

УДК 595 : 681

© 1994 г.

А. Л. Лобанов и М. Б. Дианов

## ДИАЛоговая КОМПЬЮТЕРная ДИАГНОСТИЧЕСКАЯ СИСТЕМА ВІКЕУ И ВОЗМОЖНОСТИ ЕЕ ИСПОЛЬЗОВАНИЯ В ЭНТОМОЛОГИИ

[A. L. LOBANOv a. M. B. DIANOv. DIALOGUE COMPUTER IDENTIFICATION  
SYSTEM VIKEY AND ITS USE IN ENTOMOLOGY]

Биологическая диагностика, рассматриваемая теперь как прикладная ветвь систематики, занимающаяся теорией и практикой построения определителей, оформилась в самостоятельную научную дисциплину в 60—70-х годах (Свиридов, 1973; Pankhurst, 1975, 1978). Именно в тот период работы отечественные исследователи внесли значительный вклад в разработку теоретических основ биодиагностики (Балковский, 1962; Кискин, 1966; Лобанов, 1972, 1975б; Свиридов, 1978, 1981, и др.) и методов использования компьютеров для определения таксономической принадлежности биологических объектов (Кискин и др., 1965; Лобанов, 1974, 1975а, 1983; Lobanov et al., 1981; Уткина и др., 1980, и др.). К сожалению, в следующем десятилетии число отечественных работ по биологической диагностике резко сократилось, хотя за рубежом интерес к этой теме совсем не уменьшился. Среди зарубежных публикаций можно особо выделить описания конкретных действующих компьютерных систем (Boswell, Gibbs, 1986; Dedet et al., 1990; Miller, Day, 1990; Pankhurst, 1986; Payne, 1978; Rubio, 1986; Schalk, 1993, и др.). Отечественных публикаций такого рода в последние годы почти не было, что может создать неверное впечатление о нашем отставании в этой актуальной области. Исследования в области компьютерной биологической диагностики ведутся во многих научных центрах, но, к сожалению, почти не публикуются. Как исключение можно назвать обзор Пименова (1987) и краткие сообщения о работах тартуских и севастопольских биологов и программистов (Райтвийр, Кулль, 1988; Бутаков, Лелеков, 1991).

Данной статьей мы хотим отчасти восполнить этот пробел и рассказать о современном состоянии и примерах использования оригинальной компьютерной системы «ДИАГНОСТИКА» (новое название — ВІКЕУ), являющейся по свидетельству Райтвийра и Кулля (1988) первой отечественной диалоговой компьютерной диагностической системой. Первые версии этой системы «ДИАГНОСТИКА-1» и «ДИАГНОСТИКА-2» были разработаны А. Л. Лобановым в 1974 г. в Институте биологии Коми филиала АН СССР для ЭВМ «Наири-С» (Лобанов, 1975а). В дальнейшем система переносилась на более совершенные типы ЭВМ: «Наири-2» — «ДИАГНОСТИКА-3», БЭСМ-6 и СМ-1420 — «ДИАГНОСТИКА-4» (Лобанов, 1983). Начиная с 5-й версии, созданной в Зоологическом институте РАН для персональных компьютеров IBM PC (языки программирования — FoxPro и Фортран),

система получила новое название BIKEY (от слов — Biological Identification KEYS). Последняя версия системы — BIKEY6, ориентированная на работу как с текстовыми описаниями, так и с изображениями таксонов, была разработана совместно с М. Б. Диановым.

Система BIKEY6 предназначена для автоматизации процессов работы с диагностической информацией о биологических таксонах и осуществляет следующие функции.

1. Накопление и хранение на магнитных носителях в компактной форме диагностической информации о таксонах (в том числе и изображений организмов и их фрагментов).

2. Редактирование и дополнение исходной диагностической информации в удобном для пользователя режиме.

3. Всесторонний анализ этой информации, ее логический контроль и оценка диагностической полноты.

4. Проведение оптимизированного определения объектов по созданным системой машинным ключам в режиме диалога «человек—ЭВМ» (путь определения формируется динамически в ходе диалога, заключающегося в выборе человеком наиболее удобных для него признаков с учетом рекомендаций программы, которая стремится сократить длину пути определения).

5. Автоматическое генерирование на ЭВМ оптимизированных одноходовых текстовых определителей (из одного машинного ключа можно получить неограниченное число вариантов текстовых определителей путем задания разных весов разным признакам).

6. Печатаение в виде стандартного диагноза на естественном языке данных об одном таксоне или группе таксонов (таким образом специалисту облегчается содержательный контроль накопленной диагностической информации).

7. Анализ параметров традиционных одноходовых текстовых ключей, созданных вне системы BIKEY.

В ближайшее время предполагается включить в пакет программы для анализа накопленной диагностической информации методами нумерической таксономии (оценка сходства, автоматическая классификация, построение дендрограмм и т. п.) и программы для обмена данными с системой DELTA (Dallwitz, Paine, 1986), ставшей уже фактически международным стандартом в области хранения и анализа таксономической информации.

Функционально пакет BIKEY6 состоит из трех частей, автономных, но связанных друг с другом общими файлами данных:

1) набора программ на языке СУБД FoxPro для проверки и анализа готовых текстовых одноходовых ключей;

2) единого комплекса программ на языке FoxPro для создания, анализа и последующего совершенствования машинных многоходовых ключей, исходная информация для которых хранится в виде баз данных самого распространенного и доступного формата DBF для систем dBASE, FoxBase, FoxPro, Clipper;

3) ряда программ на алгоритмическом языке Фортран-88, работающих с уже готовыми машинными ключами и выполняющих самые сложные функции — предварительной обработки изображений (программы PICFILL и др.), управляемого компьютером оптимизированного диалогового определения (программы DIAKEY и PICKEY), генерации оптимизированных текстовых одноходовых ключей для публикации (GENKEY) и печати стандартизованных описаний таксонов (DESKEY).

