

Министерство образования Республики Беларусь

Учреждение образования  
«Полоцкий государственный университет»

С. Г. Шнитко

## **ГИС В ГЕОДЕЗИИ**

Конспект лекций  
для студентов специальностей 1-56 02 01 «Геодезия»  
и 1-56 02 01 02 «Инженерная геодезия»

Новополоцк  
ПГУ  
2014

УДК 528(075.8)

ББК 26.12я73

Ш77

Рекомендовано к изданию учебно-методической комиссией  
геодезического факультета в качестве конспекта лекций  
(протокол № 6 от 15.05.2014)

#### РЕЦЕНЗЕНТЫ:

канд. техн. наук, доц., декан геодезического факультета И. П. ШЕВЕЛЕВ;  
канд. техн. наук, доц. каф. геодезии и кадастров Г. Е. ГОЛОВАНЬ

#### **Шнитко, С. Г.**

Ш77

ГИС в геодезии : конспект лекций для студентов специальностей  
1-56 02 01 «Геодезия» и 1-56 02 01 02 «Инженерная геодезия» / С. Г. Шнитко. –  
Новополоцк : ПГУ, 2014. – 68 с.  
ISBN 978-985-531-474-6.

Рассмотрены базовые сведения о геоинформационных системах, их классификация и подсистемы, основные структуры данных и принципы их организации, а также источники данных для ГИС и геопространственный анализ.

Предназначен для студентов специальностей 1-56 02 01 «Геодезия» и 1-56 02 01 02 «Инженерная геодезия».

**УДК 528(075.8)**

**ББК 26.12я73**

**ISBN 978-985-531-474-6**

© Шнитко С. Г., 2014

© УО «ПГУ», 2014

# ТЕМА 1. ОБЩИЕ СВЕДЕНИЯ О ГЕОИНФОРМАЦИОННЫХ СИСТЕМАХ

## 1.1. Понятие географической информационной системы

Появление географических информационных систем относят к началу 60-х годов XX в. Именно тогда были созданы предпосылки и условия для информатизации и компьютеризации сфер деятельности в области моделирования географического пространства и решения пространственных задач. Их разработка связана с исследованиями, проведенными университетами, академическими учреждениями, оборонными ведомствами и картографическими службами.

Впервые термин «географическая информационная система» появился в англоязычной литературе и использовался в двух вариантах, таких, как *geographic information system* и *geographical information system*, очень скоро он также получил сокращенное наименование (аббревиатуру) *GIS*. Чуть позже этот термин проник в российский научный лексикон, где существует в двух равнозначных формах: исходной – «географическая информационная система» и редуцированной – «геоинформационная система». Первая из них очень скоро стала официально-представительной, а вполне разумное стремление к краткости в речи и текстах сократило последнюю из них до аббревиатуры «ГИС». Очень кратко ГИС определялись как: *информационные системы, обеспечивающие сбор, хранение, обработку, отображение и распространение данных, а также получение на их основе новой информации и знаний о пространственно-координированных явлениях.*

Информационная система – это взаимосвязанная совокупность средств, методов, персонала, используемых для хранения, обработки и выдачи информации в интересах достижения поставленной цели.

ГИС отличается от других информационных систем тем, что оперирует с пространственными, или географическими данными.

Пространственные данные – это данные, соотнесенные к некоей пространственной системе. Такой системой может быть декартова система координат, географическая система координат, классифицированная совокупность территориальных объектов и т.п.

## 1.2. Классификация информационных систем

На рисунке 1.1 показано разделение между пространственными и непространственными информационными системами (ИС).



Рис. 1.1. Классификация информационных систем

### 1.3. Наука о ГИС

*Геоинформатика* – наука, технология и производственная деятельность по научному обоснованию, проектированию, созданию, эксплуатации и использованию географических информационных систем, разработке геоинформационных технологий, приложению ГИС для практических и научных целей.

Геоинформатика является наукой, изучающей все аспекты сбора, обработки и представления информации о свойствах объектов, процессов и явлений, происходящих на Земле.

Как и всякая наука, геоинформатика имеет основную область исследований и основной метод исследований. Определение области и метода содержится в названии геоинформатики. Термин «геоинформатика» состоит из двух частей: «гео» и «информатика».

Первая часть «гео» определяет область исследования науки – объекты и явления, происходящие на земной поверхности. В геоинформатике используют пространственно-временные данные, в которых содержится информация о пространственном положении объектов, их свойствах и времени, для которого эти свойства имели место.

«Информатика» в составе термина «геоинформатика» определяет основной метод исследования: объекты исследования изучаются на основе компьютерных технологий, при этом привлекаются данные математики, картографии, геодезии и других наук. Также информатика занимается изучением и развитием систем сбора, передачи, обработки и хранения информации с помощью автоматизированных методов обработки и автоматизированных систем.

#### **1.4. Структура ГИС**

Структуру информационной системы составляет совокупность ее частей, называемых *подсистемами*.

ГИС имеют следующие подсистемы:

1. **Подсистема сбора данных**, которая собирает и проводит предварительную обработку данных из различных источников. Эта подсистема также в основном отвечает за преобразования различных типов пространственных данных (например, от изолиний топографической карты к модели рельефа ГИС).

2. **Подсистема хранения и выборки данных**, организующая пространственные данные с целью их выборки, обновления редактирования.

3. **Подсистема манипуляции данными и анализа**, которая, выполнив различные задачи на основе этих данных, группирует и разделяет их, устанавливает параметры и ограничения и выполняет моделирующие функции.

4. **Подсистема вывода**, которая отображает всю базу данных или часть ее в табличной, диаграммной или картографической форме.

##### **Подсистема сбора данных**

Первая подсистема ГИС может быть соотнесена с первым и вторым шагом процесса картографирования – сбором данных и компиляцией (составлением) карт. Исходная информация берется из таких источников, как аэрофотосъемка, цифровое дистанционное зондирование, геодезические работы, словесные описания и зарисовки, данные статистики и т.д. Использование компьютера и других электронных устройств, например дигитайзера или сканера, позволяет проводить подготовку исходных данных для записи, или кодирования точек, линий и областей к их дальнейшему использованию. Кроме того, источниками могут быть готовые цифровые карты, цифровые модели рельефа, цифровые ортофотоснимки и многие другие.

##### **Подсистема хранения и выборки данных**

Вторая подсистема – подсистема хранения и выборки – полностью соответствует нашим представлениям о функциях компьютера как хранителя информации. В ГИС подсистема хранения и выборки позволяет делать запросы, возвращающие только нужную, контекстно-связанную ин-

формацию, она переносит акцент с общей интерпретации информации на формулирование адекватных запросов. В общих словах, эта подсистема хранит, явно, либо неявно, геометрические координаты точечных, линейных и площадных геометрических объектов и связанные с ними характеристики (атрибуты). Компьютерные методы поиска естественным образом присущи самому программному обеспечению ГИС.

#### **Подсистема манипуляции данными и анализа**

Подсистема анализа позволяет значительно упростить и облегчить анализ пространственно-связанных данных, практически исключить ручной труд и в значительной мере упростить расчеты, выполняемые пользователем. Подсистема анализа является «сердцем» ГИС. Необходимость анализа карт для выделения и сравнения картин распределения земных феноменов дала импульс для поиска новых, более удобных, быстрых и мощных методов. ГИС-анализ использует потенциал современных компьютеров, сравнения и описания информации, хранящейся в базах данных, которые дают быстрый доступ к исходным данным и позволяют агрегировать и классифицировать данные для дальнейшего анализа. Они способны комбинировать выбранные наборы данных уникальными и ценными способами.

#### **Подсистема вывода**

После выполнения анализа, нужно представить его результаты. В картографии, будь то традиционная бумажная картография или ее цифровой эквивалент (компьютерная картография) выходной продукт в целом тот же – карта. Подсистема вывода позволяет компоновать результирующие данные в любой удобной для пользователя форме.

### **1.5. Отличие ГИС от традиционных систем представления реальности**

#### **1.5.1. Отличие ГИС от традиционной карты**

Традиционный подход к картам, *парадигма сообщения* (communication paradigm), подразумевал, что карта является конечным продуктом, призванным сообщать о пространственных распределениях через использование символов, классификации и т.д.

Это – традиционный взгляд на картографию, но он ограничен, поскольку пользователю карты не доступна через карту исходная, не классифицированная информация. Другими словами, пользователь, имея только конечный продукт, не может перегруппировать данные для получения большей отдачи при изменившихся обстоятельствах или потребностях.

Альтернативный подход к картографии, который поддерживает хранение исходных данных для обеспечения возможности последующей переклассификации, выработался примерно тогда же, когда изготовители карт начали

использовать достижения компьютерной техники. При этом подходе, называемом *аналитической парадигмой* (analytical paradigm), исходные атрибутивные данные сохраняются на компьютерных носителях и отображаются исходя из нужд пользователя и с использованием пользовательских классификаций.

### **1.5.2. Отличие ГИС от САПР**

Точные картографические изображения, созданные в графическом редакторе (например, в AutoCad), нельзя именовать геоинформационной системой. Такие изображения называют цифровыми картами и рассматривают как составные элементы или результат функционирования ГИС.

*Цифровая карта* (Numerical map, Digital map) – цифровая модель поверхности, сформированная с учетом законов картографической генерализации в принятых для карт проекции, разграфке, системе координат и высот. По сути, термин «цифровая карта» означает именно цифровую модель, цифровые картографические данные.

Цифровая карта создается с полным соблюдением нормативов и правил картографирования, точности карт, генерализации, системы условных обозначений. Она служит основой для изготовления обычных бумажных, компьютерных, электронных карт, входит в состав картографической базы данных, является одним из важнейших элементов информационного обеспечения ГИС и одновременно может быть результатом функционирования ГИС.

#### **Основные отличия ГИС от САПР:**

- Важным признаком ГИС является географическая привязка объектов, что дает возможность пользоваться единым координатным пространством. Трансформирование из одной координатной системы в другую и изменения проекций можно выполнять, опираясь на особенности конечного продукта. Используя жесткую координатную привязку, можно с легкостью управлять одними и теми же слоями или объектами ГИС различного типа и масштабности. В итоге пользователю предоставляют набор деталей, которые можно собирать разными способами, а вид готовой ГИС будет определяться только его творческими способностями.

- Другой фундаментальный признак ГИС – это применение аналитической обработки. В этом случае аналитический алгоритм составляется пользователем на основании запросов. Выполнив несколько последовательных операций пространственного анализа (буферизацию, объединение, вырезание, наложение), почти всегда можно получить необходимый результат.

- К одной из наиболее значимых функций ГИС относится возможность моделирования на их основе. В принципе человеку нужно только составить серию запросов: «что произойдет, если...», и простейшая модель местности или географического объекта готова.

Таким образом, нужно различать цифровую карту, изготовленную для тиражирования на бумагу или пластик, и для ГИС.

Обычно выделяют целый ряд признаков, которые позволяют отличать цифровые карты для ГИС от цифрового макета карты для печати (табл. 1.1).

Таблица 1.1

**Отличие цифровой карты для ГИС от цифрового макета карты**

Признак	Цифровая карта для ГИС	Цифровой макет карты
Форма хранения и обработки готового продукта на ПК	Набор файлов	Один файл
Координаты объектов	Реальные пространственные или местные	Условные (в пределах отдельного изображения)
Возможность преобразования изображения из одной координатной системы в другую	Да	Нет
Проекционные преобразования	Да	Нет
Преобразование из одного формата данных в другой формат	Да	Сложно, так как трансформация сложных графических примитивов приводит к потере данных
Топологическая корректность	В большинстве случаев да	Нет
Модель представления данных	Векторная и растровая	Векторная и растровая
Форматы представления данных	Графические примитивы и атрибутивная информация в виде баз данных	Графические примитивы
Графические примитивы	Точки, линии, полиномы	Точки, линии, полигоны, текст, фигуры и группы объектов (комбинация точек, линий, полигонов и фигур)
Структура графических объектов	Несколько слоев	Может быть как несколько, так и один слой
Легенда	Как инструмент управления визуализацией объектов	Как часть карты в виде группы графических объектов
Координатная привязка объектов	Точная	Используются выноски и смещение объектов, тем самым повышая наглядность
Использование пространственных запросов и моделирования для создания принципиально нового изображения графических объектов	Да	Да
Реализация тематических карт, используя включения-отключения слоев и объектов	Не обязательно, чаще нет	Да
Соблюдение стандартов представления бумажных карт	Да	Да

## ТЕМА 2. ГЕОИНФОРМАЦИОННЫЕ СТРУКТУРЫ ДАННЫХ

### 2.1. Пространственные (географические) объекты

#### 2.1.1. Пространственные данные и объекты

Информационную основу ГИС образуют цифровые представления (модели) реальности. С появлением компьютера все множество данных разделилось на два типа: *цифровые* и *аналоговые*. Последними стали именовать данные на традиционных «бумажных» носителях, используя этот термин как антоним цифровым данным. В отличие от аналоговой, цифровая форма представления, хранения и передачи данных реализуется в виде цифровых кодов или цифровых сигналов.

Объектом информационного моделирования в ГИС является *пространственный объект (географический объект)*. Это одно из ключевых понятий геоинформатики. Он может быть определен как цифровое представление (модель) объекта реальности (местности), содержащее его местоположение и набор свойств (характеристик, атрибутов), или сам объект.

Некоторое множество цифровых данных о пространственных объектах образует *пространственные данные*. Они состоят из двух взаимосвязанных частей: *позиционной (тополого-геометрической)* и *непозиционной (атрибутивной)* составляющих, которые образуют описание пространственного положения и тематического содержания данных соответственно.

Так, например, жилая постройка на дисплее может быть представлена в виде полигона (позиционная составляющая), а в атрибутивной базе данных (непозиционная составляющая) будет содержаться информация о ее площади, почтовом адресе, количестве этажей, материале стен, типе фундамента, годе постройки и т.д.

Пространственные объекты как абстрактные представления реальных объектов и предмет информационного моделирования (цифрового описания) в ГИС разнообразны и традиционно классифицируются соответственно характеру пространственной локализации отображаемых ими объектов реальности, мерности пространства, которое они образуют, модели данных, используемой для их описания, и по другим основаниям. В рамках объектно-ориентированных моделей данные могут конструироваться в новые классы объектов, отличные от базовых или созданных ранее.

### 2.1.2. Базовые типы пространственных объектов

Базовыми (элементарными) типами пространственных объектов, которыми оперируют современные ГИС, обычно считаются (в скобках приведены их синонимы) следующие:

- **точка** (точечный объект) – 0-мерный объект, характеризуемый плановыми координатами;
- **линия** (линейный объект, полилиния) – 1-мерный объект, образованный последовательностью не менее двух точек с известными плановыми координатами (линейными сегментами или дугами);
- **область** (полигон, полигональный объект, контур, контурный объект) – 2-мерный (площадной) объект, внутренняя область, ограниченная замкнутой последовательностью линий (дуг в векторных топологических моделях (данных) или сегментов в модели «спагетти») и идентифицируемая внутренней точкой (меткой);
- **пиксел** – 2-мерный объект, элемент цифрового изображения, наименьшая из его составляющих, получаемая в результате дискретизации изображения (разбиения на далее неделимые элементы растра), элемент дискретизации координатной плоскости в растровой модели (данных) ГИС;
- **ячейка** (регулярная ячейка) – 2-мерный объект, элемент разбиения земной поверхности линиями регулярной сети;
- **поверхность** (рельеф) – 2-мерный объект, определяемый не только плановыми координатами, но и аппликацией  $Z$ , которая входит в число атрибутов образующих ее объектов; оболочка тела;
- **тело** – 3-мерный (объемный) объект, описываемый тройкой (триплетом) координат, включающей аппликату  $Z$ , и ограниченный поверхностями.

### 2.1.3. Цифровое описание пространственного объекта

Общее цифровое описание пространственного объекта включает:

- наименование;
- указание местоположения (местонахождения, локализации);
- набор свойств;
- отношения с другими объектами;
- пространственное «поведение».

Два последних элемента описания пространственного объекта факкультативны.

**Наименованием** объекта служит его географическое наименование (имя собственное, если оно есть), его условный код и/или идентификатор, присваиваемый пользователем или назначаемый системой.

В зависимости от типа объекта его **местоположение** определяется парой (триплетом) координат (для точечного объекта) или набором координат, организованным определенным образом в рамках некоторой модели данных. Это геометрическая часть описания данных, геометрия (метрика) рассматриваемых пространственных объектов, отличная от их семантики (непозиционных свойств).

