

ОРГАНИЗАЦИЯ ВЗАИМОДЕЙСТВИЯ ВИРТУАЛЬНОГО СТРОКОВОГО ПРОСТРАНСТВА ТЕХНОЛОГИЧЕСКИХ ДАННЫХ И КОММУНИКАТИВНОГО ФОРМАТА MARC

А.Н. Филиппов, А.Н. Сисюков

В статье описана структура российской версии коммуникативного формата MARC и его применение. Показан пример наложения полей данных формата в триплексное представление ВСПТД. Рассмотрен реализованный механизм макета документа для записей в формате RUSMARC.

Введение

Эффективное обеспечение информацией экспертных систем предполагает возможность взаимодействия с коммуникативными форматами, широко представленными в сети Интернет. Направления, в которых проводились работы по развитию формата ВСПТД [3, 4], были связаны также и с получением информации, представленной в Российском коммуникативном формате RUSMARC. Российский коммуникативный формат не является совершенно новой, отдельно стоящей разработкой – это российская версия Международного коммуникативного формата UNIMARC в трактовке и категориях действующих в России ГОСТов и Правил каталогизации, базирующаяся на выборе наиболее общих схем представления элементов данных. С практической точки зрения это означает, что любая запись, переданная в Российском коммуникативном формате, должна адекватно восприниматься любым программным обеспечением, о котором заявлено, что данное программное обеспечение работает с форматом UNIMARC.

Запись в формате MARC может быть принята из глобальной сети по протоколу Z39.50. Реализованный клиент может общаться с серверами в сети, поддерживающими спецификацию ISO 23950. При наличии в полученной записи полей категории «электронный документ» может быть получена ссылка на полнотекстовый вариант документа.

Структура формата RUSMARC

Российский коммуникативный формат (рис. 1), как и UNIMARC, реализует библиографическую запись в машиночитаемой форме в структуре международного стандарта ISO 2709. Каждая библиографическая запись, подготовленная для обмена, должна содержать:

- **МАРКЕР ЗАПИСИ** – состоящий из первых 24-х символов элемент записи, в котором определяются ее основные характеристики;
 - **УКАЗАТЕЛЬ**, состоящий из 3-х цифровой метки для каждого поля данных, длины поля и позиции начального символа, относящегося к первому полю данных;
 - **ПОЛЯ ДАННЫХ** переменной длины, отделенные друг от друга разделителем поля.
- Общая структура:

МАРКЕР ЗАПИСИ	УКАЗАТЕЛЬ	ПЕРЕМЕННЫЕ ПОЛЯ	%
---------------	-----------	-----------------	---

% – Разделитель записи

В дальнейшем устанавливается, что данным в полях не обязательно предшествуют индикаторы, и они не обязательно подразделяются на подполя. Российский коммуникативный формат, как и UNIMARC, использует следующие положения, установленные Международным стандартом ISO 2709.

ISO 2709 предписывает, что каждая запись начинается с 24-х символьного маркера записи. Он содержит данные, относящиеся к структуре записи, определения которых

даются в стандарте ISO 2709, а также некоторые элементы данных, выделенные для особого применения. Эти элементы данных, определяемые при применении, связаны с типом записи, ее библиографическим уровнем и положением в иерархии уровней, степенью полноты записи и использованием правил ISBD или других основанных на ISBD правил подготовки записи. Элементы данных в маркере записи предназначены прежде всего для обработки записи и только косвенно – для идентификации самих библиографических материалов.

