

ИСПОЛЬЗОВАНИЕ ДЕКЛАРАЦИИ ПРОСТРАНСТВ ИМЕН РЕСУРСОВ ЭЛЕКТРОННОГО БИЗНЕСА ДЛЯ ИДЕНТИФИКАЦИИ ПРЕДМЕТНОЙ ОБЛАСТИ

Ключевые слова: уникальный идентификатор ресурсов (URI), стандарт описания предметной области ресурсов (RDF), репозитории, пространства имен, агент-брокер

Раскрыта сущность использования и декларации пространства имен в XML-документах, освещена технология нахождения релевантных ресурсов электронного бизнеса на основе их пространств имен для автоматизированной обработки агентами-посредниками.

Для эффективности электронного бизнеса очень важно, чтобы поставщики и потребители различных товаров и услуг могли легко и быстро найти друг друга. *Электронный бизнес* - деятельность, в которой для получения прибыли используют информационные технологии и телекоммуникации.

Информационные потоки Интернет-среды сложны, динамичны и объемны, поэтому их обработка с помощью существующих программных средств недостаточно эффективна.

Эффективность выполнения поиска информации определяется тем, найдены ли необходимые пользователю ресурсы, и как много нерелевантных запросов относительно требуемых ресурсов было предложено.

К сожалению, описания ресурсов, предлагаемые их владельцами, часто оказываются нерелевантными (как правило, вследствие недобросовестной рекламы либо низкой квалификации разработчиков ресурсов). В то же время анализ терминов, которые используются в метаописании ресурса, позволяет четко определить его предметную область и конкретные возможности.

В данной работе предлагается методика анализа пространства имен, на которые ссылаются XML-документы ресурсов Интернет, и тех имен, которые непосредственно используются в XML-схемах.

На сегодняшний день такая информация о ресурсах достаточно доступна и относительно легко поддается автоматизированной обработке и в то же время позволяет вполне адекватно моделировать знания пользователей (как поставщиков, так и потребителей). Создание владельцем ресурса собственного пространства имен позволяет более точно идентифицировать ресурс, а использование стандартных пространств имен приводит к его обобщенному описанию.

Целью работы является технология нахождения релевантных ресурсов электронного бизнеса на основе их пространств имен для автоматизированной обработки агентами-посредниками.

Моделирование знаний о предметной области с помощью XML

Расширяемый язык разметки XML (eXtensible Markup Language) [1] – результат попытки создания средства для текстового выражения структурированной информации в стандартизированном виде, предназначенный для описания других языков документов. Синтаксически XML-документ – это текст, размеченный тегами, которые необходимы для подробного описания данных. Расширив множество тегов, разработчики XML достигли следующих целей [2]:

Явным образом выделили в документе структуру данных, что делает возможной дальнейшую машинную обработку документа, который при этом все еще остается понятным человеку.

Отделили данные, содержащиеся в документе, от того, каким образом документ будет представлен визуально, что дает широкие возможности для публикации документов на различных носителях — на бумаге, в Интернет, на мобильных устройствах.

Основной конструкцией языка XML являются элементы. Организуя содержимое в элементах, можно явно выделить иерархическую структуру документа. Такой документ, состоящий из вложенных друг в друга элементов, имеет *древовидную* структуру, которая является одной из наиболее подходящих абстракций для описания объектов и отношений между ними.

Имена в XML могут принадлежать различным *пространствам имен*. Различные логические схемы разных документов могут использовать одни и те же имена элементов в различных значениях. Для интерпретации этих значений необходимо указать *пространство имен* – коллекцию имен, идентифицируемых по ссылке URI (Uniform Resource Identifier – идентификатор ресурсов, позволяющий описывать и идентифицировать не только информационные ресурсы Интернет, но и предметы реального мира и общие понятия предметной области), которые используются документами XML в качестве имен типов, элементов и атрибутов. Пространство имен можно рассматривать как ресурс, из которого извлекают необходимые определения. Пространства имен позволяют разработчикам XML объединять несколько словарей в одном документе для полного описания проблемы.

Субъекты электронного бизнеса нуждаются в обмене информацией, которая у различных участников этого процесса хранится в различных форматах. Однако преобразование и объединение DTD (Document Type Declaration) XML-документов – сложная и трудоемкая задача. Альтернативным способом описания логической структуры документов XML являются XML Schemas. XML Schemas и пространства имен позволяют:

лучше организовывать словари для решения сложных проблем;

сохранять сильную типизацию данных при преобразованиях в XML;
более точно и гибко описывать словари, чем в DTD.

Чтобы различать схемы документов, каждой из них ставится в соответствие уникальный идентификатор ресурса. Две схемы будут считаться тождественными тогда и только тогда, когда их уникальные идентификаторы будут совпадать. Часто в качестве URI используются URL различных Web-сайтов.

Технологии XML обеспечивают стандартное представление данных для обработки разными приложениями без специальной дополнительной обработки информации. Для обмена XML-документами необходимо установить стандартные механизмы общего использования словарей (электронных репозиториях) предприятий.

