

Метаданные, их свойства, функции, классификация и средства представления

© М.Р. Когаловский
Институт проблем рынка РАН
Москва
kogalov@cemi.rssi.ru

Аннотация

Важную роль в современных информационных системах играет особый вид информационных ресурсов, называемых метаданными. Свойства метаданных, их состав и функции существенно зависят от технологий реализации систем, в которых они используются, особенностей описываемых ими ресурсов, а также от области применения и конкретных приложений. Метаданным посвящено огромное количество публикаций. Тем не менее, трактовка термина «метаданные» все еще не устоялась. Большинство работ посвящено обсуждению различных стандартов метаданных. При обсуждении частных видов метаданных нередко не делаются необходимые оговорки, и тем самым их свойства и функции неправомерно распространяются на общий случай. Наряду с этим встречаются ошибочные представления и исторические заблуждения. В данной работе анализируются распространенные трактовки термина «метаданные», рассматриваются общие свойства и функции метаданных, обсуждаются их классификации, а также выразительные средства их представления.

Работа поддержана грантом РГНФ 11-02-12026-в

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. Эти описания необходимы программным средствам, выполняющим указанные функции, а также пользователям для формулировки запросов, анализа данных и интерпретации их содержания. Описания такого рода называются *метаданными* и являются особым видом информационных ресурсов. Их создание часто требует значительных усилий и существенных затрат. Однако они существенно повышают цен-

ность данных, обеспечивают более широкие возможности их использования.

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, различные классификаторы, аннотации статей – все это примеры метаданных. В данной работе, однако, нас будут интересовать метаданные, используемые в компьютерных системах. В этой сфере метаданные начали использоваться уже на ранней стадии ее развития. Когда начали создаваться языки и технологии программирования, термин «метаданные» еще не существовал. Однако в исходном коде программы на языке программирования должно было содержаться описание типов данных, которыми эта программа оперирует. На их основе 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, 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, ограниченный 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 метадан-

ными. 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 также, несомненно, являются метаданными.

Появление гипертекстовых технологий, а затем и Всемирной паутины, предусматривает использование описаний 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 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 необходимо описывать свойства используемых в них данных, связанные с особенностью предметной области исследований. Вместе с тем, могут быть необходимы описания характеристик научных приборов, с помощью которых они получены, место и время регистрации этих данных 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, организации, владеющие информационными ресурсами библиотеки, ее пользователей (их профили), 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 [9, 28]. Несколько позднее была реализована концепция интегрированных словарей-справочников данных — систем управления базами данных, выполняющих традиционные функции СУБД и одновременно функции системы словаря-справочника данных.

Идея интеграции системы словаря-справочника данных и СУБД, хотя и в ограниченной форме, позднее нашла воплощение 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 SQL стали самоописываемыми — они содержат пользовательские данные вместе с описывающими их метаданными.

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. 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 объектом для многих исследований, главным образом, в связи с развитием технологий Семантического Веба, электронных библиотек и ряда других новых пластов информационных технологий. Кроме того, стала актуальной необходимость обеспечения обмена метаданными между различными системами, обеспечения interoperability и повторного использования информационных ресурсов. Все это вызвало активную деятельность по стандартизации метаданных, которой занимаются международные органы стандартизации, промышленные консорциумы, научные и другие сообщества (см., например [2, 3, 8, 11, 12, 17, 24, 25, 28, 29, 31]). В результате были созданы многочисленные стандарты описания метаданных горизонтальной и вертикальной сферы. Активно используются на практике стандарты платформы XML, Дублинское ядро, дескриптивное подмножество языка SQL, большое число схем метаданных для различных областей применения, стандарты языков концептуального моделирования, многочисленные стандарты метаданных в науке — геоданных, астрономических данных, а также научных данных в других областях исследований.

Использование термина «*метаданные*» существенно активизировалось в последние годы. Однако, к сожалению, 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. Хотя метаданным в последние годы посвящено огромное количество публикаций, 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 и других сфер деятельности, разработок электронных библиотек, разнообразных репозиторий цифровых информационных объектов, информационных систем для конкретных сфер применения. При этом во многих работах, посвященных электронным библиотекам или приложениям, основанным на стан-

дартах Семантического Веба, доминирует весьма ограниченный взгляд на функции метаданных. Метаданные рассматриваются в них лишь как средство описания контента (содержания) информационных ресурсов, хотя это только одна из многих возможных функций метаданных.

