

А.М.Лях*, С.Г.Лелеков**

**Институт биологии южных морей им. А.О.Ковалевского, г.Севастополь*

***Севастопольский государственный университет, г.Севастополь*

АДАПТИВНАЯ ЭКСПЕРТНАЯ СИСТЕМА ДЛЯ ТАКСОНОМИЧЕСКОЙ ИДЕНТИФИКАЦИИ ОРГАНИЗМОВ

В процессе таксономической идентификации исследователь создает диагноз организма, руководствуясь подсказками экспертной системы (ЭС). Этот диагноз может описывать новый для системы вид, но ЭС не сохраняет такую информацию. Предложена концепция адаптивной ЭС, которая запоминает значения признаков организма, предоставляет пользователю возможность связывать значения признаков с названиями таксонов и автоматически обучается распознаванию новых таксонов. Адаптивная ЭС располагает средствами нивелирования семантической неоднозначности таксономических диагнозов, для чего применяется теория нечетких множеств. Для наполнения адаптивной ЭС определительной информацией предложена концепция структурированных диагнозов, составленных по жесткому иерархическому плану, в который включены таксономически значимые и незначимые признаки. Информационно насыщенная структура признаков позволяет наполнять адаптивную ЭС диагнозами новых таксонов.

КЛЮЧЕВЫЕ СЛОВА: *таксономическая идентификация, определительные ключи, экспертные системы, лингвистические переменные, нечеткие множества, структурированные диагнозы*

Важным этапом большинства биологических исследований является таксономическая идентификация организмов. За таксономическим названием скрывается информация о строении, функционировании, экологии, биологии и распространении организма.

Таксономическая идентификация заключается в сопоставлении значений признаков организма с признаками таксонов, перечисленных в специальной литературе – определителях, содержащих набор определительных ключей. Для корректной идентификации необходимо, чтобы признаки организма совпали с признаками одного из таксонов, в противном случае таксономическое положение организма останется неизвестным. Ошибка исследователя приводит к неверной идентификации. Так как определительные ключи ограничены объемом содержащейся в них информации, то с их помощью невозможно идентифицировать новые таксоны.

Экспертные системы (ЭС) имеют большие возможности по сравнению с определительными ключами. Они оценивают степень сходства признаков организма с признаками таксонов и выдвигают гипотезы об их подобии. Они устойчивы к ошибкам пользователя и умеют идентифицировать организм, даже если его признаки частично совпадут с признаками таксонов. Алгоритмы ЭС подбирают таксоны похожие на исследуемый организм. Такой алгоритм определения реализован в таксономической экспертной системе «Taxex» [1], частично в системах «Superkey» [2], «IntKey» (*delta-intkey.com*) и «Lucid» (*lucidcentral.com*).

© А.М.Лях, С.Г.Лелеков, 2014

Адаптивная экспертная система. В процессе таксономической идентификации пользователь выделяет значения признаков организма и вводит их в ЭС. Таким способом он описывает организм – создает диагноз, руководствуясь подсказками системы. Главный недостаток ЭС заключается в том, что они не сохраняют эти диагнозы.

Несовпадение признаков организма и таксонов означает, что пользователь ошибся или он определяет новый для системы таксон: вид-вселенец, видовой морфотип или тератологическую форму. Экспертная система отнесет незнакомый организм к известному таксону и оценит степень совпадения признаков – выдвинет гипотезу о таксономическом положении организма. Окончательное решение о качестве определения принимает человек. Если он не согласен с гипотезами ЭС, то обратится к другим источникам, повторно выделит признаки организма и сопоставит их значения с признаками таксонов. ЭС результаты повторных определений не зафиксируют, потому что в них не предусмотрен механизм обратного отклика. Поэтому, когда исследователь снова пытается идентифицировать с помощью ЭС представителей того же вида, ему не удастся выполнить корректное определение. Процесс повторится. Научить ЭС определению новых таксонов может только эксперт – он вносит в информационную базу данных системы новую информацию.

Для того, чтобы избежать многократных определений предлагается дополнить ЭС подсистемой, которая будет запоминать значения признаков организма, информировать пользователя о том, что введенная комбинация признаков уже содержится в базе данных, и предоставит ему возможность связывать значения признаков с новым таксономическим названием. Таким образом, система усваивает информацию о новых таксонах и обучается их опознанию. Подобные ЭС относятся к классу адаптивных экспертных систем.

