

УДК: 004.052.2

Структура семантической интероперабельности для онтологии объектов железнодорожной инфраструктуры

Semantic interoperability structure for ontology of railway infrastructure objects

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

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



Аннотация

В работе представлена структура семантической интероперабельности, ориентированная на использование прикладной онтологии железнодорожной инфраструктуры. При этом установление геоинтероперабельности выходит за пределы простой возможности получить доступ к информации баз геоданных, так как оно требует больше времени на обеспечение транзакций на основе заранее определенного точного словаря базы геоданных инфраструктуры. Но самое существенное – это то, что пользователи и провайдеры должны иметь релевантное понимание семантики запросов и ответов. В работе показано, что моделирование семантики должно быть глубоко внедрено в структуру геоинтероперабельности, обеспечивая исчерпывающее описание семантической геоинтероперабельности, лежащей в основе развития семантической пространственной инфраструктуры данных.

Ключевые слова: транспорт, геоданные, семантическая геоинтероперабельность, онтологии, Semantic Web.

Abstract

The paper presents the structure of semantic interoperability focused on the use of applied ontology of railway infrastructure. At the same time, establishing geointeroperability goes beyond simply being able to access geodatabase information, as it requires more time to provide transactions based on a pre-defined, accurate infrastructure geodatabase vocabulary. But the most important thing is that users and providers must have a relevant understanding of the semantics of requests and responses. The paper shows that semantic modeling should be deeply embedded in the structure of geointeroperability, providing an exhaustive description of semantic geointeroperability, which underlies the development of semantic spatial data infrastructure.

Keywords: transport, geodata, semantic geo interoperability, ontologies, Semantic Web.



Введение

Сегодня ведущей парадигмой структурирования информационного контента являются онтологии, или иерархические концептуальные структуры [1], которые формируются аналитиком на основе изучения и структурирования потоков информации, документов, протоколов извлеченных знаний и других источников. С методической точки зрения это один из наиболее систематических и наглядных способов структурирования и формализации знаний. Онтологии особенно актуальны в прикладных областях, где от точной концептуализации спецификаций зависит безопасность функционирования, например, в транспортных системах.

Онтологии – содержательные теории, которые включают общий набор распространяемых фактов, чье основное назначение – идентифицировать определенные классы объектов и отношений, которые существуют в некоторой части предметной области. Таким образом, неформально определенные онтологии – это соглашения об общедоступной концептуализации. Формальное определение основывалось бы на том, что онтология является (возможно, неполной) аксиоматизацией допустимых прикладных моделей. Другими словами, онтология состоит из основного словаря и отношений, используемых для описания некоторых аспектов действительности, включая ряд аксиом, связанных с предполагаемым значением словаря.

Следует заметить, что специфика географического мира в достаточной мере определяет параметры создания онтологий. Чтобы адекватно представлять географический мир, необходимо иметь компьютерные представления географических знаний (в первую очередь – изображений), которые способны фиксировать не только описательные атрибуты пользовательских концепций, но также и описывать геометрические и позиционные компоненты этих концепций. Эти представления также должны фиксировать пространственные и временные зависимости между экземплярами этих концепций.

В отличие от случая обычных информационных систем большинство пространственных и временных зависимостей не представлены в ГИС и чаще всего могут просто выводиться путем использования различных географических функций. Поэтому обязательно должна быть привнесена дополнительная семантика в схему

географического приложения, семантические спецификации которой, являющиеся частью онтологии этого приложения, зафиксированы разработчиком модели. Новое поколение информационных систем должно обладать способностью обрабатывать семантическую неоднородность, возникающую в результате использования разнородных источников информации.

Использование множественных онтологий становится основной особенностью современных информационных систем, если в них предполагается поддержка семантики при интеграции информации. Онтологии могут фиксировать семантику информации, могут быть представлены в формальном языке и могут также использоваться, чтобы хранить связанные метаданные, допуская, таким образом, семантический подход к информационному интегрированию.

Онтология играет существенную роль в создании ГИС, так как она позволяет устанавливать соответствия и взаимосвязи среди различных типов геоданных – пространственных сущностей и отношений. Использование онтологий будет способствовать улучшению функционирования информационных систем, благодаря тому, что удастся избежать проблем, вызванных противоречиями между онтологиями, неявно внедренными в ГИС, конфликтами между онтологическими концепциями и реализацией, а также конфликтами между онтологиями «здорового смысла» пользователя и математическими концепциями программного обеспечения.