Перечень **свойств** соответствует атрибутам объекта, качественным и количественным его характеристикам, которые приписываются ему в цифровом виде пользователем, могут быть получены в ходе обработки данных или генерируются системой автоматически (к последнему типу атрибутов принадлежат, например, значения площадей и периметров полигональных объектов). Существует расширенное толкование понятия атрибута объекта; последнему могут быть поставлены в соответствие любые типы данных: текст, цифровое изображение, видео- или аудиозапись, графика (включая карту), что, по существу, реализуется на практике в мультимедийных электронных атласах. Под атрибутами понимаются именно содержательные, тематические (непозиционные, непространственные) свойства объектов.

Под **отношениями** понимают, прежде всего, топологические свойства (топологию). К топологическим свойствам пространственного объекта принято относить его:

- *размерность* (мерность, пространственную размерность), сообразно которой выше были выделены 0, 1, 2 и 3-мерные объекты;
- *замкнутость*, если речь идет о линейных объектах в широком смысле слова; связность; простота (отсутствие самопересечения линейных объектов и «островов» в полигоне);
- *нахождение на границе, внутри или вне полигона*;
- *признак точечного объекта*, указывающий, является ли он конечным для некоторой линии.

Примерами топологических отношений объектов являются их свойства «пересекаться» (или «не пересекаться»), «касаться», «быть внутри», «содержать», «совпадать».

Топология вместе с геометрией образует тополого-геометрическую часть описания данных, его позиционную часть. Таким образом, в самом общем виде в пространственных данных следует различать и выделять три составные части: *топологическую, геометрическую и атрибутивную* – «геометрию», «топологию» и «атрибутику» цифровой модели пространственного объекта.

Четкое разделение позиционных и непозиционных данных – историческая традиция, имеющая определенные технологические корни. Управление

атрибутивной частью данных обычно возлагается на средства систем управления базами данных (СУБД), встроенных в программные средства ГИС или внешних по отношению к ним. В наиболее яркой форме оно реализовано в так называемой векторной модели данных, атрибуты которой представлены таблицей, хранятся и управляются СУБД, поддерживающей реляционную модель данных, а их позиционная часть, связанная с атрибутивной через идентификаторы пространственных объектов, управляется другими средствами. Модели пространственных данных такого типа получили широкое распространение и наименование геореляционных. Будучи еще недавно практически единственной и став классической, геореляционная модель не выглядит достаточно изящной. Современной альтернативой является интегрированный подход, когда и атрибутивная, и тополого-геометрическая части данных хранятся и управляются в единой среде СУБД, а также объектный и объектно-реляционный подходы (и одноименные им типы моделей данных).

## **2.2. Виды компьютерных моделей пространственных объектов**

В ГИС географические объекты могут быть представлены следующими распространенными компьютерными моделями:

- *Векторные модели* – это модели географических объектов на основе представления векторов совокупностями координатных пар.
- *Растровые модели* – это модели географических объектов в виде совокупности ячеек регулярной сетки или растра.
- *Триангуляционные модели* – это модели географических объектов в виде сети смежных треугольников.

## **2.3. Векторные модели географических объектов**

Существуют два вида векторных моделей:

- *векторная нетопологическая модель;*
- *векторная топологическая модель.*

Топология – это раздел математики, изучающий идею непрерывности. Топология описывает пространственные отношения между соединенными или прилегающими векторными объектами (точками, линиями и полигонами) в ГИС.

### **2.3.1. Векторная нетопологическая модель**

Пространственная информация о географических объектах представляется наборами элементарных графических примитивов – точек, линий, областей (полигонов).

Семантическая информация, передается условными знаками (символами и цветом) и надписями (например, название улицы, этажность дома и т.д.).

Векторные модели широко применяются в САПР. Они строятся на векторах, занимающих часть пространства в отличие от занимающих все пространство растровых моделей. Это определяет их основное преимущество – на порядки меньший объем памяти для хранения и меньшие затраты времени на обработку и представление, а главное – более высокая точность позиционирования и представления данных.

При построении векторных моделей объекты создаются путем соединения точек прямыми линиями, дугами окружностей, полилиниями. Площадные объекты задаются наборами линий.

Каждое векторное образование в цифровой форме представляется координатными парами  $X, Y$ :

- Точечный географический объект представляется одной координатной парой  $X, Y$ .
- Линейный географический объект представляется последовательностью координатных пар  $X_1, Y_1; X_2, Y_2; X_3, Y_3; \dots$  сегментов полилинии.
- Полигональный географический объект представляется последовательностью координатных пар  $X_1, Y_1; X_2, Y_2; X_3, Y_3; \dots; X_1, Y_1$  сегментов полигона. В этом списке совпадение первой и последней пары координат означает, что полигон замкнутый.

При описании множества полигонов каждый отрезок границы, заключенный между двумя узловыми точками (за исключением внешней границы полигонов), будет описан в этом случае дважды (по и против часовой стрелки).

Такая модель данных для описания точечных, линейных и полигональных объектов носит наименование модели «спагетти» (рис. 2.1). Она не является эффективной с точки зрения избыточности хранимых данных и возможностей использования аналитических операций ГИС и поддерживается недорогими программными средствами настольного картографирования и ГИС.

В данной модели пространственные отношения между объектами (топология), например, положение смежных областей, подразумеваются, а не записываются в компьютер в явном виде. И все отношения между всеми объектами должны вычисляться независимо. Результатом отсутствия такого явного описания топологии выступает огромная дополнительная вычислительная нагрузка, которая затрудняет измерения и анализ. Но так как «спагетти»-модель очень сильно напоминает бумажную карту, она является эффективным методом картографического отображения и часто используется в компьютеризированной картографии, где анализ не является главной целью.



Структура данных

Объект	Номер	Положение
Точка	5	одна пара координат (x,y)
Линия	16	набор пар координат (x,y)
Область	25	набор пар координат (x,y), первая и последняя совпадают

Рис. 2.1. «Спагетти»-модель векторных данных

### 2.3.2. Векторная топологическая модель

Векторная топологическая (линейно-узловая) модель обязана своим происхождением задаче описания полигональных объектов. С ней связаны и особые термины, отражающие ее структуру; ее главные элементы (примитивы):

- промежуточная точка;
- сегмент (линейный сегмент, отрезок (прямой));
- узел;
- дуга;
- полигон (область, полигональный объект, многоугольник, контур, контурный объект), в том числе:
  - простой полигон;
  - внутренний полигон («остров», анклав);
  - составной полигон;
  - универсальный полигон (внешняя область).

В отличие от «спагетти»-модели, топологические модели (рис. 2.2), содержат топологическую информацию в явном виде. Для поддержки продвину-тых аналитических методов нужно внести в компьютер как можно больше яв-

ной топологической информации. Подобно тому, как математический сопроцессор объединяет многие специализированные математические операции, так и топологическая модель данных объединяет решения некоторых из наиболее часто используемых в географическом анализе функций. Это обеспечивается включением в структуру данных информации о смежности для устранения необходимости ее определения при выполнении многих операций.

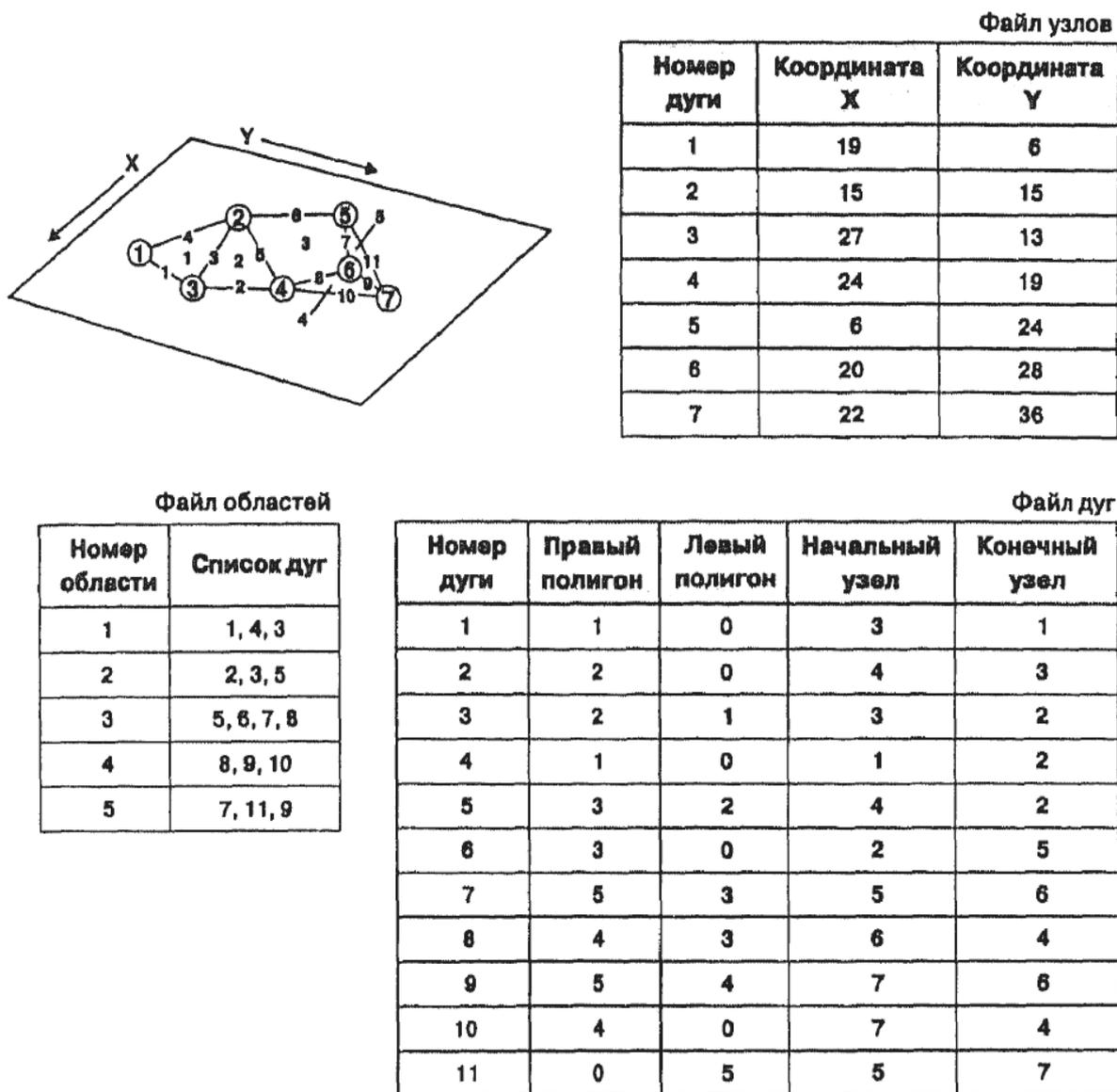


Рис. 2.2. Топологическая векторная модель данных

Топологическая информация описывается набором узлов и дуг.

**Узел (node)** – больше, чем просто точка, обычно это пересечение двух или более дуг, и его номер используется для ссылки на любую дугу, которой он принадлежит.

Каждая **дуга (arc)** начинается и заканчивается либо в точке пересечения с другой дугой, либо в узле, не принадлежащем другим дугам.

Дуги образуются последовательностями **отрезков**, соединенных промежуточными (формообразующими) точками. В этом случае каждая линия имеет два набора чисел: пары координат промежуточных точек и номера узлов. Кроме того, каждая дуга имеет свой идентификационный номер, который используется для указания того, какие узлы представляют ее начало и конец.

**Области**, ограниченные дугами, также имеют идентифицирующие коды, которые используются для определения их отношений с дугами. Каждая дуга содержит явную информацию о номерах областей слева и справа от нее, что позволяет находить смежные области. Эта особенность данной модели позволяет компьютеру знать действительные отношения между графическими объектами.

### 2.3.3. Векторизация

Векторные модели получают разными способами. Один из наиболее распространенных – *векторизация* сканированных (растровых) изображений.

Векторизация – процедура выделения векторных объектов с растрового изображения и получение их в векторном формате.

Для векторизации необходимо высокое качество (отчетливые линии и контуры) растровых образов. Чтобы обеспечить требуемую четкость линий, иногда приходится заниматься улучшением качества изображения.

При векторизации возможны ошибки, исправление которых осуществляется в два этапа:

- 1) корректировка растрового изображения до его векторизации;
- 2) корректировка векторных объектов.

### 2.3.4. Особенности векторных моделей

В векторных форматах набор данных определен объектами базы данных. Векторная модель может организовывать пространство в любой последовательности и дает «произвольный доступ» к данным.

В векторной форме легче осуществляются операции с линейными и точечными объектами, например, анализ сети – разработка маршрутов движения по сети дорог, замена условных обозначений.

Что касается точности векторных данных, то можно говорить о преимуществе векторных моделей перед растровыми, так векторные данные могут кодироваться с любой мыслимой степенью точности, которая ограничивается лишь возможностями метода внутреннего представления коор-

динат. Обычно для представления векторных данных используется 8 или 16 десятичных знаков (одинарная или двойная точность).

Не все природные явления имеют характерные четкие границы, которые можно представить в виде математически определенных линий. Это обусловлено динамикой явлений или способами сбора пространственной информации. Почвы, типы растительности, склоны, место обитания диких животных – все эти объекты не имеют четких границ.

Обычно линии на карте имеют толщину 0,4 мм и, как часто считается, отражают неопределенность положения объекта. В растровой системе эта неопределенность задается размером ячейки. Поэтому следует помнить, что в ГИС действительное представление о точности дают размер растровой ячейки и неопределенность положения векторного объекта, а не точность координат.

### 2.3.5. Форматы векторных данных

Векторные графические форматы данных представлены в таблице 2.1.

Таблица 2.1

**Векторные графические форматы данных**

Наименование формата	Описание
DXF, DWG, DGN	Форматы данных систем автоматизированного проектирования (САПР)
DX90	Формат цифровых навигационных карт
DLG	Формат данных геологической съемки США
DWF	Формат передачи графических данных по Интернету
F1M	Формат данных Роскартографии, предназначен для обмена данными
GEN	Обменный формат ГИС-пакета ARC/INFO
SHP	Формат данных ГИС-пакета ArcView (шейп-файл), описывается несколькими файлами записей с определенными расширениями: .shp – позиционные данные; .shx – индекс формы пространственных данных; dbf – атрибутивные данные и др.
TAB	Формат ГИС-пакета MapInfo; описывается файлами с определенными расширениями: .tab – текстовое описание структуры данных таблиц; .dat – табличные данные; .map – графические объекты; .ID – список указателей (индекс) на графические объекты
MIF/MID	Обменный формат ГИС-пакета MapInfo
HPGL	Формат вывода на принтер или графопостроитель
DMF	Формат ГИС-пакета Digital

### 2.3.6. Векторная модель для представления поверхностей

#### Определение модели TIN

В отличие от цифровых представлений точечных, линейных и площадных объектов, трехмерные объекты требуют особых форм представления, т.к. их местоположение описывается не только двумерными, но и высотными координатами. К наиболее распространенному типу трехмерных объектов относится топографический рельеф земной поверхности. При помощи трехмерных объектов могут быть также смоделированы карты плотности населения, атмосферного давления, влажности и т.п.

Цифровые модели рельефа позволяют по конечному набору выборочных точек определять возвышение, крутизну склона, направление ската в произвольной точке на местности. Возможно выявление особенностей местности – бассейнов рек, дренажных сетей, пиков, впадин и т.п. Такие модели широко применяются во многих процедурах ГИС-анализа: при выборе места строительства зданий и коммуникаций, в анализе дренажных сетей, видимости, при выборе маршрута движения по пересеченной местности.

Поверхности являются непрерывными феноменами в противоположность дискретным объектам, выражаемым точками, линиями и полигонами. Но существуют способы представления поверхностей, в которых используется конечное количество точек. Разные подходы к выбору узловых точек, в которых известно значение возвышения поверхности, определяют две наиболее распространенные модели данных. В геоинформационных системах поверхности обычно описываются при помощи растровых моделей и триангуляционных сетей. В растровых моделях выборочные точки расположены в узлах регулярной растровой решетки, а в триангуляционных сетях – располагаются нерегулярно так, чтобы наилучшим образом «обогнуть» поверхность (отсюда название – triangulated irregular networks – TIN).

Модель TIN (Triangulation Irregular Network – триангуляционная нерегулярная сеть) географических объектов – модель поверхности в виде сети смежных не пересекающихся треугольных граней, определенная по узлам и ребрам, которые покрывают поверхность (рис. 2.3).

