ГОСУДАРСТВЕННЫЙ СТАНДАРТ
СОЮЗА ССР

СИСТЕМА СТАНДАРТОВ ПО ИНФОРМАЦИИ,
БИБЛИОТЕЧНОМУ И ИЗДАТЕЛЬСКОМУ ДЕЛУ

ТЕЗАУРУС ИНФОРМАЦИОННО-ПОИСКОВЫЙ МНОГОЯЗЫЧНЫЙ

ПРАВИЛА РАЗРАБОТКИ

ГОСТ 7.24—80

Издание официальное

Цена 3 коп.

ГОСУДАРСТВЕННЫЙ КОМИТЕТ СССР ПО СТАНДАРТАМ
Москва

gost.e-ecolog.ru

УДК 025.48 : 006.354 Группа Т62

ГОСУДАРСТВЕННЫЙ СТАНДАРТ СОЮЗА ССР

Система стандартов по информации, библиотечному и издательскому делу

ТЕЗАУРУС ИНФОРМАЦИОННО-ПОИСКОВЫЙ МНОГОЯЗЫЧНЫЙ

Правила разработки

System of standards on information, librarianship and publishing. Multilingual information retrieval thesaurus. Guidelines for development

ГОСТ 7.24—80

Постановлением Государственного комитета СССР по стандартам от 30 декабря 1980 г. № 6175 срок введения установлен

с 01.01.82

Настоящий стандарт устанавливает правила разработки многоязычных информационно-поисковых тезаурусов (МИПТ), создаваемых в рамках международных и национальных систем научной и технической информации и включающих в качестве языка-компонента один из языков народов СССР.

Пояснения к терминам, используемым в стандарте, приведены в справочном приложении.

1. ОСНОВНЫЕ ПОЛОЖЕНИЯ

1.1. Национальные варианты МИПТ должны соответствовать следующим требованиям:

все национальные варианты МИПТ имеют равный статус;

понятие, выраженное дескриптором в одном из языков-компонентов, должно быть отражено в каждом языке-компоненте;

национальный вариант МИПТ должен учитывать требования, предъявляемые к одноязычным информационно-поисковым тезаурусам (ИПТ) соответствующим национальным стандартом;

парадигматические отношения эквивалентных дескрипторов отдельных национальных вариантов МИПТ должны максимально совпадать.

2. ОРГАНИЗАЦИЯ РАЗРАБОТКИ МНОГОЯЗЫЧНЫХ ИПТ

2.1. Разработка МИПТ осуществляется на межъязыковом уровне и на уровне отдельных языков-компонентов.

2.2. На межъязыковом уровне разрабатываются единые методические принципы и план построения МИПТ, а также согласуются рабочие материалы МИПТ.

Издание официальное Перепечатка воспрещена

Переиздание. Апрель 1988 г.

© Издательство стандартов, 1988

Единые методические принципы определяют:
назначение и применение МИПТ;
методический подход к разработке МИПТ;
общую структуру МИПТ (наличие или отсутствие вспомогательных указателей, согласованный уровень предкоординации дескрипторов и т. д.);
форму и порядок представления национальных вариантов, дополнений и изменений к ним;
правила ведения МИПТ.
Согласованию на межъязыковом уровне подлежат:
тематический охват МИПТ;
принципы систематизации лексики;
содержание классов условной эквивалентности национальных вариантов и установление эквивалентности между дескрипторами национальных вариантов;
иерархические и ассоциативные отношения для эквивалентных дескрипторов;
структура и состав кодов идентификации.

2.3. На уровне отдельных языков-компонентов разработка национальных вариантов МИПТ производится в соответствии с решением, принятым на межъязыковом уровне, и требованиями национальных стандартов на разработку одноязычных ИПТ.

3. ОСНОВНЫЕ ЭТАПЫ ПОСТРОЕНИЯ МНОГОЯЗЫЧНЫХ ИПТ

3.1. Построение МИПТ состоит из следующих этапов:
определение исходных условий;
определение тематического охвата;
формирование национальных вариантов МИПТ;
установление эквивалентности лексических единиц (ЛЕ) на языках-компонентах;
построение и согласование словарных статей;
присвоение кодов идентификации дескрипторам;
построение лексико-семантических указателей;
построение вспомогательных указателей.

3.2. Организация работы и методика создания МИПТ определяются наличием или отсутствием одноязычных ИПТ по данной тематике на языках-компонентах.

При создании МИПТ возможны следующие исходные условия:
отсутствие одноязычных тезаурусов;
наличие одного тезауруса;
наличие нескольких тезаурусов.