Основные биологические понятия, используемые при описании программ: «таксон», «ряд признаков» и «признак». Таксон — название конкретной систематической группы организмов. Ряд признаков — элемент распознавания (черта, свойство). Примеры: «цвет тела», «число окологлазных щети-

нок», «наличие пунктировки» и т. п. Признак — конкретное состояние элемента распознавания. Примеры: «черное тело», «3 окологлазные щетинки», «пунктировки нет» и т. п. Один ряд может включать 2 или более признаков. Одно определение называется диагнозом. Шагом называется цикл, состоящий из выбора ряда признаков, затем выбора признака в этом ряду и отбора из имеющих таксонов тех, которые могут иметь этот признак. Длина пути определения — число шагов, сделанных пользователем до получения однозначного ответа.

Мы считаем уместным пояснить некоторые другие, употребляемые далее термины. Диагностические ключи классифицируются по разным аспектам — основаниям классификации (Лобанов, 1972). Наиболее важные из них три: число входов, число признаков в одном ряду и форма представления ключа. Особое значение мы придаем числу входов — возможности на каждом шаге определения выбирать наиболее удобный или надежный для определяющего элемент распознавания из нескольких. Если такой возможности нет, то определитель считается одновходовым, а если она есть — многовходовым. Часто этот аспект классификации ключей смешивают со следующим — числом признаков в ряду (числом альтернатив на каждом шаге определения). Если имеются только 2 альтернативы — теза и антитеза, ключ дихотомический. Если хотя бы в одном случае таких альтернатив 3 или более, ключ считается политомическим. Последний термин часто ошибочно используется для обозначения многовходовых ключей. Это совершенно неверно и связано только с тем, что многвходовые ключи обычно бывают политомическими, ибо в отличие от традиционных текстовых одновходовых ключей форма здесь не диктует составителю строгую дихотомичность. В качестве одного из последних примеров такого смешения разных аспектов классификации ключей можно привести работу киевских исследователей (Михалевич, Некрутенко, 1992), которая, несмотря на эту ошибку, очень ценна очередным доводом в пользу перехода на компьютерную технологию при подготовке определителей.

Доводы биологов в пользу отказа от строгой дихотомии неоднократно приводились в литературе (Кискин, 1966; Свиридов, 1973), поэтому мы остановимся только на математическом и психологическом аспектах выбора числа признаков в одном ряду. Если минимизировать общее число признаков в ключе для определения заданной группы таксонов, то оптимальный средний объем ряда признаков 2.718 (Лобанов, 1983). Если измерять длину пути определения не просто числом использованных рядов признаков, а суммарным числом признаков в использованных рядах (Гамбарян, 1975), то минимум средней длины достигается при среднем объеме рядов признаков, равном 3.69. Таким образом, математический подход к оптимизации ключей рекомендует среднее число признаков в ряду, близкое к 3, — политомия, с преобладанием тритомии, но с использованием и дихотомических рядов. Это хорошо согласуется и с потребностью биологов выделять во многих случаях третий признак даже в таких казалось бы дихотомических рядах, как «наличие—отсутствие», например, для рудиментарного состояния какой-то структуры. Верхняя граница числа признаков в ряду диктуется скорее особенностями человеческого восприятия и памяти. Специальные исследования (Miller, 1956) показывают, что объем кратковременной памяти ограничивается  $7 \pm 2$  блоками сопоставляемой информации. Очевидно, что использование в одном ряду более 7—9 признаков только затруднит работу с ключом.

По форме представления можно выделить много типов ключей, но мы здесь будем говорить только о текстовых (традиционная форма публикации у отечественных зоологов) и машинных, представляющих собой специальные компьютерные формы хранения диагностической информации (Lobanov et al.,

