

УДК: 004.052.2

## АНАЛИЗ КОНЦЕПТОВ ДЛЯ ПРОЕКТИРОВАНИЯ БАЗЫ ГЕОДАНЫХ ЖЕЛЕЗНОДОРОЖНЫХ ГЕООПИСАНИЙ



Дулин С.К.

д.т.н., профессор, главный научный сотрудник, АО «НИИАС»;  
ИПИ ФИЦ ИУ РАН, E-mail: skdulin@mail.ru, Москва, Россия

аспирант, Московский физико-технический институт (МФТИ),  
E-mail: antr5@mail.ru, Москва, Россия

### Аннотация

В работе рассмотрен подход, позволяющий учитывать как ситуативный характер геоконцептов, так и семантику геообъектов при формировании базы геоданных железнодорожных геоописаний. Исследования последних лет показали, что лучшая поддержка базы геоданных обеспечивается, если концепты развиваются в соответствии со спецификой возникающих транспортных ситуаций. Чтобы сократить необходимость изменения схемы базы геоданных в геоинформационных средах, необходимо использовать структуры и функции ситуативных географических концептов. В работе представлен подход, в котором к концептам привязывается геоинформационный контекст, что допускает извлечение контекстов из базы геоданных и их интерпретацию. Это помогает раскрытию неявных аспектов геоданных, добавлению эмпирических компонентов к геоонтологии и расширению контекста геоданных, представленных в базах геоданных. Предполагается, что при формировании базы геоданных железнодорожных геоописаний могут быть приняты онтологические допущения об особенностях дорог, при этом проблемы совместимости решаются путем формализации исследований и практического опыта открытого обмена результатами исследований.

### Ключевые слова:

Геоданные; соотнесение и преобразование геоописаний; анализ геотекстов; совместный доступ

## ANALYSIS OF GEOGRAPHICAL CONCEPTS FOR DESIGNING A GEODATA DATABASE OF RAILWAY GEODESCRIPTIONS

Dulin S.K.

D.ofSci., Professor, Chief Researcher, JSC «NIIAS»; Federal Research Center «Informatics and Management» of the RAS, E-mail: skdulin@mail.ru, Moscow, Russia

Ryabtsev A.B.

Postgraduate student, Moscow Institute of Physics and Technology (MIPT), E-mail: antr5@mail.ru, Moscow, Russia

### Abstract

The paper considers an approach that allows taking into account both the situational nature of geo concepts and the semantics of geo objects when forming a geodata database of railway geo descriptions. Research in recent years has shown that the best support for a geodata database is provided if concepts are developed in accordance with the specifics of emerging transport situations. To reduce the need to change the geodata schema in geographic information environments, it is necessary to use the structures and functions of situational geographical concepts. The paper presents an approach in which a geoinformation context is linked to concepts, which allows for the extraction of contexts from a geodata database and their interpretation. This helps to uncover the implicit aspects of geodata, add empirical components to geontology, and expand the context of geodata represented in geodata databases. It is assumed that when forming a geodata database of railway geo-descriptions, ontological assumptions about the features of roads can be made, while compatibility problems are solved by formalizing research and practical experience in the open exchange of research results.

### Keywords:

Geodata; correlation and transformation of geo descriptions; analysis of geotexts; shared access

## 1. Введение

Важность определения, какие географические концепты (понятия) существуют и как представить и обработать их в компьютере, мотивировали существенный исследовательский прогресс в геоинформатике. Этот прогресс обусловил появление более полных концептуальных моделей в геоинформационных системах (ГИС) и потребовал расширения представлений, используемых для фиксации знаний предметной области. В частности, геопредставления должны пополняться недостающим знанием, так как в большинстве ГИС необходимо интерпретировать информацию, подразумевающую латентные знания. Общепринятые носители значения – имена и определения, которые мы даем концептам и категориям, фиксируют только необходимую часть значения. Расширенные представления должны стать более богатыми, чтобы уменьшить диссонанс между производимыми программными продуктами и информационными нуждами потребителей. С одной стороны это индуцирует расширения семантики геопредставлений, чем в настоящее время активно занимаются производители, а с другой стороны, это вовлекает потребителей геоинформационных ресурсов в процесс преобразования неявных значений в явные.

Онтологический подход обеспечивает переход к семантически более богатому представлению значений в вычислительных системах. В онтологических структурах большее семантическое значение достигается за счет связывания геоданных с логическими построениями описания и взаимодействия существующих в проблемной области концептов. Значение является контекстно нагруженным в таких структурах. Например, географический регион может быть по-разному концептуализирован с различных точек зрения: как геологический тип, с позиции анализа почвы, возможности создания инфраструктуры, рассмотрения условий проживания и т.д. Контекстное значение может включать связь с определенными ситуациями, например, исследование региона может быть связано с заданными целями, наблюдениями, фиксацией местоположений, событиями, действиями, спецификой проживающего там социума и т.д.