Формирование адекватного и достаточно полного представления о независимых от конкретной области применения свойствах и функциях метаданных является, на наш взгляд, актуальной проблемой. В данной работе предлагается систематический взгляд на метаданные как на информационный ресурс особого рода, анализируются распространенные трактовки этого термина, приводятся примеры метаданных, используемых в различных видах компьютерных систем, рассматриваются общие (независимые от сферы применения) свойства метаданных и их функции, выразительные средства для их представления, обсуждаются вопросы их обобщенной классификации.

2 Определения термина «метаданные»

Как уже отмечалось, сосуществуют различные точки зрения на содержание термина «метаданные». В работе [14] отмечается, что неоднозначность в понимании смысла и функций метаданных связана с доминирующими в настоящее время двумя подходами. Первый из них свойственен библиотечному сообществу. Он связан с созданием и использованием документальных ИПС, электронных каталогов библиотечных фондов, а в последние годы – с разработками репозитория 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 [14] 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.

Нужно заметить, что на достигнутой стадии развития электронных библиотек, технологии их разработки широко используют самые разнообразные достижения «компьютерных наук», которые далеко выходят за рамки предлагаемых «библиотечной наукой». Это в первую очередь связано с расширением разнообразия видов информационных ресурсов, составляющих контент современных электронных библиотек, и с расширением объема требований пользователей к функциональности этих информационных систем.

Сосуществование двух указанных подходов, а также огромное многообразие видов метаданных и сфер их использования, в значительной мере определяют разнообразие имеющихся трактовок термина «метаданные». Приведем и обсудим некоторые наиболее распространенные из них.

Определение 1. Метаданные — это данные о данных (см., например, [21, 26, 30, 37]).

Определение 2. Метаданные — это информация, которая делает данные полезными [21].

Определение 3. Метаданные — это машино-обрабатываемые данные, которые описывают некоторые ресурсы, цифровые и нецифровые [22].

Определение 4. Метаданными называется допускающая компьютерную обработку и интерпретацию человеком информация о цифровых и нецифровых объектах [31].

Определение 5. Метаданные — это структурированные данные, которые описывают характеристики некоторого ресурса [39].

Определение 6. Метаданными называется структурированная информация, которая описывает, поясняет, указывает местоположение и иным образом облегчает поиск, использование информационного ресурса, а также управление им [42].

Определение 7. Метаданные — это структурированные данные, представляющие собой характеристики описываемых сущностей для целей их идентификации, поиска, оценки, управления ими [38].

Определение 8. 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 [1].

Определение 9. Во Всемирной паутине метаданные – это слабоструктурированные данные, как правило, соответствующие согласованным моделям, обеспечивающим операционную интероперабельность в неоднородной среде [27].

Определение 10. Метаданными называется любая дескриптивная информация о других источниках данных, которая способствует организации, идентификации, представлению, определению местоположения, обеспечению интероперабельности, управлению и использованию этих данных [18].

Определение 11. 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 [26].

Определение 12. Метаданные обычно определяются как «данные о данных», которые предназначены для выражения семантики информации, следовательно, для улучшения ее поиска и выборки, понимания и использования [15].

Определение 13. Метаданные — это данные из более общей формальной системы, описывающей заданную систему данных [44].

Проанализируем кратко этот довольно представительный набор определений.

Определения 1 и 2 являются чрезмерно общими, и не дают конструктивного представления о существовании метаданных.

В определении 3 декларируется универсальный характер (цифровой и нецифровой) информационных ресурсов, которые могут описываться метаданными. Отмечается важное свойство метаданных - возможность их компьютерной обработки. Однако при этом не указывается предназначение метаданных – ориентированы ли они на поддержку системных функций или на конечного пользователя.

Напротив, в определении 4 явно указывается, что метаданные могут описывать как цифровые, так и нецифровые информационные ресурсы, и должны быть воспринимаемыми человеком.

Определение 5 имеет такой же неконструктивный характер, как и первые два, но его авторы дополнительно отмечают, что метаданные представляют собой структурированные данные. Такая точка зрения высказывается во многих источниках. Однако метаданные могут быть не только структурированными данными (см. разд. 4).

