

КИЕВСКИЙ ПОЛИТЕХНИЧЕСКИЙ ИНСТИТУТ

На правах рукописи

АЛЬ-АРЖА ВАДИА ИСКАНДЕР

гр. Ливана

УДК 681.3-371.3

ПРОБЛЕМНО-ОРИЕНТИРОВАННАЯ ВЫЧИСЛИТЕЛЬНАЯ  
СРЕДА ДЛЯ АВТОМАТИЗАЦИИ ИЗУЧЕНИЯ ИНОСТРАННОГО  
ЯЗЫКА

Специальность 05.13.13 - Вычислительные машины,  
комплексы, системы и сети.

А В Т О Р Е Ф Е Р А Т

Диссертации на соискание ученой степени  
кандидата технических наук

Киев -1992г.

04.03



00360358 (P)

Работа выполнена на к... систем и сетей Киевского политехнического института.

Научный руководитель - кандидат технических наук, профессор  
Корнейчук В. И.

Официальные оппоненты: доктор технических наук, профессор,  
в. н. с. института информатики АН Украины  
Врюхович Е. И.  
кандидат технических наук, доцент,  
с. н. с., зав. лабораторией информатики  
КПИ Пурин О. Ф.

Ведущая организация: НПО УКРЕВУЗ информатики.

Защита состоится "21" декабря 1992г. в 14<sup>30</sup> часов на заседании специализированного совета Д 068.14.09 в Киевском политехническом институте.

Отзывы на авторефераты (в двух экземплярах), заверенные печатью учреждения, просим направлять по адресу:

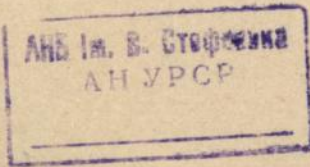
252056, г. Киев-56, проспект Победы, 37, КПИ, Ученому секретарю.

С диссертацией можно ознакомиться в библиотеке института.

Автореферат разослан "20" ноября 1992г.

Ученый секретарь  
специализированного совета  
доктор технических наук, профессор

Бузовский О. В.



**А Н Н О Т А Ц И Я**

Целью реферируемой диссертационной работы является исследование и разработка алгоритмов и структур специальных вычислительных средств, реализующих процесс накопления, свертки и распознавания иноязычной лексической информации в обучающей системе.

Для достижения поставленной цели диссертации решаются следующие основные задачи:

1. Анализ алгоритмов реализации процесса машино-организованного изучения иноязычной лексики и разработка структуры базового ядра обучающе-тренирующей вычислительной системы.

2. Разработка методов и вычислительных средств накопления и распознавания формализованных сообщений обучаемых по запросам системы обучения.

3. Разработка структур специальных лингвистических процессов, реализующих поиск данных в словарях обучающих систем для изучения иностранного языка.

4. Разработка алгоритмов, обеспечивающих взаимодействие информационных компонентов при поиске словарных статей и их практическая апробация в среде персонального компьютера.

На защиту выносятся следующие основные положения и результаты:

1. Структуры организации и способ формализации представления и распознавания конструкций, составляющих ответ обучаемого и позволяющих рационализировать процесс диагностики освоенных пользователями знаний.

2. Способы организации и структуры специализированных вычислителей для обработки лингвистической информации в словарях АОО, обучающих иностранной лексике и позволяющие снизить затраты памяти и повысить скорость поиска искомого данных, за счет компактного представления данных на иностранном языке.

3. Алгоритмы и программы, позволяющие осуществить эффективный диалог в ходе учебного перевода.

**ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ**

Актуальность темы.

Процесс обучения человека представляет собой одну из жизненно обусловленных стадий существования личности. В зависимости от целе-

внх функций, возникающих перед этой личностью, уровня развития и процесса познания, постоянно усложняющего процесс обучения (с позиции глубины, информационного объема, а также времени, отводимого для процесса усвоения), складываются различные требования к системе образования. Но какой бы сложной или простой она не была, для любого общества останется актуальной проблема взаимодействия людей на разных языках, в процессе которого и происходит аккумуляция человеческих знаний.

Поэтому процесс обучения иностранному языку никогда не потеряет своего значения. Однако, надо отметить, что для этой области характерны общие проблемы системы образования, которые обуславливают необходимость интенсификации процесса обучения, повышение степени его индивидуализации, расширение объемов информации, представляющей модели предметного взаимодействия в иной лингвистико-культурной среде.

