МЕЖГОСУДАРСТВЕННЫЙ СТАНДАРТ

Система стандартов по информации, библиотечному и издательскому делу

ТЕЗАУРУС ИНФОРМАЦИОННО-ПОИСКОВЫЙ ОДНОЯЗЫЧНЫЙ

Правила разработки, структура, состав и форма представления

Издание официальное

БЗ 3—2001/35

МЕЖГОСУДАРСТВЕННЫЙ СОВЕТ
ПО СТАНДАРТИЗАЦИИ, МЕТРОЛОГИИ И СЕРТИФИКАЦИИ
Минск

gost.e-ecolog.ru

Предисловие

1 РАЗРАБОТАН Всероссийским институтом научной и технической информации Российской Академии наук, Минпромнауки России и Межгосударственным техническим комитетом по стандартизации МТК 191 «Научно-техническая информация, библиотечное и издательское дело»

ВНЕСЕН Госстандартом России

2 ПРИНЯТ Межгосударственным Советом по стандартизации, метрологии и сертификации (протокол № 19 от 22 мая 2001 г.)

За принятие проголосовали:

Наименование государства	Наименование национального органа по стандартизации
Азербайджанская Республика	Азгосстандарт
Республика Беларусь	Госстандарт Республики Беларусь
Республика Казахстан	Госстандарт Республики Казахстан
Кыргызская Республика	Кыргызстандарт
Российская Федерация	Госстандарт России
Республика Таджикистан	Таджикстандарт
Туркменистан	Главгосслужба «Туркменстандартлары»
Республика Узбекистан	Узгосстандарт
Украина	Госстандарт Украины

3 Постановлением Государственного комитета Российской Федерации по стандартизации и метрологии от 4 сентября 2001 г. № 370-ст межгосударственный стандарт ГОСТ 7.25—2001 введен в действие непосредственно в качестве государственного стандарта Российской Федерации с 1 июля 2002 г.

4 ВЗАМЕН ГОСТ 7.25—80

© ИПК Издательство стандартов, 2001

Настоящий стандарт не может быть полностью или частично воспроизведен, тиражирован и распространен в качестве официального издания на территории Российской Федерации без разрешения Госстандарта России

МЕЖГОСУДАРСТВЕННЫЙ СТАНДАРТ

Система стандартов по информации, библиотечному и издательскому делу

ТЕЗАУРУС ИНФОРМАЦИОННО-ПОИСКОВЫЙ ОДНОЯЗЫЧНЫЙ

Правила разработки, структура, состав и форма представления

System of standards on information, librarianship and publishing. Monolingual thesaurus for information retrieval. Rules for its development, structure, composition and form of presentation

Дата введения 2002—07—01

1 Область применения

Настоящий стандарт устанавливает правила разработки, структуру, состав и форму представления информационно-поисковых тезаурусов (далее — ИПТ), ориентированных на использование лексики русского языка и разрабатываемых в рамках автоматизированных информационных систем и сетей научно-технической информации.

2 Нормативные ссылки

В настоящем стандарте использованы ссылки на следующие стандарты:

ГОСТ 7.0—99 Система стандартов по информации, библиотечному и издательскому делу. Информационно-библиотечная деятельность, библиография. Термины и определения

ГОСТ 7.47—84 Система стандартов по информации, библиотечному и издательскому делу. Коммуникативный формат для словарей информационных языков и терминологических данных

ГОСТ 7.49—84 Система стандартов по информации, библиотечному и издательскому делу. Рубрикатор ГАСНТИ. Структура, правила пользования и ведения

ГОСТ 7.67—94 (ИСО 3166—88) Система стандартов по информации, библиотечному и издательскому делу. Коды названий стран

ГОСТ 7.73—96 Система стандартов по информации, библиотечному и издательскому делу. Поиск и распространение информации. Термины и определения

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

ГОСТ 7.75—97 Система стандартов по информации, библиотечному и издательскому делу. Коды наименований языков

ГОСТ 7.77—98 Система стандартов по информации, библиотечному и издательскому делу. Межгосударственный рубрикатор научно-технической информации. Структура, правила использования и ведения

3 Определения

В настоящем стандарте применяют следующие термины с соответствующими определениями:

3.1 **лексическая единица (ЛЕ) ИПТ:** Слово, словосочетание или лексически значимый компонент сложного слова естественного языка, включенное в ИПТ в качестве дескриптора или аскриптора.

