

XML SCHEMA OF UNIMARC FORMAT VARIANT AND BIBLIOGRAPHIC RECORD IN BISIS SOFTWARE SYSTEM¹

Branko Milosavljević² and Bojana Dimić³

Abstract. BISIS library software system is based on a UNIMARC format variant called YUMARC format. An XML schema for description of the YUMARC format as well as an XML schema of the record formed in accordance with the YUMARC format are proposed in the article.

AMS Mathematics Subject Classification (2000): 68U35

Key words and phrases: XMLSchema, Bibliographic records, BISIS

1. Introduction

MARC (*Machine Readable Cataloguing*) bibliographic format [1], as a format for recording and exchange of bibliographic data in a machine-readable form, has represented a basis for defining other bibliographic MARC formats intended for the needs of library systems in different countries. The UNIMARC (*Universal Machine Readable Cataloguing*) format was one of the formats developed for European countries [2]. As a universal cataloguing format of bibliographic materials, this format enables exchange of data in a machine-readable form among national and international bibliographic institutions.

A connection between the structure of bibliographic records and the structure of an XML document was noticed in early phases of the XML technology development. One of the common points is the existence of the elements which can be mandatory or optional, repeatable or divided into subelements. Every XML structure can be formally described by using the XML schema that is defined by the XMLSchema language [3]. The rules for creating XML schema correspond to the rules for creating the structure of the bibliographic records prescribed by some of the library standards. The biggest advantage of using XML for representing bibliographic data is reflected in simpler data processing in open distribution systems. The already developed software tools for work with XML documents are also to be credited for this.

The BISIS library software system has been in the process of development at the Novi Sad University since 1992, and several versions have been made so far [4]. The system is based on the UNIMARC format, is supplemented for

¹The work is supported by the Ministry for Science, Technologies and Development of the Republic of Serbia (Project No. 144017)

²Faculty of Technical Sciences, Novi Sad

³Faculty of Sciences, Novi Sad

national use and is adapted to the BISIS library software system. This format was named YUMARC.

The development of BISIS system Version 4 based on the XML technologies is under way. The results achieved so far were published in the articles [5]-[7]. As proved in the article [5], all concepts of the UNIMARC format can be modeled by using XML Schema language. It was also proved that all concepts of bibliographic record within the UNIMARC format can be modeled by using XML Schema language was also proved. The process of modeling the concepts of bibliographic record controls using the same language was demonstrated in the article [7]. Besides, an XSLT specification for describing the controls that can not be described by XML Schema language was added.

A new XML editor for processing bibliographic materials will be implemented for the Version 4. The implementation of the editor is based on the XML schemas of YUMARC format and the bibliographic record made in this format. On the basis of the XML schemas given in [5], a modification of these schemas adapted for the software architecture of the editor for bibliographic material processing is described in this paper.

2. YUMARC format

According to YUMARC format, a bibliographic record consists of a series of fields. A field consists of a series of subfields and can be assigned two indicators, at the most. Subfields represent basic data conveyors for publication data, whereas the role of indicator is to closely determine the meaning of data in a field or appearance of data on a display. Fields are defined by a tag consisting of three numerical characters and are divided in ten blocks. The first digit in the tag indicates the block to which the field belongs (i.e., the field with the tag 200 belongs to the block 2 named Main description block). According to the YUMARC format, a field is characterized by the extent of how repeatable, mandatory and secondary it is. A field is repeatable if it can be repeated several times in a record. A field is mandatory when it must appear in a record. If a field is secondary it can appear as the content of a subfield in the field of the fourth block (Linking bibliographic record block).

Field data is separated in several subfields. Whithin a field, subfields are defined by their designation consisting of an alphanumeric character and can be repeatable and mandatory.

A subfield data can be a text or code from one of the YUMARC format code books. For some subfields, a maximum permitted length of data can be defined. Also, there are some subfields which have defined default values, i.e. values that are determined beforehand as subfield content. The code books defined by YUMARC format can apply to a subfield (*internal code books*) or to several subfields (*external code books*). The latter are sorted by types, for example, country code book, language code book, authorship code book, etc.

The block 4 is a block for linking bibliographic records. A link between the records is established by entering textual or numeric data into the corresponding fields or subfields of the block 4. This data can be an identification number of

the record with which a link is made or some new secondary fields by which a linking document is described.