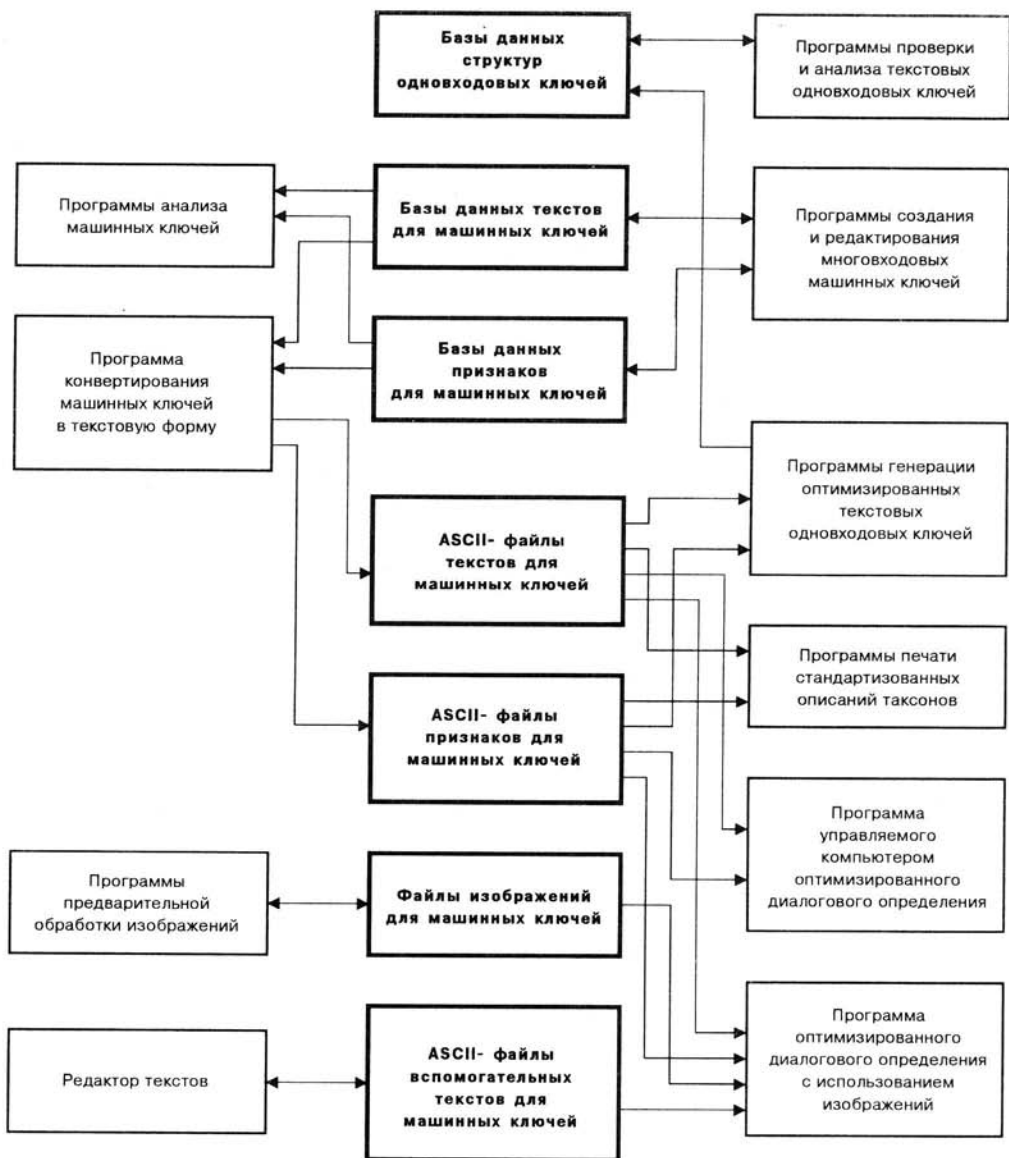


Рис. 1. Блок-схема взаимодействия основных составных частей системы VIKEY6 (в толстых рамках — файлы данных, в тонких — программы; пересечения линий связи не означают их соединения).

1981). В машинных ключах всегда будет подразумеваться самый эффективный тип диагноза — последовательный оптимальный (Лобанов, 1975а). Блок-схема связей и взаимодействия отдельных частей системы приведена на рис. 1. Ниже описываются программы каждого из трех разделов пакета VIKEY6.

## 1. ПРОГРАММЫ ДЛЯ ПРОВЕРКИ И АНАЛИЗА ГОТОВЫХ ТЕКСТОВЫХ ОДНОВХОДОВЫХ КЛЮЧЕЙ

Программы этого раздела выполняют ввод данных о ключе в базу данных, логические проверки ключа и анализ математических параметров ключа. Из-за заметных различий в форме серийных (наиболее распространенных в отечественной зоологической литературе) и скобочных (чаще используемых за рубежом) текстовых определителей, для них используются разные программы и разные структуры баз данных. Предусмотрен сокращенный вариант ввода данных о ключе, когда вводятся только номера тез и антитез (без соответствующих им текстов), а вместо названий таксонов — только пометки у соответствующих тез и антитез. Для обеих форм ключей проверяются правильность нумерации тез и антитез (отсутствие пропусков и повторов) и правильность перекрестных ссылок между тезами и антитезами. Для серийных ключей проверяются полнота ссылок и отсутствие нарушений логики дихотомического ветвления. Для скобочных, кроме этого, выполняются проверки на отсутствие повторных ссылок на один узел и на отсутствие ссылок назад. При анализе ключей вычисляются средняя длина пути определения и числовая оценка — показатель совершенства конструкции ключа (Лобанов, 1975б), изменяющийся от 10 (при полной сбалансированности ветвления дерева ключа) до 0 (при самой неоптимальной форме дерева — так называемой «гребенке»).

## 2. КОМПЛЕКС ПРОГРАММ ДЛЯ СОЗДАНИЯ, АНАЛИЗА И СОВЕРШЕНСТВОВАНИЯ МАШИННЫХ МНОГОВХОДОВЫХ КЛЮЧЕЙ

Исходные данные для системы ВІКЕУ6 готовятся в виде многовходовых политомических цифровых таблиц. Основы теории диагностических ключей и, в частности, особенности политомических ключей, были рассмотрены ранее (Лобанов, 1972). Построение такой таблицы целесообразно проводить в 3 этапа с получением одновременно трех составных частей компьютерного ключа: списка таксонов, списка рядов и признаков и, наконец, цифровой матрицы. Сначала нужно составить список таксонов (видов, родов или семейств и т. д.), с которыми будет работать машинный ключ. Затем необходимо сформировать ряды признаков, которые нужны для надежного различения всех заданных таксонов. В системе ВІКЕУ6 в одном ряду должно быть не более 9 признаков. Самый ответственный этап — составление собственно цифровой таблицы (матрицы), в которой строки соответствуют таксонам, а столбцы — рядам признаков. На пересечении строки и столбца указываются номера признаков, которые известны для данного таксона в данном ряду признаков. Отсутствие информации о признаках таксона по данному ряду признаков обозначается нулем (прочерки в таблице недопустимы). Но необходимо иметь в виду, что все программы будут трактовать эти нули как наличие у данного таксона всех признаков данного ряда.

При формировании рядов признаков и составлении цифровой таблицы необходимо придерживаться специальных рекомендаций для машинных ключей. Формулировки признаков в рядах должны быть лаконичны. Не следует усложнять признаки для устранения их перекрывания у разных таксонов. Перекрывание признаков будет учтено машинными программами при оптимизации пути определения. Наличие сложных формулировок признаков говорит о том, что ряд признаков искусственный и его нужно разделить на 2 или несколько более простых рядов. В отличие от традиционных ключей избыток рядов признаков не затрудняет, а облегчает пользование машинным ключом. Подготовленные систематиком 3 части компьютерного ключа (список таксонов, список рядов и признаков, цифровая матрица)



вводятся в две базы данных формата DBF при помощи удобной программы ввода. Для ввода больших матриц предусмотрены особые режимы, облегчающие систематику эту работу. Можно вводить набор признаков для одного таксона, или признаки одного ряда для набора таксонов, или любой прямоугольный фрагмент матрицы.

Отдельные программы предназначены для редактирования (включая и дополнения) трех составляющих ключа: названий таксонов, текстов признаков (включая изменение числа признаков в ряду) и матрицы признаков. Особое внимание уделено детальному сопоставлению пар близких таксонов. Специальная программа позволяет редактировать одновременно или все признаки такой пары, или только те ряды, где у этих двух таксонов нет отличий, или только те ряды, где они отличаются, или только те ряды, где хотя бы для одного таксона нет данных о признаках (нуль в матрице).

