

МЕЖГОСУДАРСТВЕННЫЙ
СТАНДАРТ

ГОСТ
7.24—
2007

Система стандартов по информации,
библиотечному и издательскому делу

ТЕЗАУРУС ИНФОРМАЦИОННО-ПОИСКОВЫЙ МНОГОЯЗЫЧНЫЙ

Состав, структура и основные требования
к построению

Издание официальное



Москва
Стандартинформ
2010

Предисловие

Цели, основные принципы и основной порядок проведения работ по межгосударственной стандартизации установлены ГОСТ 1.0—92 «Межгосударственная система стандартизации. Основные положения» и ГОСТ 1.2—97 «Межгосударственная система стандартизации. Стандарты межгосударственные, правила и рекомендации по межгосударственной стандартизации. Порядок разработки, принятия, применения, обновления и отмены»

Сведения о стандарте

- 1 РАЗРАБОТАН Всероссийским институтом научной и технической информации Российской академии наук
- 2 ВНЕСЕН Федеральным агентством по техническому регулированию и метрологии
- 3 ПРИНЯТ Межгосударственным советом по стандартизации, метрологии и сертификации (протокол № 31 от 8 июня 2007 г.)

За принятие проголосовали:

| Краткое наименование страны по МК (ИСО 3166) 004—97 | Код страны по МК (ИСО 3166) 004—97 | Сокращенное наименование национального органа по стандартизации |
|---|------------------------------------|--|
| Азербайджан | AZ | Азстандарт |
| Армения | AM | Армстандарт |
| Беларусь | BY | Госстандарт Республики Беларусь |
| Грузия | GE | Грузстандарт |
| Кыргызстан | KG | Кыргызстандарт |
| Молдова | MD | Молдова-Стандарт |
| Российская Федерация | RU | Федеральное агентство по техническому регулированию и метрологии |
| Таджикистан | TJ | Таджикстандарт |
| Туркменистан | TM | Главгосслужба «Туркменстандартлары» |
| Узбекистан | UZ | Агентство «Узстандарт» |
| Казахстан | KZ | Госстандарт Республики Казахстан |

4 Настоящий стандарт разработан с учетом основных нормативных положений международного стандарта ИСО 5964:1985 «Документация. Руководство по построению и разработке многоязычных тезаурусов» (ISO 5964:1985 «Guidelines for the establishment and development of multilingual thesauri», IDT)

5 Приказом Федерального агентства по техническому регулированию и метрологии от 31 октября 2007 г. № 287-ст межгосударственный стандарт ГОСТ 7.24—2001 введен в действие в качестве национального стандарта Российской Федерации с 1 июля 2008 г.

6 ВЗАМЕН ГОСТ 7.24—90

7 ПЕРЕИЗДАНИЕ. Февраль 2010 г.

Информация о введении в действие (прекращении действия) настоящего стандарта публикуется в указателе «Национальные стандарты».

Информация об изменениях к настоящему стандарту публикуется в указателе «Национальные стандарты», а текст изменений — в информационных указателях «Национальные стандарты». В случае пересмотра или отмены настоящего стандарта соответствующая информация будет опубликована в информационном указателе «Национальные стандарты»

© Стандартинформ, 2007
© СТАНДАРТИНФОРМ, 2010

В Российской Федерации настоящий стандарт не может быть полностью или частично воспроизведен, тиражирован и распространен в качестве официального издания без разрешения Федерального агентства по техническому регулированию и метрологии

Содержание

| | |
|---|---|
| 1 Область применения | 1 |
| 2 Нормативные ссылки | 1 |
| 3 Термины и определения | 1 |
| 4 Состав и структура тезауруса | 2 |
| 5 Основные требования к построению многоязычных информационно-поисковых тезаурусов | 2 |
| Приложение А (справочное) Примеры использования средств указания эквивалентности дескрипторов в одноязычных версиях | 6 |

МЕЖГОСУДАРСТВЕННЫЙ СТАНДАРТ

Система стандартов по информации, библиотечному и издательскому делу

ТЕЗАУРУС ИНФОРМАЦИОННО-ПОИСКОВЫЙ МНОГОЯЗЫЧНЫЙ

Состав, структура и основные требования к построению

System of standards on information, librarianship and publishing. Multilingual thesaurus for information retrieval.
Composition, structure and basic requirements for development

Дата введения — 2008—07—01

1 Область применения

Настоящий стандарт распространяется на многоязычные информационно-поисковые тезаурусы (далее — МИПТ) и устанавливает состав, структуру и основные требования к построению МИПТ, применяемых в информационно-поисковых системах.

