

ОРГАНИЗАЦИЯ СОТРУДНИЧЕСТВА ЖЕЛЕЗНЫХ ДОРОГ (ОСЖД)

II издание

Разработано совещанием экспертов по НТЭИ

Утверждено заседанием членов Комитета ОСЖД от 16 февраля 2004 г.

Дата вступления в силу: 2 мая 2004 г.

Примечание:

Теряет силу I издание от 1 октября 1996 г.

**О
905-3**

**РАЗРАБОТКА, ВЕДЕНИЕ И ИСПОЛЬЗОВАНИЕ
МНОГОЯЗЫЧНЫХ ИНФОРМАЦИОННО-ПОИСКОВЫХ
ТЕЗАУРУСОВ ОСЖД**

(Методические рекомендации)

Содержание

| | |
|---|----|
| 1. Введение | 3 |
| 2. Основные определения | 3 |
| 3. Этапы построения МИПТ ОСЖД | 4 |
| 4. Состав и структура МИПТ ОСЖД | 7 |
| 5. Правила использования МИПТ ОСЖД | 8 |
| 6. Требования к программно-техническим средствам для создания, использования и ведения МИПТ ОСЖД | 10 |
| 7. Экспериментальная проверка и ведение МИПТ ОСЖД..... | 10 |

1. ВВЕДЕНИЕ

1.1. Настоящая памятка разработана с целью унификации принципов разработки, использования и ведения многоязычных информационно-поисковых тезаурусов (МИПТ) ОСЖД. Настоящая памятка разработана взамен памятки О 905-3 "Основные положения по созданию, ведению и использованию международного информационно-поискового тезауруса "Транспорт".

1.2. МИПТ ОСЖД создаются в рамках международного сотрудничества стран-членов ОСЖД в соответствии с решениями совещаний экспертов по научно-технической и экономической информации (НТЭИ). Для разработки МИПТ создаются Временные рабочие группы в составе специалистов железных дорог стран-членов ОСЖД.

1.3. Памятка содержит основные положения по разработке, структуре, составу и ведению МИПТ ОСЖД, а также рекомендации по их использованию в соответствии со стандартом ИСО 2788 и 5984.

1.4. МИПТ предназначены для индексирования документов и запросов при аналитико-синтетической обработке информационных документов, включаемых в Автоматизированный банк данных (АБД) ОСЖД, а также поиска информации с целью обмена данными между автоматизированными информационными системами центров НТЭИ стран-членов ОСЖД. По своей функции они являются пособием для автоматизированного документального поиска в массивах, содержащих результаты обработки транспортной литературы другой страны или ряда стран. С их помощью запросы, сформулированные на одном языке, и существенные содержательные аспекты документов, сформулированные на другом языке, трансформируются на общий информационно-поисковый язык (ИПЯ) и, таким образом, обеспечивается их сравнимость.

1.5. Обязательным компонентом дескрипторных статей многоязычных информационно-поисковых тезаурусов являются средства установления эквивалентности дескрипторов национальных версий. Это решается путем приписывания эквивалентным дескрипторам специального цифрового четырехзначного кода соответствия.

2. ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ

2.1. **Тезаурусом** называется контролируемый, но изменяемый перечень (словарь) терминов (дескрипторов) с отражением смысловых связей между ними. Тезаурус является лексическим инструментом и основным ядром лингвистического обеспечения автоматизированных информационно-поисковых систем.

2.2. **Многоязычный информационно-поисковый тезаурус (МИПТ)** это совокупность эквивалентных дескрипторов национальных версий МИПТ, связанных средствами установления эквивалентности (кодами идентификации) лексических единиц и упорядоченных по алфавиту.

2.3. **Лексическая единица (ЛЕ)** – выбранное для использования в МИПТ слово, словосочетание или лексически значимая компонента сложного слова естественного языка.

2.4. Основными ЛЕ тезауруса являются **дескрипторы** (основные термины) и **недескрипторы (аскрипторы)** (вспомогательные термины), упорядоченные по систематическому и алфавитному принципам и содержащие указания на имеющиеся между ними смысловые отношения как иерархического (родовидового), так и неиерархического типа.