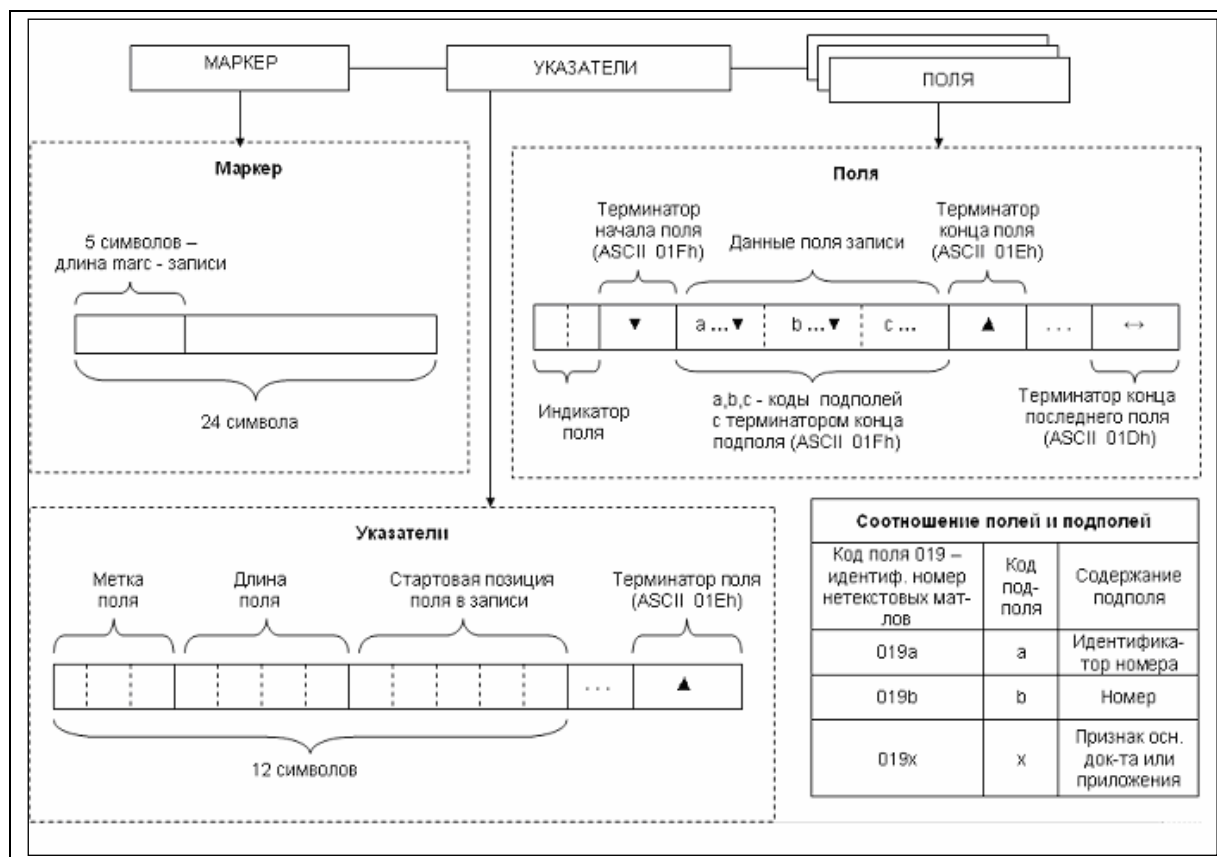


Рис. 1. Структура MARC–записи

За маркером записи следуют указатели. Каждая статья указателя состоит из трех частей – трехцифровой метки, четырехзначного числа, указывающего длину поля данных, и пятизначного числа, указывающего начальную символьную позицию. Другие символы в статье указателя не допускаются.

Более подробное описание формата RUSMARC представлено в различных источниках [1, 2, 5].

Для работы с форматом MARC применяются сложные и дорогие программные системы, приобретение которых не всегда целесообразно. В связи с этим была поставлена задача перевода данных из формата MARC в представление ВСПТД (конвертирование). В рамках этой проблемы были выработаны следующие соглашения:

- имена подполей представляются в виде nnnx, где nnn – код поля (в формате MARC), а x – код подполя;
- имена индикаторов представляются в виде Innn;
- в качестве префиксов в триplete применяется номер повторяемости поля;
- повторяемые подполя представляются как вектор значений с заданным разделителем;
- маркеру MARC-записи присваивается код поля 000.

Пример представления записи в формате RUSMARC и ВСПТД

Рассмотрим пример библиографической записи:

Митрофанов С.П., Куликов Д.Д., Миляев О.Н., Падун Б.С. Технологическая подготовка гибких производственных систем. / Под общей редакцией С.П. Митрофанова. Л.: Машиностроение, 1987. – 352 с.

