STUDIA INFORMATICA 2003
Volume 24 Number 2A (53)

Elżbieta KOSMULSKA-BOCHENEK
Politechnika Wrocławska, Wydział Informatyki i Zarządzania

REPREZENTACJA DANYCH XML W NOTACJI ASN.1

Streszczenie. Artykuł dotyczy notacji uzyskanej z połączenia dwóch standardów reprezentacji danych: ASN.1 i XML. Przedstawiono istotę notacji ASN.1 i XML oraz połączoną notację ASN.1-XML. Obie notacje dają możliwość specyfikacji danych w celu późniejszego przesłania ich przez sieć, ale różnica polega na liczbie bajtów wygenerowanych dla transferu. Przedstawiono zasady tworzenia kodu transferu dla obu języków. Przytoczony w artykule przykład ilustruje proces kodowania w obu notacjach oraz dostarcza ilościowego porównania kodu transferowego otrzymanego w obu przypadkach dla tych samych danych. W zakończeniu artykułu podano obszar zastosowań tego rodzaju standardu oraz aktualny stan prac normalizacyjnych.

Słowa kluczowe: składnia abstrakcyjna, ASN.1, reguły kodowania, XML.

REPRESENTING DATA XML USING ASN.1 NOTATION

Summary. The paper describes the notation resulting from two standards of data representation: ASN.1 and XML. The basic ideas of the two notations are outlined and benefits from combining them are presented. The both languages produce a network transfer code, but there is a difference in number of resulting bytes. The rules for encoding data structures for ASN.1 and XML are introduced. An example proving the difference in terms of volume for transfer message using rules for XML schema and ASN.1 encoding rules is presented. Some potential application domains and a current state of normalization process complete the paper.

Keywords: abstract syntax, ASN.1, encoding rules, XML.

1. Rozszerzalny język znaczników - XML

XML [4] jest standardem języka do opisu danych. O ile HTML stosuje się do przedstawienia struktury zawartości strony WWW (*World Wide Web*) w formacie odpowiednim do wyświetlenia przez przeglądarkę, o tyle XML daje możliwości opisu danych w taki sposób,

aby mogły one być interpretowane i przetwarzane przez różne aplikacje. Dane przedstawione w XML mogą więc być nie tylko wyświetlane, ale również sprowadzane na żądanie i przetwarzane bez pośrednictwa człowieka.

Dokument XML zewnętrznie przypomina dokument HTML, z tym że zawiera dodatkowo nowe pola etykiet (*tag*), określające znaczenie zawartych w nich danych. XML może być podstawą do definicji innych języków znacznikowych, uzyskanych przez specyfikację etykiet i związanych z nimi atrybutów. Przykładem takich języków jest WML (*Wireless Markup Language*), przeznaczony do specyfikacji interfejsu użytkownika w mobilnych urządzeniach lub VoxML, opracowany przez Motorolę w celu utworzenia interfejsu głosowego z aplikacjami WWW [5].

Stosowanie notacji XML umożliwia projektantom aplikacji organizowanie danych jako tekstowe pliki HTML, które mogą być zinterpretowane w każdym urządzeniu cyfrowym, od komputerów począwszy, na telefonach komórkowych skończywszy.

2. Notacja składni abstrakcyjnej - ASN.1

ASN.1 (*Abstract Syntax Notation One*), czyli notacja składni abstrakcyjnej wersji 1*) jest znormalizowanym językiem (ITU-T Rec.X680-683) służącym do definiowania składni wiadomości, jaka może być wymieniana między współpracującymi ze sobą aplikacjami, niezależnie od ich lokalnej implementacji. Jako język specyfikacji danych aplikacyjnych ASN.1 ma wiele zalet, do których można zaliczyć:

- zwięzłą notację umożliwiającą przedstawienie zarówno prostych, jak i skomplikowanych typów,
- bogaty repertuar typów,
- możliwości wprowadzania ograniczeń w zakresie wartości i rozmiaru,
- możliwości wprowadzania późniejszych rozszerzeń specyfikacji,
- możliwość używania zarówno w połączeniu z nowoczesnymi językami programowania (Java, C++) jak i starymi, takimi jak C lub COBOL.

