

**ВЗАИМНОЕ ПРЕОБРАЗОВАНИЕ
СПЕЦИФИКАЦИЙ XML И ASN.1**

Введение. Как структура представления древовидных структурированных данных нотация абстрактного синтаксиса ASN.1 [1] широко используется в протоколах коммуникации, например, SNMP согласно RFC 3411–RFC 3418 и LDAP согласно RFC 4510 – RFC 4519; протоколах безопасности X.509 (Рекомендация X.509 ITU-X); форматах данных PKCS7 (RFC 2315). ASN.1 разработан для эффективности: данные упаковываются в байты, поэтому они не удобочитаемы и ими трудно управлять. Поскольку данные ASN.1 структурированы, можно представить эту же информацию на языке разметки XML. XML малоэффективен по размерам данных, но удобочитаем и у него есть много стандартных свободных инструментов (например, XML-процессоры для парсинга и генерации, XSL-процессоры для отображения, XML-редакторы для создания). Поэтому трансляция между ASN.1 и XML допускает эффективно управлять данными ASN.1. Это дружелюбный для пользователя способ.

Кроме того, XML разработан для обмена или распределения данных по Интернету и, если данные ASN.1 перевести в XML, их можно использовать в Интернете. Так можно получить доступ к системам, основанным на ASN.1, используя Web-браузеры. Пример перевода – преобразование LDAP-сообщений в XML-документы [2].

Хотя XSD [3] и ASN.1 основаны на структуре с грамматикой метаязыка для структурированных древовидных данных, их выразительная сила не эквивалентна.

Аргументирована целесообразность трансляции ASN.1 и XML для целей eCommerce и рассмотрены детальные различия между этими двумя видами обмена. Предложены способы снижения издержек, связанных с обменом XML-документами.

Цель данной работы – исследовать возможность разработки высоконагруженных систем на основе XML, а после стабилизации обменных форматов использовать ASN.1 как обменный формат в меру его компактности и малой требовательности к ресурсам. Этот вопрос особо актуален для eCommerce и построения электронного документа, который должен обрабатывать и хранить огромное множество XML-документов / сообщений.

Далее рассмотрены следующие вопросы:

- 1) обоснование трансляции ASN.1 в XSD. Хотя выгоды и потребность в трансляции являются явными, следует их перечислить и детализировать;
- 2) обоснование трансляции XSD в ASN.1. Это недавнее веяние, также называемое Fast Web Services (быстрые Веб-услуги), развито рабочими группами бинарного символьного XML-кодирования (XML Binary Characterization) и эффективной передачи XML (Efficient XML Interchange) консорциума W3C;
- 3) применение преобразований в конкретных бизнес-сферах;
- 4) анализ гомоморфизма взаимного преобразования XSD и ASN.1.

Обоснование трансляции ASN.1 в XSD. Ныне расширяется интеграция разных приложений с использованием HTTP через Интернет, а XML стал популярен как стандартный формат обмена данными для таких проектов. XML поддерживает множество приложений, однако все еще много приложений основано на ASN.1. Для взаимодействия этих приложений необходимо преобразование данных между ASN.1 и XML.

Уже закончена деятельность по трансляции данных, используемых в каталогах. OASIS создал спецификацию языка разметки DSML 2.0 для услуг каталогизации, используемой для «выражения запросов и обновлений (результатов этих операций) в форме XML-документов» [2]. Ранняя спецификация DSML 1.0, предоставляла средства для отображения структурированной информации каталога в виде XML-документа. Основные предпосылки для перевода LDAP-сообщений в XML-документы обусловлены тем, что:

- смартфон или PDA (карманный персональный компьютер) должны получить доступ к информации каталога (т. е. к серверу), но не поддерживают LDAP-протокол, посредством которого взаимодействуют клиент с сервером;
- прикладная программа должна получить доступ к каталогу через брандмауэр, но передача трафика протокола LDAP запрещена, поскольку он лишен возможности аудита такого трафика;
- программист создает приложение, используя инструменты и методы XML-программирования, причем приложение получает доступ к каталогу.

