

Метаданные в компьютерных системах¹

© М.Р. Когаловский
E-mail: kogalov@cemi.rssi.ru
Институт проблем рынка РАН

Программирование, МАИК/Наука «Интерпериодика». 2013. Т. 39, № 4. С. 28-46

Аннотация

Важную роль в современных компьютерных системах играет особый вид информационных ресурсов, называемых метаданными. Метаданным посвящено огромное количество публикаций. В большинстве из них рассматриваются стандарты метаданных, используемых в различных областях, и весьма малочисленные публикации, в которых обсуждаются свойства и функции этого вида информационных ресурсов. Вероятно, по этой причине трактовка термина *метаданные* все еще не устоялась. Нередко в публикациях встречаются явные ошибки, при обсуждении частных видов метаданных не делаются необходимые оговорки, и тем самым их свойства и функции неправомерно распространяются на общий случай. В данной работе предлагается систематический взгляд на метаданные как на информационные ресурсы особого рода, обсуждается вопрос об определении этого термина, приводятся примеры метаданных, используемых в различных пластах информационных технологий, рассматриваются общие (независимые от сферы применения) свойства метаданных и их функции, выразительные средства для их представления, известные обобщенные классификации метаданных, а также деятельность по их стандартизации.

1. Введение

Хранение и обработка данных в компьютерных системах, обмен данными между ними и доступ к ним пользователей невозможны без явно представленных описаний свойств этих данных. Такие описания необходимы программным средствам, выполняющим указанные функции, а также пользователям для оценки возможности применения имеющихся данных в различных ситуациях, интерпретации и анализа их содержания, формулировки запросов. Описания такого рода называются *метаданными* и являются особым видом информационных ресурсов. Их создание часто требует значительных усилий и существенных затрат. Однако они существенно повышают ценность данных, обеспечивают более широкие возможности их использования.

Довольно трудно установить, кто и когда в области информационных технологий впервые предложил использовать термин *метаданные*. Авторы работы [1] утверждают, что это было сделано Г. Мили в 1967 г. в его статье [2]. Однако в этой статье интересующий нас термин вовсе не упоминается. Возможно, правы авторы статьи о метаданных в англоязычной Wikipedia [3], приписывающие первенство П. Бэгли [4].

Каков бы ни был ответ на этот вопрос, метаданные начали использоваться задолго до появления компьютерных систем и до введения этого термина в научно-техническую лексику. Библиографические описания использованных источников в публикациях, библиотечные каталоги и тематические указатели, различные классификаторы, аннотации статей – все это примеры метаданных.

В данной работе нас будут интересовать метаданные, используемые в компьютерных системах. В этой сфере метаданные начали использоваться уже на ранней

¹ Работа поддержана грантом РГНФ, проект 11-02-12026-в

стадии их развития. Когда начали создаваться языки и технологии программирования, термин *метаданные* еще не существовал. Однако в исходном коде программы на языке программирования должны были содержаться описания типов данных, которыми эта программа оперирует. Используя описания, компилятор обеспечивает требуемое для этих данных распределение оперативной памяти, оценивает правомерность выполнения над ними предусмотренных в программе операций. Если программа оперирует данными во внешней памяти, то в ее исходном коде, а для некоторых операционных систем и в языках управления заданиями, должны описываться свойства файлов, в которых хранятся эти данные. Указанные описания, несомненно, также являются примерами видов метаданных.

Рождение технологий баз данных и информационно-поисковых систем также потребовало использования метаданных. При проектировании конкретной базы данных необходимо сформировать концептуальную схему предметной области, представляя ее средствами какого-либо языка концептуального моделирования. На ее основе нужно создать описание структуры базы данных, ограничений целостности, полномочий пользователей и т.д. Для этой цели используется язык описания данных СУБД, выбранной для реализации системы базы данных. Указанное описание представляется в виде схемы создаваемой базы данных, которая, также как и концептуальная схема предметной области, является метаданными.

В ранних документальных информационно-поисковых системах каждый документ представлялся его поисковым образом, который идентифицировал этот документ, характеризовал его содержание набором дескрипторов и заменял этот документ в процессах обработки пользовательских запросов. Поисковый образ исходного документа является его метаданными. Метаданными текстовых публикаций в электронных библиотечных каталогах служат также их библиографические описания и аннотации, рубрики тематических классификаторов, к которым относится содержание данных публикаций, и другие их характеристики.

В 1960-1970 гг. были довольно широко распространены генераторы отчетов с разработанным и впервые реализованным компанией IBM входным языком PRG (Report Program Generator). Этот язык позволяет описывать формат генерируемого отчета. Такие описания, несомненно, являются метаданными.

Появление гипертекстовых технологий, а затем и Всемирной паутины, предусматривает использование гипертекстовой разметки исходного текста в гипертекстовых издательских системах и в веб-страницах. Совокупность тегов разметки также представляет собой метаданные таких ресурсов.

Создание систем, основанных на знаниях, и технологий Семантического Веба, позволило явным образом представлять и использовать для поиска и логического вывода семантику данных, а также онтологию предметной области. Для их описания в последние годы активно используются, в частности, язык RDF, языки описания онтологий RDFS, OWL и OWL2, а также профили языка OWL2. Такие описания также являются примерами метаданных.

В научных информационных системах метаданные позволяют описывать свойства используемых в них данных, связанные с особенностью предметной области исследований. Вместе с тем, часто необходимы описания характеристик научных приборов, с помощью которых они получены, мест и времени регистрации этих данных, компьютерных модельных экспериментов и т.п.

Метаданные необходимы в современных электронных библиотеках. Они описывают контент библиотеки в целом, составляющие его коллекции информационных ресурсов, отдельные ресурсы, содержащиеся в коллекциях, их классификаторы, связи между ними, организации, владеющие информационными ресурсами библиотеки, профили

их авторов и пользователей библиотеки, ряд других информационных объектов и сервисов таких систем.

Осознание необходимости метаданных привело к созданию инструментария управления этим специфическим видом информационных ресурсов. Впервые серьезное внимание проблемам управления метаданными начало уделяться в 1970-е годы в контексте информационных систем, оперирующих структурированными данными. Была предложена концепция системы словаря-справочника данных. Появился ряд систем такого рода [5, 6]. Несколько позднее была реализована концепция интегрированных словарей-справочников данных – систем управления базами данных, выполняющих традиционные функции СУБД и одновременно функции системы словаря-справочника данных. Был разработан также международный стандарт для систем словарей-справочников данных [7].

Идея интеграции системы словаря-справочника данных и СУБД, хотя и в ограниченной форме, позднее нашла поддержку в стандарте языка SQL. В нем предусмотрено представление схемы базы данных (на «логическом» уровне) в форме набора системных таблиц базы данных, которыми можно оперировать с помощью обычных средств языка. В результате базы данных, которые управляются СУБД, основанными на стандартах SQL, стали самоописываемыми – они содержат пользовательские данные вместе с описывающими их метаданными (схемой базы данных).

Развитие информационных технологий и сферы их применения привело к существенному расширению функций метаданных и их многообразия. Средства представления метаданных и управления ими созданы и развиваются как для информационных, так и для других компьютерных систем. При этом содержание метаданных, их функции и средства их представления зависят от используемых информационных технологий, функциональных возможностей и предметной области использующих их систем, природы описываемых ресурсов, контекста и характера их использования, а также от многих других факторов.

В последние два десятилетия метаданные стали привлекать большое внимание, главным образом, в связи с развитием технологий Семантического Веба, технологий электронных библиотек и ряда других новых пластов информационных технологий. Кроме того, стала актуальной необходимость обеспечения обмена метаданными между различными системами, обеспечения интероперабельности и повторного использования информационных ресурсов, интеграции данных из многих источников. Все это вызвало активную деятельность по стандартизации метаданных, осуществляемую официальными международными и национальными органами стандартизации, промышленными консорциумами, научными и другими сообществами. В результате были созданы многочисленные стандарты описания метаданных «горизонтальной» и «вертикальной сферы». Активно используются на практике стандарты платформы XML [8, 9], Дублинское ядро [10-12], дескриптивное подмножество языка SQL, большое число схем метаданных для различных сфер применения, стандарты языков концептуального и онтологического моделирования, многочисленные стандарты научных метаданных и многие другие.