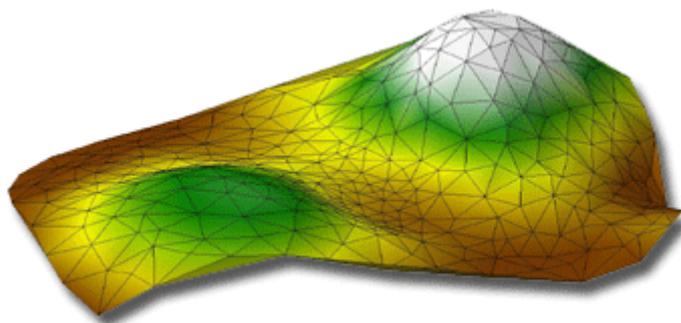


Рис. 2.3. Модель TIN

Геометрия модели TIN образуется гранями, узлами и ребрами в трехмерном пространстве.

Грань (Face) – поверхность треугольника в трехмерном пространстве.

Узел (Node) – вершина треугольника с координатами X, Y, Z.

Ребро (Edge) – сторона треугольника в трехмерном пространстве.

Каждая грань TIN является частью поверхности в 3D-пространстве.

### **Свойства модели TIN**

Модель TIN обладает следующими свойствами:

- позволяет получить точное представление о локальной части поверхности, используя переменную плотность узлов со значением Z и линии перегиба поверхности;
- является основой 3D-визуализации поверхности;
- позволяет выполнить сложный анализ поверхности (вычисление высот, уклонов, экспозиций склонов, получение изолиний поверхности, расчеты объемов, вертикальные профили по трассе линии, анализ видимости).

### **Триангуляция Делоне**

Исходными данными для построения TIN является набор точек с координатами X, Y, Z. Задача заключается в том, чтобы по этому набору точек создать сеть смежных непересекающихся треугольников.

Задача построения триангуляции по набору точек является одной из базовых в вычислительной геометрии. К ней сводятся многие другие задачи, она широко используется в машинной графике и геоинформационных системах для моделирования поверхностей и решения пространственных задач.

Наибольшее распространение в ГИС получила триангуляция Делоне (Delaunay), которая названа по имени ее автора – советского математика Бориса Николаевича Делоне (1890 – 1980). По определению Делоне три точки формируют треугольник в триангуляции тогда и только тогда, когда в окружности, описанной вокруг этого треугольника нет других точек разбиения. Каждый ограничивающий треугольник круг не содержит точек из набора внутри него.

Один из алгоритмов построения триангуляции Делоне основан на генерировании полигонов Тиссена (Thiessen) или Вороного. Для этого поверхность разбивается на области, в которых каждая точка расположена ближе всего к некоторому узлу сети – генерирующей точке. Полученные границы называют полигонами Тиссена или полигонами Вороного. Две точки соединяются линией в триангуляции Делоне, если их полигоны Тиссена имеют общую границу. Этот метод позволяет получить требуемые треугольники. Полигоны Тиссена используются также при анализе близости.

## Топология в TIN

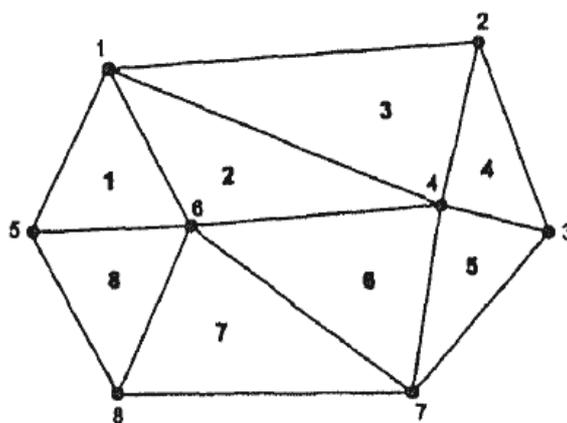
Модель TIN – это топологическая структура данных: ребра соединяются в узлах; каждый треугольник смежный с соседними треугольниками.

Топологические отношения создаются путем создания в базе данных для каждого узла указания на смежные узлы. Пространство вокруг территории представляется фиктивным узлом.

Модель TIN (рис. 2.4) в цифровом виде описывается связанными файлами – файлом вершин, файлом указателей и файлом треугольников. Для каждой точки разбиения сохраняется ее уникальный номер, координаты и список точек, с которыми она соединена прямыми (по часовой стрелке).

Точка	X	Y	Z
1	X <sub>1</sub>	Y <sub>1</sub>	Z <sub>1</sub>
2	X <sub>2</sub>	Y <sub>2</sub>	Z <sub>2</sub>
3	X <sub>3</sub>	Y <sub>3</sub>	Z <sub>3</sub>
4	X <sub>4</sub>	Y <sub>4</sub>	Z <sub>4</sub>
5	X <sub>5</sub>	Y <sub>5</sub>	Z <sub>5</sub>
6	X <sub>6</sub>	Y <sub>6</sub>	Z <sub>6</sub>
7	X <sub>7</sub>	Y <sub>7</sub>	Z <sub>7</sub>
8	X <sub>8</sub>	Y <sub>8</sub>	Z <sub>8</sub>

Файл точек



Треуголь- ник	Вершины			Соседние треугольники		
	1	2	3	1	2	3
1	1	6	5	2	8	-
2	1	4	6	1	3	6
3	1	2	4	-	4	2
4	2	3	4	3	-	5
5	4	3	7	4	-	6
6	6	4	7	2	5	7
7	6	7	8	6	-	8
8	5	6	8	1	7	-

Файл треугольников

Рис. 2.4. Топология TIN

## Этапы создания модели TIN

Модель TIN создается в следующей последовательности.

**Этап 1.** Задание множества точек  $i$  по координатам  $X_i, Y_i, Z_i$  (рис. 2.5).



Рис. 2.5. Множество точек

**Этап 2.** Построение триангуляция Делоне (рис. 2.6).

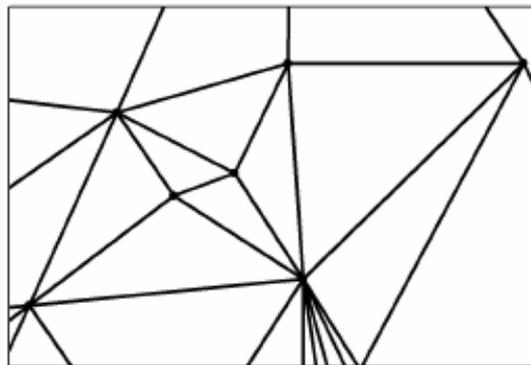


Рис. 2.6. Триангуляция Делоне

**Этап 3.** Ввод линий перегиба поверхности (Breakline) и модификация TIN с учетом линий перегиба. Линии перегиба рельефа определяют резкие изменения поверхности, такие как линия верха, низа откоса, гребни, тальвеги и др. (рис. 2.7).

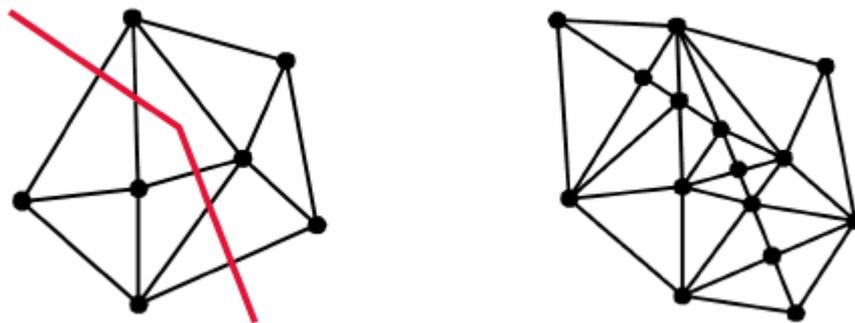


Рис. 2.7. Линии перегиба поверхности

**Этап 4.** Ввод областей исключения (Exclusion areas) с постоянным значением  $Z$  и модификация TIN с учетом полигональных объектов, например водной поверхности (рис. 2.8).

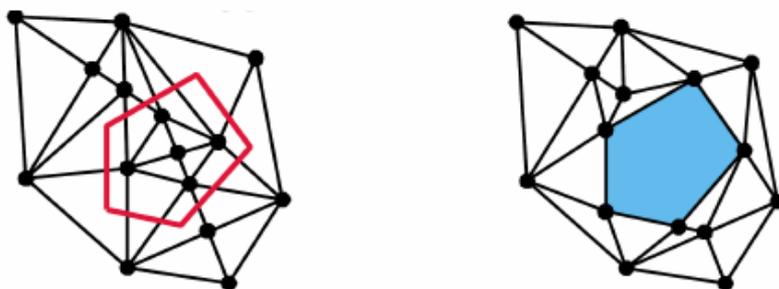


Рис. 2.8. Области исключения

**Этап 5.** Решение задач по модели TIN поверхности.

## 2.4. Растровые модели географических объектов

### 2.4.1. Концепция растровых моделей географических объектов

Земное пространство непрерывно. Для представления земного пространства используются модели данных, которые в любом случае основаны на наборах дискретных объектов.

В векторной модели дискретизация реального земного пространства выполняется посредством выделения дискретных географических объектов и отображения их отдельными пространственными элементами – точками, линиями, областями.

В растровой модели дискретизация реального земного пространства достигается путем разделения его непрерывной последовательности на множество смежных элементарных объектов – пространственных ячеек. Ячейки являются базовыми элементами растровой модели данных.

Ячейка – это наименьшая единица информации растровой модели. Под ячейкой понимается гомогенный объект, т.е. такой объект, который имеет одну характеристику или одно значение. Поэтому растровая модель дает информацию о том, что расположено в данном месте. Для сравнения векторная модель дает информацию о том, где расположен объект. Для каждой растровой модели используются ячейки одного выбранного размера. В ГИС в зависимости от способа задания ячеек различают два типа пространственных ячеек:

- пиксел (Pixel – Picture Element) – ячейка, представляющая собой минимальный неделимый далее элемент изображения;
- ячейка (Cell) – ячейка заданного размера в форме квадрата, прямоугольника, треугольника, шестиугольника или трапеции.

Наиболее распространенная форма ячейки – квадрат или прямоугольник.

Растровая модель данных в ГИС предполагает разбиение пространства (координатной плоскости) с вмещающими ее пространственными объектами на аналогичные пикселям дискретные элементы, упорядоченные в виде прямоугольной матрицы. Для цифрового описания (позиционирования) точечного объекта будет достаточно указать его принадлежность к тому или иному элементу дискретизации, учитывая, что его положение однозначно определено мерами столбца и строки матрицы (при необходимости координаты пикселя, либо его центроида или любого угла могут быть вычислены).

Пикселу присваивается цифровое значение, определяющее имя или семантику (атрибут) объекта. Аналогичным образом описываются линейные и полигональные объекты: каждый элемент матрицы получает значение, соответствующее принадлежности или непринадлежности к нему того или иного объекта (рис. 2.9).

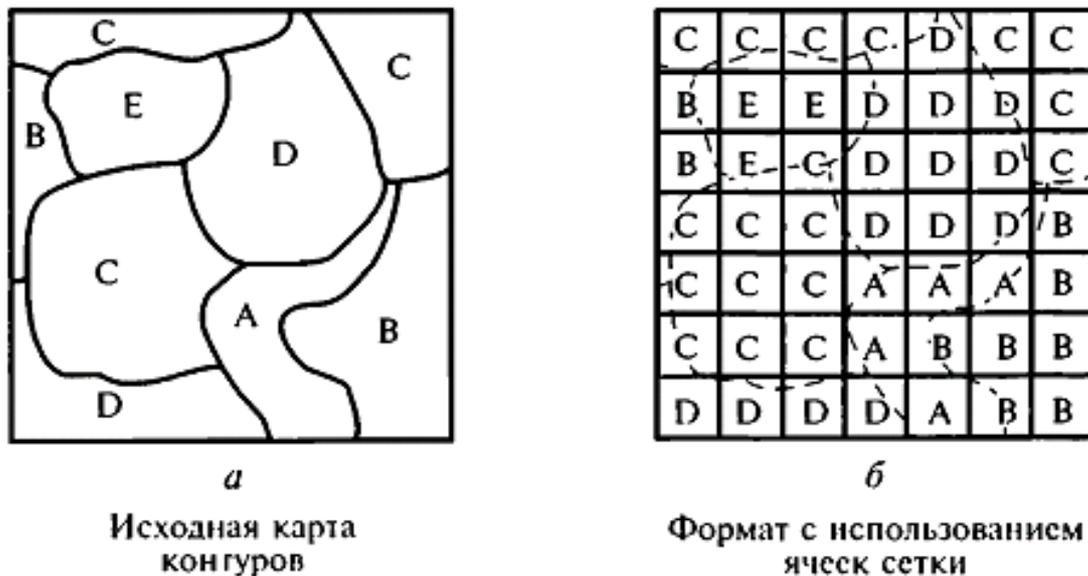


Рис. 2.9 Растровая модель

В растровых системах есть два способа добавления атрибутивной информации об объектах. Простейшим является присвоение значения атрибута каждой ячейке растра (например, индекс растительности). Но в таком варианте каждая ячейка имеет только одно значение атрибута. Второй подход – связывание каждой ячейки растра с базой данных так, что любое число атрибутов может быть присвоено каждой ячейке растра (рис. 2.10).



Рис. 2.10. Атрибуты растра

Представление исходных полигональных объектов в виде растра может показаться весьма грубым приближением их истинной формы. Однако, выбрав подходящий размер пиксела растровой модели, можно добиться пространственного разрешения (точности представления объектов), удовлетворяющего целям их цифрового описания и последующей обработки, если этому не препятствуют соображения экономии машинной памяти: двукратное увеличение разрешения ведет к четырехкратному росту объемов хранимых данных и т.д. Полученная матрица образует растровый слой с однотипными объектами; множество разнотипных объектов образует набор слоев, составляющих полное цифровое описание моделируемой предметной области. С каждым семантическим значением или кодом пиксела, кроме того, может быть связан неограниченный по длине набор (таблица) атрибутов, каждый из которых можно развернуть в производный слой, соответствующий размеру исходной матрицы.

#### 2.4.2. Характеристики растровых моделей

Для растровых моделей существует ряд характеристик:

- разрешение,
- ориентация,
- зоны,
- значение,
- положение.

**Разрешение** – минимальный линейный размер наименьшего участка отображаемого пространства (поверхности), отображаемый одним пикселем.

Пиксели обычно представляют собой прямоугольники или квадраты, реже используются треугольники и шестиугольники. Более высоким разрешением обладает растр с меньшим размером ячеек. Высокое разрешение подразумевает обилие деталей, множество ячеек, минимальный размер ячеек.

В общем случае разрешение выражают в следующих единицах:

- ppi (pixel per inch) – пиксел на дюйм,
- dpi (dot per inch) – точек на дюйм,
- lpi (line per inch) – линий на дюйм.

**Значение** – элемент информации, хранящийся в элементе растра (пикселе). Поскольку при обработке применяют типизированные данные, то есть необходимость определить типы значений растровой модели.

Тип значений в ячейках растра определяется как реальным явлением, так и особенностями ГИС. В частности, в разных системах можно использовать разные классы значений: целые числа, действительные (десятичные) значения, буквенные значения.

Целые числа могут служить характеристиками оптической плотности или кодами, указывающими на позицию в прилагаемой таблице или легенде. Например, возможна следующая легенда, указывающая наименование класса почв: 0 – пустой класс, 1 – суглинистые, 2 – песчаные, 3 – щебнистые и т.п.

**Ориентация** – угол между направлением на север и положением колонок растра.

**Зона** растровой модели включает соседствующие друг с другом ячейки, имеющие одинаковое значение. Зоной могут быть отдельные объекты, природные явления, ареалы типов почв, элементы гидрографии и т.п.

### 2.4.3. Растровое представление поверхности

Растры представляют поверхности в виде регулярной матрицы двумерных ячеек со значением  $Z$ . Каждая ячейка хранит свое значение  $Z$ . Трехмерная растровая модель поверхности представляет собой совокупность смежных блоков. Значения  $Z$  отображаются соответствующим цветом (рис. 2.11).

