

1. УДК 574.9:57.045
**ПОИСК ЗАВИСИМОСТЕЙ МЕЖДУ ПОКАЗАТЕЛЯМИ В БАЗЕ ДАННЫХ
РАСПРОСТРАНЕНИЯ МЫШЕОБРАЗНЫХ ПЛАНЕТЫ**

Звягинцева А.В., Перелыгина Н.Г.

Донецкий национальный технический университет

кафедра компьютерных систем мониторинга

E-mail: Fly7007@yandex.ru

Аннотация

Звягинцева А.В., Перелыгина Н.Г. Поиск зависимостей между показателями в базе данных распространения мышеобразных планеты. На основе данных многолетних наблюдений за видовым разнообразием мышеобразных и обработки картографической информации ареалов их распространения с использованием ГИС-технологий, создана база данных, позволившая систематизировать соответствующую информацию.

Постановка задачи

Изучение и анализ видового разнообразия является необходимым этапом внедрения биомониторинга возможных изменений природной среды. Подотряд мышеобразных отряда грызунов включает более четверти всех видов млекопитающих. К этой группе относятся крысы, мыши, полевки, лемминги, хомячки, песчанки и тушканчики. Они распространены по всему миру, за исключением, Антарктики, и встречаются почти во всех наземных биотопах. Мышиные самое обширное семейство современных грызунов и вообще млекопитающих. Оно насчитывает около 120 родов и примерно 400-500 видов. Мышеобразные являются частью уникального генофонда земной поверхности, играют важную роль в трофических цепях, в значительной мере определяют формирование и развитие природных экосистем. Они рассматриваются как ценная биологическая модель, представляют практический интерес в связи с эпидемиологическим и лесохозяйственным значением. Все исследования человечества в области токсикологии, оценки различных опасностей и создания лекарств построены на обработке опытных данных, полученных в своем большинстве при проведении экспериментов на мышах и крысах. Тем не менее, распространение мелких грызунов в природе в полной мере не исследовано, структура их сообществ недостаточно изучена [1]. Актуальность исследования видового разнообразия мышеобразных заключается в необходимости изучения ареалов обитания грызунов, их экологических и морфологических особенностей, закономерностей распространения видов в зависимости от разных показателей. Приоритетным считается изучение структурных особенностей сообществ в конкретных природных условиях с учетом их видового состава, численности, продолжительности жизни и других параметров. Эти сведения имеют несомненную прикладную ценность, так как изменения в структуре сообществ часто служат индикаторами природных нарушений.

На основе собранных данных формируются банки данных. При накоплении достаточного объема данных они используются для проведения интеллектуального анализа информации. Очень часто получаемые из различных источников данные хранятся в неупорядоченном, разрозненном виде, в связи с чем исследователь практически не имеет систематизированной и полной информации об интересующем его объекте. Поэтому на основе имеющихся данных создаются структурированные базы данных (далее, БД), хранящие информацию в соответствии с некоторым упорядочиванием, сортировкой. Данные в БД хранятся, как правило, в третьей нормальной форме: непротиворечивость данных и отсутствие избыточности в них. Это позволяет делать определенные выводы и строить системы оценки получаемых данных, что является немаловажной и актуальной задачей.

Исследования

В настоящее время в области биоразнообразия планеты существует большое количество БД, полученных исследователями разных стран [2-6]. Однако следует отметить, что имеющаяся информация чаще всего является разрозненной и не структурируемой, включает различные показатели, по которым для отдельных видов одного и того же семейства и/или рода данные либо совсем отсутствуют, либо известен только ориентировочный диапазон их значений. Кроме того, имеющаяся в справочниках или выставленная на сайтах информация характеризуется разной и не всегда удобной формой ее представления, например, в виде фотографий, рисунков и/или карт с нанесенными на них соответствующими параметрами, зонами и/или характеристиками, либо в виде цифровых значений или их диапазонов, причем такая информация дается не по всем видам. Бывает информация представлена отдельными цифрами по текстовому файлу, ссылками на другие разноязычные источники и БД. Чаще всего название вида (семейства, рода) в таких БД представлено параллельно на русском, английском языках и на латыни. Латинские названия, общие для всех стран и языков, являются уникальными для каждого вида мышеобразных и необходимы для их идентификации. Русскоязычные названия видов фигурируют только в качестве перевода. Исходя из вышесказанного очевидно, что интегрирование этой информации в одну БД на сегодняшний день актуально.

В связи с этим *целью* работы является сбор и систематизация накопленной информации, а также создание на ее основе общей БД, позволяющей применить методы интеллектуального анализа данных для поиска закономерностей биоразнообразия.