The block 9 is defined in the national framework. It is intended for entering data that are of local importance for a library network or a particular library. This block contains subfields the data of which is structured onto another level of the substructure – the level of the *subsubfields*. The subsubfields have the same characteristics as subfields, and code books for them are also available.

By using different sets of fields and subfields of the YUMARC format, it is possible to describe all types of publications: books, journals, newspapers, doctoral theses, articles, music works, maps, manuscripts, computer files, etc.

The following example illustrates processing of a monographic publication according to the YUMARC format. Say a publication with a title “Concepts of Programming Languages” is given. The author of the book is Robert W. Sebesta and its ISBN (International Serial Book Number) is 0-8053-7133-8. The publication language is English, the country of publication is the USA. As far as location data is concerned, there is one copy of this publication in the library with the call number 20111 and the accession number 000020111. The YUMARC record of this publication is shown in Figure 1.

```
001 ## [7]ba[a]i[b]a[c]m[d]0
010 ## [a]0-8053-7133-8
100 ## [b]d[c]1996[h]scr
101 0# [a]eng
102 ## [a]usa
105 ## [a]a
200 0# [a]Concepts of Programming Languages[f]Robert W. Sebesta
205 ## [a]3. izd.
210 ## [a]Reading [etc.][c]Addison-Wesley Publishing Company[d]1996
215 ## [a]xv, 634 str.[c]ilustr.[d]24 cm
700 #1 [4]070[a]Sebesta[b]Robert W.
992 ## [b]crsale9701-old
996 0# [d]<1>M<n>20111[f]000020111
```

Figure 1: A YUMARC record

The record is displayed in multiple lines, with one line for each record field. The first three characters in the record stand for a field identifier, the two subsequent ones for the values of the first and second indicator. If their values are not defined, the character # is entered. The designations of the subfields are entered between the characters [and], with a subfield content entered behind them. The subsubfield designations are entered between the characters < and >.

Now we give a description of the structure of the bibliographic record shown in Figure 1. The Block 0 is intended for identification. The necessary coded data describing the publication are entered in the field 001. Book’s ISBN is the content of the subfield *a* of the field 010. The fields of block 1 are intended for

coded information about the publication type, cataloguing language, publication language, etc. The codes are taken from one of the YUMARC format code books. The title proper of a book is entered in the subfield *a* of the field 200, and the author name in the subfield *f*. The data about the author is also entered in the field 700 whereby the second indicator of this field determines whether it is the name or the family name of the author that is entered in the subfield *a*. The field 210 is intended for edition statement and the field 215 for collation statement. The call number and accession number data of a particular copy of the publication are entered in the subfields of the field 996. The subfield *d* contains the subsubfields for structuring call number data. Hence, a location mark is entered in the subsubfield *l* whereas the number determining the place in the given location is entered in the subsubfield *n*.

3. XML Schema of the YUMARC format

A model of the YUMARC format specification is presented by the XML schema, an instance of which is a document into which the data about the elements of the format will be stored.

A graphic representation of the YUMARC format XML schema is shown in Figure 2. The root element is *format* and it has three attributes – *formatID*, *name* and *description* – which stand for format identification number, name and format description respectively.

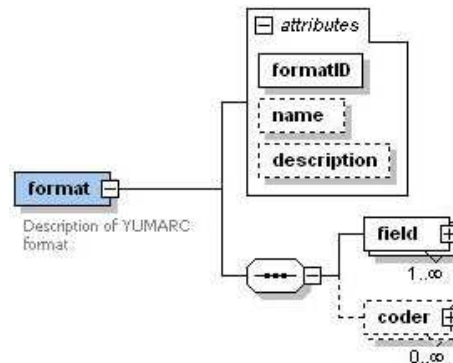


Figure 2: Root element of the XML schema of YUMARC format

The element *format* contains a series of subelements concerning the specification of format fields and a series of elements concerning external code books.

A specification of format field is modeled by the element *field*, whose detailed representation is shown in Figure 3. A field is modeled in the XML schema in such a way to contain five attributes which determine it more closely, a maximum of two subelements which define indicators and a series of subelements by which the subfields of the given field is specified.

The element *field* contains the following attributes:

- *name* – three-digit designation of the field,
- *description* – field description, i.e., description of the bibliographic elements that comprise a field according to the YUMARC format rules,
- *mandatory* – logical value that determines if a field is mandatory,
- *repeatable* - logical value that determines the repeatability of a field,
- *secondary* - logical value that determines if a field is secondary.