Решению этих проблем с использованием кибернетических технологий посвящены работы многих исследователей и ученых, таких как А. Я. Савельев, А. Н. Довгялло, А. А. Стогний, А. Ф. Чернявский, Е. Л. Юценко, Л. А. Растринин, Л. Н. Новицкий, Л. В. Нещецкий, В. Д. Самойлов, R. Atkinson, D. Bitzer, J. Carbonell, E. Coffman, L. Osin, G. Pask, S. Pipert и др.

Результаты научной деятельности их привели к решению ряда практических задач в области машинного обучения.

Количество автоматизированных систем, используемых для проведения учебных занятий, растет лавинообразно.

Так следует отметить системы типа COURSEWRITER, PLATO, TICCT, разработки фирмы APPLE (США), АОО семейства СПОК, КОНТАКТ, ДИСК. Тем не менее остается актуальной задача автоматизации обучения понимания текстов на иностранных языках, способности осуществлять профессиональную предметно-ориентированную деятельность в иной язычной и культурной среде.

Таким образом, возникает необходимость в разработке специальной системы, обладающей широкими возможностями по накоплению иноязычной лексики, способной обеспечить приемлемые временные параметры и являющейся ядром для комплекса обучающе-вычислительных средств, реализующих все концепции изучения иностранного языка.

Методы исследований базировались на основных положениях, теории цифровых вычислительных машин, теории алгоритмов, методах программированного обучения, теории вероятностей и математической ста-

истики.

Научная новизна состоит в разработке новых способов организации структур аппаратных средств и алгоритмов позволивших создать необходимого по объему поисковой части и достаточного по функциям ядра обучающе-вычислительной системы для изучения иноязычной лексики, позволяющего манипулировать со сжатой и формализованной учебной информацией, состав которой дает возможность реализовать любой из известных типов обучающего диалога (инструктивный, справочно-консультирующий, продуцирующий, контролирующий). В процессе работы над диссертацией разработаны:

- методы и структуры вычислительных средств для представления и распознавания формализованных сообщений обучаемого при отработке навыков в иноязычной лексике, отличающиеся простотой и повышенной скоростью распознавания;

- способы и структуры вычислительных средств сжатия исходной информации позволяющие хранить и расширять объемы базовой учебной информации в автоматизированной обучающей системе для изучения иностранных языков;

- алгоритмы и программы реализации ядра обучающей системы, оформленный как самостоятельная информационно-справочная многоязычная АСС.

Практическая ценность работы представляется разработкой ядра обучающе-вычислительной системы для изучения иноязычной лексики, применение которого снижает трудозатраты на программирование исходной базы и повышает степень свободы пользователя при формулировке своего ответа или запроса к системе.

- Разработкой и реализацией информационно-справочной автоматизированной обучающей системы "TRADUCTION", которая служит в качестве АРМа - справочника для специалистов, работающих с иноязычными тематическими текстами.

Реализация научных результатов работы. Результаты исследований полученные в диссертационной работе могут быть использованы при реализации вопросно-ответных систем с высокой степенью структуризации входного сообщения, а методы хранения информации могут быть применены для эффективного построения ассоциативных ЗУ в процессорах с лингвистической обработкой данных. рассмотренные на основе предложенных подходов обучающие системы различного типа могут применяться для организации учебного процесса в высших и средних учебных заведениях всех типов.

#### Апробация работы.

Основные научные результаты диссертационной работы докладывались и обсуждались на научно-технической школы-семинаре "Теория и

практика построения функционально-ориентированных вычислительных и микропроцессорных систем обработки информации", Каменец-Подольский, 19-22 июня, 1990г.; Всесоюзной научно-технической школе "Проблемы памяти ЭЕМ", Уфа, 3-8 сентября, 1990г.; Всесоюзной научно-технической школе "Устройства и системы хранения информации", Алушта, 1-8 октября, 1991г.

#### Публикации.

Основное содержание работы опубликовано в 7 печатных работах.

Структура и объем работы. Диссертационная работа состоит из введения, четырех глав, заключения, списка литературы и приложения. Основной текст диссертации изложен на 190 страницах машинного текста и содержит 3 таблицы и 16 рисунков.

Во введении основан выбор темы диссертации и сформулированы цели исследования.

В первой главе диссертации выполнен аналитический обзор существующих типов автоматизированных обучающих систем различного назначения и оценена их применимость для целей изучения иностранного языка. В соответствии с результатами анализа построена классификационная схема существующего спектра методов машинного обучения и сделан выбор составляющих ядра обучающей системы для изучения иноязычной лексики.