Авторы разработали прототип адаптивной ЭС «Таксакейс», которая размещена в сети интернет по адресу taxakeys.org. На основе системы созданы интерактивные определители некоторых представителей фауны Севастопольского региона: высших таксонов зоопланктона (автор описаний В.В.Мельников), клadoцер (Cladocera; автор описаний В.В.Мельников), копепода (Copepoda; авторы описаний А.В.Темных, М.И.Силаков), рыб семейства собачковых (Blennidae) и сельдевых (Clupeidae) (автор описаний С.А.Царин).

Адаптивная экспертная система совмещает функцию диагностической системы и таксономической базы данных. Из более 20 рассмотренных в работах [1 – 3] диагностических систем только три «Taxex», «Superkey» и «Lucid» являются настоящими экспертными системами, одна «IntKey» – частично выполняет функции ЭС, остальные – это оболочки для интерактивных компьютерных ключей. Другими словами в области биологической диагностики предложенная концепция адаптивной ЭС не имеет аналогов. Хотя еще в 80-е гг. прошлого столетия авторы статьи [4] указывали, что ЭС должны сохранять сеанс работы с пользователем для повторной идентификации. Однако в биологических ЭС никто так и не реализовал эту возможность.

Адаптация экспертных систем к естественному языку. Другой положительной стороной адаптивных систем является способность учитывать человеческую склонность к нечетким, размытым формулировкам. Неоднозначность естественного языка, применяемого экспертами для описания объ-

ектов и пользователями для ответов на вопросы экспертной системы, одна из основных проблем создания и использования компьютерных баз знаний.

Эксперт описывает организм в системе признаков, каждый из которых является лингвистической переменной, сформированной из слов естественного языка. Например «отношение длины живота к цефалотораксу у копепода» будет лингвистической переменной, если она принимает значения: «длина живота *много меньше* длины цефалоторакса», «длина живота *меньше* длины цефалоторакса», «длина живота *немного меньше* длины цефалоторакса», «длина живота *приблизительно равна* длине цефалоторакса», «длина живота *равна* длине цефалоторакса», «длина живота *больше* длины цефалоторакса». Пользователь вынужден использовать терминологию эксперта, хотя их представление о сущности того или иного значения лингвистической переменной могут отличаться.

Для работы с лингвистическими переменными используется аппарат нечетких множеств. Значения лингвистических переменных описываются нечеткими переменными, которые могут быть заданы: 1) на конечных множествах, допускающих линейный порядок с дискретными и интервальными значениями; 2) на конечных множествах, не допускающих линейный порядок.

Подобный подход полностью не решает проблему неоднозначности естественного языка: эксперт не может задать все возможные значения лингвистической переменной ввиду многообразия способов построения ее конструкций. Многообразие значений лингвистических переменных возникает при использовании слов-модификаторов, относящихся к количественным наречиям: очень, почти, едва, слегка и т.п. [5]. Возможным решением здесь является перечисление экспертом некоторых базовых значений, в терминах которых и происходит описание объекта, а пользователю предоставляется возможность использовать смысловые модификации базовых терминов. Таким образом, ЭС адаптируется к использованию нестрогих семантических конструкций естественного языка.

Структурированные диагнозы таксонов. Для успешного функционирования адаптивных ЭС необходимо, чтобы система была насыщена определительными признаками. В этом случае пользователь системы сможет описать признаки практически любого таксона и, впоследствии, сохранить их на базе знаний системы. Для упрощения работы с признаками и составления диагнозов таксонов предлагается концепция структурирования диагнозов таксонов.

Количество признаков в диагностической системе определяет ее таксономическое разрешение. Чем больше признаков в диагнозе, тем больше его информативность и выше вероятность различить похожие таксоны. Однако часто авторы определителей сознательно опускают «незначительные» признаки низших таксонов. Они полагают, что читатели изучают определитель сверху вниз: от таксонов верхнего ранга – отрядов, семейств, родов, – к таксонам нижнего – видов и подвидов. Постоянное повторение признаков высших таксонов значительно увеличивает объем определителей и, с точки зрения авторов, выглядит нецелесообразным.

Краткие диагнозы таксонов, включающие только важные признаки, ввел в систематику Карл Линей – он первым перестал употреблять глаголы

в описаниях организмов. К сожалению, краткость со временем превратилась в настоящее бедствие: диагнозы стали столь лаконичными, что распознать новые виды и составить списки синонимов к ним стало затруднительно [6]. Например, авторы работ [7 – 9] приводят в диагнозах видов диатомовых водорослей только три позиции: два размера панциря и число штрихов в 10 мкм. Даже с помощью помещенных в работы фотографий диатомовых очень сложно использовать эти диагнозы для опознания видов.

