

Конвертирование оригинального тезауруса MeSH в формат RUSMARC

Converting the MeSH authority file into RUSMARC format

А.С. Карауш

Муниципальная информационная библиотечная система г.Томска

Н.А. Мешечак

Научно-медицинская библиотека Сибирского государственного медицинского университета (г.Томск)

О.С. Колобов

Научно-техническая библиотека Томского политехнического университета

Alexander S. Karaush

Municipal Information Library System, Tomsk, Russia

Natalya A. Meshechak

Scientific-Medical Library of the Siberian State Medical University

Oleg S. Kolobov

Research Library of Tomsk Polytechnic University, Tomsk, Russia

В докладе рассмотрены проблемы стандартизации и использования авторитетных средств предметизации. Приведены результаты сравнения предметных рубрикаторов, созданных в ЦНМБ и Национальной медицинской библиотеке США. Определены задачи удаленной предметизации в рамках проекта «Открытая электронная медицинская библиотека». Представлены основные этапы конвертирования оригинального медицинского тезауруса MESH версии 2003 года Национальной медицинской библиотеки США в формат RUSMARC.

Применение тезаурусов, как и любых других авторитетных средств каталогизации, предполагает использование единых правил и форматов их подготовки и представления. Автоматизированная библиотечная информационная система, как правило, не подозревает о возможном формате хранения и схеме данных того или иного ресурса, находящегося в сети или в локальной базе данных пользователя. Именно поэтому решение вопросов стандартизации форматов для представления авторитетных файлов или записей позволит уменьшить затраты персонала на работу с авторитетными данными.

Одним из направлений проекта «Открытая электронная медицинская библиотека» являлась задача получения авторитетного файла медицинского тезауруса MeSH в формате RUSMARC для дальнейшей работы по расширению возможностей поиска и удаленной предметизации.

Тезаурус MeSH входит в комплект поставки Системы «ИРБИС» для медицинских библиотек. Однако, использование версии рубрикатора MeSH, разработанной в Государственной центральной научной медицинской библиотеке России, невозможно без дополнительных соглашений с ЦНМБ. При этом оригинальный вариант тезауруса MeSH, созданный в Национальной медицинской библиотеке США, не является коммерческим проектом и распространяется без ограничений.

При сравнении двух вариантов тезауруса MeSH - американского (оригинального) и российского, являющегося его копией (по заявлению авторов), в российском варианте были выявлены потери данных. А также, в нем была изменена структура «рубрика-подрубрика».

Поэтому, было решено осуществить автоматическое конвертирование рубрик оригинальной версии MeSH в соответствующие поля формата RUSMARC и организовать ссылки на связанный заголовок на русском языке. Таким образом, полученный вариант тезауруса MeSH является дополненным вариантом оригинального MeSH, а не конвертированным российским вариантом тезауруса.

На рисунке 1 представлена запись оригинального тезауруса MeSH в формате MARC-21, а на рисунке 2 представлена запись в формате RUSMARC с добавленными полями на русском языке.

```

<Leader>01049nz 2200325n 4500S</Leader>
<1>D030881</1>
<3>DNLM</3>
<5>20011027000000.0</5>
<8>010725 n ancnnbabn || ana b</8>
<40> $aDNLM$cDNLM</40>
<72> $aF1.$x145.$x209.$x259</72>
<150> $aDisclosure</150>
<450> $wnnna$aWhistle Blowing</450>
<450> $aDuty to Recontact</450>
<450> $aFinancial Disclosure</450>
<450> $aInformation Disclosure</450>
<450> $aWhistleblowing</450>
<450> $wnnna$aBlowing, Whistle</450>
<450> $wnnna$aBlowings, Whistle</450>
<450> $wnnna$aDisclosure, Financial</450>
<450> $wnnna$aDisclosure, Information</450>
<450> $wnnna$aDisclosures, Financial</450>
<450> $wnnna$aFinancial Disclosures</450>
<450> $wnnna$aRecontact, Duty to</450>
<450> $wnnna$aWhistle Blowings</450>
<550> $aConfidentiality</550>
<550> $aInformed Consent</550>
<667> $aTRUTH DISCLOSURE is also available</667>
<680> $iRevealing of information, by oral or written
communication.</680>
<688> $a2002; use TRUTH DISCLOSURE 1978-2001; for
WHISTLEBLOWING use TRUTH DISCLOSURE 1997-2001</688>

```

Рисунок 1. Запись рубрики оригинального MeSH в формате MARC-21

```

LDR05 n
LDR06 x
LDR07
LDR08
001 D014342
005 20021002000000.0
100 ## $a 20021002aengy0304 ba
152 ## $a AACR2 $b blc
250 ## $a Truth Disclosure
$a Truthful revealing of information, specifically when the
information disclosed is likely to be psychologically painful ("bad
300 ## $a DISCLOSURE is also available
330 ## $a DISCLOSURE is also available
450 ## $8 eng $a Disclosure, Truth
450 ## $8 eng $a Disclosures, Truth
450 ## $8 eng $a Truth Disclosures
550 ## $8 eng $a Physician-Patient Relations
550 ## $8 eng $a Professional-Patient Relations
686 ## $a F1. $x 145. $x 209. $x 259. $x 800 $z mesh
750 ## $8 rus $a ИСТИНЫ РАСКРЫТИЕ
801 # 0 $a USA $b NLM $c 20021002
801 # 0 $a RUS $b Library SSMU Tomsk $c 20031003
820 ## $a 78
830 ## $a DISCLOSURE is also available
901 ## $a F1.145.209.259.800

```

Рисунок 2. Запись рубрики MeSH в формате RUSMARC.

Основные этапы конвертирования и применения оригинального тезауруса MeSH в формат RUSMARC с учетом перевода записей на русский язык:

1. Определение формата и структуры оригинального тезауруса в MARC-21.
2. Составление списка полей RUSMARC Authority для представления предметного тезауруса, с учетом работы многоуровневых ссылок на записи и дескрипторы;
3. Изучение формата и структуры Российской версии тезауруса MeSH в Системе «ИРБИС»;
4. Подготовка алгоритма сравнения Российской версии MeSH и его оригинального варианта;
5. Перевод оригинального тезауруса MeSH и добавление рубрик на русском языке с сохранением всей информации, имеющейся в первоисточнике;
6. Конвертирование оригинального тезауруса MeSH из формата MARC-21 в формат RUSMARC Authority с учетом многоуровневой структуры тезауруса и с сохранением структуры записей дескрипторов;
7. Проверка структуры полученного тезауруса в соответствии с формальными требованиями коммуникативного формата и протокола Z39.50 с учетом программного обеспечения, установленного в библиотеках.
8. Адаптация варианта новой версии тезауруса в качестве средства предметизации в Системе «ИРБИС»;
9. Настройка конвертера Системы ИРБИС для получения корректных записей в формате RUSMARC с учетом многоуровневых ссылок.

В результате, была получена версия тезауруса MeSH, позволяющая проводить поиск и удаленную предметизацию по всем правилам, используемым в версии, созданной Национальной медицинской библиотекой США. Тезаурус содержит весь объем записей, доступных в оригинальной версии MeSH от 2003 года.