Хотя использование термина *метаданные* существенно активизировалось в последние годы, к сожалению, все еще не сложилось однозначное понимание этого термина. Широко распространенная абстрактная формула «*метаданные – это данные о данных*» не раскрывает многообразия свойств и функций этого вида информационных ресурсов. Метаданным в последние годы посвящено огромное количество публикаций. Однако в них мало внимания уделяется систематическому обсуждению их общих свойств и функций. Чаще всего обсуждаются проблемы, связанные с созданием систем метаданных для конкретных областей научных исследований, бизнеса, электронных

правительств, разработок электронных библиотек, для разнообразных репозиторий цифровых информационных объектов, информационных систем для конкретных сфер применения. При этом во многих работах, посвященных электронным библиотекам или приложениям, основанным на стандартах Семантического Веба, доминирует весьма ограниченный взгляд на функции метаданных лишь как на средство описания контента (содержания) информационных ресурсов, хотя это – только одна из многих возможных функций метаданных.

Формирование адекватного и достаточно полного представления об общих свойствах и функциях метаданных является, на наш взгляд, актуальной проблемой. В данной работе предлагается систематический взгляд на метаданные как на информационные ресурсы особого рода, обсуждается вопрос об определении этого термина, приводятся примеры метаданных, используемых в различных областях информационных технологий, рассматриваются общие (независимые от сферы применения) свойства метаданных и их функции, выразительные средства для их представления, известные обобщенные классификации метаданных, а также деятельность по их стандартизации.

2. Об определении термина *метаданные*

Как уже отмечалось, сосуществуют различные точки зрения на содержание термина *метаданные*. В работе [13] отмечается, что неоднозначность в понимании смысла и функций метаданных обусловлена сосуществованием доминирующих в настоящее время двух подходов. Первый из них сложился в библиотечном сообществе. Он связан с созданием и использованием документальных ИПС, электронных каталогов библиотечных фондов, а в последние годы – с разработками репозиторий цифровых объектов, представляющими собой основную проблематику электронных библиотек. При этом представители библиотечного сообщества имеют в виду, главным образом, текстовые информационные объекты. Истоки этого подхода авторы [13] видят в библиотечной науке (*Library Science*). В отличие от него, истоки второго подхода, по их мнению, относятся к «компьютерным наукам» (*Computer Sciences*). К этой сфере, относятся развивающиеся уже около полувека технологии баз данных и другие области информатики, связанные с управлением данными и знаниями.

Указанные подходы имеют дело с различного рода информационными ресурсами, для управления которыми в разных сферах их использования и в различных классах приложений необходимы разнообразные виды метаданных. Это многообразие видов и контекстов их использования и обуславливает в значительной мере разнообразие трактовок смысла и свойств метаданных.

Существуют и другие причины отсутствия устоявшегося единого понимания содержания термина *метаданные*. Недостаточно тщательны или компетентны авторы многих публикаций, посвященных метаданным, которые вводят в заблуждение читателей. Так, многочисленны примеры, когда авторы публикаций рассматривают частный вид метаданных и, не делая на этот счет никаких оговорок, тем самым провоцируют у читателей обобщение их свойств, присущих лишь рассматриваемому частному случаю. Разнообразные системы метаданных создаются большим количеством профессиональных сообществ в разных областях деятельности, имеющих собственные представления о существе метаданных.

В связи с большим разнообразием видов, свойств и функций метаданных дать лаконичное и корректное определение этого термина довольно непросто. Большинство встречающихся в литературе определений подчеркивает какие-либо отдельные функции или свойства метаданных и не обладает достаточной общностью. Широко употребляемое

определение «Метаданные – это данные о данных» слишком абстрактно и, кроме того, не учитывает, что в настоящее время метаданные служат для описания свойств не только данных, но и многих других видов ресурсов информационных технологий. Неудовлетворительно, на наш взгляд, и определение «Метаданные – это данные, относящиеся к какому-либо элементу данных, т.е. метаданные – это данные о данных», которое приведено в солидном энциклопедическом издании [14]. Это определение, как и предыдущее, не охватывает метаданных ресурсов, отличных от данных, использует не определенное ясным образом понятие *элемента данных*. В нем нет и ответа на важный вопрос, что означает «относящиеся». Интересно заметить, что в другом авторитетном энциклопедическом издании [15] вообще отсутствует какое-либо определение термина *метаданные*.

Приведем еще несколько определений, отражающих различные точки зрения на метаданные, их назначение и свойства:

- Метаданные – это информация, которая делает данные полезными [16].
- Метаданные – это машино-обрабатываемые данные, которые описывают некоторые ресурсы, цифровые и нецифровые [17].
- Метаданными называется допускающая компьютерную обработку и интерпретацию человеком информация о цифровых и нецифровых объектах [18].
- Метаданные – это структурированные данные, которые описывают характеристики некоторого ресурса [19].
- Метаданными называется структурированная информация, которая описывает, поясняет, указывает местоположение и иным образом облегчает поиск, использование информационного ресурса, а также управление им [20].
- Метаданные – это структурированные данные, представляющие собой характеристики описываемых сущностей для целей их идентификации, поиска, оценки, управления ими [21].
- Метаданные – это описание не только состава данных, их структуры (формата) представления, места хранения и других их признаков, но и поддерживающих их информационных систем, технологий, пользователей, методов доступа [22].
- Во Всемирной паутине метаданные – это слабоструктурированные данные, как правило, соответствующие согласованным моделям, обеспечивающим операционную интероперабельность в неоднородной среде [23].
- Метаданными называется любая дескриптивная информация о других источниках данных, которая способствует организации, идентификации, представлению, определению местоположения, обеспечению интероперабельности, управлению и использованию этих данных [24].
- Метаданные – это данные, относящиеся к некоторому элементу данных, т.е. метаданные – это данные о данных. Для элемента данных спецификация, определяющая, каким образом он может создаваться, в каких контекстах он может использоваться, как его трансформировали или как он может интерпретироваться либо обрабатываться [25].
- Метаданные обычно определяются как «данные о данных», которые предназначены для выражения семантики информации, следовательно, для улучшения ее поиска и выборки, понимания и использования [26].
- Метаданные – это данные из более общей формальной системы, описывающей заданную систему данных [27].

Нетрудно видеть, что в приведенных определениях, как и во многих других встречающихся в литературе, чаще всего акцентируются некоторые отдельные функции или свойства метаданных. Некоторые из определений предлагаются авторами в контекстах, посвященных обсуждению конкретных видов описываемых ресурсов, классов

приложений, используемых технологий. При этом не подчеркивается, что «областью действия» этих определений являются именно эти специфические случаи. В результате приписываемые метаданным свойства или функции либо категория описываемых ресурсов неправомерно обобщаются. В большинстве определений не учитывается отмеченный ранее фактор значительного расширения множества видов ресурсов, которые описываются метаданными в используемых в настоящее время информационных технологиях. Вместе с тем, некоторые определения характеризуют принципиально важные общие свойства метаданных, которые будут рассматриваться далее.

3. Ресурсы, описываемые метаданными

При рождении термина *метаданные* имелось в виду, что они предназначены для описания свойств данных. По прошествии времени разнообразие ресурсов, описание свойств которых стало необходимым в современных компьютерных системах, существенно расширилось, хотя термин *метаданные* при этом не всегда используется.

Помимо собственно данных, требуются описания свойства систем, обеспечивающих их хранение, доступ к ним и их обработку, описания характеристик пользователей (их полномочий доступа, характера информационных потребностей и т.п.). Возникли также потребности описания концептуальных схем и онтологий различных предметных областей, веб-сервисов и их интерфейсов, средств вычислительной техники, научных приборов и других технических устройств, потоков работ, бизнес-процессов, обозначений, используемых на географических картах (легенда), а также организаций – владельцев, генераторов и пользователей информационных ресурсов.

