

УДК 519 : 57

О. А. Михалевич, Ю. П. Некрутенко

НЕСКОЛЬКО ЗАМЕЧАНИЙ ОБ ИСПОЛЬЗОВАНИИ ЭВМ ДЛЯ КЛАССИФИКАЦИИ БИОЛОГИЧЕСКИХ ОБЪЕКТОВ *

Области приложения энтомологических знаний непрерывно расширяются. Они требуются специалистам по защите растений в сельском и лесном хозяйстве, при обосновании и проведении мероприятий по охране природы, в медицине, ветеринарии, пчеловодстве и шелководстве. Насекомые широко используются в физиологических, биохимических и биофизических исследованиях. Бессмысленно было бы требовать от всех специалистов, работающих с насекомыми, досконального знания многообразного мира *Insecta*. Однако им каждый раз непременно надо знать, с каким объектом они работают. Помощь в этом могли бы оказать определительные ключи, аналогичные тем, которые приведены в рассматриваемой книге. Другой важной областью применения ключей, реализованных на ЭВМ, может быть учебный процесс, когда начинающий энтомолог или любитель знакомится с определенной группой насекомых. ЭВМ поможет ему быстро находить наиболее важные для классификации данной группы признаки. Кроме того, отпадает необходимость в бесконечном перелистывании толстых определителей.

Прежде всего необходимо отметить, что для практического использования сама по себе книга Ф. Ф. Рубио не нужна. Нет смысла набирать программы вручную в том случае, если имеется ЭВМ, аналогичная использованной автором (IBM-совместимая). Он согласен любезно предоставить дискеты с записями всех программ. Если же в вашем распоряжении находится персональный компьютер (ПЭВМ) другого типа, можно также воспользоваться дисками, полученными от автора, так как язык «Бейсик» позволяет легко внести любые изменения в программу. Сложность может возникнуть в том случае, когда имеются существенные различия в составе энтомофауны района ваших исследований и Пиренейского полуострова — места исследований Ф. Ф. Рубио. Здесь потребуется существенная переработка программ как по части видового состава булавоусых чешуекрылых, так и по части определительного ключа. К этому вопросу мы еще вернемся ниже.

В настоящее время владение методами решения разнообразных задач на ПЭВМ, умение эффективно использовать их в конкретных исследованиях становится непременным компонентом образования каждого научного сотрудника. Любое биологическое исследование, если оно осуществляется на современном уровне, требует привлечения сложных математических методов, которые невозможно реализовать без использования ЭВМ. Это касается всех направлений, где ЭВМ используются в биологии: обработка данных экспериментов; планирование, организация и проведение их в управляемом режиме; моделирование биологических процессов; классификация объектов и явлений (частный случай распознавания образов).

В связи с развитием средств диалогового доступа к ЭВМ через системные или удаленные терминалы в последние 10-летия произошло коренное изменение принципов общения пользователя с ЭВМ. Получили распространение простые в работе интерактивные системы программирования, которые предусматривают, что в процессе выполнения прог-

* О книге Ф. Фернандеса-Рубио «Компьютеризованный ключ для определения испанских булавоусых чешуекрылых» (Fidel Fernandes-Rubio. Computerized key for the determination of Spanish Rhopalocera.—Eusco jaurlaritza, Gobierno; Servicio de Publicaciones, 1986.—237 р.)

раммы активность проявляется попеременно то со стороны пользователя, то со стороны ЭВМ.

Все перечисленное выше позволило Ф. Ф. Рубио разработать относительно простую систему ключей для определения булавоусых чешуекрылых (*Rhopalocera*) Пиренейского полуострова. Книга состоит из пяти семантически идентичных частей, где на испанском, баскском, французском, английском и немецком языках изложены основные принципы определения этих насекомых. Соответствующие программы различаются только языком комментариев и вопросов, но не своим программным оформлением.