The indicator specification is modeled by the element *indicator*. The cardinality of occurrence of this element within the field it relates to is 0..2, which means that the condition is modeled for that field to have two indicators at the most, but can have none. A detailed specification of the element *indicator* is presented in Figure 4.

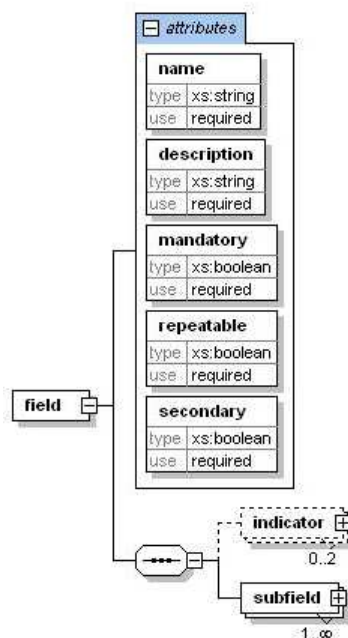


Figure 3: Model of a YUMARC format field – element *field*

The indicator is defined by three attributes of the element *indicator* and these are:

- *index* – can have value 1 or 2, depending on whether it is the first or the second indicator that is in question,
- *description* – indicator description, i.e., information carried by the indicator and described in accordance with the rules defined by the YUMARC format,

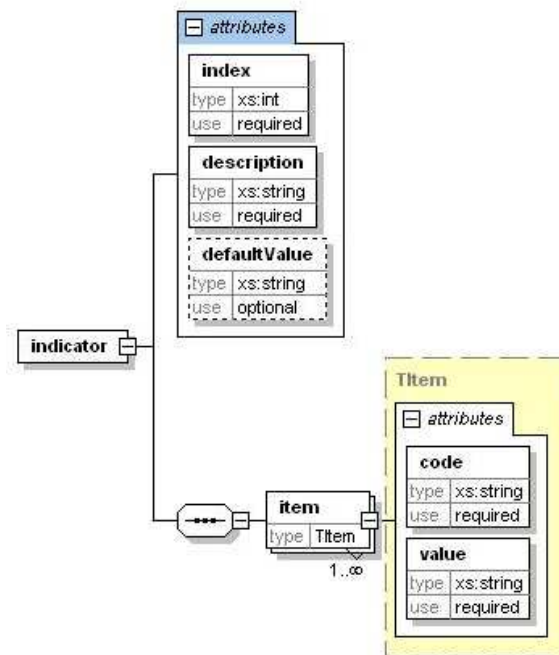


Figure 4: Model of an indicator – element *indicator*

- *defaultValue* – default value of indicator, that is, the value that the indicator will have in the bibliographic record in case of not being assigned another value; use of the attribute is optional.

The element called *indicator* is a sequence of elements called *item* by which a possible value of the indicator is modeled and is of the *TItem* type that is defined globally. The type *TItem* defines two attributes, the attribute *code* standing for the value that the indicator can have in a bibliographic record and the attribute *value* standing for the textual description of the meaning of the value.

A detailed representation of the element *subfield* is shown in Figure 5. The specification of the YUMARC format subfield is determined by six attributes. A code book of subfields modeled by the element *coder* or a series of elements that represent subsubfields defined within YUMARC format can appear as a subelement of the element that stands for the format subfield.

The attributes of the element *subfield* have the following meanings:

- *name* – subfield designation in the form of an alphanumeric character,
- *description* – description of a subfield, that is, description of the bibliographic elements that comprise a subfield in accordance with the YUMARC format rules,

- *mandatory* – logical value that determines if a subfield is mandatory,
- *repeatable* – logical value that determines repeatability of a subfield in a field,
- *length* – permitted length of data in a subfield,
- *defaultValue* – default value of a subfield, that is, the value that a subfield will have in the bibliographic record in case of not being assigned another value; use of this attribute use is optional.

