

## ИНТЕРНЕТ-ОПРЕДЕЛИТЕЛИ БИОЛОГИЧЕСКИХ ОБЪЕКТОВ. 5 ЛЕТ СПУСТЯ

А.Г. Кирейчук, А.Л. Лобанов, И.С. Смирнов, А.А. Иночкин, С.Д. Степаньянц

Разработка электронных определительных ключей и систем экологического мониторинга получает за последние годы все более интенсивное развитие, связанное с совершенствованием компьютерных технологий и возрастающей ролью оценки и сохранения биологического разнообразия при воздействии антропогенной нагрузки. В связи с распространением Интернета, становится актуальным перевод программ ввода и работы с ключами в эту среду, а также представление данных систем в Интернете для более оперативного использования в целях мониторинга экологической ситуации, управления ресурсами и в образовании. Практическое применение интерактивных определителей (компьютерных диалоговых ключей) для видов, наносящих ущерб народному хозяйству или являющихся ресурсными или полезными, или видов, которые могут быть встроены в системы экологического мониторинга, всегда имело важное значение. В условиях быстрых изменений современной биоты и вторжений чужеродных вредоносных видов животных и растений компьютерные определители приобретают особую роль. С появлением Интернета стало возможным дистанционное использование компьютерных ключей пользователями, находящимися в любой точке планеты. Своевременная идентификация тех же чужеродных вселенцев и принятие соответствующих мер, предотвращающих их нежелательное распространение, зачастую могут сэкономить колоссальные финансовые и человеческие ресурсы. Это обстоятельство вызвало новую волну интереса к интерактивным ключам.

Первые опыты применения электронно-вычислительных машин (ЭВМ) для определения таксономической принадлежности биологических объектов были осуществлены в середине 60-х годов, когда ЭВМ еще были для биологов экзотической техникой. Однако эти опыты принесли обнадеживающие результаты [1-9], которые с появлением и дальнейшим широким распространением Интернета позволили перейти к следующей задаче - разработке идеологии идеального определителя и программы для дистанционного определения [10-12]. В ходе подготовки к реализации задуманного, при поддержке РФФИ (грант N 02-07-90105), в течение 2002-2004 гг. был создан электронный Атлас жуков России и сопредельных стран, который размещен на колеоптерологическом сайте Зоологического института РАН: <http://www.zin.ru/Animalia/Coleoptera/rus/atlas.htm>, а в дальнейшем (2005 г.) была получена поддержка РФФИ специально на разработку программного обеспечения к многоходовым политомическим определителям с использованием сети Интернет. Для решения этой задачи 6 лет назад к разработке первого варианта удалось привлечь профессиональных математиков из СПбГУ. В 2007 г. была создана и функционировала в Интернете программа WebKey-X. Она имела ряд недостатков, которые долго не удавалось устранить. Через три года программистом из Москвы Александром Алексеевичем Иночкиным был получен более приемлемый вариант интернет-ключа, с работой которого можно познакомиться в Интернете: <http://www.zin.ru/projects/webkey-x/superkey>. На рисунке 1 приведен скриншот первой страницы программы Superkey.

