

ОБНАРУЖЕНИЕ И УСТРАНЕНИЕ ОШИБОК РАЗНЫХ ТИПОВ

БД ГИС подвержена ошибкам графики, атрибутов и их согласования. Хотя все они заметно различаются, в дальнейшем мы рассмотрим вначале графические ошибки, а затем, в одном разделе, ошибки атрибутов и согласования. Чаще всего ошибки атрибутов обнаруживаются из-за их несогласованности с графикой. Не всегда бывает именно так, и обнаружение чисто атрибутивных ошибок чаще всего выполняется через проверку атрибутивной БД. Хотя это тоже часть общего процесса устранения ошибок, полное описание всех возможных вариантов таких ошибок не требуется. Примеры для каждого типа ГИС заняли бы немалую часть книги, поэтому мы посмотрим, как обнаруживаются ошибки в одной-двух наиболее распространенных системах, а вы сможете изменить эти процедуру соответственно используемой вами ГИС.

Графические ошибки в векторных системах

По окончании оцифровки, векторно-топологические ГИС требуют построения топологии (если это не было частью самого процесса оцифровки). В любом случае, топология, содержащая явную информацию об отношениях графических объектов в БД, должна позволить вам идентифицировать графические ошибки некоторых типов. Одни из них будут обозначены текстовыми сообщениями, другие должны быть выявлены в результате просмотра статистики БД, отображающей количества типов и объектов, или проверки изображения для поиска ошибок, которые данная ГИС не может обнаружить сама. Вам нужно будет искать ошибки шести основных типов, соответствующих отрицаниям следующих утверждений:

1. Присутствуют все графические объекты, которые должны быть введены.
2. Не оцифровано объектов сверх того.
3. Объекты находятся на должных местах и имеют должные форму и размеры.
4. Соединены все объекты, которые должны быть соединены.
5. Все области имеют ровно одну метку для идентификации.
6. Все объекты находятся в пределах рабочей области, определенной опорными точками.

Крупная коммерческая ГИС должна быть способна обеспечивать эти общие топологические отношения, и вы можете использовать их для обнаружения ошибок. Хорошей процедурой для сравнения оцифрованных объектов и исходной карты является отображение на экране или даже вывод твердой копии. Последний позволит вам физически наложить и сравнить две карты на копировальном столе с подсветкой. Помимо этого, многие ГИС

имеют набор символов для индикации некоторых ошибок. Чтобы сэкономить время, познакомьтесь с ними до начала редактирования. Теперь мы пройдем по конкретным типам ошибок, которые мы можем найти в связи с этими шестью общими типами.

Как вы помните из нашего обсуждения векторных моделей данных, узлы - это специальные точки для индикации связи между линиями, составленными из отдельных отрезков. Узлы - это не просто точки между отрезками линии, которые показывают изменение ее направления, они имеют определенное топологическое значение. Узлы могут использоваться для обозначения пересечения двух улиц или слияния реки и озера, но они должны появляться не на каждом отрезке линии или границы полигона.

Возможны также так называемые псевдоузлы, в которых линия соединяется сама с собой или когда в узле соединяются только две линии. Поэтому первым типом ошибок, которые могут быть обнаружены, являются псевдоузлы, которые мы не намеревались создавать, то есть когда мы не трактуем линию как две самостоятельных дуги. Создание псевдоузла при отсутствии пересечения с другой линией чаще вызывается необходимостью смены значений атрибутов где-то в промежутке между двумя обычными узлами. Мелкие изолированные полигоны часто изображаются одной замкнутой на себя в псевдоузле дугой; такие псевдоузлы как правило не являются ошибочными (Рисунок 6.1). Ваша ГИС должна быть способна отмечать псевдоузлы с помощью легко различимого графического символа. При построении своей первой БД ГИС вы можете быть сбиты с толку обилием появившихся псевдоузлов. Перед тем, как паниковать, вы должны знать, что не все псевдоузлы являются ошибками, а их символы всего лишь указывают на возможные проблемы.