3.2 **класс эквивалентности ЛЕ:** Совокупность лексических единиц, различие значение которых несущественно для определения основного содержания документов в информационной системе.

3.3 Остальные термины — по ГОСТ 7.0, ГОСТ 7.73, ГОСТ 7.74.

Издание официальное

4 Правила построения информационно-поискового тезауруса

4.1 По своему построению ИПТ подразделяют на два типа:

- тезаурусы, выделяющие среди своих лексических единиц дескрипторы и аскрипторы;
- тезаурусы, все ЛЕ которых являются дескрипторами.

Построение ИПТ включает следующие этапы:

- определение тематического охвата ИПТ;
- сбор массива ЛЕ;
- формирование словника ИПТ;
- построение словарных статей и указателей;
- оформление ИПТ;
- экспертиза и регистрация ИПТ.

4.2 Определение тематического охвата ИПТ

Определение тематического охвата ИПТ осуществляется на основе Государственного рубрикатора НТИ по ГОСТ 7.49 или Межгосударственного рубрикатора НТИ по ГОСТ 7.77 путем анализа информационной потребности абонентов (потребителей). При анализе состава имеющегося или проектируемого информационного фонда и информационных потребностей абонентов отмечаются все рубрики нижнего уровня Рубрикатора, по которым осуществляется комплектование фондов или которые соответствуют запросам абонентов.

4.3 Сбор массива лексических единиц

До начала работ по сбору массива ЛЕ разработчик должен обратиться в депозитарный фонд с целью определения наличия зарегистрированных тезаурусов по заданной тематике.

При наличии таких тезаурусов проводят оценку возможности внедрения их в данной системе.

Первоначальный сбор лексики осуществляется выделением ЛЕ из представительной коллекции документов и/или запросов.

В полученный массив ЛЕ дополнительно должны быть включены соответствующие тематике ЛЕ, выделенные в соответствии с методикой построения данного тезауруса из следующих источников:

Государственного рубрикатора НТИ (ГОСТ 7.49);

рубрикатора соответствующей системы научно-технической информации;

энциклопедических, толковых и терминологических словарей и справочников;

терминологических стандартов;

классификаторов технико-экономической информации;

таблиц Универсальной десятичной классификации (УДК) и других систем классификации;

нормативных (авторитетных) словарей коммуникативных форматов.

Наименования стран и других географических единиц должны соответствовать требованиям ГОСТ 7.67, наименования языков — требованиям ГОСТ 7.75.

4.4 Формирование словника ИПТ

4.4.1 В словник ИПТ могут быть включены следующие типы ЛЕ:

одиночные слова (существительные, прилагательные, глаголы, наречия);

именные словосочетания;

лексически значимые компоненты сложных слов;

сокращения слов и словосочетаний.

4.4.2 Допускается включать словосочетания в словник, если в качестве опорного слова они содержат существительное и если выполнено одно из следующих условий:

- значение словосочетания не выводится из значений его компонентов.

Примеры

ЧЕРНЫЙ ЯЩИК,
АБСОЛЮТНО ЧЕРНОЕ ТЕЛО,
ЦАРСКАЯ ВОДКА;

- хотя бы один из компонентов словосочетания не употребляется в составе других сочетаний или употребляется всегда в другом смысле.

Примеры

ТОРГОВЛЯ НА ВЫНОС,
ЛЕГКАЯ ПРОМЫШЛЕННОСТЬ;

- для данного словосочетания в словнике ИПТ существуют полные синонимы.

Пример

НАТРИЯ ХЛОРИД = ПОВАРЕННАЯ СОЛЬ;

- данное словосочетание является устойчивым словосочетанием с именем собственным.

Примеры

ТАБЛИЦА МЕНДЕЛЕЕВА,

ЗАКОН БОЙЛЯ-МАРИОТТА;

- отдельные слова словосочетания имеют слишком широкое значение.