Пакет состоит из 9 программ. Одна из них «CLAVE» — основная, с нее каждый раз начинается работа с определителем. Здесь находятся ключи для определения видов, принадлежащих к семействам *Papilionidae* (5 видов), *Danaidae* (2), *Libytheidae* (1) и *Riodinidae* (1). Кроме того, эта программа обеспечивает автоматический переход к программам «CLP», «CLN», «CLS», «CLL-1» и «CLH», по которым производится определение бабочек других семейств. Программа «CLP» предназначена для определения бабочек семейства *Pieridae*. Их в фауне Пиренейского полуострова насчитывается 22 вида. Программы «CLN», «CLS» и «CLN» обеспечивают соответственно определение видов семейств *Nymphalidae* (41 вид), *Satyridae* (36, не учитывая видов рода *Erebia*) и *Hesperiidae* (29). Семейство *Lycanidae* представлено 70 видами. Поэтому для удобства работы пришлось создать две отдельные программы («CLL-1» — 39 видов и «CLL-2» — 31 вид).

В связи с тем, что существуют значительные трудности с определением видов рода *Erebia*, потребовалось две отдельные программы для классификации его представителей. Программа «CLE» предназначена для определения видов рода *Erebia* по характеристикам гениталий. Она связана автоматическими переходами как с головной программой «CLAVE», так и с программами «CLE-2» (определение по внешним морфологическим признакам) и «CLS» (семейство *Nymphalidae*).

Работа по любой подчиненной программе завершается переходом на основную «CLAVE». Однако пользователь по желанию может сразу обратиться к любой из программ, если ему известно, к какому семейству относится идентифицируемая особь.

Программы сами по себе организованы чрезвычайно просто. Вначале вводится ряд обозначений-символов (например, GE# = «Род», ES# = «Единственный вид» и т. п.). Эта часть программы демонстрируется пользователю только в нужные моменты. Затем ПЭВМ через дисплей знакомит пользователя с правилами диалога: «Чтобы воспользоваться этим ключом, ответьте на следующие вопросы, написав «Д», если ответ «Да»; «Н», если ответ «Нет»; соответствующий числом, если ответы упорядочены. Затем нажмите клавишу «CR». Постарайтесь не давать ошибочных ответов! Все понятно? Ответьте «Д» или «Н». После ознакомления начинается собственно определение. ПЭВМ предъявляет список соответствующих таксонов и осведомляется, знает ли пользователь, к какому из них принадлежит идентифицируемая в данный момент особь. Если таксономическая принадлежность известна, ПЭВМ переходит к более низким таксономическим уровням. Если пользователь не может сам отнести особь к определенному таксону, то начинается опрос по характерным признакам с обязательным ответом «да» или «нет». Вопросы со стороны ПЭВМ продолжаются до тех пор, пока определение не будет доведено до конца. (Здесь есть ряд исключений, о которых будет сказано в дальнейшем.) Закончив определение очередной особи, ПЭВМ делает запрос: «Будете продолжать?», имея в виду «работать с данной программой». Если следует отрицательный ответ, то ПЭВМ вежливо прощается с пользователем: «Всего хорошего!» и выключается.

Из изложенного ясно, что все программы построены на условных

переходах. Если условие выполняется («Д»), ПЭВМ осуществляет переход на один адрес программы, если «Нет» — на другой.

К удобствам, которые предоставляют рассматриваемые программы, следует отнести то, что ПЭВМ автоматически информирует пользователя как о подробной таксономической принадлежности рассматриваемой особи (подотряд, семейство, триба и т. д.), так и о сходных видах чешуекрылых, а также о синонимах, если они имеются.