Развитие методов молекулярной идентификации организмов по фрагментам ДНК или ДНК-штрихкодирование [10] привело к появлению «турбо-таксономии» [11]. Это синтез морфологического и молекулярного подходов к систематике, в котором для описания видов используются ДНК-штрихкод, краткий диагноз с несколькими значимыми морфологическими признаками вида и высококачественное изображение типового экземпляра вида. Так как морфологическое описание в данном случае ограничено небольшим числом признаков, то скорость описания новых видов зависит от скорости расшифровки фрагментов ДНК. Например, подготовка к печати работы с быстрыми (турбо) описаниями 101 вида жуков-долгоносиков заняла один год [12]. За это время традиционным способом обычно описывают 10 – 15 видов [13]. Поэтому включение в диагнозы максимального числа признаков, в том числе малоинформативных, необоснованно раздувает описания, занимает много времени и не приводит к четкому разграничению видов [13].

Однако краткий диагноз и изображение видового образца недостаточны для опознавания организма по морфологическим признакам, а ДНК-секвенирование – дорогостоящий и трудоемкий метод. Поэтому доступность полнотекстовых морфологических описаний видов актуально для многих биолого-экологических исследований. Такие описания естественны для человека, они ориентированы на человека и предназначены для человека, тогда как штрихкод создан для машин и человек им пользоваться не способен.

Одной из основных проблем систематики является отсутствие формализованной схемы описаний. Исследователь использует определитель с целью найти описание конкретного таксона. Когда у одного вида указано значение признака, а у другого нет, это может означать, что этот признак у вида отсутствует, он упомянут в диагнозе рода или автор не знает значения признака. Исследователь не знает правильного ответа, поэтому фрагментарность таксономических диагнозов затрудняет идентификацию организмов.

Структурированные диагнозы таксонов – это диагнозы, составленные по жесткому иерархическому плану, в который включены таксономические значимые и незначимые признаки таксонов. Включение максимального числа признаков повышает информативность описаний и таксономическое разрешение диагностических систем. Основываясь на большом числе признаков, заложенных в диагностическую систему, можно описывать новые таксоны. Адаптивная диагностическая система на данном этапе выступает в качестве хранилища описаний. После того, как человек ассоциирует с описаниями название таксона, система преобразует их в таксономический ключ.

Диагнозы, составленные по единому плану, легко сопоставить, увидеть различия в описаниях таксонов, обнаружить пропуски. Для некоторых групп организмов уже разработаны планы составления диагнозов. Напри-

мер, подобный иерархический структурированный план предложен для диатомовых водорослей [14] и его использовала Л.Н.Бухтиярова [15].

Структурированные диагнозы невозможно составить без унификации терминологии. Терминология должна быть однозначной и понятной каждому, иначе диагнозы, составленные разными авторами, окажутся несопоставимыми. Особенно это важно для крупных таксономических групп организмов с неустоявшейся систематикой. Отсутствие единой терминологии на ранних этапах развития систематики привело к тому, что ревизию таксонов, описанных в старых работах, невозможно выполнить без привлечения типовых образцов.

Структурированный диагноз – это готовая строка определительной таблицы ЭС. Она представляет собой прямоугольную матрицу, в которой столбцы содержат признаки таксонов, строки – названия таксонов, а ячейки – значения признаков таксонов. Структурированные диагнозы таксонов естественным образом интегрируются в экспертную систему и служат основой для идентификации организмов. Таксономические данные, накопленные в многочисленных работах, позволяют создавать структурированные диагнозы и строить определительные таблицы без помощи экспертов. Р.Е.Дринкватер [16] создал электронный ключ для определения австралийский росянок (*Drosera*), используя только опубликованные описания видов.

Немаловажным аспектом восприятия структурированных диагнозов является их визуальное представление. Иерархические ступени описаний должны быть хорошо различимы при чтении, выделены отступами или шрифтовым стилем, жирностью, цветом, разрядкой, начертанием. Такой диагноз легко читаем. Грамотно оформленное структурированное описание вида легко сравнить с другими, выделить похожие и отличающиеся признаки. Л.Н.Бухтиярова [15] опубликовала замечательные структурированные диагнозы двух видов диатомовых водорослей, оформленные с помощью абзацных отступов, жирного текста и курсива. Подобные диагнозы – большая редкость. Большинство диагнозов из старых [17 – 19] и современных [12, 20, 21] изданий напечатаны сплошным текстом, без выделения ключевых понятий – их тяжело читать и трудно сопоставить.