Пример

слово МАШИНЫ в словосочетаниях:

СТРОИТЕЛЬНЫЕ МАШИНЫ,

ЭЛЕКТРИЧЕСКИЕ МАШИНЫ;

- для данного словосочетания в словнике ИПТ существует общепринятая аббревиатура.

Примеры

ПОВЕРХНОСТНО-АКТИВНЫЕ ВЕЩЕСТВА = ПАВ,

УНИВЕРСАЛЬНАЯ ДЕСЯТИЧНАЯ КЛАССИФИКАЦИЯ = УДК,

ИНФОРМАЦИОННО-ПОИСКОВЫЙ ТЕЗАУРУС = ИПТ,

ЭЛЕКТРОННО-ВЫЧИСЛИТЕЛЬНАЯ МАШИНА = ЭВМ;

- разбиение словосочетаний на отдельные компоненты приводит к потере важных для поиска семантических связей.

Пример

разбиение ЛЕ «ЯЗЫК ПРОГРАММИРОВАНИЯ» не позволяет установить связи с такими ЛЕ, как «АЛГОЛ», «КОБОЛ», «ФОРТРАН».

Словосочетания, которые не удовлетворяют перечисленным условиям, разбивают на компоненты.

4.4.3 Падежные формы одиночных существительных следует заменять формой именительного падежа. Падежные формы опорных существительных в словосочетаниях также заменяют формой именительного падежа с одновременным изменением падежной формы согласованных определений.

4.4.4 В ИПТ должно быть принято последовательное представление ЛЕ, выраженных именами существительными и именными словосочетаниями, в формах единственного либо множественного числа в соответствии с традицией лексикографических источников, либо практикой промышленных каталогов. В предисловии к ИПТ должны быть указаны принципы выбора формы числа ЛЕ.

4.4.5 Формы прилагательных и причастий следует приводить к именительному падежу. Прилагательные и причастия в единственном числе приводят к форме мужского рода.

4.4.6 Глаголы рекомендуется заменять отглагольными существительными. Глаголы, включенные в словник, приводят к форме инфинитива.

4.4.7 Компоненты сложных слов, включаемые в словник ИПТ, должны быть широко употребляющимися частями сложных слов, общепринятых в данной специальности. Отсутствующий компонент сложного слова заменяют дефисом.

Примеры

ЭЛЕКТРО-

ГИДРО-

ВОДО-

ТЕРМО-.

4.4.8 При наличии аббревиатур в словник должны быть включены также соответствующие им полные формы.

Примеры

ИПТ — ИНФОРМАЦИОННО-ПОИСКОВЫЙ ТЕЗАУРУС,

ОКГ — ОПТИЧЕСКИЙ КВАНТОВЫЙ ГЕНЕРАТОР,

УКВ — УЛЬТРАКОРОТКОВОЛНОВЫЙ.

4.4.9 Сокращения должны соответствовать требованиям стандартов на сокращения слов в библиографических описаниях.

4.4.10 Исключения из требований 4.4 возможны только в тех случаях, когда перечисленные типы ЛЕ отсутствуют.

4.5 Построение словарных статей

4.5.1 При построении словарных статей ИПТ лексическим единицам приписывают ссылки согласно таблице 1.