2 Нормативные ссылки

В настоящем стандарте использованы ссылки на следующие межгосударственные стандарты:

ГОСТ 7.25—2001 Система стандартов по информации, библиотечному и издательскому делу. Тезаурус информационно-поисковый одноязычный. Правила разработки, структура, состав и форма представления

ГОСТ 7.74—96 Система стандартов по информации, библиотечному и издательскому делу. Информационно-поисковые языки. Термины и определения

Примечание — При пользовании настоящим стандартом целесообразно проверить действие ссылочных стандартов по указателю «Национальные стандарты», составленному по состоянию на 1 января текущего года, и по соответствующим информационным указателям, опубликованным в текущем году. Если ссылочный стандарт заменен (изменен), то при пользовании настоящим стандартом следует руководствоваться заменяющим (измененным) стандартом. Если ссылочный стандарт отменен без замены, то положение, в котором дана ссылка на него, применяется в части, не затрагивающей эту ссылку.

3 Термины и определения

В настоящем стандарте применены термины по ГОСТ 7.74, а также следующие термины с соответствующими определениями:

3.1 многоязычный информационно-поисковый тезаурус: Информационно-поисковый тезаурус, содержащий лексические единицы, взятые из нескольких естественных языков и представляющий эквивалентные по смыслу понятия на каждом из этих языков.

Примечание — Предназначен для обработки документов (запросов) и информационного поиска с целью обмена информацией на различных естественных языках.

3.2 одноязычный информационно-поисковый тезаурус: Информационно-поисковый тезаурус, содержащий лексические единицы, взятые из одного естественного языка.

3.3 язык-компонент МИПТ: Язык, на основе которого разработана какая-либо из одноязычных версий.

3.4 дескриптор МИПТ: Основная лексическая единица МИПТ, представляющая собой совокупность эквивалентных дескрипторов одноязычных версий, связанных средствами для указания эквивалентности.

4 Состав и структура тезауруса

4.1 Обязательными составными частями МИПТ являются:

- вводная часть;
- алфавитные лексико-семантические указатели, включающие средства для указания эквивалентности дескрипторов одноязычных версий (см. 4.4). Допускается в состав тезауруса вводить систематические, пермутационные, иерархические и другие указатели и списки специальных категорий лексических единиц, а также приложения, содержащие дополнительные сведения о разработке и использовании МИПТ.

4.2 Вводная часть включает титульный лист и введение, составленные согласно ГОСТ 7.25 и изложенные на одном или нескольких языках-компонентах.

На титульном листе следует указывать все языки-компоненты. Рекомендуется наименование тезауруса на титульном листе формулировать на одном языке, а дополнительно к основному титульному листу включать во вводную часть титульные листы на всех языках-компонентах. Введение должно содержать краткое описание методических принципов установления эквивалентности дескрипторов одноязычных версий, а также указание на то, что тезаурус разработан в соответствии с настоящим стандартом. Рекомендуется в состав вводной части включать полный или сокращенный перевод введения на всех языках-компонентах.

4.3 Алфавитный лексико-семантический указатель в каждой из одноязычных версий содержит перечень тезаурусных статей, расположенных в алфавитном порядке заглавных лексических единиц на соответствующем языке-компоненте. Состав и структура тезаурусных статей в лексико-семантических указателях одноязычных версий — по ГОСТ 7.25.

4.4 Средства для указания эквивалентности дескрипторов одноязычных версий реализуются введением в тезаурусную статью кода эквивалентности дескрипторов (идентификационного кода) и (или) эквивалентных дескрипторов на языках-компонентах.

При использовании идентификационных кодов в дополнение к алфавитному лексико-семантическому указателю МИПТ составляют указатель идентификационных кодов, в котором каждому коду приписывают соответствующие дескрипторы одноязычных версий.

