теловѣкѡ
себѣ|вреденѡ, другиѡ безлолезенѡ.* 2)|Безпечальный,
свободный опѢ вся-|кихѢ хлопотѢ, попеченій.|Ты (куз-
нечикѢ) скатешь и лоешь, | свободенѡ, беззаботенѡ.
М. Л.

Беззаботно. нар. 1) Безпечно. 2) Безпе-|чально. *Живу
беззаботно.*

О Б Е З З А Б О Ч И В А Ю, ешь, заботилѢ, бо-
чу,|бочивать, заботить. гл. д. просто-|нар. Свобождаю
кого опѢ заботы, опѢ|хлопотѢ. *Ежели ты дѣло это
воз-|мешь на себя, такѡ совершенно ме-
ня|обеззаботишь.*

Обеззаботиваюсь, ешься, бочуся, забо-|пился, обеззабо-
питься. гл. возвр. |просто-нар. Освобождаюсь опѢ хло-
|потѢ. *Пристроивѡ дѣтей кѡ мѣсту|обеззаботился.*

Пример 2.

Примеры выделения цветом элементов словарных статей.

ЗАБОБОНЫ, бонѢ. с. ж. множ. Бред-|ни, сказки по боль-
шей части основан-|ныя на суевѣрїи, и не достойныя
|вниманія умнаго человѣка. *Сей ли |лутѣ сласенїя, яко
ломратенѡ забо-|бонами не знаешь, что ты долженѡ
|еси Богу, государю, отгествѣ?* **Деоф.** |Прокон.
*Ты жилѡ по тѣмѡ законамѡ,|Которые писалѡ;|Смѣялся
забобонамѡ.* **М. Лом.**

Беззаботный, ная, ное. *Беззаботенѣ*,|пна, пно. **прил.** 1) Бепечный, нера-|дивый. *Беззаботный челоѡкъ себѣ|вреденѣ, другимѣ безлолезенѣ.* 2)|Бепечальный, свободный опѣ вся-|кихѣ хлопотѣ, попеченій. *Ты (кузнечикѣ) скажешь и лоешь, | свобо-денѣ, беззаботенѣ.* **М. Л.**

О Б Е З З А Б О Ч И В А Ю, ешь, забопилѣ, бочу,|бочивапѣ, за-бопипѣ. **гл. д. просто-|нар.** Свобождаю кого опѣ заботы, опѣ|хлопотѣ. *Ежели ты дѣло это воз-|мешь на себя, такѣ совершенно меня|обеззаботишь.*

Пример 3.

Результирующая таблица 1. «Заголовочные слова»

№ столбца	№ словарной статьи	заголовочное слово	словарная статья полностью
2	1	ЗАБОБОНЫ	ЗАБОБО'НЫ , бонѣ. с. ж. множ. Бред- ни, сказки по большей части основан- ныя на суевѣрїи, и не достойныя вниманїя умнаго челоѡка. <i>Сей ли лутѣ сласенїя, яко пом-ратенѣ забо- бонами не зна-ешь, что ты долженѣ еси Богу, государю, отечеству?</i> Феоф. Прокп. Ты жилѣ по тѣмѣ зако-намѣ, Которые ли-салѣ; Смѣялся забабонамѣ. М. Лом.
2	2	Забабонщина	Забабонщина , щины. с. ж. про-спонар. Рѣчи и рассказы пуспыя, основанныя на суевѣрїи. <i>Какой забабонщинѣ ты вѣришь!</i>
2	3	ЗАБОТА	ЗАБО'ТА , пы. с. ж. Попеченіе о чемѣ, пщаніе, спараніе со-

№ столбца	№ словарной статьи	заголовочное слово	словарная статья полностью
			единенное сѢ без- покойствомъ и печалію. Домашнїя заботы сокрушили его здоровье. Весь въѣз живу въ заботахъ. Межь тѣмъ на Горизонтъ заря восходитъ. И смертныхъ бдѣнїя, и тмы заботъ возводитъ. В. Петр. Ен.
3	1	Заботливый	Забоѡтливый, вая, вое. Забоѡтливъ, ва, во. прил. Спарательный, попечи- пельный; который о чемъ, или о комъ печется, забо-пишся. Заботливая хо- зяйка. Забо-тливъ о лодиченныхъ.
3	2	Заботливо	Забоѡтливо. нар. Спарательно, попечи- пельно. Этотъ хозяинъ заботливо за всемъ смотритъ.
3	3	Заботливость	Забоѡтливость, спи. с. ж. Свойсп-во по- печипельнаго, спарательнаго, пща- пельнаго челоуѣка; попечи-пель- носпъ. Съ его заботливостію всякой трудъ преодолѣть можно.
3	4	Заботный	Забоѡтный, ная, ное. Забоѡтенъ, пна, пно. прил. Съ хлопотами соединенный. Заботная долж-ность. Заботной день.
3	5	Заботно	Заботно. нар. Хлопотно, мно-годѣльно, суепливо. Всякому кре-стьянину вълѣтнюю лорцу забот-но.
3	6	Забосу	Забосу, пишь, забопилъ, пишь. гл. д. протст. Причиняю кому забопу,

№ столбца	№ словарной статьи	заголовочное слово	словарная статья полностью
			хло- попы. <i>Малрасно ты меня такую без- дѣлицею заботишь.</i>
3	7	<i>Забосусь</i>	<i>Забосусь</i> , ся, пишьяся, забопился, забо- пиписья. гл. возв. Спараяся, пекуся о комѢ или о чемѢ; забопу имѢю. <i>Заб- сусь о воспитаніи дѣтей. Добрый те- ловѣкъ о дру-гихѣ заботится.</i>
4	1	<i>Беззаботливый</i>	<i>Беззаботливый</i> , вая, вое. <i>Безза-ботливѣ</i> , ва, во. прил. Безпечный, заботливоспи неимѢющий; небрежу-щій о помѢ, чпо надобно. <i>Такѣ беззаботливѣ, то ни за темѣ не смотритѣ.</i>
4	2	<i>Беззаботливо</i>	<i>Беззаботливо</i> . нар. БезѢ хлопопѢ, без- печно. <i>Беззаботливо вѣкъ про-вож- даетѣ.</i>
4	3	<i>Беззаботливость</i>	<i>Беззаботливость</i> , спи. с. ж. прост. Безпечность, нерадивость. <i>Ты своею беззаботливостію проро-нилѣ это мѣсто.</i>
4	4	<i>Беззаботный</i>	<i>Беззаботный</i> , ная, ное. <i>Беззаб-тенѣ</i> , пна, пно. прил. 1) Безпеч-ный, нера- дивый. <i>Беззаботный теловѣкъ себѣ вреденѣ, другиѣмѣ безлозенѣ.</i> 2) Безпечальный, сво-бодный опѢ вся- кихѣ хлопопѢ, попеченій. <i>Ты (кузнечикѣ) скасеш и лоешь, свободенѣ, беззаботенѣ.</i> М. Л.
4	5	<i>Беззаботно</i>	<i>Беззаботно</i> . нар. 1) Безпечно. 2) Безпе- чально. <i>Живу беззаботно.</i>

№ столбца	№ словарной статьи	заголовочное слово	словарная статья полностью
4	6	ОБЕЗЗАБОЧИВАЮ	ОБЕЗЗАБОЧИВАЮ, ешь, забопилѢ, бочу, бочивать, забопить. гл. д. просто-нар. Свобождаю кого опѢ забопы, опѢ хлопотѢ. <i>Ежели ты дѣло это воз- мешь на себя, такѢ совершен-но меня обеззаботишь.</i>
4	7	Обеззаботиваюсь	Обеззаботиваюсь, ешься, бочуся, забо- пился, обеззабопиться. гл. возвр. простонар. Освобождаюсь опѢ хло- потѢ. <i>ПристроивѢ дѣтей кѢ мѣсту обеззаботился.</i>

Пример 4.**Результирующая таблица 2. «Значения заголовочных слов».**

№ столбца	№ словарной статьи	заголовочное слово	№ значения	значения
2	1	ЗАБОБОНЫ	1	Бред- ни, сказки по боль-шей части основан- ныя на суевѣрїи, и не достойныя вниманїя умнаго человѣка. <i>Сей ли луть сласенїя, яко помраченѢ забо- бонами не знаешь, что ты долженѢ еси Богу, государю, отетеству?</i> Ееоф. Прокон. <i>Ты жилѢ по тѣмѢ зако-намѢ, Которые ли-салѢ; Смѣялся забабонамѢ.</i>

				М. Лом.
2	2	Забабонщина	1	Рѣчи и рассказы пустыя, основанныя на суевѣрїи. <i>Какой забабонщинѣ ты вѣришь!</i>
2	3	ЗАБОТА	1	Попеченіе о чемѣ, пшчаніе, стараніе соединенное съ без-покойствомъ и печалію. <i>Домашнія заботы сокрушили его здоровье. Весь вѣкъ живу въ заботахъ.</i> <i>Межъ тѣмъ на Горизонтѣ заря восходитъ, И смертныхъ бдѣнїя, и тмы заботъ возводитъ.</i> В. Петр. Ен.
3	1	Заботливый	1	Спарательный, попечительный; копорый о чемѣ, или о комѣ печется, забопипся. <i>Заботливая хозяйка. Заботливъ о подчиненныхъ.</i>
3	2	Заботливо	1	Спарательно, попечительно. <i>Этотъ хозяинъ заботливо за всемъ смотритъ.</i>
3	3	Заботливость	1	Свойство попечительнаго, спарательнаго, пшца-птельнаго челоѣка; попечитель-носсть. <i>Съ его заботливостію всякой трудъ преодолѣть можно.</i>
3	4	Заботный	1	Съ хлопотами соединенный. <i>Заботная должность. Заботной день.</i>
3	5	Заботно	1	Хлопотно, мно-годѣльно, суепливо. <i>Всякому крестьянину вѣлѣнную поручаю.</i>

3	6	<i>Забосу</i>	1	Причиняю кому заботу, хло- попы. <i>Малрасно ты меня такою без-дѣлицею заботишь.</i>
3	7	<i>Забосусь</i>	1	Спараюсь, пекуся о комѢ или о чемѢ; заботу имѣю. <i>Заб-тусь о воспитаніи дѣтей. Добрый те-ловѣкъ о другихъ заботит- ся.</i>
4	1	<i>Беззаботлив- вый</i>	1	Безпечный, заботливо- спи неимѣющий; небрежущій о помѢ, чпо надобно. <i>Такъ без- заботливъ, что ни за темъ не смотритъ.</i>
4	2	<i>Беззаботливо</i>	1	БезѢ хлопотѢ, без- печно. <i>Без- заботливо вѣкъ лрвож- даетъ.</i>
4	3	<i>Беззаботли- вость</i>	1	Безпечность, нерадивость. <i>Ты своею беззаботливостію проро- нилъ ето мѣсто.</i>
4	4	<i>Беззаботный</i>	1	Безпечный, нера- дивый. <i>Безза- ботный теловѣкъ себѣ вреденъ, други мъ безлолезенъ.</i>
4	4	<i>Беззаботный</i>	2	Безпечальный, свободный опѢ вся- кихъ хлопотѢ, попече- ній. <i>Ты (кузнечикъ) скасешь и поешь, свободенъ, беззабо- тенъ. М. Л.</i>
4	5	<i>Беззаботно</i>	1	Безпечно.
4	5	<i>Беззаботно</i>	2	Безпе- чально. <i>Живу беззабот- но.</i>
4	6	<i>О БЕЗЗА- БОЧИВАЮ</i>	1	Свобождаю кого опѢ заботы, опѢ хлопотѢ. <i>Ежели ты дѣло это воз- мешь на себя, такъ совершенно меня обеззаботишь.</i>
4	7	<i>Обеззаботива- юсь</i>	1	Освобождаюсь опѢ хло- потѢ.

				Пристроивъ дѣтей къ мѣсту обеззаботился.
--	--	--	--	---

Пример 5.

Таблица 3. Грамматические характеристики заголовочных слов.

№ столбца	№ словарной статьи	заголовочное слово	Варианты	Часть речи	Род	Число	Гл. характеристики	Пометы
2	1	ЗАБОБОНЫ	бонѣ.	с.	ж.	множ.		
2	2	Забобнищина	щины.	с.	ж.			про- спонар.
2	3	ЗАБОТА	пы.	с.	ж.			
3	1	Заботливый	вая, вое. Забот- ливѣ, ва, во.	прил.				
3	2	Заботливо		нар.				
3	3	Заботливость	спи.	с.	ж.			
3	4	Заботный	ная, ное. Заботенѣ, пна, пно.	прил.				
3	5	Заботно		нар.				
3	6	Забосу	пишь, за- бopilъ, пипись.	гл.			д.	просп.
3	7	Забосусь	ся, пишься, забopился, забо- пиписья.	гл.			возв.	
4	1	Беззаботливый	вая, вое. Беззабот-	прил.				

			лвѣ, ва, во.					
4	2	Беззаботливо		нар.				
4	3	Беззаботлив- вость	спи.	с.	ж.			просп.
4	4	Беззаботный	ная, ное. Беззабо- тенѣ, пна, пно.	прил.				
4	5	Беззаботно		нар.				
4	6	ОБЕЗЗА- БОЧИВА Ю	ешь, забо- пилѣ, бо- чу, бочивап- ь, забо- пипись.	гл.		д.		просто- нар.
4	7	Обеззаботиваюсь	ешься, бо- чуся, забо- пился, обеззабо- пипись.	гл.		возвр.	про- спонар.	