Таблица 1 — Типы и значения ссылок

Тип ссылки	Обозначение на русском языке	Символьное обозначение	Значение ссылки	Аналог на английском языке
1 Ссылка от аскриптора к эквивалентному дескриптору	см.	→	смотри	USE
2 Ссылка от дескриптора к эквивалентному аскриптору	с	=	синоним	UF (used for)
3 Ссылка от аскриптора к нескольким альтернативно заменяющим его дескрипторам	иа	=:	используй альтернативно	—
4 Ссылка от аскриптора к заменяющей его комбинации дескрипторов	ик	=+	используй комбинацию	—
5 Ссылка от дескриптора к вышестоящему дескриптору	в	<	выше	BT (broader term)
6 Ссылка от дескриптора к вышестоящему родовому дескриптору	вр	:<	выше—род	BTG (broader term generic)
7 Ссылка от дескриптора к вышестоящему дескриптору, обозначающему целое	вц	— <	выше—целое	BTP (broader term partitive)
8 Ссылка от дескриптора к нижестоящему дескриптору	н	>	ниже	NT (narrower term)
9 Ссылка от дескриптора к нижестоящему видовому дескриптору	нв	>:	ниже—вид	NTG (narrower term generic)
10 Ссылка от дескриптора к нижестоящему дескриптору, обозначающему часть	нч	> —	ниже—часть	NTP (narrower term partitive)
11 Ссылка от дескриптора к ассоциативно связанному дескриптору	а	— (черта)	ассоциация	RT (related term)
12 Ссылка от дескриптора к аскриптору, который заменяется данным дескриптором при альтернативном выборе (обратная ссылка к «иа»)	са	:=	сравни альтернативный выбор	—
13 Ссылка от дескриптора к аскриптору, который заменяется комбинацией, включающей данный дескриптор (обратная ссылка к «ик»)	ск	+=	сравни комбинацию	—
14 Техническая обратная ссылка	ср	:	сравни	—
15 Уточнение значения и области применения	лп	/ ... /	лексическое примечание	SN (scope note)

4.5.2 Ссылки отмечают связи данной ЛЕ с другими и являются результатом выполнения следующих операций:

- устранения неоднозначности ЛЕ;
- установления отношений эквивалентности;
- выбора дескриптора, представляющего класс эквивалентности при индексировании (для ИПТ 1-го типа, различающих дескрипторы и аскрипторы);
- установления иерархических и ассоциативных отношений дескрипторов.

4.5.3 Неоднозначность ЛЕ устраняется релятором или лексическим примечанием в зависимости от характера неоднозначности и способа использования ИПТ.

4.5.3.1 Релятор является частью ЛЕ и поясняет ее значение, относя ее к определенной понятийной категории или предметно-тематической области. Допускается релятор представлять сокращением.

Примеры

СТАБИЛИЗАТОР (ЭЛЕКТРОТЕХН.),
СТАБИЛИЗАТОР (ХИМ.),
СТАБИЛИЗАТОР (АВИАЦИЯ),
СМАЗКА (ПРОЦЕСС),
СМАЗКА (ВЕЩЕСТВО).

Допускается в качестве реляторов использовать кодовые обозначения, смысл которых поясняется во введении к ИПТ.

Примеры

СТАБИЛИЗАТОР (45),
СТАБИЛИЗАТОР (61),
СТАБИЛИЗАТОР (55.47),

где в качестве реляторов использованы коды соответствующих рубрик Государственного рубрикатора НТИ.

4.5.3.2 Лексическое примечание не является частью ЛЕ и представляет собой поясняющий текст на естественном языке, отличающий значение ЛЕ от других значений, не рассматриваемых в данном ИПТ.

Примеры

МЛЕКОПИТАЮЩИЕ (виды на суше),
АЭРОПЛАН (только для документов до 1940 г.),
СПЛАВЫ МЕДИ (сплавы, в которых медь является основным компонентом).

4.5.4 ЛЕ объявляются эквивалентными в ИПТ, если замена одной ЛЕ на другую не приводит к изменению смысла текста, существенному для поиска информации.

4.5.4.1 Эквивалентность устанавливается между:

- абсолютными синонимами, в частности — между аббревиатурами и полными формами ЛЕ.

Примеры

ИПТ = ИНФОРМАЦИОННО-ПОИСКОВЫЙ ТЕЗАУРУС,
ГЕОМЕТРИЯ РИМАНА = РИМАНОВА ГЕОМЕТРИЯ,
НАГРЕВ = НАГРЕВАНИЕ,
КОНУСНЫЙ = КОНИЧЕСКИЙ = В ФОРМЕ КОНУСА;

- стилистическими синонимами (если различие в сфере употребления несущественно для задач ИПТ)

Примеры

ПНЕВМОНИЯ = ВОСПАЛЕНИЕ ЛЕГКИХ,
АДГЕЗИЯ = ПРИЛИПАНИЕ,
ЛЕНТА ЛИПКАЯ = ЛЕНТА КЛЕЙКАЯ;

- относительными синонимами (если случаи несовпадения значений несущественны для задач ИПТ)

Примеры

СТОЛ = ДИЕТА = ПИТАНИЕ,
БЮРО = КОНТОРА = ФИРМА,
ВИНТ = БОЛТ.