2.5. **Дескриптор (д)** – ЛЕ информационно-поискового тезауруса (ИПТ), предназначенная для использования в поисковых образах документов и запросов на всех этапах обработки и поиска информации.

2.6. **Недескриптор (аскриптор)** – лексическая единица ИПТ, которая в поисковых образах документов и запросов подлежит замене на дескриптор или комбинацию дескрипторов.

2.7. **Модификатор (м)** – несамостоятельный дескриптор, который может применяться только в сочетании с дескриптором. Модификаторы используются для описания общего положения вещей, например, различных действий, свойств, размеров.

2.8. **Словарная статья** – элемент алфавитной части тезауруса, содержащий заглавную ЛЕ и лексические единицы, связанные с ней родовидовыми и (или) иерархическими отношениями.

2.9. **Релятор** является частью ЛЕ и поясняет ее значение, относя к определенной понятийной категории или предметно-тематической области. Устраняет полисемию терминов тезауруса.

3. ЭТАПЫ ПОСТРОЕНИЯ МИПТ ОСЖД

3.1. **Определение тематического охвата МИПТ** – осуществляется на основе анализа документопотока, обрабатываемого Центрами научно-технической и экономической информации (НТЭИ) стран-членов ОСЖД, а также Рубрикатора тематических направлений комплектования международного распределенного банка данных.

3.2. **Сбор массива лексических единиц** – осуществляется выделением ЛЕ из:

- документальных банков данных, формируемых Центрами НТЭИ, по критерию частоты их повторяемости в банках данных;
- вторичных источников лексики – энциклопедических, толковых и терминологических словарей и справочников по транспорту и техническим наукам, терминологических баз данных, различных систем классификации (УДК, МПК и т.д.).

3.3. **Формирование словника МИПТ.**

3.3.1. В словник МИПТ включены следующие типы ЛЕ:

- одиночные слова (существительные, прилагательные, наречия);
- именные словосочетания;
- аббревиатуры;
- принятые сокращения слов.

3.3.2. Дескрипторы и аскрипторы приведены к стандартной грамматической форме – именительный падеж единственного числа. Имена существительные в косвенных падежах, а также другие части речи – имена прилагательные, причастия, предлоги, союзы используются только в качестве компонентов дескрипторов-словосочетаний, либо как модификаторы.

3.3.3. Лексический состав МИПТ содержит:

- отраслевую железнодорожную терминологию и терминологическую лексику;
- лексику других отраслей транспорта;
- терминологию и терминологическую лексику смежных отраслей;
- общенаучную лексику;
- общелитературную лексику.

3.3.4. С содержательной точки зрения в МИПТ можно выделить несколько основных типов лексических единиц. Они могут обозначать:

- научные дисциплины;
- теоретические понятия, категории, методы, модели и т.п.;
- элементы производственной инфраструктуры, хозяйствующие субъекты;
- виды деятельности;
- технические средства, машины, механизмы, а также их узлы и детали;
- технологические процессы и их элементы;
- субъекты и объекты научного исследования;
- общественные процессы и явления.

3.4. Построение словарных статей

В МИПТ используются два типа словарных статей.

3.4.1. **Словарная статья дескриптора** содержит собственно дескриптор, перечень дескрипторов, связанных с ним парадигматическими отношениями, и недескрипторы, заменяемые данным дескриптором при индексировании или поиске.

Дескриптор, являющийся в иерархическом отношении вышестоящим по отношению к заглавному, снабжается в словарной статье ссылкой "в" (выше).

Например: **Стальная шпала Д**
в Шпала

Дескрипторы, нижестоящие по отношению к заглавному, вводятся в словарную статью со ссылкой "н" (ниже).

Например: **Перевозочный документ Д**
н Накладная

Кроме того, если все эти отношения не позволяют обеспечить однозначное понимание дескриптора, то в словарной статье могут дополнительно использоваться реляторы - краткие пометы, которые приводятся при дескрипторе в круглых скобках и являются его составной частью. Чаще всего они используются для выделения в качестве самостоятельного дескриптора одного из значений полисемичного термина.