Здесь же сделан вывод о необходимости компактного представления данных, входящих в состав словаря, справочника, переводной информации и контрольной базы. В связи с этим проведен обзор существующих средств архивации и сжатия данных и принято решение о необходимости создания специальных вычислительных компонентов, управляющих режимами поиска лексической информации по сжатому тексту или специальным моделям.

Во второй главе сконцентрированы решения, связанные с рационализацией представления контрольной информации в информационной среде для изучения иноязычной лексики. Разработаны структуры специализированных вычислителей, позволяющие значительно сократить объемы памяти, необходимые для хранения служебной информации и снизить дублирование информации в базе контролируемых конструкций.

Третья глава диссертации посвящена исследованиям в области построения словарей иноязычной лексики. Рассмотрены различные подходы к поиску данных в этих словарях и предложены методы и структуры вычислителей, позволяющих снизить объем хранения и время поиска перевода за счет компактного представления словарных единиц на исходном

языке.

В четвертой главе рассмотрена спроектированная с учетом полученных в ходе исследований результатов информационно-справочная АОС "TRADUCTION", позволяющая обучаемому получить перевод на любом иностранном языке, а также справку по порядку, форме и особенностях использования и перевода данной лексической единицы. Предложена модель, реализации стратегии обучения с помощью подобной АОС иноязычной лексики.

В приложениях предоставлены листинги и примеры работы программных средств, а именно: АОС "TRADUCTION" и программы моделирования стратегии обучения.

#### СОДЕРЖАНИЕ РАБОТЫ.

Анализ существующих АОС общего назначения и систем, предназначенных для изучения иностранного языка, позволил сформулировать перспективные направления в построении специализированных машинных обучающих систем, позволяющих создать комплексное средство поддержки изучения иноязычной лексики, используемое на любых этапах и стадиях обучения. Это позволило оценить возможность представления учебной информации в сжатом виде, для чего проанализированы действующие архиваторы и существующие стандартные утилиты. Было установлено, что наиболее эффективным подходом к решению этой проблемы является комбинированная программно-аппаратная реализация средств сжатия. Выявлен основной недостаток известных средств, заключающийся в необходимости предварительной раскодировки информации. В результате предложена структура системы для поддержки процедуры изучения иностранного языка, отличающаяся от известных наличием компонентов, способных провести поиск информации в свернутых массивах и обеспечить единое методическое и информационное поле с родовидовыми связями, создающее возможность реализации адаптивного сценария осуществления режима (консультации, обучения, справки, контроля). Структура такой системы приведена на рис. 1 и содержит следующие компоненты:

дидактический блок, обеспечивающий фиксацию статистических характеристик процесса обучения: количество обращений к словарю, число ошибочных ответов, время продвижения по предложенному сценарию, количество обращений за грамматической помощью, наименований понятий, разделов, тем и других структурных единиц учебного материала, на которых было допущено максимальное число ошибок и/или число обращений

Структура проблемно-ориентированного ядра вычислительной среды для автоматизации изучения иноязычной лексики