Аспекты влияния этих ситуаций без сомнения останутся неявными, ограничивая создание моделей производителями информационного контента, но все равно с течением времени они индуцируют расширение явно представленных знаний, обогащая контекст геопредставлений. Это расширенное понятие контекста связано с развитием концептов и теорий, обеспечивая дополнение и совершенствование онтологических структур. С точки зрения создания баз геоданных, ситуативный контекст позиционирует концепт как в пределах схемы взаимосвязанных геоданных, так и в пределах сети описанных, связанных и логически определенных концептов. [1]

Соответственно, работа посвящена расширению представления географических понятий

(концептов) за счет учета ситуативного влияния. В связи с этим вводится понятие ситуативной категории, представляющей это расширение для баз геоданных.

## 2. Базы геоданных

И эволюционная, и ситуативная точки зрения на развитие геоонтологии связаны с существенными усилиями по реализации проекта базы геоданных, так как схема базы геоданных, основанной на развивающихся концептах, должна самостоятельно развиваться, приводя к перманентному обслуживанию и сложностям использования. Одно из возможных решений этой проблемы – проектирование схемы базы геоданных, основанное на метаконцептных, таких как онтологии, модели, ситуативные концепты и т.д., а не на зафиксированных концептах предметной области. Схема базы геоданных, разработанная на этом принципе, имеет очевидные преимущества будучи динамическим репозиторием и регистром для компонентов знаний и их отношений и обладает возможностью включать расширенное ситуативное представление концептов. Это предполагает, что, по крайней мере, некоторые геоданные являются динамическими и контекстуализированными, то есть организация метаконцептов для схемы базы геоданных обеспечивает:

1. Обнаружение и развитие: статическое представление геоданных не приспособлено для изучения или обнаружения нового знания и не способствует уточнению интерпретации существующих геоданных. Моделирование открытой системы, ориентированной на геопространственные знания, со статическими ограничениями проблематично, а реализация таких ограничений в статичной схеме базы геоданных неизбежно приведет к непрерывным корректировкам схемы, неточностям и рассогласованности.

2. Общность и спецификацию: концепты и теории, вообще говоря, располагаются в диапазоне от общности до ограниченной применимости к определенной предметной области. Этот диапазон общности предлагает, что чем менее общий концептуальный уровень, тем более вероятна эволюция концептов. Концептуальное изменение на более общих уровнях достаточно редкое явление и может быть расценено как сдвиг парадигмы, тогда как неустойчивость в менее общих концептах может быть оценена как продолжающееся изучение и приобретение знания о некотором регионе. Например, общий концепт «вокзал» может остаться надолго установленным в пределах определенной перспективы, но детализированный концепт «самый большой вокзал г. Москвы» ситуативен и может эволюционировать, изменяясь при реконструкции вокзалов, либо с расширением территории города.

3. Контекстуализацию и использование опыта: концепты становятся все более контекстными и многомерными. Они могут быть ориентированы

на цели и функции и использовать историческое и ситуативное знание. Использование контекстного и исторического опыта могло бы объяснить, например, существование разнообразных описаний и классификаций каждого региона, разработанных специалистами в области наук о Земле, или почему некоторые люди считают расположенный рядом лесной массив лесом, в то время как другие видят его как рощу. При этом разработка принципов концептуального подбора могла заложить основу когнитивной инфраструктуры для использования специфики ситуации и опыта.

Предпосылка, что концептуальная неустойчивость влечет за собой изменение схемы, является практическим результатом сопровождения схемы базы геоданных, основанной на концептах, выбранных из предметной области на определенном уровне абстракции. Из пяти уровней абстракции, рассматриваемых при проектировании базы геоданных, наиболее существенными в рамках данной темы являются следующие три: *эпистемологический* уровень, который содержит правила структуризации концептов и примитивы типа кортежей, отношений, объектов, классов, атрибутов и т.д.; *концептуальный* уровень, который содержит концепты, идентифицированные в предметной области, их свойства, отношения и ограничения; и *лингвистический* уровень, который содержит описания реальных данных и отношений. Эти уровни используются при проектировании базы геоданных нисходящим способом: на первом уровне формируются эпистемологические структуры; на втором уровне выбираются концепты предметной области.

При этом эпистемологические структуры используются тремя способами: как (1) независимая от технологии управления геоданными концептуальная схема, (2) связанная с технологией проектирования логическая схема, и, наконец, как (3) физическая схема, определенная для технологии поддержки экземпляров геоданных системой аппаратных средств/программного обеспечения. И, наконец, третий уровень – уровень геоданных, которые позиционируются в физической системе как описания конкретных объектов. Таким образом, схема, разработанная на основе статических концептов, идентифицированных на концептуальном уровне, неизбежно будет иметь тенденцию к изменению.