Две программы разработаны специально для проверок машинных ключей. Одна из них проверяет отличия каждого таксона от всех остальных, перебирая возможные пары. До тех пор, пока есть неотличимые пары, ключ, строго говоря, не может служить определителем и нуждается в редактировании. После устранения неотличимых пар эта же программа используется для создания «запаса прочности» в ключе, т. е. гарантии наличия для любой пары таксонов двух или более отличающих их признаков. Истинно многовходовым ключ может считаться только при существовании такого запаса. Вторая программа обнаруживает в ключе явные ошибки и противоречия, заключающиеся в использовании в характеристике таксона признака, отсутствующего в соответствующем ряду, и в наличии в ряду признака, который не используется ни в одной характеристике. В связи с тем что программы следующего раздела системы используют текстовые файлы специального формата, отдельная программа предназначена для конвертирования данных из баз данных в ASCII-файлы. Сервисная программа печати позволяет распечатывать в удобной для систематика форме все составные части машинного ключа, а программа помощи пользователю обеспечивает контекстно-зависимые экраны «Help», содержащие краткие указания по работе с системой на разных этапах создания и анализа ключа.

Система VIKEY6 отличается от системы VIKEY5 наличием специальной программы, которая образует названия файлов с изображениями, определяет их возможные экранные размеры и выдает этот перечень как руководство систематику для работы над рисунками. Эти файлы обязательны. Вместе с их списком программа выдает перечень файлов с дополнительными текстами (комментарии к рядам признаков и справки о таксонах), которые факультативны. Второе отличие 6 версии заключается в более заметных ограничениях на длину и число строк в формулировках рядов признаков и самих признаков (не более 5 строк длиной до 36 символов). Это ограничение связано с необходимостью одновременного показа на экране большого числа текстов и рисунков.

### 3. ПРОГРАММЫ НА АЛГОРИТМИЧЕСКОМ ЯЗЫКЕ ФОРТРАН-88 ДЛЯ РАБОТЫ С ГОТОВЫМИ МАШИННЫМИ КЛЮЧАМИ

Фортран — старейший из алгоритмических языков, активно используемых при разработке прикладного программного обеспечения для научных исследований. Существование огромного массива написанных на Фортране программ для больших ЭВМ и библиотек тщательно отработанных математических алгоритмов, а также безупречная реализация в компиляторах этого языка числовых операций, делающая этот язык вне конкуренции в области вычислений (вопреки бытующим неверным представлениям здесь ему значительно уступают новейшие версии Паскаля и даже Си) оставляют

до сих пор Фортран в ряду реально используемых на персональных компьютерах языков. Для компенсации недостатков Фортрана в области обработки изображений и экранной графики М. Б. Диановым были созданы на языке ассемблера библиотеки процедур быстрого вывода на экран как для текстового, так и графического режима, а также процедур поддержки манипулятора «мышь». Наличие этих библиотек позволило авторам сохранить использование Фортрана для разработки всех программ пакета VIKEY6.

### 3.1. Предварительная обработка изображений (программы PICFILL и другие)

Полученные сканированием изображения сохраняются в одном из стандартных форматов для дальнейшей обработки. Изображения, выводимые на экран, должны быть подобны и иметь строго определенные размеры. Поэтому оригинальные изображения делаются на специально подготовленных типовых бланках. При использовании ранее сделанных изображений в ходе сканирования проводится масштабирование фрагментов изображений и приведение рисунков к заданным размерам. Когда качество полученных экранных изображений не удовлетворяет поставленным требованиям, выполняется их ретуширование средствами графического редактора. Важным требованием к цветным изображениям, иллюстрирующим признаки, является правильная цветопередача. Для ее достижения черно-белое контурное изображение, полученное на сканере, специалист-систематик может раскрасить прямо на экране при помощи специально разработанной для этого М. Б. Диановым программы PICFILL, в которой точная цветопередача достигается подбором 16 цветов из 262144 возможных оттенков (для видеоадаптера VGA). Тотальные изображения обычно получают сканированием на цветном сканере с последующей обработкой мощными графическими редакторами. Для нестандартных операций был написан ряд специализированных программ и утилит. На последнем этапе экранные изображения конвертируются в разработанный М. Б. Диановым формат для использования непосредственно в программе диалоговой диагностики. При его разработке основным требованием был быстрее вывод на экран, иногда в ущерб размеру файла. Однако используемый метод компрессии не уступает использованному в стандарте РСХ.

### 3.2. Управляемое компьютером оптимизированное диалоговое определение (программы DIAKEY и PICKEY)

Самые привлекательные для пользователя в пакете VIKEY6 программы — диалоговые политомические многоходовые определители: DIAKEY (работающая в текстовом режиме) и PICKEY (работающая в графическом режиме и активно использующая иллюстрации). Программа PICKEY представляет собой дальнейшее развитие программы DIAKEY и полностью совместима с ней по используемым файлам текстовых данных. В обеих программах использован один и тот же алгоритм определения. Отличия заключаются только в том, что программа DIAKEY совсем не использует изображения (но зато она может работать на всех типах видеоадаптеров, начиная с CGA) и позволяет работать с более длинными строками описания рядов признаков и признаков (72 символа в строке). При этом и число строк в одном описании может достигать 9. Для подготовки таких более обширных текстов следует использовать соответствующие программы пакета VIKEY5. Но так как в программе PICKEY есть возможность вывода на экран дополнительных текстов описаний еще большего объема, то при наличии файлов с изображениями и подходящего компьютера эту программу, конечно, стоит предпочитать предыдущей, и поэтому далее описывается только она. Исполь-

зование иллюстраций и графического интерфейса предполагает легкое освоение и использование программы широким кругом пользователей-биологов, имеющих лишь начальные навыки работы на персональном компьютере. Все средства работы с графикой и организации удобного пользовательского интерфейса с использованием «мыши» разработаны для программы **PCKEY** М. Б. Диановым.