Репозитории [3] могут включать основанные на XML механизмы каталогов, требования по вопросам конфигурации, структур баз данных, знаний, средства моделирования UML, а также словари, которые разъясняют общие и специфические отношения, синонимы и неоднозначное представление терминов в зависимости от контекста. Репозитории становятся де-факто стандартом определенных предметных областей. Предприятия используют:

- внутренние концепции;
- концепции, присущие другим предприятиям их определенных предметных областей;
- концепции, присущие другим предприятиям других государств или группы государств;
- глобальные концепции.

Все концепции, кроме глобальных, представляются по-разному, в зависимости от деловых потребностей, поэтому системы ведения электронного бизнеса на основе XML должны учитывать все эти расхождения и стандартно обрабатывать XML-документы.

Стандарт описания предметной области ресурсов RDF.

Для описания предметной области ресурсов Консорциумом W3C в рамках проекта семантической интерпретации информационных ресурсов Интернет Semantic Web предложен стандарт RDF (Resource Description Framework) [4, 5], который поддерживают многие ведущие производители программного обеспечения и поставщики контента. RDF представляет собой модель описания метаданных. Этот язык использует XML-синтаксис.

Стандарт RDF включает две основные части - способ описания ресурсов и способ задания схем, по которым описывается ресурс. Способ описания ресурсов [6] определяет модель описания ресурсов и связей между ними. Способ задания схем (RDF Schema - RDFS [7]) предназначен для задания структуры предметной области. В RDF можно описывать как структуру ресурса, так и связанную с ним предметную область.

RDF описывает ресурсы в виде ориентированного размеченного графа – каждый ресурс может иметь свойства, которые в свою очередь также могут быть ресурсами или их коллекциями. Базовый строительный блок в RDF — это тройка «объект — атрибут — значение», причем RDF позволяет менять местами объекты и значения.

Агенты в электронном бизнесе

Агенты в электронном бизнесе позволяют реализовывать бизнес-логику обработки ресурсов. Они могут использоваться как посредники между поставщиками и потребителями.

Агенты избавляют от необходимости использования жестко заданных ограничений, что облегчает адаптацию программных средств к новым задачам. *Программные агенты* – это программные сущности, способные действовать автономно, выполняя задачи, поставленные перед ними пользователями или другими программными агентами.

Агент получает от потребителя (потребителем может быть одиночный клиент или компания) запрос, содержащий описание продукции, которую этот потребитель хочет купить. После этого агент ищет среди Web-ресурсов информацию об этой продукции и поставщиках, ее предлагающих. Агент стремится предоставить потребителю только наиболее релевантные ссылки. В процессе взаимодействия агента с потребителем параметры запроса (цена, время доставки, гарантийные обязательства и т.д.) могут быть уточнены и скорректированы в зависимости от предложений поставщиков. Агентно-ориентированная технология предоставляет механизм для обработки обновлений в содержании и структуре бизнес-транзакций.

Нами предлагается использовать для описания продукции, которую хочет приобрести потребитель, имена пространств имен и набор ключевых слов, которые описывают предметную область пользователя. На рис.1 показан процесс обработки запроса потребителя агентом-брокером.

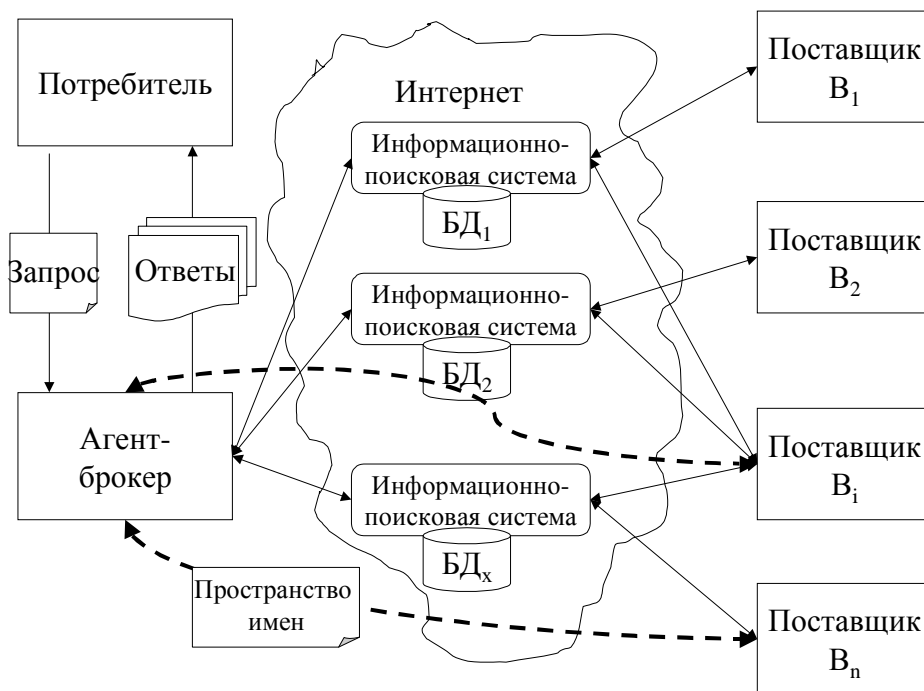


Рис.1. Процесс обработки запроса потребителя агентом-брокером.