В определениях 6 и 7 метаданные также трактуются как структурированные данные. Но эти определения более конструктивны по сравнению с предыдущими, поскольку в них предлагается некоторая конкретизация функций метаданных.

В определении 8 делается акцент на функции описания представления данных. Кроме того, важно, что в нем расширяется «область действия» метаданных. В соответствии с этим определением, описываемым ресурсом могут быть не только собственно данные, но и поддерживающие их системы, используемые в них технологии, пользователи рассматриваемых данных и др.

Авторы определения 9 полагают, что метаданные в Вебе — это слабоструктурированные данные. На наш взгляд, однако, эти метаданные могут быть и неструктурированными, например, текстами на естественных языках. В этом определении акцентируется важная функция метаданных — обеспечение интероперабельности в неоднородных средах информационных ресурсов, описываемых метаданными.

В определении 10 также отмечается эта функция наряду с другими существенными функциями метаданных - обеспечение организации, идентификации, поиска (определения местоположения) и, вообще, управления описываемыми данными. При этом авторы не ограничиваются представлением о метаданных как о структурированных данных.

В определении 11 описываемым ресурсом является *элемент данных*. Поскольку элемент данных – категория, свойственная структурированным данным, данное определение относится лишь к метаданным, описывающим структурированные данные. Нужно заметить при этом, что метаданные элементов данных – лишь только составная часть метаданных структурированных данных. В рассматриваемом определении не охватываются такие, например, аспекты их описания, как структурные характеристики данных, ограничения целостности, полномочия доступа, семантика данных и другие их свойства.

В определении 12 указывается еще одна важная функция метаданных – описание семантики информационных ресурсов, которое является основой для реализации ряда существенных функций поддерживающих их систем.

Наконец, важное свойство метаданных отмечается в весьма лаконичном определении 13. Описание свойств данных и других ресурсов любой природы может осуществляться лишь в терминах более абстрактной системы понятий. Так, в системе базы данных описание хранимых данных осуществляется не в терминах значений этих данных, а в таких терминах, как хранимая запись, элемент данных, ключ записи, указатель, тип хранимой записи, ключ индексирования, область памяти, и др., т.е. в системе терминов модели данных среды хранения базы данных.

Завершая анализ приведенных определений термина «*метаданные*», можно констатировать, что они характеризуют разные точки зрения на метаданные, акцентируют отдельные их свойства и/или функции. Некоторые из них являются слишком абстрактными и в силу этого малопродуктивными. Другие относятся к частному виду метаданных и поэтому не носят общего характера. Хотя такие определения предлагаются авторами в конкретном контексте, в ряде случаев их специфика явным образом в определении не оговаривается. В результате приписываемые метаданным свойства или функции либо категория описываемых ресурсов неправомерно обобщаются.

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. Помимо собственно данных, требуются описания систем, обеспечивающих их хранение, доступ к ним и их обработку, описания характеристик пользователей (их полномочий доступа, их профиля, определяющего, в частности, характер их информационных потребностей и т.п.). Возникли также потребности описания концептуальных схем и онтологий, разнообразных сервисов, средств вычислительной техники, научных приборов и других технических устройств, потоков работ, бизнес-процессов, а также организаций – владельцев, генераторов и пользователей информационных ресурсов.

Нужны описания не только структурированных, но и неструктурированных 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* [23]. Бизнес-правила характеризуют «правила игры» в бизнесе, своего рода руководство 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 [20] 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 наряду с метаданными информационных ресурсов будем обсуждать далее и метаданные ресурсов произвольного вида.

4 Свойства метаданных

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

Относительный характер метаданных. 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 являются данными.

Многоуровневость метаданных. Как уже отмечалось, что описание свойств данных и любого другого ресурса может осуществляться только 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. При этом метаданные, как и данные, являются информационными ресурсами. Для них также могут существовать метаданные, представляющие собой мета-метаданные по отношению к исходным описываемым ресурсам. Такая иерархия метаданных может включать любое число уровней. Так, стандарт Meta Object Facility (MOF) [33] консорциума OMG предусматривает три уровня, спецификация стандартов Дублинского ядра [12, 24] – два уровня: предназначенный непосредственно для практических целей и обеспечивающий его описание. Спецификации более высоких уровней иерар-

хии метаданных обозначаются терминами «*мета-метаданные*», «*мета-мета-метаданные*» и т.д.