Программа **PCKEY** реализует активный графический интерфейс и рассчитана на обязательное использование манипулятора «мышь». Основные элементы интерфейса — фрейм (прямоугольный фрагмент экрана) с текстом, фрейм с изображением и кнопка. Каждый элемент может быть активным: активный фрейм окружен красной рамкой, кнопки активизируются при нажатии. При нажатии исполнительной клавиши мыши (левой) в пределах активного фрейма реализуется действие. При нажатии отменяющей клавиши мыши (правой) производится откат к предыдущему состоянию. В любом месте программы есть реакция на клавишу  $\langle \text{Escape} \rangle$  — выход и  $\langle \text{F1} \rangle$  — «Help» (выдача контекстнозависимой помощи).

Процесс диалогового определения экземпляра на компьютере при помощи этой программы состоит из шагов. На каждом шаге определяющий должен выбрать наиболее понятный ему и доступный для наблюдения на данном экземпляре ряд признаков, а затем из признаков этого ряда выбрать тот, который лучше всего соответствует свойствам этого экземпляра. Шаг заканчивается отбором из списка тех таксонов, которые имеют выбранный признак (на первом шаге отбор идет из полного списка таксонов, включенных в определитель). Если таких таксонов 2 или более, программа переходит к следующему шагу. На каждом шаге ряды предлагаются программой в порядке убывания их диагностической ценности, вычисляемой для данного шага по оригинальному алгоритму (Лобанов, 1974). Желательно использовать ряды, выданные машиной первыми, но окончательное решение принимает определяющий, находя компромисс между надежностью, доступностью и ценностью ряда.

Запустив программу, определяющий видит первый рабочий экран (рис. 2). Он заполнен текстовыми фреймами с названиями рядов признаков, фреймом с иллюстрацией к текущему ряду признаков, фреймом с названием ключа и фамилией автора и фреймом с количеством таксонов в списке. Иллюстрация к текущему ряду признаков предназначена только для лучшего понимания пользователем признака. Среди фреймов с названиями рядов признаков необходимо выбрать желательный путем перемещения курсора по текстовым фреймам. При этом происходит смена поясняющих иллюстраций в графическом фрейме. Когда число доступных рядов признаков больше 7 и фреймы с их названиями не помещаются на экране, программа показывает края фреймов, размещенных за пределами экрана. Поместив курсор на край невидимого фрейма, можно произвести скроллинг фреймов (он возможен как вверх, так и вниз). Поместив курсор на фрейм с наименованием ряда признаков и нажав левую клавишу мыши, пользователь может вызвать на экран справочный фрейм с дополнительными обширными текстовыми пояснениями к этому ряду. Указав курсором на фрейм с названием ключа и фамилией автора и нажав левую клавишу мыши, пользователь вызывает текстовый фрейм с пояснениями ко всему ключу. Возврат к определению одинаков для работы с любым фреймом программы: следует нажать правую кнопку мыши.

Сделав выбор ряда признаков, определяющий переводит курсор на графический фрейм с иллюстрацией к выбранному ряду и нажимает левую клавишу мыши. Программа переходит ко второму рабочему экрану (рис. 3). Этот экран разделен на несколько одинаковых областей (от двух до девяти), которые заполнены иллюстрациями к признакам выбранного ряда. пользо-



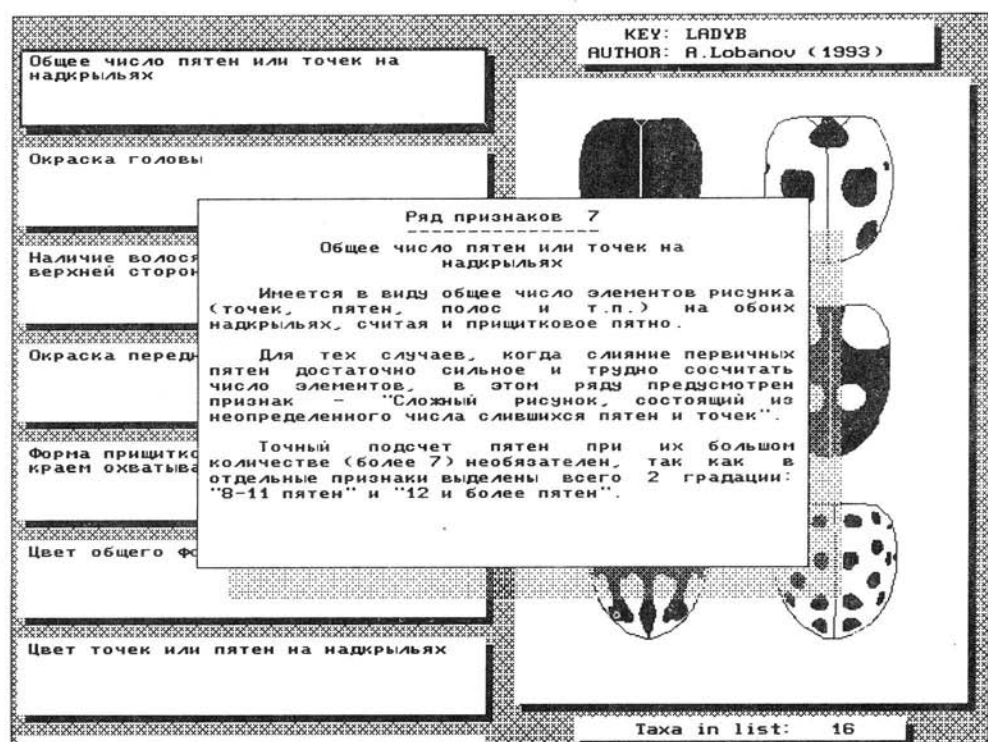


Рис. 2. Копия первого рабочего экрана программы PICKEY — момент выбора ряда признаков.

