

должающиеся издания и др.) регламентируется, ГОСТ 7.9 – 95 «Реферат и аннотация. Общие требования». Состав аспектов содержания, которые являются специфичными как для аннотаций на литературно-художественные издания, так и для аннотаций на все остальные виды изданий (не литературно-художественные), приводится в ГОСТ 7.86-2003 «Издания. Общие требования к издательской аннотации».

4. В отличие от аннотирования документов на традиционном (бумажном) носителе информации, имеющем многовековую историю, аннотирование электронных документов и электронных ресурсов – представляет собой новое направление в библиотечно-информационной деятельности, методика которого еще только складывается. Специфичность аннотирования электронных документов и электронных ресурсов во многом обусловлена их «электронной природой»: физической формой объекта аннотирования (электронный ресурс локального или удаленного доступа), видом и объемом электронного ресурса, данными о физическом носителе электронного документа или ресурса, системными требованиями и др. Вместе с тем, общая методика аннотирования вполне применима и к электронным документам (ресурсам). Игнорирование общей методики аннотирования имеет своим негативным следствием появление неинформативных, бесполезных для пользователей и библиотечно-информационных работников вторичных документов. Создание работоспособных методов аннотирования электронных документов (ресурсов) невозможно без опоры на общую методику аннотирования и обогащение ее новыми приемами, обусловленными электронной природой этих документов и потребностями пользователей.

А. Шульгина

Онтология для управления контентом

Рассматривается соотношение между рубрикаторм и современной формой его существования – онтологией. Кратко описана история термина в философском знании, обозначающим сущностные свойства объектов. Выделены основные инструменты построения онтологий для предметной области - экземпляры или индивиды, классы, их таксономическая иерархия, атрибуты и отношения. Приведены программные средства для автоматизированной разработки онтологий, их слияния и анализа. Показаны сильные стороны онтологий, существенно дополняющие свойства линейных рубрикаторов.

Ключевые слова: онтология, рубрикаторм, предметное поле, таксономическая иерархия, атрибуты, отношения, свойства.

Онтологии – это современные формы существования рубрикаторов. Понятие «онтология» сейчас ассоциируется с базами данных, аналитикой, информационными технологиями и системами, но оно прошла долгий путь, прежде чем стала незаменимым помощником в организации огромных массивов знаний. Сам термин «онтология» был подарен нам философией, где он означал «науку о бытии, о сущем». Вопросами бытия занимались такие известные дре-

внейшие философы, как Платон, Аристотель, позже Декарт и Иммануил Кант. Тогда, с помощью онтологического анализа, реальный мир разделялся на составляющие и классы объектов и определение их онтологий, или же совокупности фундаментальных свойств, которые определяют их изменения и поведение. Так как у фундаментальных и естественных наук не хватало инструментов для охвата такого объема информации, онтологии были заимствованы прикладной наукой.

Открывателем «онтологий» как формальных явных описаний терминов предметной области и отношений между ними можно по праву считать Томаса Грубера, который в 1993 году в своей работе, посвященной взаимодействию интеллектуальных систем между собой и с человеком, упомянул этот термин. Интеллектуальными системами называются программы, которые моделируют некоторые аспекты интеллектуальной деятельности человека. Основная идея Т. Грубера состояла в том, чтобы дать возможность интеллектуальным системам обмениваться между собой заложенными в них знаниями о мирах задач. Если изначально в интеллектуальной системе знания о мире могут быть закодированы как угодно, то для обмена этими знаниями с другой интеллектуальной системой необходимо предоставить конкретное описание этих знаний. Это описание должно быть в достаточной степени формальным, чтобы быть понятным другой системе, а также описания систем должны быть написаны на одном языке. Кроме того, описание должно быть понятно также и человеку.

Онтологию по Т. Груберу можно представить как описание декларативных знаний, сделанное в виде классов с отношением иерархии между ними. К этому описанию, которое предназначено для чтения человеком, присоединено описание в канонической форме, созданное для чтения машинами. Каждая интеллектуальная система может предоставлять несколько таких описаний, соответствующих различным областям хранящихся в ней декларативных знаний и, таким образом, выступает как хранилище библиотеки онтологий. Т. Грубер представлял, что интеллектуальные системы будут выступать как библиотеки онтологий и свободно обмениваться онтологиями между собой. При этом библиотеке онтологий вовсе не обязательно быть интеллектуальной системой. Ей достаточно просто предоставлять сервис по передаче онтологий по требованию.

Современные онтологии строятся по большей части одинаково, независимо от языка написания. Обычно они состоят из экземпляров, понятий, атрибутов и отношений. Экземпляры или индивиды – это основные, нижеуровневые компоненты онтологии.