Нужны описания не только структурированных, но и неструктурированных, а также слабоструктурированных данных. С помощью метаданных должны описываться различного рода ограничения, которым должны удовлетворять данные, доступ к ним и их использование. Метаданные в бизнесе описывают специфические для этой сферы деятельности ограничения, называемые *бизнес-правилами* [28]. Бизнес-правила характеризуют «правила игры» в бизнесе, своего рода руководства для бизнес-деятельности.

Следует упомянуть также использование метаданных для описания разрабатываемых программных систем и процессов их функционирования. Некоторые такие метаданные формируются CASE-инструментами и представляются диаграммами на языке UML. Работа [29] напоминает об использовании метаданных для описания программ, их фрагментов, различного рода процессов и их компонентов, сред функционирования систем, совокупностей событий и отдельных событий, уже отмечавшихся ранее ограничений, людей и их ролей в ИТ-системах, организаций, их подразделений, индивидуумов или индивидуумов в определенной роли.

Перечисленными примерами отнюдь не исчерпывается множество видов ресурсов, описываемых метаданными. Поэтому в дальнейшем обсуждении, если не оговаривается, что имеются в виду метаданные информационных ресурсов, предполагаются метаданные ресурсов произвольной природы.

В связи с распространением такого более широкого толкования термина *метаданные* и разнообразия их свойств стало правомерным использование термина *метаинформация*.

4. Примеры метаданных в различных областях информационных технологий

В данном разделе приведем некоторые примеры видов метаданных, используемых в различных областях информационных технологий и в некоторых конкретных предметных областях.

В технологиях баз данных:

- Концептуальные схемы предметных областей
- Схемы баз данных
- Описания междууровневых отображений схем в системах баз данных с многоуровневой архитектурой.

В технологиях виртуальной интеграции данных:

- Локальные схемы источников данных
- Глобальные схемы
- Описания отображений между локальными схемами интегрируемых источников и глобальной схемой
- Описания отображений между онтологиями локальных источников и общей онтологией
- Характеристики регистрации источников в посредниках в системах виртуальной интеграции данных.

В технологиях текстового поиска:

- Идентификаторы текстовых документов
- Наборы значений индексирующих атрибутов документов
- Библиографические описания документов
- Аннотации публикаций
- Каталоги коллекций документов
- Наборы ключевых слов документов
- Рубрики классификаторов для документов
- Наборы значений элементов метаданных Дублинского ядра, описывающие содержание документов
- Индексы ISBN для монографий.

В веб-технологиях:

- Гипертекстовая разметка веб-страниц
- Наборы имен и значений параметров тегов META в веб-страницах
- Разметка фрагментов веб-страниц средствами микроформатов (hCard, hReview, hProduct, hRecipe и др.)
- Описания типов XML-документов (DTD)
- XML-схемы для типов XML-документов
- RDF-спецификации ресурсов
- Описания онтологий на языке OWL или OWL2
- Семантические аннотации веб-страниц или их фрагментов.

В CASE-технологиях:

- UML-диаграммы спецификаций разрабатываемых систем

В технологии веб-сервисов:

- Описания интерфейсов веб-сервисов средствами языка WSDL
- Описание характеристик веб-сервисов в регистре UDDI.

В виртуальной обсерватории:

- Наборы характеристик наблюдаемых участков звездного неба
- Наборы характеристик телескопов
- Период времени проведения наблюдения.

В системах управления транспортом:

- Схемы маршрутов городского пассажирского транспорта
- Карты метрополитенов
- Совокупности обозначений для конкретных объектов на картах.

В системах баз данных временных рядов экономических показателей:

- Описания методик исчисления показателей либо порядка проведения их натуральных измерений
- Интервалы времени проведения измерений (начальные и конечные годы, кварталы, месяцы и т.д.)
- Периодичность рядов экономических показателей (годовой, квартальный, месячный, дневной)
- Единицы измерения показателей
- Наименования отраслей хозяйства, к которым относятся данные показатели.

В информационном пространстве Соционет [30, 31], основанном на технологии открытых архивов OAI [32, 33]:

- Описания информационных объектов
- Описания семантических связей между информационными объектами
- Описания коллекций информационных объектов
- Профили авторов информационных объектов или пользователей архива
- Профили организаций–поставщиков информационных ресурсов
- Описания интегрированных в систему открытых архивов.

Приведенные примеры убедительно демонстрируют существование большого разнообразия видов метаданных, а также областей информационных технологий, в которых они используются.

5. Свойства метаданных

По свойствам и назначению различаются метаданные «горизонтальной» сферы («универсальные») и «вертикальной» сферы («специализированные»). «Универсальные» метаданные, в отличие от метаданных «вертикальной» сферы, могут использоваться в разных областях информационных технологий и в разнообразных приложениях. В данной работе обсуждаются именно метаданные первой из этих категорий.

Конкретные состав и свойства метаданных зависят от особенностей использующей их системы, от характера описываемых метаданными ресурсов, от используемых информационных технологий, потребностей пользователей системы и от многих других факторов. В этом разделе рассматриваются общие свойства метаданных, не специфичные для какого-либо конкретного случая их применения или вида описываемых ресурсов.

Явное представление описания свойств ресурса. Метаданные описывают свойства различных ресурсов *явным образом*. Они представляются в форме спецификаций на специальных языках, текстовых документов, графических схем, таблиц в базах данных, информационных объектов в различных ИТ-продуктах, например, в репозиториях метаданных CASE-инструментов или открытых архивов и др. Вместе с тем, некоторые свойства рассматриваемых ресурсов могут подразумеваться и не представлять явными описаниями. Подобное подразумеваемое описание не может рассматриваться как метаданные.

Относительный характер разделения информационных ресурсов на данные и метаданные. Разделение информационных ресурсов на данные и метаданные является *относительным*. Информационные ресурсы, играющие роль метаданных в одних случаях,

являются данными в других, и наоборот. Действительно, совокупность тегов гипертекстовой разметки веб-страницы, название публикации или ее аннотация, содержащаяся в ее тексте, могут использоваться в качестве метаданных, описывающих перечисленные ресурсы. Вместе с тем, все они являются составными частями соответствующих ресурсов и тем самым на стадии использования ресурсов являются данными.

Способность описывать разнообразные виды ресурсов. Как уже отмечалось, метаданные могут использоваться для описания свойств не только данных, но и ресурсов многих других видов. В данной статье, однако, основное внимание уделяется метаданным, служащим для описания информационных ресурсов.

Метаданные представляются средствами более абстрактной системы понятий по отношению к системе понятий описываемого ресурса. Метаданные описывают данные и любые другие ресурсы в терминах более абстрактной системы понятий. Такие системы понятий и их связей воплощают модели более абстрактного уровня (метамодели) по отношению к моделям, в терминах которых представлены описываемые ресурсы. Так, в системе базы данных описание хранимых данных осуществляется не средствами значений этих данных, а в таких категориях, как типы хранимых записей и элементов данных, ключи записей, указатели, т.е. в системе понятий модели данных среды хранения базы данных. Семантика текстовых документов может описываться в терминах классов, атрибутов классов и отношений между классами, определенных в онтологии соответствующей предметной области.

Метаданные независимые и зависимые от области применения. Метаданные первой из этих категорий («универсальные» метаданные) обычно описываются выразительными средствами стандартов метаданных «горизонтальной» сферы (стандартов, не ориентированных на какую-либо конкретную область применения). Метаданные второй категории, напротив, применяются в конкретных областях применения, зависят от специфики предметной области и характеризуются большим разнообразием. Существуют, например, собственные системы метаданных в различных областях научных исследований, в музейном деле, в образовательной сфере и т.д. Для представления метаданных в некоторых областях применения разработаны специальные стандарты (стандарты «вертикальной» сферы).

Зависимость состава метаданных от вида описываемых ресурсов. Состав метаданных, описывающих те или иные ресурсы, зависит от вида этих ресурсов. Действительно, метаданные реляционной базы данных, например, описывают, из каких таблиц она состоит, структуру каждой таблицы, типы данных, содержащихся в каждом столбце таблицы, ограничения целостности, ограничения доступа и т.д. Для другого вида ресурсов – аудиофайла формата *mp3* – состав метаданных, представляемых его тегами совсем иной – названия альбома и дорожек, число дорожек, жанр, авторы и исполнители, год выпуска и другие характеристики.