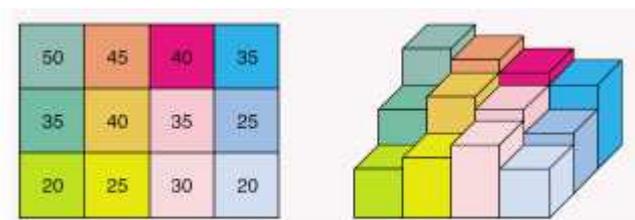


Рис. 2.11. 2D- и 3D-растровое представление поверхности

С растровым представлением поверхности связано сеточное представление поверхности. Каждую ячейку растра идентифицируют местоположением и значением  $Z$ . Для определенных типов данных значение ячейки представляет собой измеренное значение в центральной точке ячейки.

Для сеточного представления поверхности базовой структурой является сетка, образованная параллельными и взаимно перпендикулярными линиями, проходящими через центральные точки ячеек растра. В ГИС для отображения поверхности на основании такой сетки используются две основные модели – грид и латтис.

Грид (Grid) – модель в виде коллекции центральных точек ячеек со значениями  $Z$ , расположенных регулярно через горизонтальные интервалы (в отличие от формата ESRI Grid). Грид передает форму поверхности точками в трехмерном пространстве. Если значениями  $Z$  являются высоты местности, грид представляет собой цифровую модель высот (Digital elevation model – DEM).

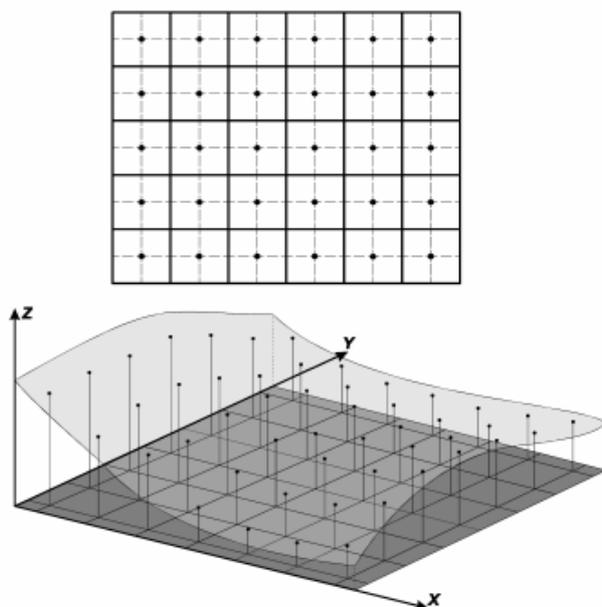


Рис. 2.12. Грид в 2D и 3D представлении

#### 2.4.4. Недостатки и преимущества растровых моделей

По сравнению с векторными моделями растровые модели обладают следующими недостатками:

- географические объекты характеризуются менее точной информацией о местоположении и размерах;
- разрывы непрерывности передаются недостаточно хорошо;
- точные местоположения точек вершин, дна теряются;
- растры требуют больших объемов памяти.

Представление географических объектов растровыми моделями имеет следующие преимущества:

- растр отображает непрерывно охватываемую территорию;
- растровые данные проще для обработки и обеспечивают более высокое быстродействие;
- ввод растровых данных менее трудоемкий.

#### 2.4.5. Форматы растровых данных

Для хранения растровых данных используются многие форматы; некоторые из них представлены в таблице 2.2.

Таблица 2.2

**Форматы растровых данных**

Название	Характеристика формата
BMP (Microsoft Windows/IBM® OS/2® Bitmap)	Битовый двоичный формат растровых графических файлов, предусматривающий 4, 8 и 24 бита на точку.
TIFF (Tagged Image File Format)	Формат для создания и обмена изображениями высокого качества; поддерживает цветные (до 24 бит), черно-белые изображения и градации серого; использует разные типы сжатия данных с учетом фотометрических свойств изображения.
TIFF-LZW	TIFF-LZW ERDAS Imagine – формат сжатия растровых данных, основанный на алгоритме Lempel Ziff & Welsh, поддерживается при наличии специальной библиотеки. Коэффициент сжатия изображения, представленного в формате TIFF, составляет 1:7,7
GeoTIFF	Расширение формата TIFF для передачи изображений, имеющих пространственную привязку, включает информацию о системе координат и проекции, параметрах геометрической коррекции.
JPEG (Joint Photographic Experts Group)	Формат для сжатия изображений с коэффициентом более 25:1 с потерей качества; позволяет передавать до 16 млн. цветов (до 32 бит); распространен для изображений в Интернете.
JFIF (JPEG File Interchange Format)	Формат файлов для хранения и передачи изображений, сжатых по алгоритму JPEG.
PCX	Распространенный формат графических файлов для цветных изображений, используемый в большинстве графических редакторов. Изображение сжимается до 1:1,5.
ERDAS IMAGINE	Файлы разработаны, используя программное обеспечение обработки изображений IMAGINE. Файлы ERDAS IMAGINE могут хранить непрерывные и дискретные одноканальные или многоканальные данные.
IMG (Image)	Формат сканированных с высоким разрешением изображений. Графический файл в пакете GEM. Используется во многих растровых ГИС-пакетах, например, ERDAS.
ER Mapper	ER Mapper файлы разработаны, используя программное обеспечение обработки изображений ER Mapper.
ESRI Grid	Формат, разработанный фирмой ESRI, поддерживающий 32-битовые целые и действительные растровые сетки. Целые гриды предназначены для представления дискретных данных, действительные гриды используются для представления непрерывных явлений.

#### 2.4.6. Файл геопривязки растровых данных

Файлы растровых данных в выбранном формате представляют изображение, в котором положение каждой ячейки фиксируется координатами изображения – номером ряда и номером колонки. Эти данные не содержат информацию о координатах ячеек в принятой системе координат для реального земного пространства, и поэтому их недостаточно для представления географических объектов.

Для представления географических объектов необходимо установить преобразование координат изображения в координаты реального мира. Специализированные форматы, такие как ERDAS MAGINE, BSQ, BIL, BIP, GeoTIFF, Grid хранят геопространственную информацию в заголовке файла изображения. Растровые данные, представленные в общих форматах, необходимо совместно использовать с отдельными файлами геопривязки – World-файлами, которые содержат информацию о преобразовании координат изображения в координаты реального мира. Файл геопривязки имеет то же название, что и файл изображения, но его расширение дополнено в конце буквой w (World).

Например, для файла изображения map.tif файл геопривязки будет map.tifw или map.tfw.

Файл геопривязки представляется в формате ASCII; он содержит информацию для преобразования координат изображения в координаты реального земного пространства (табл. 2.3).

Таблица 2.3

**Пример содержания файла геопривязки**

Аргументы	Пример значений аргументов
Размер пиксела по оси X	0.423566666650000
Коэффициенты поворота	0.000000000000000 0.000000000000000
Размер пиксела по оси -Y	-0.423566666650000
Координаты центра верхнего левого пиксела	79000.000000000000000 71000.000000000000000

## **ТЕМА 3. ОРГАНИЗАЦИЯ ДАННЫХ В ГЕОИНФОРМАЦИОННЫХ СИСТЕМАХ**

### **3.1. Определение, значение и задачи системной организация данных**

Организацией данных называют процесс сведения разнородных данных и моделей в единую логически непротиворечивую модель, которую в дальнейшем можно будет эффективно применять в различных технологиях анализа и управления.

Собранные данные могут храниться в виде наборов или файлов. Кроме того, при сборе данные могут организовывать связанные совокупности, называемые моделями. Для того, чтобы разнородные данные и модели можно было обрабатывать в одной системе, они должны быть упорядочены и сведены к единой информационной модели, в которой будут дополнять друг друга. Результатом организации данных является создание такой информационной модели, которая позволяет организовать эффективное хранение в базе данных и эффективную обработку в информационных системах при различных технологиях. Организация данных придает качественно новые свойства геоданным. Именно она дает возможность использовать географические данные при решении широкого круга прикладных задач управления, анализа, логистики, планирования, проектирования, прогнозирования, использования ресурсов, мониторинга и др.

### **3.2. Принципы организации данных в ГИС**

Существует множество структур данных, используемых для представления географических объектов, в зависимости от потребностей в данных и их последующего использования. Наибольшее распространение получили два общепринятых принципа организации моделей пространственных данных:

- 1) принцип послойной организации информации (его часто называют классическим);
- 2) объектно-ориентированный принцип организации данных.

#### **3.2.1. Послойный принцип организации информации**

Реальный мир состоит из множества географических объектов. Декомпозиция реального мира может быть выполнена путем стратификации - процедуры, которая разбивает множество географических объектов на страты или

слои, удобные для обработки и анализа. В результате изучаемая область реального мира представляется набором слоев взаимосвязанных данных.

Послойный принцип организации информации заключается в том, что данные о территории организуются (раслаиваются) в виде набора тематических слоев. Слой состоит из однородных данных, объединяемых общей тематикой. Например, в один слой выносятся все объекты гидрографии, или все шоссейные дороги, или все, относящееся к растительному покрову (рис. 3.1).

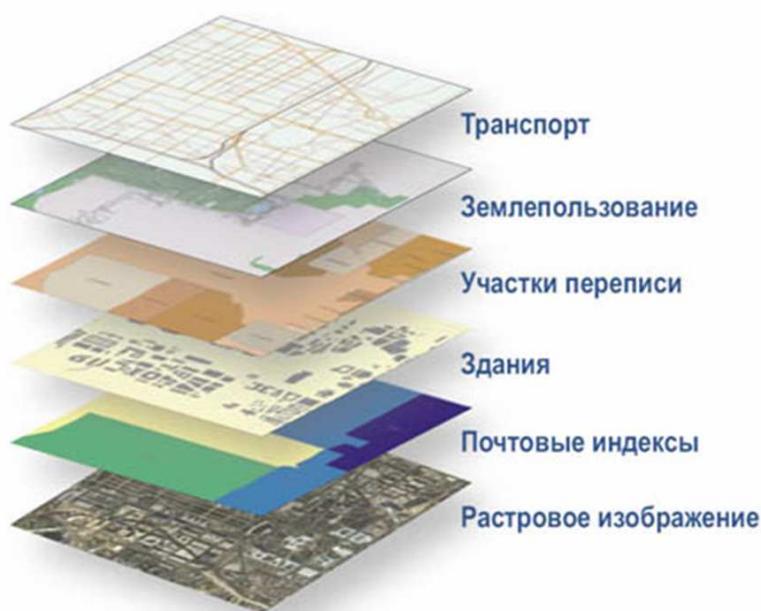


Рис. 3.1. Послойный принцип организации информации

### 3.2.2. Объектно-ориентированный принцип организации данных

Другой принцип появился относительно недавно и связан с объектно-ориентированным подходом, типовым для современного программирования. При этом группировка объектов соответствует их логическим взаимосвязям. Объектно-ориентированный принцип организации данных в ГИС фокусирует внимание не столько на общих свойствах объектов (моделируемых через деление на слои в предыдущем подходе), сколько на их положении в какой-либо сложной иерархической схеме классификации и на взаимоотношениях между объектами. В силу этого удобно отображаются различные родственные и генетические отношения между объектами, отношения соподчиненности, функциональные связи между объектами.

Такой подход ближе к структуре человеческого мышления. Он эффективен, когда необходимо использование логических взаимосвязей объектов, но мало полезен при непрерывном распределении в пространстве признаков (рельеф, удельное содержание полезного ископаемого, загрязнение почвы тяжелыми металлами).

### 3.3. Виды моделей организации данных

Принципы организации данных определяют соответствующие модели организации данных:

- 1) геореляционная модель организации данных;
- 2) объектно-ориентированная модель организации данных;
- 3) объектно-реляционная модель организации данных.

Широкое распространение получили модели данных ESRI: а) геореляционная модель организации данных в виде моделей данных «Шейпфайл» и «Покрытие», б) объектно-ориентированная модель организации данных в виде модели данных «База геоданных». На рисунке 3.2 представлена иерархия моделей данных ESRI от общего верхнего уровня моделей географических объектов до нижнего специального уровня организации данных.



Рис. 3.2. Модели данных ESRI

На рисунке 3.2 модель данных «База геоданных» представлена как ветвь векторной объектно-ориентированной модели, которая в ней играет ведущую роль. Фактически база геоданных содержит модели данных «Шейпфайл» и «Покрытие», растровые и триангуляционные модели данных.

### 3.4. Геореляционная модель данных

#### 3.4.1. Сущность геореляционной модели данных

В базе данных, спроектированной как реляционная модель данных, данные хранятся как наборы таблиц (называемых отношениями), которые логически ассоциированы друг с другом с помощью общих атрибутов. Отдельные записи хранятся как строки таблиц, в то время как атрибуты хранятся в виде колонок. Каждая колонка может содержать атрибутивные данные только одного типа: дату, текстовую строку, числовые данные и т.п. Таблицы обычно стандартизуются для минимизации дублирования.

ГИС содержит два типа данных – пространственные и семантические.

Пространственные данные географических объектов хранятся в отдельных таблицах пространственных данных в виде последовательности координатных пар X, Y.

Атрибутивные данные географических объектов организуются в таблицы атрибутивных данных. Число записей в таблицах атрибутов равно числу графических объектов в двоичных файлах.

Отношения между географическими объектами делаются явными с помощью топологии, которая также представляется соответствующими таблицами.

Сущность этой модели заключается в раздельном хранении значений координат и атрибутивных данных. Она основана на геометрическом типе объекта и отображает мир в виде наборов точек, линий и полигонов. Координаты каждого объекта с уникальным идентификационным номером, хранятся в двоичных файлах. Атрибутивные значения и описание топологии хранятся в таблицах реляционной СУБД (рис. 3.3).

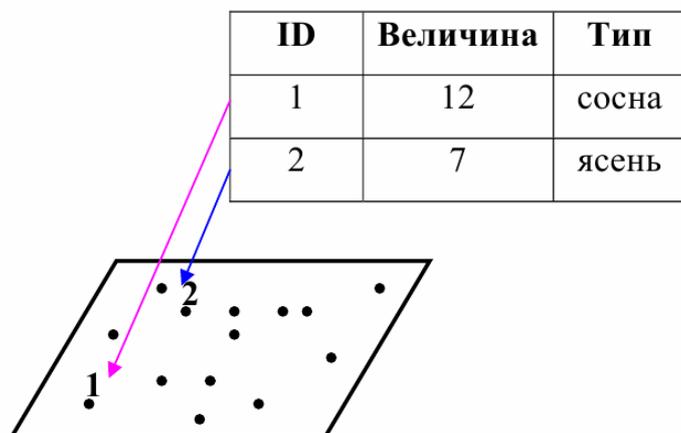


Рис. 3.3. Принцип геореляционной модели

Записи связаны с геометрией посредством идентификационного номера объекта (Identifier – ID). Модель географических данных представляет географические объекты как набор взаимосвязанных пространственных и атрибутивных данных. При этом ГИС осуществляет совместное согласованное управление целостной информацией объектов, распределяемой между файловой системой и базой данных.

Таким образом, геореляционная модель данных определяется следующими условиями:

1) между записями в таблицах пространственных данных, которые отображают модели географических объектов (точками, линиями, полигонами), и записями в таблице атрибутов устанавливается отношение «один-к-одному»;

2) связь между географическим объектом и записью в таблице атрибутов поддерживается через единственный уникальный номер – идентификатор объекта;

3) идентификатор хранится в двух местах: в файлах географических объектов, содержащих пары координат X, Y, и в соответствующих записях таблицы атрибутов географических объектов.

### **3.4.2. Модель данных «Шейпфайл»**

Модель данных «Шейпфайл» представляется цифровым форматом Shapefile.

Формат Шейпфайл (Shapefile) – это цифровой векторный формат ESRI для хранения пространственной и связанной семантической / атрибутивной информации о географических объектах.

Этот формат не приспособлен для хранения топологической информации. Формат Shapefiles создан для ArcView GIS; он может использоваться в ARC/INFO, ArcGIS.

Формат Shapefile содержит набор файлов с одинаковым названием, но разным расширением. Эти файлы делятся на обязательные и факультативные (дополнительные). Обязательными файлами являются три файла с расширением .shp, .shx, .dbf, так как они содержат базовые данные:

- Файл формы с расширением .shp (shape file) – это главный файл, который хранит географические объекты в его собственной записи как список координатных пар X, Y.

- Файл индекса формы .shx ускоряет вычерчивание всех пространственных объектов в шейпфайле.

- Файл атрибутов с расширением .dbf (dBASE file) хранит атрибутивную информацию о пространственных объектах в .shp файле как таблицу атрибутов в формате dBASE.