В отличие от разработки механизмов версионирования, которые управляют последовательными изменениями схемы, попробуем сконцентрироваться здесь на концептуальном проекте схемы, основанном на концептах самого общего описательного уровня, которые являются, по-видимому, более устойчивыми и не провоцируют изменение схемы. Такие концепты могли бы быть получены на онтологическом уровне, который позволит увеличить значимость эпистемологических или концептуальных элементов, используя их со многими концептуально-логическими системами или с онтологиями верхнего, проблемно-ориентиро-

ванного и прикладного уровней, располагаясь, тем самым, от общего до все более и более конкретного уровня. Однако, хотя общие онтологические концепты могут быть идентифицированы в любой прикладной области, природа открытых систем связана с потребностью моделировать непредсказуемые отношения и свойства, которые не могут быть полностью предопределены, что может привести к сложной сетевой структуре схемы, которую трудно использовать и поддерживать. В действительности вряд ли можно рассчитывать на глобальную регулярность структуры базы геоданных в открытых системах.

Чтобы преодолеть эти ограничения следует использовать примитивы более абстрактного *эпистемологического уровня* как основу для проекта схемы. Можно предположить, что концепты предметных областей инкапсулируют контексты человеческого наблюдения и интерпретации геопространственных данных и тем самым дополняют проект базы геоданных. То есть предлагается разработать технологически независимую *концептуальную* схему для концептов и взаимодействия геоданных, которая могла бы быть логически и физически адаптирована к базам геоданных или к другим приложениям, которые используют концепты онтологических геосистем. [2]

Изменение уровня абстракции для схемы базы геоданных при таком подходе включает как детализацию от концептов до геоданных, так и восходящее абстрагирование от геоданных до концептов. Рассмотрение двунаправленных отношений между геоданными и концептами отличаются от однонаправленного нисходящего геопространственного подхода, который использует пространственные и/или временные конструкции на эпистемологическом или онтологическом уровнях. Можно отметить такое же отличие и от негеографических метапредставлений, в которых связь между данными и концептами является главным образом однонаправленной.

### 3. Атрибутивные концепты

Постоянный интерес к концептам базируется на двух основных аспектах: философском и когнитивном, которые поддерживают, соответственно, логические и ментальные представления концептов. В обоих аспектах концепты обладают *представлением* и *расширением*: расширение относится к группе объектов (экземпляров концепта), которые иллюстрируют концепт, в то время как представление относится к основному значению, инкапсулированному в концепт. Это положение широко используется при моделировании баз данных. В этом разделе мы исследуем *представление* и *расширение*, чтобы разработать конструктивное представление базы данных для структуры концепта.

Важная роль представления концепта – спецификация свойств концепта (включая его признаки, функции, ограничения, отношения к другим концептам) и обеспечение классификации

для выделения объектов, которые являются примерами концепта. Например, концепт «лес» имеет свойства, типа названия, размера, формы, густоты, функции отдыха, коммерческой функции, и т.д., и она должна быть различима (возможно, не уникально) от других лесных массивов, основанных на спецификациях размера, функции, и т.д. Это контрастирует с традиционным понятием «экземпляра», в котором для концепта просто обеспечивается идентификация, но не общие свойства объектов, которые представляет концепт. Мы придерживаемся использования концепта, в котором *представление* обладает свойствами.

В глобальном отношении расширение группирует проявления концепта посредством механизма классификации, например, коллекция всех лесов в мире. Поскольку участники расширения отражают свойства представления, оно может быть расценено как определение всего «пространства возможности» для участников расширения, тем самым расширение обозначает то разнообразие, с которым связан концепт.

Существуют некоторые терминологические различия среди исследовательских сообществ: в классическом понимании расширение концепта называют классом, и очень общий концепт, типа субстанции Аристотеля, называют категорией; тогда как в когнитивном сообществе категория относится к расширению концепта. Кроме того, процесс определения расширения концепта упоминается как классификация в классическом смысле, и категоризация в когнитивном смысле. Мы будем использовать классическое обозначение класса, чтобы использовать группировку элементов, включающих расширение концепта, а также создание элемента заданного класса и классификацию, чтобы обозначить процесс объединения элементов в класс при расширении. Понятие «термин» используется для определения отдельного объекта, который мог бы быть помещен в класс, включая и геопространственный объект, в то время как пример объекта относится к объекту, который был помещен в класс.

Заметим, что создание объекта (экземпляра концепта) и классификация традиционно различаются: оба подразумевают добавление объекта к существующему классу, но создание объекта подразумевает создание нового термина, подобно созданию объекта в объектно-ориентированном программировании, тогда как классификация подразумевает размещение существующего объекта в класс, подобно распознаванию объекта при дистанционном зондировании, когда изображение классифицируется через маркировку его пикселей.

Следует различать два способа развития представления, причем оба могут изменить свойства концепта: при первом способе, который называется кластеризацией, коллекция объектов является отдельной группой, и содержание разрабатывается для каждой группы.