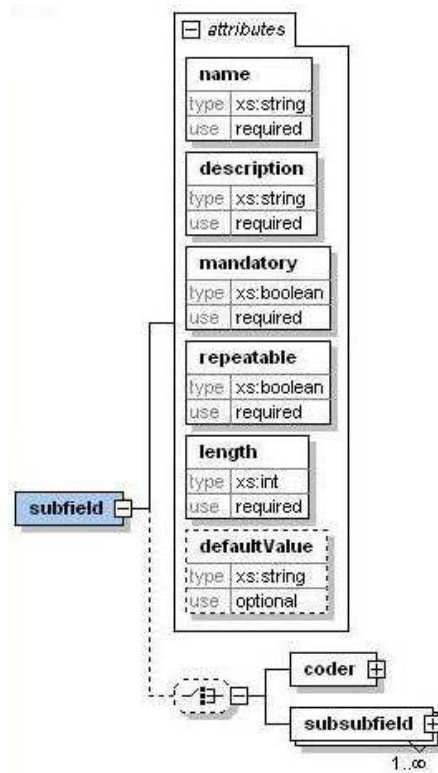


Figure 5: Model of subfield specification – element *subfield*

The code book of subfields is modeled by the element *coder*, a detailed presentation of which is shown in Figure 6. The attribute *external* is of the logical type and has the value *true* if the subfield to which the code book is related assumes values from an external code book. In that case, the attribute *type* has a numerical value that is allocated to the corresponding external code book. In the above case, the sequence of elements *item* the use of which is optional will not appear. In case of an internal code book being related to a subfield, the attribute *external* will have a value *false*, whereas the attribute

type will not have a defined value. Such a case is prescribed by the schema because the use of the attribute *type* is optional. In case of an internal code book for a subfield, a sequence of subelements *item* of the globally defined type *TItem* appears within the element *coder*.

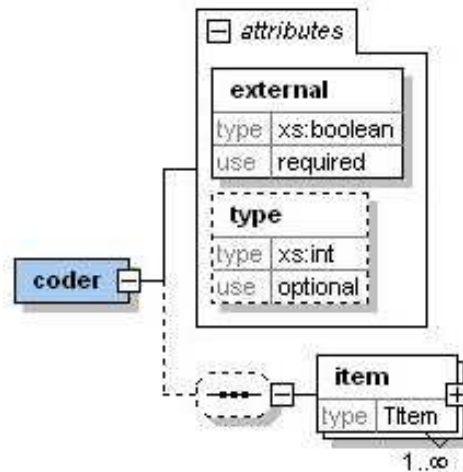


Figure 6: Model of the subfield code book – element *coder*

Specification of the YUMARC format subsubfield is identical to the specification of the subfield. The only difference in the models of these two elements consists in the fact that subsubfields as subelements can have only an element that relates to the subsubfield code book and cannot have additional substructure levels.

A graphical presentation of the external code book model that occurs as a subelement of the root element *format* is shown in Figure 7. The external code book is modeled by the element *coder* that represents an extension of the type *TCoder* with the attributes *external*, *type* and *name*. The type *TCoder* is defined as a sequence of code book items represented by the element *item*, described earlier in the article. The attribute *external* for the external code book obtains the value *true*. The attribute *type* represents a unique numerical value for the external code book that serves to reference from the subfield that will take the code book's value. The textual description of the external code book is modeled by the attribute *name*.

3.1. XML document of the YUMARC format

An XML document that represents occurrence of the XML schema described in the previous section contains the YUMARC format specification.

A portion of the XML document from the YUMARC format specification that relates to the specification of the field 200 is shown in Listing 1. The attribute *name* contains a field name, whereas the attribute *description* contains

