

УДК 004.045:574.9

А.Д. Василенко, Я.И. Шаховская, А.В. Звягинцева
Донецкий национальный технический университет, г. Донецк
кафедра компьютерных систем мониторинга
E-mail: rattram@gmail.com, Vasilenko.lyoha@yandex.ru

АНАЛИЗ АЛГОРИТМОВ ОБРАБОТКИ КАРТОГРАФИЧЕСКОЙ ИНФОРМАЦИИ ПРИ МОНИТОРИНГЕ РАСПРОСТРАНЕНИЯ ВИДОВ

Аннотация

Василенко А.Д., Шаховская Я.И., Звягинцева А.В. Анализ алгоритмов обработки картографической информации при мониторинге распространения видов. Выполнен аналитический обзор интернет и литературных источников баз данных с характеристиками видов живых организмов. Произведен анализ и сравнение алгоритмов обработки растровых изображений карт. Выбран оптимальный алгоритм для нахождения площади ареалов распространения видов на земном шаре на основе картографической информации.

***Ключевые слова:** характеристики видов, алгоритмы обработки растровых изображений, площадь ареалов распространения видов.*

Актуальность

На сегодняшний день живые организмы распространены по всему земному шару, включая как материковую, так и водную часть, и их распространение зависит от определенных факторов – как параметров окружающей среды, так и самих характеристик живого организма и влияния антропогенного фактора. Поэтому исследование ареалов распространения видов является одной из ключевых задач, представляющей большой интерес с научной точки зрения. Информация о численном значении площади позволяет определить зависимости между различными параметрами и спрогнозировать развитие определенных видов под влиянием различных условий. Имея оптимальный автоматизированный алгоритм обработки растровых изображений карт, становится возможным обрабатывать большое количество карт с целью увеличения базы данных ареалов обитания видов.

Цель работы заключается в создании наиболее полной и достоверной базы данных характеристик биоразнообразия, а также в обосновании и выборе алгоритма нахождения площади ареала распространения видов на земном шаре.

Анализ и сбор данных. Существует множество различных сайтов, содержащих информацию о названии видов на русском, английском и латинском языках, о характеристиках видов, об ареалах их распространения и

т.д. Чаще всего информация представлена на любительских, научных и познавательных сайтах и оформлена либо в виде текста, либо таблиц Excel, либо средствами MySQL или просто в виде хаотичного представления картинок с изображением внешнего вида животного, а также карт или схем с нанесенными ареалами их распространения на земном шаре. Сопоставимость данных по биоразнообразию с различных сайтов может представлять интерес, однако следует отметить разрозненность этих данных, разные форматы их представления и, чаще всего, разные методики поиска и обработки соответствующей информации. Все это говорит о возникновении больших трудностей при объединении информации с различных сайтов, а иногда о невозможности ее осуществления в принципе. Кроме того, для большинства видов чаще всего отсутствует числовая информация о площадях ареалов их распространения на земном шаре, что создает новую задачу нахождения площади ареалов, используя имеющуюся на сайтах картографическую информацию. Следует также отметить, что каждый сайт зачастую предоставляет свой тип карты. Для того, чтобы найти карту с определенным видом животного, обычно название вида на языке самого сайта или латинском (как общепризнанном во всем мире) языке записывается в строку поиска, если на сайте предусмотрена такая функция. Однако, когда возникает необходимость найти большое количество карт, метод набора вручную оказывается громоздким и чаще всего занимает много времени.

В основу данной работы была положена БД с сайта [1]. Она содержит информацию по более чем 4000 видам по двадцати характеристикам (возраст зрелости, вес особи, продолжительность жизни, период лактации и т.д.), а также их русские, английское и латинское названия. Общий объем базы данных составляет восемьдесят тысяч значений. Также к данной БД в процессе работы нами добавлена цифровая база данных изображений каждого вида и графическое представление карты его ареала распространения. В последствии для каждого вида также будет рассчитана площадь ареалов его распространения. Таким образом, объем база данных будет составлять 84000 значений и около 10000 изображений.

Методы картирования ареалов. Представление об ареале распространения какого-либо вида животного предоставляется возможным после нахождения картографического изображения этого самого ареала. Существуют различные методы картирования ареалов, такие как точечный, контурный и сеточный методы.

Одним из наиболее распространенных является точечный метод: местонахождение каждого обнаруженного вида обозначается точкой на карте. С помощью данного метода можно точно определить пункт сбора, поимки особей того или иного вида либо его регистрации. Он является наиболее эффективным, если нужна карта распространения редкого или просто малоизученного вида. Точечный метод хорош также в начале исследования

географического распространения животного. Однако этот метод имеет ряд недостатков, таких как, к примеру, малопоказательность, так как требует известной доли воображения, чтобы представить себе форму и площадь ареала. Кроме того, животные встречаются не в определенной точке, а распространены на ареалах определенной площади.