Также важна возможность публикации информации в Интернете. Однако ASN.1 невозможно представить в Интернете в «чистом» виде, а трансляция его в XML-документ и применение к нему XSLT дает возможность его представить конечному пользователю посредством Веб-браузера.

По этим соображениям можно сказать, что трансляция ASN.1 в XML.

Обоснование трансляции XSD в ASN.1. С точки зрения разработчика приложения XML имеет множество преимуществ. Соответственно возникает вопрос о целесообразности трансляции XML в бинарное представление ASN.1. Отметим выгоды бинарного кодирования, ассоциированные с ASN.1 для кодирования XML-документа.

1. Бинарный формат не является проперитарным (не требующим авторских прав) и его поддерживает большое количество инструментов.
2. Информация о дереве подчиненности данных и их структурированных фрагментов не хранится в бинарном кодировании.
3. Утверждают, что главный недостаток бинарного кодирования состоит в сложности чтения человеком (если оно не преобразовано в разметку XML), однако это не требуется для многих приложений, когда большинство сообщений никогда не читают конечные пользователи.
4. XML-документы могут передать устройства, имеющие слабый процессор и ограниченную полосу пропускания. А системы обработки с большим количеством транзакций, медленной полосой пропускания и слабыми процессорами с малым объемом памяти не подходят для применения сложных алгоритмов сжатия данных. В среднем норма сжатия XML-документов лучше, чем для таких механизмов, как zip.
5. Приложения не замедляются, поскольку отсутствует потребность в создании словаря на лету (кодер и декодер сгенерированы раз и навсегда).
6. Применение функций безопасности к данным, определенным согласно, хорошо отработано, а нехватка повторных текстовых последовательностей в ASN.1, предотвращение избыточности и общая текстовая ориентация правил упакованного PER-кодирования снижают уязвимость функций защиты для угроз, по сравнению с прямым шифрованием XML.

Общие соображения для новых приложений. При создании высоконагруженных приложений встает вопрос об использовании ASN.1 или XML. Рекомендации на этот счет следующие:

- использовать XML, ввиду его удобств для новых продуктов;
- впоследствии, когда стабилизируются XSD-схемы обмена XML-документами, использовать рекомендации раздела D.1 из ISO/IEC 8825-5 [4] для преобразования XSD-схемы в ASN.1 с одной из ASN.1 кодировок.

Дальнейшее развитие этих рекомендаций будет изменяться в свете рекомендации W3C «Формата эффективного обмена XML (EXI) 1.0» [5].

EXI. Формат эффективного обмена XML – компактное, эффективное представление XML. Он одновременно улучшает производительность и уменьшает требования к полосе пропускания, не ставя под угрозу эффективное использование таких ресурсов как срок службы аккумулятора, размер кода, мощность процессора и объем памяти.

В EXI использованы грамматический подход, допускающий эффективное кодирование, алгоритмы прямого кодирования и небольшой набор типов данных. Следовательно, процессоры EXI относительно просты, их можно реализовать на оборудовании с ограниченными возможностями.

Поскольку EXI – «информированная» схема, то в формате использована доступная информация схемы для компактности и повышения производительности, но обработка не зависит от точных, полных или текущих схем. Поддерживая произвольные расширения схемы и ее производные, формат эффективно работает с частичными схемами или в отсутствии любой схемы. Непосредственно формат не зависит от конкретного языка схемы или формата данных схемы.

EXI – результат исследований рабочих групп W3C по бинарному представлению (ХВС – XML Binary Characterization) и эффективному обмену XML (EXI – Efficient XML Interchange). ХВС делегировали исследование затрат и выгод альтернативной формы XML и сформулировать путь объективной оценки потенциала замены формата для XML. Основываясь на рекомендациях ХВС, перед EXI стояли задачи сначала измерить, оценить и сравнить разные технологии XML (используя разработанные ХВС метрики), а затем сформулировать рекомендацию для спецификации формата W3C. Результаты обеих групп получены в полном объеме и отображены в [5].