На основе имеющихся БД построена единая БД, охватывающая множество показателей и графические изображения. При этом в качестве атрибутивных (численных) данных в базу внесена информация, характеризующая среднюю продолжительность жизни, количество выводков, интервал между выводками, вес, размеры особи, продолжительность беременности, период лактации, интенсивность метаболизма и др. по каждому из 300 видов подсемейства мыши. Общее количество внесенных численных значений в БД – более 7 тыс. Следует отметить, что в процессе исследования возникали трудности, связанные, прежде всего, с недостаточной изученностью отдельных видов мышеобразных и соответственно недостаточным количеством данных по таким показателям, как количество особей в группе, продолжительность жизни в природе, послеродовые темпы роста и т.п.

В виду отсутствия в цифровом виде в существующих БД площади ареалов распространения большинства мышеобразных и наличия общей БД карт с нанесенными ареалами обитания видов [7], в процессе сбора пространственных данных осуществлялся поиск наиболее популярными поисковыми системами Google и Yandex картографической информации с изображениями ареалов распространения исследуемых видов. Найденные растровые изображения были представлены в виде материков или отдельных их участков с выделенными каким-либо цветом на них ареалами распространения тех или иных видов. Собранный картографическая информация оцифровывалась и заносилась в БД в виде численного значения площади для каждого из ареалов распространения всех видов мышей в отдельности и общей площади распространения каждого из видов. Также в БД помещена векторная карта мира с множеством слоев со всеми нанесенными на нее ареалами распространения всех анализируемых видов млекопитающих.

В БД для каждого анализируемого вида также занесены фотографии с изображениями их внешнего вида и растровые изображения ареалов их распространения, где хранятся в виде файлов в формате *.png. с достаточно хорошим сжатием изображения с минимальной потерей качества. В зависимости от разрешения растрового изображения выделены следующие условные категории качества представления изображений: малого (257x311 – 469x587 точек), среднего (472x727 – 1060x1083) и большого (2000x2090 точек) разрешений.

Для реализации поставленных задач в работе использовались картографический, статистический, аналитический методы, а также методы обработки и хранения цифровых данных. Статистический метод применен для определения моды, медиана, минимального, максимального и среднего значения продолжительности жизни, послеродовых темпов роста, веса, размера особи и т.п., а также для определения видов распределения различных параметров и нахождения взаимосвязи между ними. На основе аналитического метода проведен тщательный анализ данных по мышьеобразным с целью их всестороннего изучения и выявления существующих закономерностей. В качестве основного метода исследования использовался картографический, отличающийся надежностью и эффективностью. С его помощью осуществлена пошаговая оцифровка и обработка растровых изображений, а именно определение границ ареалов распространения видов и площадей соответствующих полигонов.

В настоящее время существует ряд методик “снятия” цифровой информации с графических объектов, среди которых выделяют методики получения численных значений с использованием графических редакторов и методики анализа и представления количественных данных на основе использования специальных программ для оцифровки графических изображений.

Для создания БД, статистической и графической обработки информации в работе использована система *MapInfo*, совмещающая преимущества обработки информации, которыми обладают БД (в том числе язык запросов *SQL*), и наглядность карт, списков и графиков. Это позволило спроектировать БД мышьеобразных и наполнить её текстовыми, графическими и численными данными с возможностью определения параметров графических объектов (установление границ и определение площади ареала обитания).

Выбранная методика обработки графических данных представляет собой пошаговый процесс оцифровки растровых изображений и их обработку: определение границ областей и площадей полигонов. В качестве растрового изображения использовались карты (например, рис. 1) с нанесенными на них полигонами (ареалами распространения определенного вида млекопитающего), которые визуально выделялись на фоне карты. Далее осуществлялась привязка данных изображения к векторной карте мира [8], после чего растровое изображение становилось оцифрованным (векторным). На векторной карте мира выбирались точки, широта и долгота которых были крайними точками на растровой карте, то есть выполнялась привязка крайних точек карты к координатам на векторной карте. Для чего добавлялись четыре точки, так как изображения, как правило, представляются в виде геометрических фигур типа квадрат, прямоугольник.