3.2.1. Отсутствие одноязычных тезаурусов по данной тематике на языках-компонентах требует создания одноязычного ИПТ на одном из языков-компонентов с последующим построением соответствующих национальных вариантов МИПТ или одновременного создания двух или нескольких национальных вариантов МИПТ.

При этом необходимо руководствоваться требованиями ГОСТ 7.25—80, а также разд. 1 и 2 настоящего стандарта.

3.2.2. При наличии одного одноязычного тезауруса по данной тематике лексика этого тезауруса берется за основу МИПТ и переводится на другие языки-компоненты. Для обеспечения точности перевода необходимо переводить дескрипторы с учетом всей словарной статьи. При переводе возможны различные степени эквивалентности терминов двух (или более) языков (см. п. 3.3).

Данный метод применяется также в случае добавления дополнительного языка к уже существующему МИПТ.

В процессе создания МИПТ должен дополняться лексикой, используемой в документах и информационных запросах по данной тематике.

3.2.3. При наличии нескольких тезаурусов на языках-компонентах используется метод совмещения и слияния.

В ходе совмещения и слияния нескольких существующих одноязычных ИПТ они могут подвергаться изменениям.

3.3. Основным этапом построения МИПТ является установление эквивалентности между дескрипторами национальных вариантов.

3.3.1. При установлении эквивалентности могут встретиться следующие степени соответствия:

полная эквивалентность, например:

русск.	франц.
Амперметр	Ampéremétre
Лингвистический анализ	Analyse linguistique
Патент	Document brevet

неполная эквивалентность

русск.	франц.
Лифт Подъемник	Ascenseur

англ.	франц.
Авторское право — Copyright Author's right	Copyright Droit d'auteur
Дерево — Tree Wood	Arbre Bois

отсутствие эквивалента, например:

русск.	франц.
Лингвистическое обеспечение	—
Научная организация труда	—
—	Bande dessinée

Для каждого дескриптора, входящего в национальный вариант МИПТ, необходимо установить эквивалент на каждом языке-компоненте, стремясь к полной эквивалентности. В сложных случаях следует формулировать рабочие определения дескрипторов.

3.3.2. При неполной эквивалентности возможны следующие пути уточнения значения лексической единицы и установления эквивалентности:

а) добавление в скобках пояснительных помет или дефиниций для расширения, сужения или определения значения дескриптора, например:

англ.	франц.	русск.
Tree	Arbre	Дерево (растение)
Wood	Bois	Дерево (материал)

англ.	русск.
Translation	Перевод (документ)
Translating	Перевод (процесс)

Пояснительные пометы к дескрипторам могут различаться в разных языках. В частности, дескриптор в одном языке может пониматься неоднозначно и требовать пояснительной пометы, а его эквивалент в другом языке может не иметь пояснительной пометы;

б) использование условных синонимов словарной статьи для уточнения значения дескриптора, например:

русск.	франц.
Решетка	Grille
Изгородь	clôture
Ограда	

в) использование в качестве дескрипторов словосочетаний вместо неоднозначно понимаемых компонентов, например:

русск.	франц.	англ.
Жесткость воды	Dureté de l'eau	Hardness of water
Жесткость излучения	Dureté de radiation	Hardness of radiation
Жесткость конструкций	Rigidité de la structure	Stiffness of structure
Кран водопроводный	Robinet	Cock
Кран подъемный	Grue	Crane

Если понятия, различимые в одном языке, оказываются неразличимыми в другом, установление полной эквивалентности не проводится, например:

русск.	англ.	франц.
Авторское право	Copyright Author's right	Copyright Droit d'auteur
Ленточный конвейер	Band conveyer Belt conveyer	

3.3.3. В случае отсутствия лексического эквивалента в одном из языков-компонентов возможны следующие решения:

а) создание новой ЛЕ путем прямого заимствования или калькирования иностранного слова или словосочетания; при этом сле-

дует руководствоваться принятыми в данном языке правилами образования новых слов, например:

рус. Файл от англ. File;
рус. Железная дорога от фр. Chemin de fer;

б) представление дескриптора логической суммой нескольких дескрипторов, например:

русск.	франц.	англ.
Реферирование	Élaboration+résumé analytique	
Информационно-логическая система	Systéme d'information+intelligence artificielle	Information System+Artificial Intelligence

в) использование в качестве эквивалента дескриптора-описания, например:

франц.	русск.
Systéme de plancher	Способ введения горных работ с использованием проходческого подвесного потолка

англ.	русск.
Corer	Машина для удаления сердцевины плода

г) использование в качестве эквивалента дескриптора с более широким или близким значением, например:

франц.	русск.
Affiliation (часть библиографического описания, обозначающая место работы автора документа)	Библиографическое описание

д) изъятие дескриптора из МИПТ—в том случае, когда составители МИПТ уверены, что изъятый дескриптор не будет использоваться при индексировании и/или поиске.