Например: **Сцепка (передача усилия)**
Сцепка (поездное устройство)

В ряде случаев использование релятора позволяет выделить одно из значений термина в качестве отдельного дескриптора.

Например: **Социальное обеспечение**
Социальное обеспечение (услуги)

Релятор может быть также представлен сокращением.

Например: **Область (геогр.)**

Структура словарной статьи дескриптора представлена следующими элементами:

| Элемент словарной статьи | | | Значение элемента | Обязательность элемента |
|--------------------------|----------------|----|--|-------------------------|
| в н с | о и s.a. | BT | Заглавный дескриптор | обязательный |
| | | NT | Вышестоящие дескрипторы | факультативный |
| | | UF | Нижестоящие дескрипторы | факультативный |
| | | | Недескрипторы (синонимы), заменяемые при индексировании заглавным дескриптором | факультативный |

Пример словарной статьи дескриптора:

| | | |
|--|---|---|
| Крыша подвижного состава D 2762 в Кузов подвижного состава 2753 н Раздвижная крыша 7141 н Раздвижная стена 7143 н Скатывающаяся крыша 6702 с Люк крыши подвижного состава 3033 | Fahrzeugdach D 2762 o Fahrzeugaufbau 2753 u Schiebedach 7141 u Schiebewand 7143 u Rolldach 6702 s.a. <i>Fahrzeugdachoeffnung</i> 3033 | Rolling stock roof D 2762 BT Car body 2753 NT Sliding roof 7141 NT Sliding wall 7143 NT Roller roof 6702 UF <i>Manhole in vehicle roof</i> 3033 |
|--|---|---|

3.4.2. Словарная статья недескриптора содержит только отсылку к соответствующему дескриптору или комбинации дескрипторов и имеет следующий вид:

| Элемент словарной статьи | | | Значение элемента | Обязательность элемента |
|--------------------------|----|-----|---|-------------------------|
| см. | b. | USE | <i>Недескриптор (аскриптор)</i> | обязательный |
| | | | Используй = отсылка к заменяющему дескриптору | обязательный |

или

| | | | | |
|----|------|-----|---|--------------|
| ик | b.k. | USE | <i>Недескриптор (аскриптор)</i> | обязательный |
| | | | Используй комбинацию = отсылка к заменяющей комбинации дескрипторов | обязательный |

Пример словарной статьи недескриптора:

| | | |
|---|--|--|
| <i>Использование отходов тепла С 0053</i> см Рекуперация энергии 2312 | <i>Abwaermeverwertung S 0053</i> b. Energierueckgewinnung 2312 | <i>Waste heat utilization S 0053</i> USE Power regeneration 2312 |
|---|--|--|

или

| | | |
|---|---|---|
| <i>Отопление поезда С 9543</i> ик Отопление + Поезд 3936+9513 | <i>Zugheizung S 9543</i> b.k. Heizung + Zug 3936+9513 | <i>Train heating S 9543</i> USE Heating + Train 3936+9513 |
|---|---|---|

3.4.3. Обязательным компонентом словарных статей МИПТ являются средства установления эквивалентности дескрипторов национальных версий. Это решается путем приписывания эквивалентным дескрипторам специального цифрового четырехзначного идентификационного кода соответствия, который обеспечивает идентификацию понятий на принятых в тезаурусе языках.

3.5. Экспертиза и регистрация МИПТ

Разработанные МИПТ ОСЖД, включая их последующие переиздания, проходят экспертизу на соответствие их стандарту и депонируются в соответствующих депозитарных фондах. Разработчикам выдается сертификат соответствия.

4. СОСТАВ И СТРУКТУРА МИПТ ОСЖД

4.1. В состав МИПТ входит вводная часть, алфавитный лексико-семантический указатель и систематический указатель.

4.2. **Алфавитный лексико-семантический указатель** – основной указатель МИПТ. Представляет собой упорядоченную последовательность словарных статей МИПТ и формируется путем расположения их в алфавитном порядке заглавных ЛЕ.