Ресурсы, описываемые метаданными. Как уже отмечалось (см. разд. 3), метаданные могут использоваться для описания свойств не только данных, но и многих других ресурсов различных видов.

Автономные/встроенные метаданные. Метаданные могут быть *автономными* (*отчужденными*) от описываемого ресурса или *встроенными* в него. Примеры автономных метаданных: схемы баз данных (они отчуждены от описываемого ими наполнения базы данных); DTD, описывающие тип XML-документов и хранимые в Вебе вне конкретных экземпляров XML-документов; репозиторий метаданных открытого архива в технологии открытых архивов OAI [35, 41]; полнотекстовый индекс коллекции текстовых документов электронной библиотеки. Примеры встроенных метаданных: HTML-разметка веб-страницы; разметка веб-страницы средствами микроформатов; DTD, специфицированное в теле XML-документа; аннотация статьи, содержащаяся в ее тексте; глоссарий, содержащийся в тексте спецификации официального стандарта. Если неструктурированные метаданные встроены в текстовый документ, то для компьютерной обработки необходимо извлечь их из текста. Методам решения этой сложной задачи посвящен в последнее время ряд исследований. Информационные ресурсы со встроенными описывающими их метаданными естественно называть *самоописываемыми*.

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

Метаданные независимые и зависимые от предметной области. Метаданные первой из этих категорий обычно описываются выразительными средствами стандартов горизонтальной сферы (стандартов, не ориентированных на какую-либо конкретную область применения). Метаданные второй категории характеризуются большим разнообразием. Существуют собственные системы метаданных в различных областях научных исследований, в музейном деле, в образовательной сфере и т.д.

Системные и пользовательские метаданные. Метаданные могут быть предназначены для *компьютерной обработки*, связанной с исполнением системных функций (системные метаданные). Они могут быть также предназначены для *пользователей* информационной системы. Метаданные, предназначенные для пользователей, могут служить, напри-

мер, для информирования пользователей о свойствах информационных ресурсов, свойствах системных сервисов, контекста использования системы или предоставляют пользователю инструкцию о порядке выполнения требуемых ему операций. Существуют системы, служащие для поддержки всей совокупности метаданных информационной системы, называемые системами словарей-справочников данных [9, 28]. Их компонент *справочник данных* управляет метаданными, предназначенными для системы, а *словарь данных* — метаданными, предназначенными для пользователей.

Структурированные и неструктурированные метаданные. Для поддержки системных функций, как правило, используются структурированные метаданные. Метаданные, предназначенные для пользователей, могут быть как структурированными, так и неструктурированными, например, они могут быть обычными текстами. В отличие от структурированных метаданных, для неструктурированных метаданных не существует каких-либо общепринятых стандартов их представления. Структурированные метаданные привлекают большое внимание. Им посвящена масса публикаций. Структурированные метаданные, так же как и неструктурированные, могут описывать и структурированные, и не структурированные данные.

Уровень гранулярности описания ресурсов. 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.

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

Степень формализованности метаданных. 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 интенционал описывается средствами логики. Язык OWL для описания онтологий Веба основан на дескриптивной логике.

Зависимость метаданных от информационной архитектуры системы. Состав и содержание метаданных зависит от того уровня информационной архитектуры информационной системы, которому принадлежит ресурс, описываемый метаданными. Действительно, метаданные, описывающие веб-страницы на уровне хранения, представляют со-

бой 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 для 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. В то же время, метаданные на «логическом» уровне представления данных описывают их видение пользователем. В реляционной системе базы данных метаданные описывают состав таблиц базы данных, свойства каждой таблицы, связи между таблицами, полномочия пользователей и т.п., но не организацию хранения базы данных.