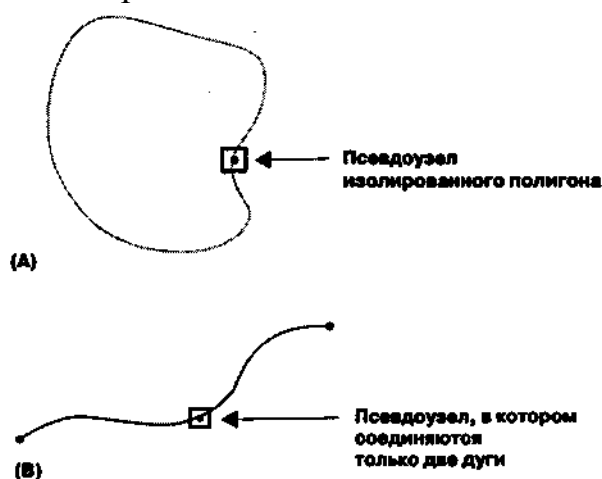


Рисунок 6.1. Псевдоузлы. Два типа псевдоузлов, которые могут быть ошибками.

Псевдоузлы, которые не являются результатом намеренного создания изолированного полигона (в том числе одного полигона внутри другого) обусловлены чаще всего ошибками оператора дигитайзера. Другими

словами, вы либо пытались создать незамкнутую фигуру, но поместили курсор не туда, куда надо, либо вы пытались создать полигон, который соединен с другими полигонами (т.е. имеет связывающую с ними дугу), но нажали не ту кнопку, что требовалось. В качестве средства избежать ошибочных псевдоузлов вы можете пронумеровать ваши точки при подготовке карты или использовать специальный код или символ для обозначения мест, в которых находятся действительно необходимые псевдоузлы. Полезно использовать цифровой код, который совпадает с числами на кнопках курсора дигитайзера, которые соответствуют узлам.

Если программа сообщает, что ваше покрытие содержит один или более псевдоузлов, то для исправления ошибок можно использовать подготовленную карту. Во-первых, вам нужно определить, являются ли псевдоузлы на самом деле ошибками. Законные псевдоузлы (т.е. такие, которые присутствуют для определенной цели) могут быть проигнорированы.

Ошибочные узлы могут быть удалены или перемещены для восстановления корректности. В коммерческих системах это делается обычно легко.

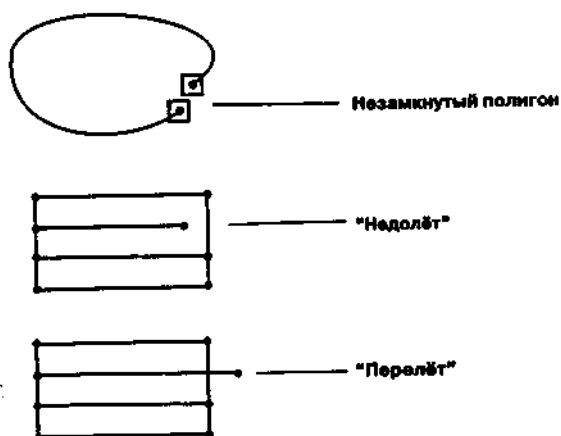


Рисунок 6.2. Ошибки узлов. Ошибочные висящие узлы трех основных типов: при незамкнутости границы полигона; когда дуга не достигает объекта, к которому должна быть присоединена; когда дуга пересекает объект, к которому должна быть присоединена.

Другая обычная ошибка, называемая висящим узлом, может быть определена как узел на ни с чем не соединенном конце линии (рисунок 6.2). Возможны три вида ошибок, создающих висящие узлы: незамыкание границы полигона; "недолёт" (undershoot), т.е. неприсоединение дуги к объекту, к которому она должна быть присоединена; «перелёт" (overshoot), при заходе дуги за объект к которому она должна быть присоединена. В одних случаях причиной ошибки может быть неправильное положение курсора дигитайзера, в других - установка недостаточной величины расстояния неразличимости точек. Правильная установка этой величины является одним способом избежать этой проблемы, подготовка карты -

другим. Обычно легче находить "перелёты", чем "недолёты". Если вам свойственно создавать висящие узлы, то уж лучше первые, чем вторые. Хотя точная оцифровка была бы еще лучше, такое решение оказывается лучше для тех, кто часто совершает эти ошибки.