Пример 6.

Таблица 4. Характеристики значений.

№ столбца	№ слов. статьи	заголовоч- ное слово	№ значения	Толкование	Примеры	Цитата	Источник
2	1	ЗАБОБО- НЫ	1	Бред- ни, сказки по большей части ос- нован- ные на суевѣрїи, и не дос-		Сей ли луть сласенїя, яко ломра- тенѣ за- бо-	Феоф. Прокоп.

				пойныя вниманія умнаго че- ловѣка.		бонами не зна- ешь, что ты долженъ еси Бо- гу, госу- дарю, отечест- ву?	
2	1	ЗАБОБО- НЫ	1	Бред- ни, сказки по большей части ос- нован- ныя на суевѣрїи, и не дос- пойныя вниманія умнаго че- ловѣка.		Ты жилъ по тѣмъ законамъ, Ко- торые ли- салъ; См- ѣлся забобо- намъ.	М. Лом.
2	2	Забабонщи- на	1	Рѣчи и раз- сказы пус- тые, осно- ванные на суевѣрїи.	Какой забабон- щинѣ ты вѣришь!		
2	3	ЗАБОТА	1	Попеченіе о чемъ, пщаніе, спараніе со- единенное съ без- покойствомъ и печалію.	Домаш- нія заботы сокрушили его здоро- вье. Весь вѣкъ живу въ заботахъ.	Межь тѣмъ на Горизонтѣ заря вос- ходитъ. И смерт- ныхъ бдѣнія, и	В. Петр. Ен.

					тѣмъ заботѣ возводитѣ.	
3	1	Заботливый	1	Спарательный, попельный; копорый о чемѣ, или о комѣ печется, заботипся.	Заботливая хоз- зяка. Заботливѣ о лодси- неннихѣ.	
3	2	Заботливо	1	Спарательно, попельно.	Этотѣ хозяинѣ заботливо за всемѣ смот- ритѣ.	
3	3	Заботливость	1	Свойство по- печипельнаго, спарательнаго, пща- пельнаго челоуѣка; попельнось.	Сѣ его заботливостію вся- кой трцѣ пре- одолѣть можно.	
3	4	Заботный	1	Сѣ хлопопа- ми соединенный.	Заботная долж- ность. Заботной день	.
3	5	Заботно	1	Хлопопно, многодѣльно, суепливо.	Всякому крестья- нину вѣ дѣтнюю	

					лорц за- ботно.		
3	6	Заботу	1	Причиняю кому заботу, хло- попы.	Малрасно ты меня такою без- дѣлицею заботишь.		
3	7	Заботуся	1	Спараюся, пекуся о комѣ или о чемѣ; заботу имѣю.	Забо- туся о восли- таніи дѣтей. Добрый те- ловѣкъ о другихъ заботит- ся.		
4	1	Беззабот- ливый	1	Безпечный, забопливо- спи неимѣющі й; небрежущій о помѣ, чпо надобно.	Такъ без- заботливъ, сто ни за темъ не смот- ритъ.		
4	2	Беззабот- ливо	1	Безъ хло- потъ, без- печно.	Беззабот- ливо въкъ лрочно ж- дастъ.		
4	3	Беззабот- ливость	1	Безпечность, нерадивость.	Ты сво- ею беззабо- тливо- стію про- ронишь ето мѣсто .		
4	4	Беззабот-	1	Безпечный,	Беззабот-		

		ный		нера- дивый.	ный те- ловѣкъ себѣ вреден ѡ, дру- гилѡ без- полезенѡ.		
4	4	Беззабот- ный	2	Безпечальный, свободный опѣ вся- кихѣ хло- попѣ, попе- ченій.		Ты (куз- нечикѣ) скатаешь и лоешь, свободенѡ, беззабо- тенѡ.	М. Л.
4	5	Беззаботно	1	Безпечно.			
4	5	Беззаботно	2	Безпе- чально.	Живу беззабот- но.		
4	6	О Б Е З - З А Б О Ч И - В А Ю	1	Свобождаю кого опѣ заботы, опѣ хлопопѣ.	Ежели ты дѣло это воз- мешь на себя, такѡ со- вершенно ме- ня обеззаб- отишь.		
4	7	Обеззабси- ваюсь	1	Освобождаюсь опѣ хло- попѣ.	Присстро- ивѡ дѣтей кѡ мѣсту обе- ззаботил- ся.		

Задачи конструкторского этапа

Перед студентами в форме заданий курсовой работы были поставлены первые три задачи конструкторского этапа: 1) разработка лингвистической базы данных; 2) разработка технологического процесса заполнения базы данных и автоматизированный ввод словарных статей; 3) исследование эффективности технологического процесса заполнения базы данных. Им были даны следующие рекомендации по их выполнению.

Задание 3. Перед началом разработки лингвистической базы данных отрезка САР необходимо ознакомиться с методическими материалами: 1) М.И.Чернышева. «Состав и структура Словаря академии Российской». (файл Чернышева.doc). 2) Предисловие. ... (файл Предисловие.doc). Для более глубокого понимания проблемы создания баз данных гнездовых словарей рекомендуется ознакомиться со статьей А.Н.Тихонов. «Первый гнездовой словарь и его роль в развитии русской лексикографии» (файл Тихонов.doc). Структура разработанной базы данных должна отражать особенности заданного отрезка САР. Среда разработки базы данных Microsoft Access (2003).

Задание 4. При разработке технологического процесса заполнения базы данных должны быть выделены и описаны этапы, процедуры и отдельные операции. Технологический процесс должен быть представлен в виде схемы с пояснениями.

Задание 5. Исследование эффективности технологического процесса ввода информации состоит в определении основных характеристик — ресурсных затрат и ограничений на выполнение этапов, процедур и отдельных операций. Данное исследование может проводиться в процессе реализации предыдущих заданий. Результаты могут быть получены методами сплошного или выборочного исследования.

Перечисленные задания выполнили не все студенты. Далее приводятся несколько вариантов выполнения этих заданий.

Вариант 1.

Исполнитель: студент группы ИУ5-103 Н.С.Петров

Исходным материалом является отрезок первого тома Словаря Академии Российской (столбцы 203 — 302), представляющий собой совокупность из 717 словарных статей.

Разработка лингвистической базы данных

В результате проведенного анализа заданного отрезка САР и методической литературы, была разработана лингвистическая база данных, структура которой отражает особенности заданного отрезка САР. Структуру разработанной базы отражают инфологическая и даталогические схемы базы данных.

Инфологическая схема лингвистической базы данных разработана в нотации Чена и представлена на рисунке 1.

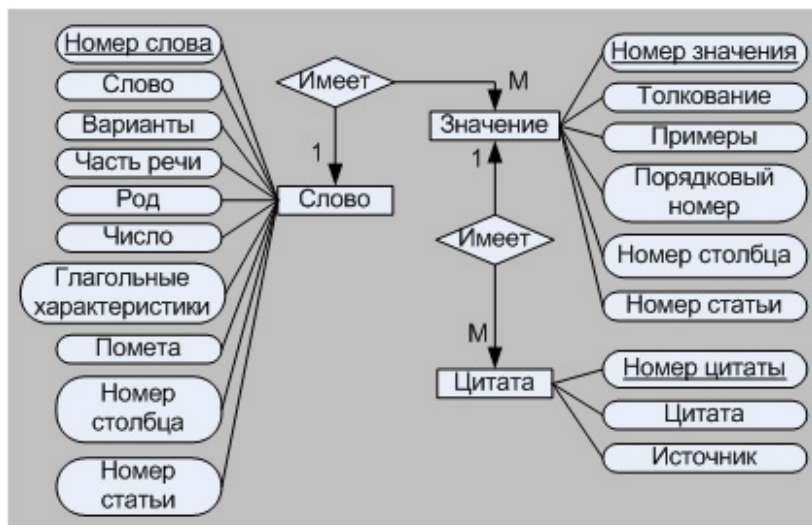


Рисунок 1. Инфологическая схема лингвистической базы данных

По инфологической модели лингвистической базы данных была разработана даталогическая схема в формате Microsoft Access. Даталогическая схема представлена на рисунке 2.



Рисунок 2. Даталогическая схема лингвистической базы данных

База данных разработана в среде Microsoft Access 2003 и представляет собой совокупность трех связанных таблиц, описывающих структуру словарной статьи САР.

Разработка технологического процесса заполнения базы данных и автоматизированный ввод словарных статей

Для заполнения разработанной лингвистической базы данных был разработан технологический процесс, включающий все этапы, процедуры и операции, регламентирующие порядок заполнения базы данных.

Схема технологического процесса представлена на рисунке 3.

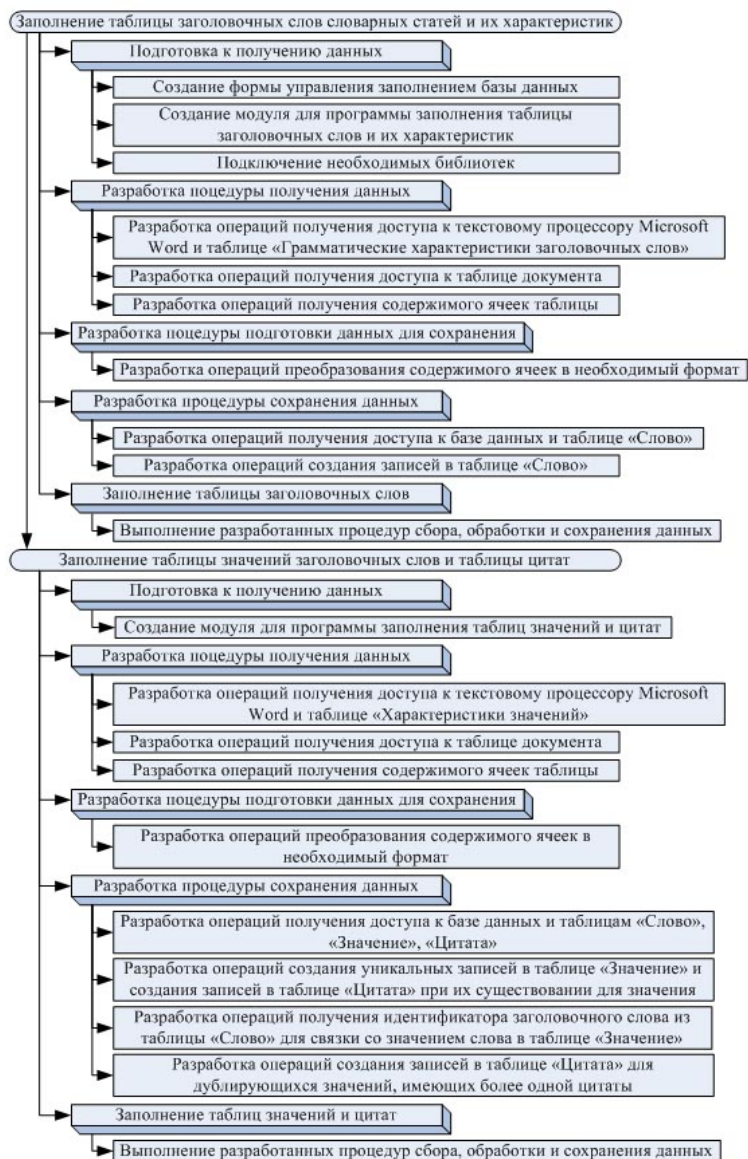


Рисунок 3. Схема технологического процесса

Пояснения к схеме технологического процесса.

1. Заполнение таблицы заголовочных слов словарных статей и их характеристик.
 - 1.1. Подготовка к получению данных.
 - 1.1.1. Создание формы управления заполнением базы данных.

Создается форма Access для управления этапами создания заполнения базы данных: База данных → Формы → Создание формы в режиме конструктора.

Создаются две кнопки: Панель элементов → Кнопка (Кнопка1: «Заполнение БД Заголовочными словами и их Грамматическими характеристиками (Таблица Слово)», Кнопка2: «Задание Заголовочным словам значений и цитат (Заполнение Таблиц "Значение" и "Цитата")»).
 - 1.1.2. Создание модуля для программы-обработчика событий кнопки «Заполнение БД Заголовочными словами и их Грамматическими характеристиками (Таблица Слово)».
 - 1.1.3. Подключение необходимых библиотек.

Microsoft Visual Basic → Tools → References → Выбор пункта Microsoft Word 11.0 Object Library → Нажатие кнопки Ок.
- 1.2. Разработка процедуры получения данных из таблицы «Грамматические характеристики заголовочных слов».
 - 1.2.1. Разработка операций получения доступа к текстовому процессору Microsoft Word и таблице «Грамматические характеристики заголовочных слов».

В теле программы-обработчика создается код, осуществляющий запуск редактора Word и открытие в нем документа «Грамматические характеристики заголовочных слов». Для этого используется технология ADO и объектная модель редактора Word.
- 1.2.2. Разработка операций получения доступа к таблице документа и к строкам данной таблицы.

В теле программы-обработчика создается код, получающий доступ к первой таблице документа и указанной строке в ней.

- 1.2.3. Разработка операций получения содержимого ячейки каждой строки таблицы, кроме первой.

В теле программы-обработчика создается код, осуществляющий получение содержимого указанной ячейки текущей строки.

- 1.3. Разработка процедуры подготовки полученных данных для сохранения.