Явное и неявное представление метаданных. Многие свойства различных ресурсов информационных и других систем описываются явным образом представленными метаданными. Эти метаданные поддерживаются в форме текстовых документов, иллюстраций или изображений, сообщений по электронной почте, в электронных таблицах, в базах данных и в различных ИТ-инструментах, например, в репозиториях метаданных CASE-инструментов или открытых архивов. Вместе с тем, некоторые их свойства могут быть не представлены в системах явными описаниями. Такая ситуация имеет место, например, в научных публикациях. Для содержащихся в них ссылок на использованные источники не указывается семантика [6] этих связей.

Средства 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 (см. разд. 7).

В данном разделе мы ограничились рассмотрением только основных свойств метаданных, которые не зависят от предметной области, в которой они применяются. Как уже отмечалось, разнообразие метаданных, используемых в конкретных областях применения, чрезвычайно велико. Во многих направлениях научных исследований существуют собственные, специфичные для них системы метаданных. Составляющие их виды метаданных позволяют описывать многообразие свойств исследуемых объектов, явлений, процессов. Например, это — характеристики измерительных приборов, обеспечивающих получение научных данных, разнообразных компьютерных и иных моделей, в соответствии с которыми вычислены или оценены значения используемых данных, характеристики публикаций, источников информационных ресурсов и многое другое.

5 Функции метаданных

Метаданные выполняют разнообразные функции в системах, их использующих. Конкретные их

функции и состав существенным образом зависят от информационных технологий, на которых базируется система, от ее функциональности, свойств поддерживаемых в ней информационных ресурсов, способов их организации в системе, особенностей задач их обработки и от многих других факторов.

Рассмотрим кратко основные функции, выполняемые метаданными, не исчерпывающие всех их возможностей. Приведенные ниже примеры выразительных средств представления метаданных, выполняющих обсуждаемые функции, относятся к числу «универсальных», не специфичных для какой-либо конкретной предметной области или деятельности какого-либо профессионального сообщества.

Описание моделей предметной области. Важнейшая функция метаданных - описание абстрактных моделей предметной области. С разработки этих моделей начинается проектирование информационных систем. Это - концептуальная схема предметной области, дополняемая в системах семантического уровня онтологией [5]. Такие модели не зависят от технологий реализации разрабатываемой информационной системы. Для спецификации концептуальной схемы используются языки концептуального моделирования различного уровня формальности, 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 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 источников данных. Эти функции метаданных выполняют, в частности, схемы баз данных, определения типа XML-документов, XML-схемы для репозитория информационных объектов, онтологии источников информации. Метаданные этого вида представляются, соответственно, языками описания данных (например, дескриптивный подязык языка SQL, язык описания объектов ODL объектных баз данных), подязык языка XML и язык XML-схем, языки описания онтологий (например, RDFS, OWL2).

Описание 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. В последнее время для представления таких метаданных часто используются описатели, представленные в форме XML-документов, библиотечных каталожных записей различных диалектов стандарта MARC. Эти метаданные описывают структурные свойства описываемых информационных объектов, различные другие их характеристики. В «интеллектуальных» системах описывается семантика информационных объектов (см. далее).

Описание семантики информационных источников, информационных объектов или их фрагментов. 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). Для текстовых документов в электронных библиотеках и коллекций информационных ресурсов электронных библиотек, для структурных элементов баз данных, для веб-страниц, а также для электронных библиотек или других информационных репозиториях в целом используют также *семантическое аннотирование*, заключающееся в ассоциировании с описываемым ресурсом некоторой информации (*семантической аннотации*), характеризующей его семантику [36]. Семантическая аннотация может быть встроена в описываемый ресурс или автономна от него. Одним из простейших средств встроеного семантического аннотирования, используемого для описания семантики фрагментов веб-страниц, являются так называемые *микроформаты* [32]. 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.

Онтологии, как и другие средства представления знаний, используются также для автономного формального описания семантики информационных источников. В настоящее время для этого чаще всего служат разработанные консорциумом W3C языки RDF, RDFS, OWL, OWL2 и его профили.

Следует отметить, что вопреки мнению, высказанному в [13] 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 хотя бы потому, что для любых информационных ресурсов, структурированных и неструктурированных, могут использоваться семантические аннотации.

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 СУБД. В среде Веба разметка хранимых HTML-файлов описывает, каким образом веб-браузер должен формировать представление веб-страниц на экране пользовательского компьютера и какие другие файлы из Интернет нужно для этого использовать.