Висящие узлы обычно обозначаются программой графическим символом, отличным от используемого для псевдоузлов. Кроме того, если висящий узел обусловлен незамкнутым полигоном, то ГИС предупредит вас сообщением о числе замкнутых полигонов в базе данных; если оно отличается от того, что вы насчитали при подготовке карты перед оцифровкой, то вам будет ясно, что нужно поискать такие висящие узлы. Опять же, исправления довольно просты. В случаях "недолётов" узел передвигается, или "пристегивается" к объекту, с которым он должен быть соединен. "Перелёты" исправляются определением должной точки пересечения и "обрезанием" линии так, чтобы она соединялась там, где следует. В случае открытого полигона вы просто передвигаете один из узлов до соединения с другим. Чаще всего ГИС сама устранит после этого лишний узел.

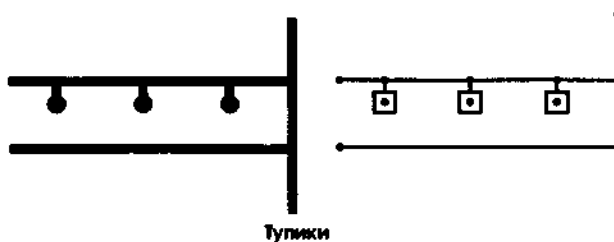


Рисунок 6.3. Допустимые висящие узлы. В данном случае показывают тупики вдоль проезжей улицы в жилой зоне.

Как и в случае с псевдоузлами, некоторые висящие узлы вводятся в БД ГИС намеренно. Чаще всего эти узлы служат указателями особой ситуации на конце линии или дуги. Например, вы могли бы использовать узлы для указания положений тупиковых площадок в жилой зоне (Рисунок 6.3) или истоков рек. То есть, мы видим, что висящие узлы могут быть законными объектами в БД, а не только ошибками.

При оцифровке полигонов вы должны указывать метку - точку внутри каждого из них, которая служит для связи с атрибутами и выбора места отображения текстовой информации об этом полигоне. Нужна одна и только одна такая точка. В связи с этим возможны ошибки двух типов: отсутствующие метки и лишние метки (Рисунок 6.4). И те и другие чаще всего обязаны потере контроля в процессе оцифровки. Хотя хорошая подготовка карты уменьшит число ошибок с метками, чаще всего эта проблема вызывается беспорядочной организацией работ, перерывами в процессе оцифровки или усталостью. К счастью, они очень легко обнаруживаются и исправляются добавлением меток там, где их не хватает, и удалением лишних там, где они есть.

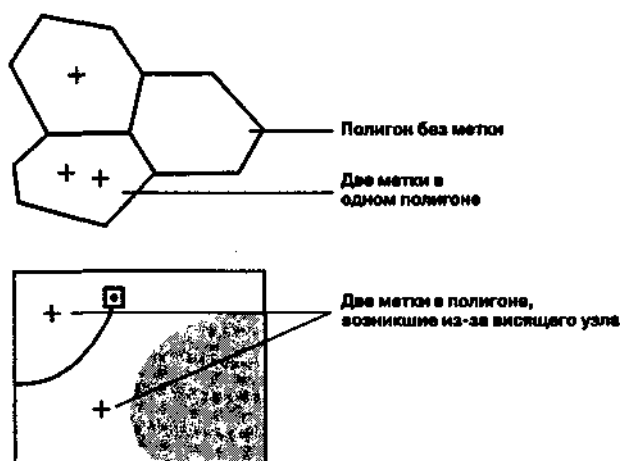


Рисунок 6.4. Ошибки с метками. Внутри каждого полигона должна быть одна точка, к которой присоединяются атрибуты. Ошибки возникают, когда метки нет или когда их больше одной.