- 1.3.1. Разработка операций преобразования содержимого ячеек в приемлемый для сохранения вид.

В теле программы-обработчика создается код, преобразующий данные, извлеченные из ячейки таблицы, в необходимый формат путем отсечения служебных символов.

- 1.4. Разработка процедуры сохранения данных.

- 1.4.1. Разработка операций получения доступа к разработанной базе данных и таблице «Слово».

В теле программы-обработчика создается код, открывающий для текущей базы данных таблицу «Слово».

- 1.4.2. Разработка операций создания записей в таблице «Слово» для каждой строки таблицы «Грамматические характеристики заголовочных слов».

Данные, извлеченные из каждой ячейки определенной строки таблицы документа и необходимым образом отформатированные, вносятся в таблицу «Слово» лингвистической базы данных как новая запись.

- 1.5. Заполнение таблицы заголовочных слов.

- 1.5.1. Выполнение разработанных процедур сбора, обработки и сохранения данных нажатием кнопки «Заполнение БД Заголовочными словами и их Грамматическими характеристиками (Таблица Слово)».

Происходит выполнение вышеописанного кода, заполняющего таблицу «Слово» лингвистической базы данных. Для каждой строки таблицы документа (кроме первой) создается соответствующая запись в таблице базы данных.

2. Заполнение таблицы значений заголовочных слов и таблицы цитат.

2.1. Подготовка к получению данных.

2.1.1. Создание модуля для программы-обработчика событий кнопки «Задание Заголовочным словам значений и цитат (Заполнение Таблиц "Значение" и "Цитата")».

Создается обработчик нажатия Кнопки2: Меню кнопки → Обработка событий.

2.2. Разработка процедуры получения данных из таблицы «Характеристики значений».

2.2.1. Разработка операций получения доступа к текстовому процессору Microsoft Word и таблице «Характеристики значений».

В теле программы-обработчика создается код, осуществляющий запуск редактора Word и открытие в нем документа «Характеристики значений». Для этого используется технология ADO и объектная модель редактора Word.

2.2.2. Разработка операций получения доступа к таблице документа и к строкам данной таблицы.

В теле программы-обработчика создается код, получающий доступ к первой таблице документа и указанной строке в ней.

2.2.3. Разработка операций получения содержимого ячейки каждой строки таблицы, кроме первой.

В теле программы-обработчика создается код, осуществляющий получение содержимого указанной ячейки текущей строки.

2.3. Разработка процедуры подготовки полученных данных для сохранения.

2.3.1. Разработка операций преобразования содержимого ячеек в приемлемый для сохранения вид.

В теле программы-обработчика создается код, преобразующий данные, извлеченные из ячейки таблицы, в необходимый формат путем отсечения служебных символов.

2.4. Разработка процедуры сохранения данных.

- 2.4.1. Разработка операций получения доступа к разработанной базе данных и таблицам «Слово», «Значение», «Цитата».

В теле программы-обработчика создается код, открывающий для текущей базы данных таблицы «Слово», «Значение», «Цитата».

- 2.4.2. Разработка операций создания записей в таблице «Значение» для каждого впервые встретившегося значения заголовочного слова и создания записей в таблице «Цитата» в случае, если для данного значения существует цитата.

Данные, извлеченные из каждой ячейки каждой строки таблицы документа (кроме первой) проверяются на существование идентичного толкования в таблице «Значение» базы данных. Если данного толкования для указанного заголовочного слова еще нет в таблице «Значение», то эти данные вносятся в таблицу базы данных как новая запись. Если для данного толкования существует в таблице документа цитата, то она вносится в таблицу «Цитата» базы данных.

- 2.4.3. Разработка операций создания записей в таблице «Цитата» для каждого значения, встречающегося не один раз для данного заголовочного слова, и имеющего более одной цитаты.

В случае, если для выбранного из таблицы документа толкования уже есть идентичный вариант в таблице «Значение» базы данных, то поле цитаты для данного толкования вносится в таблицу «Цитата» базы данных.

- 2.4.4. Разработка операций получения уникального идентификатора необходимого заголовочного слова из таблицы «Слово» для связи со значением данного слова в таблице «Значение».

При внесении новой записи в таблицу «Значение», происходит поиск указанного заголовочного слова в таблице «Слово» базы данных. У найденной записи извлекается ключ и вносится в соответствующую запись таблицы «Значение».

- 2.5. Заполнение таблиц значений и цитат.

- 2.5.1. Выполнение разработанных процедур сбора, обработки и сохранения данных нажатием кнопки «Задание Заголовочным словам значений и цитат (Заполнение Таблиц "Значение" и "Цитата")».

Происходит выполнение вышеописанного кода, заполняющего таблицы «Значение» и «Цитата» лингвистической базы данных. Для каждой строки таблицы документа (кроме первой) создается соответствующая запись в таблицах базы данных.

Программная реализация заполнения лингвистической базы данных

```
Private Sub Кнопка0_Click()
On Error GoTo Err_Кнопка0_Click

Dim objWord As Word.Application
Dim docWord1 As Word.Document

Dim tblWord1 As Word.Table

Dim rowWord1 As Word.Row

Dim cnx As ADODB.Connection
Dim cmd As ADODB.Command
Dim rst As Recordset
Dim db As Database

Dim strNumCol As String
Dim strArticle As String
Dim strWord As String
Dim strVar As String
Dim strPart As String
Dim strKind As String
Dim strNum As String
Dim strVerb As String
Dim strMark As String

Dim intA As Integer

Set cnx = CurrentProject.Connection
Set cmd = New ADODB.Command

Set cmd.ActiveConnection = cnx
cmd.CommandType = adCmdText

Set objWord = New Word.Application
Set docWord1 = objWord.Documents.Open("E:\Univer\ИУ5-103\Филлипович\Табл.
Грамматические характеристики заголовочных слов.doc", , True)
Set tblWord1 = docWord1.Tables(1)

Set db = CurrentDb()
Set rst = db.OpenRecordset("Слово")

Set rowWord1 = docWord1.Tables(1).Rows(1)

Do
Set rowWord1 = rowWord1.Next
```

```

rst.AddNew
strNumCol = rowWord1.Cells(1).Range.Text
strNumCol = Left(strNumCol, Len(strNumCol) - 2)
rst.Fields("Номер столбца").Value = Val(strNumCol)

strArticle = rowWord1.Cells(2).Range.ListFormat.ListString
strArticle = Left(strArticle, Len(strArticle) - 0)
rst.Fields("Номер статьи").Value = Val(strArticle)

strWord = rowWord1.Cells(3).Range.Text
strWord = Left(strWord, Len(strWord) - 2)
rst.Fields("Слово").Value = strWord

strVar = rowWord1.Cells(4).Range.Text
strVar = Left(strVar, Len(strVar) - 2)
rst.Fields("Варианты").Value = strVar

strPart = rowWord1.Cells(5).Range.Text
strPart = Left(strPart, Len(strPart) - 2)
rst.Fields("Часть речи").Value = strPart

strKind = rowWord1.Cells(6).Range.Text
strKind = Left(strKind, Len(strKind) - 2)
rst.Fields("Под").Value = strKind

strNum = rowWord1.Cells(7).Range.Text
strNum = Left(strNum, Len(strNum) - 2)
rst.Fields("Число").Value = strNum

strVerb = rowWord1.Cells(8).Range.Text
strVerb = Left(strVerb, Len(strVerb) - 2)
rst.Fields("Глагольные характеристики").Value = strVerb

strMark = rowWord1.Cells(9).Range.Text
strMark = Left(strMark, Len(strMark) - 2)
rst.Fields("Помера").Value = strMark

'cmd.CommandText = "INSERT INTO Слово Values (" & Val(strNumCol) & "," & _
'Val(strArticle) & "," & strWord & "," & strVar & "," & strPart & "," & strKind & _
'"," & strVerb & "," & strMark & ")"
'cmd.Execute
rst.Update
Loop Until rowWord1.IsLast

MsgBox ("База заполнена!")

Set rst = Nothing
docWord1.Close
objWord.Quit
Set docWord = Nothing
Set objWord = Nothing
Set cmd = Nothing

Exit_Кнопка0_Click:
Exit Sub

Err_Кнопка0_Click:
MsgBox Err.Description
Resume Exit_Кнопка0_Click

```

```
End Sub

Private Sub Кнопка2_Click()
On Error GoTo Err_Кнопка2_Click

    Dim rst1 As Recordset
    Dim rst2 As Recordset
    Dim rstt As ADODB.Recordset

    Dim objWord As Word.Application
    Dim docWord As Word.Document
    Dim tblWord As Word.Table
    Dim rowWord1 As Word.Row

    Dim strNumCol As String
    Dim strArticle As String
    Dim strNum As String
    Dim strCom As String
    Dim strExam As String
    Dim strCit As String
    Dim strSour As String

    Set objWord = New Word.Application
    Set docWord = objWord.Documents.Open("E:\Univer\ИУ5-103\Филиппович\Табл.
Характеристики значений.doc", , True)
    Set tblWord = docWord.Tables(1)

    Set db = CurrentDb()
    Set rst1 = db.OpenRecordset("Значение")
    Set rst2 = db.OpenRecordset("Цитата")
    Set rstt = New ADODB.Recordset
    rstt.ActiveConnection = CurrentProject.Connection

    Set rowWord1 = docWord.Tables(1).Rows(1)

Do
    Set rowWord1 = rowWord1.Next

    'Номер столбца
    strNumCol = rowWord1.Cells(1).Range.Text
    strNumCol = Left(strNumCol, Len(strNumCol) - 2)

    'Номер словарной статьи
    strArticle = rowWord1.Cells(2).Range.ListFormat.ListString
    strArticle = Left(strArticle, Len(strArticle) - 0)

    If Len(strArticle) <= 0 Then

        strArticle = rowWord1.Cells(2).Range.Text
        strArticle = Left(strArticle, Len(strArticle) - 2)

    End If

    'Номер значения
    strNum = rowWord1.Cells(4).Range.Text
    strNum = Left(strNum, Len(strNum) - 2)

    'Толкование
    strCom = rowWord1.Cells(5).Range.Text
```

```
strCom = Left(strCom, Len(strCom) - 2)

'Примеры
strExam = rowWord1.Cells(6).Range.Text
strExam = Left(strExam, Len(strExam) - 2)

'Цитата
strCit = rowWord1.Cells(7).Range.Text
strCit = Left(strCit, Len(strCit) - 2)

'Источник
strSour = rowWord1.Cells(8).Range.Text
strSour = Left(strSour, Len(strSour) - 2)

rstt.Open "SELECT [Номер значения] FROM Значение " & _
" WHERE [Номер столбца] = " & strNumCol & _
" AND [Номер статьи] = " & strArticle & _
" AND [Порядковый номер] = " & strNum

'Если значение слова встречается впервые
If rstt.RecordCount <= 0 Then

    rst1.AddNew

    rst1.Fields("Номер столбца").Value = Val(strNumCol)
    rst1.Fields("Номер статьи").Value = Val(strArticle)
    rst1.Fields("Порядковый номер").Value = Val(strNum)
    rst1.Fields("Толкование").Value = strCom
    rst1.Fields("Примеры").Value = strExam

    rstt.Close
    rstt.Open "SELECT [Номер слова] FROM Слово " & _
    " WHERE [Номер столбца] = " & strNumCol & _
    " AND [Номер статьи] = " & strArticle

    rstt.MoveFirst
    rst1.Fields("Номер слова").Value = rstt.Fields("Номер слова").Value

'Если Цитата не пуста, то заполняем таблицу "Цитата"
If Len(strCit) > 0 Then

    rst2.AddNew
    rst2.Fields("Цитата").Value = strCit
    rst2.Fields("Источник").Value = strSour
    rst2.Fields("Номер значения").Value = rst1.Fields("Номер значения").Value
    rst2.Update

End If

rstt.Close
rst1.Update

'Если уже имеется такое значение
Else

'Если Цитата не пуста, то заполняем таблицу "Цитата"
If Len(strCit) > 0 Then

    rst2.AddNew
    rst2.Fields("Цитата").Value = strCit
```

```
rst2.Fields("Источник").Value = strSour
rst2.Fields("Номер значения").Value = rstt.Fields("Номер значения").Value
rst2.Update
rstt.Close

End If

End If

Loop Until rowWord1.IsLast

'MsgBox (rst.Fields("Номер значения").Value)
MsgBox ("База заполнена!")

Set rst1 = Nothing
Set rst2 = Nothing
Set rstt = Nothing

docWord.Close
objWord.Quit
Set docWord = Nothing
Set objWord = Nothing
Set tblWord = Nothing
Set rowWord1 = Nothing

Exit_Кнопка2_Click:
Exit Sub

Err_Кнопка2_Click:
Set rst1 = Nothing
Set rst2 = Nothing
Set rstt = Nothing

docWord.Close
objWord.Quit
Set docWord = Nothing
Set objWord = Nothing
Set tblWord = Nothing
Set rowWord1 = Nothing

MsgBox Err.Description
Resume Exit_Кнопка2_Click

End Sub
```

Для апробации лингвистической базы данных разработана форма, предоставляющая удобный интерфейс просмотра записей базы данных по всем словарным статьям (рисунок 4). При открытии файла базы данных происходит автоматическое открытие данной формы.