После рассмотрения основной структуры книги Ф. Ф. Рубио необходимо обсудить те вопросы, которые были поставлены в предыдущем тексте. Они касаются прежде всего соотношения между дихотомическими (иерархическими) и политомическими (образными) ключами. Из изложенного выше вполне ясно, что Ф. Ф. Рубио в своей работе использовал традиционные («книжные» — напечатанные в книгах) дихотомические ключи. Они испытаны многолетней практикой систематиков и вполне себя оправдывают для целей обучения и начального ознакомления с предметом. Эти ключи удобны и удовлетворительны в тех случаях, когда высшие таксоны насчитывают малое число таксонов низшего ранга, и тогда, когда одновходовые иерархические ступени четко выражены (теза — антитеза). Сравнение объектов и дополнение системы новыми таксонами в них затруднено. (Это один из основных недостатков в системе программ, составленных наподобие рассматриваемых.) Система замкнута. Для того, чтобы ввести в нее новые виды, в некоторых случаях потребуется относительно небольшая переработка программ. Благодаря некоторым особенностям языка программирования «Бейсик», это можно сделать. Но если в систему потребуется ввести новый род, не говоря уже о трибе (подсемействе или семействе), то необходимо будет в корне переделывать весь пакет программ. Как уже упоминалось, предложенная система программ является замкнутой. В частности, никоим образом не предусмотрены случаи, когда какую-либо особь не удается определить. Скажем, в руки исследователя попала бабочка, ранее не отмеченная в фауне полуострова. В других случаях отдельные особи настолько отличаются от «типа», хранящегося в памяти машины, что их тоже невозможно будет определить. ЭВМ должна «честно признаваться в этом». Программы же, составленные Ф. Ф. Рубио, не позволяют ей поступить таким образом. В сложных случаях, когда в ходе определения иногда возможны ошибки (обусловленные нечеткостью проявления признака), такие ключи мало чем могут помочь. Сложности возникают и в тех случаях, когда ряд таксономически значимых признаков по каким-либо причинам отсутствует (ископаемые остатки, материалы из погадок и т. п.). Во всех подобных случаях прекрасно себя зарекомендовали политомические ключи-идентификаторы объекта.

Разные типы политомических ключей, предусматривающих несколько альтернатив выбора соответствующих признаков и группировок объектов, были предложены еще в 60-е годы. В отличие от дихотомических ключей, политомические учитывают одинаковое число аспектов. Однако в этом случае производится свободное их сопоставление (комбинаторика) вне иерархической зависимости в целостной системе признаков. Цифровое кодирование в информационно-поисковой системе позволяет максимально унифицировать, систематизировать и минимизировать систему признаков. Это в свою очередь позволяет создавать многовходовые (многоаспектные) системы, облегчает проведение логических операций (типа сравнений, объединений и т. п.) над объектами и признаками и дает возможность широко использовать ЭВМ.

В настоящее время политомические ключи на основе баз данных применяются в ботанике, зоологии, медицине, ветеринарии и др. областях науки. Только за период с 1960 г. по 1975 г. число публикаций, где использовали политомический подход, превысило 300 названий.

К вопросу о типах используемых ключей непосредственно примыкает вопрос о количестве признаков, требуемых для идентификации тех или

иных объектов. К примеру, в обсуждаемой книге для определения 22 видов белянок (Pieridae), населяющих Пиренейский полуостров, с использованием традиционных подходов и признаков, Ф. Ф. Рубио пришлось использовать 22 признака. Программа, предназначенная для выполнения этой работы, включает 173 оператора. Однако из теории информации (Вентуаль, 1969) известно, что при наличии q объектов, заданных бинарными описаниями, для их безошибочного распознавания необходимо как минимум $p_{min} = \ln q / \ln 2$ признаков. То есть для определения 22 видов бабочек семейства Pieridae, живущих на Пиренейском полуострове, может оказаться достаточным всего 5 признаков. В данном случае мы не ставили перед собой задачу предложить оптимальный ключ, но совершенно ясно, что ключ, использованный Ф. Ф. Рубио, явно избыточен.

В связи с этим необходимо вернуться к нашему замечанию в скобках относительно «книжных» дихотомических ключей. Это замечание имело целью подчеркнуть, что существует принципиальное различие к определению объектов у «человека с определителем-книгой» и у «человека за определителем-ЭВМ». Если в первом случае человек вынужден непрерывно перелистывать страницы — иначе он просто не может работать с книгой, то нельзя же заставлять его без конца «листать память ЭВМ» (как это будут делать любые дихотомические ключи, реализованные на ЭВМ; лучший пример — программы Ф. Ф. Рубио). Вычислительные машины могут и должны работать по-другому. Человеку, действительно, трудно, а порою он совершенно не способен быстро пересмотреть многочисленные комбинации из нескольких признаков и выбрать из них ту, которая соответствует определенному таксону. Для ЭВМ же перебор комбинаций — родная и излюбленная стихия. Ей потребуется всего несколько микросекунд, чтобы по множеству признаков, характеризующих данный объект, выяснить тот таксон (вид), к которому этот объект принадлежит. Следовательно, быстрые и безошибочные определители, ориентированные на ПЭВМ, должны строиться только на основе политомических ключей (баз данных). Даже в тех случаях, когда исследователю приходится работать с поврежденными объектами (ряд признаков отсутствует), потребуется очень небольшая избыточность ключа.