Другой тип ошибок чаще всего встречается, когда программа использует векторную модель, в которой каждый полигон имеет свою отдельную границу. В таких случаях вы должны оцифровывать общие линии границ полигонов более одного раза. Невозможность поместить курсор дигитайзера точно в требуемой позиции для каждой точки на этой линии часто приводит к возникновению последовательности крошечных полигонов, называемых осколочными, или рукавными полигонами (*sliver polygons*) (Рисунок 6.5). Они могут возникать также в результате операций наложения (см. Главу 12), или когда каждая из двух смежных исходных карт имеет свою проекцию (см. далее в этой главе). Мы ограничимся рассмотрением осколочных полигонов, возникающих в процессе ввода.



Рисунок 6.5. Осколочные полигоны. Они возникают из-за плохой оцифровки вдоль общих границ, где линия должна вводиться более одного раза.

Сильно нерегулярные границы особенно подвержены такой "кудрявости" (*fuzziness*) оцифровки.

Конечно, легче всего избежать получения осколочных полигонов при вводе, если использовать ГИС, которая не требует двойной оцифровки линий. Но иногда вы и сами можете случайно ввести одну линию дважды,

получив все тот же осколочный полигон. При этом вы можете получить также и висящий узел, поскольку была создана ненужная линия. Простое удаление линии в таком случае решит обе проблемы.

Поиск осколочных полигонов в отсутствие висящего узла более труден. Один из способов - сравнить число введенных в компьютер полигонов с Числом полигонов на исходной карте. Но даже если вы знаете, что осколочные полигоны где-то есть, часто их очень трудно найти. Обычно приходится просматривать изображение в поисках подозрительных границ полигонов, а затем увеличивать его, чтобы увидеть осколочные полигоны. В некоторых случаях вам может понадобиться несколько шагов увеличения. Увы, часто вы, пока не увидите их, не знаете, что перед вами, простая линия или осколочные полигоны.

Иногда, когда имеется последовательность совсем крошечных полигонов, но нет висящего узла, можно увеличить расстояние неразличимости точек при вводе. Если этот параметр можно изменять в подсистеме редактирования, программа автоматически удалит осколочные полигоны.

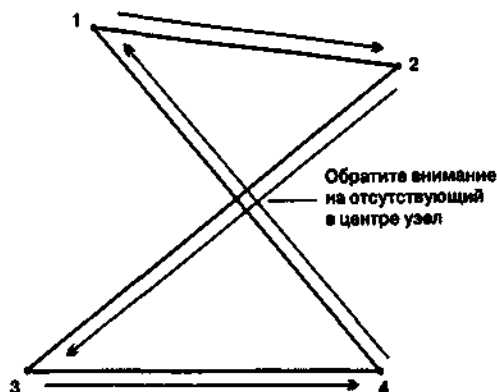


Рисунок 6.6. Странные полигоны. Пример создания странного полигона в результате неправильной последовательности ввода точек. Хотя графически мы имеем два полигона, точка, в которой линии пересекаются, не имеет узла.

Отдельной проблемой, связанной с полигонами, является создание "странных" полигонов, у которых не хватает узлов (Рисунок 6.6). В этом случае полигон является графическим артефактом, который выглядит настоящим полигоном с отсутствием одного или нескольких узлов. Обычно это случается, когда пересекаются два или более участков границы. Наиболее частой причиной такой ошибки является точка, введенная в неправильном месте или в неправильной последовательности. Например, у нас есть прямоугольник, для определения которого требуется только четыре точки (Рисунок 6.6). Его можно оцифровать, вводя сначала левую верхнюю точку, затем правую верхнюю, затем правую нижнюю, затем левую нижнюю, и закончить в левой верхней, откуда начали. Однако (возможно потому, что вы не пронумеровали точки при подготовке карты) вместо этого

вы идете от левой верхней к правой верхней, затем, по ошибке, к левой нижней, потом к правой нижней и возвращаетесь к левой верхней, откуда начали.

Хотя сами точки были введены правильно, ваш полигон больше похож на песочные часы, чем на прямоугольник. Фактически, он выглядит как два треугольника, соединенных средней точкой. Однако, центральная точка не вводилась и не является узлом.