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 уникальным образом идентифицируют отдельные строки или, соответственно, некоторые их множества. XML-документы при наличии DTD или XML-схемы идентифицируются их встроенными идентификаторами. В электронных библиотеках текстовых документов могут использоваться 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 Вебе в качестве уникальных идентификаторов веб-страниц используются указатели URL или URI [43]. Первый из них идентифицирует страницу по ее местоположению в Интернет. В ГИС точки земной поверхности идентифицируются их широтой и долготой. Наконец, нужно упомянуть индексы УДК и ISBN, идентификаторы цифровых объектов DOI, элементы метаданных Дублинского ядра. Существуют и другие примеры видов метаданных, используемых для идентификации информационных ресурсов.

Управление структурированными данными. Это комплекс задач, связанных с обеспечением эффективного хранения и обработки структурированных данных, а также доступа к ним. Чаще всего для решения этих задач используют технологии баз данных. Метаданные, представленные в виде схемы баз данных, обеспечивают механизмы СУБД в процессе ее функционирования необходимой информацией об организации базы данных, ограничениях целостности данных и управлении доступом.

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 может проверить корректность их структуры. Онто-

логия SQL-системы базы данных в системах доступа к данным, основанная на онтологиях [4], позволяет обнаруживать неполноту и противоречивость данных в базе данных.

Предоставление пользователям сведений об описываемых ресурсах. Метаданные могут быть предназначены не только для поддержки системных функций, но и для информирования пользователей о свойствах информационных ресурсов, сервисов, которые обеспечивают доступ к ним и о других описываемых ресурсах. Ориентированные на пользователя метаданные могут быть как структурированными, так и 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.

Управление доступом. 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.

Описание характеристик пользователей и их информационных потребностей. Во многих информационных системах имеется возможность поддерживать сведения о пользователях и их информационных потребностях, необходимые для учета их работы и сбора статистических данных об использовании системных ресурсов. Совокупность таких метаданных называется *профилем пользователя*.

Обеспечение обнаружения и поиска информационных ресурсов. Это — одна из главных функций метаданных, предусматривающая использование их в критериях поиска. При этом могут использоваться не только идентифицирующие, но и другие виды метаданных. 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 метаданные (например, в профилях пользователей), характеризующие информационные потребности пользователей, позволяют обеспечить избирательное распространение информации средствами системы.

Тематическая систематизация информационных ресурсов. Эта функция выполняется путем формирования каталогов и предметных указателей содержащихся в системе информационных ресур-

сов, индексирования ресурсов с помощью рубрикаторов.

Обеспечение анализа данных. Эта функция характерна для систем многомерных или реляционных баз данных и хранилищ данных и связана с использованием методов OLAP. Используемые метаданные содержатся в схемах баз данных.

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. Рассмотрим эту функцию на примере создания и функционирования системы виртуальной интеграции информационных ресурсов из множества источников. Здесь необходимы метаданные, которые описывают: локальные схемы интегрируемых источников, глобальную схему, отображения между локальными схемами и глобальной схемой. В системах семантической интеграции данных необходимы также метаданные, включающие онтологии локальных источников, общую онтологию системы интеграции, а также отображения онтологий. В системах виртуальной интеграции данных с архитектурой адаптеров-посредника необходимы также метаданные, которые описывают регистрационные данные источников в посреднике – центральном компоненте архитектуры систем интеграции данных.

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

Наукометрия в научных электронных библиотеках. Идентифицирующие метаданные (автор, организация, название работы), описания рубрик классификаторов и ссылок цитирования в описаниях научных публикаций в электронных библиотеках позволяют оценивать цитируемость работ 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. Благодаря 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 [6], можно получать многоаспектную наукометрическую и науковедческую информацию.

Рассмотренные в этом разделе функции метаданных используются наиболее часто. Однако в некоторых компьютерных системах они могут выпол-

нять и иные специфические функции, обсуждение которых не входит в задачу этой работы.

6 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 (разд. 4, 5) 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 [42]. 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 [43] или DOI [16]) 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 [42] 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 [40], 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 [20] 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 сохранения. Предназначены для управления 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, данные аутентификации 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 поиска и др.