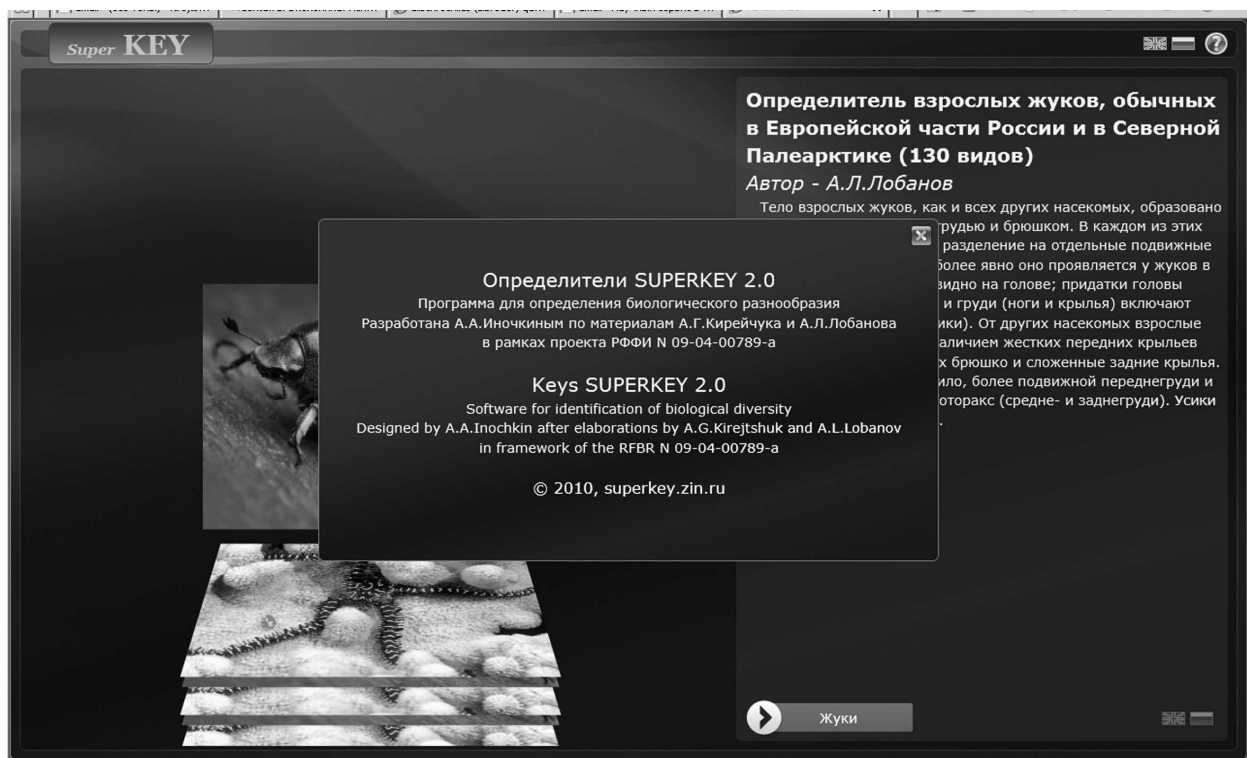


Рис. 1. Скриншот первой страницы Программы Superkey

Интерактивная программа для определения биологических объектов Superkey использует современную и перспективную СУБД Microsoft SQL Server. Клиентская программа написана на основе технологии Silverlight, что позволило создать удобный динамичный интерфейс. Возможность инсталляции приложения на клиентский компьютер позволяет уменьшить время запуска приложения и оптимизировать коммуникационные расходы, а возможность запуска приложения в любом из популярных браузеров обеспечивает простой путь для использования определителя. Выбор таксонов производится на основе расчета вероятности совпадения выбранных пользователем признаков с наиболее полным набором признаков таксона. Это позволяет проводить поиск таксонов даже при ошибках пользователя в выборе и описании некоторых признаков. Программа предполагает многовходовость - возможность начать идентификацию таксона с любого удобного пользователю признака. Описываемый интерактивный определитель позволяет использовать диагностические признаки, расположенные в виртуальной "книге" в зависимости от выбранных пользователем предпочтений по фильтрации и сортировке, в двух маленьких окнах для выбора, которые расположены справа от "книги". Если указана сортировка "по диагностической ценности", порядок признаков меняется после каждого очередного шага определения. Предпочтения по фильтрации и сортировке могут меняться пользователем на любом этапе определения. Предполагается, что программа может включать базы данных на различных языках и осуществлять переход из таблиц на одном из языков в таблицы на другом языке на каждом шаге определения. В сравнении с прежней (WebKey-X) новая программа отличается простотой в использовании, значительной надежностью и отличается быстротой. Немаловажным обстоятельством явилось также то, что Superkey имеет оригинальный и не перегруженный дизайн, более дружелюбный для пользователей, не знакомых с особенностями программирования и конкретными базами данных. Это определило отказ от продолжения работы над WebKey-X в пользу Superkey.

