

Лекция 3

Организация данных в ГИС. Географические и атрибутивные данные. Источники данных для ГИС. Типы данных, используемых в гидрометеорологических ГИС.

Цель лекции: получить понятие об организации данных в ГИС, типах и источниках данных

Краткое содержание лекции. Применительно к характеризующей нами сфере «данные» можно рассматривать и определять в трех контекстах: вне автоматизированной среды использования, внутри ее и в среде ГИС. В первых двух контекстах под «данными» понимаются либо факты, некие известные вещи (из которых могут быть выведены заключения), либо сведения, подготовленные для компьютерной обработки. Под «данными» в среде ГИС понимаются «объекты о явлениях реального мира; результаты наблюдений и измерений этих объектов. Элемент данных содержит три главные компоненты: атрибутивные сведения, которые описывают сущность (семантику), характеристики, переменные, значения и тому подобные его квалификации; географические сведения, характеризующие его положение в пространстве относительно других данных; временные сведения, описывающие момент или период времени, для которого предоставляются данные». «Данные», выступают как сырье, которое путем обработки можно превратить в информацию, т. е. данные — это как бы строительный элемент в процессе создания информации. Они рассматриваются как объект обработки и основа для получения информации.

Данные - это совокупность сведений, зафиксированных на определенном носителе в форме, пригодной для постоянного хранения, передачи и обработки. Преобразование и обработка данных позволяет получить информацию

В практическое понимание «информации» в настоящее время в основном включаются «процессы обмена разнообразными сведениями между людьми, человеком и автоматом — актуальная информация, процессы взаимодействия объектов неживой природы — потенциальная информация, степень сложности, организованности, упорядоченности той или иной системы». Такое понимание основывается на существовании в современной науке нескольких парадигм, которые с разных сторон стараются объяснить факты и явления информационного порядка.

Информация - это результат преобразования и анализа данных. Отличие информации от данных состоит в том, что данные - это фиксированные сведения о событиях и явлениях, которые хранятся на определенных носителях, а информация появляется в результате обработки данных при решении конкретных задач. Например, в базах данных хранятся различные данные, а по определенному запросу система управления базой данных выдает требуемую информацию. Применительно к ГИС под информацией понимается совокупность сведений, определяющих меру наших знаний об объекте.

Существуют и другие определения информации, например, информация – это сведения об объектах и явлениях окружающей среды, их параметрах, свойствах и состоянии, которые уменьшают имеющуюся о них степень неопределенности, неполноты знаний.

Определений «знания» также много, как и определений «информации». Так, Л. Бриллюен считает, что информация отлична от знания, «для которого у нас нет количественной меры». Д.И.Блюменау приводит 10 (!) определений понятия «знание» различных авторов, и сам определяет информацию как знание, включенное непосредственно в коммуникативный процесс. Последнее предложение необходимо дополнить: включенное субъективно в субъективный процесс, т. е. знания — это интерпретация информации. Однако интерпретация в известном смысле не ограничивается знанием и этот ряд полностью будет, вероятно, выглядеть следующим образом: информация — знание — мысль (гипотеза).

«Знания» в философском их понимании — отражение семантических аспектов окружающей действительности в мозгу человека или даже в технической системе. Отметим

также историческую последовательность привлечения данных, информации и знаний в геоинформатике. Так, вначале появились банки данных, позднее оформились географические информационные системы и, наконец, появились системы, основанные на знаниях, — интеллектуальные системы. Наиболее общее определение: знание – результат познания действительности, получивший подтверждение в практике. Научное знание отличается своей систематичностью, обоснованностью и высокой степенью структуризации.

Знания – это зафиксированная и проверенная практикой обработанная информация, которая использовалась и может многократно использоваться для принятия решений.

Знания – это вид информации, которая хранится в базе знаний и отображает знания специалиста в конкретной предметной области. Знания – это интеллектуальный капитал. Формальные знания могут быть в виде документов (стандартов, нормативов), регламентирующих принятие решений или учебников, инструкций с описанием решения задач.

Неформальные знания – это знания и опыт специалистов в определенной предметной области. Необходимо отметить, что универсальных определений этих понятий (данных, информации, знаний) нет, они трактуются по-разному. Информационные системы можно рассматривать как эффективный инструмент получения знаний.

Принятия решений осуществляются на основе полученной информации и имеющихся знаний. Принятие решений – это выбор наилучшего в некотором смысле варианта решения из множества допустимых на основании имеющейся информации.

Для решения поставленной задачи фиксированные данные обрабатываются на основании имеющихся знаний, далее полученная информация анализируется с помощью имеющихся знаний. На основании анализа, предлагаются все допустимые решения, а в результате выбора принимается одно наилучшее в некотором смысле решение. Результаты решения пополняют знания. В зависимости от сферы использования информация может быть различной: научной, технической, управляющей, экономической и т.д.

Возвращаясь непосредственно к геоинформационным системам, важно подчеркнуть их способность хранить и обрабатывать пространственные, или географические, данные, что и отличает ГИС от иных информационных систем. Распространено мнение, утверждающее тождественность понятий «географические информационные системы» и «пространственные (пространственно-координированные, пространственно распределенные) информационные системы», т.е. слово «географические» в данном контексте имеет смысл не обозначения науки, а характеристики пространственности. При таком подходе нельзя поставить в один ряд с географическими системами геологические, геофизические и другие системы, что также встречается в литературе. Они все являются пространственными, а следовательно, и географическими. Естественно и то, что ГИС объединяет в единую систему пространственную информацию и информацию других типов для решения пространственных задач.