4.5.4.2 Допускается установление эквивалентности также между ЛЕ, различными по значению, но семантически связанными, в тех случаях, когда отождествление этих понятий полезно для функционирования информационной системы.

Примеры

УСТОЙЧИВОСТЬ = НЕУСТОЙЧИВОСТЬ,
ТОРГОВЛЯ = ПРОДАЖА,
РЕКА = РУЧЕЙ,
МАСЛО = СМАЗКА.

4.5.5 Выбор дескрипторов проводится при формировании ИПТ первого типа.

4.5.5.1 Для выбора дескрипторов в классах эквивалентности выбирают одну ЛЕ в качестве представителя всего класса.

Если в классе эквивалентности есть существительные, то представителем класса объявляется одно из них.

Если в классе эквивалентности более одного существительного, а также в тех случаях, когда существительные отсутствуют, критериями выбора представителя являются:

- полнота выражения значений данного класса ЛЕ;
- соответствие стандартам и рекомендациям по научно-технической терминологии;
- краткость и понятность;
- при прочих равных условиях наибольшая частота использования (распространенность) в документах и запросах.

4.5.5.2 Каждая ЛЕ получает статус дескриптора или аскриптора.

Статус аскриптора присваивают в следующих случаях:

- ЛЕ входит в класс эквивалентности, но не является представителем класса;
- значение ЛЕ целесообразно при поиске информации представить как пересечение значений других ЛЕ, получивших статус дескриптора;
- значение ЛЕ целесообразно при поиске информации представить как объединение значений других ЛЕ, несовместимых в одном поисковом образе.

В остальных случаях ЛЕ получает статус дескриптора.

4.5.5.3 Словарные статьи ЛЕ, получивших статус аскриптора, должны быть снабжены ссылками на заменяющие их дескрипторы.

Примеры

1) Аскриптор — синоним дескриптора

Языковедение	Фарси
см. ЯЗЫКОЗНАНИЕ	см. ПЕРСИДСКИЙ ЯЗЫК

2) Аскриптор, заменяемый комбинацией дескрипторов

Русские тексты	Алфавитно-цифровые знаки
ик РУССКИЙ ЯЗЫК ТЕКСТЫ	ик БУКВЫ ЦИФРЫ

3) Аскриптор, заменяемый несовместимыми дескрипторами

Искусственные языки

иа ИНФОРМАЦИОННО-ПОИСКОВЫЕ ЯЗЫКИ
ЯЗЫКИ МЕЖДУНАРОДНОГО ОБЩЕНИЯ
ЯЗЫКИ ПРОГРАММИРОВАНИЯ

4.5.5.4 Если статус аскриптора получают ЛЕ, представляющие класс эквивалентности, то все ЛЕ, входящие в этот класс, получают статус аскриптора, и в их словарных статьях должны быть идентичные ссылки на заменяющие дескрипторы.

Пример

Персидские тексты	Тексты на фарси
ик ПЕРСИДСКИЙ ЯЗЫК ТЕКСТЫ	ик ПЕРСИДСКИЙ ЯЗЫК ТЕКСТЫ

4.5.5.5 Статьи дескрипторов снабжаются ссылками на заменяемые ими аскрипторы.

Примеры

1) ЯЗЫКОЗНАНИЕ	ПЕРСИДСКИЙ ЯЗЫК
с лингвистика языковедение	с фарси
2) РУССКИЙ ЯЗЫК	ТЕКСТЫ
ск русские тексты	ск персидские тексты русские тексты тексты на фарси
3) ИНФОРМАЦИОННО-ПОИСКОВЫЕ ЯЗЫКИ	ЯЗЫКИ ПРОГРАММИРОВАНИЯ
са искусственные языки	са искусственные языки

4.5.6 Установление иерархических и ассоциативных отношений

4.5.6.1 Для дескрипторов устанавливаются парадигматические отношения, отражающие лексико-семантические связи между понятиями, выражаемыми дескрипторами. Связь указывают путем внесения в дескрипторную статью ссылки, включающей обозначение согласно таблице ссылок и связанный дескриптор.