4.5 Систематические, иерархические и пермутационные указатели строятся по ГОСТ 7.25 на языках-компонентах и содержат средства для указания эквивалентности дескрипторов одноязычных версий. Примеры использования средств для указания эквивалентности дескрипторов приведены в приложении А.

4.6 Состав и структура одноязычных версий должны соответствовать следующим требованиям:

- в каждой одноязычной версии должны содержаться все дескрипторы, необходимые для представления онтологии предметной области с целью поиска и обмена информацией между системами, использующими различные языки;
- классы условной эквивалентности лексических единиц одноязычных версий должны соответствовать друг другу по объему выражаемых понятий;
- соответствующие друг другу дескрипторы отдельных одноязычных версий должны иметь по возможности максимально совпадающие парадигматические отношения.

Одноязычные версии допускается дополнять дескрипторами, не являющимися необходимыми для межязыкового обмена и отсутствующими в других одноязычных версиях; при этом указывают иерархическую подчиненность их какому-либо дескриптору МИПТ.

Одноязычные версии могут включать парадигматические отношения, отсутствующие в других одноязычных версиях, если они на данном языке-компоненте уточняют понятия, выраженные дескриптором МИПТ, и не противоречат его пониманию в других языках-компонентах.

5 Основные требования к построению многоязычных информационно-поисковых тезаурусов

5.1 Содержание работ по формированию одноязычных версий определяют наличием или отсутствием одноязычных тезаурусов по данной тематике на языках-компонентах.

5.1.1 При отсутствии одноязычных тезаурусов по данной тематике на языках-компонентах, как правило, сначала создают одну одноязычную версию и на ее основе строят остальные.

Возможна также совместная разработка одновременно всех или нескольких одноязычных версий при взаимном согласовании состава дескрипторов и их парадигматических связей.

5.1.2 При наличии тезауруса по данной тематике на одном из языков-компонентов, как правило, его берут за основу разрабатываемого МИПТ. Другие одноязычные версии формируют подбором эквивалентов дескрипторов этого тезауруса на языках-компонентах. При этом возможны различные степени семантического соответствия эквивалентов дескрипторов на различных языках.

5.1.3 При наличии нескольких исходных тезаурусов по данной тематике на языках-компонентах разработка МИПТ ведется подбором эквивалентных дескрипторов в данных тезаурусах (см. 5.2.1 и 5.2.2). Дескрипторы, не имеющие эквивалентов в отдельных одноязычных версиях, переводят на соответствующие языки и включают в одноязычные версии, в которых этот дескриптор первоначально отсутствовал. При этом следует вносить в исходные тезаурусы изменения, необходимые для согласования одноязычных версий.

5.1.4 Рекомендуется одну из одноязычных версий выбирать в качестве базовой и устанавливать эквивалентность дескрипторов других одноязычных версий по отношению к дескрипторам базовой версии.

5.2 При установлении эквивалентности дескрипторов различных одноязычных версий необходимо различать на разных языках-компонентах следующие степени эквивалентности терминов:

- полная;
- неполная;
- частичная;
- отсутствие эквивалентного термина.

5.2.1 При наличии в языках-компонентах полностью эквивалентных терминов их считают представителями одного дескриптора МИПТ.

5.2.2 При отсутствии в языках-компонентах полных эквивалентов для выражения одного и того же понятия в качестве дескриптора МИПТ в одноязычных версиях используют неполные и частичные эквиваленты.

Неполными эквивалентами являются термины, для которых объемы выражаемых ими понятий пересекаются.

Частичными эквивалентами являются термины, для которых объем понятия, выражаемого одним эквивалентом, входит в объем понятия, выражаемого другим эквивалентом.

5.2.2.1 Если различие в объеме понятий между неполными или частичными эквивалентами существенно для данного МИПТ, то их можно использовать для представления дескриптора МИПТ в разных одноязычных версиях, например:

ru: АННОТАЦИЯ
fr: RÉSUMÉ

5.2.2.2 Если различие в объеме понятий между неполными или частичными эквивалентами существенно, то при представлении дескриптора МИПТ в одноязычных версиях значение используемого термина уточняют одним из следующих способов:

- добавлением релятора (см. 5.2.2.3);
- добавлением примечания (см. 5.2.2.4);
- включением термина в уточняющее словосочетание с определяющими словами (см. 5.2.2.5);
- введением условно эквивалентных дескрипторов в тезаурусную статью дескриптора одноязычной версии (см. 5.2.2.6);
- использованием комбинации двух или более терминов (см. 5.2.2.7).