Агент-брокер С получает от потребителя А ключевые слова, характеризующие искомый продукт, и имена пространств имен (взятые из репозиторий или созданных самим пользователем), которые характеризуют предметную область, к которой принадлежит искомый продукт.

Агент-брокер С переадресует ключевые слова из запроса потребителя А к одной или нескольким глобальным либо специализированным поисковым машинам Интернет. Получив от них ответы, он отбрасывает повторяющиеся и неактуальные ссылки, а затем отфильтровывает те из оставшихся ссылок, которые соответствуют поставщикам искомой продукции. После этого агент обращается к Web-сайтам поставщиков и определяет пространства имен, которые используются для XML-описаний этих ресурсов. Затем он сравнивает пространства имен потребителя и поставщиков.

Анализаторы XML-документов позволяют получить информацию, относящуюся к пространству имен: базовое имя узла, префикс пространства имен, идентификатор URI пространства имен, который соответствует префиксу пространства имен узла и т.д. Например, в определении пространства имен

`xmlns:catalog="http://www.e-torg_it.com/my_name_catalog.dtd"`

`xmlns` – зарезервированное ключевое слово для декларации пространства имен, `catalog` – префикс пространства имен, а `"http://www.e-torg_it.com/my_name_catalog.dtd"` – URI этого пространства имен.

Проанализировав описание пространства имен, можно сформировать множества терминов предметной области с учетом их иерархии (т.е.получить набор множеств – «термины 1-го уровня», «термины 2-го уровня», ..., «термины i-го уровня»). Сравнивая близость таких наборов множеств (с учетом иерархии), можно количественно оценить релевантность запроса потребителя ресурсам, которые предлагает поставщик.

$$R(A, B) = \sum_i \sum_j \frac{|a_i \cap b_j|}{|a_i \cup b_j|} * C_{i_j}, \quad (1)$$

где $|X|$ – количество элементов в множестве x , a_i – множество терминов i -го уровня иерархии потребителя А, b_j – множество терминов j -го уровня иерархии поставщика В, C_{i_j} – коэффициент оценки близости терминов разных уровней иерархии, $0 \leq C_{i_j} \leq 1$, причем $C_{i_j} = 1$ для $i = j$.

Релевантность запроса потребителя ресурсам, предлагаемым поставщиком оценивается по формуле (1). После этого агент рекомендует потребителю продукцию поставщиков, предлагающих наиболее релевантные ресурсы.

Например,

$a_1 = \{\text{“Компьютер”}\}$,

$a_2 = \{\text{“Принтер”, “Сканер”, “Монитор”, “Модем”}\}$,

$b_1 = \{\text{“Телекоммуникации”}\}$,

$b_2 = \{\text{“Телефон”, “Факс”, “Модем”}\}$,

$$C = \begin{vmatrix} 1 & 0.5 \\ 0.5 & 1 \end{vmatrix}.$$

Тогда

$$\begin{aligned} R(A, B) &= 1 * \frac{|0|}{|\{\text{"Компьютер"}, \text{"Телекоммуникации"}\}|} + 0.5 * \frac{|0|}{|a_1 \cup b_2|} + 0.5 * \\ &* \frac{|0|}{|a_2 \cup b_1|} + 1 * \frac{|\{\text{"Модем"}\}|}{|a_2 \cup b_2|} = \\ &= 0 + 0 + 0 + 1 * \frac{1}{6} = \frac{1}{6} \end{aligned}$$

Таким образом, близость пространств имен потребителя А и поставщика В оценивается как 1/6.

Выводы

В данной статье предлагается технология формализованного описания предметной области, в которой специализируются поставщики товаров и услуг электронного бизнеса (в частности, с помощью создания собственного уникального пространства имен), для его автоматизированной обработки агентами-посредниками электронной коммерции.

1 <http://www.w3.org/TR/1998/REC-xml-19980210>, Extensible Markup Language (XML) 1.0, W3C Recommendation 10.02.1998

2 А.Валиков Технология XSLT. СПб.: ИХВ-Петербург, 2002. – 544 с.

3 Д.Мартин, М. Бирбек, М.Кэй, Б.Лозген, Д.Пиннок, С. Ливингстон, П.Старк, К.Уильяме, Р.Андерсон, С.Мор, Д.Балилес, Б.Пит, Н.Озу XML для профессионалов. Изд.»Лори», 2001. –864 с.

4 <http://www.w3.org/TR/rdf-syntax-grammar/>, RDF/XML Syntax Specification (Revised), W3C Working Draft 25 March 2002

5 <http://www.w3.org/TR/rdf-mt/>, RDF Model Theory, W3C Working Draft 29 April 2002

6 <http://www.w3.org/TR/PR-rdf-syntax>, RDF syntax, W3C Recommendation

7 <http://www.w3.org/TR/PR-rdf-schema>, RDF Schema, W3C Working Draft