Состав метаданных для описания ресурсов конкретной создаваемой системы должен проектироваться. Точно так же, как при создании информационных систем разработчик определяет состав поддерживаемых в них пользовательских информационных ресурсов, с учетом потребностей пользователей и системных механизмов должен проектироваться и состав метаданных, описывающих ресурсы системы. При использовании стандарта метаданных состав метаданных определяется спецификацией этого стандарта.

Зависимость состава метаданных в информационной системе от уровня информационной архитектуры, к которому относятся описываемые ресурсы. Действительно, метаданные, описывающие веб-страницы на уровне хранения,

представляют собой совокупность тегов гипертекстовой разметки, которую можно видеть в исходном коде страницы. В то же время, на «логическом» уровне, в том виде, как браузер воспроизводит эту страницу на экране компьютера, с этой страницей ассоциируются совсем иные метаданные, например, ее URL и описание ее контента в различного рода каталогах пользователей Веба. В системах баз данных метаданные среды хранения описывают структуру хранимых данных и ее свойства, индексы, характеристики пространства памяти и другие элементы среды хранения. Метаданные на «логическом» уровне представления данных описывают видение базы данных пользователем. В реляционной системе базы данных метаданные «логического» уровня описывают состав таблиц базы данных, структуру каждой таблицы, ограничения целостности, триггеры, полномочия пользователей и др., но не организацию хранения базы данных.

Возможность описания ресурса с различной степенью гранулярности. Метаданные могут описывать информационные и другие ресурсы на любом требуемом уровне их гранулярности. Например, в электронных библиотеках используются метаданные, описывающие коллекции информационных ресурсов в целом, а также отдельные входящие в них информационные объекты. В системах баз данных схема базы данных описывает не только структуру базы данных в целом, но и свойства ее структурных компонентов – в реляционных базах данных отдельных таблиц, данных, содержащихся в их столбцах и т.д. Описание монографии на веб-сайте включает библиографическое ее описание и может включать или не включать аннотацию, оглавление и другие компоненты метадаанных.

Автономные и встроенные метаданные. Метаданные могут быть *автономными* (отчужденными) от описываемого ресурса или *встроенными* в него. Примеры *автономных метадаанных*: схемы баз данных (они отчуждены от описываемого ими наполнения базы данных); DTD [8], описывающие тип XML-документов и хранимые в Вебе вне конкретных экземпляров XML-документов; репозитории метадаанных открытых архивов в технологии открытых архивов OAI [32, 33]; полнотекстовые индексы коллекций текстовых документов электронных библиотек. Примерами *встроенных метадаанных* могут служить: HTML-разметка веб-страниц; разметка веб-страниц средствами микроформатов; DTD, специфицированные в теле XML-документов; аннотации статьи, содержащиеся в их тексте; глоссарии, содержащиеся в текстах спецификаций официальных стандартов. Встроенными метадаанными являются также теги мультимедийных файлов. Информационный ресурс со встроенными описывающими его метадаанными естественно называть *самоописываемым* по отношению к тому набору свойств ресурсов, которые этими метадаанными описываются.

Метаданные независимые и зависимые от контента описываемых информационных ресурсов. Различаются метаданные, *независимые от контента* описываемого информационного ресурса и *зависимые от контента*. Примерами метадаанных, независимых от контента, могут служить: дата создания и место хранения текстового документа; тип и имя файла, в котором он содержится; URL веб-страницы. Примеры метадаанных, зависимых от контента описываемого ресурса: аннотация текста; его статистические характеристики (частотные характеристики вхождений слов словаря, длина текста в литеррах и т.п.). Некоторые из таких метадаанных могут продуцироваться автоматически из текста данного документа, другие формируются вручную.

Метаданные могут быть структурированными, неструктурированными и слабоструктурированными данными. Вопреки сложившимся представлениям, метаданные могут быть в различной степени структурированными. Для поддержки системных функций, как правило, используются структурированные метаданные. Метаданные, предназначенные для пользователей, могут быть как структурированными,

так и неструктурированными либо слабоструктурированными, например, они могут быть обычными текстами. Они могут быть и слабоструктурированными – например, представленными в виде веб-страницы. В отличие от структурированных метаданных, для неструктурированных метаданных не существует каких-либо общепринятых стандартов их представления. Наибольшее внимание привлекают структурированные метаданные. Им посвящена масса публикаций.

Метаданными могут описываться свойства как структурированных, так и неструктурированных и слабоструктурированных данных. Во многих публикациях утверждается, что метаданные могут существовать только для структурированных данных. Однако в действительности метаданными можно снабдить информационные ресурсы любого из указанных видов. Каждый из этих видов информационных ресурсов в различных приложениях может описываться любым видом метаданных – структурированными, неструктурированными либо слабоструктурированными.

Статические и динамические метаданные. Метаданные могут быть как статическими, так и динамическими. Так, в системах баз данных относительно неизменна схема базы данных. В дескрипторных ИПС неизменными остаются поисковые образы документов. В то же время каталоги коллекций в электронной библиотеке изменяются при включении в них новых информационных объектов.

Формализованные и неформализованные метаданные. Для представления метаданных могут служить в различной степени формализованные выразительные средства – естественные языки, полуформальные языки, например, набор элементов метаданных Дублинского ядра, а также формальные языки. В дедуктивных базах данных, например, интенционал описывается средствами логики. На дескриптивной логике основан язык OWL для описания онтологий Веба.

Семантические метаданные. Существует особый вид метаданных, которые служат для описания семантики ресурсов. Такие метаданные называются *семантическими*. Они могут быть формализованными или неформализованными.

Многоуровневость метаданных. Метаданные, как уже отмечалось, представляют собой особый вид информационных ресурсов. Может существовать необходимость и их описания метаданными более высокого уровня. Иерархия метаданных может включать любое число уровней. Так, стандарт Meta Object Facility (MOF) [34] консорциума OMG предусматривает три уровня, спецификация стандартов Дублинского ядра [10, 12] – два уровня: предназначенный непосредственно для практических целей и обеспечивающий его описание в спецификации стандарта. Еще один пример. Онтология предметной области представляет собой ее метаданные. Но если использовать онтологию для семантического аннотирования схемы базы данных, то в этой роли онтология представляет собой мета-метаданные. Метаданные более высоких уровней иерархии принято обозначать терминами *мета-метаданные, мета-мета-метаданные* и т.д.

Системные и пользовательские метаданные. Метаданные могут быть предназначены для поддержки функционирования системных механизмов (*системные метаданные*) либо для информирования пользователей информационной системы о свойствах описываемых ресурсов или инструктирования их по правилам использования различных системных сервисов (*пользовательские метаданные*). Существуют системы, служащие для поддержки всей совокупности метаданных информационной системы, называемые *системами словарей-справочников данных* [5, 6]. Их компонент *справочник данных* управляет метаданными, предназначенными для системы, а *словарь данных* – метаданными, предназначенными для пользователей.

Возможность использования различных средств представления метаданных. Для представления метаданных могут использоваться разнообразные выразительные

средства: естественные и искусственные языки, наборы элементов метаданных (схемы метаданных), графические схемы и диаграммы, а также другие средства.

Выше были рассмотрены основные свойства метаданных, которые не зависят от области их применения. Разнообразие специфических систем метаданных, используемых в конкретных областях, чрезвычайно велико. Это касается, прежде всего, научных исследований. Научные метаданные позволяют описывать многообразие свойств исследуемых объектов, явлений, процессов, характеристики измерительных приборов, обеспечивающих получение научных данных, разнообразных компьютерных и иных моделей, в соответствии с которыми вычислены или оценены значения используемых данных, публикации, источники информационных ресурсов и многое другое. Такие специфические метаданные обладают особыми свойствами, которые представляют интерес в соответствующем контексте.

6. Функции метаданных

Метаданные выполняют разнообразные функции в системах, их использующих. Конкретные функции и состав метаданных, как и их свойства, существенным образом зависят от информационных технологий, на которых базируется компьютерная система, от ее функциональности, свойств поддерживаемых в ней информационных ресурсов, способов их организации в системе, особенностей задач их обработки, потребностей пользователей и от многих других факторов. Рассмотрим кратко основные функции, выполняемые метаданными.