Кластеризация является слабо инновационным способом в том смысле, что содержания являются

результатом различных комбинаций фиксированного набора свойств. Второй способ – категоризация – достаточно инновационен, так как в этом случае выводятся новые свойства, что приводит к новым концептам и в конечном счете к новым экземплярам. Например, при анализе изображения, образцы часто обнаруживаются благодаря кластеризации различных аспектов фиксированных свойств пикселей; а в категоризации образцы стимулируют рассмотрение дополнительных свойств, приводящих к новым типам объектов. Кластеризация поэтому близко связана с индукцией или последовательным развитием образца, а категоризация связана с абдукцией или творческим актом генерации новых структур концептов, поэтому термин категория относится к представлениям концепта, связанным с развитием его содержания. Ключевые термины перечисленной спецификации представлены в таблице 1.

Таблица 1

*Ключевые термины спецификации*

Термин	Семантика термина
Концепт	Элементарное знание для понимания сущности или явления окружающей среды
Экземпляр концепта	Конкретная (реальная) сущность, которая соответствует концепту
Представление концепта	Обозначение (определение) и свойства концепта: атрибуты, допустимые действия (функциональность), правила и т.д.
Расширение (набор экземпляров концептов)	Группа экземпляров концепта, иллюстрирующих (описывающих) концепт
Класс концептов	Группа концептов, объединенных по некоторым признакам
Категория	Группа экземпляров концепта, используемая для расширения или уточнения представления концепта
Пример концепта	Экземпляр концепта, принадлежащий классу концептов
Ситуативная категория	Группа экземпляров концепта, используемая для расширения (лополнения) или уточнения примеров концепта или представления концепта
Классификация	Размещение существующих экземпляров концептов по классам
Создание экземпляра	Создание представления концепта на основе концептуально подобных экземпляров концепта из ситуативной категории
Кластеризация	Внедрение/ эксплуатация на ограниченном (пилотном) участке
Категоризация	Создание представления концепта на основе концептуально различных экземпляров концепта из ситуативной категории

Рис. 1 изображает фрагмент схемы базы данных в нотациях языка UML (Unified Modeling Language), который показывает традиционные отношения между концептом и его расширением. Здесь агрегация как ассоциация при отношении между целым и его частями представлена ромбиком на блоке класса и линией, идущей от этого ромбика к содержащемуся классу. На рисунке использованы мультипликаторы со следующим содержанием (Табл. 2).

Таблица 2

Мультипликаторы

Нотация	Содержание
0..1	Ноль или один экземпляр
1..1	Обязательно один экземпляр
0..* или *	Ноль или более экземпляров
1..*	Один или более экземпляров

Рис. 1 изображает фрагмент схемы базы данных в нотациях языка UML (Unified Modeling Language), который показывает традиционные отношения между концептом и его расширением. Здесь агрегация как ассоциация при отношении между целым и его частями представлена ромбиком на блоке класса и линией, идущей от этого ромбика к содержащемуся классу. На рисунке использованы мультипликаторы со следующим содержанием (Табл. 2).

Отметим, что семантика этих отношений, вообще говоря, согласуется с когнитивными и философскими традициями, связывая специфическое расширение с единственным концептом и обуславливая абстрактность тех концептов, которые не должны иметь расширение (например, «качество»).

Кроме того, логически возможно для одного и того же списка объектов находиться во взаимно-однозначном соответствии с расширениями существенно отличных концептов, подразумевая, что возникающие концепты могут быть классифицированы многими способами, реализуя различные пути развития концептов.

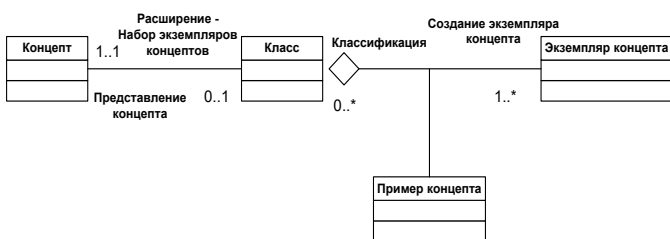


Рисунок 1. Традиционные отношения между концептом и его расширением

#### 4. Географические концепты

Географические концепты качественно отличаются от атрибутивных концептов главным образом пространственно-временным характером категоризируемых объектов. Некоторые объекты в геопространственной области не могут быть отнесены к материальным объектам, поскольку они могут быть:

- продуктом социальных соглашений и норм, типа геополитических границ;
- историческими факторами, влияние которых все еще проявляется на настоящее (например, предшествующее наводнение);
- интерпретируемыми объектами, типа объектов, выделенных из непрерывных данных;

- созданными человеком объектами, которые обнаруживаются лишь частично или косвенно, или могут быть обнаружены на несоответствующем масштабе; метеорологическими проявлениями, идентичность которых зависит от распознавания их составляющих и/или топологических отношений.

Такие геопространственные объекты не могут быть полностью наблюдаемы, но могут быть косвенно описаны. Эта косвенность обязывает проявить особое внимание к:

- обоснованию: ситуации, обеспечивающей распознавание возникновения концепта, и облегчающей его понимание;
- выводу: процессу, идентифицирующему возникновение или изучение концепта в возникшей ситуации;
- обнаружению: факту, что много концептов оказываются неизвестными заранее. Например, специалисты в области наук о Земле, изучающие новую территорию, создают не только новые пространственно-временные зоны, но также пересматривают или изобретают новые региональные концепты.