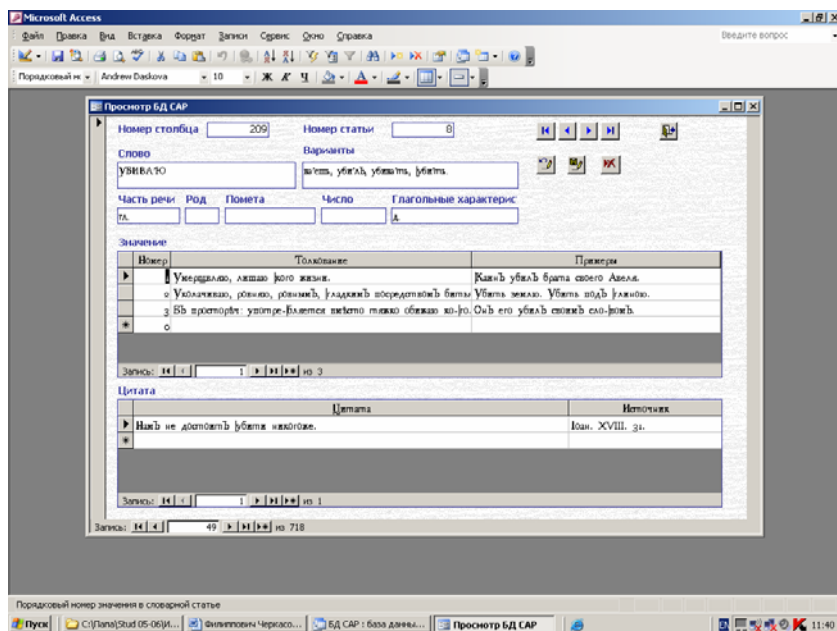


Рисунок 4. Экранная форма лингвистической базы данных.

Исследование эффективности технологического процесса заполнения базы данных

При реализации технологического процесса ввода информации в лингвистическую базу данных были подсчитаны временные затраты на каждую операцию и возможные ограничения на их выполнение. В результате были определены суммарные временные затраты и ограничения на выполнение процедур и этапов техпроцесса.

Итоговые значения временных затрат и ограничений, характеризующих эффективность технологического процесса, приведены в таблице 5. В столбце ограничения описываются программные средства, необходимые для выполнения соответствующей операции, процедуры или этапа.

Этап	Процесс	Операция	Временные затраты, мин	Ограничения
Заполнение таблицы заголовочных слов словарных статей и их характеристик	Подготовка к получению данных	Создание формы управления заполнением базы данных	2	Наличие MS Access как клиента автоматизации
		Создание модуля для программы заполнения таблицы заголовочных слов и их характеристик	1	Наличие среды программирования Visual Basic for Applications
		Подключение необходимых библиотек	2	Наличие объектных библиотек ADO версии 2.1 и MS Word как сервера автоматизации
		Итого	5	Наличие MS Access, Word, среды VBA, библиотек ADO 11.
	Разработка процедуры получения данных	Разработка операций получения доступа к текстовому процессору Microsoft Word и таблице «Грамматические характеристики заголовочных слов»	10	Наличие среды VBA и MS Word

		Разработка операций получения доступа к таблице документа	5	Наличие среды VBA и MS Word
		Разработка операций получения содержимого ячеек таблицы	15	Наличие среды VBA и MS Word
		Итого	30	Наличие среды VBA и MS Word
	Разработка процедуры подготовки данных для сохранения	Разработка операций преобразования содержимого ячеек в необходимый формат	10	Наличие среды VBA
		Итого	10	Наличие среды VBA
	Разработка процедуры сохранения данных	Разработка операций получения доступа к базе данных и таблице «Слово»	5	Наличие MS Access и среды VBA
		Разработка операций создания записей в таблице «Слово»	15	Наличие MS Access и среды VBA
		Итого	20	Наличие MS Access и среды VBA
	Заполнение таблицы заголовочных слов	Выполнение разработанных процедур сбора, обработки и сохранения данных	10	Наличие MS Access и среды VBA

		Итого	10	Наличие MS Access и среды VBA
	Итого	75	Наличие MS Access, Word, среды VBA, библиотек ADO 2.1.	
Заполнение таблицы значений заголовочных слов и таблицы цитат	Подготовка к получению данных	Создание модуля для программы заполнения таблиц значений и цитат	1	Наличие среды программирования Visual Basic for Applications
		Итого	1	Наличие среды программирования Visual Basic for Applications
	Разработка процедуры получения данных	Разработка операций получения доступа к текстовому процессору Microsoft Word и таблице «Характеристики значений»	5	Наличие среды VBA и MS Word
		Разработка операций получения доступа к таблице документа	5	Наличие среды VBA и MS Word
		Разработка операций получения содержимого ячеек таблицы	15	Наличие среды VBA и MS Word

		Итого	25	Наличие среды VBA и MS Word
	Разработка процедуры подготовки данных для сохранения	Разработка операций преобразования содержимого ячеек в необходимый формат	5	Наличие среды VBA
		Итого	5	Наличие среды VBA
	Разработка процедуры сохранения данных	Разработка операций получения доступа к базе данных и таблицам «Слово», «Значение», «Цитата»	5	Наличие среды VBA и MS Access
		Разработка операций создания уникальных записей в таблице «Значение» и создания записей в таблице «Цитата» при их существовании для значения	15	Наличие среды VBA и MS Access
		Разработка операций создания записей в таблице «Цитата» для дублирующихся значений, имеющих более одной цитаты	5	Наличие среды VBA и MS Access

		Разработка операций получения идентификатора заголовочного слова из таблицы «Слово» для связки со значением слова в таблице «Значение»	5	Наличие среды VBA и MS Access
		Итого	30	Наличие среды VBA и MS Access
	Заполнение таблиц значений и цитат	Выполнение разработанных процедур сбора, обработки и сохранения данных	15	Наличие MS Access и среды VBA
		Итого	15	Наличие MS Access и среды VBA
	Итого	76	Наличие MS Access, Word, среды VBA	
	Итого	151	Наличие MS Access, Word, среды VBA, библиотек ADO 2.1.	

Таблица 5. Временные затраты и ограничения на выполнение этапов, процедур и операций технологического процесса

Таким образом, для реализации технологического процесса ввода информации в лингвистическую базу данных необходим временной ресурс 2 часа, 31 минута. Ограничением является невозможность реализации всего технологического процесса без наличия программных

средств Microsoft Access, Word, среды разработки Visual Basic for Applications, библиотек ADO 2.1.

Вариант 2.

Исполнитель: студент группы ИУ5-103 М.Р.Маркус.

Исходным материалом является отрезок второго тома Словаря Академии Российской (столбцы 103-202) представляющий собой совокупность из 610 словарных статей.

Разработка лингвистической базы данных

В качестве структуры лингвистической базы данных был выбран формат хранения языка XML. Он обеспечивает нужную структурированность и обладает практически неограниченными возможностями по увеличению детализованности структуры словаря.

XML-документ словаря имеет следующую структуру:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
- <!-- Created with Dictionary Parse v 1.0 -->
<dictionary>
  <item></item>
  ...
  <item></item>
</dictionary>
```

Где элементы вида <item> представляют собой словарные статьи.

Структура элемента <item>