Описание абстрактных моделей предметной области информационной системы. С разработки этих моделей начинается проектирование информационных систем. Это - концептуальная схема предметной области, дополняемая в системах семантического уровня онтологией [35]. Абстрактные модели не зависят от технологий реализации разрабатываемой системы. Для спецификации концептуальной схемы используются языки концептуального моделирования различного уровня формализованности, например, диаграммы UML или формальные языки представления знаний, основанные на математической логике. Тем самым концептуальная схема может в различной степени удерживать семантику предметной области. Онтологии также могут представляться различными средствами – от традиционных классификаторов до формальных языков описания онтологий. В последние годы языки описания онтологий OWL и OWL2 консорциума W3C стали иногда использоваться для описания концептуальной схемы предметной области.

Описание даталогических моделей в системах баз данных и в других репозиториях структурированных данных. В отличие от абстрактных моделей, даталогические модели предметной области, используемые в технологиях баз данных и управления структурированными данными, зависят от технологий реализации системы. Такие модели в системах баз данных представляются схемами базы данных различных уровней архитектуры системы базы данных – концептуальной схемой базы данных, внутренней схемой, внешними схемами.

Описание представления данных на разных уровнях информационной архитектуры. В системах баз данных метаданные описывают представление хранимых данных (внутренняя схема базы данных), «логическое» представление полной базы данных (концептуальная схема базы данных) и представления для пользователей или приложений необходимых им фрагментов базы данных или полной базы данных (внешние схемы). Такие метаданные определяются на языках описания данных используемых СУБД. В среде Веба разметка хранимых HTML-файлов описывает, какой формируемый

веб-браузером вид должна иметь на экране пользовательского компьютера веб-страница и какие другие файлы из Интернета нужно для этого использовать.

Описание источников данных. Эти функции метаданных выполняют, в частности, схемы баз данных, определения типа XML-документов и XML-схемы для репозитория XML-документов, онтологии источников информации, репозитории метаданных открытых архивов, каталоги веб-сайтов в поисковых машинах Веба, каталоги коллекций информационных объектов в электронных библиотеках. Метаданные этого вида представляются языками описания данных (например, дескриптивный подязык языка SQL, язык описания объектов ODL объектных баз данных), подязык языка XML и язык XML-схем, языками описания онтологий (например, RDFS, OWL2) и другими средствами.

Описание отдельных информационных объектов. Эта функция необходима в электронных библиотеках, системах открытых архивов и других репозиториях информационных объектов с гранулярностью доступа к их информационным ресурсам на уровне отдельного объекта. Для представления таких метаданных часто используются описатели, представленные в форме XML-документов, библиотечных каталожных записей в различных диалектах стандарта MARC. Эти метаданные описывают структурные свойства описываемых информационных объектов, различные другие их характеристики. В «интеллектуальных» системах описывается семантика информационных объектов (см. далее). Например, может использоваться семантическая аннотация отдельных страниц веб-сайтов.

Описание семантики источников информации, отдельных информационных объектов или их фрагментов. Как уже отмечалось, метаданные, используемые для этой цели, называются *семантическими*. В простейших случаях семантика информационных источников или отдельных информационных объектов описывается с помощью рубрикаторов, например, российских ГРНТИ или ББК, либо международных УДК или JEL (Journal of Economic Literature Classification System). Другим более многоаспектным, но неформальным средством описания семантики является набор элементов метаданных Дублинского ядра (Dublin Core, DC). Как и другие средства представления знаний, для формального описания семантики информационных источников или предметной области системы могут использоваться онтологии. В настоящее время для представления онтологий чаще всего служат разработанные консорциумом W3C языки RDF, RDFS, OWL, OWL2 и его профили.

Для описания семантики текстовых документов и коллекций информационных ресурсов электронных библиотек, структурных элементов баз данных, веб-страниц, а также электронных библиотек или других информационных репозиториях в целом используют также *семантическое аннотирование*. *Семантическая аннотация* - это ассоциируемые с описываемым ресурсом семантические метаданные [36]. Семантическая аннотация может быть встроена в описываемый ресурс или автономна. Одним из простейших средств встроеного семантического аннотирования, используемого для описания семантики фрагментов веб-страниц, являются так называемые *микроформаты* [37]. Некоторые из них поддерживаются популярными поисковыми машинами Веба. Например, Яндекс поддерживает микроформаты: hCard - для разметки контактной информации (адресов, телефонов и т. д.); hRecipe - для описания кулинарных рецептов; hReview - для разметки рецензий, отзывов; hProduct - для разметки описаний товаров.

Для семантического аннотирования могут использоваться различные средства - от естественных до формальных искусственных языков. Могут использоваться, например, управляемые словари и тезаурусы, таксономии и формальные онтологии. Частным случаем семантического аннотирования является *онтологическое аннотирование*. В этом случае описание семантики ресурса осуществляется в терминах некоторой онтологии.

Следует отметить, что вопреки мнению, высказанному в [38] и ряде других публикаций, метаданные могут описывать семантику не только структурированных, но и *неструктурированных данных*. В указанной работе утверждается: «Структурированные данные обладают метаданными, которые позволяют нам понять, что они означают. Для неструктурированных данных не существует какого-либо эквивалента». Это утверждение ошибочно. Действительно, что для любых информационных ресурсов, структурированных и неструктурированных, могут существовать семантические данные, например, в форме семантической аннотации.

Идентификация описываемых ресурсов. Метаданные описывают, какие встроенные в экземпляры ресурсов их элементы или ассоциированные с ними атрибуты обеспечивают их идентификацию. Примерами встроенных идентифицирующих элементов являются первичные и вторичные ключи строк таблиц в системах баз данных. Их значения идентифицируют строки уникальным образом или, соответственно, некоторые их множества. Элементы XML-документов при наличии DTD или XML-схемы идентифицируются их встроенными идентификаторами. В электронных библиотеках текстовых документов также используются их уникальные идентификаторы. Часто они генерируются системой, недоступны в явном виде пользователям, и служат для ссылок на эти документы в генерируемых системой каталогах. Для идентификации текстовых документов могут также использоваться наборы значений дескрипторов или ключевых слов. В качестве уникальных идентификаторов веб-страниц используются указатели URL или URI [39]. В геоданных точки земной поверхности идентифицируются их широтой и долготой. Следует упомянуть также индексы УДК, ISBN и ISSN, идентификаторы цифровых объектов DOI, элементы метаданных Дублинского ядра. Существуют и другие примеры видов метаданных, используемых для идентификации информационных ресурсов.

Предоставление пользователям сведений об описываемых ресурсах. Метаданные могут быть предназначены для информирования пользователей о свойствах и правилах использования информационных ресурсов, сервисов, которые обеспечивают доступ к ним, и о других описываемых ресурсах (*пользовательские метаданные*). Ориентированные на пользователя метаданные представляются в виде, воспринимаемом человеком, и могут быть как структурированными, так слабоструктурированными или неструктурированными данными. Последние могут представляться в виде текстов на естественных языках, схем, чертежей, фрагментов карт местности и др. Метаданные рассматриваемого назначения могут описывать генезис описываемых ресурсов, связанные с ними предположения, способы доступа и условия их использования, их семантику, используемую терминологию, а также иные сведения. Примерами таких метаданных являются инструкции для пользователей по работе с системой, оглавление или аннотация книги.

Описание ресурсов, предназначенное для поддержки функционирования системных механизмов. Метаданные могут быть предназначены для поддержки выполнения системных функций (*системные метаданные*). Чаще всего это структурированные данные, представленные в формате, удобном для компьютерной обработки.

Верификация данных. Метаданные структурированных данных (например, схема базы данных) позволяют контролировать их форматы и типы данных, проверять соблюдение ограничений целостности. Эти задачи решают механизмы СУБД. Для XML-документов с заданным определением типа документов (DTD) используемый XML-процессор проверяет корректность их структуры. Онтология SQL-системы базы данных в системах доступа к данным, основанная на онтологиях [40], позволяет обнаруживать неполноту и противоречивость данных в базе данных.