Учет перечисленных моментов должен повысить надежность распознавания ситуации, зависящей от географических концептов, при этом очевидно, что конкретизация и категоризация достаточно релевантны процессу географического обнаружения.

Схема, показанная ниже на рис. 2 включает терминологические различия и отражает ситуативный подход к обеспечению поддержки означивания как концептов, так и их примеров. Эволюция концептов смоделирована в схеме разделением понятия концепта на идентификатор концепта и на состояние концепта, являющегося собственно концептом. Аналогичный подход используется и для представления экземпляров концептов. Следовательно, изменения в представлении концепта или его расширении могли бы рассматриваться как новые состояния концепта, а изменения в свойствах конкретного экземпляра концепта, такие, как пространственное или временное описание, или классификация, были бы таким же образом отмечены как различные состояния экземпляра концепта. Концепты могут таким образом быть охарактеризованы как обладающие постоянной идентичностью, связанной с множественными изменениями состояний, состоящих из классов, категорий и свойств; аналогично, экземпляры концептов с постоянной идентичностью могли бы быть охарактеризованы как обладающие множественными состояниями свойств.

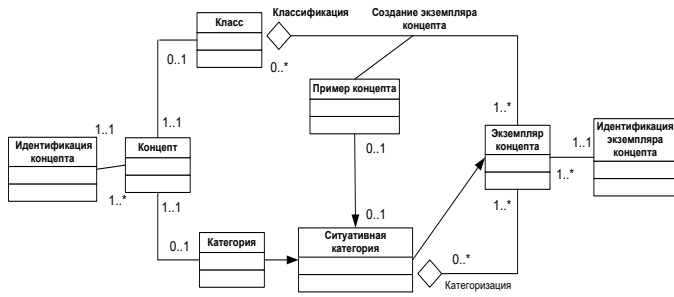


Рисунок 2. Модель данных, представляющая взаимодействия между концептами и поддерживающая ситуативную категорию.

Запрос, соответствующий схеме, описанной выше, позволяет извлечь из базы геоданных контекст для географических концептов или характеристик. Обеспечение такой контекстной информации имеет несколько преимуществ, включая: коммуникацию, при которой большее значение может быть передано или стимулировано; обоснование, при котором можно обеспечить запись и обоснование для научных и других полученных данных; онтологию, при котором эмпирическое и динамическое измерение может быть добавлено к геопространственной онтологии; и обобщение, так что обобщение и абстрагирование типа картографических обобщений может быть связано с детализированными источниками и согласовано с концептуальными структурами.

Контекст определен ниже как рекурсия ситуативного концепта, характеризующего концепт или его конкретный пример.

Обозначим через  $X$  концепт,  $o$  – экземпляр концепта,  $i$  – пример концепта и  $e$  – ситуативную категорию как группу  $\{e_j\}$  экземпляров концепта, тогда:

$$\begin{aligned}
 o.Context() &\rightarrow o \\
 i.Context() &\rightarrow (X.i, X.i.o, X.Context(), X.i.e.Context()) \\
 X.Context() &\rightarrow (X, X.e.Context()) \\
 e.Context() &\rightarrow (e, e_1.Context(), \dots, e_n.Context())
 \end{aligned}$$

Эта рекурсивная формулировка приводит к расширяющейся сети ситуативных концептов для базового концепта или его экземпляра. Такие сети могут быть представлены оргграфом без петель, который описывает геопространственный случай или дополнительно контекстуализирует концепт. Однако такие графы являют собой достаточно сложные конструкции. Упрощенный контекст, который можно назвать интерпретацией, уменьшает эту сложность, ограничивая широту контекста единственным путем через граф.

$$\begin{aligned}
 o.Interp() &\rightarrow o \\
 i.Interp() &\rightarrow (X.i, X.i.o, X.Interp(), X.i.e.Interp()) \\
 X.Interp() &\rightarrow (X, X.e.Interp()) \\
 e.Interp() &\rightarrow (e, e_j.Interp())
 \end{aligned}$$

Основная задача баз геоданных – хранение структуры и объяснение «концептов» модуля карты, которые маркируют области на тысячах карт. Также важны структура и научное объ-

яснение этих концептов и областей. Эти элементы собираются главным образом из легенд цифровых геологических карт; информация об определенных геопространственных проявлениях также существенна, но это второстепенный вопрос, поскольку такую информацию часто трудно получить.