Все недостатки точечного метода компенсирует контурный – он обозначает границы ареала на контурной карте. Крайние точки распространения животных соединяются сплошной линией, получая замкнутое изображение. Таким образом, получается какая-то обозначенная площадь, где распространен определенный вид животного.

Также во многих странах используется сеточный метод, который также называют методом формальных квадратов или растровым методом. Поле карты разбивают на квадраты со стороной 50 км (масштаб при необходимости может меняться). Если в пределах квадрата обнаружены особи изучаемого вида, он закрашивается или отмечается условным знаком. Прочие квадраты не обозначаются.

Создание скриптового кода для поиска и обработки картографической информации. Для поиска карт более 4000 видов животных был написан скрипт на языке PHP с использованием технологии Google API. Элементы разработанного скрипта показаны на рисунке 1.

```

for ($i=0; $i<2; $i++)
{
for ($j=0; $j<25; $j++)
$u = 0;
$word="$search[$i] areas of distribution site:$saits[$j]";
$n=$i+1;
$keyforparse = $word;
$url =
"http://ajax.googleapis.com/ajax/services/search/images?v=1.0&q=".urlencode($key
forparse)."";
$ch = curl_init();
curl_setopt($ch, CURLOPT_URL, $url);
curl_setopt($ch, CURLOPT_RETURNTRANSFER, 1);
$json_data = curl_exec($ch);
curl_close($ch);
$response = json_decode($json_data, true);
$d = urldecode($dir[$i]);
$i1 = $i + 1;
for($counter=0; $counter < 4; $counter++)
{
echo("".$response['responseData']['results'][$counter]['url']."<br/>");
$a[$counter] = $response['responseData']['results'][$counter]['url'];
$randname = "jpg1/{ $i1 }/map{ $counter }.png";

```

```
copy(${counter},${randname});  
$u = $u+1;  
$fp = fopen("jpg1/{$i1}/source.txt", "a");  
$mytext = "${counter}\n";  
$test = fwrite($fp, $mytext);  
fclose($fp);  
}
```

Рисунок 1 – Элементы скрипта для поиска карт

В данном скрипте считываются названия видов животных из файла search.txt, содержание которого представлено на рисунке 2.

```
Drosophila melanogaster  
Apis mellifera  
Lasius niger  
Bicyclus anynana  
Homarus americanus  
Acipenser brevirostrum
```

Рисунок 2 – Содержание файла search.txt

На следующем шаге название из файла записывается в переменную \$search[\$i], затем подставляется в строку \$word="\$search[\$i] areas of distribution site:\$saits[\$j]".

Если задать поиск по изображениям в поисковой системе Google с запросом «<название_вида_животного> areas of distribution», то зачастую находится не карта, а произвольное изображение. Программными средствами очень сложно оценить, что именно будет найдено, поэтому необходимо к строке добавить «site:\$saits[\$j]», где \$saits[\$j] – переменная, содержащая один из адресов сайтов, которые находятся в файле s.txt. Таким образом, для каждого вида животного ищется карта на каждом из 26 анализируемых сайтов.

После того, как сработал поиск, скрипт обрабатывает найденное изображение и «вытаскивает» его адрес, записывает этот адрес в файл source.txt, находящийся в папке с названием вида животного, а также сохраняет само изображение, находящееся по этому адресу, в ту же папку.

С использованием данного скрипта было найдено примерно 5000 карт по более чем 4000 видам. Карты с ареалами распространения разных видов, имеют разный масштаб, разный способ нанесения информации, различный формат и пр. (рис. 3), поэтому для всех карт нельзя создать универсальный алгоритм, имеет смысл разбить карты на группы, и для каждой группы разрабатывать отдельный алгоритм расчета числового значения площади ареала распространения вида.

В ходе выполнения работы найдено два вида изображения карт: векторные и растровые. Так как большинство изображений карт растровые и уже известно много алгоритмов обработки для векторных изображений, были рассмотрены алгоритмы обработки растровых изображений карт.

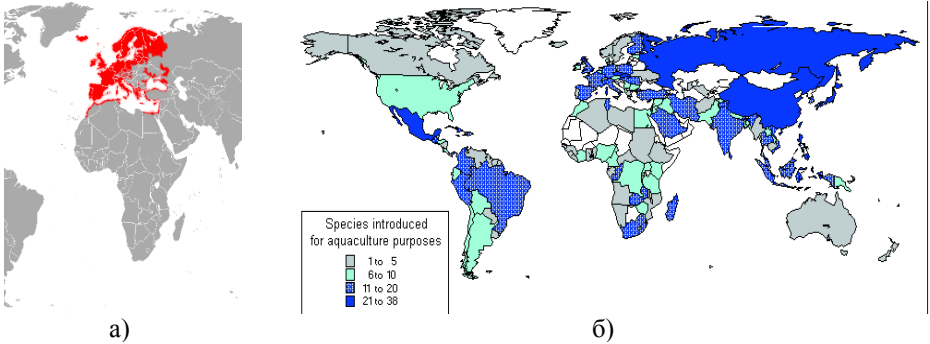


Рисунок 3 – Пример растровых изображений карт с нанесением ареалов обитания видов:

а) залитыми областями; б) залитыми и заштрихованными областями

Прежде чем приступить к работе с картами, приходилось их анализировать. Анализ осуществлялся с целью разбиения найденных карт на группы, после чего определялся вид предварительной обработки карт и используемых алгоритмов для нахождения площади ареалов распространения видов. В процессе анализа карт все они были сгруппированы в три группы: карты с залитыми ареалами, с заштрихованными ареалами, с нечетко выделенными границами и размытым цветом заливки ареала.

Анализ алгоритмов предварительной обработки растровых карт. Так как найденные карты в большей своей части сильно отличались друг от друга по форме, масштабу, цвету, а также типу заливки, была проведена их предварительная обработка.

Для предварительной обработки найденных растровых изображений карт были использованы программы Adobe Photoshop, Easy Trace Pro и Illustrator.

Алгоритм обработки изображения с помощью программы Adobe Photoshop является достаточно трудоемкий и его нельзя автоматизировать. Данный алгоритм основан на разбиении цветов на классы и распознании их на картинке. Цвета преобразовываются в цифровые и наша карта становится векторной и теперь ее можно проанализировать, используя алгоритмы, рассмотренные ниже. В программе можно улучшить видимость границ и штриховки ареалов, поменять яркость цветов. С помощью некоторых манипуляций инструментами Adobe Photoshop карты можно привести к одному типу, а именно залить все ареалы одним цветом или создать штриховку одного типа.

Алгоритм обработки изображения с помощью программы Easy Trace Pro позволяет в автоматическом режиме векторизировать растровое изображение. С недавнего времени в программу были добавлены новые атрибуты, позволяющие работать с рельефом. Так же данный пакет может открывать и

обрабатывать в определенной последовательности большое количество карт в автоматическом режиме [2].

Алгоритм обработки изображения с помощью программы Adobe Illustrator позволяет работать с векторными изображениями. Данная программа в автоматическом режиме векторизует растровое изображение. Но так как программа не предназначена для работы с картами, имеет место значительная ошибка в распознании программой цветов выделенных ареалов. Для обработки одной карты требуется 1 – 2 минуты [3].

Анализ алгоритмов для нахождения площади ареалов распространения видов на земном шаре. В ходе выполнения работы с целью нахождения площади ареала распространения определенного вида проанализировано шесть алгоритмов обработки изображения растровых карт:

- 1) алгоритм обработки изображения с помощью программы MapInfo (основан на привязке растровой карты к векторной);
- 2) алгоритм обработки изображения с помощью программы ImageMagick (основан на подсчете количества пикселей изображения);
- 3) алгоритм с использованием уже известных площадей стран;
- 4) алгоритм обработки изображения с помощью создания нейросети;
- 5) алгоритм обработки изображения с помощью программы ArcViewGIS (основан на создании нового слоя поверх имеющейся карты);
- 6) алгоритм нахождения площади с помощью Postgis/Geography (основан на заполнении базы данных);

Рассмотрим эти алгоритмы подробнее.

Алгоритм обработки изображения с помощью программы MapInfo позволяет осуществить привязку растрового изображения карты к векторной за счет стандартных инструментов. При этом создается слой с выделенным ареалом обитания животного, площадь которого программа вычисляет автоматически при нажатии пользователем по вкладке «Геоинформация». Каждая карта обрабатывается отдельно, так как карты обрабатываются вручную. Обработка одной карты в целом занимает от 5 до 30 минут [4].

Алгоритм обработки изображения с помощью программы ImageMagick работает через консоль. Основная идея этого алгоритма основана на подсчете пикселей растрового изображения карты. Программа выдает количество пикселей цвета, которым закрасен ареал, после чего это количество пикселей умножается на количество километров, занимаемое одним пикселем и собственно получаем площадь ареала. Каждая карта обрабатывается отдельно, так как программа не может открывать одновременно несколько изображений. Имеется возможность полуавтоматизировать процесс, создав цикл в коде программы, позволяющий открывать карты по очереди. Т.к. масштабы анализируемых карт колеблются в значительных пределах, то количество километров, занимаемых одним пикселем, будет различным. В связи с чем нет необходимости автоматизировать нахождение площади ареала [5].