Описание ограничений доступа к информационным ресурсам, обеспечение информационной безопасности. Метаданные позволяют описывать ограничения доступа и полномочия пользователей на доступ к информационным ресурсам или на выполнение иных операций используемой системы. Эти полномочия и ограничения проверяются системными механизмами при попытке пользователя выполнить защищенную операцию. В системах баз данных ограничения доступа описываются в схеме базы данных. Цифровые подписи, сертификаты безопасности, открытые и закрытые ключи, а также другие средства аутентификации пользователей и подтверждения подлинности сообщений, также представляют собой метаданные, служащие для обеспечения информационной безопасности.

Описание характеристик пользователей и их информационных потребностей. Во многих информационных системах поддерживаются сведения о пользователях и их информационных потребностях, необходимые для учета их работы, сбора статистических данных об использовании системных ресурсов, а также для проверки их полномочий на запрашиваемые действия системы. Совокупность таких метаданных называется *профилем пользователя*.

Управление структурированными данными и другими информационными ресурсами. Это комплекс задач, связанных с обеспечением эффективного хранения и обработки информационных ресурсов, а также доступа к ним. Так, метаданные, представленные в виде схемы баз данных, обеспечивают механизмы СУБД в процессе ее функционирования необходимой информацией об организации базы данных, ограничениях целостности данных и управления доступом.

Обнаружение и поиск информационных ресурсов. Это - одна из важных функций метаданных, предусматривающая использование их в критериях поиска. При этом могут использоваться не только идентифицирующие, но и другие виды метаданных, например, полнотекстовые индексы в системах текстового поиска. Обнаружение требуемых информационных ресурсов может также осуществляться путем навигации в каталогах, предметных указателях, рубрикаторах. Использование семантических метаданных в процессе поиска позволяет существенно снизить уровень информационного шума.

Распространение информационных ресурсов. Поддерживаемые в информационной системе профили пользователей характеризуют их информационные потребности, и на этой основе может осуществляться избирательное распространение информации средствами системы.

Тематическая систематизация коллекций информационных ресурсов. Эта функция выполняется путем формирования каталогов и предметных указателей содержащихся в системе информационных ресурсов, индексирования ресурсов с помощью тематических рубрикаторов.

Обеспечение интероперабельности и повторного использования информационных ресурсов. Техническая и/или семантическая интероперабельность и на этой основе повторное использование информационных ресурсов может обеспечиваться путем унификации представления метаданных средствами открытых стандартов. Например, база данных, описанная схемой, представленной на дескриптивном подязыке языка SQL или на языке описания объектов ODL стандарта ODMG, может повторно использоваться другой СУБД, поддерживающей стандарт SQL или, соответственно, ODMG.

Интеграция данных из многих источников. Существование метаданных, описывающих интегрируемые источники данных, – необходимое условие создания систем интеграции. В последние годы активно разрабатываются системы виртуальной интеграции данных из множества источников. Для их разработки необходимы метаданные, которые

описывают: локальные схемы интегрируемых источников, глобальную схему, отображения между локальными схемами и глобальной схемой. В системах семантической виртуальной интеграции данных необходимы также метаданные, включающие онтологии локальных источников, общую онтологию системы интеграции, а также отображения локальных онтологий в общую. В системах виртуальной интеграции данных с архитектурой адаптеров-посредника необходимы также метаданные, которые описывают регистрационные данные источников в посреднике – центральном компоненте архитектуры таких систем.

Наукометрия в научных электронных библиотеках и индексах цитирования.

Идентифицирующие метаданные (автор, организация, название работы), описания рубрик классификаторов и связей цитирования в метаданных научных публикаций в электронных библиотеках позволяют оценивать цитируемость публикаций отдельных авторов или организаций, в которых они работают, анализировать тематическую структуру контента библиотеки, формировать рейтинги востребованности и цитируемости работ по авторам и по отдельным публикациям. При поддержке классификации связей между публикациями, авторами и публикациями, организациями и авторами, и осуществляемой на ее основе семантической структуризации контента научных электронных библиотек с поддержкой метаданных, описывающих сформированную семантическую структуру [41], можно получать дополнительную многоаспектную наукометрическую и науковедческую информацию.

Защита авторских прав на интеллектуальную собственность. Эту функцию поддерживают метаданные, предоставляющие информацию об авторских правах на описываемые ресурсы.

Представленное выше обсуждение функций и свойств метаданных показывает, что следующие часто встречающиеся в литературе утверждения являются заблуждениями либо справедливы лишь отчасти:

- Метаданные могут быть только у структурированных данных.
- Метаданные – сами являются структурированными данными.
- Метаданные для Веба являются слабоструктурированными данными.
- Функция метаданных - описание семантики информационных ресурсов.
- Семантические метаданные могут использоваться только для структурированных данных.

Рассмотренные в этом разделе функции метаданных используются наиболее часто. Однако в некоторых компьютерных системах они могут выполнять и иные специфические функции, обсуждение которых не входит в задачу этой работы.

8. Классификации метаданных

Возможны различные подходы к классификации метаданных. Имеют право на существование классификации метаданных по их функциям, по уровням семантической абстрактности, по их свойствам, по уровням информационной архитектуры, к которым относятся описываемые ресурсы, а также по ряду других критериев. Выше фактически были рассмотрены детальные классификации метаданных, независимых от предметной области применения, по их функциям, а также по их свойствам. Однако в литературе чаще всего предлагаются агрегированные функциональные классификации.

Пожалуй, наиболее популярна классификация, рассмотренная в [20]. С небольшими вариациями она приводится и во многих других источниках. В этой классификации различаются *описательные, структурные и административные метаданные.*

Описательные метаданные описывают контент ресурса (например, это - набор значений элементов метаданных Дублинского ядра), библиографические его данные (если это – публикация), аннотацию, идентификаторы ресурса (например, URI [39] или DOI [42]) и т. п.

Структурные метаданные характеризуют общую структуру ресурса и ее компоненты, объем и другие подобные свойства описываемого ресурса.

Административные метаданные описывают даты создания и обновления ресурса, кем он создан или модифицирован, кто владелец прав на этот ресурс, полномочия доступа пользователей, сведения об имеющихся версиях и/или копиях ресурса, где они хранятся и другие сведения, необходимые для администрирования и управления ресурсом.

Авторы отчета [20] полагают при этом, что административные метаданные состоят из нескольких подмножеств, два из которых иногда рассматриваются как самостоятельные классы метаданных. Это – метаданные управления правами на интеллектуальную собственность, а также метаданные, описывающие различные аспекты деятельности, связанной с сохранением ресурсов.

В материалах проекта [43], посвященного созданию цифровых архивов, предлагается та же классификация. Однако в состав административных метаданных авторы включают: технические метаданные (описание ИТ-аспектов информационных ресурсов и используемых информационных технологий), метаданные управления правами (декларации прав на ресурс и ограничения доступа к нему), а также метаданные цифрового происхождения (сведения о создании и последующей обработке цифрового ресурса, в том числе, подробности об ответственности за каждое событие на протяжении его жизни).

В работе [29] предлагается несколько более расширенная классификация, предусматривающая следующие классы метаданных.

Административные метаданные. Служат для управления и администрирования электронными коллекциями и другими информационными ресурсами. Предоставляют информацию о сборе данных, их местоположении, правах и способах репродуцирования ресурсов, условиях легального доступа.

Описательные метаданные. Используются для идентификации и описания свойств коллекций и связанных с ними информационных ресурсов. К таким метаданным относятся каталожные записи, информация для помощи при поиске, описания версий ресурсов, специализированные индексы, кураторская информация, гипертекстовые связи между ресурсами, аннотации создателей и пользователей.

Метаданные цифрового сохранения. Предназначены для управления цифровым сохранением коллекций и других информационных ресурсов. К этому классу метаданных относятся документация о физических условиях сохранения ресурсов, о предпринятых действиях для физического сохранения, о цифровых версиях ресурсов, об изменениях, которые были произведены во время их оцифровки и сохранения.