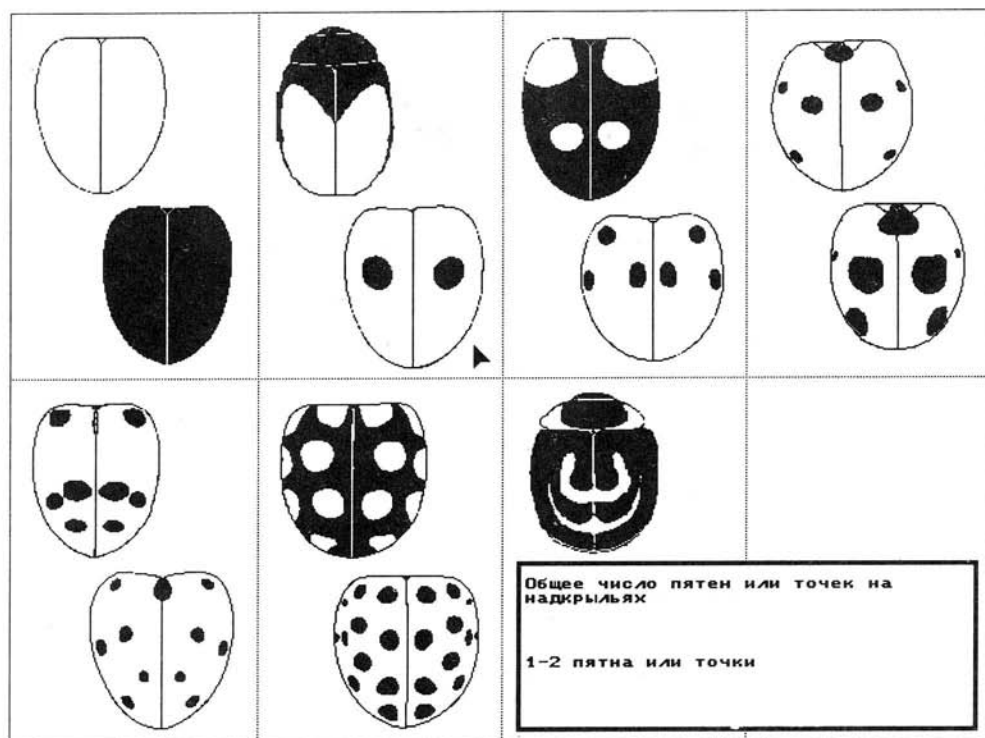


Рис. 3. Копия второго рабочего экрана программы PICKEY — момент выбора признака.

ватель внимательно изучает нужный элемент распознавания у определяемого экземпляра и выбирает то изображение, которое наиболее адекватно определяемому экземпляру. Программа предоставляет возможность контролировать правильное понимание иллюстрации по текстовому описанию выбранного признака, помещенному на перемещающемся текстовом фрейме. Для отказа от использования непонятного ряда признаков достаточно нажать правую кнопку мыши. Произведя выбор, определяющий переводит курсор на соответствующее изображение и нажимает левую кнопку мыши. Программа производит удаление из матрицы определения строк тех таксонов, у которых не может быть введенного признака. Если оставшихся таксонов 2 или более, программа делает переоценку доступных рядов признаков и снова переходит к первому рабочему экрану.

Пользователь повторяет перечисленные операции до окончания определения. Находясь на первом рабочем экране, определяющий может узнать названия таксонов, оставшихся в списке возможных. Для этого необходимо переместить курсор на фрейм «Список таксонов», сделав его активным, и нажать левую клавишу мыши. Эта возможность позволяет использовать программу в качестве справочника (списки таксонов с заданным сочетанием признаков).

При верном выборе признаков определение завершается показом тотального черно-белого и, если есть возможность, цветного изображения организма, а также текстового описания таксона (третий рабочий экран — рис. 4). Сначала демонстрируется тотальное черно-белое изображение и полное название таксона. При нажатии любой кнопки поверх этого изображения появляется фрейм с текстом краткого описания таксона и тремя кнопками. Нажатие левой клавиши мыши при указании на кнопки вызывает

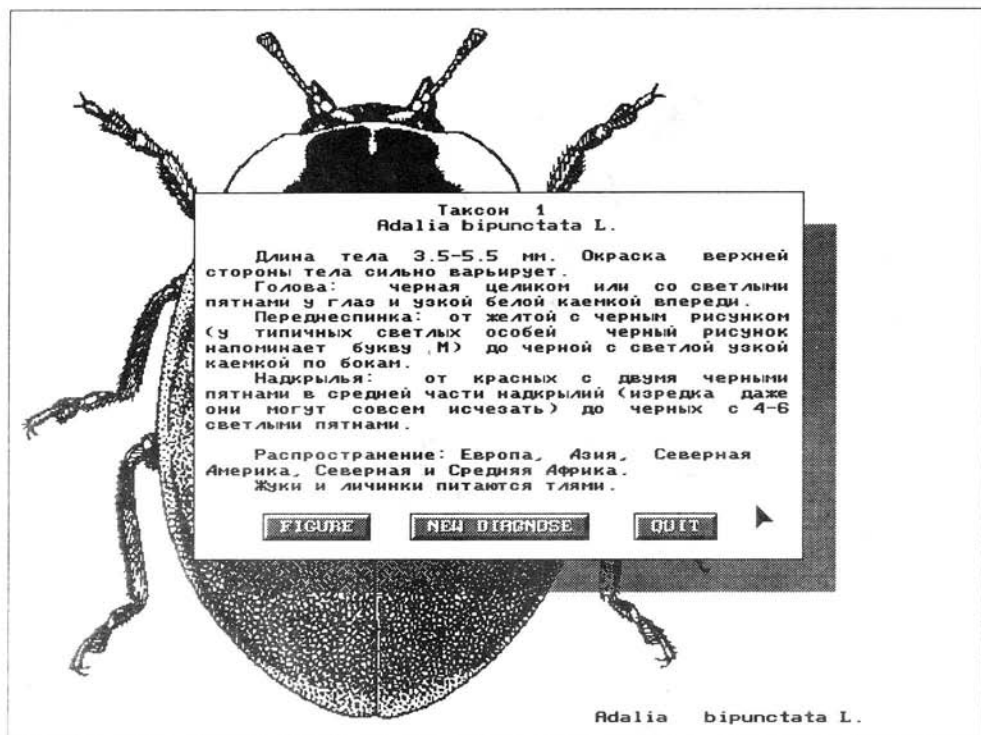


Рис. 4. Копия третьего рабочего экрана программы PICKEY — конец диагноза.