Алгоритм с использованием уже известных площадей стран основан на использовании уже известных площадей стран мира. Площадь ареалов обитания животных будет находиться на основе суммы площадей стран, где обитает изучаемый вид. В случае, если площадь распространения изучаемого вида будет занимать территорию не всей страны, то расчет площади ареала обитания может производиться с помощью приблизительных вычислений, исходя из части занимаемой площади. Данный алгоритм предусматривает определенную погрешность, которая возрастает с уменьшением разности площади между страной и определяемым ареалом обитания вида. Таким образом, данный алгоритм не предполагает автоматизации нахождения площади.

Алгоритм обработки изображения с помощью создания нейросети является трудоемкий и занимает много времени. Он основан на обучении ПЭВМ распознавать растровые изображения карт. Машина сама распознает залитые определенным цветом участки карты и принимает их за ареалы обитания живого организма и с помощью определенных алгоритмов подсчитывает площадь ареала. Сам процесс обучения нейросети может занять до полугода, в связи с чем применение данного метода рационально лишь в случаях использования большого количества карт одного типа. Погрешность вычисления площади ареала будет зависеть от качества используемых карт, в частности от четкости границ и заливки ареалов обитания [6].

Алгоритм обработки изображения с помощью программы ArcViewGIS подобен алгоритму обработки изображения с помощью программы MapInfo. Программа ArcViewGIS основана также на оцифровке растрового изображения. По сравнению MapInfo, ArcViewGIS не осуществляет привязку растровой карты к векторной. В ArcViewGIS создается слой поверх загруженной растровой карты, на котором выделяется ареал распространения вида и программа автоматически рассчитает площадь выделенного ареала. Каждая карта обрабатывается отдельно, так как соответствующая работа может выполняться только вручную. Время работы с одной картой в целом составляет от 10 до 30 минут [7].

Алгоритм нахождения площади с помощью Postgis/Geography позволяет уменьшить ошибку в расчетах нахождения площади ареала распространения вида, так как учитывается кривизна поверхности Земли. Алгоритм основан на добавление в базу данных Postgis/Geography векторных карт и последующего нахождения площади ареалов распространения вида с помощью написания на РНР соответствующих команд в базу. Время реализации всего алгоритма зависит от количества и размера загружаемых карт. В среднем на загрузку и обработку одной карты требуется 2 – 3 минуты. Погрешность расчетов зависит от достоверности исходных карт и нанесенных границ самих ареалов распространения вида. Процесс загрузки и обработки карт полностью автоматизирован [8].

С целью анализа и эффективности применения указанных выше алгоритмов для определения площади ареалов распространения видов было проанализировано с использованием обработки изображения с помощью программы MapInfo, imageMagick по семь карт из каждой трех групп. На данный момент было определено, что наиболее оптимальный для обработки растровых изображений карт с целью нахождения площади ареалов распространения видов является алгоритм обработки изображения с помощью средств MapInfo. Так как в сравнении с другими алгоритмами данный алгоритм минимизирует погрешность нахождения площади нужного ареала обитания животного.

Выводы. Таким образом, был разработан скрипт, позволяющий осуществить поиск всех существующих карт с нанесенными ареалами распространения видов на земном шаре. Скрипт облегчил поиск карт, учитывая, что ручной поиск для более чем 4000 видов занял бы значительно более длительное время, чем понадобилось скрипту. В каждой папке с названием вида животного хранится как минимум 1 карта с ареалом распространения, если карта данного вида вообще была создана. В ходе выполнения работы были проанализированы следующие алгоритмы: алгоритм обработки изображения с помощью программы MapInfo, с помощью программы ImageMagick, алгоритм с использованием уже известных площадей стран, алгоритм обработки изображения с помощью создания нейросети, алгоритм обработки изображения с помощью программы ArcViewGIS и алгоритм нахождения площади с помощью Postgis/Geography. В процессе анализа найденных карт все они были сгруппированы в три группы: карты с залитыми ареалами, с заштрихованными ареалами, с нечетко выделенными границами и размытым цветом заливки ареала. На данный момент было выяснено, что наиболее оптимальным для обработки растровых изображений карт является алгоритм обработки изображения с помощью программы MapInfo так как, в сравнении с другими проанализированными алгоритмами, данный алгоритм минимизирует погрешность нахождения площади ареала распространения вида.

Список литературы

1. <http://genomics.senescence.info/species/> (01.04.2013).
2. <http://www.easytrace.com/>(04.04.2014).
3. <http://www.adobe.com/ua/products/illustrator.html>(04.04.2014).
4. <http://www.mapinfo.com/>(03.03.2014).
5. <http://gis-lab.info/qa/im-countpixels.html>(15.03.2014).
6. <http://www.statsoft.ru/home/textbook/modules/stneunet.html>(25.03.2014).
7. <https://www.arcgis.com/features/>(28.03.2014).
8. <http://workshops.boundlessgeo.com/postgis-intro/geography.html>(01.04.2014).