Предпринимались также попытки создания обобщенной *классификации научных метаданных*. Так, предлагаемая в работе [19] 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, процессы исследовательской деятельности, ее инфраструктурные и другие ресурсы, связанные с наукой и научной деятельностью.

7 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 обеспе-

чивают 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 пласт разнообразных языков. К их числу относятся языки описания данных СУБД, концептуального моделирования, описания онтологий, бизнес-процессов, потоков работ. К этой категории относятся, например, дескриптивный подязык языка SQL, язык описания объектов ODЛ, язык описания интерфейсов IDЛ стандарта CORBA, языки консорциума W3C: OWL, OWL2, RDF, RDFS, язык XML-схем. Сюда же следует отнести языки разметки, схемы метаданных, визуальные языки, которые рассматриваются ниже, и многие другие языки 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 [12, 17, 24] — стандарта 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 — 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, созданные различными профессиональными сообществами для многих сфер применения. 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 [42]. Основным средством представления метаданных в этом языке являются графические диаграммы нескольких типов, которые на стадии проектирования позволяют описать различные аспекты создаваемой сложной программной системы.

8 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 органами стандартизации, промышленными консорциумами, различными профессиональными сообществами. Разработано большое число стандартов метаданных как независимых от сферы применения, так и предназначенных для специфических применений. 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 «общих» метаданных можно найти в [2, 11, 42]. Среди стандартов второй группы значительное место занимают многочисленные стандарты научных метаданных, созданные во многих областях исследований.

9 Заключение

В данной 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. Именно это обстоятельство послужило поводом для подготовки данной статьи.

Множество определений термина «метаданные» существует не случайно. Причина не только в тех факторах, которые отмечались выше, но и в том, что довольно сложно дать качественное 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 видов приложений по необходимости потребуют создания и использования новых видов метаданных. В связи с этим будут, естественно, создаваться и новые выразительные средства для их представления. Однако обсуждаемые в данной статье свойства и функции метаданных будут присущи и вновь создаваемым их видам.

Литература

- [1] 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.

- [2] Коголовский М.Р. Энциклопедия технологий баз данных. М.: Финансы и статистика, 2002. 800 с.
- [3] Коголовский М.Р. Перспективные технологии информационных систем. М.: ДМК Пресс, Компания АйТи, 2003. 288 с.
- [4] Коголовский М.Р. Доступ к реляционным базам данных, основанный на онтологиях / Программирование, МАИК/Наука «Интерпериодика». 2012. № 4.
- [5] Коголовский М.Р., Калинин Л.А. Концептуальное и онтологическое моделирование в информационных системах // Программирование. МАИК «Наука» / Интерпериодика. 2009. № 5.
- [6] Коголовский М.Р., Парин 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.
- [7] Коголовский М.Р., Парин С.И. Метрики онлайн-пространств // Экономика и математические методы. 2008. Вып. 2.
- [8] Коголовский М.Р., Хохлов Ю.Е. Стандарты XML для электронного правительства. М.: Институт развития информационного общества, 2008. 416 с.
- [9] Леонг-Хонг Б., Плагман Б. Системы словарей-справочников данных / Пер. с англ.: Предисловие В.М. Савинкова. М.: Финансы и статистика, 1986. 311 с.
- [10] Парин С.И., Ляпунов В.М., Пузырев Р.Л. Система Соционет как платформа для разработки научных информационных ресурсов и онлайн-сервисов // Электронные библиотеки. 2003. Том 6. Выпуск 1.
- [11] Хохлов Ю.Е., Арнаутов С.А. Обзор форматов метаданных. http://www.elbib.ru/index.phtml?env_page=methodology/metadata/md_review/md_review.html
- [12] ANSI/NISO Z39.85 - The Dublin Core Metadata Element Set.
- [13] Blur R. The Information Oriented Architecture. The Blur Group, 2011. <http://www.bloorgroup.com/>
- [14] Burnett K., Kwong Bor Ng, Park S. A comparison of the two traditions of metadata development / J. 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/~kbng/publications/1999JASISPark.pdf>
- [15] Corcho O. Ontology based document annotation: trends and open research problems / Intern. Journal of Metadata, Semantics and Ontologies. - Volume 1, Issue 1, January 2006. http://oa.upm.es/5638/2/IJMSO_Corcho_FinalVersionPrintedInJournal.pdf