Как вы можете догадаться из приведенного примера, простой способ избежать данной проблемы - пронумеровать вводимые точки. Но даже если вы этого не сделаете, ее можно избежать, установив единое правило оцифровки полигонов. Например, вы можете вводить их, двигаясь вдоль границы всегда по часовой стрелке. Это предохранит вас от пропуска необходимых узлов. Кстати, многие пользователи используют этот же подход при определении последовательности оцифровки частей карты за несколько сессий. Эту хорошую привычку стоит перенять.

Обнаружение странных полигонов трудно, но не невозможно. Простейший метод состоит в выделении узлов и отображении их совместно с полигональным покрытием. Области, которые по-видимому должны иметь узлы, но не имеют их, будут отличаться от оцифрованных правильно. Исправление ошибки состоит в перемещении линий в должные положения, организуя тем самым узлы в правильной последовательности. Иногда легче просто удалить ошибочные линии и использовать подсистему редактирования для повторного ввода точек в правильной последовательности.

Ошибки, которые мы рассматривали до сих пор, - наиболее простые для поиска векторные ошибки; как правило, вы можете сделать необходимые исправления, не выводя карту на печать. Более досадные графические проблемы стоят под номерами 1, 2, 3 и 6 в списке, приведенном выше.

Проблемы с пропущенными, лишними, смещенными или деформированными объектами легче всего обнаруживаются в результате вывода цифрового покрытия в том же масштабе, что использовался при вводе (Рисунок 6.7). Если вы наложите исходную и выведенную карты друг на друга на копировальном столе с подсветкой, то сможете увидеть проблемные участки.

Почти все эти ошибки обусловлены недостатком подготовки карты или неудачной организацией работы, хотя перерывы в работе и усталость всегда будут играть свою роль в их появлении. Исправление их облегчается маркировкой проблемных областей на карте, лучше с точным указанием сущности проблемы и способа ее исправления.

Если объект пропущен, отметьте его, указывая по порядку точки, линии и полигоны, которые должны быть оцифрованы, включая любые другие сведения, относящиеся к положениям узлов, и прочую топологическую информацию, которая может понадобиться. Лишние

объекты должны быть помечены для удаления. Для тех объектов, которые вышли за границы рабочей области, установленной опорными точками, точки должны быть удалены и введены заново. Деформированные или смещенные объекты обычно могут быть выбраны по отдельности и перемещены, без переоцифровки.

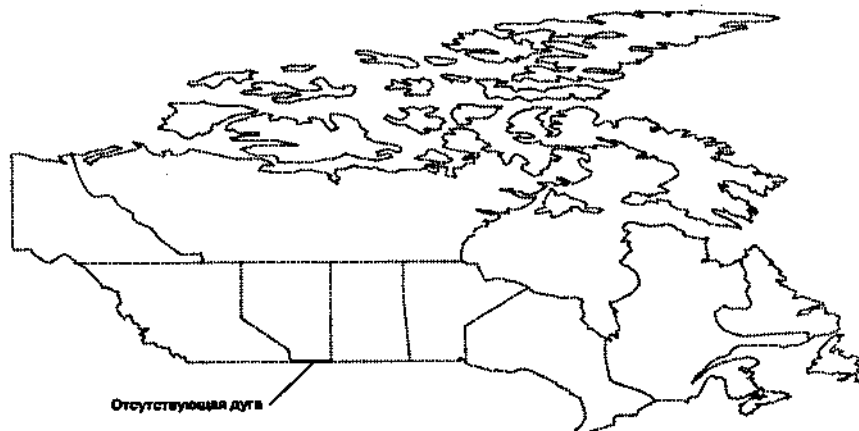


Рисунок 6.7. Отсутствующая дуга. Иллюстрация сравнения с исходной картой для определения обычных графических ошибок. Заметьте пропущенную дугу в цифровой версии этой карты Канады.

Перед тем как покинуть векторные ошибки, нужна пара предостережений. Во-первых, помните, что модифицируя объекты, вы можете изменить также и их пространственные отношения, которые вы ввели первоначально. Большинство ГИС потребуют ввести одну или более команд для подтверждения изменений. Скорее всего, нужно будет вызвать процедуру перестроения топологии на основе новых данных. Во-вторых, как это ни очевидно, вы должны сохранить вашу новую карту. Впрочем, забыв однажды сделать это после нескольких часов редактирования, вы не захотите повторения ошибки.

