

**Использование тезауруса MeSH
в лингвистическом обеспечении базы данных «MedArt»
Using MeSH Thesaurus in the Linguistic Support
of the «MedArt» Database**

Л. А. Шамардина

*Научно-медицинская библиотека Сибирского государственного
медицинского университета, Томск, Россия*

L. A. Shamardina

*Scientific-Medical Library of the Siberian State Medical University,
Tomsk, Russia*

В докладе приведена информация об организационных аспектах и трудностях использования тезауруса MeSH в корпоративных проектах.

Основной целью создания любой библиографической базы данных (БД) несомненно является повышение качества информационного обеспечения пользователей. Для обеспечения точности и полноты поиска в БД при минимальной затрате сил и времени необходимо адекватное лингвистическое обеспечение ресурса. БД «MedArt» представляет собой результат распределенной росписи российских медицинских журналов усилиями 15 медицинских библиотек (проект «Сводный каталог периодики и аналитики по медицине») [<http://ucm.sibtechcenter.ru/>]. Для корпоративных библиотечных проектов характерна проблема разного подхода участников к индексированию документов и просто использование в библиотеках разных лингвистических средств. В проекте «Сводный каталог периодики и аналитики по медицине» для решения данной проблемы участникам изначально ставились следующие условия:

- индексирование с использованием тезауруса MeSH;
- применение единой методики индексирования.

Все библиотеки – участницы проекта ведут каталогизацию в АБИС Ирбис и индексирование документов осуществляется посредством встроенного в АБИС Ирбис автоматизированного тезауруса MeSH. В качестве методического руководства по индексированию используется адаптированный перевод методики индексирования Национальной Медицинской Библиотеки США (MEDLARS-MEDLINE INDEXING MANUAL NLM USA) [<http://www.nlm.nih.gov/mesh/>].

Структура тезауруса MeSH

Тезаурус MeSH представлен в двух видах:

- алфавитный список терминов с перекрестными ссылками;
- иерархическая структура («дерево»), имеющая 11 уровней. Первый уровень иерархии включает 14 основных категорий (Анатомические термины, Болезни, Химические вещества и лекарства и т. д.).

Тезаурус MeSH содержит несколько различных типов терминов:

- дескрипторы (характеризуют предмет или содержание);
- квалификаторы (используются только с дескрипторами и предоставляют собой средства группировки документов, которые связаны со специфическим аспектом предмета);
- типы публикаций;
- географические дескрипторы (включают континенты, государства, области и другие географические объекты);
- другие термины (синонимы или семантически связанные термины, которые являются перекрестными ссылками к дескрипторам).

Основные принципы индексирования

Согласно правилам индексирования по MeSH для каждого документа определяются главные дескрипторы, которые в дальнейшем можно использовать для составления различных библиогра-

фических указателей; неглавные дескрипторы, обеспечивающие дополнительные возможности для поиска; а также обязательные характеристики, необходимые для отражения объекта, методов исследования.

Сложные понятия могут быть описаны тремя различными способами: адекватный дескриптор, сочетание дескриптор/квалификатор, координация двух и более отдельных дескрипторов. Каждая категория MeSH имеет свои особенности индексирования, которые подробно описаны в руководстве MEDLARS.

Таким образом, использование тезауруса MeSH при индексировании требует постоянного обучения и высокой квалификации сотрудников.

Лингвистическая группа проекта

Для гармонизации правил индексирования в рамках проекта была создана лингвистическая группа, в которую входят ведущие специалисты нескольких библиотек – участниц проекта. В сферу деятельности этой группы входит:

- обсуждение и корректировка методического руководства по индексированию;
- принятие методических решений по часто встречающимся «трудным» случаям индексирования.

Организация работы по индексированию

Для оптимальной организации процесса индексирования необходимо учитывать количество специалистов-индексаторов в библиотеке, их квалификацию, уровень технического оснащения и другие аспекты. Поэтому однозначно регламентировать этот процесс невозможно и каждая библиотека-партнер строит работу с учетом своей специфики. НМБ СибГМУ выступает в проекте в качестве и партнера, и координатора, что предполагает два уровня организации работы:

1. Как библиотеки-партнера

Наряду с остальными партнерами проекта НМБ СибГМУ расписывает журналы различной тематики. Работа по индексированию организована следующим образом:

- за каждым сотрудником закреплен список журналов, которые он индексирует;
- журналы разделены по уровню сложности и по специальностям;
- журналы повышенной сложности индексируют только сотрудники с медицинским образованием по своей или смежным специальностям;
- если в журнале попадаются отдельные статьи по профилю другого сотрудника, то они передаются ему на индексирование;
- для часто повторяющихся «трудных» случаев индексирования принимается определенное методическое решение;
- для устранения разночтений названия фармакологических препаратов, не имеющиеся в MeSH, вносятся в специальный словарь с близким термином MeSH.

2. Как координатора

2.1. Приведение библиографических записей к унифицированному виду:

- при импорте библиографических записей содержимое полей 606 конвертируется в текущую версию MeSH библиотеки-координатора;
- если в импортированной записи имеются дескрипторы MeSH, они проверяются на соответствие правилам индексирования и при необходимости редактируются;
- после редактирования библиотеке-партнеру составляется подробное письмо с указанием ошибок и конкретными примерами;
- если однотипные ошибки повторяются библиотекой-партнером систематически, записи с некорректным индексированием возвращаются на доработку;
- если в импортированной записи не содержатся дескрипторы MeSH, она индексируется;
- экспортируются библиотекам-партнерам отредактированные записи с дескрипторами MeSH текущей версии библиотеки-координатора.

- 2.2. разработка методической документации;
- 2.3. организация обучения методике индексирования.

Проблемы, связанные с использованием тезауруса MeSH

Трудности в использовании тезауруса в корпоративных проектах многократно возрастают в связи с необходимостью согласования работы по индексированию среди большого количества участников. Можно выделить следующие проблемы, с которыми столкнется любая библиотека независимо от того, работает она индивидуально или в проекте.

Организационные

1. разный уровень специалистов в библиотеках-партнерах;
2. разные версии MeSH (в том числе печатные) у библиотек-партнеров;
3. нерегулярное обновление тезауруса.

Технические

1. проблема реиндексирования библиографических записей после обновления версии тезауруса;
2. отсутствие функции поиска с использованием MeSH в АБИС Ирбис;
3. отсутствие в АБИС Ирбис отдельного поля для обязательных характеристик, поэтому приходится вносить их в поле, предназначенное для неглавных дескрипторов.

Связанные с качеством тезауруса:

1. неудобство построения иерархической структуры. Например, названия фармацевтических препаратов стоят в дереве по своей химической структуре, а не фармакологической группе;
2. отсутствие терминов, отражающих национальные особенности медицинской предметной области (особенно это касается категорий здравоохранение, географические дескрипторы);
3. некорректное представление тезауруса в АБИС Ирбис;
4. грамматические ошибки.

Выводы

По текущим результатам нашей работы с тезаурусом MeSH можно говорить о том, что использование переводных тезаурусов не может обеспечить высокого качества индексирования документов даже при строгом следовании оригинальной методике индексирования.