- [16] DOI (Digital Object Identifier). <http://www.doi.org/>
- [17] Dublin Core Metadata Element Set, Version 1.1: Reference Description, 2003-06-02. <http://dublincore.org/documents/dces/>
- [18] 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/>
- [19] Galhardas H., Simon E., Tomasic A. A Framework for Classifying Scientific Metadata. INRIA, 1998. <http://www.aai.org/Papers/Workshops/1998/WS-98-14/WS98-14-015.pdf>
- [20] 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
- [21] Grotschel M., Lugger J. Scientific Information System and Metadata. Konrad-Zuse-Zentrum für Informationstechnik, Berlin. <http://www.zib.de/groetschel/pubnew/paper/groetschelluegger1999.pdf>
- [22] Halshofer B. and Klas W. A Survey of Techniques for Achieving Metadata Interoperability /ACM Computing Surveys, Vol. 42, No. 2, Article 7, February 2010.
- [23] Inmon, William H. Business metadata: the quest for business clarity /W.H. Inmon, Bonnie O'Neil, Lowell Fryman. Elsevier Inc., 2008. – 314 p.
- [24] ISO 15836:2009. Information and documentation - - The Dublin Core metadata element set.
- [25] ISO/IEC Information technology – Information Resource Dictionary System (IRDS) Framework. International Standard ISO/IEC 10027:1990, 1990.
- [26] Jeusfeld M.A. Metadata. In: Encyclopedia of Database Systems, Springer, 2009. – pp. 1723-1724. <http://www.springerlink.com/content/h241167167r35055/>
- [27] Lagose C. Metadata for the Web. Cornell University. CS 431 – March 2, 2005.
- [28] Lefkovits H.C. Data Dictionary Systems. Q. E. D. Information Sciences. – Inc. Wellesley Massachusetts, 1977. – 491 p.
- [29] Metadata Standards. http://www.chin.gc.ca/English/Standards/metadata_intro.html
- [30] Metadata. U.K. Office for Library and Information Networking (UKOLN). <http://www.ukoln.ac.uk/metadata/>
- [31] Metadata Standards and Applications. Introduction: Background, Goals, and Course Outline. ALCTS. http://www.loc.gov/catworkshop/courses/metadatastandards/pdf/MSA_Instructor_Manual.pdf
- [32] Microformats. <http://microformats.org/>
- [33] OMG Meta Object Facility (MOF) Core Specification. Version 2.4.1. OMG Document Number: formal/2011-08-07
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/>
- [34] Open Archives Initiative (2000). <http://www.openarchives.org/>
- [35] 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>
- [36] Scientific Data Management. <http://personal.cscs.ch/~mvalle/sdm/scientific-data-management.html>
- [37] 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>
- [38] Taylor C. An Introduction to Metadata. The University of Queensland, Australia. <http://www.library.uq.edu.au/papers/ctmeta4.html>
- [39] The Making of America II. <http://sunsite3.berkeley.edu/MOA2/>
- [40] The Open Archives Initiative Protocol for Metadata Harvesting. <http://www.openarchives.org/OAI/openarchivesprotocol.html>
- [41] Understanding metadata. National Information Standards Organization (NISO), 2004. <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>
- [42] Uniform Resource Identifier (URI): Generic Syntax. <http://tools.ietf.org/html/rfc3986>
- [43] Wikipedia. Метаданные. <http://ru.wikipedia.org/wiki/Метаданные>

Metadata, their Properties, Functions and Classifications

Mikhail Kogalovsky

In modern information systems the important role is played with a special kind of the information resources named by metadata. Properties of metadata, their composition and functions depend essentially on technologies used for implementation of systems in which they are used. They depend also on features of resources being described as well as on area of use and concrete applications. The huge quantity of publications is devoted to metadata. Nevertheless, the treatment of the term "metadata" still has not stabilized. The majority of publications are devoted to discussion of concrete standards of metadata. In many publications necessary provisos are not done and as a result properties and functions of special kinds of metadata are extended groundlessly on the general case. There are erroneous points of view and historical mistakes. In the article the widespread treatments of the term "metadata" are analyzed, the general properties and functions of metadata are considered; their classifications and expressive means for their representation are discussed.