4.5.6.2 Основными типами связей являются:

- род — вид,
- часть — целое,
- причина — следствие,
- сырье — продукт,
- административная иерархия,
- процесс — объект,
- функциональное сходство,
- процесс — субъект,
- свойство — носитель свойства,
- антонимия.

4.5.6.3 Отношения в ИПТ (исключая синонимию, рассмотренную в 4.5.5) могут быть разделены на два класса: иерархические и ассоциативные.

4.5.6.4 Иерархические отношения обладают свойствами транзитивности и антисимметричности, которые могут быть использованы при избыточном индексировании в интересах повышения эффективности информационного поиска. Предпочтительно указывать связи между дескрипторами как отношения иерархического вида, если они обладают этими свойствами. Применяемые в ИПТ иерархические отношения могут дифференцироваться на отдельные виды.

4.5.6.5 В ИПТ должно фиксироваться наличие иерархического отношения дескрипторов указанием связей в соответствии с таблицей ссылок.

4.5.6.6 Родовидовая связь устанавливается между двумя дескрипторами, если объем понятия нижестоящего дескриптора входит в объем понятия вышестоящего дескриптора.

Примеры

МЕБЕЛЬ	СТУЛЬЯ	СТОЛЫ
нв СТУЛЬЯ	вр МЕБЕЛЬ	вр МЕБЕЛЬ
СТОЛЫ		

4.5.6.7 Связь часть — целое устанавливается между двумя дескрипторами в том случае, если нижестоящий дескриптор обозначает компонент объекта, обозначаемого вышестоящим дескриптором.

Примеры

АВТОМОБИЛИ	КУЗОВ	ШАССИ
нч КУЗОВ	вц АВТОМОБИЛИ	вц АВТОМОБИЛИ
ШАССИ		

4.5.6.8 Если для одного дескриптора можно указать более одного непосредственно вышестоящего дескриптора, то в иерархических отношениях должны быть установлены связи со всеми дескрипторами.

Примеры

ТЕЛЯТА
в КРУПНЫЙ РОГАТЫЙ СКОТ
МОЛОДНЯК ЖИВОТНЫХ

КОМНАТА
вр ПОМЕЩЕНИЕ
вц ДОМ.

4.5.6.9 При установлении иерархических отношений должны быть указаны связи со всеми нижестоящими дескрипторами независимо от аспекта деления. Аспект деления может быть указан в примечании при ссылке.

Пример

АНТЕННЫ
н ВЫСОКОЧАСТОТНЫЕ АНТЕННЫ (диапазон)
НИЗКОЧАСТОТНЫЕ АНТЕННЫ (диапазон)
ШИРОКОПОЛОСНЫЕ АНТЕННЫ (диапазон)
ПАРАБОЛИЧЕСКИЕ АНТЕННЫ (конструкция)
ТЕЛЕСКОПИЧЕСКИЕ АНТЕННЫ (конструкция)
ШТЫРЕВЫЕ АНТЕННЫ (конструкция)
ЩЕЛЕВЫЕ АНТЕННЫ (конструкция).

4.5.6.10 При необходимости иерархические связи в дескрипторных статьях могут быть указаны на два и более уровней. В таких случаях уровни иерархии нумеруют относительно заглавной ЛЕ.

Пример

ПОЛК
в1 ДИВИЗИЯ
в2 АРМИЯ
н1 БАТАЛЬОН
н2 РОТА
БАТАРЕЯ

4.5.6.11 Ассоциативное отношение является объединением отношений, не входящих в иерархические отношения или в отношения синонимии. Допускается включать в ассоциативное отношение все виды отношений, кроме синонимии и отношения род — вид.