Standard ASN.1 nie nakłada ograniczeń na formę zewnętrzną zapisu składni. Podstawą opisu składni w ASN.1 są *znaki* – należące do zdefiniowanego zbioru – które pogrupowane tworzą *elementy* ASN.1 oraz ich *nazwy*. Szczegółowy opis standardu języka można znaleźć w [7].

---

*) Standard istnieje od roku 1984 i jest uaktualniany co 4 lata, ale w nazwie pozostawiono "1", co sugeruje trwałość i aktualność koncepcji języka oraz jego podstawowych mechanizmów.

Od swojego pierwotnego zastosowania, jakim była specyfikacja wiadomości w protokole poczty w środowisku OSI, krąg aplikacji wykorzystujących ASN.1 znacznie się poszerzył. ASN.1 jest obecnie językiem specyfikacji protokołów zarządzania siecią telefonii komórkowej, transmisji głosu i obrazu w Internecie, technik kryptograficznych i wielu innych.

W module ASN.1 mogą znajdować się wartości zapisane w formacie XML, a tym samym dane te zostaną wyświetlone przy użyciu przeglądarki XML. Każda definicja typu ASN.1 może być zapisana za pomocą znaczników. Przyjęto zasadę użycia jako znaczników nazw typów zdefiniowanych w ASN.1 i identyfikatorów specyfikowanych przez użytkownika. Włączenie notacji XML do tradycyjnej ASN.1 jest przedstawione w propozycjach rekomendacji ITU-T X.680 – 683 (ISO/IEC 8824-1) i X.681 (ISO/IEC 8824-2) - wersje obu dokumentów nie są jeszcze zatwierdzone.

3. Reguły kodowania

ASN.1 jako język opisu składni abstrakcyjnej umożliwia definiowanie struktur danych oraz nadawanie im wartości, ale nie uwzględnia sposobu reprezentacji danych dla ich transmisji. W zasadzie może być użyty w tym celu dowolny mechanizm kodowania, który gwarantowałby jednoznaczną interpretację transmitowanej informacji zarówno po stronie wysyłającej, jak i odbierającej. Od przyjętej metody kodowania wymaga się więc ścisłego związku z przenoszoną składnią, ponieważ obaj partnerzy transmisji powinni mieć określoną wiedzę o wymienianych strukturach danych. Dlatego też dla ASN.1 opracowano zbiór reguł służących do kodowania zdefiniowanych składni, który ze względu na swój ścisły związek z ASN.1 można uważać za pewne przedłużenie lub uzupełnienie norm tego języka. Z notacją ASN.1 zgodne są cztery, opracowane i znormalizowane przez ISO i ITU-T, składnie transferu [1,2].

Reguły kodowania określają, w jaki sposób dane powinny być zakodowane dla transmisji i są one niezależne od typu maszyny, języka programowania czy też reprezentacji tych danych w programie aplikacyjnym. Zbiór reguł kodowania dostarcza wzorców reprezentacji każdego elementu - zdefiniowanego w ASN.1 – w formie ciągu bajtów. W ten sposób każda struktura aplikacyjna podlega linearyzacji koniecznej do fizycznej transmisji danych. Zgodnie z normą podstawowych reguł kodowania (BER - *Basic Encoding Rules*) na reprezentację wartości abstrakcyjnego typu danych języka ASN.1 składają się trzy pola:

– pole identyfikatora,
– pole długości zawierające rozmiar wartości (w oktetach),
– pole zawartości, w którym umieszczana jest aktualna wartość.

Dodatkowo, jeśli jest to konieczne, występuje jeszcze jedno pole, zwane polem końca oktetów. Każde pole wyrównane jest do całkowitej liczby oktetów.