Экземпляры могут представлять собой как физические объекты (люди, дома, планеты), так и абстрактные (числа, слова). Строго говоря, онтология может обойтись и без конкретных объектов. Однако, одной из главных целей онтологии является классификация таких объектов, поэтому они также включаются.

Понятия или классы – абстрактные группы, коллекции или наборы объектов, которые описывают понятия предметной области. Например, класс автомобилей представляет все автомобили. Конкретные автомобили – экземпляры этого класса. Класс может иметь подклассы, которые представляют более конкретные понятия, чем надкласс. Например, мы можем разделить класс всех автомобилей на грузовые, пассажирские, грузопассажирские и специальные. Они могут включать в себя экземпляры, другие классы, либо же сочетания и того, и другого. Пример: понятие «люди», вложенное понятие «человек». Чем является «человек» – вложенным понятием, или экземпляром (индивидом) – зависит от онтологии.

Объекты в онтологии могут иметь атрибуты. Каждый атрибут имеет, по крайней мере, имя и значение, и используется для хранения информации, которая специфична для объекта и привязана к нему.

Значение атрибута может быть сложным типом данных. Важная роль атрибутов заключается в том, чтобы определять зависимости (отношения) между объектами онтологии. Обычно отношением является атрибут, значением которого является другой объект. Атрибуты описывают свойства классов и экземпляров. На практике разработка онтологии включает:

- определение классов в онтологии;
- расположение классов в таксономическую иерархию (подкласс – надкласс);
- определение слотов и описание допускаемых значений этих слотов;
- заполнение значений слотов экземпляров.

После этого мы можем создавать базу знаний.

Составление описания декларативного знания обычно требует большой работы и определенных навыков. Для обозначения этой работы, а также ее результата, Грубер придумал и ввел в обиход специальный термин «концептуализация». Описание он называл «спецификацией». Таким образом, определение значительно упростилось, и онтологию можно определить как спецификацию концептуализации, имеющей вид: $O = \{C, R, A\}$, где O – онтология, C – совокупность концептов предметной области, R – совокупность отношений между ними, A – набор аксиом (законов и правил, которые описывают законы и принципы существования концептов).

Как уже говорилось ранее, онтологии – это новейшие и удобнейшие формы рубрикаторов. Как выглядели первые рубрикаторы? Как вообще классифицировали самые первые документы? Египтяне делали клинописные таблицы с названиями, средневековые монахи делали записи о наличии книги в огромных талмудах. Им не нужно было классифицировать документы по темам, потому что в основном книги были прикованы к полкам, для защиты от воровства, а, следовательно, не меняли своего места с момента появления в монастырской библиотеке. До XVIII века книги, стоящие на полках, размещались по алфавиту с выделением некоторых отраслевых разделов, пока в 1876 году один из виднейших американских библиотекведов М. Дьюи (1851–

1931), который был президентом американской ассоциации и редактором журнала, разработал десятичную классификацию (ДК). В системе Дьюи вся совокупность знаний делилась на 10 основных классов, а отрасли знания и дисциплины, в свою очередь, подразделялись и обозначались десятичным методом. Этой системе более 100 лет, и она до сих пор применима во многих библиотеках и информационных центрах.

Рождение Универсальной десятичной классификации (УДК) в начале XX века связано с именами бельгийских юристов Поля Отле и Анри Лафонтена. Фундаментом новой системы стала Десятичная классификация М. Дьюи, который бескорыстно предоставил П. Отле и А. Лафонтену права по использованию и модификации своей системы для создания всеобъемлющего каталога опубликованных знаний. В течение долгих лет эта работа велась в рамках Международной федерации по информации и документации. В настоящее время УДК является интеллектуальной собственностью специально организованного международного Консорциума УДК, объединяющего основных издателей таблиц УДК на разных языках.

Крупнейшей универсальной классификацией документов в нашей стране ныне является библиотечно-библиографическая классификация – ББК. Но, так как идеальных, полно отражающих все аспекты знаний систем не бывает, УДК и ББК дорабатывается до сих пор. Все эти классификации удобны, но, благодаря развитию информационного обеспечения 60-х–70-х гг. XX века, все больший интерес стал вызывать вопрос разработки систем, ориентированных на работу со структурированными данными. В этот период были разработаны основы и модели организации фактографических данных, отработаны программно-технические решения по накоплению и физическому хранению таких данных, продуманы и реализованы специальные языки запросов к базам данных и решен целый ряд других задач по эффективному управлению большими объемами структурированной информации. В результате основу информационного обеспечения деятельности предприятий и организаций к началу 90-х годов составили фактографические информационные системы, вобравшие в себя в совокупности колоссальный объем структурированных данных. Но, тем не менее, даже ввиду отлаженной работы таких информационных систем, в фирмах все равно оставался ряд неструктурированных документов, работу с которыми невозможно было автоматизировать. Соответственно, у специалистов появились потребности в системах, ориентированных на работу с неструктурированной или слабоструктурированной информацией, и в 70-х годах возникла отдельная ветвь программного обеспечения систем управления базами данных.