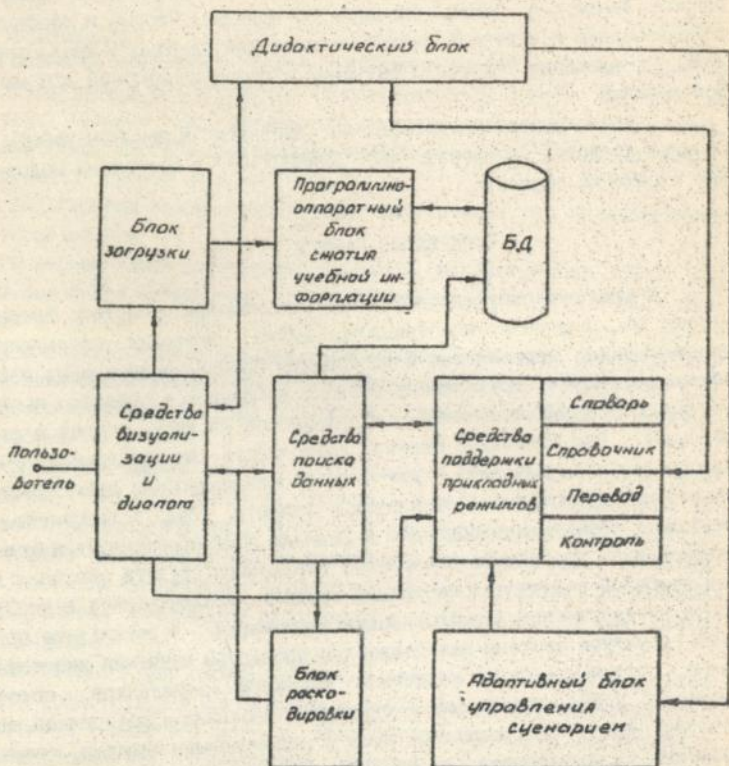


Рис. 1.

за разъяснением и т. п. ;

программно-аппаратный блок сжатия учебной информации и собственно базу данных (БД), реализующие и поддерживающие хранение данных в форме, обеспечивающей размещение значительных по объему справочников, словарей, инструктивного текстового материала, графически и с такой организационной структурой, которая позволяет устанавливать логические, структурные, ассоциативные и иные связи, необходимые для создания гипертекстовых режимов манипулирования в локальном физическом объеме;

средства поддержки прикладных режимов, создающих единометодическое поле и общую информационную среду при организации самостоятельного изучения иностранного языка на основе обеспечения пользователя словарным минимумом, консультацией по грамматическим и лексическим справочникам, подсказкой при переводе контроля при тренажировании собственных навыков;

адаптивный блок управления сценарием выбранного режима, позволяющий создать для обучаемого комфортные условия взаимодействия за счет движения по пути, наиболее выгодному для пользователя в контексте времени, темы, грамматических особенностей и т. п. ;

средство поиска, обеспечивающих эффективный поиск данных в свернутых форматах без раскодировки и визуализации;

блок раскодировки, средства визуализации и диалога и блок фрагментации, обеспечивающие контакт с пользователем (автором курса или обучаемым) и передачу информации: служебной, вспомогательной, учебной.

Система из обычной информационно-поисковой превращается в средство дидактического воздействия на обучаемого, когда в ее составе появляются средства, позволяющие распознавать вводимое обучаемым сообщение на предмет его истинности, т. е. на соответствие некоторому соглашению в модели знаний. Любая автономная информационно-обучающая система массового применения, как уже указывалось, должна содержать информационную, поисково-предъявительную и обучающую части, последняя из которых содержит аппарат анализа и оценки ответов обучаемых.

С точки зрения реализации естественно-языкового интерфейса человек-ЭВМ, что чрезвычайно важно для систем изучения иноязычной лексики, АОС можно рассматривать как вопросно-ответную систему. Основой для такого заключения является их сходство при анализе основных компонентов. Так в состав вопросно-ответной системы обычно включают : монитор, управляющий работой; базу знаний, содержащую ин-

формацию о модели предметной области и модели управления формирования ответа на запрос; лингвистический процессор или обработчик запросов, формирующий обращение к базе данных; база данных, содержащая информацию о фактах в выбранной предметной области.

Важной компонентой развитых вопросно-ответных систем является средство семантического анализа естественно-языковых текстов.

В понимании естественно-языковых текстов весьма важным является также проблема сегментации текстов, т. е. разбиение текста на такие минимальные смысловые конструкции, к которым применимы базовые смысловые формулы. В естественном языке минимальной смысловой конструкцией является предложение. Однако предложение само может включать как соподчиненные, так и независимые смысловые части, т. е. может быть составной конструкцией.

Диалог обучаемого с АОС предполагает вопросно-ответную ситуацию, когда системе естественно ожидать в ответе обучаемого раскрытия определенного смысла, заданного вопросом, ограниченного объема лексем в ответе, с большой точностью соответствующих лексемам, ожидаемым по заданному вопросу.

Итак, в АОС задача смыслового анализа текстов решается в условиях конкретного контекста. Любая предметная область содержательного состоит из конкретных понятий и отношений между этими понятиями. Но множество конкретных понятий и отношений можно разбить на конечное число типов понятий и типов отношений. Каждое осмысленное предложение предметной области можно перевести в текст, составленный из типов понятий и типов отношений, без детального учета грамматических особенностей, отображая каждое понятие или отношение в соответствующий тип.

В предметной области, очер-едными задачами научения иностранному языку, могут быть построены связи-отношения перекрывающие все проблемные области. Ограничимся информационно-обучающей частью этих отношений, базирующихся на словаре.