5.2.2.3 Добавление релятора, который является частью дескриптора, применяют в тех случаях, когда для более полного соответствия содержанию дескриптора необходимо ограничивать объем понятия, выражаемого термином, например:

| | |
|------------------|-------------------|
| англ. | русс. |
| TREE | ДЕРЕВО (РАСТЕНИЕ) |
| WOOD (MATERIAL) | ДЕРЕВО (МАТЕРИАЛ) |
| WOOD (LANDSCAPE) | ЛЕС (ЛАНДШАФТ) |

5.2.2.4 Добавление примечания, которое не является частью дескриптора, используют в случаях уточнения объема понятия, выраженного соответствующими дескрипторами на разных языках-компонентах, например:

| | |
|-----------------------|--|
| франц. AFFILIATION | русс. МЕСТО РАБОТЫ (как элемент библиографического описания, обозначающий место работы автора публикации) |
|-----------------------|--|

5.2.2.5 Включение термина в уточняющее словосочетание применяют в тех случаях, когда требуется ограничивать объем понятия и данное словосочетание устойчиво употребляется в документах, например:

| | |
|--------------------------|-------------------------|
| англ. CRANE (MACHINE) | русс. ПОДЪЕМНЫЙ КРАН |
|--------------------------|-------------------------|

5.2.2.6 Введение условно эквивалентных дескрипторов в тезаурусную статью дескриптора одноязычной версии применяют в тех случаях, когда они необходимы для однозначного представления понятия в данной версии, например:

| | |
|------------------------------------|-----------------|
| англ. GRAPH UF diagram chart | русс. ГРАФИК |
|------------------------------------|-----------------|

5.2.2.7 Если исходный дескриптор обозначает понятие, не выраженное в виде единого термина на другом языке-компоненте, то допускается использовать в одноязычной версии в качестве эквивалента комбинацию двух и более дескрипторов одноязычной версии, например:

| | |
|--|---|
| франц. CARTE PERFORÉE+SELECTION MANUELLE REFERER ANALYTIQUE+ÉLABORATION | русс. ПЕРФОКАРТА РУЧНОГО ОБРАЩЕНИЯ РЕФЕРИРОВАНИЕ |
|--|---|

5.2.3 При отсутствии лексического эквивалента в одном из языков-компонентов допускаются решения, приведенные в 5.2.3.1—5.2.3.5.

5.2.3.1 Создание нового термина переводом или калькированием иноязычного дескриптора, например:

| | | |
|-------------------------|----------------------------|--------------------------|
| англ. FIVE-YEAR PLAN | франц. PLAN QUINQUENNAL | русс. ПЯТИЛЕТНИЙ ПЛАН |
|-------------------------|----------------------------|--------------------------|

5.2.3.2 Использование в качестве эквивалента словосочетания, например:

| | |
|----------------|---|
| англ. CORER | русс. МАШИНА ДЛЯ УДАЛЕНИЯ СЕРДЦЕВИНЫ ПЛОДА |
|----------------|---|

5.2.3.3 Заимствование иноязычного термина, например:

| | |
|---------------|---------------|
| англ. FILE | русс. ФАЙЛ |
|---------------|---------------|

5.2.3.4 Использование в качестве эквивалента термина с близким значением, например:

| | |
|--------------------------------|--------------------------------------|
| франц. OUTILS LINGUISTIQUES | русс. ЛИНГВИСТИЧЕСКОЕ ОБЕСПЕЧЕНИЕ |
|--------------------------------|--------------------------------------|

| | |
|-------------------|--------------------|
| англ. TEENAGER | русс. ПОДРОСТОК |
|-------------------|--------------------|

5.2.3.5 Использование в качестве эквивалента термина, более широкого по значению и не входящего в лексический состав данной тематической области, например в МИПТ по библиотечному делу:

| | |
|---------------------|--|
| франц. ANIMATEUR | русс. ОРГАНИЗАТОР КУЛЬТАССОВОЙ РАБОТЫ |
|---------------------|--|

5.3 По окончании разработки МИПТ представляют для регистрации в национальный и/или международный центр ведения информационно-поисковых языков. МИПТ представляют в составе всех его одноязычных версий.