4.2.1. В Алфавитном указателе представлены все лексические единицы – дескрипторы и недескрипторы – с полными словарными статьями. Включение в Тезаурус недескрипторов существенно повышает его семантическую силу. С одной стороны, включение недескриптора в словарную статью дескриптора позволяет более точно и однозначно описать семантику дескриптора, отразить входящие в него компоненты. С другой стороны, пользователь может обратиться к информационной системе по любым известным ему терминам, а если эти термины зафиксированы в МИПТ в качестве запрещенных для использования в информационной системе, то получить отсылку к релевантному дескриптору.

Например:

| | | |
|--|---|---|
| <i>Преобразователь С 8414</i> см Трансформатор 8148 | <i>Umformer S 8414</i> b. Transformator 8148 | <i>Converter S 8414</i> USE Transformer 8148 |
|--|---|---|

4.2.2. В Алфавитном лексико-семантическом указателе полностью описана семантика дескрипторов и все парадигматические отношения релевантного дескриптора.

4.2.3. Упорядочение ЛЕ внутри словарной статьи также осуществляется по алфавиту соответствующей языковой версии.

4.2.4. Заглавные ЛЕ в словарных статьях выделены жирным шрифтом и сдвигом влево относительно начала незаглавных ЛЕ.

4.2.5. Недескрипторы в словарных статьях даются курсивом.

4.2.6. Алфавитный лексико-семантический указатель используется как при индексировании документов, так и при поиске для того, чтобы найти релевантный дескриптор или проверить его значение в информационной системе. При поиске он может служить также средством для изменения стратегии поиска.

4.3. **Систематический указатель** содержит только дескрипторы без словарных статей, упорядоченные по рубрикам Рубриката тематических направлений комплектования международного распределенного банка данных (далее – Рубрикат), представляющего собой иерархическую классификационную систему.

4.3.1. Для предоставления возможности быстрого тематического поиска документов по отраслям железнодорожного транспорта и видам деятельности содержание каждого документа, вводимого в систему, описывается рубриками Рубриката в дополнение к дескрипторному описанию содержания. Главная функция Рубриката – упорядочение документов по этим рубрикам.

4.3.2. Каждый дескриптор в Систематическом указателе отнесён к одной рубрике.

4.3.3. Упорядочение рубрик в Систематическом указателе производится в порядке возрастания их цифрового кода. Внутри каждой рубрики упорядочение дескрипторов производится по алфавиту соответствующей языковой версии.

Фрагмент систематического указателя:

| Код рубрики | Код дескр. | Русский язык | Немецкий язык | Английский язык |
|--------------------|------------|---------------------------------------|--------------------------------|--------------------------------|
| 73.29.71.31 | | Контактная сеть | Fahrleitungsnetz | Catenary |
| 73.29.71.31 | 2717 | Контактная сеть | Fahrleitung | Catenary |
| 73.29.71.31.11 | 4496 | Изолятор | Isolator | Insulator |
| 73.29.71.31.11 | 2718 | Подвеска контактного провода | Fahrleitungsaufhaengung | Contact wire suspension |
| 73.29.71.31.13 | 2708 | Контактный провод | Fahrdraht | Contact wire |
| 73.29.71.31.19 | 7403 | Линейное секционирование [конт. сеть] | Streckentrennung [Fahrleitung] | Line sectionalizing [catenary] |
| 73.29.71.31.19 | 7301 | Питающий провод | Speiseleitung | Feeder |
| 73.29.71.31.19 | 7438 | Смена систем тока | Systemwechsel | Current system change |
| 73.29.71.31.23 | 2719 | Монтаж контактной сети | Fahrleitungsbau | Catenary installation |

4.3.4. Систематический указатель используется для поиска релевантной запросу рубрики или дескрипторов в пределах определенной рубрики, представляющей интерес для индексатора или пользователя, т. е. сначала индексатор находит подходящую рубрику, а затем, в случае необходимости, ищет в ней релевантные дескрипторы.

5. ПРАВИЛА ИСПОЛЬЗОВАНИЯ МИПТ ОСЖД

5.1. МИПТ используются для индексирования документов при создании национальных баз и банков данных и при международном обмене научнотехнической информацией.

5.2. С помощью индексов МИПТ (кодов рубрик, дескрипторов, кодов дескрипторов) передается содержание документов. Переданное содержание, в первую очередь, служит критерием поиска информации.