```
<item word="" col="" base=""> ..... //word – номер слова
..... //col – номер колонки
..... //base – основа
<type val="base"> или <type val="main"> ..... //тип слова
..... //base – славянского языка
..... //main – славенороссийского языка
  <def></def> ..... //собст. слово
  <variant></variant> ..... //варианты написания
  <grammar></grammar> ..... //грамматические характеристики
  <text></text> ..... //текст определения
  <example></example> ..... //пример употребления
  <quote></quote> ..... //цитата-пример
  <source></source> ..... //источник цитаты
  <meaning num=""></meaning>
..... //значение слова num – номер значения
```

Все теги являются необязательными. Тэг <meaning> может включать в себя тэги <text>, <quote>, <example> и <source>.

Разработка технологического процесса заполнения базы данных и автоматизированный ввод словарных статей

Для заполнения разработанной лингвистической базы данных был разработан технологический процесс, включающий все этапы, процедуры и операции, регламентирующие порядок заполнения базы данных.

Схема технологического процесса представлена на рисунке 2.

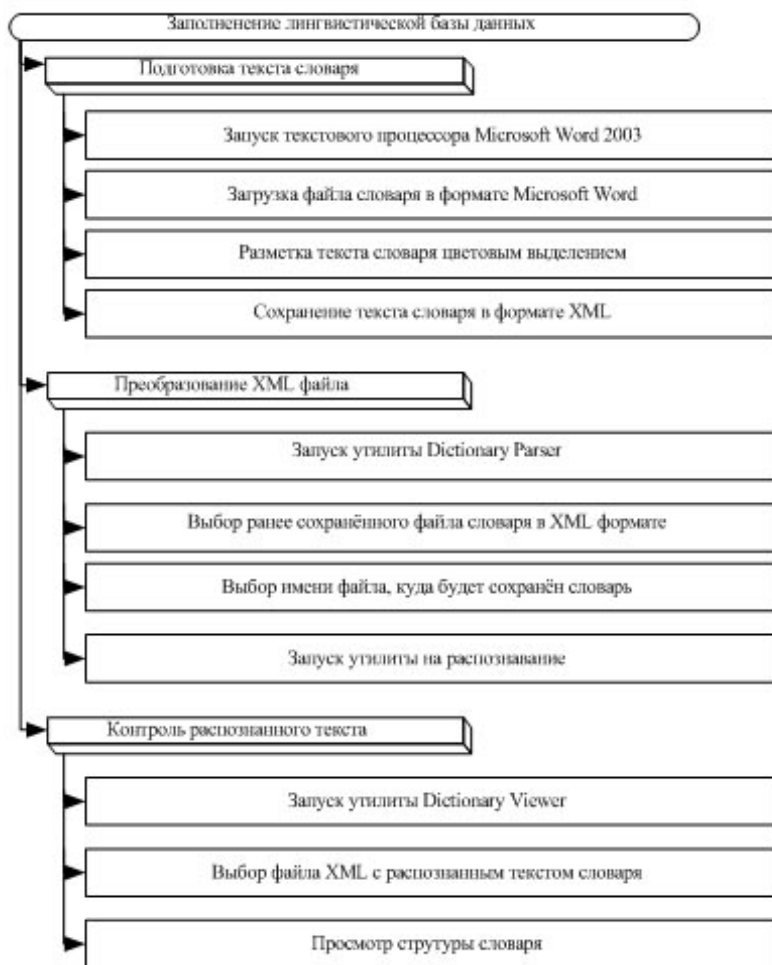


Рисунок 5. Схема технологического процесса

Пояснения к схеме технологического процесса:

1. Заполнение лингвистической базы данных
 - 1.1. Подготовка текста словаря.
 - 1.1.1. Запуск текстового процессора Microsoft Word 2003
 - 1.1.2. Загрузка файла словаря. (Файл-Открыть)

- 1.1.3. Разметка текста цветовым выделением — согласно заданию, определения выделяются жёлтым, грамматические характеристики — зелёным, цитаты — бирюзовым, источники цитат — красным цветом.
- 1.1.4. Сохранение файла словаря в формате XML. (Файл-Сохранить как...-XML Документ)
- 1.2. Преобразование XML файла
 - 1.2.1. Запуск утилиты Dictionary Parser
 - 1.2.2. Выбор ранее сохранённого файла словаря. (Open File)
 - 1.2.3. Выбор имени файла, куда будет сохранён словарь. (Save To)
 - 1.2.4. Запуск распознавания. (Parse). Дождаться появления надписи "Parsing complete".
- 1.3. Контроль текста
 - 1.3.1. Запуск утилиты Dictionary Viewer
 - 1.3.2. Выбор распознанного файла словаря.
 - 1.3.3. Просмотр словаря с помощью встроенных средств и поиска.

Программная реализация заполнения лингвистической базы данных.

Для автоматизации процесса переноса данных словаря в базу данных на основе XML был разработан комплекс утилит, позволяющих формировать базу данных на основе словаря и просматривать данную базу данных.

1. Назначение программного продукта.

Утилиты Dictionary Parser/Viewer предназначены для извлечения данных из файлов «Словаря Академии Российской» формата Microsoft Word 2003 и преобразования их в XML файл специального формата, а также для просмотра результирующей базы данных.

2. Подготовка к работе программы.

Утилиты не требуют специальной установки, достаточно просто скопировать их на жёсткий диск. Для работы программы необходим размеченный цветовым выделением файл словаря, сохранённый с помощью программы Microsoft Word 2003 как XML файл.

Цветовое выделение осуществляется следующим образом: определения выделяются жёлтым, грамматические характеристики — зелёным, цитаты — бирюзовым, источники цитат — красным цветом.

Для корректного отображения информации на компьютере пользо-

вателя должны быть установлены шрифты Andrew Daskova Set.

2.1. Требования к техническому обеспечению.

- Процессор не ниже Intel Pentium II
- Объём оперативной памяти не ниже 256 МБ.
- Монитор, поддерживающий разрешение не ниже 800x600.
- Видеокарта с 4 Мб видеопамати.
- Мышь, клавиатура.

2.3. Требования к программному обеспечению:

- Microsoft Windows XP Home Edition
- .Net Framework 1.1
- Microsoft Word 2003

3. Запуск программы.

Для работы с утилитами достаточно запустить соответствующий .exe файл.

4. Работа с программой.

4.1. Dictionary Parser

Окно утилиты для преобразования словаря в xml файл (Dictionary Parser) выглядит следующим образом: (рисунок 6).

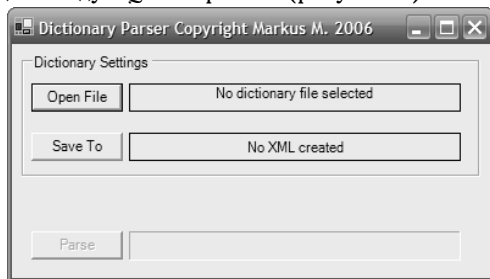


Рисунок 6. Вид окна программы Dictionary Parser.

Для работы с программой необходимо выбрать файл словаря, сохранённый с помощью Microsoft Word 2003 в формат XML (Open File) и указать имя файла, куда будет сохранён преобразованный словарь (Save To)

После этого для создания файла словаря в формате xml необходимо нажать кнопку Parse и дождаться появления надписи «Parsing Complete» (рисунок 7)

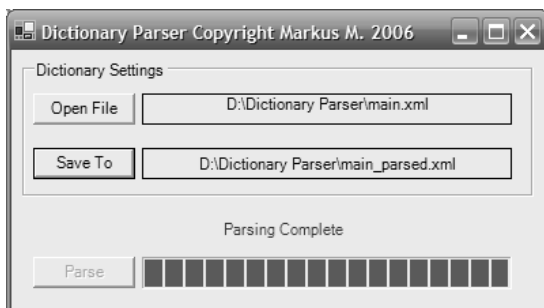


Рисунок 7. Вид окна программы после завершения преобразования.

4.2. Dictionary Viewer

Окно программы для просмотра сохранённого с помощью утилиты Dictionary Parser словаря выглядит следующим образом (рисунок 8):

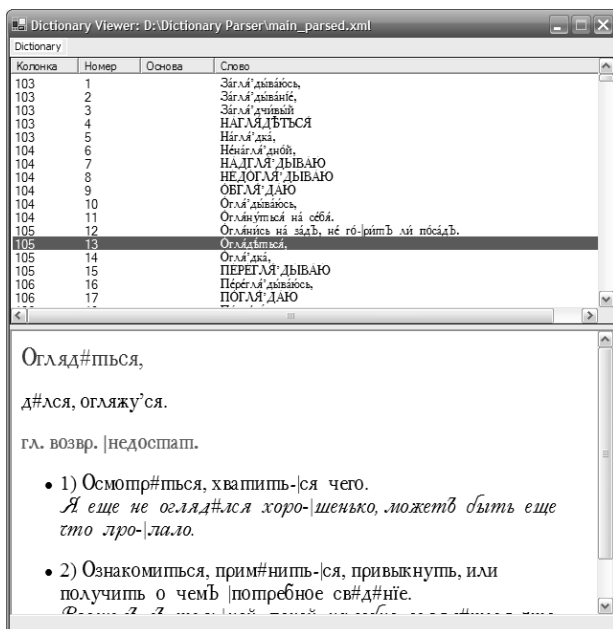


Рисунок 8. Вид окна программы Dictionary Viewer.

Для просмотра словаря необходимо выбрать файл словаря, сформированный утилитой Dictionary Parser (Dictionary-Open...). После того, как программа загрузит файл, его можно будет просматривать щёлкая

на заголовочных словах в списке. В нижней части окна программы появится словарное определение слова².

Для поиска нужного слова нужно воспользоваться встроенными в утилиту средствами поиска (Dictionary-Search Word...) (рисунок 9):

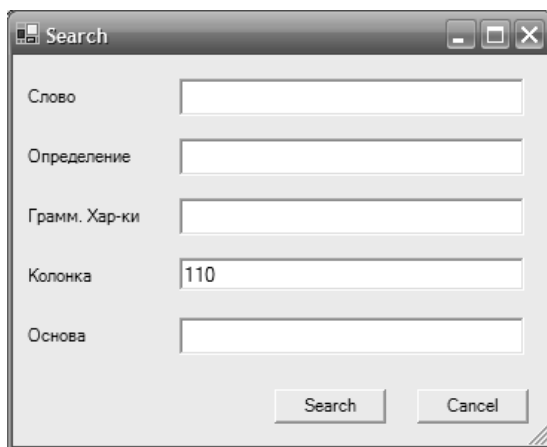


Рисунок 9. Окно параметров поиска.

В этом окне задаются параметры фильтра отображения слов. После нажатия на кнопку Search список заголовочных слов будет отфильтрован согласно параметрам. Для загрузки полного списка слов следует сбросить установки фильтра (Dictionary-Reopen).

Исследование эффективности технологического процесса заполнения базы данных

В процессе реализации технологического процесса ввода информации в лингвистическую базу данных были подсчитаны временные затраты на каждую операцию технологического процесса и возможные ограничения на их выполнение. В результате были определены суммарные временные затраты и ограничения на выполнение процедур и этапов техпроцесса.

² В данной версии программы некорректно отображаются символы шрифта Andrew Daskova. Это связано с ошибкой в Microsoft Internet Explorer 6.0

Этап	Процесс	Операция	Времен- ные за- траты, мин	Ограниче- ния
Заполнение лингвистической базы данных	Подготовка текста словаря	Запуск текстового процессора Microsoft Word 2003	1	Наличие ПО
		Загрузка словаря	1	Наличие файла словаря
		Разметка текста цветовым выделением	На часть словаря объёмом в 610 статей — 200 мин.	
		Сохранение файла словаря в формате XML	1	Версия Microsoft Word не ниже 2003
	Преобразование XML файла	Запуск утилиты Dictionary Parser	1	Наличие утилиты и установленного MS . Net Framework 1.1
		Выбор ранее сохранённого файла словаря	1	Наличие файла размеченного словаря
		Выбор имени файла, в который будет сохранён словарь	1	Наличие места на диске

Этап	Процесс	Операция	Времен- ные за- траты, мин	Ограниче- ния
		Запуск распозна- вания и распозна- вание.	На часть словаря объёмом 610 ста- тей — 10- 15 секунд	При кор- ректно со- хранённом файле сло- варя в фор- мате XML
	Контроль текста	Запуск утилиты Dictionary Viewer	1	Наличие утилиты и установ- ленного MS . Net Framework 1.1
		Выбор ранее со- хранённого файла словаря	1	Наличие XML фай- лы распо- знанного словаря
		Просмотр словаря	-	
Итого:			~209 минут	

Таблица 6. Временные затраты на выполнение этапов технологического процесса.

Разработка программы позволила существенно сократить время, необходимое для переноса данных из словаря в базу данных (для сравнения — обработка данных «вручную» и распределение по таблицам занимает порядка 7-8 часов).

В качестве требований для работы комплекса программ выступает наличие на компьютере у пользователя следующего ПО: Microsoft Word 2003, .NET Framework.

Вариант 2.

Исполнитель: студент группы ИУ5-104 К.Е.Иовков.

Исходным материалом является отрезок второго тома Словаря Академии Российской (столбцы 1103-1200) представляющий собой совокупность из XXX словарных статей.

Подготовка и декомпозиция словарных статей

Подготовка к вводу статей в базу данных и их декомпозиция осуществлялась с помощью написанного на встроенном в текстовый редактор MS Word языке Visual Basic Script макроса. Данный макрос осуществлял автоматизированную маркировку цветом автоматически выделяемых элементов словарной статьи и их перенос в соответствующие столбцы сводных таблиц. Процесс осуществлялся итеративно: при нажатии соответствующей кнопки на форме (рисунок 10) программа формировала новые строки в четырёх таблицах (расположенных в отдельных документах) и заполняла их данными, после чего выделялся следующий фрагмент текста.

Рисунок.10. Форма макроса для подготовки и декомпозиции словарных статей.

Выделение предложений осуществлялось автоматически; при этом дополнительными кнопками данное выделение расширялось или сужалось на единицу предложения или текста. При наличии ошибок был предусмотрен старт маркировки с текущей позиции курсора (текущего словарного слова).