Шейпфайлы являются простыми, поскольку они хранят примитивные геометрические типы данных: точечные, линейные и полигональные. Эти примитивы имеют ограниченное использование без каких-либо признаков для указания того, что они представляют.

Таким образом, таблица записей будет хранить пространственные объекты / атрибуты для каждой примитивной формы в шейпфайле. Формы (точечные, линейные, полигональные), а также данные атрибутов могут создавать бесконечное множество представлений о географических данных. Представление предоставляет возможности для мощного и точного вычисления.

### **3.4.3. Модель данных «Покрытие»**

Покрытие (Coverage) – это геореляционная модель, которая имеет векторный топологический формат данных.

Покрытие содержит пространственные и атрибутивные данные географических объектов. Оно использует набор следующих классов пространственных объектов для представления географических объектов:

- Точка (Point) – используется для представления точечных пространственных объектов или пользовательских идентификаторов ID-полигонов. Точка определяется координатной парой X, Y.
- Дуга (Arc) – используется для представления линейных пространственных объектов или границ полигонов. Дуга определяется последовательностью координатных пар X, Y начального узла, промежуточных вершин, конечного узла. Дуги топологически связываются через их конечные точки (узлы). Один линейный объект может быть образован многими дугами.
- Узел (Node) – представляет конечные точки дуг или пересечение линейных объектов. Узел имеет уникальный идентификатор. Узел может быть топологически связан с набором дуг, которые соединены одна с другой.
- Путь (Route) – линейный пространственный объект, составляющий одну или несколько дуг или части дуг.
- Секция (Section) – дуга или часть дуги, которая используется для определения пути или создания путевых блоков.
- Полигон (Polygon) – представляет площадные объекты. Полигоны топологически определяются серией дуг, которые формируют их границы, включая дуги, определяющие острова внутри. Пользовательские идентификаторы ID-полигонов представляются точками внутри границ.
- Регион (Region) – совокупность полигонов, представляющих географический объект.

- Аннотация (Annotation) – текст, используемый для обозначения объектов. Аннотации не имеют топологических связей с другими объектами, не используются для аналитических целей.

- Реперная точка (Tic) – регистрационная точка, определяющая положение известной точки на земной поверхности, для которой известны координаты реального земного пространства. Реперные точки позволяют регистрировать и трансформировать координаты покрытия. Рекомендуемое количество реперных точек – 4 и более.

- Охват покрытия (Coverage Extent) – минимальный прямоугольник, ограничивающий покрытие, который представляет территориальный охват покрытия. Охват покрытия определяется предельными координатами  $X_{max}$ ,  $X_{min}$ ,  $Y_{max}$ ,  $Y_{min}$  его элементов (рис. 3.4).

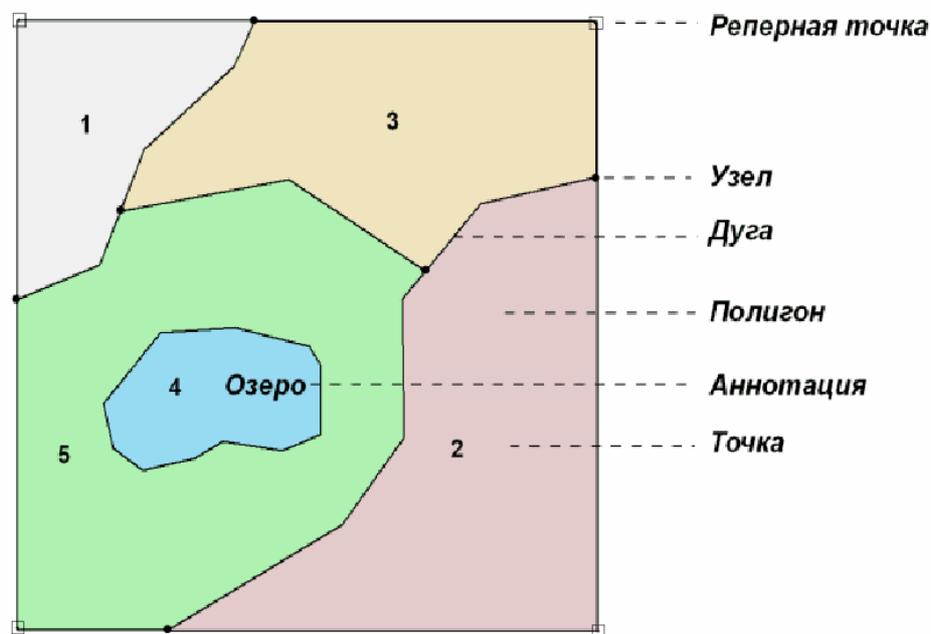


Рис. 3.4. Классы пространственных объектов в покрытии

Описательные данные для классов пространственных объектов хранятся в соответствующих таблицах атрибутов. Связывание пространственных объектов и атрибутов обеспечивается следующими базовыми положениями:

- пространственные объекты в покрытии существуют в отношении один-к-одному с соответствующими записями в таблице атрибутов пространственных объектов;

- ArcGIS поддерживает связь между пространственными объектами и атрибутами посредством уникального идентификатора, назначенного каждому объекту;

- порядковый номер пространственного объекта физически хранится в двух местах покрытия: в файлах, содержащих пространственные данные для каждого пространственного объекта (координатные пары) и с соответствующей записью в таблице атрибутов пространственных объектов. ArcGIS автоматически создает и поддерживает эти связи.

#### **3.4.4. Преимущества и недостатки геореляционной модели данных**

Геореляционная модель имеет следующие преимущества:

- простая структура таблиц, которые легко читать;
- интуитивный, простой пользовательский интерфейс;
- наличие множества инструментов для конечных пользователей (например, макросов и скриптов);
- простота изменения и добавления новых привязок, данных и записей;
- простота использования таблиц, описывающих географические элементы с общими атрибутами;
- возможность привязки таблиц, описывающих топологию, необходимую для ГИС-анализа;
- прямой доступ к данным, обеспечивающий их быструю и эффективную обработку;
- независимость данных от приложения;
- наличие больших объемов ГИС-данных в этом формате.

Геореляционная модель имеет следующие недостатки:

- ограниченное представление реального мира;
- ограниченную гибкость управления запросами и данными;
- медленный последовательный доступ;
- трудность моделирования сложных отношений данных, поскольку для этого часто необходимы квалифицированные прикладные программисты баз данных;
- необходимость выражения сложных отношений в виде процедур в каждой программе, которая обращается к базе данных.

## **ТЕМА 4. СБОР И ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ГЕОГРАФИЧЕСКИХ ДАННЫХ**

Основу информационного обеспечения ГИС составляют источники пространственных данных. Информационное обеспечение ГИС остается крайне трудоемким делом. Затраты на информационное обеспечение геоинформационных проектов достигают 90% от их общей стоимости.

Об объеме пространственных данных говорят следующие результаты специально проведенных исследований: до 70% всех данных, составляющих информационные ресурсы наций, регионов и ведомств, имеют пространственную привязку или могут быть координированы, получив статус пространственных.

В современном обществе геоинформация рассматривается как национальный ресурс. Она является также товаром, который создают и покупают. Геоинформация служит основой управления в тех организациях, где нужна актуальная информация, связанная с пространственными структурами. Она сочетает в себе все свойства информации и информационных продуктов, в этом ее уникальность и ценность.

Сбор информации решает две задачи: получение первичной геоинформации и частичное преобразование ее в геоданные.

### **4.1. Источники географических данных**

Наиболее распространенные источники пространственных данных следующие:

- аналоговые топографические карты и планы;
- цифровые топографические карты и планы;
- данные топографических съемок, измерений электронными тахеометрами и приемниками глобальной системы позиционирования;
- материалы дистанционного зондирования Земли;
- базовые наборы геопро пространственных данных;
- интернет-источники цифровых пространственных данных;
- общегеографические и тематические карты;
- землеустроительная документация;
- градостроительная документация;

Наиболее распространенные источники семантических данных – таблицы и текстовые документы.

#### **4.1.1. Топографические карты и планы аналоговые**

Топографическая карта представляет собой уменьшенное и обобщенное изображение земной поверхности, созданное на единых математической основе и оформлении, передающее размещение и свойства основных природных и социально-экономических объектов местности. Топографические карты и планы, которые традиционно создавались на бумажной основе, представляют собой аналоговые модели. Топографические карты являются первичными данными, источником данных для производных карт, например, тематических.

Недостатками карт выступает то, что они ограничены в двумерных статических представлениях и всегда отображаются в определенном масштабе. Карта всегда является графическим представлением на определенном уровне детализации. На карте масштаб определяет пространственное разрешение графического представления. Чем меньше масштаб, тем меньше деталей карта может показать. С другой стороны, точность представления данных ограничивает диапазон соответствующих масштабов. Выбор надлежащего масштаба карты является одним из первых и наиболее важных шагов. Карты больших физических размеров должны быть разрезаны на листы для удобства пользования. Недостатком карт является также то, что они ограничены в отображении семантики объектов из-за ограниченного места на листе карты.

Для перевода аналоговых данных в цифровую форму выполняют оцифровку (цифрование – digitizing) аналоговых топографических карт. Для этого современные ГИС-пакеты имеют встроенные инструменты. Различают оцифровку а) дигитайзером и б) по дисплею (интерактивную или автоматизированную).

#### **4.1.2. Топографические карты и планы цифровые**

Цифровые топографические карты создаются путем:

- 1) оцифровки аналоговых топографических карт;
- 2) обработки материалов топографических съемок, геодезических измерений электронными тахеометрами и приемниками глобальной системы позиционирования;
- 3) обработки материалов дистанционного зондирования Земли (аэрофотоснимков, космических снимков);
- 4) комбинированным способом.

#### **4.1.3. Данные топографических съемок, измерений электронными тахеометрами и приемниками глобальной системы позиционирования**

В базу данных ГИС теперь можно загружать трехмерные данные любой инструментальной съемки, измерений электронными тахеометрами, GPS-приемниками и выполнять в ней все традиционные вычисления, необходимые для корректировки этих данных и создания координатных точек с известными уровнями ошибки. Для получения наилучшего значения точки можно произвести корректировку разрозненных данных методом наименьших квадратов.

Измерения координатной геометрии (COGO) могут быть также объединены с данными съемки, их можно обрабатывать одним и тем же способом. В базу данных геосъемки также можно включить координаты, полученные GPS-приемниками. Обработка данных съемки и вычисления выполняются в базе геоданных в том же координатном пространстве, что и обработка других данных.

Процесс передачи данных от полевых станций, обработки данных инструментальной съемки, последующей передачи обработанных данных в системы COGO и САПР для отрисовки и структуризации и затем передачи в ГИС для интеграции с другими данными сегодня может быть реализован в одной базе геоданных в едином координатном пространстве.

Новая возможность, существенная с точки зрения обеспечения точности базы данных ГИС, – интеграция данных геодезической съемки с местоположением ГИС-объектов на карте. Может быть установлена привязка между координатами, полученными из данных съемки, и точками на объектах. После этого объекты можно переместить в правильное положение и в таком виде сохранить их в базе данных. Можно выбрать допуски замыкания, алгоритмы конфигурирования и пакетную обработку по уравниванию. Можно добавить целиком новые ГИС-объекты, определенные по данным геодезических измерений, а также повысить точность существующих ГИС-данных. Ошибки местоположений новых объектов могут быть представлены эллипсами погрешностей. Теперь можно провести количественное сравнение допусков относительной и абсолютной ошибки, указанных в описаниях информационных продуктов, с точностью размещения объектов.

#### **4.1.4. Материалы дистанционного зондирования Земли**

Дистанционное зондирование Земли (ДЗЗ) — наблюдение поверхности Земли авиационными и космическими средствами, оснащенными различными видами съемочной аппаратуры. Рабочий диапазон съемочной аппаратуры составляет от долей микрометра (видимое оптическое излучение) до метров

(радиоволны). Методы зондирования могут быть пассивные, то есть использующие естественное отраженное или вторичное тепловое излучение объектов на поверхности Земли, обусловленное солнечной активностью, и активные, использующие вынужденное излучение объектов, инициированное искусственным источником направленного действия.

Материалы дистанционного зондирования Земли, являются одним из основных источников данных для геоинформационных систем. Это, прежде всего материалы аэрофотосъемки и космической съемки.

### **Аэрофотоснимки**

Аэрофотосъемка выполняется в основном для топографического картографирования территории, также применяются в геологии, в лесном и сельском хозяйстве.

Современные методы обработки материалов аэрофотосъемки – цифровая фотограмметрическая обработка, позволяющая определять геометрические, количественные и другие свойства объектов на поверхности земли по фотографическим изображениям, получаемым с помощью летательных аппаратов любых видов. В настоящее время обработка полученных изображений ведется с помощью компьютерной техники и программ для фотограмметрической обработки изображений. При этом дополнительно выполняются коррекции перспективы, дисторсии и иных оптических искажений, цветокоррекция полученных снимков.

Фотографическое изображение преобразовывается в цифровую форму путем сканирования. При этом изображение делится на определенное количество пикселей равных площадей. Каждый такой площадной объект содержит достаточную информацию в отношении и плотности цвета. В цифровой фотограмметрии точность получения результатов возрастает с увеличением разрешения сканирования. Чем меньше размер пиксела, тем точнее результат.

Широко используемым продуктом цифровой обработки являются ортофотопланы. Ортофотоплан – это фотографический план местности на точной геодезической опоре, полученный путем аэрофотосъемки с последующим преобразованием аэроснимков (из центральной проекции в ортогональную) на основе эффективного метода их дифференциального ортофототрансформирования.

### **Космические снимки**

Сбор данных для ГИС на базе космических снимков высокого разрешения становится все более популярным благодаря увеличившемуся числу спутников и архивов данных.

Космические съемки выполняют с использованием либо фотографических, либо сканерных систем. Космическая съемка осуществляется специальными приборами – датчиками. Датчики могут быть пассивными и активными, причем пассивные датчики улавливают отраженное или испускаемое естественное излучение, а активные способны сами излучать необходимый сигнал и фиксировать его отражение от объекта. Наряду с этим по-прежнему используются космические системы с панхроматическим фотооборудованием и многоспектральными фотокамерами, обеспечивающими высокое разрешение и геометрическую точность.

#### **4.1.5. Интернет-источники цифровых пространственных данных**

Интернет является неоценимым источником поиска и получения цифровых данных. В Интернете можно найти множество пространственных и табличных данных, а также снимков. Хорошими адресами для начального поиска данных являются, например, <http://earth.google.com/>, <http://www.geographynetwork.com/data/> или <http://www.scanex.ru/ru/data/>. В сети Интернет доступ к пространственной информации предоставляют картографические и геоинформационные Web-сервисы.

При определении того, стоит ли приобретать цифровые данные, необходимо знать о происхождении, пространственном охвате, надежности данных, масштабе данных, разрешении, точности и формате данных, актуальности и обновляемости, условиях и стоимости получения, приобретения. Для этого следует получить метаданные (данные о данных) в форме словаря данных или отчета о качестве данных от их поставщика. Метаданные должны содержать соответствующую общую информацию о данных. Развитое программное обеспечение ГИС содержат инструменты формализованного описания метаданных.

#### **4.1.6. Общегеографические и тематические карты**

Общегеографические карты масштаба от 1:200000 до 1:1000000 и мельче карты содержат разнообразные сведения о рельефе, гидрографии, почвенно-растительном покрове, населенных пунктах, хозяйственных объектах, путях сообщения, линиях коммуникаций, границах.

Тематические карты создаются по определенной предметной области. Среди них выделяют карты природы, населения, экономики и др.

#### **4.1.7. Документация землеустройства**

Документация землеустройства – это утвержденные в установленном порядке текстовые и графические материалы, которыми регулируется ис-

пользование и охрана земель государственной, коммунальной и частной собственности, а также материалы обследования и разведки земли, авторского надзора, за выполнением проектов и тому подобное. Документация землеустройства разрабатывается в виде программы, схемы, проектов, специальной тематической карты, атласов, технической документации.

К технической документации относятся кадастровые планы, индексные карты. Кадастровые планы содержат границы земельных участков (недвижимой собственности), которые не отображаются на топографических планах. Идентификация земельных участков выполняется путем присвоения кадастровых кодов (номеров). Для уникальной идентификации земельных участков создаются индексные карты.

#### **4.1.8. Градостроительная документация**

Градостроительная документация – это утвержденные текстовые и графические материалы, которыми регулируется планирование, застройка и другое градостроительное использование территории. Документация обязательно состоит из текстовых и графических материалов.