При формировании процедуры обратной связи, т. е. взаимодействия с обучаемым, в АОС для обучения иноязычной лексики возможны следующие дидактические ситуации:

- распознавание слов, в виде перевода одного из предложенных, вставки в контекст предложения, подбора синонима и т. п;
- ввод ответа в виде законченной фразы по предложенному перечню слов-семантических единиц, когда порядок этих единиц определяет достижение определенного контекста;

- свободное конструирование ответа, где контекст складывается из самостоятельно подобранных семантических единиц, участие которых в предложении означает распознавание обучаемым ситуации;

- комбинацию из перечисленных.

Введем систему отношений, моделирующих изложенные выше дидактические ситуации, которые могут порождаться на основе хранения в АОС в качестве базы-словаря.

Для распознавания семантической единицы-слова, адекватного словарному эквиваленту, система должна различить набор символов  $\{s_i, i \in \overline{I}\}$ , где  $I$  - максимально возможная длина слов в словаре,  $s_i \in \{\theta\}$ , где  $\theta$  - соответствующий алфавит.

Тогда, "правая" часть словаря может быть представлена, как набор  $\{s_i, i \in \overline{I}\}_j$ , где  $j \in \{\overline{J}\}$ , а  $J$  - мощность словаря.

Представленные выше дидактические процедуры, учитывая пояснение к ним, могут моделироваться двумя основными отношениями, а именно, отношением перестановки семантических единиц, обозначаемую как  $\&$  и операцией фиксации структуры предложения, обозначаемую как  $\perp$ . Тогда, представленные дидактические ситуации могут быть формализованы следующим образом:

1. Распознавание  $\{s_i, i \in \overline{I}\}_j, j \in \overline{J}$  либо  $\bigvee_{j \in \{l, m, n\}} \{s_i, i \in \overline{I}\}_j$ , где  $\bigvee$  - означает "ИЛИ", а  $l, m, n$  - номера синонимов в словаре.

2. Распознавание конструкции:  $\bigwedge_{j \in \{\overline{J}\}} \{s_i, i \in \overline{I}\}_j$ , где каждая семантическая единица имеет фиксированное место.

3. Распознавание конструкции:  $\&_{j \in \{\overline{J}\}} \{s_i, i \in \overline{I}\}_j$ , где семантические единицы только опеределают контекст.

4. Распознавание комбинированной ситуации типов:

$\perp_{j \in \{\overline{J}\}} (\& \{s_i, i \in \overline{I}\}) ; \&_{j \in \{\overline{J}\}} (\perp \{s_i, i \in \overline{I}\})$ .

Следует отметить, что распознавание одной семантической единицы, т.е. фактически 1-й ситуации, сводится к реализации конструкции во втором случае, если смысловая конструкция состоит из одного символа  $s_i$ .

Порождение вопросительных конструкций в АОС сопряжено с созданием достаточно объемного текстового блока, в котором фактически присутствуют семантические единицы из уже один раз записанного слс-

вара. Поэтому задача свертки вопросительных конструкций представляется актуальной с позиции экономии памяти, а построение соответствующего аппарата свертки и распознавания, как "эталонной" информации, так и ответа, представляется целесообразным с позиции упрощения процедуры анализа. Последняя операция может существенно отразиться и на времени реакции системы. Предлагается аппаратно организованные средства реализующие варианты свертки и распознавания вопросительных конструкций в АОС для изучения иноязычной лексики. В реферируемой работе разработаны схемотехнические решения, реализующие описанные те ситуации, позволяющие сократить аппаратные затраты на кодирование ответа обучаемого, а так же облегчить обучаемому и преподавателю диагностику причин ошибок.

Ошибка в формировании ответа на запрос машины непосредственно в элементе ответа, каковым является слово, приводит к неверному заключению о знаниях обучаемого и сводит к минимуму эффект применения АОС в обучении языком. Поскольку возможность такой ошибки исключить нельзя, необходимо попытаться предохранить обучаемого от ситуаций, связанных со случайностью, порождающей однократную ошибку при вводе ответа. В любом естественном языке однократная ошибка может быть трех видов: неправильная буква, пропущенная буква, лишняя буква. Можно предложить следующий способ учета такого рода ошибок при ответе обучаемого. Пусть  $n$  - максимальное число символов, которые может содержать ответ. Обозначим слова, выбираемые из словаря для сравнения

$$A = a_1 a_2 \dots a_n,$$

а вводимые - как

$$B = b_1 b_2 \dots b_n.$$

Если число букв в словах  $A$  и  $B$  меньше  $n$ , то оставшиеся позиции заполняются символами пробелов. Результат сравнения  $A$  и  $B$  представим в виде матрицы сравнения  $C$ , элементы которой

$$C_{i,j} = \begin{cases} 1, & \text{если } a_i = b_j \\ 0, & \text{если } a_i \neq b_j \end{cases}$$

где  $i, j = 1, 2, \dots, n$ .