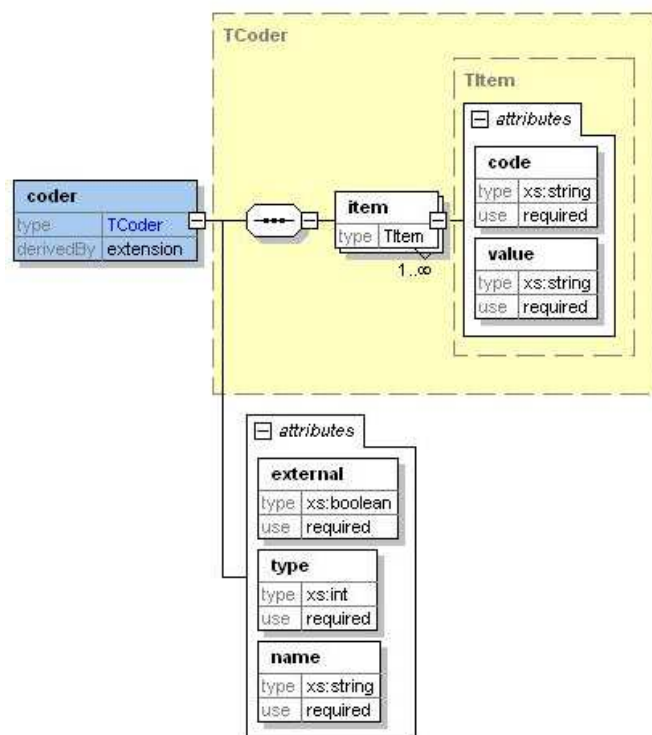


Figure 7: Model of the external code book

the text “Title proper and statement of responsibility” that represents description of the field 200. The attribute *mandatory* has the value *true*, which means the field is mandatory. The defined first indicator of the field 200 is presented in the given XML segment by the nested element *indicator* with the attribute value *index* that equals 1. Description of the indicator is the attribute value *description* of the element *indicator*. The default value of the indicator in question is 0, and it is recognized by the attribute value *defaultValue* of element *indicator*.

The elements *subfield*, containing the specification of subfield of the field 200 according to YUMARC format, appear as the subelements of the element by which the field 200 is specified. The subfield *z* of the field 200, described as “*Language of the parallel title*”, is not mandatory nor repeatable, and cannot have length bigger than 3. This subfield is coded, which is presented in the XML document by the nested element *coder*. The value of the attribute *external* of the element *coder* is *true*, which means that the subfield takes the values from the external code book. Identification of the external code book for the subfield is accomplished on the basis of the value of the element *coder*’s attribute *type*.

```

<field name="200" description="Stvarni naslov i podaci o odgovornosti" mandatory="true"
repeatable="false" secondary="true">
  <indicator index="1" description="Indikator važnosti glavnog stvarnog naslova"
defaultValue="0">
    <item code="0" value="Naslov nije značajan"/>
    <item code="1" value="Naslov je značajan"/>
  </indicator>
  <subfield name="a" description="Glavni stvarni naslov" mandatory="true"
repeatable="false" length="0"/>
  <subfield name="b" description="Opšta oznaka građe" mandatory="false"
repeatable="false" length="0"/>
  <subfield name="c" description="Glavni stvarni naslov drugog autora"
mandatory="false" repeatable="false" length="0"/>
  <subfield name="d" description="Uporedni stvarni naslov" mandatory="false"
repeatable="false" length="0"/>
  <subfield name="e" description="Podnaslov" mandatory="false" repeatable="false"
length="0"/>
  <subfield name="f" description="Prvi podatak o odgovornosti" mandatory="false"
repeatable="false" length="0"/>
  <subfield name="g" description="Drugi i sledeći podatak o odgovornosti"
mandatory="false" repeatable="true" length="0"/>
  <subfield name="h" description="Oznaka dela" mandatory="false"
repeatable="false" length="0"/>
  <subfield name="i" description="Stvarni naslov dela" mandatory="false"
repeatable="false" length="0"/>
  <subfield name="v" description="Broj sveske" mandatory="false"
repeatable="false" length="0"/>
  <subfield name="z" description="Jezik uporednog stvarnog naslova"
mandatory="false" repeatable="false" length="3">
    <coder external="true" type="1"/>
  </subfield>
</field>

```

Listing 1.

A part of the XML document of the format specification relating to the specification of the subfield *a* of the field 105 that in its turn takes values from an internal code book is shown on Listing 2. The internal code book in the given example is represented by the element *coder* that is a subelement of the element *subfield* for the specification of subfield *a*. Since an internal code book is in question, the attribute *external* has the value *false*. The elements *item*, the attributes of which describe the item of the internal code book, appear as subelements of the element *coder*. The code *d* has description “grafiki prikazi” that appears as a value of the attribute *value* of the corresponding element *item*.

```

<field name="105" description="Polje kodiranih podataka: knjige" mandatory="false"
repeatable="false" secondary="false">
  <subfield name="a" description="Kodovi za ilustracije" mandatory="false"
repeatable="false" length="1">
    <coder external="false">
      <item code="a" value="ilustracije"/>
      <item code="b" value="zemljopisne karte"/>
      <item code="c" value="portreti"/>
      <item code="d" value="grafički prikazi"/>
      <item code="e" value="planovi"/>
      <item code="f" value="table (listovi, koji sadrže slikovnu
građu...)" />
      <item code="g" value="note"/>
      <item code="h" value="faksimili"/>
      <item code="i" value="grbovi"/>
      <item code="j" value="genealoške table"/>
    </coder>
  </subfield>
</field>

```

Listing 2

4. XML Schema of the YUMARC format bibliographic record

XML schema of a bibliographic record represents the model of a record representation in the XML document form. Graphical representation of an XML schema is shown in Figure 8.