Функциональные требования формата EXI подготовлены рабочей группой ХВС на основе анализа желательных свойств кодирования высокой эффективности для XML. Полученные из набора случаев практического использования эти свойства идентифицированы рабочей группой ХВС по направлениям: метаданные в системах радиопередачи; массивы с плавающей запятой в энергетике; модели сжатия, сериализации и передачи графиков X3D; Веб-службы для мобильного оборудования; Веб-службы в Интранет предприятия; электронные документы; FIXML как обработка ценных бумаг; мультимедийные XML-документы для мобильных телефонов; корпоративная бизнес-коммуникация; сжатие мгновенных сообщений XMPP; продолжительное хранение XML-документов; обработка бизнес-сфер и знаний; роутинг XML, основанный на контенте, и подписка на издание; роутинг Веб-служб; интероперабельность военной информации; обработка датчиков и коммуникация; SyncML для синхронизации данных; суперкомпьютеры и Грид-системы.

В значительной мере формат основан на результатах измерений, выполненных группой для оценивания производительности (преимущественно эффективности обработки и компактности) разных реальных форматов. Формат EXI основан на эффективном XML, включая, например, базовый подход эвристической грамматики, алгоритм сжатия и энтропию результатов кодирования.

EXI совместим с XML на уровне информационного набора XML, а не на уровне синтаксиса XML. Это позволяет ему инкапсулировать эффективный альтернативный синтаксис и грамматику для XML, минимизируя воздействие на прикладную интероперабельность XML.

Эффективность EXI. Исследования компактности представления XML-документов выполнены в рамках разработки формата EXI. К сожалению, в [5] не указаны затраты времени и ресурсов на создание компактного представления всеми из перечисленных способов.

Измерение проходило в два этапа, на первом – сравнили результаты: XML, gzipped XML и эффективного XML (рис. 1). На втором этапе сравнили результаты: XML, ASN.1 PER и эффективного XML (рис. 2).

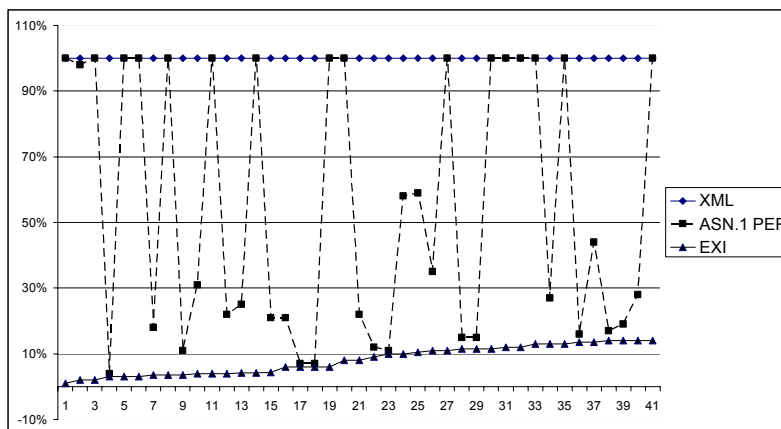


РИС. 1. Сравнение компактности XML, gzipped XML и эффективного XML

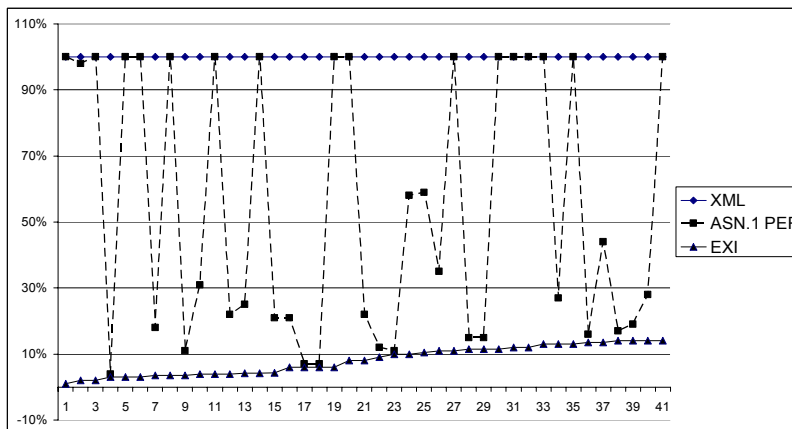


РИС. 2. Сравнение компактности XML, ASN.1 PER и эффективного XML

Использование правил кодирования ASN.1 для XML основаны на разделении схем, используемых для определения структуры сообщения (метаданных) и непосредственного контента сообщений / документов. В таблице приведены схемы XML и сравнение их с ASN.1, как схемами, использующими правила бинарного кодирования.