Если слова  $A$  и  $B$  совпадают, то все элементы матрицы  $C$ , стоящие на главной диагонали будут равны единице. Для представления всех вариантов однократных ошибок типа неправильная, пропущенная или лишняя буква требуется определить значения элементов матрицы  $C$ , стоящих на главной и двух ближайших, параллельных ей диагоналях. Остальные элементы полагаем равными нулю:

$$C = \begin{pmatrix} C_{11} & C_{12} & 0 & \dots & 0 & 0 \\ C_{21} & C_{22} & C_{23} & \dots & 0 & 0 \\ 0 & C_{32} & C_{33} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & & C_{n-1, n-1} & C_{n-1, n} \\ 0 & 0 & 0 & & C_{n, n-1} & C_{n, n} \end{pmatrix}$$

Тогда результат сравнения можно представить для совпадения слов:  $S1 = C_{11} C_{22} \dots C_{nn} = \bigwedge_{\kappa=\lambda} C_{\kappa\kappa}$ ; для неправильной буквы:

$$S2 = \bigvee_{i=\lambda} (C_{11} \dots C_{i-1, i-1} C_{i, \lambda} C_{i+1, i+1} \dots C_{nn}) = \bigvee_{i=\lambda} (\bigwedge_{\kappa=\lambda, i \neq \kappa} C_{\kappa\kappa});$$

$$\text{для лишней буквы: } S3 = \bigvee_{i=\lambda} (C_{11} \dots C_{i-1, i-1} C_{i, i+1} \dots C_{n, n-1}) = \\ = \bigvee_{i=\lambda} [( \bigwedge_{\kappa=\lambda} C_{\kappa\kappa} ) ( \bigwedge_{\kappa=i+1} C_{\kappa, \kappa-1} )];$$

$$\text{для пропущенной буквы } S4 = \bigvee_{i=\lambda} (C_{11} \dots C_{i-1, i-1} C_{i, i+1} \dots C_{n-1, n}) = \\ = \bigvee_{i=\lambda} [( \bigwedge_{\kappa=\lambda, i < \kappa} C_{\kappa\kappa} ) ( \bigwedge_{\kappa=i+1} C_{\kappa-1, \kappa} )]$$

Поиск можно считать результативным, если

$$S = \overline{S1VS2VS3VS4} = 1$$

То есть, если при просмотре элементов словаря будут обнаружены слова, обращающие при сравнении в "1" функции S2, S3, S4, то при условии S1=0 необходимо проанализировать адреса этих слов на список адресов эталонных элементов ответа. Этот же подход можно использовать при поиске перевода или словарных статей в автоматизированном словаре АСС при условии ошибочного формирования вопроса. Схема такого устройства разработана в реферируемой работе.

Особенностью АСС для изучения иностранного языка является необходимость поддерживать большие объемы текстовой информации, которые невозможно или функционально нецелесообразно переводить в долговременную память, т.е. организовывать оверлейную структуру.

Способ представления информации в АСС во многом предопределен методом поиска этой информации в АСС.

Для решения задачи поиска слова в словаре существуют три основных подхода: последовательный поиск "в лоб"; сравнение слов с использованием алфавитного или числового порядка (бинарный поиск); поиск на основе ключ-адресных преобразований (хэширование); поиск, основанный на представлении слова как последовательности символов и построения посимвольного дерева словаря.

Недостатки перечисленных методов (нерациональное использование памяти, наличие лишних ветвей в алгоритме поиска информации и др.) могут быть устранены. Его можно сравнить с автоматом, имеющим несколько уровней состояний, причем каждому уровню однозначно соот-

ответствует количеству букв слова, введенных обучаемым на данный момент. Важно, что переход из одного состояния в другое на одном и том же уровне невозможен (каждое место в слове может быть занято только одной буквой одновременно), а переход в состояние следующего уровня ("перешагнуть" через уровень также невозможно) осуществляется лишь в том случае, если существует связь между уже введенным набором букв и соответствующей буквой. В противном случае обучаемому на этом же уровне предъявляется сообщение об отсутствии в лексическом запасе АСС слова, начинающегося с введенного набора букв. Таким образом, количество состояний, пройденных автоматом во время поиска слова, равно количеству букв введенного слова. Такой алгоритм реализует по сути поиск, основанный на построении посимвольного дерева словаря.