#### **4.1.9. Таблицы и текстовые документы**

Источниками семантических данных могут быть таблицы. Информационный продукт не всегда представляет собой карту. Он также может включать список цифр, таблицу или отчет. Таблица или список могут также составлять целый информационный продукт. Все эти элементы входят в категорию так называемых табличных данных. Электронные таблицы, базы данных и таблицы текстовых файлов являются частью широкой категории табличных данных. Каждая таблица должна иметь название, соответствующие заголовки колонок, типичные записи и указание на источник этих данных. Если существуют установленные форматы отчетов, обязательно определите их на данном этапе. Обычно отчеты можно автоматически конвертировать в нужные форматы.

К текстовым документам предметной или тематической информации относятся технические, экономические, статистические, социологические, демографические и другие виды данных; данные государственной статистической отчетности. Текстовая информация более повествовательного характера часто является важной составляющей информационного продукта. Текстовые файлы могут быть представлены в форматах Adobe.PDF, Microsoft Word или в формате .txt.

## **4.2. Характеристики данных**

Географические данные имеют множество характеристик. Существенными для сбора данных являются масштаб, разрешение, картографическая проекция, допуск ошибки.

### **4.2.1. Масштаб**

Масштаб данных отражает детальность (определенный объем информации) и относительную точность отображения объектов на карте. Чем больше масштаб, тем более точен и подробен набор данных.

В некоторых случаях может потребоваться хранение данных в базе в нескольких масштабах. Если в базе данных представлено больше одного масштаба, это должно быть соответствующим образом документировано в метаданных.

Масштаб влияет как на стоимость, так и на точность конечной базы данных.

### **4.2.2. Разрешение**

Пространственное разрешение определяется как размер самых мелких элементов, поддающихся нанесению на карту или выборке в данном масштабе. Разрешение карты напрямую связано с ее масштабом. По мере уменьшения масштаба карты разрешение уменьшается; границы элементов должны быть сглажены, упрощены или просто не должны быть отображены. Существуют минимальные размеры полигона и длина линии, которые можно отобразить в данном масштабе. Элементы с более низкими разрешениями сливаются с окружающими данными, преобразуются в точку или удаляются.

Данные должны иметь минимально достаточное разрешение, чтобы на их основе можно было создавать информационные продукты.

Например, данные о земельных участках в городе должны иметь высокое разрешение, а Web-приложение, отображающее транспортные маршруты между штатами, должно быть мелкомасштабным.

Разрешение также влияет на ошибку данных.

### **4.2.3. Картографическая проекция**

Картографическая проекция преобразует трехмерную поверхность Земли в плоскую двумерную поверхность. Этот процесс создает картографические искажения, относящиеся к расстоянию, площади, форме геометрической фигуры или направлению. В результате, все плоские карты имеют некоторую степень пространственного искажения. Тип используемой проекции определяет степень и тип искажения на карте. Конкретная картографическая проекция может сохранить неискаженным одно свойство за счет других или, в качестве компромисса, отобразить несколько свойств с

уменьшенной точностью. Необходимо знать искажение, которое нанесет наименьший ущерб качеству вашей базы данных.

Даты (Datum) – важный аспект, связанный с проекцией. Даты – это набор базовых параметров и контрольных точек, используемых для точного задания трехмерной формы Земли. Даты задают начало отсчета и ориентацию для линий широты и долготы. Даты дают базовую ссылку для определения местоположений на поверхности Земли. Для эффективного использования пространственных данных, полученных с карты, нужно знать проекцию этой карты и даты.

#### **4.2.4. Допуск ошибки**

Допуск ошибки определяет, какой размер ошибки допустим, а какой – недопустим. Поскольку ошибка связана с разрешением и масштабом, она также напрямую привязана к стоимости: уменьшение процента ошибки стоит денег.

Существует четыре типа ошибок.

1) ссылочная ошибка относится к ошибке в определении или привязке ссылок. Например, правильно ли указаны соответствующие названия улиц и номера домов;

2) топологическая ошибка происходит при разрыве в необходимых привязках, например, если полигоны не замкнуты или уличные сети разорваны;

3) относительная ошибка — это ошибка в расположении двух объектов по отношению друг к другу. Например, важно знать местоположение коллектора по отношению к владению и дороге;

4) абсолютная ошибка — это ошибка в определении реального положения какого-либо объекта в мире.

### **4.3. Предварительная обработка исходных данных**

#### **4.3.1. Назначение**

Множество данных геоинформационных систем собирается с помощью различных технических средств, технологий и в разных условиях. Данные отражают различные характеристики и свойства.

Они могут иметь различные размерности, разное количество значащих цифр, точность и т.д. Для того чтобы разнородные данные и модели можно было обрабатывать совместно, они должны быть предварительно упорядочены и представлены в форме, воспринимаемой единой системой данных.

Каждый вид данных в зависимости от степени или параметров соответствия системным требованиям может требовать выполнения одной или нескольких процедур обработки, или не требовать таковых.

Наибольшее применение находят следующие виды процедур предварительной обработки исходных данных: первичная обработка, локализация геогра-

фических объектов, оцифровка, трансформирование данных, конвертирование данных, унификация, классификация, идентификация, стратификация и другие.

#### **4.3.2. Первичная обработка**

При первичной обработке используются общие и специальные средства подготовки данных.

##### **Общие средства обработки**

Общими могут быть следующие средства подготовки данных.

**Фильтрация данных.** Загруженные (или предварительно обработанные) данные могут быть избыточными по количеству записей или содержанию отдельных полей с точки зрения задач пользователя. Для удаления избыточных по ряду выделенных признаков записей может быть использована процедура фильтрации данных.

**Нормализация данных.** Использование отдельных инструментов требует предварительной нормализации данных, то есть приведения интервалов значений отдельных множеств к интервалам с предопределенными границами. Ряд продуктов имеют инструменты, которые реализуют процедуру нормализации.

**Удаление противоречий.** В анализируемом наборе данных можно выделить входные и выходные поля. Значения входных полей полностью определяют значения выходных полей. Противоречия возникают, если есть группы записей, в которых значения во входных полях полностью совпадают, а в выходных – различаются.

**Удаление дубликатов.** В данных могут встречаться записи с одинаковыми входными факторами и одинаковыми выходными, т.е. дубликаты. Такие данные несут избыточность. Удаление дубликатов осуществляется с помощью специальной процедуры.

**Разделение.** Процедура разделения позволяет разделить исходный набор данных на несколько непересекающихся подмножеств. Если необходимо использовать простые принципы разделения данных, процедура разделения эффективнее, чем параллельное применение нескольких процедур.

**Интерполяция.** Если в столбце некоторые данные отсутствуют в силу каких-либо причин, обосновывается один из двух путей работы с данными: а) убрать из обработки все строки, которые содержат пропущенные данные; б) заполнить пропущенные данные, используя процедуру интерполяции.

##### **Специальные средства обработки**

Примеры специальных средств подготовки данных:

- первичная обработка данных ДЗЗ включает радиометрическую калибровку, географическую привязку, геометрическую коррекцию изображений и др.

- первичная обработка GPS данных включает сохранение данных с приемника GPS в файл журнала (Log File), фильтрацию входных данных GPS по времени, расстоянию или отклонению.

#### **4.3.3. Локализация географических объектов**

Установление пространственных отношений между процессами, явлениями и объектами на поверхности Земли является одной из главных задач геоинформационных систем. Эта проблема решается посредством локализации географических объектов.

Пространственной локализацией географических объектов называют процесс установления пространственных отношений между характеристиками геоданных путем соотнесения разных видов информации к одному из наиболее устойчивых ее видов. Этим видом являются координаты в обще-земной или локальной системе координат. По существу, локализация в геоинформатике означает процесс привязки данных к системе координат реального земного пространства.

#### **4.3.4. Оцифровка**

Оцифровка (Digitizing) - синоним цифрование, дигитализация, – есть процесс представления объекта, изображения, документа, звука или сигнала (как правило, аналогового сигнала) посредством дискретного набора точек или элементов. Другими словами, это процесс перевода аналоговых данных в цифровую форму, доступную для существования в цифровой компьютерной среде. В результате оцифровки получается так называемое «цифровое изображение» объекта или «цифровая форма» сигнала. Оцифровка текстов используется для преобразования изображения текстов в цифровую форму с помощью оптического распознавания символов.

Оцифровка имеет огромное значение для вычислительного процесса. Дигитализация является основным способом представления изображений в формате, пригодном для передачи и компьютерной обработки.

Оцифровка происходит в два этапа:

1. Дискретизация – а) разделение на аналоговые сигналы и на регулярные интервалы времени (частоты дискретизации); б) чтение значений сигнала в точке.

2. Квантование – округление до фиксированного набора чисел (например, целых).

В географических информационных системах дигитализация есть инструмент цифрового представления географических объектов в растровой или векторной форме. Дигитализация является инструментом создания электронных карт из различных географических изображений или путем

оцифровки традиционных бумажных аналоговых карт. Различают следующие две технологии оцифровки бумажных карт.

1) оцифровка с помощью графического устройства – дигитайзера – и ручного обвода объектов карты (tablet-based digitizing);

2) оцифровка путем сканирования оригиналов для получения растровых копий и последующей векторизации растровой картографической подложки по дисплею манипулятором «мышь» (on-screen digitizing).

Векторизацией называется также процесс конвертирования растровой графики в векторную графику с использованием программного и компьютерного обеспечения. По степени автоматизации различают ручную (manual), полуавтоматическую (semi-automated) и автоматическую (automatic) дигитализацию. Процессы дигитализации обслуживаются программными средствами, которые называются графическими редакторами.

#### **4.3.5. Трансформация данных**

Трансформация данных (Data transformation) – перенесение данных с одного носителя данных на другой без существенного изменения содержащейся в них информации. Трансформация данных позволяет по заданным наперед правилам изменения структуры данных, строить из одних экземпляров объектов данных другие экземпляры. Чаще всего трансформация используется для предварительной обработки информации без построения конкретной модели.

В геоинформационных системах часто возникает необходимость решения задач:

- преобразования данных между координатными системами (преобразование сдвигом, преобразование сдвигом и поворотом, аффинное преобразование, полиномиальные преобразования, преобразование «резиновый лист»);
- преобразования данных между различными форматами данных (формат данных есть внутренняя структура файла, которая позволяет использовать компьютер).

#### **4.3.6. Унификация**

Разнообразие технологий и методов сбора данных порождает разнообразие типов данных, которые впоследствии необходимо обрабатывать. Обрабатывать множество различных данных неудобно и неэффективно. Для упрощения процесса обработки, хранения и обмена разнородные данные приводят к структурному единому виду, который используется при последующей обработке информации. Такие данные называют унифицированными. Процедура сведения разнородных видов и структур данных к единому виду и структуре называется унификацией. В ходе унификации данных осуществляется построение единой структуры информационной модели.

#### **4.3.7. Классификация**

В процессе организации данных осуществляется их классификация. Она позволяет относить различные модели и их характеристики к разным классам, подклассам и типам, что дает возможность систематизировать исходные наборы данных и использовать свойства классов при анализе конкретных геоинформационных данных. Это одна из важнейших функций организации данных, геоданных. Таким образом, организованные данные – это классифицированные данные.

#### **4.3.8. Идентификация**

Идентификация – это процедура присвоения наборам данных специальных кодов – идентификаторов. Она обеспечивает взаимно однозначное соответствие между данными и идентификаторами, исключает повторение одинаковых идентификаторов, то есть обеспечивает непротиворечивость моделей и их характеристик.

#### **4.3.9. Стратификация**

В ГИС один из общепринятых принципов организации пространственной информации называют послойным принципом. Суть его состоит в том, что данные о территории организуются в виде набора тематических слоев, отвечающих конкретным потребностям и содержащих небольшое число типов однородных данных, объединяемых общей тематикой. Например, слой почв, слой гидрографической сети, слой транспортной сети, слой зданий, слой инженерных сетей и так далее.

Стратификация – это создание слоев, которые могут иметь несколько значений в разных областях применения. Стратификация данных в ГИС – это процедура, которая разбивает систему данных на страты или слои, удобные для обработки и анализа.

Стратификация – это не просто организация данных, а создание инструмента анализа и обобщения данных на разных территориальных уровнях. Это процесс задания некой структуры для эффективной работы геоинформационной системы.

Пространственные объекты характеризуются тем, что имеют графическую форму представления. Графическая форма представления пространственных объектов должна решать две задачи: показывать взаимное расположение объектов и связи между ними (топология); давать возможность количественной оценки геометрических характеристик объектов и их положения в выбранной системе координат (метрика). Графическая форма представления пространственных объектов является в то же время основой для стратификации данных.

## ТЕМА 5.ГЕОПРОСТРАНСТВЕННЫЙ АНАЛИЗ

### 5.1. Определение и задачи геопространственного анализа

Геопространственный анализ – это процесс поиска пространственных закономерностей в распределении географических данных и взаимосвязей между объектами.

Преимущество геоинформационной методологии состоит в том, что ГИС позволяет идентифицировать, поддерживать и управлять пространственными связями между топологическими объектами, представляющими объекты реального мира, создавать новые объекты, связи, связывать новые атрибуты.

Наиболее общие задачи геопространственного анализа следующие:

- a) анализ местоположения объектов – поиск, где размещаются объекты (Mapping where things are);
- b) анализ распределения числовых показателей – выявление, где больше, где меньше (Mapping the most and least);
- c) построение карт плотности – картографирование плотности (Mapping density);
- d) поиск объектов внутри области – поиск того, что внутри (Finding what's inside);
- e) анализ окружения – поиск того, что рядом (Finding what's nearby);
- f) анализ пространственных изменений – картографирование изменений (Mapping change).

### 5.2. Классификация аналитических средств ГИС

Базовые аналитические средства можно сгруппировать в 4 обширные категории:

- 1) функции измерений, выбора данных, классификации;
- 2) оверлейные функции;
- 3) функции окрестности;
- 4) функции связности.

### 5.3. Функции измерений

Функции измерения (Measurement functions), как и выбора данных и классификации, позволяют анализировать данные без выполнения существенных изменений. Они часто используются в начале анализа.

Функции геометрических измерений включают вычисления местоположения, длин линий, расстояний между двумя объектами и площади отдельных объектов.

### **5.3.1. Измерения на векторных данных.**

#### **Определение местоположения**

Свойство «местоположение» пространственных объектов описывается координатами, которые всегда хранятся в базе данных в виде списка координатных пар. В ГИС используются разные инструменты получения координат, в том числе внешне простейшее средство – курсором. Для некоторых аналитических задач представляет интерес определение координат особых точек. Например:

- определение координат точки пересечения двух прямых;
- определения координат центроидов и центров;
- определение длин линий;
- определение площади полигона.

### **5.3.2. Измерения на растровых данных**

Измерения на растровых данных проще из-за регулярности ячейки. Размер площади ячейки постоянен и определяется разрешающей способностью ячейки.

Местоположение отдельной ячейки определяется точкой геопривязки растра, разрешением ячейки, номером колонки и номером ряда ячейки на растре. Точкой геопривязки растра может быть нижний левый или верхний левый угол растра; это условие учитывается программным продуктом.

Расстояние между двумя ячейками растра есть функция местоположения их средних точек и разрешения ячеек. В растровой модели в ГИС определение длин вертикальных или горизонтальных линий проводится путем подсчета количества ячеек, по которым проходит линия, и умножением их на размер одной ячейки. Если линия ориентирована по диагоналям ячеек, то необходимо выполнить произведение количества ячеек на размер ячейки и на  $\sqrt{2}$ .

Площадь выбранных объектов на растре вычисляется как произведение количества ячеек на площадь отдельной ячейки.

## **5.4. Функции выбора данных**

При исследовании пространственного набора данных, прежде всего можно выбрать определенные пространственные объекты, чтобы временно ограничить область исследования, используя функции выбора данных (Retrieval functions). Такие выделения могут быть сделаны на пространственном основании или на основании атрибутивных данных, связанных с

пространственными объектами. Средством выбора данных являются запросы пространственного выбора (Spatial selection queries). Запросы пространственного выбора можно комбинировать или выполнять в некоторой последовательности для получения конечного результата.

#### **5.4.1. Интерактивный пространственный выбор данных**

Интерактивный пространственный выбор данных (Interactive spatial selection) выполняется путем указания на объекте или вычерчивания пространственных объектов на дисплее. После этого выбранные пространственные объекты индицируются на карте. Интерактивно определенные объекты называются селектированными или выбранными объектами. Затем ГИС выбирает пространственные объекты в активных слоях данных, которые накладываются на выбранные объекты.