Несмотря на все замечания, приведенные ранее, справедливость требует подчеркнуть, что основные принципы, использованные Ф. Ф. Рубио для реализации дихотомического ключа, могут быть рекомендованы к применению для решения задач подобного рода. Это прежде всего касается общего оформления программ. При наличии ПЭВМ с ограниченной памятью оперативного запоминающего устройства вполне логичным будет подразделение программы на ряд подпрограмм и их взаимная увязка. Определенные удобства состоят и в том, что в зависимости от уровня подготовки пользователя, он может начинать определение с таксона любого уровня или же быстро перейти на требуемый уровень. Такой подход стал возможным благодаря тому, что ЭВМ сразу же задает прямые вопросы: «Вам известно, к какому семейству (трибе, роду) принадлежит данная особь?» Иногда же машина даже экзаменует определяющего: «Вы уверены, что данная особь принадлежит к семейству (трибе, роду)...?». Если полной уверенности нет, следуют уточняющие вопросы, после которых может оказаться, что ЭВМ помочь не в силах.

Книга прекрасно оформлена, но, несмотря на то, что клише были изготовлены с ЭВМ, в них вкрались некоторые небольшие ошибки, которые могут привести к сбоям в работе машины, если при вводе программ оператор вовремя их не заметит, хотя они и очевидны.

Появление книги Ф. Ф. Рубио следует приветствовать хотя бы потому, что это первый опыт применения ЭВМ для идентификации чешуекрылых. В ней на практике показано, что даже при наличии относительно скромных персональных компьютеров стопочка тоненьких магнитных дисков вполне может заменить целую библиотеку многотомных определителей. Кроме того, она позволила уяснить, что существует наущная не-

обходимость в разработке специально ориентированных систем управления базами данных (СУБД) — только на их основе может быть осуществлено быстрое и безошибочное определение любых объектов на ЭВМ. Безусловно, диахотомические ключи, реализованные на ЭВМ, могут и должны сохраняться. Они могут облегчить обучение начинающих, хотя для этого они должны быть снабжены иллюстративным материалом, реализованным на ЭВМ. Специалисты же должны работать с самыми быстрыми и удобными СУБД. Возможно, что для перехода к этим СУБД придется пересмотреть всю систему диагностических признаков, используемых при классификации тех или иных объектов (задача для энтомологов). Кроме того, необходимо разработать систему отбора признаков, которые в совокупности позволяли бы дифференцировать (классифицировать) любые наборы объектов данного класса с наибольшей скоростью и точностью, но при минимальных затратах машинного времени и времени специалистов (задача для математиков и программистов).

В связи с бурным развитием микропроцессорной техники любая исследовательская лаборатория и научная группа может приобрести персональную ЭВМ для решения задач рассмотренного типа и множества других. В этой ситуации биологу, безусловно, необходимо иметь отчетливое представление об области применения ЭВМ и способах решения задач на них, основных приемах программирования, особенностях работы программиста и оператора, временных и материальных затратах на разработку программ и решение задач на ЭВМ.

Інститут зоології АН України
(252601 Київ)

Получено 10.12.91

Декілька зауважень про використання ЕОМ для класифікації біологічних об'єктів.
Михалевич О. А., Некрутенко Ю. П.— Вестн. зоол., 1992, № 3.— Аналітичний огляд концептуальних підвалин виданого в Іспанії «Комп'ютеризованого ключа для визначення денних лускокрилих Іспанії» Ф. Фернандеса-Рубіо, як прецедента для створення подібних вітчизняних підручників по різних групах тварин.