соответствующие действия. Кнопка «Рисунок» вызывает показ по очереди цветного и черно-белого тотального изображения. Кнопка «Новый диагноз» вызывает переход к новому определению. Кнопка «Конец» — выход из программы.

В процессе определения могут возникнуть две особые ситуации, обрабатываемые программой PICKEY. При выборе непредусмотренного автором ключа набора признаков (когда указанный набор не соответствует ни одному из таксонов, включенных в определитель), программа выдает соответствующее сообщение и дает возможность вернуться на предыдущий шаг определения или начать определение с самого начала. Если ключ незавершен автором и возникает ситуация, когда для дифференциации близких таксонов в ключе нет больше признаков — программа сигнализирует об этом и показывает список неразличимых таксонов.

Для полноценной работы программы PICKEY необходимы следующие файлы данных: 1) файл с текстовыми данными ключа в формате ASCII; 2) файл с цифровыми данными ключа в формате ASCII; 3) сборные рисунки — иллюстрации к рядам признаков; 4) иллюстрации к каждому признаку ряда; 5) черно-белые тотальные рисунки таксонов; 6) по возможности — цветные тотальные рисунки таксонов; 7) файлы с информацией о таксонах в формате ASCII; 8) файл с информацией о ключе в формате ASCII; 9) файлы с дополнительной информацией о рядах признаков в формате ASCII. Общий объем этих файлов для демонстрационного ключа 16 видов божьих коровок (16 рядов признаков) — 2030 килобайт (после упаковки архиватором PKZIP v2.04 — 1030 килобайт). Для работы с программой PICKEY и этим ключом необходимы все 143 файла, но при использовании только программы DIAKEY число нужных файлов снижается до 2, а их суммарный объем — до 7 килобайт (2 килобайта после архивации). Исполняемые модули самих программ занимают всего 43 (DIAKEY) и 55 (PICKEY) килобайт.

Обе программы (DIAKEY и PICKEY) не являются в строгом смысле экспертными системами, но имеют многие черты, характерные для последних, и успешно конкурируют с ними в удобстве, скорости и надежности процесса определения. Важное преимущество этих программ перед настоящими экспертными системами — простота передачи экспертом (специалистом-систематиком) своих знаний системе. В системах VIKEY5 и VIKEY6 для этого не нужны долгие диалоги с компьютером и поиски экспертом неочевидных для него ответов на неожиданные вопросы системы. Систематик сам формирует удобную для него схему рядов признаков и передает системе свои знания о таксонах в рамках этой своей схемы в виде матрицы, которую он может заполнять без компьютера, в удобной для него обстановке и в произвольной последовательности. Пригодность этой матрицы для диагностики контролируется системой, но опять-таки без навязывания эксперту конкретных вопросов в тех случаях, когда информации для построения ключа еще мало. Такой «мягкий» способ наполнения машинного определителя необходимой информацией немаловажен для успешной работы с системой тех зоологов, которые критически относятся к некоторым аспектам компьютеризации своей творческой деятельности.

### 3.3. Генерация оптимизированных текстовых одноходовых ключей для публикации (программа GENKEY)

В основе алгоритма программы лежит предложенный ранее способ последовательного построения дерева оптимальных диагнозов (Лобанов, 1983). Для каждого таксона в ключе моделируется пошаговый оптимальный диагноз. Если характеристики таксона по части рядов признаков поли-

модальны (содержат несколько возможных признаков в одном ряду), то такие диагнозы проводятся для каждого возможного сочетания признаков. Пути всех смоделированных диагнозов постепенно образуют в памяти ЭВМ дерево (для его компактной записи разработано специальное матричное представление). После окончания процесса формирования этого дерева подсчитываются и печатаются математические характеристики полученного одноходового ключа, а затем выводится в текстовой файл сам ключ в форме, пригодной для публикации. В программе предусмотрен механизм быстрого изменения пользователем оценок весов отдельных рядов признаков. На эти оценки (перед окончательным выбором одного ряда признаков из нескольких возможных) умножаются вычисляемые программой на каждом шаге диагностические ценности рядов. Изменяя значения весов, исследователь получает возможность влиять на ход построения ключа и генерировать из одного набора данных произвольное число вариантов одноходовых ключей, зачастую существенно отличающихся друг от друга.

### 3.4. Печать стандартизованных описаний таксонов (программа DESKEY)

Для окончательной проверки накопленной в машинном ключе информации разработана специальная программа, которая выводит в текстовой файл и на экран сведения о каждом таксоне в виде стандартизованных текстовых описаний. Опыт показывает, что ошибки в характеристиках таксонов, оставшиеся незамеченными при интерактивной проверке на экране, легче всего обнаруживаются именно при неторопливом чтении систематиком таких описаний, где нет никаких кодов и сокращений.

## 4. ЗАКЛЮЧЕНИЕ

Пакеты программ ВІКЕУ5 и ВІКЕУ6 могут использоваться для построения ключей любых групп животных и растений, но особую ценность представляют для энтомологов-систематиков, которые значительную часть своих усилий затрачивают на создание новых определителей и корректировку уже имеющихся для включения в них новой информации. Компьютерная технология может дать таким ученым большую экономию времени. В то же время программы DIAKEY и PISKEY могут быть очень полезны прикладным биологам, так как позволяют максимально упростить процедуру определения таксономической принадлежности организмов, что особенно важно для экспресс-диагностики вредных насекомых в сети станций защиты растений, в службах карантина растений и защиты леса, а также для определения насекомых — индикаторов загрязнения окружающей среды. В Зоологическом институте РАН система ВІКЕУ успешно использована для создания определителей отрядов насекомых Палеарктики, видов наиболее обычных жуков кокциnellид, видов отдельных родов жуков усачей и жужелиц, а также семейств и видов разных групп кишечнорастных.

Авторы считают своим долгом поблагодарить специалистов, способствовавших своими советами и практической помощью созданию первых версий системы «ДИАГНОСТИКА», В. С. Никифорова и Б. Г. Новаковского, а также руководителей, при неизменной поддержке которых эта система развивалась и совершенствовалась, О. А. Скарлато, А. Ф. Алимова, Г. С. Медведева.