Пространственные данные связаны с атрибутивными данными. Выбор пространственных объектов посредством этих связей приводит к выбору записей в таблицах. И наоборот, выбор записей в таблицах приводит к выбору пространственных объектов.

#### **5.4.2. Пространственный выбор по атрибутивным условиям**

Пространственные объекты могут быть выбраны путем формирования условий выбора на основе атрибутов пространственных объектов (Spatial selection by attribute conditions). Эти условия формулируются в форме запросов на языке структурированных запросов (Structured Query Language – SQL) (если данные находятся в реляционной базе данных) или на специфическом языке программного обеспечения (если данные находятся в ГИС непосредственно). Выражения запросов могут быть простыми и комбинированными.

Для одного условия типа «где есть пространственные объекты с ...?» создается простое выражение, которое содержит последовательно название атрибута, оператор вычисления – арифметический (+, –, \*, /) или сравнения (=, <>, >, >=, <, <=, Like), значение атрибута. Например, условие «где есть земельные участки с площадью меньше 0,05 га» трансформируется в выражение запроса: **«площадь» < 0,05**.

Когда для выбора используется более одного критерия, создается составное выражение, которое содержит названия атрибутов, операторы вычислений, значения атрибутов и логические операторы (And, Or, Not). Например, запрос «где есть земельные участки с площадью менее 0,05 га и тип использования земель 'жилая застройка'» содержит два выражения с логическим оператором: **«площадь» < 0,05 And «тип» = 'жилая застройка'**.

### **5.4.3. Пространственный выбор на основании топологических отношений**

Пространственный выбор на основании топологических отношений (Spatial selection using topological relationships) может выполняться определенным оператором. Выбор пространственных объектов зависит от их местоположения относительно других пространственных объектов. Наиболее общими являются следующие типы запросов:

- Выбор пространственных объектов, которые находятся внутри объектов (Selecting features that are inside selection objects). Этот тип запроса использует отношение включения (containment) между пространственными объектами. Очевидно, что полигоны могут включать полигоны, линии и точки, а линии могут включать линии и точки; другие включения невозможны.

- Выбор пространственных объектов, которые пересекают (Selecting features that intersect). Необходимый для выбора оператор будет выделять пространственный объект, который геометрически использует совместно общую часть объекта-источника.

- Выбор пространственных объектов, смежных с выбранными объектами (Selecting features adjacent to selection objects). Смежность есть отношение встречи, которое выражает то, что пространственные объекты совместно используют границы. Оно применяется только для линейных или полигональных объектов.

- Выбор пространственных объектов по их удалению (Selecting features based on their distance). В качестве средства выбора пространственных объектов этот тип запроса использует функцию расстояния. Такой выбор может быть поиском внутри определенного расстояния от заданных объектов, на заданном расстоянии или больше заданного расстояния.

## **5.5. Функции классификации**

### **5.5.1. Цели классификации по атрибутам**

Эффективным инструментом пространственного анализа в ГИС являются функции автоматизированной классификации объектов по значениям их атрибутов (Classification functions). Функции классификации по атрибутам – это техника целенаправленного извлечения деталей из множества начальных данных с целью выявления закономерностей в пространственном распределении объектов и их визуализации.

Классификации по атрибутам могут быть:

- простыми, созданными на основе одного критерия (тип ландшафта);
- сложными, созданными на основе многих критериев одного покрытия (высота, количество осадков, экологические показатели и т.п.);

- комбинированными, созданными на основе многих параметров разных покрытий.

Набор начальных данных может быть результатом некоторой классификации; в таком случае говорят о повторной классификации (Reclassification). В процессе классификации входное множество значений атрибутов может быть оставлено неповрежденным.

В классификации векторных данных возможны два результата:

1. Входные объекты могут стать выходными объектами в новом слое данных с дополнительно назначенной категорией. Другими словами, пространственная протяженность оригинальных пространственных объектов не изменилась.

2. Выходные объекты получены в результате объединения соседних пространственных объектов с близкими значениями атрибутов. Такая постобрабатывающая функция называется пространственным слиянием, агрегированием или растворением.

### **5.5.2. Методы автоматизированной классификации по атрибутам**

При автоматизированной классификации (Automatic classification) по атрибутам пользователь определяет атрибут классификации, метод классификации и число классов.

Выбор метода классификации опирается на оценку распределения данных посредством гистограммы. На горизонтальную ось гистограм выносятся значения, а на вертикальную – частота их появления в пределах данной выборки. Для автоматизированной классификации по атрибутам программа ArcMap предоставляет возможность использовать один из шести стандартных методов классификации:

- **Равный интервал.** Метод равных интервалов (Equal interval) основан на следующем принципе классификации: каждый класс имеет равный диапазон значений. Разность между максимальным и минимальным значением одинакова для каждого класса. Программа вычитает минимальное значение в наборе данных из максимального значения. Полученное значение делит на заданное число классов. Затем получает предельное значение для первого класса путем прибавления результата деления к самому меньшему значению выборки. Таким же образом устанавливаются интервалы для остальной части классов. Равные интервалы более просты для понимания, так как диапазон для каждого класса одинаков.

- **Заданный интервал.** Метод заданных интервалов (Defined interval) позволяет пользователю определить интервал, на который будет разделен весь диапазон значений атрибута. В отличие от схемы равных интервалов, где пользователь определяет количество интервалов, здесь необходимо указать значения интервалов. ГИС автоматически определяет количество классов на основании этого интервала.

- **Квантиль.** Метод квантиль (Quantile) создает равное число объектов в каждом классе. ГИС упорядочивает объекты по принципу изменения их атрибута в интервале от максимального до минимального значения и суммирует их количество. Затем делит общее количество объектов на число классов, которые задал пользователь. После этого присваивает первым по порядку объектам значения самого низкого класса, пока этот класс не будет заполнен, затем перемещается к следующему классу, и т.д.

- **Естественная разбивка.** Метод естественной разбивки (Natural breaks) основан на следующем принципе классификации: границы классов определяются там, где имеется резкий перепад между группами значений. По Jenks ГИС автоматически определяет максимальное и минимальное значение для каждого класса, используя математическую процедуру, которая анализирует резкие изменения в данных. Данная процедура выбирает интервалы, которые лучше всего группируют близкие значения, и максимизирует различия между классами.

- **Геометрический интервал.** Метод геометрического интервала (Geometrical interval) основывается на интервалах классов, которые имеют геометрическую серию. Геометрическая серия – это последовательность значений, где каждое последующее значение получается умножением предыдущего значения на геометрический коэффициент. Алгоритм создает эти геометрические интервалы, минимизируя сумму квадратов элементов на класс. Это гарантирует, что диапазон каждого класса имеет приблизительно одно количество значений в каждом классе и что изменения между интервалами довольно последовательные.

- **Стандартное отклонение.** Метод стандартного отклонения (Standard deviation) основан на следующем принципе классификации: каждый класс определен величиной отклонения от среднего по выборке. ГИС сначала находит среднее по выборке, разделив сумму всех значений на общее число объектов. После этого вычисляется среднеквадратическое отклонение путем вычитания среднего из каждого значения и возведения разности в квадрат. Полученные значения суммируются и делятся на число объектов. Из полученного выражения извлекается корень.

## 5.6. Оверлейные функции

### 5.6.1. Определение и общая характеристика оверлейных функций

Оверлей (Overlay) – это совместная обработка наложения двух или более исходных слоев одной географической области, в результате которой создается производный слой с новыми географическими данными как комбинация топологических сегментов исходных географических данных. Оверлей – это мощное средство анализа множества разноименных и разнотипных пространственных объектов.

Существует два основных пути выполнения оверлейных операций – на векторных моделях и на растровых моделях географических объектов. Геоинформационные системы предоставляют возможность использовать также комбинированный путь. Выбор метода зависит, прежде всего, от целей анализа, от того, какие данные уже существуют, от требуемой точности анализа, сложности операций. Оверлейные операции могут приводить к отличающимся результатам. Каждый путь оверлейного анализа имеет свою специфику.

В системе, основанной на векторных моделях, топологические оверлейные операции являются более сложными, чем в системе, основанной на растровых моделях. Так как пространственные данные хранятся как точки, линии и/или полигоны, они требуют относительно сложных геометрических операций, чтобы вывести пересечение полигонов и создать новые узлы и дуги с объединенными значениями атрибутов.

Элементами оверлейных операций являются входной слой, оверлейный слой, выходной слой (рис. 5.1). Наложение пространственных объектов входных слоев позволяет разделять их на топологические сегменты и комбинировать из этих сегментов новые объекты в зависимости от цели анализа.

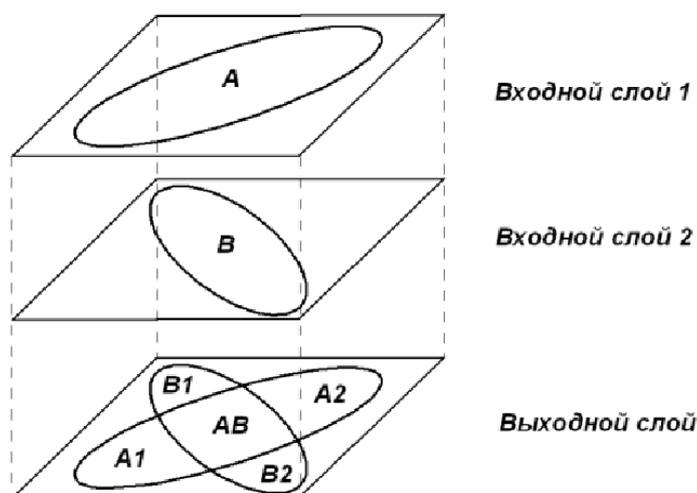


Рис. 5.1. Топологическое наложение векторных слоев

Новые полигоны создаются на пересечении полигонов входного и оверлейного слоев. Наложение линейного объекта на полигональный объект разделяет его на два новых полигональных объекта. Новые объекты хранятся в выходном слое; входной слой не изменяется. Атрибуты пространственных объектов в оверлейном слое присваиваются соответствующим новым пространственным объектам вместе с атрибутами объектов входного слоя.

Во многих случаях требуются манипуляции с более чем двумя слоями векторных данных для достижения цели анализа. Операции выполняются ступенчатым способом: два входных слоя обрабатываются, чтобы сформировать производный слой; этот промежуточный слой затем обрабатывается с третьим слоем, чтобы сформировать следующий промежуточный слой, и так далее до достижения желательного результирующего слоя карты.

В системе, основанной на растровых моделях, топологические оверлейные операции проще, чем в системе, основанной на векторных моделях. Каждая ячейка растрового слоя связана с одним соответствующим географическим местоположением. Это делает ее удобной для комбинирования характеристик многих слоев в одном слое. Обычно каждой характеристике присваиваются многие значения, позволяя пользователю математически комбинировать слои и назначать новые значения каждой ячейке в выходном слое.

### **5.6.2. Булева алгебра в топологическом наложении**

Для реализации топологического наложения в ГИС используется алгебра логики.

В алгебре логики истинностные значения высказываний принято обозначать числами 1 (истина – true) и 0 (ложь – false). Каждой логической операции соответствует функция, принимающая значения 1, 0. Такие функции называются функциями алгебры логики или булевыми функциями.

Чтобы определить, является ли определенное состояние истинным или ложным, в пространственном анализе используются логические операторы Булевой алгебры, которые обозначаются AND (И), OR (ИЛИ), NOT (НЕ) в текстовом формате, и соответственно  $\cap$ ,  $\cup$ ,  $\neg$  в символьном формате.

Два входных слоя топологического наложения можно рассматривать как два набора данных – набор А и набор В. Для них определяются следующие базовые логические операции:

- логическая операция конъюнкция  $A \cap B$  – определяет пересечение двух наборов данных, идентифицирующее те сущности, которые принадлежат и набору А, и набору В (истинно А и В);

• логическая операция дизъюнкция  $A \cup B$  – определяет объединение двух наборов данных, идентифицирующее те сущности, которые принадлежат или набору  $A$ , или набору  $B$  (истинно  $A$  или  $B$ );

• логическая операция отрицание  $A \neg B$  – определяет разность двух наборов данных, идентифицирующая те объекты которые принадлежат  $A$ , но не  $B$  (истинно не  $B$ ).

Эти соотношения можно визуализировать с помощью диаграмм Венна (рис. 5.2).

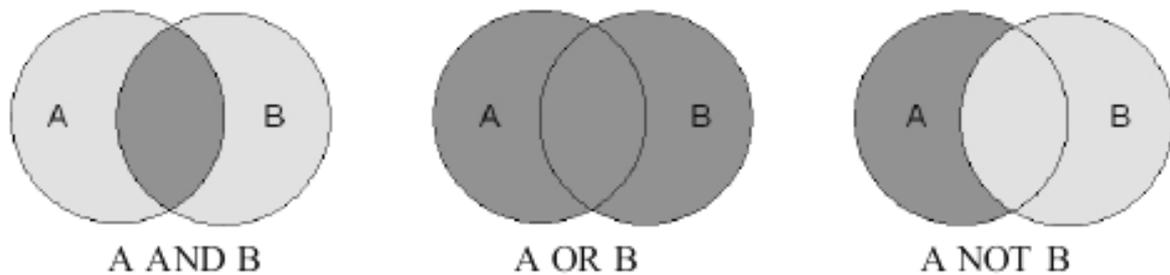


Рис. 5.2. Диаграммы Венна

Используя базовые логические операции можно описать сложные логические функции.

Булева алгебра применяется в вычислении или моделировании новых объектов в топологической оверлейной обработке для систем, основанных на векторных и растровых моделях. Эти операции могут применяться ко всем типам данных – булевым, относительным, интервальным, порядковым или номинальным.

### **5.6.3. Векторные оверлейные операторы.**

#### **Классификация векторных оверлейных операций**

Топологические векторные оверлейные операции можно классифицировать по двум основаниям:

- 1) по элементам, которые содержатся в слоях для наложения (или по слоям, содержащим точечные, линейные или полигональные элементы);
- 2) по типу операции (например, генерирование выходного слоя операциями соединения, пересечения или другими булевыми операциями).

#### **Алгоритмы векторных оверлейных операций**

Из возможных комбинаций наложений точечных, линейных или полигональных пространственных объектов одного слоя на полигональные

пространственные объекты другого слоя основными являются трое следующих алгоритма обработки:

- **«Точка в полигоне» (Point-in-Polygon)**. Операция идентифицирует полигон, в который падает каждая точка. Результатом наложения является множество точек с дополнительными атрибутами полигона, в пределах которого находятся точки.

- **«Линия в полигоне» (Line-in-Polygon)**. Операция идентифицирует полигон, который содержит каждую линию или сегменты линии. Результатом наложения «Линия в полигоне» является новый слой, содержащий линии с дополнительными атрибутами полигонов, в которые падают линии.

- **«Полигон на полигоне» (Polygon-on-Polygon)**, иными словами «Оверлей полигонов» (Polygon Overlay). Операция совместно обрабатывает накладываемые полигоны от двух входных слоев, чтобы создать новые полигоны в выходном слое. Результатом анализа «Полигон на полигоне» является выходной слой, содержащий новые полигоны с атрибутами от каждого из двух оверлейных полигонов.

### **Базовые оверлейные операций векторных моделей**

Векторная оверлейная обработка использует предварительно упомянутые булевы логические операции AND, OR, и NOT, чтобы определить пересечение (Intersection) или объединение (Union) двух слоев, и дополнение (Complement) одного слоя в систему.

Векторная оверлейная обработка также использует булевы отношения включения (Inclusion (обозначение  $\subset$ )) и исключения (Exclusion (обозначение  $\not\subset$ )), чтобы определить, которая часть слоя или содержится, или не содержится в пределах другого слоя в системе.

Например, рассмотрим два слоя A, B в ГИС.

Операция « $A \subset B$ » определяет часть слоя A, которая содержится в слое B.

Операция « $A \not\subset B$ » определяет часть слоя A, которая не содержится в слое B.

Существуют многочисленные векторные оверлейные операции, однако все они являются производными от четырех базовых операций: Union, Intersection, Inclusion, Exclusion.

**Объединение.** Объединение (Union) аналогично булеву логическому оператору OR (ИЛИ), где все элементы от обоих входных слоев будут присутствовать в выходном слое. Этот инструмент строит новый класс пространственных объектов комбинированием пространственных объектов и атрибутов каждого класса пространственных объектов. Объекты входного

слоя разбиваются пересекающими их объектами оверлейного слоя. Атрибуты объектов входного слоя содержат атрибуты объектов входного и оверлейного слоя. Операция объединения двух слоев может быть представлена в булевой алгебраической форме: Слой 1  $\cup$  Слой 2 = Слой 3.