Можно заметить, что существующие концептуальные схемы баз геоданных могут быть весьма полезны для формирования онтологий, так как они – формальные документы, которые были разработаны, чтобы зафиксировать взгляд специалиста на некоторую предметную область реального мира. Поэтому существующие концептуальные схемы могут использоваться, чтобы создать приблизительные онтологии, в то время как существующие онтологии могут использоваться, чтобы генерировать концептуальные схемы с помощью или без помощи опытного разработчика модели [2, 3].

Имеются три различных уровня абстракций, как для онтологий, так и для концептуальных моделей (рис. 1) [4].

Первый уровень формальный, на котором в конструкции онтологий и концепту-



альных схем включены абстрактные концепции. В случае онтологий на этом уровне мы имеем абстракции формальных особенностей сущностей, такие как понятия времени и пространства. В случае концептуальных схем на этом уровне мы находим основные идеи, взятые из концептуальных моделей данных, т. е. понятия, которые широко используются в концептуальном моделировании данных: объекты, поля и связи.

Когда мы сопоставляем содержание формального уровня с географическими приложениями, мы получаем второй уровень абстракции, или уровень типов геоданных.

На этом уровне онтологии описывают словарь, который используется, чтобы представить действительное содержание знания о свойствах геоданных. На этом уровне явно присутствует онтология географических понятий, которая описывает географическое пространство, географические объекты и явления географического пространства.

На прикладном уровне онтологии более конкретны, являясь уточнением специализаций уровня типов геоданных, и формируются в соответствии с концепциями поля знания, заданного приложением. На этом уровне онтология – это способ осмысления набора концепций, позволяющего совместно и согласованно использовать его группой пользователей. В концептуальном моделировании данных на прикладном уровне примитивы концептуальной модели объединены, чтобы образовать удобочитаемые диаграммы, из которых можно получить детальные потребности приложения по организации данных. Следовательно, примитивы модели данных используются, чтобы определить концептуальную схему приложения, перемещаясь от уровня типов геоданных до прикладного уровня.

Представление пользовательских онтологий прикладной области в этом контексте рассматривается как существенная часть фиксации концепций информационного пространства. Исследование онтологического статуса пространственных типов данных [5] – наиболее актуальное направление в геоинформатике. В [6] впервые было введено понятие ГИС, управляемой онтологиями, призванной обеспечивать пользователей географической информацией возможностью достигнуть соглашения по основным сущностям географического мира. Идея управления с помощью онтологий заключается в том, что существенная часть географического знания зафиксирована

Тип представления / Вид уровня	Онтологии	Концептуальные модели
Формальный уровень	Абстракция формальных особенностей научных предметов: геометрия, география, время, пространство	Понятия концептуального моделирования: объекты (цели), области (поля), отношения
Уровень логических типов данных прикладной области	Онтология географических видов: представление, местоположение, топология, географическая информатика	Концептуальная модель и нотация: классы, пространственные отношения, пространственные ограничения целостности
Прикладной уровень	Онтология инфраструктуры объектов ОАО «РЖД»: станции, перегоны и др.	Концептуальная схема БГД: атрибутивное описание сущностей ОАО «РЖД»

Рисунок 1. Уровни онтологий

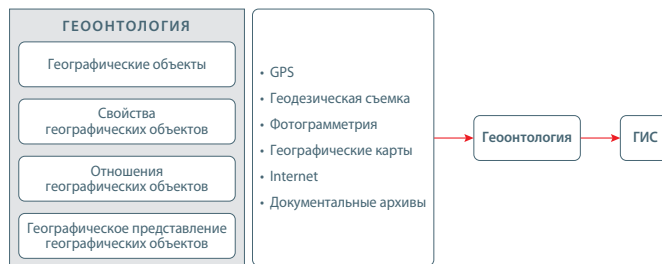


Рисунок 2. Формирование геоонтологии

процедурами, которые извлекают информацию из пространственных наборов данных, т. е. для этого необходимо создать онтологии не только для объектов некоторой области, но также и для намеченных действий, которые выражены процедурами, применимыми к набору данных, предназначенному для извлечения знания.