Закключение. Адаптивные экспертные системы выполняют две функции: диагностической экспертной системы и информационной базы с описаниями таксонов. В этом их основное отличие от компьютерных диагностических систем и таксономических баз данных.

Человек учит адаптивную экспертную систему распознавать новые таксоны, для этого он соотносит таксономические описания, сохраненные системой в процессе определения, с названиями новых для системы таксонов.

Адаптивные экспертные системы удобны в полевых исследованиях. Они помогают участникам экспедиции определить вид организма или, если определение невозможно, запомнить его признаки. Впоследствии можно извлечь сохраненные признаки, определить вид организма и обучить систему опознавать новые таксоны.

Адаптивные экспертные системы позволяют частично преодолевать проблемы межчеловеческого общения, вызванные семантической неоднозначностью естественного языка.

С помощью структурированных диагнозов удобно сравнивать таксоны

и создавать определительные ключи. Структурированные диагнозы являются основой для построения определительных таблиц, они естественным образом вписываются в формат данных, используемых адаптивной ЭС.

Авторами разработан прототип адаптивной экспертной системы «Таксакейс» (*taxakeys.org*). На ее основе созданы интерактивные определители представителей кладоцер (Cladocera), копепоид (Copepoda), высших таксонов кормового зоопланктона, рыб семейства собачковых (Blennidae) и сельдевых (Clupeidae) Севастопольского региона.

Благодарности. Исследование выполнено при финансовой поддержке РФФИ и Правительства города Севастополя в рамках научного проекта № 14-44-01566 «р_юг_а». Авторы выражают благодарность В.В.Мельникову, С.А.Царину, А.В.Темных и М.И.Силакову за предоставленные материалы.

СПИСОК ЛИТЕРАТУРЫ

1. *Lelekov S., Lyakh A.* TAXEX: TAXonomic EXpert system. History of development and technology of identification // Proc. of 'Ocean Biodiversity Informatics': an int. conf. on marine biodiversity data management (Hamburg, Germany, 29 November – 1 December, 2004).– Paris: UNESCO, VLIZ, BSH, 2007.– P.113-120. (IOC Workshop Report, 202) (VLIZ Special Publication, 37).
2. *Курейчук А.Г., Лобанов А.Л., Смирнов И.С., Иночкин А.А., Степаньянц С.Д.* Интернет-определители биологических объектов. 5 лет спустя // Междунар. суперкомпьютерная конф. «Научный сервис в сети Интернет: экзафлопсное будущее», 19-24.09 2011 г., Новороссийск.– М.: Изд-во МГУ, 2011.– С.449-453.
3. *Dallwitz M.J.* Programs for interactive identification and information retrieval [Электронный ресурс] // DELTA – DEscription Language for TAXonomy. URL: <http://delta-intkey.com/www/idprogs.htm> (дата обращения 31.12.2014).
4. *Petrini O., Rusca C.* Knowledge-based expert systems in mycology: Tolerance is a necessary attribute // *Mycologist*.– 1989.– 3, № 3.– P.128-130.
5. *Лелеков С. Г.* Метод учета влияния модификаторов на нечеткие значения лингвистических переменных // Искусственный интеллект.– 2013.– № 4 (62).– С.454-462.
6. *Боркин Л.Я.* Карл Линней (1707-1778) как зоолог // Тр. ЗИН. Приложение № 1. Вид и видообразование. Анализ новых взглядов и тенденций / Под ред. акад. А.Ф.Алимова и С.Д.Степаньянц.– С.-Пб., 2009.– С.9-78.
7. *Генкал С.И.* Атлас диатомовых водорослей планктона реки Волги / Отв. ред. И.В. Макарова.– С.-Пб.: Гидрометеиздат, 1992.– 128 с.
8. *Генкал С.И., Вехов В.Н.* Диатомовые водоросли водоемов Русской Арктики: архипелаг Новая Земля и остров Вайгач / Отв. ред. В.Г.Девяткин.– М.: Наука, 2007.– 64 с.
9. *Генкал С.И., Трофимова И.С.* Диатомовые водоросли планктона Ладожского озера и водоемов его бассейна / Отв. ред. В.Г.Девяткин.– Рыбинск: Изд-во ОАО «Рыбинский дом печати», 2009.– 72 с.
10. *Шнеер В.С.* ДНК-штрихкодирование видов животных и растений – способ их молекулярной идентификации и изучения биоразнообразия // Журн. общей биологии.– 2009.– 70, № 4.– С.296-35.
11. *Butcher B.A., Smith M.A., Sharkey M.J., Quicke D.L.J.* A turbo-taxonomic study of Thai *Aleiodes* (*Aleiodes*) and *Aleiodes* (*Arcaleiodes*) (Hymenoptera: Braconidae: Rogadinae) based largely on COI barcoded specimens, with rapid description of 179 new species // *Zootaxa*.– 2012.– 3457.– P.1-22.
12. *Riedel A., Sagata K., Surbakti S., Tänzler R., Balke M.* One hundred and one new spe-