4.5.7 В целях обеспечения ведения ИПТ и индексирования документов ссылки, фиксирующие связи ЛЕ в словарных статьях, должных быть взаимными, т.е. для каждой ссылки, указывающей связь заглавной ЛЕ с другой ЛЕ, в словарной статье другой ЛЕ должна быть обратная ссылка. Если нецелесообразно использовать обратную ссылку при поиске информации, то следует применять технологическую обратную ссылку «сравни», обеспечивающую ведение ИПТ.

4.5.7.1 В дескрипторной статье указывают все синонимы заглавного дескриптора. Для каждого синонимичного аскриптора указывают один заменяющий дескриптор.

4.5.7.2 Для каждой пары дескрипторов, связанных недифференцированной иерархической связью (ссылки «выше» и «ниже») или иерархической связью по отношению род — вид (ссылки «выше — род» и «ниже — вид»), в статье вышестоящего дескриптора должна быть ссылка на нижестоящий, а в статье нижестоящего — на вышестоящий.

4.5.7.3 Для дескрипторов, связанных иерархической связью по отношению часть — целое, допустимо давать иерархическую ссылку только от вышестоящего к нижестоящему или наоборот. В этих случаях обратная иерархическая ссылка заменяется неиспользуемой при поиске технической ссылкой «ср».

4.5.7.4 Для указания дескрипторов в статьях аскрипторов на ссылки «иа» и «ик», используются соответственно ссылки «са» («сравни альтернативный выбор») и «ск» («сравни комбинацию»).

4.6 Построение лексико-семантического указателя

4.6.1 Лексико-семантический указатель является упорядоченной последовательностью словарных статей ИПТ и формируется путем расположения их в алфавитном порядке заглавных ЛЕ.

Устанавливается следующий приоритет символов при алфавитном расположении ЛЕ:

- пробел и знаки препинания, кроме дефиса (приоритеты всех знаков препинания и их сочетаний считаются одинаковыми),
- русские буквы,
- цифры,
- латинские буквы,
- буквы других алфавитов и специальные символы.

При этом ЛЕ, написанные через дефис, занимают то же алфавитное место, что и при слитном написании.

4.6.2 Идентификаторы могут быть выделены в отдельные списки, упорядоченные по алфавиту и размещенные после основного перечня словарных статей.

4.7 Построение систематического указателя

4.7.1 Систематический указатель является перечнем дескрипторов, сгруппированных согласно принятой в ИПТ рубрикации.

4.7.2 Систематические указатели ИПТ подразделяют на три типа:

- тематические,
- категориальные,
- смешанные.

4.7.3 При построении систематического указателя тематического и смешанного типов в его тематической части следует использовать рубрики Межгосударственного рубрикатора НТИ или рубрикатора конкретной АСНТИ, совместимого с Межгосударственным рубрикатором НТИ.

4.7.4 При построении систематического указателя категориального и смешанного типов в его категориальной части следует использовать следующие общие категории:

- названия дисциплин и отраслей деятельности;
- предметы, материалы;
- методы, процессы, операции, явления;
- свойства, величины, параметры, характеристики;

- отношения, структуры, модели, законы, правила, абстрактные понятия.

4.7.5 В систематическом указателе категориального типа каждый дескриптор должен быть отнесен только к одной рубрике.

4.7.6 Внутри каждой рубрики дескрипторы располагают в алфавитном порядке.

4.8 Построение указателя иерархических отношений

4.8.1 Иерархический указатель отражает полную структуру иерархических отношений.

4.8.2 Иерархический указатель представляет собой перечень списков дескрипторов, причем каждый список начинается с дескриптора, не имеющего вышестоящих. После каждого дескриптора приведены непосредственно нижестоящие дескрипторы с указанием их уровня в иерархии путем применения нумерации либо графического обозначения уровня.

Пример

АРМИЯ	АРМИЯ
. ДИВИЗИЯ	1 ДИВИЗИЯ
. . ПОЛК	2 ПОЛК
. . . БАТАЛЬОН	3 БАТАЛЬОН
. . . . РОТА	4 РОТА

4.9 Построение пермутационного указателя

4.9.1 Пермутационный указатель предназначен для поиска ЛЕ по отдельным словам, входящим в текст ЛЕ, в том числе и по тем, которые стоят не в начале ЛЕ.