4. АВТОМАТИЗАЦИЯ ПОСТРОЕНИЯ МНОГОЯЗЫЧНЫХ ИПТ

4.1. Построение МИПТ с использованием ЭВМ позволяет автоматизировать следующие процессы:

алфавитную сортировку;

проверку взаимности и непротиворечивости отсылок;

составление указателей (лексико-семантических, перестановочных, иерархических отношений и др.);

распечатку в требуемых форматах.

4.2. Автоматизированное построение МИПТ должно допускать его представление в трех основных формах:

в памяти ЭВМ («машинный тезаурус»), в этом случае предполагается наличие программ, позволяющих выводить тезаурус или его фрагменты на экран видеотерминала или на печать;

в виде комплекта словарей и указателей, напечатанных с помощью ЭВМ и отражающих текущее состояние машинного тезауруса;

на машиночитаемом носителе в коммуникативном формате.

Допускается автоматизировать и основные (содержательные) процессы построения МИПТ, т. е. установление и коррекцию связей между ЛЕ. Тезаурус, полученный с применением таких методов, обязательно должен проходить процедуру опробования.

4.3. Наличие машинного тезауруса и соответствующих программ обеспечивает:

автоматическую проверку ЛЕ по списку входов в тезаурус;

автоматическое кодирование ЛЕ запросов;

автоматическое пополнение поисковых предписаний благодаря использованию семантических связей, зафиксированных в ИПТ;

замену ЛЕ на любом из языков-компонентов на ЛЕ на другом языке-компоненте.

5. ВЕДЕНИЕ МНОГОЯЗЫЧНЫХ ИПТ

5.1. Основной задачей ведения МИПТ является поддержание разработанного МИПТ в состоянии, позволяющем с достаточной эффективностью проводить информационный поиск в ходе эксплуатации ИПС.

При ведении МИПТ следует учитывать все его первоначальные варианты.

Ведение МИПТ в подсистемах международных систем научной и технической информации должно осуществляться с учетом принятых в данных подсистемах правил.

Ведение МИПТ включает:

внесение в МИПТ всех изменений, обеспечивающих соответствие лексики и парадигматики МИПТ поисковому массиву ИПС с учетом динамики последнего;

поддержание согласованности всех национальных вариантов МИПТ;

совершенствование структуры МИПТ, его лексического состава и парадигматики.

5.2. Ведение МИПТ включает операции, затрагивающие все национальные варианты МИПТ:

изменение лексического состава:

внесение новых ЛЕ, включая использование определенной комбинации дескрипторов, имеющихся в МИПТ, для выражения нового понятия;

удаление ЛЕ;

изменение статуса *ЛЕ* (замена дескриптора на недескриптор и наоборот);

изменение парадигматических отношений:

установление новых парадигматических отношений;

ликвидация уже зафиксированных парадигматических отношений;

изменение типов зафиксированных парадигматических отношений:

ведение указателей:

исправление ошибок;

периодическое переиздание МИПТ (каждые 5 лет или раньше, при существенном изменении МИПТ).

ПРИЛОЖЕНИЕ
Справочное

ПОЯСНЕНИЯ К ТЕРМИНАМ, ИСПОЛЬЗУЕМЫМ В СТАНДАРТЕ

Термин	Пояснение
Многоязычный информационно-поисковый тезаурус (МИПТ)	Совокупность одноязычных тезаурусов по одной и той же тематике — национальных вариантов МИПТ, каждый из которых дополнен средствами связи с другими национальными вариантами МИПТ. Многоязычный тезаурус отражает семантические отношения между лексическими единицами (ЛЕ) национальных вариантов ИПТ и предназначен для обработки и поиска информации с целью обмена информацией на языках-компонентах
Язык-компонент	Естественный язык, на котором составлен национальный вариант МИПТ
Дескриптор МИПТ	Множество эквивалентных дескрипторов национальных вариантов МИПТ, связанных одним кодом идентификации
Лексика МИПТ	Совокупность всех лексических единиц, входящих в национальные варианты МИПТ

Редактор *Р. С. Федорова*
Технический редактор *Э. В. Митяй*
Корректор *М. М. Герасименко*

Слано в наб. 06.05.88 Подп. в печ. 06.07.88 0,5 усл. п. л. 0,5 усл. кр.-отт. 0,45 уч.-изд. л. Тираж 10 000 Цена 3 коп.

Ордена «Знак Почета» Издательство стандартов, 123840, Москва, ГСП, Новопресненский пер., д. 3.
Типография Издательства стандартов, ул. Даряус и Гирено, 39. Зак. 1617.

gost.e-ecolog.ru