Ошибки атрибутов в растровых и векторных системах

Как говорилось ранее, ошибки атрибутов, включая ошибки согласования атрибутов и графики, - одни из наиболее трудных для обнаружения. Это обусловлено тем, что ГИС не знает, какие атрибуты корректны, а какие - нет. Поскольку атрибуты векторных объектов и ячеек растра значительно различаются от приложения к приложению, и поскольку для атрибутов нет эквивалента топологии, то нет и правил, по которым ГИС могла бы проверить достоверность ввода. То есть, нет явно выраженных правил утверждающих, что определенный атрибут встречается в определенной закономерности по отношению к своим соседям. Если бы было иначе, то многое из того, что мы делаем в аналитических операциях геоинформационной системы, было бы излишним. На самом деле, именно

поиск таких закономерностей чаще всего и стимулирует анализ. Возможно, после нескольких десятилетий исследований, мы сможем вычислить некоторые из них, но пока нам приходится сравнивать атрибуты цифровой БД с исходной картой для выявления большинства возможных ошибок атрибутов.

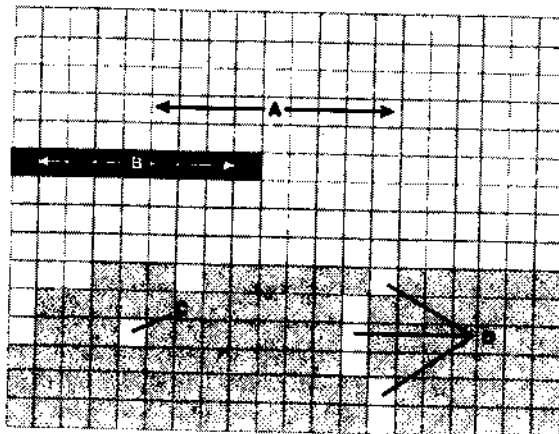


Рисунок 6.8. Ошибки атрибутов растра. Обычные ошибки атрибутов растра, определяемые по тому, как они искажают изображение: А — пропущенный ряд; В - неправильные или смещенные атрибуты (выглядят как один или более рядов существенно различных значений); С — одиночные неправильные атрибуты; D — ошибки атрибутов вдоль границ областей (вызываемые чаще всего проблемами оцифровки).

Пропуск атрибутов (*missing attributes*), возможно, - единственная ошибка атрибутов, которая может быть обнаружена без прямого сравнения с исходной картой. В случае растра они встречаются в виде потери целых рядов или колонок или частей рядов или колонок ячеек растра. Они могут быть обнаружены потому, что известность очертаний исходной карты предупредит нас об отсутствии некоторого числа ячеек растра, существенно изменяющее весь вид карты (Рисунок 6.8). Пропуск рядов или колонок в растре чаще всего вызывается смещением позиции при наборе значений ячеек растра на клавиатуре; он редко встречается при использовании сканера.

Если пропущен один или несколько рядов данных, ваша карта окажется короче, чем должна быть. Визуально это расхождение труднее заметить, но программа может сообщить, сколько рядов, колонок и ячеек растра содержится в БД. Эту информацию можно тогда сравнить с легко рассчитываемыми величинами для полной БД. Конкретные положения таких потерянных рядов обычно легко обнаружить, так как будет видна разделительная линия между одной частью карты и другой. Исправление ошибки можно сделать через выгрузку растровых данных в текстовый файл и внесение в него с помощью текстового редактора недостающих значений с перенумерацией рядов. Можно также использовать подсистему

редактирования ГИС для замены значений небольшого числа смещенных из-за пропуска ячеек или рядов, если они находятся в конце карты.

В векторных ГИС пропущенные атрибуты обычно вызываются просто тем, что ничего не было включено в таблицы атрибутов для отдельных точек, линий или полигонов. Это можно обнаружить при просмотре табличной информации или одновременном отображении объектов и их атрибутов на экране. Пропущенные атрибуты будут просто отсутствовать рядом с соответствующими объектами. Эти ошибки легко исправляются вводом должных значений атрибутов для выбранных объектов.