Для реализации метода "дерева" при построении поисковой части словаря, составляющего основу обучающей системы для изучения иностранного языка, предлагается структурировать информацию о символах, входящих в состав загружаемых слов. На основе изложенного подхода разработана структура вычислителя, позволяющая значительно сократить затраты памяти на хранение исходного лексического запаса АСС.

Итак, метод "дерева букв" делит буквы всех слов на несколько уровней, каждый из которых содержит буквы, стоящие в слове на месте, совпадающим с номером уровня. В результате, начальные буквы как последовательности начальных букв, общие для нескольких слов, упоминаются лишь один раз. Но тогда слова с одинаковыми окончаниями сокращению не поддаются. Для повышения степени сжатия поисковой части словаря можно предложить новый подход, основанный на установлении структурно-логической связи между набором пространственно распределенных символов.

Учебный перевод сталкивается с проблемой поддержки тематического словаря, лексический запас в котором представлен фактически неупорядоченными словами. В этом случае эффективность "дерева" значительно снижается и сжатие словаря практически обуславливается только начальными буквами слов. Так, для учебного лексического запаса по теме: "Погода" предлагается следующий набор: Ветер, видимость, град, гром, гроза, давление, дождь, жара, молния, мороз, радуга, снег, туман, туча.

Используя предложенную методику дерево "словаря", приведенных лексических единиц будет выглядеть аналогично рис. 2. а. При этом количество занятых ячеек будет равно 63. Предлагается придать алгоритму свертки прототипов лексических единиц свободу в 4-х направлениях

плоскости. В этом случае степень сжатия может быть повышена за счет использования соседних символов, входящих в суффиксы и окончания иных лексических единиц. Такой вариант представления приведенного выше список представлен на рис. 2 б. Он требует 46 ячеек памяти, т. к. в 1,32 раза меньше, чем в устройстве построенном по способу "дерева". Для реализации предложенного способа сверток "дерева" предположим, что ячейка накопителя кроме информации о символе содержать служебную информацию о начале слова, конце слова и занятости ячейки символом.

Структура устройства, реализующего поиск прототипа лексической единицы для выбора кода перевода, представлена в реферируемой работе. Заполнение накопителя прототипов лексических единиц производится экспертом с помощью блока визуализации.

Для моделирования полученных в ходе диссертационных исследований результатов и их практической реализации построена оригинальная информационно-справочная АСС, получившая название "TRADUCTION".

Для реализации системы выбран достаточно мощный язык С, использование которого делает ППП открытым для внесения соответствующих времени модификаций, а также позволяет при необходимости подключить мощности известных систем управления базами данных, таких как d Vista (в которой оговариваются шагом входа для Си реализаций), СУБД ODA, в которой язык программирования создан на основе объектно-ориентированного языка Си++; СУБД O<sub>2</sub>, в которой имеется прекомпилятор с языка CO<sub>2</sub> в язык Си и др.

Система предназначена для функционирования в среде MS DOS версии 3.3 и выше и ориентирована на ПЭВМ класса IBM PC XT/AT и совместных с ними модели, что обеспечивает массовость ее применения. Целью функционирования системы является рационализация процесса учебного перевода за счет использования тематически организованных словарей, грамматического справочника с гипертекстовыми связями и развитой помощи в процессе перевода.

Система обеспечивает четыре режима работы:

собственно перевод с одного языка на другой (при этом обрабатываются текстовые файлы, подготовленные стандартными текстовыми редакторами);

формирование базы данных, содержащей справочную информацию по каждому из иностранных языков;

генерация словаря перекрестных ссылок между различными видами справочной информации (по каждому слову или словосочетанию доступно



3 вида справочной информации);

взаимодействие с операционной системой ПЭВМ (чтение и сохранение файлов, просмотр файлов на экране и их распечатка, временный выход в операционную систему). Система положительно зарекомендовала себя в ходе испытаний.

#### ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ.

1. Рассмотрены особенности машинных методов обучения применительно к изучению иноязычной лексики, что позволило сформулировать требования к структуре базового состава информационно-справочной автоматизированной обучающей системы (АИСОС). Одним из основных отличий АИСОС является применение компонента поиска учебных данных в свернутой информационно-методической среде, что позволяет значительно расширить возможности системы, используя его совместно с иными типами (информационно-обучающими, гипертекстовыми, экспертными и т. п.) обучающими компонентами.

2. Сформулирована методика при анализе обращения обучаемого к системе, отличающаяся от известных упрощенным подходом к оценке поведения обучаемого и позволяющая хранить в свернутом виде возможные (эталонные) конструкции ответов.

3. Исследованы и разработаны алгоритмы и структуры специализированных вычислителей предназначенных для свертки, хранения, поиска и сравнения эталонных конструкций возможных сообщений обучаемых.

4. Исследованы и разработаны вычислительные устройства, которые позволяют свернуть исходное представление словарных статей и сократить в 2-4 раза представление в машине информации, являющейся по мнению обучаемого не содержательной.

5. С учетом полученных рекомендаций по компоновке структуры АИСОС разработана и реализована для ПЭВМ класса IBM PC/AT на языке СИ обучающая система, функции которой позволяют реализовать весь спектр возникающих учебных ситуаций в процессе учебного перевода. Система работает в 4-х основных режимах, реализует собственно перевод с одного языка на другой; формирование базы данных, содержащей справочную информацию по каждому из языков; генерацию словаря перекрестных ссылок с уровнями вложенности 3; взаимодействие с операционной системой. Построенная марковская модель оценки динамики знаний и проведенный эксперимент показал высокую адекватность полученных результатов прогнозируемому уровню. Для проведения анализа поведения

модели в среде ТУРБО-ЕЕЯСИКа разработана программа моделирующая уровень знаний обучаемых и интенсивность забывания при различных стратегиях организации учебного процесса.

Основное содержание диссертационной работы представлено в следующих публикациях:

1. А. С. 1649568 (СССР). Электронный словарь для изучения иностранного языка В. И. Корнейчук, А. Ю. Михайлюк, Я. Г. Чхор, А. В. Искандер и В. П. Сидоренко. -Опубликовано в В. И. , 1991г. , N 18.

2. Городничий А. О. , Аль-Аржа В. И. Структура и организация памяти автономных информационно-обучающих систем массового применения //Вест. Киев. полит. ин-та. Сер. Научно-метод. -Киев: "Лыбидь". -1991. - N15. С. 23-25.

3. Городничий А. О. , Аль-Аржа В. И. Структура и организация памяти автономных информационно-справочных систем массового применения //Теория и практика построения функционально ориентированных вычислительных и микропроцессорных систем обработки информации: Тез. докл. научно-технической школы семинара, Каменец-Подольский, 19-22июня, 1990г. - Каменец-Подольский: В. И. , - 1990 - с.117-118.

4. Городничий А. О. , Прусс Г. О. ; Аль-Аржа В. И. Организация памяти и методы сжатия информационно-справочных систем // Устройство и системы хранения информации: Тез. докл. Всесоюз. Конф. , Алушта, 1-8 окт. 1991. - Симферополь, В. И. , -1991. -с. 46-47.

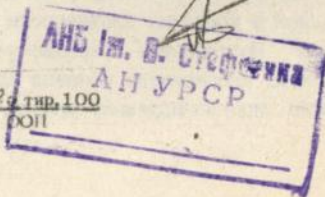
5. Городничий А. О. , Прусс Г. О. , Аль-Аржа В. И. Уплотнение информации в информационно-обучающих системах //Вестник КПИ: Сер. научно-метод. - 1992. - вып. 16 (в печати).

6. Корнейчук В. И. , Городничий А. О. , Аль-Аржа В. И. Структура и организация памяти автономных информационно-справочных систем массового применения //Проблемы памяти ЭВМ: Тез. докл. Всесоюз. ИТ школа, Уфа, 3-8 сент. , -1990. - Уфа: В. И. , -1990. - 107с.

7. Корнейчук В. И. , Михайлюк А. Ю. , Аль-Аржа В. И. Марковская модель процесса изучения иностранного языка // Вестник КПИ: Научно-метод. сер. - 1991. - Вып. 15. - с.11-14.

Подписано к печати 16/1/92 Зак. 308 тир. 100

Размножено ГВИ Минстата Украины



001

196282

AB 26.029