Модели географического мира должны сводиться друг к другу, а формальное описание геоинформационных сущностей может обеспечить создание онтологий географических видов (рис. 2).

Структура семантической геоинтероперабельности

В [7] было введено понятие интероперабельности, означающей способность клиентов одной системы использовать части другой системы без специальных условий. Здесь с позиции информатики основная суть определения – это способность использовать информацию, полученную в результате обмена.





Рисунок 3. Структура семантической геоинтероперабельности

В свою очередь, географическая интероперабельность – это способность информационных систем к 1) свободному обмену всех видов пространственной информации о Земле и об объектах и явлениях на, выше и ниже поверхности Земли; 2) совместному сетевому использованию программного обеспечения, предназначенному для управления такой информацией». Семантическая геоинтероперабельность заключается в обеспечении согласованного понимания смысла геоданных, например в процессе взаимодействия пользователей геопорталов при решении задач, требующих совместной обработки и использования ими геоданных, или при решении задач концептуального поиска в базах геоданных. Семантическая геоинтероперабельность принципиально отличается от синтаксической [7], так как в последнем случае речь идет только об обеспечении согласованной обработки и обмена геоданными на нескольких уровнях взаимодействия (сетевом, обменных форматах, вызовах удаленных процедур, запросов к базам геоданных).

Геоинтероперабельность можно сравнить с межличностным процессом коммуникации. Попробуем описать это в контексте взаимодействия двух агентов, т.е., пользователя (П) и провайдера геоданных (ПД), взаимодействующих на основе географической информации (рис. 3). П хотел бы получить информацию о некоторых географических

характеристиках местности, пусть это дорожная сеть в некотором регионе. П посылает запрос дорожной информации к ПД по интернету (каналу коммуникации), используя собственный словарь. Как только запрос достигает ПД, ПД интерпретирует его, чтобы идентифицировать и собрать имеющуюся информацию для ответа на запрос П (например, станция Северная, Восточный полигон и т.д.; пространственные пиктограммы, содержащие одномерные и двумерные объекты).

В свою очередь, когда П получил ответ на свой запрос от ПД, он интерпретирует его, чтобы определить, удовлетворяет ли ответ посланному запросу. Геоинтероперабельность между двумя агентами существует, если посланный запрос удовлетворен. Здесь два агента используют свои собственные словари, чтобы выразить абстракции реальных явлений. Эта ситуация показывает четыре различных выражения действительности, обозначенных через ПО, ПКМГ, ФКЗ и ППУС. Эти выражения связаны друг с другом как части процесса коммуникации. Во-первых, есть топографическая действительность (ПО), поскольку она существует на данный момент и предъясняется П. Это референт для П.

Во вторых, П разработал свое собственное когнитивное представление ПО, т.е. ПКМГ как симбиоз ее наблюдения и структуры справочной информации, т.е. набора правил >>>

и знаний, используемых для абстрагирования явлений. Оно составлено на основе существенных свойств (например, геометрический объект, временные и описательные свойства, поведение и отношения), которые объединяются и структурируются в понятия (т.е. символизируются). Следовательно, понятие – это упрощенное представление реального явления или его части, которое соответствует символу восприятия. Таким образом ПКМГ состоит из онтологии, созданной П.

Так как понятия – это только теоретические представления реальных явлений, П не может передать их непосредственно. Следовательно, понятия должны быть преобразованы в физические формы (например, звуки, биты и байты, и т.д.), чтобы образовать сообщение. Этот процесс называется операцией кодирования. С ее помощью П преобразует понятия топографической действительности в знаки, такие как слова, сокращения, пунктуации, символы и пиктограммы, организованные определенным способом, например, «мост через реку». Эта часть коммуникации обозначена как ФКЗ и относится к организации знаков. После того, как сообщение закодировано, оно может быть отправлено по каналу коммуникации в сторону ПД. После передачи сообщение теряет свое значение и становится набором данных, передающихся с целью взаимодействия. Когда ПД получает сообщение, он начинает операцию декодирования, которая является необходимой для интерпретации сообщения. Так же, как и П, ПД полагается на свой собственный набор понятий, т.е. ППУС, чтобы определить соответствующее значение полученному сообщению. В нашем примере это соответствует «мост через реку». Здесь ППУС является онтологией ПД.