Интернет-определитель Superkey выглядит на экране как раскрытая "книга", на страницах которой помещены признаки объектов, включенных в базы данных. В процессе определения используются диагностические признаки, отобранные и расположенные в "книге", в зависимости от выбранных пользователем предпочтений. Два маленьких окна для выбора групп признаков, которые расположены справа от "книги". Если указана сортировка "по диагностической ценности", порядок признаков меняется после каждого очередного шага определения и на первые места выдвигаются признаки, которые теоретически должны скорее привести к окончанию диагноза. Предпочтения по фильтрации и сортировке признаков могут меняться пользователем на любом этапе определения. Для перехода к другим признакам (вперед-назад) нужно навести мышкой на любой внешний угол листов "книги" с состояниями того или иного признака, которому посвящена данная страница, нажать левую кнопку мыши и потянуть за "уголок страницы". Состояние признака выбирается и отменяется кликом левой кнопки мыши на галочку в левом верхнем углу изображения этого состояния. Над "книгой" в небольшом окошке "найти признак" можно для быстрого поиска ввести термин или его часть, которые входят в формулировку признака, находящегося в "книге" при выбранных предпочтениях, т.е. заданных указанными выше способами фильтрации и сортировки, а для поиска любого из имеющихся диагностических

признаков среди всего набора таблицы "все признаки".

Каждый шаг определения отображается справа внизу от маленьких окон с фильтрами предпочтения признаков, в виде строк: признак, его состояние и маленькой иконки с изображением этого состояния. Каждый шаг определения уменьшает число оставшихся таксонов и среди этих оставшихся таксонов может оказаться тот, к которому относится определяемый экземпляр. Первые из наиболее вероятных таксонов появляются в нижнем ряду изображений в порядке уменьшения вероятности соответствия признаков определяемого объекта с введенными признаками слева направо. При каждом дальнейшем шаге определения этот ряд изменяется в зависимости от вновь установленной (пересмотренной программой) вероятности.

При достижении такого уровня определения, когда в нижнем ряду остаются лишь немногие из вероятных таксонов с совпадением по всем использованным признакам, справа в иконках нижнего ряда появляются изображения таксонов с затемненным фоном, в которых не все использованные состояния признаков совпадают. Если определенный таксон остается единственным со 100% совпадением всех выбранных состояний признаков, то он обводится зеленой рамкой. Для того чтобы просмотреть все соответствия и несоответствия по состояниям признаков следует открыть кликом левой кнопки мыши по соответствующей иконке "карточку таксона" с крупным изображением представителя этого таксона и перечислением всех его признаков, среди которых использованные совпадающие признаки засвечены зеленым фоном и несовпадающие засвечены красным и последние всегда перенесены в конец списка. В "карточке таксона" в верхнем правом углу находятся белая, зеленая и красная кнопки (белая - для просмотра всех признаков; зеленая – для просмотра выбранных и совпадающих признаков, а красная – для выбранных несовпадающих признаков), а в левом нижнем углу находится кнопка поиска веб-страниц по этому таксону в "Google".

Исправление любого шага определения может производиться по расположенному справа от "книжки" списку состояний признаков, выбранных на каждом из шагов. При наведении мышью на формулировку признака или состояния этого признака в правом верхнем углу появляется небольшая кнопка с красным крестом, кликнув на которую можно отменить выбор того или иного шага. Это также влечет изменение списка таксонов, отображенных иконками в нижнем ряду таксонов, с учетом оставшихся признаков. На рисунке 2 приведен скриншот завершающего этапа определения жука программой Superkey.