Kodowanie według zasad BER jest bardzo elastyczne (wiele opcji) i nie wprowadza żadnych ograniczeń dla obu partnerów transmisji, jednak ze względu na to, że bazuje ono na strukturach bajtowych o zmiennych długościach, jest mało oszczędne w wykorzystaniu pamięci oraz czasu pracy procesora. Dla nowoczesnych aplikacji, szczególnie tych, które wymagają autoryzacji transmisji (podpis elektroniczny) oraz szybkiej wymiany danych, tego rodzaju składnia transferu nie powinna być zastosowana.

Wymagania dotyczące efektywności kodowania i dekodowania danych transferowych stały się głównym powodem, dla którego norma składni transferu dla ASN.1 przewiduje, poza BER, jeszcze inne sposoby kodowania. Są to: wyróżnione reguły kodowania (DER – *Distinguished Encoding Rules*) i kanoniczne reguły kodowania (CER – *Canonical Encoding Rules*). Reguły kodowania CER i DER są wyspecjalizowanymi podzbiorami BER. Oznacza to, że dekoder BER jest w stanie interpretować składnię CER i DER, ale nie odwrotnie – dekodery wyspecjalizowane dla CER lub DER nie są w stanie poprawnie zdekodować ciągów bitów zakodowanych według bazowych reguł BER. Reguły DER mają zastosowanie, między innymi, w transmisji procedur uwierzytelniania (X.500). Kodowanie zgodne z CER ma obecnie mniejsze zastosowanie, aczkolwiek wymuszone było również wymaganiami bezpieczeństwa w poczcie elektronicznej (X.400).

Kolejnym krokiem w optymalizacji kodu transferu struktur ASN.1 było opracowanie reguł kodowania upakowanego (PER – *Packed Encodings Rules*), które charakteryzuje się bardzo zwartą reprezentacją kodowanych wartości. W tej notacji transferu identyfikatory nie są nigdy przesyłane, natomiast długości i wartości mogą również nie być przesyłane, jeżeli tylko znane są obu partnerom aplikacji. Dopuszcza się dwie formy generacji kodu: z wyrównaniem do pełnego oktetu oraz bez takiego wyrównania (mniejsza liczba transferowanych bitów).

4. Reguły kodowania danych zapisanych w notacji ASN.1-XML

Konwersji między notacjami ASN.1 i XML można dokonać przez zastosowanie standardowych reguł. ASN.1 daje możliwość specyfikacji typów struktur wiadomości oraz wartości danych. Aby zamienić strukturę danych ASN.1 na format XML należy utworzyć znaczniki początku i końca oraz obudować nimi daną wartość. Konwersja formatu XML na strukturę ASN.1 polega na przypisaniu tej strukturze stosowego identyfikatora ASN.1 i wyliczeniu długości. Na przykład prosta produkcja ASN.1 niech ma postać:

Nazwisko ::= [2] IMPLICIT OCTET STRING

Przy nadanej wartości "Kowalski" zostanie zakodowana w notacji XML jako (etykieta jest ignorowana):

<Nazwisko> <OCTET_STRING>Kowalski</OCTET_STRING></Nazwisko>

W rekomendacji X.693 [4] proponowane są dwa zbiory reguł kodowania: podstawowy i kanoniczny. Każda implementacja zgodna z regułami kanonicznymi jest równocześnie zgodna z regułami podstawowymi. W przypadku dekodowania przyjmuje się, że zgodność implementacji z podstawowymi regułami czyni ją automatycznie zdolną do odbioru wartości ASN.1 zakodowanych według reguł kanonicznych.

Dokument XML zawierający zakodowane wartości ASN.1 składa się formalnie z dwóch części: prologu XML, który nie musi zawsze występować, oraz właściwego dokumentu XML, który zawiera zakodowaną wartość typu ASN.1. Dokument taki nie może zawierać innych elementów, które są właściwe dla języka XML, takich jak na przykład instrukcjie przetwarzania czy komentarze. Dokument XML powinien być reprezentowany ciągiem oktetów, otrzymanych w wyniku zastosowania kodowania UTF-8. Element dokumentu XML, zawierający wartości ASN.1, ma przypisany typ "*XMLTypedValue*" i podlega pewnym ograniczeniom w stosunku do pełnej możliwości notacji ASN.1.