В то время, когда ПД придал значение сообщению, он инициирует поиск и сбор информации, которая позволяет выполнить запрос (т.е. Крымский мост, Строгинский мост и т.д.). В результате эта информация кодируется и помещается в канал коммуникации как данные для П. Эта часть коммуникации обозначена как ФКЗ.

Когда П получает ответ на свой запрос, он инициирует декодирование сообщения, идентифицирует понятия, которые он знает, чтобы придать значение сообщению, и, наконец, оценивает, отвечает ли сообщение правильно на его начальный запрос. Эта часть коммуникации обозначена как ППУС. Успешное взаимодействие соответствует установлению геоинтероперабельности между двумя агентами П и ПД.

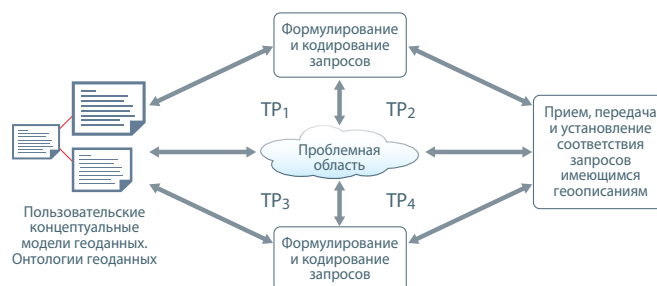


Рисунок 4. Семиотика в структуре семантической геоинтероперабельности

Структура семантической геоинтероперабельности включает четыре различных случая семиотического треугольника Фреге [8], изображенные на рис.4.

Первый треугольник возникает как результат отношений между действительностью (ПО), онтологией П – (ПКМГ) и закодированным запросом (ФКЗ). Второй треугольник – отношения между действительностью (ПО), закодированным запросом (ФКЗ) и онтологией ПД (ППУС). Третий треугольник появляется в результате отношений между действительностью (ПО), онтологией ПД – (ППУС) и закодированным ответом (ФКЗ). Наконец, четвертый треугольник – это отношения между действительностью (ПО), закодированным ответом (ФКЗ) и онтологией П – (ПКМГ). Все это показывает, что моделирование семантики глубоко внедрено в структуру геоинтероперабельности и, таким образом, обеспечивает исчерпывающее описание семантической геоинтероперабельности в целом, которая лежит в основе развития семантической пространственной инфраструктуры данных и геосемантической сети Geosemantic Web.

Семантическая геоинтероперабельность через SDIs

Геоинтероперабельность представляет собой основу для развития и реализации пространственных инфраструктур данных (Spatial Data Infrastructures (SDIs)) [9]. Цель SDIs состоит в том, чтобы координировать полезный обмен и совместное использование географической информации с использованием соответствующих сервисов. SDIs – это средства разработки функциональной совместимости для географической информации. SDIs составлены из пяти >>>

элементов: методики, технологии, стандарты, человеческие ресурсы и релевантные действия, требуемые для сбора, обработки, управления, доступа, поставки и использования географической информации. SDIs были разработаны на основе референсной модели для Open distributed processing (RM-ODP) [10]. RM-ODP состоит из структуры пяти разделов: раздел предприятия, информационный раздел, вычислительный раздел, технический раздел и раздел технологии. В контексте SDIs раздел предприятия описывает цели и возможности, методики, обязанности и бизнес-процессы SDIs. Информационный раздел, по существу, посвящен информации, доступной через SDIs и необходимой для семантической геоинтероперабельности. Вычислительный раздел касается функциональной декомпозиции SDIs в сервисы с интерфейсами и операциями. Этот раздел представляет большой интерес для определения семантических компонентов и сервисов. Технический раздел главным образом связан с взаимодействием между данными, сервисами и системными взаимосвязями. Наконец раздел технологии относится к определенно выбранной технологии для реализации SDIs.