Концепты модуля карты и их описательный язык используются гетерогенно: оба типа терминов могут иметь различные описания в различных географических областях и даже среди людей, и различные термины часто имеют идентичные значения [1,2]. Решения этой проблемы разрабатывают и в нисходящих, и в восходящих режимах: нисходящие нормативные определения разрабатываются для многих из терминов; в то же самое время в восходящем подходе существующие термины вводятся в базы данных в целях обнаружения эмпирической регулярности в использовании терминов и координирования их с определениями. Цель состоит в том, чтобы сохранить локальные термины и характеристики, поскольку они применяются непосредственно к картам, а также собрать эти термины в однородную систему в целях функциональной совместимости. Следовательно, ситуативный характер информации как таковой и динамический характер геопространственной онтологии требуют создания системы, которая может обеспечить обработку контекстов, чтобы облегчить понимание концептов и в конечном счете объектов карты.

## 5. Проблема соотношения геоописаний

Геоописания объектов, формируемые на основе географических концептов, характеризуются большим объемом, сложностью, взаимозависимостями и динамизмом. Эти описания могут существовать в несопоставимых форматах из-за наличия разнообразных источников данных и программных систем. Соотнесение геоописаний в согласованную и унифицированную форму считается критическим мероприятием для успешного принятия решений на их основе [3]. К сожалению, большинство разработанных в последнее время программных средств функционируют как автономные системы с ограниченной возможностью совместного использования геопространственной информации с другими системами, что привело к возникновению так называемых «островов информации» и появлению противоречивых моделей геоданных в несопоставимых программных продуктах.

В большинстве реализаций ГИС до настоящего времени геоданные сохранялись и обрабатывались в персональных или ведомственных базах геоданных, которые ограничивали совместное использование и редактирование. Возрастающие требования к совместной обработке геоданных для различных приложений выявили острую потребность в масштабируемости ГИС и созданию различных моделей соотношения геоописаний, которые при этом должны гарантировать

согласованность и целостность геоданных и объединить различные форматы во всестороннее и непротиворечивое геоописание.

При соотношении геоописаний необходимо учитывать явные и латентные перекрестные ссылки и отношения между различными уровнями детализации, обеспечивающими взаимодействия между сорасположенными или накладывающимися объектами [3]. Это позволит обеспечить функциональную совместимость и эффективное совместное использование геоданных при решении задач разной детализации.

Отдельной проблемой представляется преодоление мультилингвальности геоописаний, когда требуется построить согласованное унифицированное геоописание на основе интеграции информации из геоописаний с использованием различных естественных языков. На рис. 3 перечислены задачи, требующие решения при соотношении геоописаний.

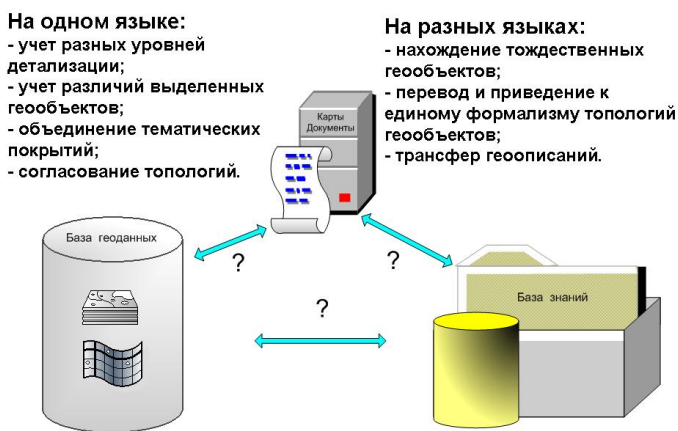


Рисунок 3. Задачи соотношения геоописаний

Соотношение геоописаний позволяет:

- повысить информативность;
- лучше осмыслить геоописания;
- оценить качество геоописаний;
- улучшить согласованность экспертов;
- выработать способы сопоставления геообъектов;
- апробировать формализмы трансфера геоописаний;

База геоданных (ГБД) содержит геоданные в четырех представлениях [3], требующие специальных методов и приемов для их отображения в реляционной среде:

- векторные данные для представления пространственных объектов,
- растровые данные для представления непрерывных полутонных изображений, сеточных тематических данных и поверхностей,
- нерегулярные триангуляционные сети (TINs) для представления поверхностей,
- адреса и локаторы (locators) для нахождения географического положения.

Помимо этой информации при формировании геоинформационного портала в ГБД необходимо обеспечить хранение и сопровождение огромного объема атрибутивной информации, включающей разнообразные тексты и документы. В рамках системы концептов А.А. Лютого геоинформаци-

онный портал можно представить как электронную библиотеку с геотекстами на разных естественных языках.

Таким образом, геоинформационный портал представляется как семиотически неоднородный по языковой принадлежности его текстов. Наряду с моноязычными вербальными текстами в нем могут присутствовать поли- и кроссязычные тексты. Аналогично и геотексты в нем могут быть моногеоязычными, поли- и кроссгеоязычными. Соответственно, одной из центральных задач здесь является извлечение релевантной картографической информации из этих геотекстов. На рис. 4 показана примерная схема из блоков, участвующих в интеллектуальном анализе геоданных, представленных в виде геотекстов.