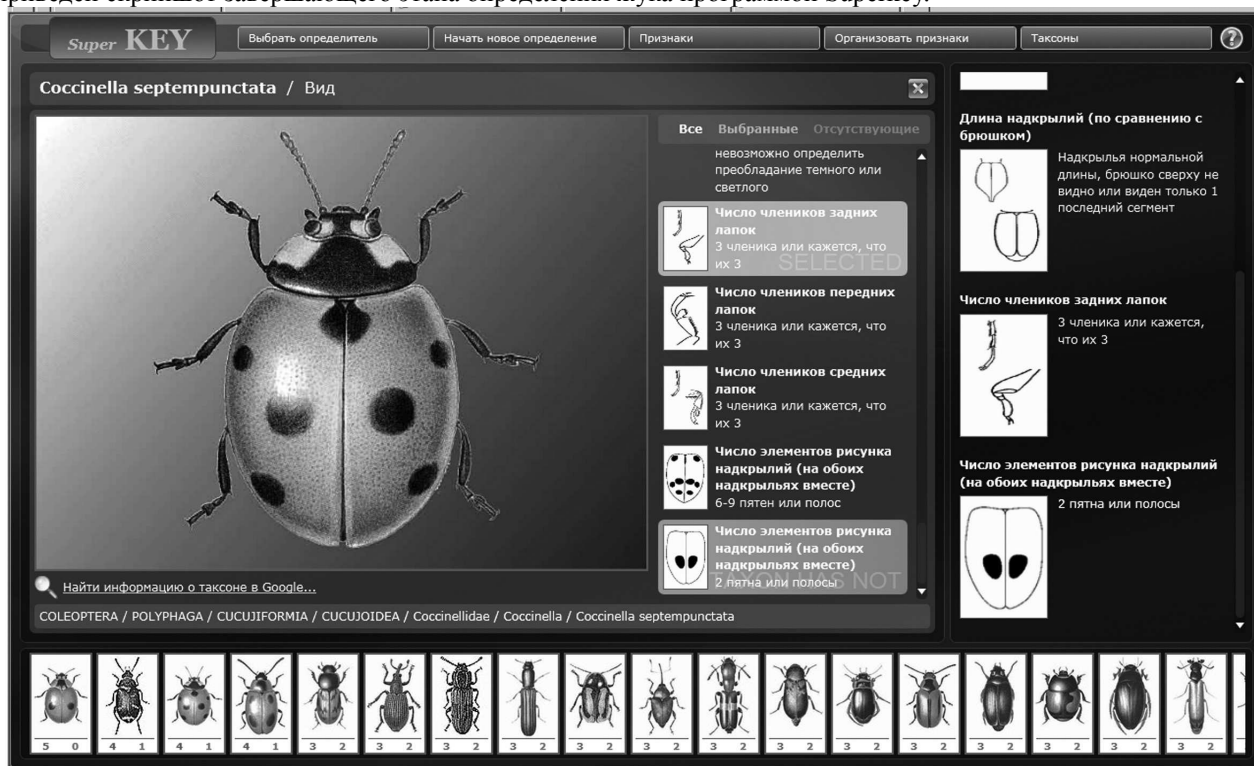


Рис. 2. Скриншот завершающего этапа определения Программой Superkey

В 2010 году система выставлена на портале ЗИН РАН с двумя базами данных, позволяющими определять некоторых обычных в России представителей отряда жуков и офиур (змеихвосток) арктических морей России (<http://www.zin.ru/projects/webkey-x/superkey/>). Сейчас создаются и наполняются базы данных по таксонам различных групп животных, объединяющие информацию о систематической иерархии таксонов, различных состояний диагностических признаков у каждого из них и иллюстрациями, включающими рисунки и фотографии, как в этих базах данных, так и в Интернете. В качестве основы для проверки определяемых объектов будут использоваться атласы видов на сайтах портала ЗИН РАН. Продолжает развиваться крупнейший в мире Интернет-атлас по жесткокрылым насекомым <http://www.zin.ru/Animalia/Coleoptera/eng/atlas.htm>. В

дальнейшем планируется интеграция баз данных определителя с уже имеющимися Интернет-страницами атласов на портале ЗИН РАН, а также создание новых страниц атласов по другим группам животных и интеграция их с информационными ресурсами портала ЗИН РАН и другими источниками, чтобы в результате определения наблюдатель получал максимальный объем полезной информации относительно таксона, к которому принадлежит определяемый объект. Предполагается также, что создаваемая система станет наиболее полным и разносторонним справочным источником по группам животных, для которых будут созданы соответствующие определители, интегрирующие всю информацию по каждой из групп.

Интернет-определитель Superkey является универсальной программой, которая может определять любые объекты (как живые, так и неживые), различающиеся по признакам, которые могут быть формализованы и помещены в соответствующие базы данных. Прямо со страниц определителя с помощью поисковика "Google" можно получить дополнительную информацию об определенном вами таксоне в Интернете, а также, помимо этого, программа работает с использованием всех включенных в нее баз данных и может быть интерактивно связана с базами других Интернет сайтов. Superkey тестирован на базах данных, включающих более 900 таксонов, которые диагностируются более 340 признаками. В России и в других странах не существует аналогов подобного программного обеспечения, при этом программа Superkey. При этом программа Superkey предусматривает ряд полезных функций, которые до сих не использовались при разработке программ для определения биоразнообразия. К настоящему времени, выпущены на лазерных дисках только определители для антарктических морских ежей и тропических голотурий Австралии, а также для определения семейств жуков и их личинок. Составление компьютерных определителей очень трудоемкая работа, требующая усилий многих специалистов, но созданный Интернет-определитель облегчает решение различных задач тысячам пользователей и таким образом оправдывает затраты на его создание.