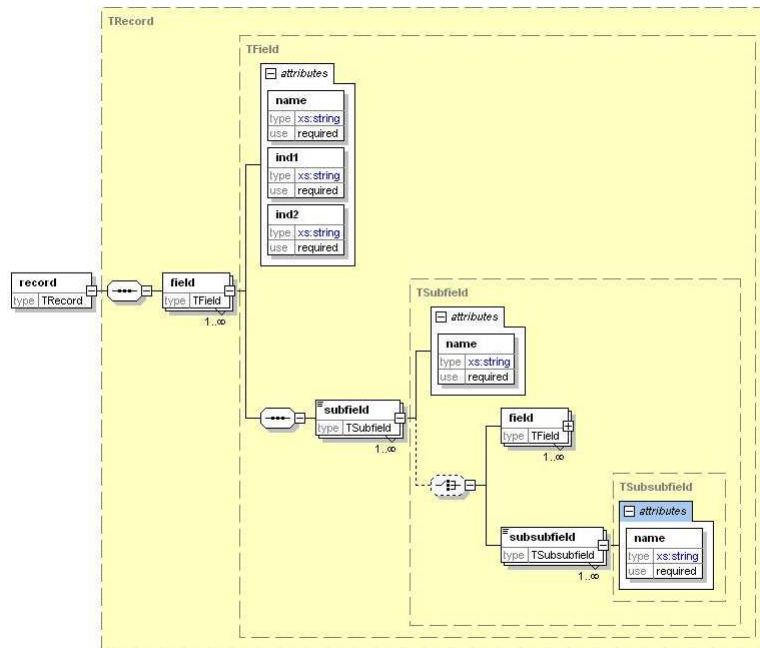


Figure 8: Graphical representation of the XML schema of bibliographic record in accordance to YUMARC format

The root element of the XML document by which a bibliographic record is represented is the element called *record* of the composite type *TRecord*. The type *TRecord* is defined as a sequence of elements *field* by which the field of the record is modeled. The element *field* is of the composite type *TField* and has three defined attributes as well as a sequence of the elements *subfield*. The attribute *name* of the element *field* models the field name, whereas the attributes *ind1* and *ind2* correspond to the values of the first and second indicator of the field.

The record subfields are represented by the element *subfield* and this element is of the composite type *TSubfield*. The subfield name is placed in the attribute *name*, and a subfield can contain other fields or subsubfields as subelements. The fields that occur as a content of subfields are secondary fields. Since they have same properties as ordinary fields in a record, these fields are also of the type *TField*. The subsubfield of a record is modeled by the element *subsubfield*. The subsubfield name will appear as a value of the attribute *name* of the element

subsubfield.

4.1. XML document of bibliographic record

The XML document of bibliographic record described in Section 2.2 is shown in Listing 3. The XML representation of a bibliographic record is to be found between the starting and ending tag of the element *record*.

Fields in a record are sorted by tags. The second field in a record is the field 010, and its tag appears as the content of the attribute *name* of the second element *field*. As this field has no defined indicators, the values of attributes *ind1* and *ind2* are blank characters. The field 010 in the given bibliographic record contains a subfield *a* with the value *0-8053-7133-8*. The given subfield is represented by the element *subfield* that appears as the subelement of the element by which the field 010 is represented. The content of the subfield is entered between the starting and ending tag of the element *subfield* whose attribute *name* equals *a*.

The value of the first indicator of the field 200 equals 0, which is represented in this XML document of the bibliographic record with the attribute *ind1="0"* of the element *field* with the value of the attribute *name* that equals 200.