Рисунок 1. Пример растрового изображения

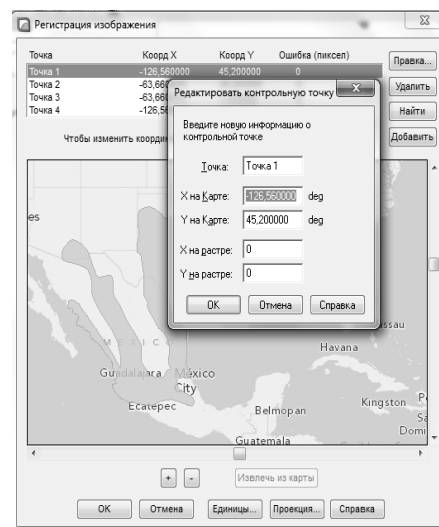


Рисунок 2. Окно редактирования контрольной

точки

Далее обозначались границы ареала распространения млекопитающего, что позволило определить площадь данного ареала, в квадратных километрах, населенную тем или иным видом (рис. 3).

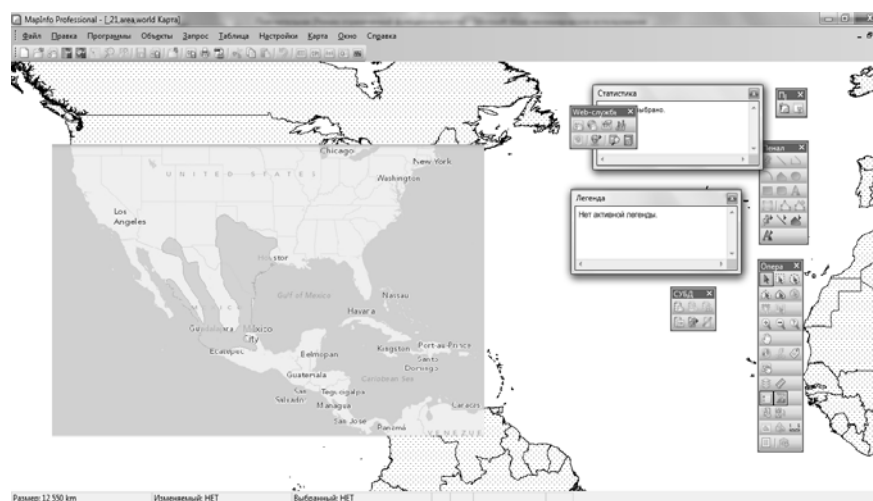


Рисунок 3 – Результат векторизации изображения

По аналогии проведена обработка порядка 300 карт, характеризующих распространение мышеобразных на Земном шаре. Прделанная работа стала основой создания в среде *MapInfo* БД по мышеобразным, позволяющей не только хранить, но и наполнять её цифровыми, текстовыми и графическими данными.

Наполнение БД происходило в следующей последовательности. Вначале выбиралось в виде полигона оцифрованное изображение ареала распространения вида, к которому впоследствии привязывалась атрибутивная информация (рис. 4), вносимая вручную. При выборе точки на пересечении ареалов распространения мышеобразных, обитающих на территории с определенной широтой и долготой, отображался список площадей ареалов их распространения (рис. 5).

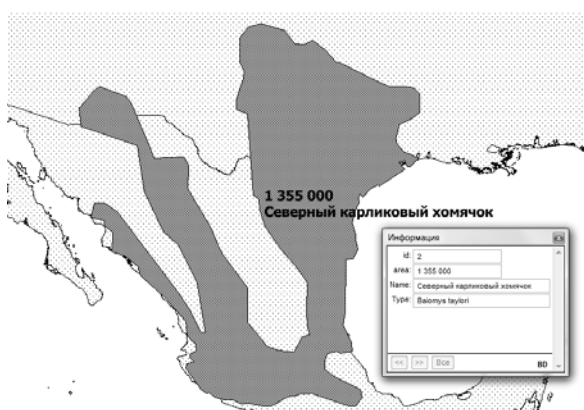


Рисунок 4 - Привязка цифровой и текстовой информации

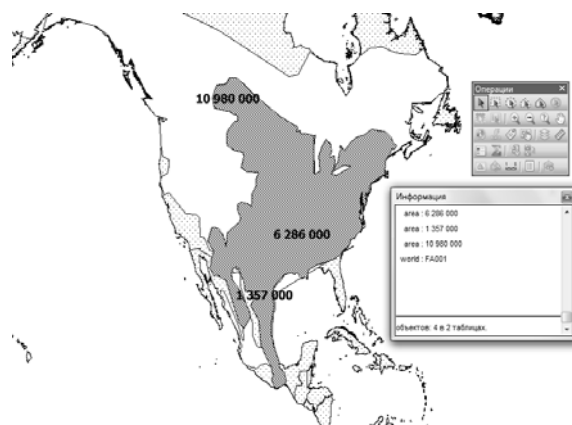


Рисунок 5 - Ряд видов мышеобразных, обитающих в выбранной зоне

Полученное векторное изображение, находящееся в созданной БД, представляет собой один из слоев изображений проекта. Причем пользователь имеет возможность выбирать слои по своему усмотрению, которые он хотел бы отобразить или не отобразить на экране. Для выбора отображаемых слоев необходимо на панели операций выбрать инструмент управления слоями и указать на тот слой, который необходимо отобразить.