- Балковский Б. Е. О повышении диагностической значимости признаков, используемых для определения растений // Бот. журн. 1962. Т. 47. № 9. С. 1309—1314.
- Бутиков Е. А., Лелеков С. Г. Диалоговая система определения объектов на основе графического интерфейса. М.: ICSTI, 1991. С. 353—357.
- Гамбарян П. П. Взвешивание признаков и их корреляция // Биол. журн. Армении. 1974. Т. 27. № 4. С. 97—100.
- Кискин П. Х. Методы диагностики животных и растений на основе политомического принципа // Политомический принцип определения животных и растений. Кишинев, 1966. С. 19—40.
- Кискин П. Х., Печерская И. Н., Печерский Ю. Н. Автоматизация диагностического поиска сортов винограда на ЭВМ «Минск-1» // Виноделие и виноградарство СССР. 1965. № 1. С. 21—22.
- Лобанов А. Л. Логический анализ и классификация существующих форм диагностических ключей // Энтотом. обзор. 1972. Т. 51, вып. 3. С. 668—681.
- Лобанов А. Л. Оценка диагностической ценности рядов признаков в многоходовых определителях, рассчитанных на использование ЭВМ // Тезисы докладов VI Коми республиканской молодежной научной конф. Сыктывкар, 1974. С. 125—126.
- Лобанов А. Л. Результаты экспериментов с биологическими диагностическими системами на базе ЭВМ «Наири-С» // Биологические исследования на Северо-Востоке европейской части СССР. (Ежегодник Института биологии Коми филиала АН СССР). Сыктывкар, 1975а. С. 162—168.
- Лобанов А. Л. Математический аппарат для расчета, оценки и сравнения конструктивных // Зоол. журн. 1975б. Т. 54, вып. 4. С. 485—497.
- Лобанов А. Л. Принципы построения определителей насекомых с использованием электронных вычислительных машин // Автореф. дис. ... канд. биол. наук, Л.: ЗИН АН СССР, 1983. 19 с.
- Михалевич О. А., Некрутенко Ю. П. Несколько замечаний об использовании ЭВМ для классификации биологических объектов // Вестник зоол. 1992. № 3. С. 79—83.
- Пименов М. Г. Математические методы и вычислительная техника в систематике высших растений // Итоги науки и техники. Ботаника. Т. 8, вып. 2. М.: ВИНТИ, 1987. 96 с.
- Райтвийр А., Куль К. Диалоговый ключ для определения грибов, реализованный на персональном компьютере // Микология и фитопатология. 1988. № 1. С. 111—117.
- Свиридов А. В. Проблема соотношения биологической диагностики и систематики // Журн. общ. биол. 1973. Т. 34. № 6. С. 900—906.
- Свиридов А. В. О некоторых актуальных вопросах теории идентификации биологических объектов с помощью ключей // Науч. докл. высш. школы. Биол. науки. 1978. № 10. С. 15—28.
- Свиридов А. В. Некоторые проблемы и перспективы теории идентификации биологических объектов с помощью ключей // Сб. трудов Зоол. музея МГУ. 1981. Т. 19. С. 18—37.
- Уткина Л. И., Уткин И. С., Благодатская В. М. О построении с помощью ЭВМ одноходовых ключей, предназначенных для видовой идентификации биологических объектов // Журн. общ. биол. 1980. Т. 41, № 2. С. 303—313.
- Boswell K., Gibbs A. J. The VIDE data bank for plant viruses // Development and applications in virus testing, 1986. Association of Applied Biologists. P. 283—287.
- Dallwitz M. J., Paine T. A. User's guide to DELTA system. A general system for coding taxonomic description // Division of Entomology. Report N 13 (Third Edition). 1986. Canberra: CSIRO. 93 p.
- Dedet J. P., Lebbe J., Vignes R. C. I. P. A.: A multinational programme for computer identification of phlebotomine sandflies of America // Bull. Soc. Fr. Parasitol. 1990. Vol. 8, suppl. N 2. P. 1131.
- Lobanov A. L., Schilow W. F., Nikritin L. M. Zur Anwendung von Computern für die Determination in der Entomologie // Deutsch. Entomol. Z. 1981. N. F. 28, H. 1—3. S. 29—43.
- Miller G. A. The magical number seven, plus or minus two // Psychol. Rev. 1956. Vol. 63. P. 81—92.
- Miller M. G., Day E. R. Interactive taxonomy: name that bug in three touches or less // Amer. Ent. 1990. Vol. 36, N 3. P. 219—224.
- Pankhurst R. J. (ed.). Biological identification with computers // The systematic association special volume, N 7. London, 1975. 333 p.
- Pankhurst R. J. Biological identification. The principles and practice of identification methods in biology, London: Edward Arnold, 1978. 104 p.
- Pankhurst R. J. A package of computer programs for handling taxonomic databases // CABIOS. 1986. N 2. P. 33—39.
- Payne R. W. GENKEY. A program for constructing and printing identification keys and diagnostic tables. Harpenden: Rothamsted Experimental Station, 1978. 54 p.



Rubio F. F. Computerized key for the determination of Spanish Rhopalocera. Madrid: Servicio de Publicaciones, 1986. 237 p.

Schalk P. H. Systematic goes multimedia: computerizing biological knowledge // ASC Newsletter. 1992. Vol. 20, N 3. P. 117—119.

Зоологический институт РАН,  
Санкт-Петербург.

Поступила 15 VI 1993.

#### SUMMARY

The last versions of a dialogue computer system BIKEY, which was constantly being improved for 20 years are described in detail. The system permits to make quickly multientry polychotomous keys to any groups of animals and plants, to use them as information systems on characters of the taxa, to build on their basis numerous variants of traditional optimized monoentry keys, to make formalized descriptions of taxa. Specific attention is paid to the program PICKEY — dialogue computer identification key based on illustrations. The system BIKEY can be useful both for entomologists-systematics and applied experts, working in networks of stations of plant protection, in services of plant quarantine, forest protection and protection of a environment.