ТАБЛИЦА

Язык схемы	Правила кодирования
XML Schema; RelaxNG; Schematron	XML
ASN.1	BER; DER; CER; PER; XER

Язык схемы описывает структуру сообщения, а правила кодирования используют для описания экземпляров сообщений. По аналогии с объектно-ориентированном программировании (ООП), схема связана с концепцией класса объектов (например, определение Java-класса), а правила кодирования использовались бы для создания экземпляра этого класса.

Как видно из результатов оценки компактности, EXI весьма эффективен даже по сравнению с ASN.1 PER.

Применение преобразований в конкретных сферах. Важность настоящего исследования – техническая, юридическая и организационная возможность внедрения в Украине инфраструктуры электронных счетов (eInvoice). Количество XML-документов в этой сфере приближается к 30 миллиардам ежегодно, по аналогии с государственной статистикой Швейцарии.

Согласно Национальной концепции развития телекоммуникаций в Украине взят курс на широкополосной Интернет на основе мобильной связи. Имея меньшую пропускную способность по сравнению с оптоволоконными сетями, эта технология предоставляет неоценимые преимущества наращивания аудитории пользователей Интернет, ввиду снижения стоимости единицы трафика и энергозатрат, задействования потенциала телефонных сетей и т.п. Учитывая эти факты, критичны размер XML-документа eInvoice и скорость его обработки.

Представлены оценки размера UBL eInvoice XML-сообщения и времени сериализации / десериализации в XML, ASN.1 BER и ASN.1 PER (рис. 3). Как показано на графиках, использование ASN.1 целесообразно согласно Приложению D [6] как простой и стандартизированный способ построения обмена электронными документами.

Анализ гомоморфизма взаимного преобразования XML и ASN.1. При трансляции стандартного XML-документа в ASN.1 и обратно наилучшим образом решается вопрос при изоморфности такого преобразования.

Преобразования XML в ASN.1. Языком схемы XML является XSD, процесс преобразования XSD в ASN.1 регламентирует [6] со следующими ограничениями:

- игнорируют все `annotation`. Назначение этого тега – интерпретация человеком;
- игнорируют все **identity-constraint definition**;
- игнорируют все **notation declaration**;
- игнорируют все компоненты схемы как фасеты **fundamental facet (ordered, bounded, cardinality, numeric)** определений **simple type definition**;
- игнорируют определение **identity-constraint definition** свойств, **substitution group exclusion** и **disallowed substitution** объявлений **element declaration**;
- игнорируют свойства **final**, **abstract** и **prohibited substitutions** определений **complex type definition**;
- игнорируют контентны **process contents** свойств групповых символов **wildcard**;
- игнорируют фасеты **fundamental facet** свойств и фасеты **final** определений **simple type definition**;
- игнорируют все ограничения **value constraint**, имеющиеся в любых определениях **type definition** или **attribute declaration**, чье определение **type definition** есть или `xsd:QName`, или определением **simple type definition**, полученным с `xsd:QName` или `xsd:NOTATION`;
- игнорируют все определения **attribute group definition**;
- игнорирование перечисленных компонентов XSD-схемы связано с невозможностью или ненужностью их представления в ASN.1.

Очевидно, имеем гомоморфное преобразование XML в ASN.1.