The subfield *d* of the field 996 contains subsubfields that are represented by the elements *subsubfield* within the starting and ending tags of the corresponding element *subfield*. The first subsubfield has the name *l* that can be read from the value of the attribute *name*. The content of the subsubfield named *M* is placed between the starting and ending tags of the corresponding element *subsubfield*.

```

<record>
  <field name="001" ind1=" " ind2=" ">
    <subfield name="7">ba</subfield>
    <subfield name="a">i</subfield>
    <subfield name="b">a</subfield>
    <subfield name="c">m</subfield>
    <subfield name="d">0</subfield>
  </field>
  <field name="010" ind1=" " ind2=" ">
    <subfield name="a">0-8053-7133-8</subfield>
  </field>
  <field name="100" ind1=" " ind2=" ">
    <subfield name="b">d</subfield>
    <subfield name="c">1996</subfield>
    <subfield name="h">scr</subfield>
  </field>
  <field name="101" ind1="0" ind2=" ">
    <subfield name="a">eng</subfield>
  </field>
  <field name="102" ind1=" " ind2=" ">
    <subfield name="a">usa</subfield>
  </field>
  <field name="105" ind1=" " ind2=" ">
    <subfield name="a">a</subfield>
  </field>
  <field name="200" ind1="0" ind2=" ">
    <subfield name="a">Concepts of Programming Languages</subfield>
    <subfield name="f">Robert W. Sebesta</subfield>
  </field>
  <field name="205" ind1=" " ind2=" ">
    <subfield name="a">3. izd.</subfield>
  </field>

```

Listing 3. XML document of a bibliographic record (part 1)

```

<field name="210" ind1=" " ind2=" ">
  <subfield name="a">Reading [etc.]</subfield>
  <subfield name="c">Addison-Wesley Publishing Company</subfield>
  <subfield name="d">1996</subfield>
</field>
<field name="215" ind1=" " ind2=" ">
  <subfield name="a">xv, 634 str.</subfield>
  <subfield name="c">illustr.</subfield>
  <subfield name="d">24 cm</subfield>
</field>
<field name="700" ind1=" " ind2="1">
  <subfield name="4">070</subfield>
  <subfield name="a">Sebesta</subfield>
  <subfield name="b">Robert W.</subfield>
</field>
<field name="992" ind1=" " ind2=" ">
  <subfield name="b">crsale9701-old</subfield>
</field>
<field name="996" ind1="0" ind2=" ">
  <subfield name="d">
    <subsubfield name="1">M</subsubfield>
    <subsubfield name="n">2011</subsubfield> </subfield>
    <subfield name="f">00002011</subfield>
  </field>
</record>

```

Listing 3. XML document of a bibliographic record (part 2)

5. Conclusion

An XML schema of the document containing specification of all the elements of the YUMARC format has been created. The advantage of such a representation of data about the format is reflected in the simple processing of these data that is based on the character of structuredness of an XML document. The use of the data about format is made easier with the existing software tools for XML document manipulation. Contained in the XML document of YUMARC format are all the characteristics of the format elements on the basis of which it is possible to carry out a quality control of bibliographic the records. The continuation of the research initiated in this article consists in the consideration of possible software solutions for carrying out the control of bibliographic record correctness. Such software solutions would have an XML document as an input document containing a specification of bibliographic record control.

In the practical implementation of the library information systems, the most salient are the YUMARC format subsets which refer to a particular type of publication. The presented idea of using the format specification model implies the creation of a unique XML document which contains the overall format, whereby for each type of publication a list of format elements related to that type is created. As to the mentioned solution, it is necessary to implement the system that would – based on data about the entire format and the list of elements related to the publication type – single out a subset of the format specification for a particular publication type.

References

- [1] MARC Format Overview. Network Development and MARC Standards Office, Library of Congress, 08/10/2006. <http://www.loc.gov/marc/status.html>.

- [2] UNIMARC Manual: Bibliographic Format/ International Federation of Library Associations and Institutions. IFLA Universal Bibliographic Control and International MARC Programme, New Providence, London 1994. <http://www.ifla.org/VI/3/p1996-1/sec-uni.htm>
- [3] XMLSchema, <http://www.w3.org/XML/Schema> (30/10/2006)
- [4] Surla, D., Konjović, Z., Overview of the Development of the Library Information System BISIS, Conference Proceedings: International Conference on Distributed Library Information Systems, Ohrid, 2004, 13-18.
- [5] M Jakšić. Modelling of Bibliographical Standards in XML Technology. *Softw. Pract. Exper.* 34 (2004) 1051-1064.
- [6] Vidaković J, Racković M. Generating content and display of library catalogue cards using XML technology. *Softw. Pract. Exper.* 36 (2006) 513-524.
- [7] Budimir G, Surla D. Quality control system of XML bibliographic records. *Novi Sad J. Math.* Vol. 34 No. 1 (2004) 107-130.

Received by the editors December 26, 2006