5. Przykład

Przedstawiony przykład ma na celu zwrócenie uwagi na różnice wielkości kodu transferu uzyskanego przy przesyłaniu rekordu danych reprezentowanych notacją ASN.1 (BER/PER) oraz notacją znaczników XML. Ze względu na ograniczoną objętość tekstu nie przedstawiono procedur generacji kodu transferu z notacji ASN.1.oraz nie przedstawiono wynikowych ciągów bajtów, jako że istotna jest ich liczba,a nie wartości. Otrzymany kod jest wynikiem zastosowania reguł BER oraz PER [1,2] do przedstawionego rekordu tekstowego. Przykłady tworzenia takiego kodu można znaleźć w [6].

Ogólna struktura rekordu zawierającego dane osobowe przedstawiona za pomocą standardowych typów notacji składni abstrakcyjnej ASN.1 (na podstawie ITU-T Rec. X.680, Annex C i X.693) ma postać:

```
PersonnelRecord ::= [APPLICATION 0] SET
{       name Name,
        title VisibleString,
        number EmployeeNumber,
        dateOfHire Date,
        nameOfSpouse Name,
        children SEQUENCE OF ChildInformation DEFAULT {}
}
ChildInformation ::= SET
{       name Name,
        dateOfBirth Date
}
```

```
Name ::= [APPLICATION 1] SEQUENCE
{       givenName VisibleString,
        initial VisibleString,
        familyName VisibleString
}
EmployeeNumber ::= [APPLICATION 2] INTEGER
Date ::= [APPLICATION 3] VisibleString -- YYYY MMDD
```

Po nadaniu przykładowych wartości rekord ten przybiera formę:

```
{ name {givenName "John", initial "P", familyName "Smith"},
        title "Director",
        number 51,
        dateOfHire "19710917",
        nameOfSpouse {givenName "Mary", initial "T", familyName "Smith"},
        children
        { {name {givenName "Ralph", initial "T", familyName "Smith"} ,
                dateOfBirth "19571111"},
          {name {givenName "Susan", initial "B", familyName "Jones"} ,
                dateOfBirth "19590717" }
        }
}
```

Kod transferu takiego rekordu przy zastosowaniu reguł kodowania BER zawiera 136 oktetów (jeśli określona jest długość rekordu) lub 161 oktetów (gdy długość wynika ze stosowania znacznika końca rekordu). Zastosowanie reguł PER znacznie skraca ciąg bajtów transferowych i kod wynikowy ma długość: 94 oktety (dla wersji z wyrównaniem) lub 84 oktety (wersja PER bez wyrównania).

Zapis tych samych danych jako rekord XML przedstawia się następująco:

```
person ::=
        <PersonnelRecord>
                <name>
                        <givenName>John</givenName>
                        <initial>P</initial>
                        <familyName>Smith</familyName>
                </name>
                <title>Director</title>
                <number>51</number>
                <dateOfHire>19710917</dateOfHire>
                <nameOfSpouse>
                        <givenName>Mary</givenName>
                        <initial>T</initial>
                        <familyName>Smith</familyName>
                </nameOfSpouse>
                <children>
                        <ChildInformation>
                                <name>
                                        <givenName>Ralph</givenName>
                                        <initial>T</initial>
                                        <familyName>Smith</familyName>
                                </name>
                                <dateOfBirth>19571111</dateOfBirth>
                        </ChildInformation>
                        <ChildInformation>
                                <name>
                                        <givenName>Susan</givenName>
                                        <initial>B</initial>
                                        <familyName>Jones</familyName>
                                </name>
                                <dateOfBirth>19590717</dateOfBirth>
                        </ChildInformation>
```

```
</children>
</PersonnelRecord>
```

Kod transferu takiej struktury wymaga 603 oktetów, z czego ok. 80 procent bajtów transferowych przeznaczonych jest na przesyłanie znaczników.