**Пересечение.** Пересечение (Intersection) аналогично булеву логическому оператору AND (И), вычисляет геометрическое пересечение входных и оверлейных объектов. Объекты или части объектов, общие для всех слоев и/или классов объектов, будут записаны в результирующий класс. Объекты входной карты, которые не покрыты объектами оверлейной карты, игнорируются. Атрибуты полигонов выходной карты содержат атрибуты полигонов входной и оверлейной карт. Операция пересечения двух слоев двух полигональных слоев может быть представлена в булевой алгебраической форме: Слой 1  $\cap$  Слой 2 = Слой 3.

**Включение.** Операция включения (Inclusion) определяет часть оверлейного слоя, который содержится во входном слое. Выходной слой будет содержать все элементы первого входного слоя, а также все элементы второго входного слоя, существующие в географическом пространстве первого входного слоя. Включение использует входной слой как шаблон, в котором объекты оверлейного слоя отсекаются по ребрам и объединяются. Операция включения двух слоев двух полигональных слоев может быть представлена в булевой алгебраической форме: (Слой 2  $\subset$  Слой 1)  $\cup$  Слой 1 = Слой 3.

**Исключение.** Операция исключения (Exclusion) «A  $\not\subset$  B» определяет часть входного слоя, которая не содержится в оверлейном слое. Выходной слой будет содержать только те элементы первого входного слоя, которые не являются географическим пространством второго входного слоя. Исключение использует оверлейный слой как шаблон, которым объекты входного слоя отсекаются по его ребрам. Операция исключения двух слоев двух полигональных слоев может быть представлена в булевой алгебраической форме: Слой 1  $\not\subset$  Слой 2 = Слой 3.

#### 5.6.4. Растровые оверлейные операторы

Векторные оверлейные операторы приводят к большей точности результата, но они геометрически сложные, что иногда может быть следствием относительно медленной работы оператора. Оверлейные программы с растрами (Raster overlay operators) не имеют этого недостатка, так как большинство их выполняют вычисления «ячейки на ячейке». Руководящим принципом является сравнение или комбинирование значений характеристик местоположения, получаемых из двух слоев.

Обработка растров выполняется посредством языка представления операций на растрах. Такой язык используется в калькуляторе растров (Raster Calculator). Он позволяет вычислить новые растры на основании существующих, используя ряд функций и операторов.

Программы ГИС имеют широкий диапазон операторов и функций, которые могут использоваться в исчислении растров:

- арифметические операторы;
- операторы сравнения и логические;
- условные выражения;
- комплексные выражения.

## **5.7. Функции окрестности**

### **5.7.1. Определение окрестности**

Под окрестностью (Neighborhood) географического объекта понимается окружающее его пространство. Во многих случаях пригодность некоторого местоположения для определенных целей зависит не только от находящегося в нем, но и от того, что есть возле него. Поэтому, геоинформационные системы должны предоставить возможность оценить пространство, прилегающее к определенному месту, выполнить анализ близости (Proximity).

Функции окрестности (Neighborhood functions) определяют характеристики близости в окрестностях местоположения.

Чтобы выполнить анализ окрестности, необходимо:

- 1) установить, какое целевое местоположение представляет интерес и какая его пространственная протяженность;
- 2) решить, как определить окрестность для каждой цели;
- 3) определить, какие характеристики должны вычисляться для каждой окрестности.

В естественных средах возникают различные случаи расположений, окрестностей и особенностей соседства. Современные программные продукты ГИС позволяют выполнить анализ окрестности как на векторных, так и на растровых моделях.

### **5.7.2. Операции окрестности в векторных моделях**

Вычисления близости используют геометрическую дистанцию для определения окрестности одного или более географических объектов. Наиболее общим и полезным методом является генерирование буферных зон (Buffer Zone Generation). Другим методом, базирующимся на геометрической дистанции, является генерирование полигонов Тиссена (Thiessen Polygon Generation).

### **5.7.3. Операции окрестности в растровых моделях**

Операции окрестности в растровых моделях становятся особенно важным средством, когда ситуация требует в большей мере анализа зависимостей между локализациями, чем интерпретации характеристик в индивидуальных локализациях.

Используются следующие функции:

- функции локальной статистики;
- функции распространения;
- функции распределения.

## **5.8. Функции связности**

Функции связности позволяют выполнять сетевой анализ.

### **5.8.1. Определение и характеристика сети**

В геоинформационных системах сеть (Network) – это система связанных точками пространственных линейных объектов.

Сети представляют географические объекты для транспортирования вещества, энергии, информации. Например, сети дорог – автомобильных, железных; улично-дорожная сеть города; сети маршрутного транспорта, в том числе пассажирского; сети инженерной инфраструктуры города – водоснабжения, водоотведения, теплоснабжения, газоснабжения электро-снабжения; гидрографическая сеть водотоков и т.д. По сетям может транспортироваться почти все: люди, машины, вещи, продукция, загрязнение воды в реках, телефонные сообщения в телефонной сети и пр.

На сетях решаются множество различных транспортных задач на основе сетевого анализа (Network analysis):

- управление инфраструктурой и ее развитием на основе базы данных на все объекты транспортного процесса (анализ транспортных потоков и пассажиропотоков, планирование и анализ маршрутной сети, составление и анализ отчетов по ДТП, реструктурирование маршрутов, поддержка эксплуатации систем энергоснабжения, сигнализации и связи);
- управление парком подвижных средств и логистика (нахождение оптимального маршрута, задача коммивояжера, транспортная задача – полномасштабная организация перевозок различных грузов из многих источников по многим адресам, мультимодальная транспортировка);
- управление движением (слежение за транспортными средствами с помощью GPS, диспетчеризация, увязка расписаний с другими видами транспорта).

Сетевой анализ может выполняться на векторных или растровых данных. Сети на векторных данных имеют следующие особенности:

- В векторных данных линейные пространственные объекты естественно ассоциируются с элементами географических объектов и позволяют обрабатывать типичные транспортные характеристики, такие как пропускная способность и затраты на единицу. Кроме того, сетевой анализ на векторных данных позволяет получить более точные результаты.

- Сеть определяется ее топологически связанными элементами. Каждый линейный пространственный объект (дуга) имеет начальную и конечную точки. Дуги присоединяются одна к другой в узлах, формируя связность (Connectivity) сети. Связность сети делает возможным выполнение сетевого анализа. Связность сети используется также для определения меры сложности сети.

- Важнейшей особенностью любой сети является ее направленность. По направленности сети разделяют на транспортные и инженерные. Транспортные сети – это ненаправленные сети. Это означает, что движение по линии принципиально может быть в прямом и обратном направлении, хотя организация движения может предусматривать одностороннее движение. Транспортные сети моделируют, используя сетевые наборы данных. Инженерные сети – это направленные сети. Потоки вещества перемещаются в одном направлении. Инженерные сети моделируют, используя геометрические сети.

- Сети могут быть одноуровневыми или многоуровневыми. Для многих приложений сетевого анализа используются одноуровневые (плановые) сети, рассматриваемые в двумерном пространстве, например, водные потоки. Плановые сети легче в обработке, так как подчинены топологическим правилам. Многоуровневые сети не рассматриваются как плановые, так как имеют многоуровневые пересечения, например, тоннели, путепроводы, подземные переходы.

Математически сети описываются теорией графов, а решение сетевых задач выполняется средствами линейного программирования. Программными пакетами ГИС поддерживаются различные функции пространственного анализа на сетях. Базовыми классическими функциями являются:

- нахождение лучшего пути, т.е. генерируется путь наименьших затрат на сети между парой определенных мест на основании геометрических и атрибутивных данных;

- разделение сети, т.е. элементам сети (узлам или сегментам) назначаются различные местоположения, используя predetermined критерии.

Продвинутой сетевой анализ использует специальную модель данных – геометрическую сеть. Геометрическая сеть – это коллекция связан-

ных ребер (Edges) и соединений (Junction) с правилами связности, которые используются для представления и моделирования поведения общей сетевой инфраструктуры в реальном мире. Ребра и соединения могут быть простыми и сложными. Примерами ребер являются оси дорог, трубопроводы, средние линии рек. Примерами сложных соединений являются перекрестки, переключатели, трансформаторы, распределительные станции.

Геометрические сети предоставляют расширенные возможности сетевого анализа, моделирования географических объектов для транспортирования, обеспечивают более совершенный уровень управления инженерной инфраструктурой.

### **5.8.2. Нахождение лучшего пути**

Нахождение лучшего маршрута (Finding the best route) или нахождение оптимального пути (Optimal path finding) используется, когда требуется найти путь наименьших затрат между двумя узлами в сети. Решение задачи обеспечивает алгоритм Дейкстры (E. Dijkstra).

Лучший маршрут может быть наискорейшим, кратчайшим или определенным в соответствии с заданными затратами. Цель заключается в нахождении последовательности связанных линий от узла-источника до узла-назначения.

Проблемы, связанные с поиском оптимального пути, могут иметь добавочное требование: дополнительные узлы, которые должны посещаться в пути. Это требование может иметь упорядоченный и неупорядоченный характер. В упорядоченном поиске оптимального пути определена последовательность, в которой эти добавочные узлы должны посещаться; в неупорядоченном поиске оптимального пути не определена последовательность посещения дополнительных пунктов.

Функция затрат может быть простой: например, она может определяться как общая длина всех линий на пути. Функция затрат может быть более комплексной, например, аргументами которой будут не только длины линий, но и пропускная способность, максимальный тариф перевозки и другие особенности. Возможен также детальный учет затрат на повороты в каждом узле – при входе на узел по одной линии и выходе с узла по другой, а также запрещенных (разрешенных) направлений движения.

### **5.8.3. Разделение сети**

Разделение сети (Network partitioning) имеет целью назначение линиям и/или узлам сети некоторого количества целевых местоположений

взаимоисключающим способом. К этой группе относятся функции распределения сети и трассирования.

### **Распределение сети**

Обычно целевые расположения играют роль центра обслуживания для сети. Это может быть любой вид сервиса: поликлиническое обслуживание, школьное образование, снабжение водой. Тип разделения сети относительно центров обслуживания известен как «Распределение сети».

Пусть имеется ряд целевых местоположений, которые функционируют как ресурсные центры. При распределении сети (Network allocation) проблема заключается в том, чтобы разделить сеть на части, которым следует назначить определенный центр обслуживания исключительно. В простой проблеме распределения сервисный центр назначается тем линиям или сегментам, по которым он самый близкий или находится в пределах заданного расстояния. В сложной проблеме распределения возникает необходимость учета дополнительных факторов, например: а) мощность центра (количество посещений поликлиники, количество школьников, киловатты); б) потребление ресурсов, которое может изменяться между линиями или сегментами линии. Следует также иметь в виду, что некоторые улицы отличаются количеством происшествий, проживающих там детей, объектов промышленности с высоким потреблением электричества, смежных земельных участков.

Сеть обслуживания (Service net) любого центра – это подмножество дистрибьюторской сети, фактически связанная часть сети. Существуют различные методы отнесения сети линий или их сегментов к определенному центру. В ArcGIS 9.x эта задача решается функцией «Область обслуживания» (Service area).

### **Трассирование**

Сетевой анализ включает трассирование (Tracing). Функция «Трассирование» по геометрической сети создает связанные наборы элементов сети, в соответствии с некоторым условием.

При помощи инструментов трассирования можно найти:

- все элементы сети, которые лежат вверх или вниз по течению от заданной точки на сети;
- общие затраты всех элементов сети, которые лежат вверх по течению от заданной точки на сети;
- общие пространственные объекты, которые выше по течению от заданной точки;

- все объекты, которые связаны или не связаны с данной точкой через сеть;
- петли, которые могут быть результатом маршрутов между точками на сети.

Трассирование выполняется, когда нужно узнать, какая часть сети условно соединяется с выбранным узлом на сети, известным как начало трассы. Для узла или линии быть условно соединенным означает, что существует путь от узла/линии к началу трассы, и путь соединения выполняет ряд условий. Какими являются эти условия, – зависит от приложения. Они могут включать направление пути, вместимость, длину, потребление ресурсов вдоль пути, и так далее. Условие обычно – это логическое выражение, например:

- путь должен направляться от узла/линии к началу трассы;
- его вместимость (определяемая как минимальная вместимость линий, которые составляют путь) должна быть ниже установленного порога;
- длина пути не должна превышать установленной максимальной длины.

Трассирование может быть полезной функцией для решения многих проблем, связанных с инженерными сетями, например: определение вентилях вверх по течению при авариях в распределительных инженерных сетях.

## ЛИТЕРАТУРА

1. Шипулин, В. Д. Основные принципы геоинформационных систем : учебн. пособие / В. Д. Шипулин ; Харьк. нац. акад. гор. хоз-ва. – Харьков : ХНАГХ, 2010. – 337 с.
2. Демерс, Майкл Н. Географические Информационные Системы. Основы / Майкл Н. Демерс ; пер. с англ. – М. : «Дата+», 1999.
3. Геоинформатика: учеб. для студ. вузов / Е. Г. Капралов [и др.] ; под ред. В. С. Тикунова. – М. : Издат. центр «Академия», 2005. – 480 с.
4. Геоинформатика / А. Д. Иванников [и др.]. – М. : «МАКС Пресс», 2001 – 349 с.
5. Бугаевский, Л. М. Геоинформационные системы : учебн. пособие для вузов / Л. М. Бугаевский, В. Я. Цветков. – М., 2000. – 222 с.

## ОГЛАВЛЕНИЕ

ТЕМА 1. ОБЩИЕ СВЕДЕНИЯ О ГЕОИНФОРМАЦИОННЫХ СИСТЕМАХ .....	3
1.1. Понятие географической информационной системы .....	3
1.2. Классификация информационных систем .....	3
1.3. Наука о ГИС .....	4
1.4. Структура ГИС .....	5
1.5. Отличие ГИС от традиционных систем представления реальности .....	6
ТЕМА 2. ГЕОИНФОРМАЦИОННЫЕ СТРУКТУРЫ ДАННЫХ .....	9
2.1. Пространственные географические объекты .....	9
2.2. Виды компьютерных моделей пространственных объектов .....	12
2.3. Векторные модели географических объектов .....	12
2.4. Растровые модели географических объектов .....	22
ТЕМА 3. ОРГАНИЗАЦИЯ ДАННЫХ В ГЕОИНФОРМАЦИОННЫХ СИСТЕМАХ .....	29
3.1. Определение, значение и задачи системной организации данных .....	29
3.2. Принципы организации данных в ГИС .....	29
3.3. Виды моделей организации данных .....	31
3.4. Геореляционная модель данных .....	32
ТЕМА 4. СБОР И ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ГЕОГРАФИЧЕСКИХ ДАННЫХ .....	37
4.1. Источники географических данных .....	37
4.2. Характеристики данных .....	43
4.3. Предварительная обработка исходных данных .....	44
ТЕМА 5. ГЕОПРОСТРАНСТВЕННЫЙ АНАЛИЗ .....	49
5.1. Определение и задачи геопространственного анализа .....	49
5.2. Классификация аналитических средств в ГИС .....	49
5.3. Функции измерений .....	49
5.4. Функции выбора данных .....	50
5.5. Функции классификации .....	52
5.6. Оверлейные функции .....	55
5.7. Функции окрестности .....	60
5.8. Функции связности .....	61
Литература .....	66

*Учебное издание*

ШНИТКО Сергей Геннадиевич

## ГИС В ГЕОДЕЗИИ

Конспект лекций  
для студентов специальностей 1-56 02 01 «Геодезия»  
и 1-56 02 01 02 «Инженерная геодезия»

Редактор *И. Н. Безборцкая*  
Дизайн обложки *Е. Н. Бурсовой*

---

Подписано в печать 30.12.14. Формат 60×84 1/16. Бумага офсетная. Ризография.  
Усл. печ. л. 3,95. Уч.-изд. л. 3,12. Тираж 30 экз. Заказ 1599.

---

Издатель и полиграфическое исполнение:  
учреждение образования «Полоцкий государственный университет».

Свидетельство о государственной регистрации  
издателя, изготовителя, распространителя печатных изданий  
№ 1/305 от 22.04.2014.

ЛП № 02330/278 от 08.05.2014.

Ул. Блохина, 29, 211440, г. Новополоцк.