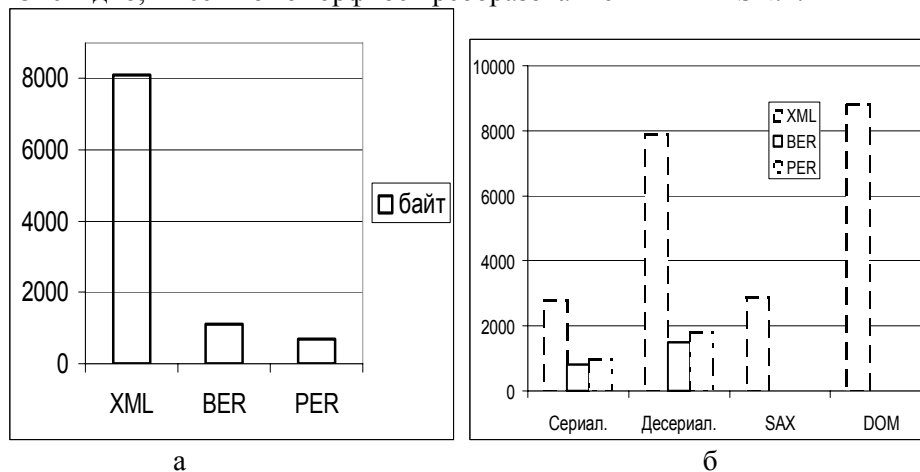


РИС. 3. Размер файла XML, ASN.1 BER и ASN.1 PER (а);
 время сериализации и десериализации XML, ASN.1 BER и ASN.1 PER (б)

Преобразования ASN.1 в XML. Для промышленного использования наиболее подходят по свойствам канонические правила XML-кодирования CXER [6], обладающего свойством однозначной трансляции ASN.1 в XML. Недостаток XER (в частности CXER) – невозможность использования списков и атрибутов XML при трансляции. Использование BASIC-XER и EXTENDED-XER нецелесообразно ввиду неоднозначной трансляции.

Значит, преобразование ASN.1 и XML гомоморфно, ввиду наличия в XML свойств, которые невозможно представить или однозначно транслировать в ASN.1. Перечень XSD компонентов, трансляция которых невозможна в ASN.1, указана в 3.1. При трансляции из ASN.1 в XML требуется использовать дополнительные механизмы, наряду с CXER, для однозначной трансляции в XML.

Использование схемы трансляции ASN.1 в XML. На рис. 4 показана схема согласно положениям D.1 [5]. Для примера преобразуем XML-документ UBL eInvoice, схема которого имеет следующий вид:

```

...
<xsd:element name="Invoice" type="InvoiceType"> </xsd:element>
  <xsd:complexType name="InvoiceType">
    <xsd:sequence>
      ...
      <xsd:element ref="cbc:ID" minOccurs="1" maxOccurs="1"> </xsd:element>
      ...
    </xsd:sequence>
  </xsd:complexType>
  <xsd:complexType name="cbc:ID">
    <xsd:simpleContent>
      <xsd:extension base="xsd:token">
        <xsd:attribute name="schemeName" type="xsd:string"
use="optional"> </xsd:attribute>
      </xsd:extension>
    </xsd:simpleContent>
  </xsd:complexType>

```

```

</xsd:complexType>
  На этапе «Отображение XSD в ASN.1» имеем ASN.1-нотацию:
  /* xml version="1.0" encoding="UTF-8" */
  Notarget-ns
  DEFINITIONS XER INSTRUCTIONS AUTOMATIC TAGS ::=
  BEGIN
  IMPORTS
    Token, String, AnyURI
    FROM XSD;
  IDType ::= SEQUENCE {
    schemeName      [ATTRIBUTE] XSD.String OPTIONAL, }
  ENCODING-CONTROL XER
    GLOBAL-DEFAULTS MODIFIED-ENCODINGS
    GLOBAL-DEFAULTS CONTROL-NAMESPACE
    "http://www.w3.org/2001/XMLSchema-instance" PREFIX "xsi"
  END
  