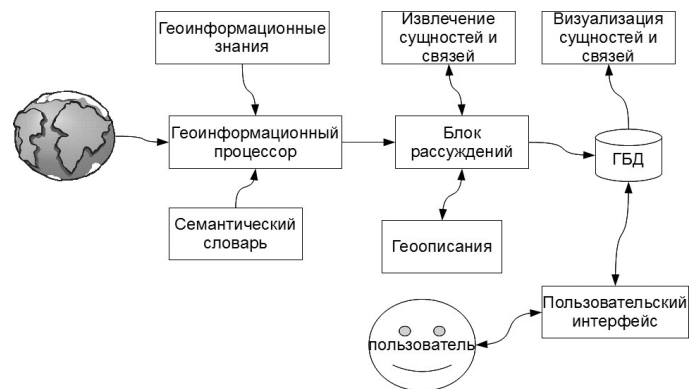


Рисунок 4. Извлечение картографической информации из геотекстов

Пространственная информационная системная среда формирования геоописаний может быть представлена объединением ГИС с внешними программами и специальными прикладными программами, обеспечивающими выборку необходимых геоданных. Особую важность здесь представляет инструмент взаимодействия с репозиторием геоданных, который позволяет адекватно пополнять существующие геоописания за счет соотношения с другим геоописанием того же объекта (Рис. 5).

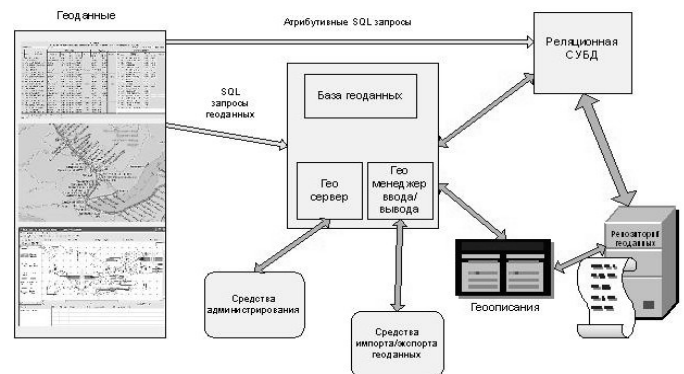


Рисунок 5. Схема формирования геоописаний

## 6. Трансфер геоописаний

Под задачей трансфера геоописаний будем понимать процесс анализа и соотношения геоописаний на двух или более геоязыках для извлече-

ния и сопоставления релевантной информации. Этот процесс является крайне необходимым при пополнении репозитория геоданных, так как позволяет получить наиболее подробное и разностороннее представление описываемых в системе геообъектов.

Можно заметить, что задача трансфера геоописаний сильно коррелирует с задачей перевода текстов на естественном или искусственном языке на основе межязыкового трансфера. Межязыковой трансфер применяется либо при переводе сильно формализованных текстов на естественном языке, либо при переводе текстов на искусственном языке [3]. Выбранный авторами настоящей работы геоязык, на котором сделано геоописание, является сильно формализованным искусственным языком, что позволило методы и средства, реализующие межязыковой трансфер, успешно применить и при трансфере геоописаний.

Одним из подходов к реализации трансфера, предложенным в работе [4], является использование в качестве аналитической машины языка функционального программирования. Это позволяет решить задачу проектирования аналитической машины, допускающую грамматические правила трансфера геоописаний. Такие грамматические правила необходимо задавать в виде функций в математическом смысле.

Следует отметить, что аналитическая машина, в основе которой лежит функциональный подход, ориентирована, в первую очередь, на анализ нормализованных геотекстов. Под нормализованным геотекстом понимается такой геотекст, который был обработан синтаксическим и/или морфологическим анализатором.

Подход к описанию правил грамматики, как к заданию математической функции (с ее областью определения и областью значения), позволил применить парадигму функционального программирования при технической (программной) реализации аналитической машины. Другими словами, имея грамматику геоязыка (заданную в виде функций) и среду функционального программирования (например, Erlang или Haskell [4]), мы получили аналитическую машину, позволяющую проводить анализ геотекстов согласно заданной грамматике. Кроме того, использование среды функционального программирования как аналитической машины позволило проектировщику применить такие технические средства, как атомы языка и кортежи для представления группирования атрибутивных характеристик.

Остановимся на задании атрибутивных характеристик. В случае использования кортежей и атомов функционального языка проблема обработки атрибутивных характеристик перестала быть заданием базы геоданных, что позволяет переложить эту проблему на язык функционального программирования. Таким образом, обработка и контроль целостности атрибутивных характеристик выполнена на уровне функционального языка, а не на уровне геоинформационной мо-

дели. Это преимущество значительно упростило построение аналитической машины для обработки геоописаний, так как позволило вынести атрибутивные характеристики за рамки модели.