Программа способна выполнить маркировку и декомпозицию на тексте аналогичного формата, при этом необходимо сделать некоторые поправки в коде самого макроса.

Результаты данной декомпозиции не отличаются по своей структуре от аналогичных результатов, полученных вручную, после минимального форматирования четырёх результирующих таблиц.

Результирующие таблицы впоследствии были отформатированы также с использованием макроса, преобразующего переносы строк в пробелы с целью последующего их использования в качестве источника строк для будущей базы данных.

Разработка лингвистической базы данных

При разработке лингвистической базы данных были использованы следующие средства: Microsoft Excel 2002; Microsoft SQL Server 2000; Microsoft Access 2002.

В результате анализа предметной области и требований к результирующим структурам, содержащим форматированные словарные статьи, была разработана логическая схема данных, вид которой в среде разработки MS Access представлен на рисунке 11:

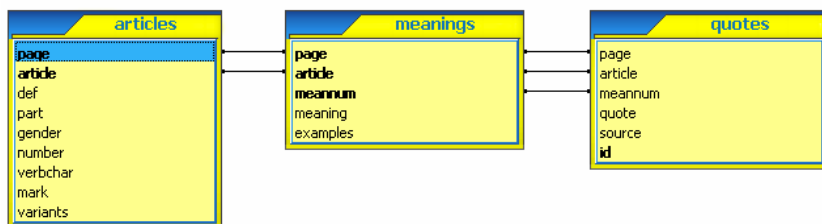


Рисунок 11. Логическая схема данных

База данных соответствует третьей нормальной форме (при условии, что структура данных (т.е. схема таблиц MS Word) была определена требованиями, сформулированными в задании на курсовую работу) и состоит из трёх таблиц:

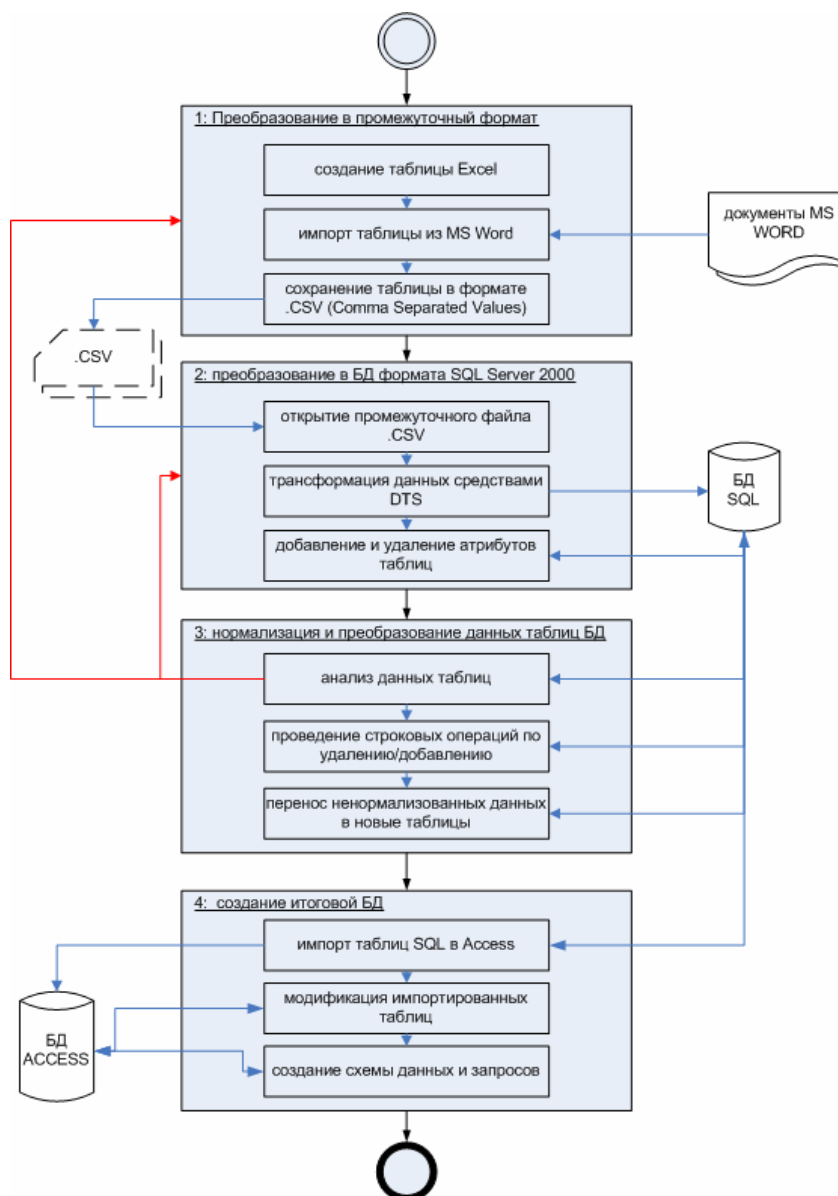
1) Таблица словарных слов и их атрибутов - meanings. Содержит поля:

- page (номер столбца);
- article (номер словарной статьи);

- def (заголовочное слово);
 - part (часть речи);
 - gender (род);
 - number (число);
 - verbchar (глагольные характеристики);
 - mark (пометы);
 - variants (варианты).
- 2) Таблица значений словарного слова - meanings. Содержит поля:
- page (номер столбца);
 - article (номер словарной статьи);
 - meannum (номер значения);
 - meaning (значение слова);
 - examples (примеры).
- 3) Таблица цитат. Содержит поля:
- page (номер столбца);
 - article (номер словарной статьи);
 - meannum (номер значения);
 - quote (цитата);
 - source (источник);
 - id (уникальный идентификатор цитаты).

Разработка технологического процесса заполнения базы данных

В результате работы по автоматизации декомпозиции словарных статей были получены четыре таблицы формата MS Word — входные данные следующего техпроцесса:



На приведённой схеме красной стрелкой указано направление перехода в случае, если анализ показал, что сформированная таблица БД содержит грубые несоответствия типов атрибутов, повлёкшие потери данных, или же структура таблицы БД не соответствует структуре таблицы MS Word.

Технологический процесс заполнения БД основан на обработке только двух таблиц (таблицы вариантов и таблицы цитат), так как содержащиеся в них данные с информационной точки зрения полностью покрывают данные из исходных статей.

Основные этапы технологического процесса:

1. Преобразование в промежуточный формат
 - 1.1. Создание таблицы Excel;
 - 1.2. Импорт таблицы из MS Word. Импорт осуществляется копированием через буфер обмена;
 - 1.3. Сохранение таблицы в формате .CSV (Comma Separated Values).
2. Преобразование в БД формата SQL Server 2000. Происходит для каждого отдельного промежуточного файла (т.е. для каждой таблицы).
 - 2.1. Открытие промежуточного файла .CSV. Устанавливается соответствие разделителей как разделителей строк и столбцов.
 - 2.2. Трансформация данных средствами DTS. Осуществляется создание таблицы в БД SQL Server с атрибутами соответствующего типа и копирование данных из промежуточного файла в созданную таблицу. Осуществляется командой Import Data для активной базы данных.
 - 2.3. Добавление и удаление атрибутов таблиц. Осуществляется коррекция типов атрибутов в зависимости от содержащихся значений, добавляются новые атрибуты таблицы с целью создания третьей нормальной формы для БД.
3. Нормализация и преобразование данных таблиц БД
 - 3.1. Анализ данных таблиц. Создание ряда запросов по проверке соответствия данных их оригинальному виду (т.е. сравнение с таблицами исходных документов), выявления неточностей импортирования и повтор всей операции импортирования при выявлении ошибок.

- 3.2. Проведение строковых операций по удалению/добавлению. В целях нормализации данных из таблицы удаляются избыточные данные — строки, не содержащие в себе новой информации, а также некоторые атрибуты (столбцы);
- 3.3. Перенос ненормализованных данных в новые таблицы. Все заново созданные столбцы таблиц (без каких-либо данных) заполняются соответствующими значениями столбцов других таблиц в целях нормализации схемы данных.
4. Создание итоговой БД
 - 4.1. Импорт таблиц SQL в Access. С помощью Data Transformation Services в SQL Server таблицы добавляются в базу данных Access. Определяется соответствие типов атрибутов импортируемой и итоговой таблиц.
 - 4.2. Модификация импортированных таблиц. Проводится переименование объектов и добавление комментариев, создаются ключи для импортированных таблиц.
 - 4.3. Создание схемы данных и запросов. Создаётся схема данных и итоговые запросы, позволяющие вывести данные в таком виде, в каком они представлены в базовых таблицах MS Word.

Исследование эффективности технологического процесса заполнения базы данных

Данное исследование производилось в процессе выполнения предыдущих заданий.

Наименование	Ресурсные затраты (время выполнения операции)	Ограничения
Декомпозиция и маркировка исходных статей с помощью макроса	13 часов	Существует отлаженный и настроенный на конкретный документ макрос
Непосредственное и автоматизированное форматирование	2 x 1 мин.	Существует отлаженный макрос

результатирующих таблиц		форматирования таблиц
Преобразование данных в промежуточный формат:		Документы отформатированы с помощью макроса форматирования или не содержат переносов строк в ячейках
Создание таблицы Excel	2 x 1 мин.	--/--
Импорт таблицы из MS Word	2 x 2 мин.	--/--
Сохранение таблицы в формате .CSV	2 x 1 мин.	--/--
Преобразование в БД формата SQL Server 2000:	42 мин.	Оператор знаком с операцией Import Data в SQL Server 2000
Открытие промежуточного файла .CSV	2 x 1 мин.	--/--
Трансформация данных средствами DTS	2 x 5 мин.	--/--
Добавление и удаление атрибутов таблиц	2 x 15 мин.	--/--, оператор знает основы манипулирования таблицами средствами SQL Enterprise Manager
Нормализация и преобразование данных таблиц БД:	1 час 30 мин.	Оператор обладает достаточными знаниями Transact — SQL и нормализации данных, данные не содержат грубых логических/структурных ошибок, допущенных при декомпозиции ста-

		тей
Анализ данных таблиц	2 x 15 мин.	--//--
Проведение строковых операций по удалению/добавлению	2 x 10 мин.	--//--
Перенос ненормализованных данных в новые таблицы	2 x 20 мин.	--//--
Создание итоговой БД:	38 мин.	Оператор знаком с MS Access
Импорт таблиц SQL в Access	3 x 4 мин.	--//--, операцией Export Data в SQL Server 2000
Модификация импортированных таблиц	3 x 2 мин.	Оператор знаком с MS Access
Создание схемы данных и запросов	20 мин.	--//--
Сумм.	16 часов	

Таблица 7. Исследование эффективности технологического процесса.

Вариант 4.**Исполнитель: студент группы ИУ5-101 Е.О.Жмакина.**

Исходным материалом является отрезок второго тома Словаря Академии Российской (столбцы 001-102) представляющий собой совокупность из 610 словарных статей.

Разработка лингвистической базы данных

Структура базы данных:

Среда разработки базы данных Microsoft Access (2002).

База содержит 4 таблицы, информация в которых соответствует данным таблиц 1 - 4, разработанных ранее в пунктах 1 и 2:

- Таблица_1_Заголовочные слова
- Таблица_2_Значения заголовочных слов
- Таблица_3_Грамматические характеристики заголовочных слов
- Таблица_4_Характеристики значений

Для удобства введения и редактирования информации в базе имеются 4 одиночных рабочих формы:

- Форма_1_Заголовочные слова.
- Форма_2_Значения заголовочных слов.

- Форма_3_ Грамматические характеристики заголовочных слов.
- Форма_4_ Характеристики значений.

На рисунках 12-15 представлены экранные формы базы данных База_данных_001-102.mdb.

Рисунок 12. Форма «Форма_4_ Характеристики значений».

Рисунок 13. Форма «Форма_1_ Заголовочные слова».

Значения заголовочных слов

№ столбца № словарной статьи

Заголовочное слово

№ значения

Значения заголовочных слов

Есть начертание четвертой вЪ [Славенороссийской азбукѣ буквы, единия изъ огласныхъ называемя Глаголь. — г подѣ пишомѣ вЪ чер-ковномъ числениѣ означаетъ три или зѣле; а со слѣдующимъ знакомъ [и] вЪ нѣзу подписанымъ три ты-сячи, или зооо. Вуква сѣя въ об-щей азыка употребленіи двоякомъ [образѣ] произношится, иногда мяг-ко, а иногда твердо.

Запись: 756

Рисунок 14. Форма «Форма 2 Значения заголовочных слов».

Грамматические характеристики заголовочных слов								
№ статьи	№ столбца статьи	Заголовочное слово	Варианты	Части речи	Род	Число	Главные характеристики	Пометы