```

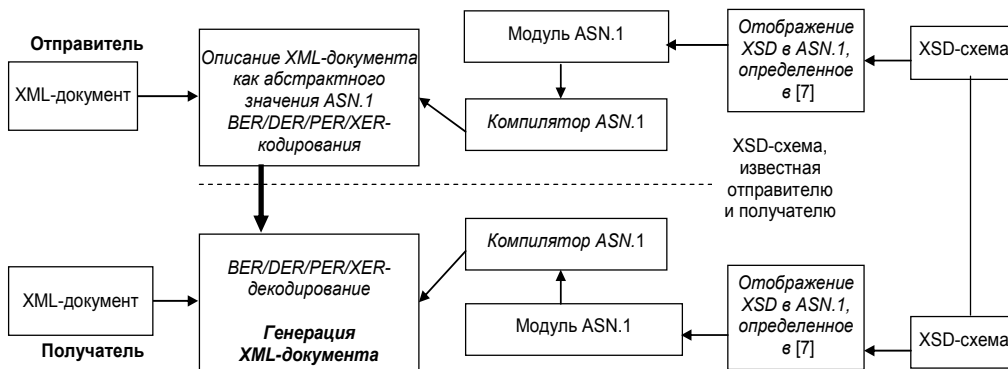


РИС. 4. Передача XML-документа с отображением XSD в ASN.1

На основе этой нотации разработан «Модуль ASN.1» и «Компилятор ASN.1». Соответствующие части передаются отправителю и получателю, производящим обмена XML-документами с бинарным кодированием ASN.1.

Заключение. На основе потенциала использования ASN.1 как обменного формата в высоконагруженных системах, изначально построенных на XML, обосновано такое использование, в первую очередь снижением размера сообщений и ресурсов, необходимых для обработки ASN.1-сообщений. На примерах показаны конкретные метрики уменьшения размера и времени сериализации / десериализации XML-сообщений UBL eInvoice при использовании ASN.1-кодирования BER и PER. Несмотря на стандартизацию преобразований между ASN.1 и XML, а также наличие готового инструментария, гомоморфное преобразование затрудняет перевод, хотя он возможен и целесообразен. Имеем единственный аналог продуктивного сужения гомоморфизма (до квазиизоморфизма), описанный в стандартах привязки языков программирования как обратное внутреннее преобразование для маршалинга типов данных в распределенных приложениях. Этот важное перспективное направление работ.

Для систем электронного документооборота пригодна схема трансляции ASN.1 в XML, описанная в D.1 [6]. Но оптимальный вариант снижения издержек на обработку XML-документов – использование формата EXI. Несмотря на его новизну (2008 г.), в основу этого формата положены результаты, отображенные в настоящей работе. В сравнении с gzipped XML и ASN.1 PER, формат EXI значительно компактнее представляет XML-документ и является интероперабельным форматом, сохраняя все свойства XML-документа.

Пока формат EXI как рекомендация W3C имеет узкий круг пользователей и не прошел испытания временем. А инструментарий ASN.1 «отточен» десятилетиями использования и стандартизирован (даже в Украине), что делает его приемлемым кандидатом на обменный формат электронного документооборота.

А.О. Мелашченко

ВЗАЄМНЕ ПЕРЕТВОРЕННЯ СПЕЦИФІКАЦІЙ XML І ASN.1

Аргументовано доцільність трансляції ASN.1 і XML для цілей eCommerce і розглянуті детальні розходження між цими двома видами обміну. Запропоновано способи зниження витрат, пов'язаних з обміном XML-документами.

A. Melashchenko

MUTUAL TRANSFORMATION OF XML AND ASN.1 SPECIFICATIONS

The appropriateness of translation ASN.1 and XML for the eCommerce purposes is argued, and detailed distinctions between these two kinds of exchange are considered. Means of decrease of the costs connected with the XML-document exchange are proposed.

1. *ДСТУ ISO/IEC 8824-1:2008* Інформаційні технології. – Нотація абстрактного синтаксису (ASN.1) – Ч. 1: Специфікація базової нотації.
2. *Directory Services Markup Language v 2.0*, 2001.
3. *W3C. XML Schema Part 0: Primer, Part 1: Structures, Part 2: Datatypes*, 2004.
4. *ДСТУ ISO/IEC 8825-5:2008* Інформаційні технології. – ASN.1. Правила кодування. – Ч. 5: Відображення визначень W3C XML-схем в ASN.1.
5. *W3C Last Call working «Efficient XML Interchange (EXI) Format 1.0»*, 2008.
6. *ДСТУ ISO/IEC 8825-4:2008* Інформаційні технології. – ASN.1. Правила кодування. – Ч. 4: Специфікація правил кодування XML (XER).

Получено 15.05.2009

Об авторе:

Мелашченко Андрей Олегович,

младший научный сотрудник, аспирант

Института кибернетики имени В.М. Глушкова НАН Украины.