На представленной на рис. 6 схеме видно, что трансфер геоописаний проходит почти те же стадии обработки, что и межязыковой трансфер. Как и любой язык, геоязык имеет синтаксис и морфологию, соответственно, в системе трансфера геоописаний наличие морфологического и синтаксического анализатора является обязательным условием. После обработки отмеченными анализаторами получается нормализованный геотекст. Такой геотекст подается на вход системе функционального программирования, которая, используя функциональные грамматические правила, представляет геоописание в функциональном виде. Дальнейшая обработка полученного геоописания сводится либо к синтезу геоописания на другом геоязыке, либо к пополнению и обновлению данных в репозитории геоданных.

Таким образом, в результате работы системы трансфера геоописаний, пользователь геоинформационной системы получает возможность сопоставления геоописаний на разных геоязыках с возможностью пополнения и обновления системного репозитория геоданных.

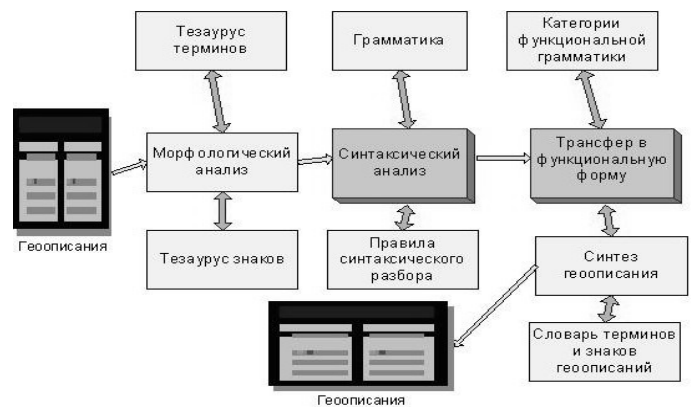


Рисунок 6. Функциональная схема трансфера геоописаний

## Заключение

В работе представлен подход к построению ГИС, позволяющий учитывать как ситуативный характер геоконцептов, так и семантику геообъектов. При моделировании геоконцептов использовались понятия концепта, класса и категории в терминах географических описаний с учетом развития ситуаций и структуры для представления ситуативных компонентов.

Можно указать основные черты предлагаемого подхода:

- необходимо, чтобы геопространственные среды, содержащие знание, формировались на основе абстрактной схемы базы геоданных, смоделированной на эпистемологических примитивах, типа концептов, а не на концептах конкретной области. Разработка схемы и сложность при этом уменьшаются, однако такие гибкость и абстракция требуют более строгой

онтологической техники представления знаний, которая может не соответствовать принятым действиям в существующих геопространственных дисциплинах, но которые, тем не менее, могут привести к лучшей семантике и улучшенной функциональной совместимости баз геоданных;

- целесообразно использовать четыре функции, интегрирующие концепты: классификация, кластеризация, реализация и категоризация;

- введение ситуативной категории приводит к усовершенствованию терминологии: группа экземпляров концепта поддерживает расширение семантики концепта, позволяя формировать категорию концепта; а группа примеров, иллюстрирующих концепт, позволяет сформировать его класс. Это вносит определенный вклад в понятие динамической геопространственной онтологии;

- расширение географического контекста реализуется как многоуровневый набор релевантных ситуативных концептов, связанных с базовыми географическими концептами.

Можно использовать этот механизм, чтобы проследить за развитием концепта, стимулируя выявление неявного знания, связанного с концептом;

- схема базы геоданных разрабатывается с учетом моделирования взаимодействия между концептами, примерами и ситуативными концептами, представляя реализацию схемы в контексте симбиоза базы геоданных и геологических онтологий;

- для согласования геоконцептов [5] используется механизм трансфера геоописаний, позволяющий выполнить соотнесение геоописания на разных геоязыках, а также пополнять и обновлять репозиторий геообъектов. Для технической реализации системы трансфера геоописаний выбрана среда функционального программирования, что позволило решить задачу трансфера геоописаний аналогичным с трансфером межъязыковых текстов образом.

## Список литературы

1. Rodríguez M. A., Egenhofer M. J. Comparing geospatial entity classes: an asymmetric and context dependent similarity measure. – *International Journal of Geographical Information Science*, 18(3): 2004, pp. 229–256.
2. Brodaric, B. and Hastings, J. An Object Model for Geologic Map Information. In: *Proceedings of the Spatial Data Handling 2002 Symposium*, Ottawa, Canada, 2002, pp. 113-117.
3. Ермаков П.В., Кожунова О.С. Применение средств функционального программирования в задачах представления языковых и межъязыковых структур // Сборник докладов конференции «Первая школа молодых ученых ИПИ РАН». – М.: ИПИ РАН, 2010. – С.15-22.
4. Дулин С.К., Ермаков П.В. Функционально-вычислимые грамматики в задачах анализа текстов на естественном языке – М.: ВЦ РАН, 2011. – 22 с.
5. Dulina N., Kozhunova O. Information monitoring system: a problem of linguistic resources consistency and verification // Третья международная конференция «Проблемы кибернетики и информатики» (PCI'2010) – Баку: Elm, 2010. – С. 56-58.