Технические метаданные. Включают, в частности, описание функционирования систем, документацию по оборудованию и программному обеспечению, технологическую информацию о проведенной оцифровке информационных ресурсов, данные аутентикации и обеспечения информационной безопасности (ключи шифрования, пароли).

Метаданные об использовании. Описывают характер использования ресурсов, в том числе, повторного использования, версии ресурсов, авторские права на ресурсы, включают журнал регистрации поисковых запросов и др.

Предпринимались также попытки создания обобщенной *классификации научных метаданных*. Так, предлагаемая в работе [44] классификация включает следующие классы научных метаданных:

- Описание трансформаций данных (описание научных моделей, компьютерных программ и др.)
- Описание семантики данных
- Описание структуры данных
- Описание атрибутов данных
- Описание значений данных.

Эта классификация, к сожалению, является довольно узкой. Она не включает метаданные, описывающие научные приборы, исследовательские эксперименты и другие процессы исследовательской деятельности, ее инфраструктурные и иные ресурсы, связанные с наукой и научной деятельностью.

9. Средства представления метаданных

Для представления метаданных могут использоваться естественные языки, а также различные искусственные языки, такие как языки разметки, схемы метаданных, графические средства. Рассмотрим кратко указанные категории выразительных средств.

Естественные языки. Такие языки обладают наиболее богатыми выразительными возможностями по сравнению с другими средствами представления метаданных. Однако они не обеспечивают однозначности и строгости интерпретации метаданных, представленных их средствами. Такие неструктурированные метаданные в малой степени пригодны для компьютерной обработки и предназначены, главным образом, для пользователей. На естественных языках представляются, например, аннотации публикаций, различные сведения об описываемых ресурсах и их авторах.

Искусственные языки. Это большой набор разнообразных языков. К их числу относятся языки описания данных СУБД, языки концептуального моделирования, описания онтологий, бизнес-процессов, потоков работ. К этой категории относятся, например, дескриптивный подязык языка SQL, язык описания объектов ODL, язык описания интерфейсов IDL стандарта CORBA, языки консорциума W3C: OWL, OWL2, RDF, RDFS, язык XML-схем. Сюда же следует отнести языки разметки, микроформаты, схемы метаданных, визуальные языки, которые рассматриваются ниже, таблицы, графы специального вида и многие другие языки, используемые в различных областях информационных технологий.

Языки разметки. Среди этих искусственных языков наиболее популярны XML, HTML, XHTML, Tex (язык разметки математических текстов) и его оболочка LaTeX. Во многих системах метаданные представляются в виде XML-документов с определенной XML-схемой.

Схемы метаданных. Средства этой категории представляют собой особый вид искусственных языков, который стал в последние годы чрезвычайно популярным. Прототипом для их разработок в значительной мере стал набор элементов метаданных Дублинского ядра [10-12] - стандарт структурированных метаданных с *нетипизированными* значениями элементов. Попытки типизации значений элементов метаданных Дублинского ядра привели к дополнению в его спецификацию описателей типов для значений некоторых его элементов, называемых *квалификаторами*.

Термин *схема метаданных* широко используется в литературе и, по сути, является синонимом термина *набор элементов метаданных*. Схема метаданных? Таким образом, это набор элементов метаданных, каждый из которых обладает некоторым именем и семантикой, принимает значения с установленной семантикой, иногда – значения из управляемого словаря. Существуют схемы метаданных для многих сфер применения, созданные различными профессиональными сообществами. Метаданные описываемых

ресурсов, выраженные средствами схем метаданных, часто кодируются в виде XML-документов.

Визуальные языки. Визуальные языки давно начали использоваться в CASE-технологиях. Наиболее популярным в этой области в настоящее время стал язык объектного анализа и проектирования UML [45]. Основным средством представления метаданных в этом языке являются графические диаграммы нескольких видов, которые на стадии проектирования позволяют описывать различные аспекты создаваемой сложной программной системы.

10. Стандартизация метаданных

Для обеспечения интероперабельности и повторного использования метаданных важное значение имеет стандартизация средств их представления. Деятельность в этой области активно проводится с давнего времени официальными международными и национальными органами стандартизации, промышленными консорциумами, различными профессиональными сообществами. Разработано большое число стандартов метаданных как независимых от сферы применения (стандартов «горизонтальной» сферы), так и предназначенных для специфических применений (стандарты «вертикальной» сферы). К первой группе относятся, например, дескриптивный подязык языка SQL, язык описания объектов ODL консорциума ODMG, стандарты консорциума OMG: язык UML, язык описания интерфейсов CORBA IDL, Common Warehouse Model (CMW); стандарты консорциума W3C: XML Schema, RDF, RDFS, OWL, OWL2 с его профилями и язык описания интерфейсов веб-сервисов WSDL; набор элементов Дублинского ядра (Dublin Core, DC), поддерживаемый Директоратом Дублинского ядра; язык моделирования бизнес-процессов BPMN. Другие примеры стандартов «универсальных» метаданных можно найти в [20, 46, 47].

Среди стандартов «вертикальной» сферы значительное место занимают многочисленные стандарты научных метаданных, созданные для многих областей исследований.

11. Заключение

В этой работе предпринята попытка представить богатый мир метаданных, важнейшего вида информационных ресурсов. Анализ обширного массива публикаций показывает, что существование метаданных, их свойства и возможные функции пока еще не осознаны многими специалистами в достаточной мере.

Множество определений термина *метаданные* существует не случайно. Причины не только в тех факторах, которые обсуждались выше, но и в том, что довольно сложно дать адекватное и лаконичное определение этого термина.

Проблематика метаданных в компьютерных системах является непреходящей. В последние годы все более активно используются семантические метаданные. Разработки новых технологий и новых категорий приложений по необходимости потребуют создания и использования новых видов метаданных. В связи с этим будут создаваться новые языки и другие выразительные средства для их представления. Однако обсуждаемые в данной статье свойства и функции «универсальных» метаданных будут присущи и вновь создаваемым их видам.

Литература

[1] Fry J., Jeris D.W. Toward a Formulation and Definition of Data Reorganization. SIGMOD Workshop 1974. Ann Arbor, Michigan, p. 83-100.