Неправильные значения атрибутов бывает очень трудно обнаружить, как в растровых, так и в векторных системах. В растровых ГИС, появление их как отдельных ячеек или коротких вертикальных или горизонтальных отрезков чаще всего обусловлено ошибками при наборе на клавиатуре, если используется этот метод ввода. Неправильно закодированная ячейка выглядит среди окружающих ячеек как находящаяся "не в своей тарелке". Большая разница в значениях атрибутов хорошо заметна. Когда неправильные атрибуты обнаруживаются на больших площадях, они, скорее всего, - результат ввода неправильного значения атрибута при групповом или блочном кодировании. Если они случаются как непрерывные отрезки или полосы неправильных значений атрибутов, то большинство программ позволят вам воспользоваться тем же методом кодирования для изменения этих ячеек.

Отдельные ячейки раstra могут быть выбраны и изменены индивидуально.

На растровых изображениях, имеющих мало относительно однородных областей (например, необработанные топографические карты), неправильные значения будут плохо заметны на двухмерном виде, не имея возможности создавать нарушения однородности. В таких случаях трехмерный вид поверхности будет иметь необычно высокие пики или слишком глубокие провалы. Хотя эти аномалии могут быть ошибками, их следует проверить, так как возможны и реальные аномалии. Чаще всего такие выбросы случаются в отдельных ячейках раstra, поэтому их легко можно выбрать и исправить в интерактивном режиме.

В растровых системах неправильные атрибуты могут также встречаться вдоль границ площадных объектов. В таких случаях типичным виновником являются либо неудачный выбор алгоритма оцифровки, либо невнимательность оператора при определении кодов атрибутов вдоль этих границ. Ловушка здесь состоит в том, что неправильные значения чаще всего совпадают с соседней областью, создавая тем самым впечатление, что они правильны. Вам нужно будет сравнить формы областей на цифровой карте с оригинальными формами введенной карты. Исправление данной ошибки обычно состоит в решении, какой из двух смежных областей действительно принадлежит эта ячейка? Когда это выяснено, каждая ячейка может быть выбрана и изменена, как и раньше.

Неправильные атрибуты может быть труднее обнаружить в векторных

системах, нежели в растровых, поскольку в этом случае обычно требуется хорошее знание исходной карты, ее атрибутов и их распределений. Если вы используете кодирование, которое, к примеру, заменяет реальные названия или значения числовыми кодами, то есть много шансов ввести неправильное число. В таких случаях коды не будут соответствовать табличной информации в других частях вашей БД или в словаре данных. Программа должна быть способна отметить такие несоответствия. Возьмем для примера числовое кодирование названий отдельных видов растений для точечного покрытия. В то время как оно освобождает пользователя от необходимости побуквенно точного набора названий видов на клавиатуре при формировании запросов, появляется возможность ошибочного ввода кодов.

Активный словарь данных, конечно, может обнаружить коды, не соответствующие какому-либо виду, но часто правильные коды отпечатываются в нашем подсознании, и мы можем ввести неправильный, но существующий, то есть вполне допустимый с точки зрения программы, код. Единственным способом предупреждения таких ошибок является проверка каждого введенного кода. Выявление ошибок такого типа требует сверки всех кодов с оригиналом. Работа утомительная, но позволяет обнаружить большинство таких ошибок. Выбирая объекты-нарушители, вы легко можете изменять их атрибуты, как и раньше, в интерактивном режиме.

Обычным источником ошибок атрибутов вышеупомянутого типа является тривиальный пропуск хотя бы одного значения при их наборе на клавиатуре, то есть чаще всего проблема не в том, что введены неправильные атрибуты, а в том, что атрибуты смещены, то есть поставлены в соответствие не своим объектам. Во многих случаях такие смещенные коды встречаются систематически. Например, вам может быть свойственно сбиваться при наборе кодов. Это знание может дать вам подсказку для поиска таких ошибок, особенно когда таблицы распечатываются и сравниваются с исходными данными. Можно также использовать сравнение созданной цифровой карты с исходной бумажной. Если атрибуты вводились с использованием пометок на самих объектах, это будет наилучшим способом обнаружения таких ошибок.