Первые три из указанных разделов особенно важны с точки зрения семантической геоинтероперабельности. Именно в разделе предприятия цель достижения семантической геоинтероперабельности должна быть четко сформулирована. Раздел предприятия должен идентифицировать любой репозиторий, который должен участвовать в SDIs, а также поставщиков данных, которые помогут пользователям и провайдерам в поиске соответствующих данных, релевантных их словарю и семантике. Вряд ли все пользователи могут знать заранее точный словарь и семантику, используемую геоинформационными источниками, доступными в SDIs.

Поэтому пользователи должны быть в состоянии взаимодействовать с SDIs, используя свой собственный словарь и находить данные, которые соответствуют их определенной цели. Следовательно, информационный раздел должен включать необходимые информационные компо-

ненты, чтобы обеспечить семантические запросы (т.е. запросы, сделанные в словаре пользователя и правильно интерпретируемые информационным сервером). Поэтому онтологии должны быть объединены как часть информационного раздела, обеспечивая фундаментальное знание для рассуждения, интерпретации запроса и выдачи соответствующих ответов. Вычислительный раздел нуждается также в определенном внимании, чтобы можно было решать проблемы определения семантических интерфейсов, связанных с онтологиями, включая операции и функции логического вывода, которые помогают интерпретации запросов и ответов. Этот раздел включает также кодирование онтологий, позволяющее устанавливать с помощью интерфейса связь с семантическими сервисами.

Учет семантики на раннем этапе разработки SDIs облегчает их дизайн и позволяет идентифицировать онтологии и требуемые семантические сервисы.

Заключение

Семантическая геоинтероперабельность стала областью активных научных исследований в течение последних двух десятилетий. Описание структуры семантической геоинтероперабельности в процессах совместной обработки географической информации важно для многих прикладных областей. Моделью семантики здесь является классическая триада (семиотический треугольник Фреге), включающая, знак (символ), концепт (осмысление) и денотат (референт). Познание напрямую связано с семантикой, где перцепционные символы, сохраненные в человеческой памяти, составляют набор понятий, используемых людьми. Информатика заимствовала понятие онтологии из философии для описания понятий, которые необходимы для развития семантической геоинтероперабельности. Онтологии – это часть жизненного цикла данных, которые должны быть разработаны и сделаны доступными с соответствующими спецификациями. ■



Список литературы

1. Грегер С.Э., Поршнева С.В. Построение онтологии архитектуры информационной системы // Фундаментальные исследования. 2013. № 10 (часть 11) С. 2405-2409.
2. Jian H Qin, Maja Žumer, Xiaoguang Wang, Wei Fan. Conceptual models and ontological schemas for semantically sustainable digital libraries // Proceedings of the ACM/IEEE Joint Conference on Digital Libraries (JCDL '20). 2020.P. 441–442.
3. Verdonck M., Gailly F., Pergl R., Guizzardi G., Martins B., Pastor O. Comparing traditional conceptual modeling with ontology-driven conceptual modeling // An empirical study. Information Systems. 2019. V.81, P. 92 – 103.
4. Дулин С.К., Поповидченко В.Г. Структура представления онтологии геоданных. М.: ВЦ РАН, 2007. 23 с.
5. Fonseca F., et al. Ontologies and Knowledge Sharing // Urban GIS. Computer, Environment and Urban Systems, 2000. № 24, Vol. 3. P. 232-251.
6. Fonseca F., Egenhofer M., Davis C. Semantic Granularity in Ontology-Driven Geographic Information Systems // Annals of Mathematics and Artificial Intelligence. 2002. V.36. P. 121–151.
7. Дулин С.К., Дулина Н.Г. От согласованности геоданных к семантической геоинтероперабельности. М.: ВЦ РАН, 2014. 28 с.
8. Фреге Г. Смысл и денотат // Семиотика и информатика. Вып. 35. М.: Языки русской культуры, 1997. С. 352–379.
9. J. Hjelmager, H. Moellering, A. Cooper, T. Delgado, A. Rajabifard, P. Rapant, D. Danko, M. Huet, D. Laurent, H. Aalders, A. Iwaniak, P. Abad, U. Düren, A. Martynenko: An initial formal model for spatial data infrastructures // Int. J. Geogr. Inf. Sci. 2008. V. 22, P. 1295–1309.
10. ISO/IEC 10746:1998 Information technology – Open distributed processing – Reference model: Overview (International Organization for Standardization, Geneva 2017).