Каждая одноязычная версия МИПТ может быть представлена как одноязычный тезаурус с иноязычными эквивалентами.

Национальные и международные центры ведения информационно-поисковых языков предоставляют заинтересованным лицам сведения о зарегистрированных МИПТ на условиях, определяемых разработчиком.

Форма представления одноязычных версий — по ГОСТ 7.25 и 4.4.

5.4 Процесс ведения МИПТ сводят к ведению его одноязычных версий и к обеспечению их согласованности. Изменения вносят одновременно во все одноязычные версии после их согласования на уровне языков-компонентов.

Рекомендуется не реже чем через пять лет обновлять сведения о МИПТ, зарегистрированных по 5.3.

Приложение А
(справочное)

Примеры использования средств указания эквивалентности дескрипторов в одноязычных версиях

Примеры**1 Использование цифровых идентификационных кодов, не зависящих от языка:**

STAŁA - дескриптор на польском языке

086210 - идентификационный код

a CHARAKTERYSTYKA
INDEKS
MODUL
STAŁY
WSKAZNIK (LICZBA)

2 Использование эквивалентов дескрипторов на другом языке:

КОНСТАНТА - дескриптор на русском языке

STAŁA - дескриптор на польском языке

c Постоянная
a КОЭФФИЦИЕНТ
МОДУЛЬ
ПОКАЗАТЕЛЬ
ХАРАКТЕРИСТИКА

3 Использование эквивалентов дескрипторов на других языках и идентификационных кодов:

БЕНЗИНЫ - дескриптор на русском языке

fre: ESSENCE MOTEUR - дескриптор на французском языке

eng: GASOLINE - дескриптор на английском языке

pol: BENZYNA - дескриптор на польском языке

011420 - идентификационный код

в НЕФТЕПРОДУКТЫ СВЕТЛЫЕ
ТОПЛИВО ЖИДКОЕ
ТОПЛИВО КАРБЮРАТОРНОЕ
н БЕНЗИНЫ АВИАЦИОННЫЕ
БЕНЗИНЫ АВТОМОБИЛЬНЫЕ
БЕНЗИНЫ АЛКИЛИРОВАННЫЕ

4 Использование совместного расположения статей на двух языках:

| | |
|--------------------------|----------------------|
| русский язык | немецкий язык |
| АМАЛЬГАМА СЕРЕБРА | SILBERAMALGAM |

| | |
|------------------|-------------------|
| в МЕТАЛЛЫ | OB METALLE |
|------------------|-------------------|

| | |
|------------------------|-------------------------|
| н КУПРОАРКВЕРИТ | UB CUPROARQUERIT |
|------------------------|-------------------------|

УДК 02:006.354

МКС 01.140.20

Т62

Ключевые слова: многоязычный информационно-поисковый тезаурус, дескриптор, лексическая единица, эквивалентность дескрипторов, идентификационный код дескриптора

Редактор *Н.В. Таланова*
Технический редактор *В.Н. Прусакова*
Корректор *В.И. Варенцова*
Компьютерная верстка *Л.А. Круговой*

Подписано в печать 10.03.2010. Формат 60×84 $\frac{1}{8}$. Бумага офсетная. Гарнитура Ариал.
Печать офсетная. Усл. печ. л. 1,40. Уч.-изд. л. 0,80. Тираж 85 экз. Зак. 70.

ФГУП «СТАНДАРТИНФОРМ», 123995 Москва, Гранатный пер., 4.
www.gostinfo.ru info@gostinfo.ru

Набрано во ФГУП «СТАНДАРТИНФОРМ» на ПЭВМ.

Отпечатано в филиале ФГУП «СТАНДАРТИНФОРМ» — тип. «Московский печатник», 105062 Москва, Лялин пер., 6.