1	1	Г.						
1	2	ГА'	га', га'.					
1	3	ГАБА'	ба'.	с.	ж.			
1	4	ГАБАЬ	ба.	с.	ж.			Ним. Зри.
1	5	ГАБЪ	ба.	с.	м.			См. рязоуотпр.
1	6	Гѣхъсьѣ	гѣхъса, гѣхънут	гл.			ср.	
1	7	Гѣхъше	гѣхъ.	с.	ср.			См.
1	8	ГАГАРА	ра.	с.	ж.			

Рисунок 15. Форма «Форма_3_Грамматические характеристики заголовочных СЛОВ».

Разработка технологического процесса заполнения базы данных и автоматизированный ввод словарных статей.

При разработке технологического процесса заполнения базы данных были выделены этапы, процедуры и отдельные операции, представленные в таблице 6.

Этапы	Время мин	Процедуры	Время мин	Операции	Время мин
Подготовка словарных ста- тей к вводу в базу.	601	Установле- ние славян- ских шриф- тов.	3	Открытие папки Settings/Control Panel/Fonts. Открытие диа- логового окна Add Fonts: File/Install New Font.	1,5
				Для установки новых шрифтов выбрать файлы *.ttf из папки F:\шрифты раз- даточного мате- риала (CD- диск) и нажать кнопку ОК.	1,5
		Выделение цветом эле- ментов сло- варных ста- тей.	228	Открытие файла F:\2_0001- 0102г.RTF. Создание копии файла 2_0001- 0102г.RTF: «Исходный текст.RTF». В дальнейшем работа с копией документа.	1,5
				Редактирование информации в файле для удо- бного вида.	15
				Выделение цве- том элементов словарных ста-	210

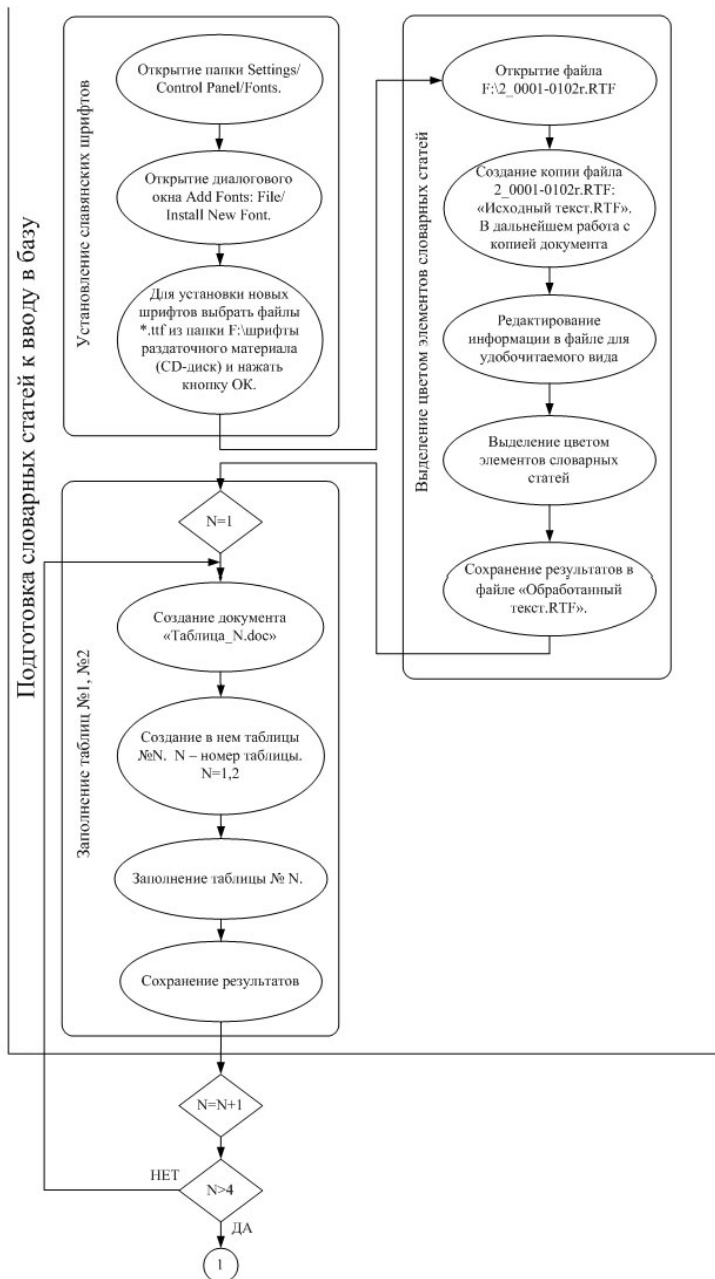
				тей.	
				Сохранение результатов в файле «Обработанный текст.RTF».	1,5
		Заполнение таблиц №1, №2 (операции повторяются циклически для таблиц 1 и 2).	370	Создание документа «Таблица_N.doc». Создание в нем таблицы №N. N – номер таблицы. N=1,2.	5*2
				Заполнение таблицы № N. Сохранение результатов.	180*2
Декомпозиция словарных статей	365	Создание текстового файла.	1	Создание документа «Таблица_N.doc».	1
		Заполнение таблиц. (операции повторяются циклически для таблиц 3 и 4).	364	Создание в «Таблица_N.doc» таблицы №N. N – номер таблицы. N=3,4.	2*2
				Заполнение таблицы № N. Сохранение результатов.	180*2
Создание базы данных в Microsoft Excel 2002.	281	Создание новой книги в Excel.	1	Создание новой книги в Excel «Таблицы.xls».	1
		Импорт данных из таблицы формата	280	Открытие таблицы №N из файла «Таблица_N.doc».	5*4

		Microsoft Word в таблицу Excel (операции повторяются циклически для таблиц 1,2,3,4).		Выделение всей области данных таблицы. Копирование ее в буфер.	
				Открытие чистого листа книги и переименование его в «Таблица_N». Установление курсора в ячейку A1. Копирование данных таблицы из буфера.	5*4
				Редактирование данных в Excel таблице. Сохранение изменений в книге «Таблицы.xls».	60*4
Создание базы данных в Microsoft Access 2002	220	Создание базы в Microsoft Access.	1	Создание новой базы в Microsoft Access под названием «База данных_001-102.mdb».	1

		Импорт Excel таблицы в Microsoft Access 2002. (операции повторяются циклически для таблиц 1,2,3,4).	100	Работа с диалоговым окном «Импорт». Указать тип файла — Excel формат. Выбрать файл «Таблицы.xls» Выбрать лист «Таблица_N» Ввести источник и требуемые параметры в условиях импорта базы из Excel.	5*4
				Редактирование данных в таблице Microsoft Access «Таблица_N». Сохранение изменений в базе.	20*4
		Создание форм базы данных. (операции повторяются циклически для форм 1,2,3,4).	120	Создание формы «Форма_N» с помощью мастера. Редактирование формы «Форма_N» в режиме конструктора. Сохранение изменений в базе.	30*4
ИТОГО		Затрачено времени на все операции: 1467 минут = 24 часов 27 минут			

Таблица 6. Технологический процесс заполнения базы данных.

Ниже на рисунке приведена схема описанного выше технологического процесса заполнения базы данных.



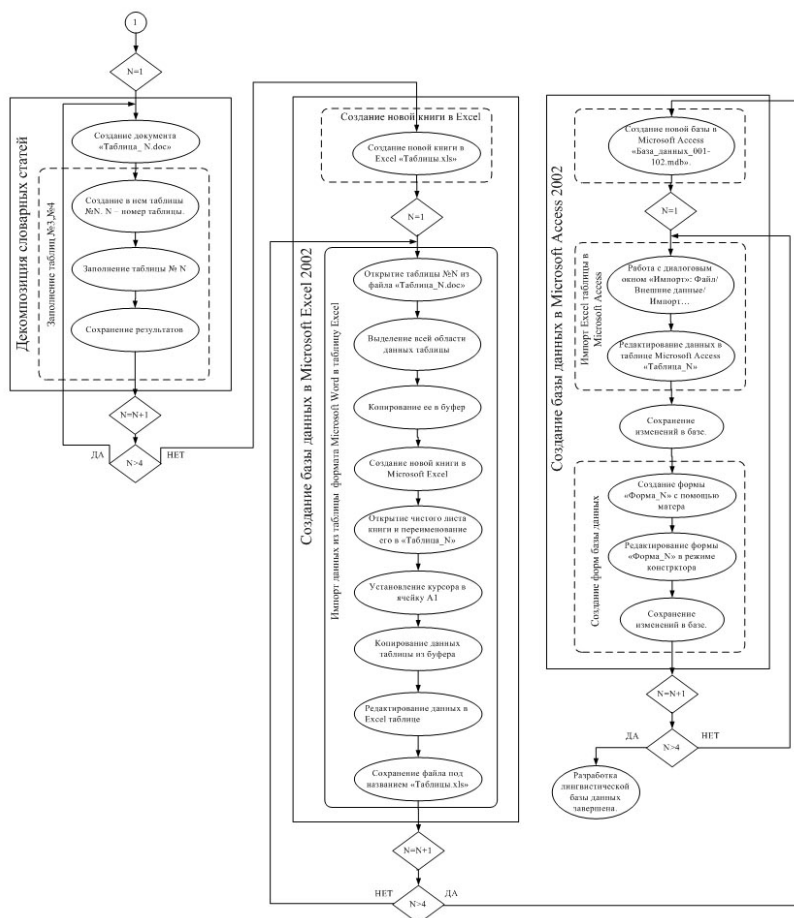


Рисунок 16. Схема технологического процесса заполнения базы данных.

Исследование эффективности технологического процесса заполнения базы данных

Исследование эффективности технологического процесса ввода информации состоит в определении основных характеристик — ресурсных затрат и ограничений на выполнение этапов, процедур и отдельных операций. Данное исследование проводилось в процессе реализации предыдущих заданий данной курсовой работы.

В таблице 6 представлены результаты исследования эффективности

технологического процесса заполнения базы данных. Результаты были получены путем фиксирования временных затрат на последовательное выполнение операций в процессе заполнения базы данных.

Вариант 5.

Исполнитель: студент группы ИУ5-103 А.В.Дон.

Исходным материалом является отрезок третьего тома Словаря Академии Российской (столбцы 103-202) представляющий собой совокупность из 610 словарных статей.

Разработка лингвистической базы данных

Разработана база данных в СУБД Microsoft Access 2003. База данных нормализована до третьей нормальной формы.

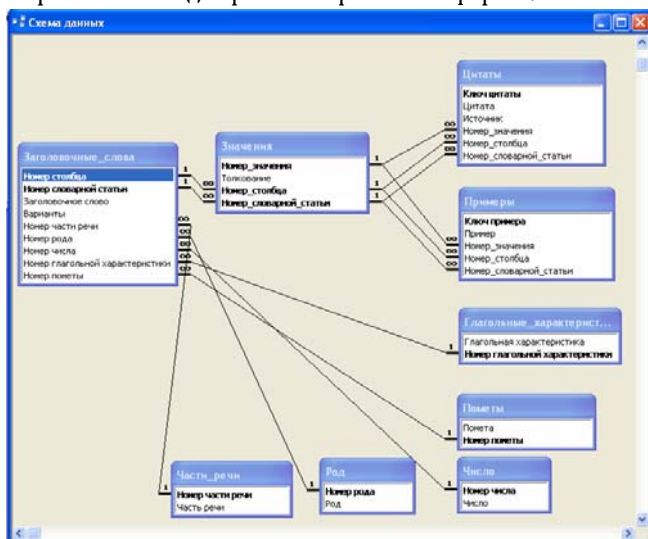


Рисунок 17. Схема базы данных.

Разработка технологического процесса заполнения базы данных и автоматизированный ввод словарных статей

Разработан технологический процесс заполнения базы данных, содержащий строго определённое количество ручных и автоматизированных операций.

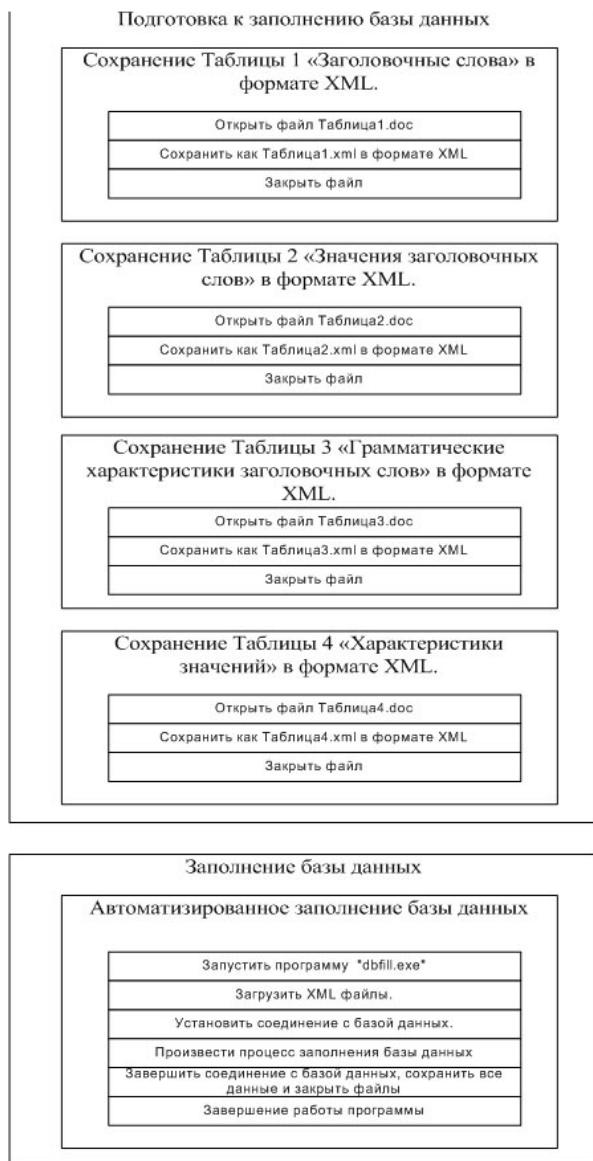


Рисунок 19. Схема технологического процесса.

Технологический процесс заполнения базы данных включает в себя следующие этапы:

1. Подготовка к заполнению базы данных

На этапе «Подготовка к заполнению базы» производятся следующие процедуры и операции:

1.2. Сохранение Таблицы 1 «Заголовочные слова» в формате XML.

- Открыть файл Таблица1.doc
- Сохранить как Таблица1.xml в формате XML
- Закрыть файл

1.2. Сохранение Таблицы 2 «Значения заголовочных слов» в формате XML.

- Открыть файл Таблица2.doc
- Сохранить как Таблица2.xml в формате XML
- Закрыть файл

1.3. Сохранение Таблицы 3 «Грамматические характеристики заголовочных слов» в формате XML.

- Открыть файл Таблица3.doc
- Сохранить как Таблица3.xml в формате XML
- Закрыть файл

1.4. Сохранение Таблицы 4 «Характеристики значений» в формате XML.

- Открыть файл Таблица4.doc
- Сохранить как Таблица4.xml в формате XML
- Закрыть файл

2. Заполнение базы данных

Этап «Заполнение базы данных» содержит следующие процедуры и операции:

2.1. Автоматизированное заполнение базы данных

Запустить программу автоматизированного заполнения базы данных “dbfill.exe”. Появится окно программы:

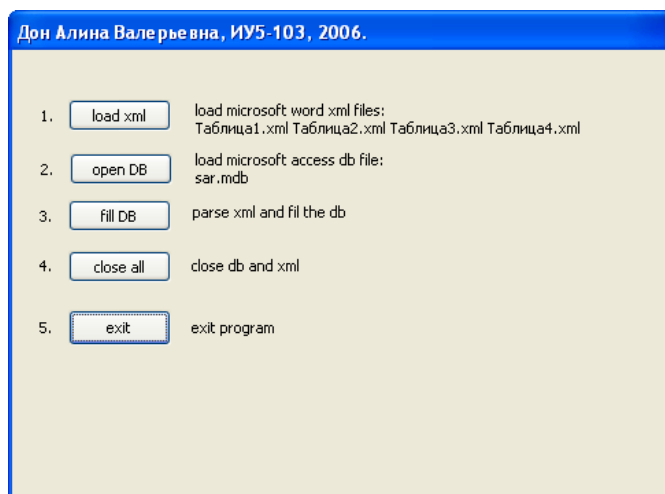


Рисунок 20. Окно программы автоматизированного заполнения базы данных.

- Нажать кнопку 1 — загрузить XML файлы.
- Нажать кнопку 2 — установить соединение с базой данных.
- Нажать кнопку 3 — произвести процесс заполнения базы данных.
- Нажать кнопку 4 — завершить соединение с базой данных, сохранить все данные и закрыть файлы.
- Нажать кнопку 5 — завершение работы программы.

Исследование эффективности технологического процесса заполнения базы данных

Исследование производилось в процессе предыдущих заданий.
Установленные ресурсные затраты и ограничения (таблица 7):

Наименование	Ресурсные затраты	Ограничения
Разработка схемы базы данных	1 человеко-неделя	
Разработка техпроцесса заполнения базы данных и программы для автоматизированного заполнения базы данных	3 человеко-недели	

Подготовка к заполнению базы данных	4 человеко-минуты	необходимо 8 Мб на диске
Сохранение Таблицы 1	1 человеко-минута	необходимо ещё 2 Мб на диске
Сохранение Таблицы 2	1 человеко-минута	необходимо ещё 2 Мб на диске
Сохранение Таблицы 3	1 человеко-минута	необходимо ещё 2 Мб на диске
Сохранение Таблицы 4	1 человеко-минута	необходимо ещё 2 Мб на диске
Заполнение базы данных	3-6 минут	необходимо ещё 10 Мб на диске, 25 Мб оперативной памяти
Запуск программы	0,5 мин	необходимо 8 Мб на диске, 4 Мб оперативной памяти
Загрузка XML файлов	0,5 мин	операционная система: Windows 2000/XP необходимо 20 Мб оперативной памяти
Установление соединения с базой данных	0,5 мин	необходимо 23 Мб оперативной памяти
Процесс заполнения базы данных	1-4 мин	необходимо 25 Мб оперативной памяти и ещё 2 Мб на диске
Завершение соединения с базой данных, сохранить все данные и закрыть файлы	0,5 мин	
Завершение работы программы	0 мин	

Таблица 7. Ресурсные затраты и ограничения.

Вариант 5.

Исполнитель: студент группы ИУ5-103 О.М.Антонюк.

Исходным материалом является отрезок третьего тома Словаря Академии Российской (столбцы 1103-1202) представляющий собой совокупность из 610 словарных статей.

Разработка лингвистической базы данных

Разработка базы данных для заданного отрезка САР. Среда разработки базы данных Microsoft Access (2003).

Разработка технологического процесса заполнения базы данных и автоматизированный ввод словарных статей

Этапы технологического процесса:

1. Подготовка аппаратных и программных средств
2. Подготовка отрезка САР
3. Создание базы данных
4. Заполнение базы данных

1. Этап «Подготовка аппаратных и программных средств» состоит из следующих процедур, которые подразделяются на операции:

- 1.1. Подготовка компьютера
 - 1.1.1. Воспользоваться своим либо компьютером в аудитории университета.
 - 1.1.2. Включить компьютер
- 1.2. Подготовить общесистемные программные средства
 - 1.2.1. Установить ОС, если отсутствует.
 - 1.2.2. Установить Microsoft Word, если не установлен
 - 1.2.3. Установить Microsoft Access, если не установлен

2. Этап «Подготовка отрезка САР» состоит из следующих процедур, которые подразделяются на операции:

- 2.1. Открыть Microsoft Word и установить шрифты
- 2.2. Выделение цветом элементов словарных статей
 - 2.2.1. Выделить желтым цветом заголовочные слова
 - 2.2.2. Выделить зеленым цветом грамматические характеристики и пометы
 - 2.2.3. Выделить бирюзовым цветом цитаты
 - 2.2.4. Выделить красным цветом шифр источника
- 2.3. Составление таблицы 1. «Заголовочные слова»
 - 2.3.1. Вычерчивание таблицы 1

- 2.3.2. Заполнение заголовка таблицы 1
- 2.3.3. Копирование данных из отрезка CAP и вставление их в соответствующие поля таблицы 1
- 2.4. Составление таблицы 2. «Значения заголовочных слов»
 - 2.4.1. Вычерчивание таблицы 2
 - 2.4.2. Заполнение заголовка таблицы 2
 - 2.4.3. Копирование данных из отрезка CAP и вставление их в соответствующие поля таблицы 2
- 2.5. Составление таблицы 3. «Грамматические характеристики заголовочных слов»
 - 2.5.1. Вычерчивание таблицы 3
 - 2.5.2. Заполнение заголовка таблицы 3
 - 2.5.3. Копирование данных из отрезка CAP и вставление их в соответствующие поля таблицы 3
- 2.6. Составление таблицы 4. «Характеристики значений»
 - 2.6.1. Вычерчивание таблицы 4
 - 2.6.2. Заполнение заголовка таблицы 4
 - 2.6.3. Копирование данных из отрезка CAP и вставление их в соответствующие поля таблицы 4
- 3. Этап «Создание базы данных» состоит из следующих процедур:
 - 3.1. Открыть Microsoft Access
 - 3.2. Создание таблицы 1
 - 3.3. Создание таблицы 2
 - 3.4. Создание таблицы 3
 - 3.5. Создание таблицы 4
 - 3.6. Создание связей между таблицами
 - 3.7. Создание форм
- 4. Этап «Заполнение базы данных» состоит из следующих процедур, которые подразделяются на операции:
 - 4.1. Заполнение таблицы 1 базы данных
 - 4.1.1. Заполнить вручную поле «ключ» в таблице 1 в Microsoft Access
 - 4.1.2. Скопировать таблицу 1 из Microsoft Word
 - 4.1.3. Вставить таблицу 1 в соответствующие поля в Microsoft Access
 - 4.1.4. Исправить ошибки при вставке таблицы 1
 - 4.1.5. Сохранить таблицу 1
 - 4.2. Заполнение таблицы 2 базы данных

- 4.2.1. Заполнить вручную поле «ключ» в таблице 2 в Microsoft Access
- 4.2.2. Скопировать таблицу 2 из Microsoft Word
- 4.2.3. Вставить таблицу 2 в соответствующие поля в Microsoft Access
- 4.2.4. Исправить ошибки при вставке таблицы 2
- 4.2.5. Сохранить таблицу 2
- 4.3. Заполнение таблицы 3 базы данных
 - 4.3.1. Заполнить вручную поле «ключ» в таблице 3 в Microsoft Access
 - 4.3.2. Скопировать таблицу 3 из Microsoft Word
 - 4.3.3. Вставить таблицу 3 в соответствующие поля в Microsoft Access
 - 4.3.4. Исправить ошибки при вставке таблицы 3
 - 4.3.5. Сохранить таблицу 3
- 4.4. Заполнение таблицы 4 базы данных
 - 4.4.1. Заполнить вручную поле «ключ» в таблице 4 в Microsoft Access
 - 4.4.2. Скопировать таблицу 4 из Microsoft Word
 - 4.4.3. Вставить таблицу 4 в соответствующие поля в Microsoft Access
 - 4.4.4. Исправить ошибки при вставке таблицы 4
 - 4.4.5. Сохранить таблицу 4

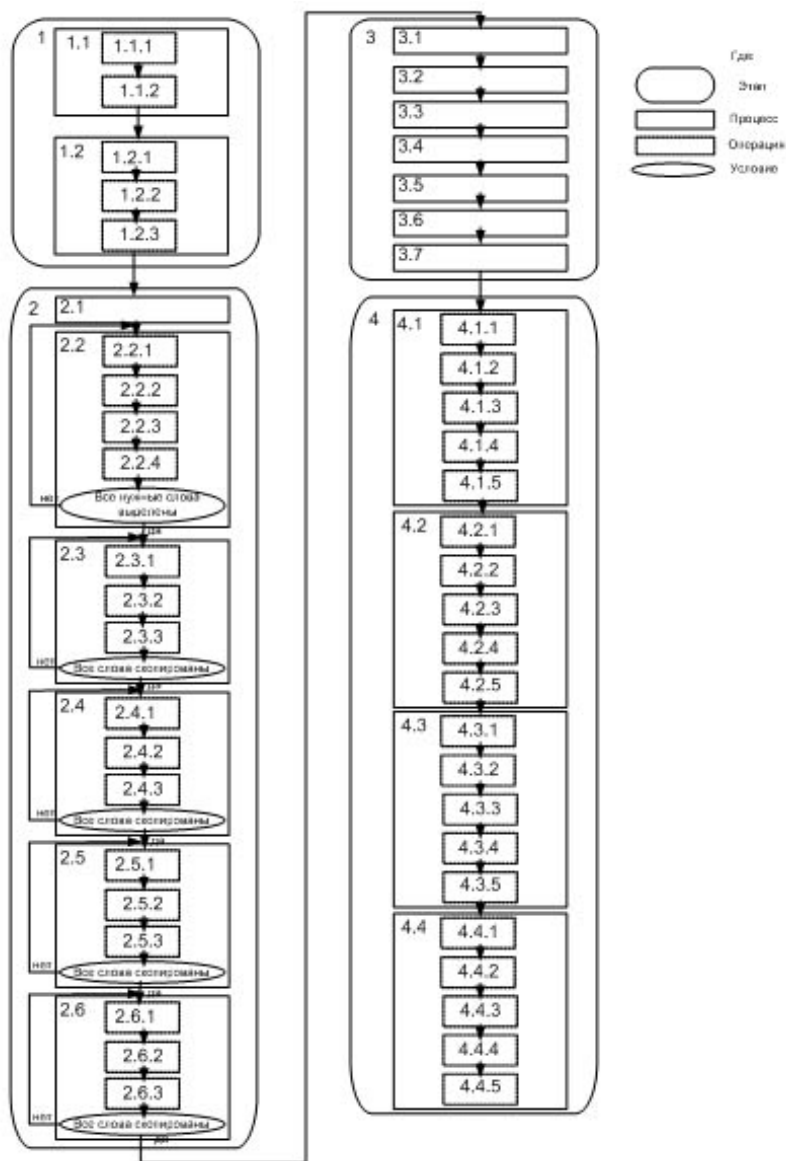


Рисунок 21. Схема технологического процесса.

Исследование эффективности технологического процесса ввода информации

Распределение временных затрат по этапам, процедурам и операциям приведено в таблице 8:

Название этапа	Временные затраты по этапу	Название процедуры	Временные затраты по процедурам	Название операции	Временные затраты по операциям
Подготовка аппаратных и программных средств	0 мин	Подготовить компьютер	0 мин	Воспользоваться своим либо компьютером в аудитории университета	0 мин
				Включить компьютер	0 мин
		Подготовить общесистемные программные средства	0 мин	Установить ОС, если отсутствует	0 мин
				Установить Microsoft Word, если не установлен	0 мин
				Установить Microsoft Access, если не установлен	0 мин
Подготовка отрезка САР	904 мин	Открыть Microsoft Word и установить	2 мин	-	
		Выделить цветом элементы словарных статей	90 мин	Выделить желтым цветом заголовочные слова	20 мин
				Выделить зеленым цветом грамматические характеристики и пометы	35 мин
				Выделить бирюзовым цветом цитаты	20 мин
				Выделить красным цветом шифр источника	15 мин
		Составление таблицы 1.	200,5 мин	Вычерчивание таблицы 1	20 мин
				Заполнение заголовка	0,5

		«Заголовочные слова»		таблицы 1	мин
				Копирование данных из отрезка САР и вставление их в соответствующие поля таблицы 1	180 мин
		Составление таблицы 2. «Значения заголовочных слов»	180,5 мин	Вычерчивание таблицы 2	20 мин
				Заполнение заголовка таблицы 2	0,5 мин
				Копирование данных из отрезка САР и вставление их в соответствующие поля таблицы 2	160 мин
		Составление таблицы 3. «Грамматические характеристики заголовочных слов»	235,5 мин	Вычерчивание таблицы 3	25 мин
				Заполнение заголовка таблицы 3	0,5 мин
				Копирование данных из отрезка САР и вставление их в соответствующие поля таблицы 3	210 мин
		Составление таблицы 4. «Характеристики значений»	195,5 мин	Вычерчивание таблицы 4	15 мин
				Заполнение заголовка таблицы 4	0,5 мин
				Копирование данных из отрезка САР и вставление их в соответствующие поля таблицы 4	180 мин
		Создание базы данных	95 мин	Открыть Microsoft Access	0,5 мин
				Создание таблицы 1	15 мин
				Создание таблицы 2	15 мин
				Создание таблицы 3	15 мин
				Создание таблицы 4	15 мин
				Создание свя-	4,5 мин

		зей между таб- лицами	мин		
		Создание форм	30 мин	-	
Запол- нение базы данных	152 мин	Заполнение таблицы 1 базы данных	40,5 мин	Заполнить вручную поле «ключ» в таблице 1 в Mi- crosoft Access	30 мин
				Скопировать таблице 1 из Microsoft Word	1 мин
				Вставить таблицу 1 в со- ответствующие поля в Microsoft Access	1 мин
				Исправить ошибки при вставке таблицы 1	8 мин
				Сохранить таблицу 1	0,5 мин
		Заполнение таблицы 2 ба- зы данных	40,5 мин	Заполнить вручную поле «ключ» в таблице 2 в Microsoft Access	30 мин
				Скопировать таблице 2 из Microsoft Word	1 мин
				Вставить таблицу 2 в соответствующие поля в Microsoft Access	1 мин
				Исправить ошибки при вставке таблицы 2	8 мин
				Сохранить таблицу 2	0,5 мин
		Заполнение таблицы 3 ба- зы данных	32,5 мин	Заполнить вручную поле «ключ» в таблице 3 в Microsoft Access	30 мин
				Скопировать таблице 3 из Microsoft Word	1 мин
				Вставить таблицу 3 в со- ответствующие поля в Microsoft Access	1 мин
				Исправить ошибки при вставке таблицы 3	0 мин
				Сохранить таблицу 3	0,5

					мин
		Заполнение таблицы 4 ба- зы данных	38,5 мин	Заполнить вручную поле «ключ» в таблице 4 в Microsoft Access	30 мин
				Скопировать таблице 4 из Microsoft Word	1 мин
				Вставить таблицу 4 в соответствующие поля в Microsoft Access	1 мин
				Исправить ошибки при вставке таблицы 4	6 мин
				Сохранить таблицу 4	
ИТОГО: 1151 минута = 19 часов 11 минут					

Таблица 8. Распределение временных затрат.

Основная литература

1. Тихонов А.Н. Первый гнездовой словарь и его роль в развитии русской лексикографии. // Словарь Академии Российской 1789-1794. Т. 5. — М.: МГИ им. Е.Р. Дашковой. 2005.

2. Чернышева М.И. Состав и структура Словаря Академии Российской. // Словарь Академии Российской 1789-1794. Т. 2. — М.: МГИ им. Е.Р. Дашковой. 2002.

3. Филиппович Анна. Электронная версия Словаря Академии Российской 1789-1794 годов. // Роль книгоиздания в развитии международных научных и культурных контактов: Материалы международной научной конференции (Москва, 21-23 сентября 2005 г.) / Сост. В.И.Васильев, М.А. Ермолаева, А.Ю. Самарин. — М.: Наука, 2005. — С. 293-296.

4. Филиппович А.Ю. Информационная технология создания электронного издания Словаря Академии Российской 1789-1794 гг. // Современные информационные технологии и письменное наследие: от древних рукописей к электронным текстам [Текст]: материалы междунар. науч. конф. (Ижевск, 13-17 июля 2006 г.) / отв. ред. В.А.Баранов. — Ижевск: Изд-во ИжГТУ, 2006. — С. 174-178.

Дополнительная литература

5. Григорьев Ю. А., Ревунков Г. И. Банки данных: Учеб. для вузов. — М.: Изд-во МГТУ им. Н. Э. Баумана, 2002. — 320 с.: ил.

6. Литвин П., Гетц К., Гунделой М. Разработка настольных приложений а Access 2002. Для профессионалов. — СПб.: Питер; К.: Издательская группа BHV, 2002. — 1008 с.: ил.

7. Бумфрей Ф., Диренцо О., Дакетт Й. и др. XML. Новые перспективы WWW. Пер. с англ. — М.: ДМК, 2000, - 688 с.: ил.

8. Шилдт Г. С#: учебный курс. — СПб.: Питер; К.: Издательская группа BHV, 2002. — 512 с.: ил.

9. Коннолли Т., Брегг К., Страчан А. Базы данных: проектирование, реализация и сопровождение. Теория и практика, 2-е изд. : Пер. с англ. — М. : Издательский дом «Вильямс», 2001. — с. 237-239.