В качестве источников данных для формирования ГИС служат:

- картографические материалы (топографические и общегеографические карты, карты административно-территориального деления, кадастровые планы и др.). Сведения, получаемые с карт, имеют территориальную привязку, поэтому их удобно использовать в качестве базового слоя ГИС. Если нет цифровых карт на исследуемую территорию, тогда графические оригиналы карт преобразуются в цифровой вид.

- данные дистанционного зондирования (ДДЗ) все шире используются для формирования баз данных ГИС. К ДДЗ, прежде всего, относят материалы, получаемые с космических носителей. Для дистанционного зондирования применяют разнообразные технологии получения изображений и передачи их на Землю, носители съемочной аппаратуры (космические аппараты и спутники) размещают на разных орбитах, оснащают разной аппаратурой. Благодаря этому получают снимки, отличающиеся разным уровнем обзорности и детальности отображения объектов природной среды в разных диапазонах

спектра (видимый и ближний инфракрасный, тепловой инфракрасный и радиодиапазон). Все это обуславливает широкий спектр экологических задач, решаемых с применением ДДЗ.

К методам дистанционного зондирования относятся и аэро- и наземные съемки, и другие неконтактные методы, например гидроакустические съемки рельефа морского дна. Материалы таких съемок обеспечивают получение как количественной, так и качественной информации о различных объектах природной среды.

- материалы полевых изысканий территорий, включают данные топографических, инженерно-геодезических изысканий, кадастровой съемки, геодезические измерения природных объектов, выполняемые нивелирами, теодолитами, электронными тахеометрами, GPS приемниками, а также результаты обследования территорий с применением геоботанических и других методов, например, исследования по перемещению животных, анализ почв и др.

- статистические данные содержат данные государственных статистических служб по самым разным отраслям народного хозяйства, а также данные стационарных измерительных постов наблюдений (гидрологические и метеорологические данные, сведения о загрязнении окружающей среды и т. д)).

- литературные данные (справочные издания, книги, монографии и статьи, содержащие разнообразные сведения по отдельным типам географических объектов).

В ГИС редко используется только один вид данных, чаще всего это сочетание разнообразных данных на какую-либо территорию.

Любая ГИС работает с различными типами данных, описывающих изучаемые объекты: пространственными, атрибутивными, тематическими, метрическими и семантическими. Эти данные являются наиболее дорогостоящим элементом системы и занимают значительный объем памяти.

Форматы данных определяют способ хранения информации на жестком диске, а также механизм ее обработки. Модели данных и форматы данных определенным образом взаимосвязаны.

Существует большое количество форматов данных. Можно отметить, что во многих ГИС поддерживаются основные форматы хранения растровых данных (TIFF, JPEG, GIF, BMP, WMF, PCX), а также GeoSpot, GeoTIFF, позволяющие передавать информацию о привязке растрового изображения к реальным географическим координатам, и MrSID - для сжатия информации. Наиболее распространенным среди векторных форматов является - DXF.

Все системы поддерживают обмен пространственной информацией (экспорт и импорт) со многими ГИС и САПР через основные обменные форматы: SHP, E00, GEN (ESRI), VEC (IDRISI), MIF (MapInfo Corp.), DWG, DXF (Autodesk), WMF (Microsoft), DGN (Bentley).

Довольно часто для эффективной реализации одних компьютерных операций предпочитают векторный формат, а для других растровый. Поэтому, в некоторых системах реализуются возможности манипулирования данными в том и в другом формате, и функции преобразования векторного в растровый, и наоборот, растрового в векторный форматы.

Совокупность цифровых данных о пространственных объектах образует множество пространственных данных и составляет содержание баз географических данных, определяет принципы построения информационного обеспечения ГИС.

База данных (БД) – совокупность данных организованных по определенным правилам, устанавливающим общие принципы описания, хранения и манипулирования данными

Выявление географических объектов и явлений и последующий выбор адекватного представления данных о них являются составной частью процесса, именуемого проектированием базы данных (БД). В ГИС пользователь рассматривает реальный мир через призму тематической базы данных. Измерения и выборки, содержащиеся в базе данных, должны как можно полнее и точнее соответствовать предмету исследования и его основным характеристикам. Представление данных должно учитывать типы их возможных

преобразований. К созданию БД ГИС предъявляются высокие требования, связанные с пространственной формой организации и представления данных

В процессе проектирования БД обычно выделяют три основных уровня: концептуальный, логический и физический. Концептуальный уровень не зависит от имеющихся аппаратных и программных средств. Для БД ГИС он связан с концептуальной моделью географических данных и включает: описание и определение рассматриваемых объектов; установление способа представления географических объектов в базе данных; выбор базовых типов пространственных объектов — точки, линии, ареалы, ячейки раstra; решение вопроса о способе представления размерности и взаимосвязей реального мира в БД (например, следует ли показать здание в виде ареала или точки). На концептуальном уровне определяется и содержание базы данных, в свою очередь определяемое сутью явления, характером его пространственного распространения и задачами, для которых создается БД. Здесь следует выделить задачи создания одной или серии карт, комплексного картографирования, создания синтетических карт для многоцелевого и многократного использования. Логический уровень определяется имеющимися программными средствами и практически не зависит от технического обеспечения. Он включает разработку логической структуры элементов базы данных в соответствии с системой управления базами данных (СУБД), используемой в программном обеспечении. Наиболее распространенными логическими структурами — моделями БД и их СУБД — являются иерархическая, сетевая, реляционная и объектно-ориентированные.

Создание БД и обращение к ней (по запросам) осуществляется с помощью системы управления базами данных (СУБД).