Выводы и дальнейшие исследования

Таким образом, на первом этапе исследования были разработаны ГИС-модели распространения видов мышеобразных и создана БД картографической и атрибутивной информации. Анализ распространения видов мышеобразных позволит систематизировать информацию об особенностях видового биоразнообразия с использованием ГИС-моделей. Планируется, что дальнейшие исследования будут направлены на поиск закономерностей распространения видов мышеобразных в зависимости от климатических условий. Это возможно при совмещении созданной БД по распространению мышеобразных с БД Всемирного климатического центра [9-11], где хранятся архивы климатических данных повторного анализа. Архив данных Всемирного климатического центра позволяет построить осредненные во времени поля климатических показателей для всей планеты. Подобное осреднение можно провести за 15 лет на широтно-долготной сетке земного шара с ячейкой в 1x1 градус. В архиве хранится около 80 климатических показателей, в частности, температура, относительная влажность, скорость и направление ветра, освещенность, осадки, высота над уровнем моря и т.д. Объем данных архива составляет несколько терабайт.

Из анализа следует, что на сегодня на поверхности Земли можно выделить более 400 ареалов по 300 видам семейства мышиных. В результате одной ГИС-модели ареала распространения видов мышеобразных может быть поставлена в соответствие атрибутивная информация из 80 осредненных климатических показателей. С помощью применения регрессионного анализа возможно определение наиболее значимых климатических показателей, которые влияют на ареалы распространения вида. Кроме этого из БД может быть взята атрибутивная информация, количественно характеризующая каждый вид. То есть объединение созданной БД с базой данных осредненных показателей Всемирного климатического центра позволит подойти к созданию моделей распространения видов на территории планеты. Предполагается, что в данном направлении и будут проводиться дальнейшие исследования.

Список литературы

1. Полевые исследования экологии наземных позвоночных животных / Интернет ресурс. – Режим доступа: [www/ URL: http://zoomet.ru/novikov_4_2-a-b.html](http://zoomet.ru/novikov_4_2-a-b.html).
2. База данных: возраст и продолжительность жизни животных / Интернет ресурс. – Режим доступа: [www/ URL: http://genomics.senescence.info/species/](http://genomics.senescence.info/species/).
3. Журналы Кэмбриджа / Интернет ресурс. – Режим доступа: [www/ URL: http://journals.cambridge.org](http://journals.cambridge.org).
4. Систематика современных млекопитающих / Интернет ресурс. – Режим доступа: [www/ URL: http://zmmu.msu.ru/personal/pavlinov/mam_world/mam_world.htm](http://zmmu.msu.ru/personal/pavlinov/mam_world/mam_world.htm).
5. FLORANIMAL – растения и животные / Интернет ресурс. – Режим доступа: [www/ URL: http://www.floranimal.ru/](http://www.floranimal.ru/).
6. Классификация животных / Интернет ресурс. – Режим доступа: [www/ URL: http://www.apus.ru/](http://www.apus.ru/).
7. Красный список исчезающих видов / Интернет ресурс. – Режим доступа: [www/ URL: http://maps.iucnredlist.org/](http://maps.iucnredlist.org/).
8. Большая подборка векторных карт мира энциклопедия / Интернет ресурс. – Режим доступа: [www/ URL: http://ivan-off.com/?p=190](http://ivan-off.com/?p=190).
9. Глобальные климатические архивы повторного анализа / Интернет ресурс. – Режим доступа: [www/ URL: http://www.reanalysis.org/atmosphere/comparison-table](http://www.reanalysis.org/atmosphere/comparison-table).
10. Всемирные спутниковые данные MODIS по атмосфере 2000-2012 / Интернет ресурс. – Режим доступа: [www/ URL: http://modis-atmos.gsfc.nasa.gov/](http://modis-atmos.gsfc.nasa.gov/).
11. Более 80 климатических показателей. – Режим доступа: http://data-portal.ecmwf.int/data/d/interim_full_moda/; <http://nomads.ncdc.noaa.gov/data.php?name=access#CFSR-data>.