Представление записи в формате RUSMARC:

000: \$00755naa2 2200373 450
001: \$1:1-0039129
005: \$20070124150255.8
100: \$a20061015a19939999m a0rusa02|||||ca
101:0 \$arus
102: \$aRU
106: \$ar
110: \$aafaaaar0az0
200: 1\$a Технологическая подготовка гибких производственных систем. \$gПод общей редакцией С.П. Митрофанова
210: \$aЛ.\$c Машиностроение \$d1987
215:: \$a352 с.
700: 1\$a Митрофанов \$bС.П. (*)
701: 1\$a Куликов \$bД.Д.
701: 1\$a Миляев \$bО.Н.
701: 1\$a Падун \$bБ.С.

Поясним фрагмент записи (*). Поле 700 содержит в форме точки доступа имя лица, имеющего первичную интеллектуальную ответственность (первое или единственное лицо в заголовке, содержащем имя индивидуального автора). Индикатор со значением «1» означает, что имя лица записано под фамилией (родовым именем, отчеством и т. д.). Подполе «а» содержит фамилию со значением «Митрофанов». Подполе «b» содержит инициалы со значением «С.П.».

Представление записи в формате ВСПТД:

\$0.000='01255naa2 2200373 450'; \$0.001'1:1-0039129';
\$0.005='20070124150255.8'; \$0.100a='20061015a19939999m a0rusa02|||||ca';
\$0.101a='rus'; \$0.102a='RU'; \$0.106a='r';
\$0.110a='afaaaar0az0'; \$0.200a='Технологическая подготовка гибких производственных систем'; \$0.200g='Под общей редакцией С.П. Митрофанова'; \$0.210a='Л.';
\$0.210c='Машиностроение'; \$0.210d=1987; \$0.215a='352 с'; \$0.700a='Митрофанов';
\$0.700b='С.П.'; \$0.701a='Куликов'; \$0.701b='Д.Д.'; \$1.701a='Миляев';
\$1.701b='О.Н.'; \$2.701='Падун'; \$2.701b='Б.С.'; \$0.I200a='1'; \$0.I700='1'; \$0.I701='1';
\$1.I701='1'; \$2.I701='1';

Организация макетов документов в формате RUSMARC

При извлечении информации из сети Интернет, представленной в формате MARC, возникает необходимость представления содержания записей в удобной для прочтения форме. Для этих целей разработан язык макетов документов. Макет документа состоит из заголовка макета и тела макета.

Заголовок макета. Номер документа указывается в первой строке (управляющей) после символа #. Далее в этой же строке указываются в триплексном формате параметры документа. Главные из них следующие:

\$Q.W=nn – ширина документа, 5<nn<128, по умолчанию – 60 (в символах).

\$Q.H=nn – высота документа для экранного редактирования, $3 < nn < 23$, по умолчанию – 10 (в строках).

\$Q.Z='Заголовок документа';

Пример первой (управляющей) строки документа:

#1 \$Q.W=48;\$Q.H=16;\$Q.Z='Осн. карточка';

Здесь #2 – номер макета 2; ширина документа – 48 символов; высота документа – 16 строк; заголовок документа – «Осн. карточка».

Тело макета документа. Тело документа может условно состоять из одной или двух частей – левой и основной. Если документ состоит из двух частей, то в конце макета размещается следующая конструкция:

\$LEFT: \$Q.W=w;

Содержание

\$ENDLEFT:

Задается левая часть документа. \$Q.W=w; - ширина левой части документа.

Базовый синтаксис

Любой текст, который необходимо отобразить в макете, задается в кавычках “. Для задания текста из марковского поля указывается его код со знаком "\$". Т.е. последовательно указываются марковские коды реквизитов со знаком "\$", участвующих в описании документа в той последовательности, которая характерна для данного документа.

Код, указанный с новой строки, означает его запись с новой строки. Однако если в конце какой-либо строки макета документа стоит \$CL (Continue Line), то следующая строка считается продолжением текущей. Между кодами в произвольном порядке может быть указано символьное выражение в кавычках.