5.3. Процесс индексирования происходит, в основном, в два этапа: определение основных компонентов содержания документа и формулировка ключевых слов на базе естественного языка; перевод ключевых слов в элементы информационно-поискового языка, т.е. в рубрики и дескрипторы МИПТ.

5.4. Для индексирования документов и запросов применяются только дескрипторы, модификаторы, рубрики.

5.5. Модификаторы (*М*) - несамостоятельные дескрипторы, должны применяться только в сочетании с дескрипторами. Они используются для отображения общего положения вещей. Их не следует ставить на первом месте.

Например: *Взаимосвязь М*

5.6. Недескриптор - синоним (*С*) имеет помету "см" (смотри), которая дает ссылку на дескриптор, применяемый при индексировании:

Например: *Двигатель С*

см Тяговый двигатель

или помету "ик" (используй комбинацию), которая дает ссылку на сочетание дескрипторов, применяемое при индексировании:

Например: *Вес поезда С*

ик Поезд + Масса груза

5.7. Тезаурус обеспечивает следующие возможности поиска релевантных дескрипторов в АБД ОСЖД:

| Тип поиска | Вид указателя |
|---|--|
| 1. Поиск дескрипторов по словам и словосочетаниям, известным индексатору или пользователю, уточнение их значения в информационной системе | Алфавитный лексико-семантический указатель |
| 1. 2. Поиск рубрик и дескрипторов в рубриках Рубрикатора тематических направлений комплектования АБД ОСЖД | Систематический указатель |

6. ТРЕБОВАНИЯ К ПРОГРАММНО-ТЕХНИЧЕСКИМ СРЕДСТВАМ ДЛЯ СОЗДАНИЯ, ИСПОЛЬЗОВАНИЯ И ВЕДЕНИЯ МИПТ ОСЖД

6.1. Для создания машинной версии МИПТ используются персональные компьютеры типа IBM и имеющиеся в распоряжении составителей МИПТ стандартные профессиональные текстовые процессоры (редакторы текстов) и программы создания и ведения терминологических баз данных.

6.2. Используемые программы для текстообработки должны обеспечивать:

ввод словаря терминов;

просмотр словаря с возможностью выдачи частоты использования терминов в базе данных;

ввод новых, корректировку, либо удаление существующих терминов словаря;

сортировку по алфавиту соответствующей языковой версии и по тематике;

печать словаря терминов на отдельных, принятых в МИПТ, языках.

7. ЭКСПЕРИМЕНТАЛЬНАЯ ПРОВЕРКА И ВЕДЕНИЕ МИПТ ОСЖД

7.1. Испытание МИПТ в практическом индексировании является основой для их обязательного применения и дальнейшего развития.

Целью испытания является:

- . исследование по международной применимости;
- . анализ относительной полноты дескрипторов и недескрипторов;
- . оценка правильности и полноты понятийных отношений.

На основе результатов экспериментальной проверки МИПТ следует их постоянная переработка.

7.2. МИПТ ОСЖД «Железнодорожный транспорт» разрабатывается сроком на пять лет. Изменения в тезаурусы вносятся редакционной группой экспертов по НТЭИ ОСЖД по разработке тезаурусов.

7.3. Имеется возможность исключать некоторые понятия или включать новые. Страны-члены ОСЖД, участвующие в международном обмене информацией, имеют право давать свои предложения ведущим странам по разработке тезауруса и копии всем членам рабочей группы по тезаурусу.

Возможны следующие предложения:

а). Предложения по дополнению:

- определение понятия;
- обоснование предложения;
- перевод понятия на языки, принятые в МИПТ.

б). Предложения по исключению:

- обоснование предложения на основе статистических показателей повторяемости.

7.4. Предложения дорог обсуждаются на совещаниях экспертов по НТЭИ (редакционной группой по тезаурусу – ведущие дороги ЛДЗ, ПКП, РЖД). Предложения, согласованные и принятые редакционной группой (ведущими дорогами) по тезаурусу, включаются в последующие издания тезаурусов.

Необходимые изменения принимаются и в национальных изданиях МИПТ центров НТЭИ стран-членов ОСЖД.