- [2] Mealy G.H. Another Look at Data. Proc. AFIPS'67, Fall Joint Computer Conference, p. 525-534. <http://dl.acm.org/citation.cfm?id=1465682> [Обращение: 2012-12-06].
- [3] Wikipedia. Metadata. <http://en.wikipedia.org/wiki/Metadata> [Обращение: 2012-12-06]
- [4] Bagley P. Extension of Programming Language Concepts. Philadelphia: University City Science Center, November 1968. <http://www.dtic.mil/dtic/tr/fulltext/u2/680815.pdf> [Обращение: 2012-12-06].
- [5] Леонг-Хонг Б., Плагман Б. Системы словарей-справочников данных /Пер. с англ.: Предисловие В.М. Савинкова. М.: Финансы и статистика, 1986. 311 с.
- [6] Lefkovits H.C. Data Dictionary Systems. Q. E. D. Information Sciences. – Inc. Wellesley Massachusetts, 1977. – 491 p.
- [7] ISO/IEC Information technology – Information Resource Dictionary System (IRDS) Framework. International Standard ISO/IEC 10027:1990.
- [8] Когаловский М.Р. Перспективные технологии информационных систем. М.: ДМК Пресс, Компания АйТи, 2003. 288 с.
- [9] Когаловский М.Р., Хохлов Ю.Е. Стандарты XML для электронного правительства. М.: Институт развития информационного общества, 2008. 416 с.
- [10] ANSI/NISO Z39.85 - The Dublin Core Metadata Element Set.
- [11] Dublin Core Metadata Element Set, Version 1.1: Reference Description, 2003-06-02. <http://dublincore.org/documents/dces/> [Обращение: 2012-12-06]
- [12] ISO 15836:2009. Information and documentation - The Dublin Core metadata element set.
- [13] Burnett K., Kwong Bor Ng, Park S. A comparison of the two traditions of metadata development. Journal of the American Society for Information Science. Special issue on integrating multiple overlapping metadata standards. Vol. 50, Issue 13, 1999, pp.1209-1217. <http://comminfo.rutgers.edu/~kbnng/publications/1999JASISPark.pdf> [Обращение: 2012-12-06]
- [14] Ling Liu, Ozsu M. Tamer (eds.). Encyclopedia of Database Systems. Springer, 2009. 748 p.
- [15] Ralston A., Reilly E.D., Hemmendinger D. (eds.). Encyclopedia of Computer Science, 4th edition. John Wiley & Sons Ltd, 2003. - 2034 p.
- [16] Grotschel M., Lugger J. Scientific Information System and Metadata. Konrad-Zuse-Zentrum fur Informationstechnik, Berlin. <http://www.zib.de/groetschel/pubnew/paper/groetschelluegger1999.pdf> [Обращение: 2012-12-06]
- [17] Halshofer B. and Klas W. A Survey of Techniques for Achieving Metadata Interoperability. ACM Computing Surveys, Vol. 42, No. 2, Article 7, February 2010 .
- [18] Metadata Standards and Applications. Introduction: Background, Goals, and Course Outline. ALCTS. http://www.loc.gov/catworkshop/courses/metadastandards/pdf/MSA_InstructorManual.pdf [Обращение: 2012-12-06]
- [19] Taylor C. An Introduction to Metadata. The University of Queensland, Australia. <http://www.library.uq.edu.au/papers/ctmeta4.html>
- [20] Understanding metadata. National Information Standards Organization (NISO), 2004. <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf> [Обращение: 2012-12-06]
- [21] Task Force on Metadata. Summary Report. American Library Association. Committee on Cataloging: Description and Access. June 1999. <http://www.libraries.psu.edu/tas/jca/ccda/tf-meta3.html>
- [22] Воройский Ф.С. Информатика. Новый систематизированный словарь-справочник (Вводный курс по информатике и вычислительной технике в терминах). - 2-е изд., перераб. и доп. М.: Либерия, 2001. 536 с.
- [23] Lagoze C. Metadata for the Web. A necessary Evil? CS 431 – March 2, 2005. Cornell University. http://www.cs.cornell.edu/Courses/cs431/2005sp/lectures/lecture_3_02_05.pdf [Обращение: 2012-12-06].
- [24] Feng L., Brussee R., Blanken H. and Veenstra M. Languages for Metadata. In: Multimedia Retrieval. Data-Centric Systems and Applications, Springer, 2007, 23-51. <http://www.springerlink.com/content/m276p88003533q86/> [Обращение: 2012-12-06]
- [25] Jeusfeld M.A. Metadata. In: Encyclopedia of Database Systems, Springer, 2009. – pp. 1723-1724. <http://www.springerlink.com/content/h241167167r35055/> [Обращение: 2012-12-06]

- [26] Corcho O. Ontology based document annotation: trends and open research problems. *International Journal of Metadata, Semantics and Ontologies*. - Volume 1, Issue 1, January 2006. http://oa.upm.es/5638/2/IJMSO_Corcho_FinalVersionPrintedInJournal.pdf [Обращение: 2012-12-06]
- [27] Википедия. Метаданные. <http://ru.wikipedia.org/wiki/Метаданные> [Обращение: 2012-12-06]
- [28] Inmon, William H. Business metadata: the quest for business clarity. In: W.H. Inmon, Bonnie O'Neil, Lowell Fryman. Elsevier Inc., 2008. – 314 p.”
- [29] Gill T., Gilliland A.G., Whalen M., Woodley M.S. Introduction to Metadata. Online Edition, Version 3.0. Ed. by Murtha Baca. Getty Research Institute, 2008. - 96 p. http://www.getty.edu/research/publications/electronic_publications/intrometadata/index.html [Обращение: 2012-12-06]
- [30] Когаловский М.Р., Паринов С.И. Метрики онлайн-информационных пространств // *Экономика и математические методы*. 2008. Том 44, Вып. 2. С. 108-120.
- [31] Паринов С.И., Ляпунов В.М., Пузырев Р.Л. Система Соционет как платформа для разработки научных информационных ресурсов и онлайн-сервисов // *Российский научный электронный журнал «Электронные библиотеки»*. 2003. Том 6. Выпуск 1. <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2003/part1/PLP> [Обращение: 2012-12-06]
- [32] Open Archives Initiative (2000). <http://www.openarchives.org/> [Обращение: 2012-12-06]
- [33] The Open Archives Initiative Protocol for Metadata Harvesting. <http://www.openarchives.org/OAI/openarchivesprotocol.html> [Обращение: 2012-12-06]
- [34] OMG Meta Object Facility (MOF) Core Specification. Version 2.4.1. OMG Document Number: formal/2011-08-07.
- [35] Когаловский М.Р., Калиниченко Л.А. Концептуальное и онтологическое моделирование в информационных системах // *Программирование. МАИК “Наука”/Интерпериодика*. 2009. № 5. С. 3-25.
- [36] Oren E., Moller K.H., Scerri S., Handschuh S., and Simtek M. What are Semantic Annotations? <http://www.siegfried-handschuh.net/pub/2006/whatissemannot2006.pdf> [Обращение: 2012-12-06]
- [37] Microformats. <http://microformats.org/> [Обращение: 2012-12-06]
- [38] Blur R. The Information Oriented Architecture. The Blur Group, 2011. <http://www.bloorgroup.com/> [Обращение: 2012-12-06]
- [39] Uniform Resource Identifier (URI): Generic Syntax. <http://tools.ietf.org/html/rfc3986> [Обращение: 2012-12-06]
- [40] Когаловский М.Р. Системы доступа к данным, основанные на онтологиях // *Программирование, МАИК/Наука «Интерпериодика»*. 2012. № 4. С. 55-77.
- [41] Когаловский М.Р., Паринов С.И. Семантическое структурирование контента научных электронных библиотек на основе онтологий // *Сборники Президентской библиотеки им. Б.Н. Ельцина. Серия «Электронная библиотека»*. Санкт-Петербург: ФГБУ «Президентская библиотека имени Б.Н. Ельцина», 2011. - Вып.2: Электронная библиотека: современные технологии интеграции информационных ресурсов. С. 26-45.
- [42] DOI (Digital Object Identifier). <http://www.doi.org/> [Обращение: 2012-12-06]
- [43] The Making of America II. <http://sunsite3.berkeley.edu/MOA2/>
- [44] Galhardas H., Simon E., Tomasic A. A Framework for Classifying Scientific Metadata. INRIA, 1998. <http://www.aaai.org/Papers/Workshops/1998/WS-98-14/WS98-14-015.pdf> [Обращение: 2012-12-06]
- [45] OMG Unified Modeling Language (OMG UML) Infrastructure. Version 2.4.1. OMG Document Number: formal/2011-08-05. <http://www.omg.org/spec/UML/2.4.1/Infrastructure/PDF/> [Обращение: 2012-12-06]
- [46] Когаловский М.Р. Энциклопедия технологий баз данных. М.: Финансы и статистика, 2002. 800 с.
- [47] Хохлов Ю.Е., Арнаутов С.А. Обзор форматов метаданных. http://www.elbib.ru/index.phtml?env_page=methodology/metadata/md_review/md_review.html [Обращение: 2012-12-06]

Metadata in Computer Systems

M.R. Kogalovsky

E-mail: kogalov@gmail.com

Market Economy Institute of RAS

Keywords: metadata, metadata function, metadata property, metadata classification, metadata description language, metadata schema, metadata standard

Abstract

The important role in modern computer systems is assigned to a special kind of the information resources named by metadata. The slew of publications is devoted to metadata. In the majority of them standards of the metadata used in various areas are considered. Publications contained discussion of properties and functions of this kind of information resources are rather occasional. Probable for this reason the treatment of the term metadata still has not stabilized. Quite often apparent errors occur in publications, at discussion of private kinds of metadata necessary comments are not done and as a result their properties and functions are extended wrongfully on the general case. In the paper the systematic view on metadata as on information resources of a special sort is offered. The problem of this term definition is discussed. Examples of the metadata used in various areas of information technologies are considered. The general (applications independent) properties of metadata and their functions, means for their representation, the known generalized classifications of metadata, and also their standardization activity are discussed.