Работа поддержана грантами РФФИ NN 05-07-90179-а, 05-07-90354-а, 09-04-00789-а; проектом N 4 подпрограммы "Изучение и исследование Антарктики" ФЦП "Мировой океан" и программой Президиума РАН "Биоразнообразие".

#### ЛИТЕРАТУРА:

1. Лобанов А.Л. Принципы построения определителей насекомых с использованием электронных вычислительных машин. - Автореферат диссертации на соискание ученой степени канд. биол. наук. Л.: ЗИН АН СССР, 1983. С. 1-19.
2. Лобанов А.Л., Дианов М.Б. Компьютерная диагностическая система BIKEY и возможности ее использования в защите растений // Защита растений в условиях реформирования агропромышленного комплекса: экономика, эффективность, экологичность. Всеросс. съезд по защите растений. Тезисы докл. 1995. С. 548-549.
3. Лобанов А.Л., Смирнов И.С. Место и роль информационных технологий в исследованиях Зоологического института РАН // Фундаментальные зоологические исследования: Теория и методы. (По материалам Международной конференции "Юбилейные чтения, посвященные 170-летию Зоологического института РАН", проходившей 23-25 октября 2002 г.), М.-СПб.: Товарищество научных изданий КМК. 2004: 283-318 (резюме на англ. яз.).
4. Свиридов А.В. Типы биодиагностических ключей и их применение. М., Зоомузей МГУ. 1994. 111 с.
5. Dallwitz M.J. A flexible computer program for generating identification keys // Syst. Zool., 1974, 23: 50-57.
6. Dallwitz M.J., Paine T.A. and Zurcher E.J. Interactive identification using the Internet // In: "Towards a global biological information infrastructure - challenges, opportunities, synergies, and the role of entomology", European Environment Agency Technical Report, 70. EEA, Copenhagen, 2002: 23-33.
7. Lobanov A.L., Stepanjants S.D., Dianov M.B. 1996. Dialogue computer system BIKEY as applied to diagnostics of Cnidaria (illustrated an example of hydroids of the genus Symplectoscyphus) // Scientia Marina (Special volume: Advances in Hydrozoan Biology; S.Piraino, J.Bouillon et al. (eds.)). V. 60, N 1. P. 211-220.
8. Pankhurst R.J. (ed.). Biological Identification with Computers. Academic Press, London. 1975. 333 pp.
9. Pankhurst R.J. Practical taxonomic computing. Cambridge University Press, Cambridge, 1991. 202 pp.
10. Лобанов А.Л., Кирейчук А.Г., Смирнов И.С., Дианов М.Б., Граничин О.Н. Интернет и интерактивные определители биологических объектов // Научный сервис в сети ИНТЕРНЕТ: Труды Всероссийской научной конференции (19-24 сентября 2005 г., г. Новороссийск). - М.: Изд-во МГУ, 2005. с. 132-134.
11. Лобанов А.Л., Кирейчук А.Г., Смирнов И.С., Граничин О.Н., Вахитов А.Т., Дианов М.Б. 2006. К реализации идеального интерактивного определителя в Интернете для биологических объектов // Научный сервис в сети ИНТЕРНЕТ: Труды Всероссийской научной конференции (18-23 сентября 2006 г., г. Новороссийск). - М.: Изд-во МГУ, 2006. 202-204.
12. А.Т. Вахитов, О.Н. Граничин, А.Г. Кирейчук, А.Л. Лобанов. 2009. Параллельный алгоритм обучения для интерактивного политомического определителя биологических видов // Научный

сервис в сети ИНТЕРНЕТ: масштабируемость, параллельность, эффективность” г. Новороссийск, 21- 26 сентября 2009 г. Труды Всероссийской суперкомпьютерной конференции. Издательство Московского университета, 2009, 332-334.