Любая классификация, как известно, начиналась с алфавитного списка предметов. Для удобства ей был присвоен цифровой или буквенно-цифровой код. Но в такой перечислительной классификации есть существенный недостаток: отсутствие фиксированных семантических связей и отношений между всеми предметными рубриками. Именно здесь нам на помощь приходят онто-

логии, которые основываются на связях между предметами, аспектами, атрибутами и др. Кроме того, онтологии являются оптимальным инструментом поиска и представления информации, что очень пригодится в программировании и вопросах интернета.

В этой связи перспективным направлением в информатике становится проектирование онтологий. В интернете достаточно доступных программ для моделирования предметного поля, например, одна из первых – программа Protégé Стенфордского Университета.

Онтология любой предметной области всегда создается с конкретной целью. Это не просто описание и классификация информации. Они используются для поддержания и развития электронных библиотек. Описывая ресурсы, библиотекари разрабатывают сложные схемы классификации и правила каталогизации для создания метаданных. Метаданные в свою очередь, увеличивают эффективность работы. Онтологии представляют основу для поиска, которая создает интеллектуальный и исторический контекст. Поиск осуществляется с помощью большого количества разнообразных сервисов. Следовательно, онтологии являются и средством управления контентом базы данных. Сервисы работают без участия человека. Значит, метаданные должны быть удобными для машин, занимающихся поиском. Помимо программ создания онтологий, есть еще очень много программных средств для работы с онтологиями, которые можно объединять между собой. Существуют программы для анализа онтологий.

Именно онтологии позволяют представить данные так, чтобы они стали пригодными для машинной обработки. Предоставляется возможность провести прямую связь между новыми понятиями, с которыми система ранее не встречалась, и описаниями уже известных классов, отношений, свойств и объектов реального мира.

Когда Сеть только развивалась, разработчики столкнулись с проблемой сложного поиска ресурсов поисковыми системами. Это произошло потому, что интернет страницы писались для людей, и электронные машины далеко не всегда могли грамотно распознать требуемую информацию. Так появилась Semantic Web – на страницах появилось описание специально для машин, что дало возможность компьютерам обрабатывать гипертекстовые страницы, предоставляющие информацию людям, и автоматически получать данные. Элементы в Semantic Web получают идентификаторы, называемые «Uniform Resource Identifier» или URI. Каждый сетевой элемент рассматривается как ресурс, который получил определенное имя. Но простой идентификации не хватает для того, чтобы понять, как можно воспользоваться этим ресурсом. Значит, URI бесполезны, пока мы не прибавим к ним описания. На этом этапе в дело вступает онтология. Описания являются основой для семантической разметки, выводя онтологии в ранг определяющего фактора. Таким образом, ком-

пьютер способен «понимать» смысл данных, обращаясь к онтологиям за информацией.

Раньше, отсутствие стандартных, единых языков затрудняло использование разных онтологий внутри взаимосвязанных систем. Было довольно сложно повсеместно использовать онтологии, поскольку существенные различия в синтаксисе и семантике создавали труднопреодолимый барьер. Каждая онтология работала сама на себя, не допуская внедрения информации других систем. Для обмена онтологической информацией был разработан стандартный язык – Web Ontology Language (OWL). Этот язык позволяет устанавливать связи между онтологиями и соединять их. Помимо него существует еще много языков, описывающих онтологии.

Одной из сильных сторон онтологий являются их потенциальные свойства для решения таких важных задач как разделение знаний и их повторное использование. Это заключение основывается на предположении о том, что если онтология явно определена для работающих с ней агентов как общий ресурс, то этот ресурс можно разделять между агентами и многократно использовать. Сейчас специалисты занимаются поисками новых структурных решений внутренней организации онтологии, способов доступа к хранящейся в ней информации, новых методов вывода и представления ее для пользователя. В данный момент необходимо развивать методы взаимодействия пользователя с онтологиями. Вовлечение компьютерных систем, частью которых являются онтологии, позволяет рассматривать всемирную сеть как организованное и структурированное пространство знаний, что, возможно, приведет к использованию информации в сети на новом уровне. Онтологии будут всегда применимы в любой деятельности, составляют основу анализа предметного поля в аналитической деятельности. Возможно, вскоре ни одна информационная фирма не сможет обходиться без системы, базирующейся на онтологических связях. Кроме того, онтологии имеют широкое применение в таких сферах, как моделирование бизнес-процессов, аналитика и библиография, семантическая сеть и искусственный интеллект.