В макете документа можно использовать средства для отображения условности печати того или фрагмента документа. Например: 'n' – не печатать предыдущий фрагмент макета документа при отсутствии последующего реквизита, заданного марковским кодом; 'r' – задает повторяемость; 'p' – не печатать последующий фрагмент макета документа при отсутствии предыдущего реквизита; 'd' – декодирование, если в записи хранится код реквизита, то будет сформировано значение по коду; \$CL – следующая строка есть продолжение предыдущей; \$NL – перевод на новую строку (антипод \$CL); \$NR – указатель номера поля при распечатке повторяемых полей (применяется внутри квадратных скобок в конструкции вида: :r[...\$NR...])

Пример: пусть задан список авторов, которые требуется распечатать с нарастающей нумерацией:

701 #0\$aИванов\$bИ.С

701 #0\$aПетров\$bА.Н

701 #0\$aСидоров\$bЯ.Г.

Применение конструкции вида:

:r[\$NR"- "\$701a#|r-r :a " " \$701b#|r-r :a " "]

приведет к формированию в документе следующей строки:

1– Иванов, И.С. 2 – Петров, А.Н. 3 – Сидоров, Я.Г.

При выводе повторяемых значений по группе различных полей, рассматривая их как реляционные структуры (таблицы), можно применять следующую конструкцию:

:r["текст1" \$cod1#xr-r "текст2" \$cod2#xr-r..."текстN"...\$NL],

где cod1 и x описаны выше, а 'r-r' – указатель на повторяемость с нарастанием. При этом можно ввести внутри квадратных скобок \$NL – перевод строки. Допустимо локальное условие внутри повторяемых полей:

```
:r[$cond: ...условие... $endcond: $cod1#xr-r "текст2" $cod2#xr-r..."текстN"...$NL]
```

Здесь условие может быть таким же, как и при описании \$COND: (следует иметь в виду, что начало и конец локального условия пишутся с маленькой буквы). Часто применяется следующая конструкция для определения повторяемости полей ...

```
$codn# -r-r...
```

Эта конструкция обеспечит вывод повторяемых полей с повторяемыми подполями, выведенными с разделителем ‘-‘. Знаки пунктуации между полями определяются в тексте (."текстN"). Реализован также механизм формирования неравномерной повторяемости поле.

Организация выполнения макета по условию

В системе предусмотрено выполнение различных вариантов макета по условию:

```
$COND: условие n
```

```
текст макета n
```

```
$ENDCOND:
```

```
$COND: условие 1
```

```
текст макета 1
```

```
$ENDCOND:
```

```
$COND: условие 2
```

```
текст макета 2
```

```
$ENDCOND:
```

Некоторые функции условия:

PRE (\$Code) – присутствие значения,

ABS (\$Code)– отсутствие значения,

EQU (\$Code=Value)- может осуществляться также проверка на равенство, неравенство, включение, не включение поля заданному значению:

VPRE(\$cod), VABS(\$cod) – проверка наличия/отсутствия кода подполя в текущей словарной таблице. В этом случае код cod может иметь формат: с*, где с – начальная часть кода. Символ ‘*’ указывает на проверку наличия/отсутствия всех подполей, имеющих заданную начальную часть кода.

```
$COND: EQU($100a[1,2]="ab")
```

LENDIGIT – выдает число цифр в значении подполя.

! Пример: IF(PRE(\$016) & LENDIGIT(\$016)!=10) THEN MSGBOX(Неверный номер ISBN - количество цифр!=10) .

DATE Допускается фильтрация по дате:

1. в заданном диапазоне: DATE(\$005 :19970101-19981231);

2. вне диапазона: DATE(\$005 :19970101<>19981231).

Максимальная глубина условия вложения не ограничена.

Макросы

Некоторые повторяющиеся фрагменты макета можно вынести в отдельные макросы. Макросы располагаются после описания всех макетов. Начало списка макросов – \$MACROS:, конец списка макросов – \$ENDMACROS:.

Все макросы находятся внутри области макросов. Имя макроса пишется с первой позиции и имеет вид \$name. Конец макроса – //. Вызов макроса в макете может быть применен в любом месте и имеет вид \$INC: \$name.