- cies of *Trigonopterus* weevils from New Guinea // Zookeys.– 2013.– 280.– P.1-150.
13. Riedel A., Sagata K., Surbakti S., Tänzler R., Balke M. Integrative taxonomy on the fast track – towards more sustainability in biodiversity research // Frontiers in Zoology.– 2013.– 10:15.
 14. Предложения для стандартизации терминологии и диагнозов по диатомовым водорослям // Бот. журн.– 1977.– 62, № 2.– С.192-213.
 15. Бухтіярова Л.М. Морфологічні особливості нових для України Bacillariophyta з гідротопів правобережного лісостепу. II. Види *Gomphonema* Ehrnb. // Modern Phytomorphology.– 2013.– 3.– С.231-240.
 16. Drinkwater R.E. Insights into the development of online plant identification keys based on literature review: an exemplar electronic key to Australian *Drosera* // Bioscience Horizons.– 2009.– 2, № 1.– P.90-96.
 17. Дьяконов А.М. Морские звезды морей СССР.– М.-Л.: Изд-во АН СССР, 1950.– 204 с. (Определители по фауне СССР, издаваемые Зоологическим институтом АН СССР; № 34).
 18. Линдберг Г.У. Личинкоядные рыбы Средней Азии.– М.-Л.: Изд-во АН СССР, 1947.– 56 с. (Определители по фауне СССР, издаваемые Зоологическим институтом АН СССР; № 25).
 19. Световидов А.Н. Рыбы Черного моря.– М.-Л.: Наука, 1964.– 552 с. (Определители по фауне СССР, издаваемые Зоологическим институтом АН СССР; № 86).
 20. Васильева Е.Д. Рыбы Черного моря. Определитель морских, солоноватоводных, эвригалинных и проходных видов с цветными иллюстрациями, собранными В.Г.Богородским.– М.: Изд-во ВНИРО, 2007.– 238 с.
 21. Round F.E., Crawford R.M., Mann D.G. The diatoms: biology and morphology of the genera.– Cambridge University Press, 1990.– 747 p.

Матеріал поступив в редакцію 05.12.2014 г.

АНОТАЦІЯ У процесі таксономічної ідентифікації дослідник створює діагноз організму, керуючись підказками експертної системи (ЕС). Цей діагноз може описувати новий для системи вигляд, але ЕС зберігає таку інформацію. Запропоновано концепцію адаптивної ЕС, яка запам'ятовує значення ознак організму, надає користувачеві можливість пов'язувати значення ознак з назвами таксонів і автоматично навчається розпізнаванню нових таксонів. Адаптивна ЕС має засобами нівелювання семантичної неоднозначності таксономічних діагнозів, для чого застосовується теорія нечітких множин. Для наповнення адаптивної ЕС визначальних інформацією запропонована концепція структурованих діагнозів, складених за жорстким ієрархічним планом, в який включені таксономічески значущі і незначимі ознаки. Інформаційно насичена структура ознак дозволяє наповнювати адаптивну ЕС діагнозами нових таксонів.

ABSTRACT During taxonomic identification researcher guided by expert system (ES) and constructs organism diagnosis. This diagnosis can describe a new taxon, but ES doesn't save this information. The conception of adaptive expert system that saves values of organism characters, allows the user to associate characters values with taxonomic names and automatically learns to recognize new taxa, has been proposed. Adaptive ES has the means to leveling semantic ambiguity of taxonomic diagnosis using methods of the theory of fuzzy sets. To fill adaptive ES by identification information a concept of structured diagnosis has been proposed. Structured diagnosis composed by rigid hierarchical plan that includes taxonomically significant and insignificant characters. Information rich structure of characters allow the user to fill adaptive ES with diagnosis of new taxa.