6. Zakończenie

Atrakcyjność zwięzłej notacji kodu transferu dla wartości wyspecyfikowanych w ASN.1 oraz popularność języka XML spowodowały, że podjęto próbę połączenia zalet obu tych języków. Notacja wartości XML stanowi bazę dla nowego zbioru reguł kodowania dla ASN.1, które zgodnie z intencją twórców powinny być opublikowane w tym roku jako nowy standard (ITU-T Rec. X.693 | ISOIEC 8825-4) pod nazwą Reguły kodowania XML (XER).

Zaprezentowany w p. 5 przykład wyraźnie ilustruje różnice wielkości kodu transferu tekstowego rekordu reprezentowanego notacją ASN.1 oraz XML. Z porównania otrzymanych wielkości wynika, że optymalny transfer uzyskany z notacji ASN.1 (84 oktety) może zmniejszyć kod wygenerowany z notacji XML (603 oktety) prawie dziesięciokrotnie. W niektórych przypadkach różnica długości kodu jest nawet stukrotna. W rezultacie aplikacja wymaga mniejszego czasu transferu, zmniejsza się ruch w sieci oraz obciążenie łączy.

Mimo że ASN.1 ma zastosowanie w wielu różnych środowiskach sieciowych, XER przewidywany jest głównie dla aplikacji internetowych (protokoły stosu TCP/IP) typu klient serwer. Obecnie do potencjalnych dziedzin zastosowań kodowania dokumentów XML w notacji ASN.1 można zaliczyć między innymi: systemy wyszukiwania informacji (np. protokół Z39.50), dostęp do katalogów (protokół LDAP), pocztę elektroniczną (SNMP), systemy medyczne oraz telekomunikacyjne [5].

LITERATURA

1. ITU-T Recommendation X.690 (2002) | ISO/IEC 8825-1:2002, Information technology – ASN.1 encoding rules: Specification of Basic Encoding Rules (BER), Canonical Encoding Rules (CER) and Distinguished Encoding Rules (DER).
2. ITU-T Recommendation X.691 (2002) | ISO/IEC 8825-2:2002, Information technology – ASN.1 encoding rules: Specification of Packed Encoding Rules (PER).
3. ITU-T Recommendation X.693 (2002) | ISO/IEC 8825-4:2002, Information technology – ASN.1 encoding rules: Specification of Packed Encoding Rules (PER).
4. http://www.w3.org.
5. http://asn1.elibel.tm.fr.

6. Kosmulska-Bochenek E.: Wymiana informacji w heterogenicznych systemach sieciowych – język ASN.1. Oficyna Wydawnicza Politechniki Wrocławskiej. Wrocław 2002.

Recenzent: Dr inż. Marcin Skowronek

Wpłynęło do Redakcji 24 marca 2003 r.

Abstract

The Extensible Markup Language (XML) is a text oriented encoding developed by W3C. It is rapidly gaining popularity and is being adopted as the underlying mechanism for many Web applications. Some drawback of this mechanism is a volume of transfer bytes. Messages specified in ASN.1 can take advantage of the compact encoding (such as Packed Encoding Rules).

XML value notation for ASN.1 types makes possible to define values using XML markup in which the element tags are derived from the ASN.1 type names and identifiers present in ASN.1 specification. BER or PER generated messages described by ASN.1 are a small fraction of the size of messages described by XML, so BER or PER generated messages described by ASN.1 can travel more rapidly then XML messages described using other schemas. The example shows the differences in size of transfer code using ASN.1 (BER, PER) notation and XML (performance improves by a factor almost 100).

Adres

Elżbieta KOSMULSKA-BOCHENEK: Politechnika Wrocławska, Wydziałowy Zakład Informatyki, Plac Grunwaldzki 9, 50-370 Wrocław, Polska, kosmulska@ci,pwr.wroc.pl .