4.9.2 Пермутационный указатель представляет собой перечень списков, каждый из которых соответствует одному из знаменательных слов в составе ЛЕ и включает все ЛЕ, содержащие это знаменательное слово. Перечень списков упорядочен по алфавиту знаменательных слов, а внутри списки упорядочены по тексту ЛЕ.

4.9.3 Если число словосочетаний в ИПТ невелико, то допускается указывать пермутации словосочетаний непосредственно в лексико-семантическом указателе. При этом словосочетания в канонической форме приводятся в качестве дескрипторов, а другие варианты — в качестве синонимичных аскрипторов.

4.10 Автоматизация построения ИПТ

4.10.1 Построение тезаурусов с использованием ПЭВМ позволяет автоматизировать следующие процессы построения ИПТ:

- частотный анализ,
- коррекцию статей,
- алфавитную сортировку словника,
- проверку взаимности и непротиворечивости ссылок,
- составление указателей,
- распечатку в требуемых формах.

4.10.2 Автоматизированное построение ИПТ должно допускать его представление в коммуникативном формате по ГОСТ 7.47 на машинных носителях информации и в виде комплекта указателей, напечатанных на бумаге с помощью компьютера и отражающих текущее состояние машинного тезауруса.

4.10.3 Состав, структура и форма представления указателей, напечатанных с помощью компьютера, должны соответствовать 4.11.

4.11 Структура, состав и форма представления тезауруса

В состав ИПТ входит вводная часть, основная часть (лексико-семантический указатель) и дополнительные части (систематический, пермутационный, иерархический и др. указатели и списки специальных категорий ЛЕ).

Обязательными составными частями являются вводная часть и лексико-семантический указатель.

Допускается в состав ИПТ вводить приложения, содержащие дополнительные сведения о разработке и использовании ИПТ.

4.11.1 Вводная часть включает титульный лист и введение.

4.11.1.1 На титульном листе должны быть приведены:

- наименование ИПТ, включающее термин «информационно-поисковый тезаурус» и указывающее область его применения;
- наименование организации-разработчика и дополнительные сведения об авторстве ИПТ;
- сведения о переиздании;
- место и год создания или издания ИПТ.

Оформление титульного листа приведено в приложении А.

4.11.1.2 Введение должно содержать следующее:

- цель создания и область применения тезауруса, описанные кодами и наименованиями рубрик Межгосударственного рубрикатора НТИ;
- ссылки на источники, использованные для сбора лексики ИПТ;
- ссылки на нормативные и методические документы, использованные при составлении ИПТ;
- описание порядка составления тезауруса, включая обоснование представительности использованных источников лексики;
- описание состава и структуры ИПТ;
- перечень отношений между ЛЕ и методические основания для их установления;
- перечень всех символов и специальных сокращений, допущенных для представления ЛЕ;
- порядок алфавитного расположения ЛЕ (расположение букв разных алфавитов, небуквенных символов и др.);
- количественные характеристики тезауруса (общее число статей, число дескрипторов, аскрипторов и др.);
- описание состава и формы представления дополнительных данных в словарных статьях;
- абзац следующего содержания: «Тезаурус подготовлен в соответствии с ГОСТ 7.25».

4.11.1.3 Введение к последующим изданиям (версиям) ИПТ дополнительно должно содержать следующее:

обоснование необходимости составления новой версии тезауруса;

указание на характер внесенных изменений.

4.11.2 Форма представления дескрипторных и аскрипторных статей в лексико-семантическом указателе

4.11.2.1 Дескрипторная статья состоит из заглавного дескриптора, списка дескрипторов и аскрипторов, семантически связанных с ним, с обозначением видов связи.

4.11.2.2 Аскрипторная статья состоит из аскриптора и заменяющих его при обработке и поиске информации дескрипторов или комбинации дескрипторов.

4.11.2.3 Словарная статья может также включать:

- частоту использования дескриптора;
- кодовый номер дескриптора;
- код дескриптора по систематическому указателю;
- классификационные индексы;
- дополнительные семантические и лексикографические пометы;
- иноязычные эквиваленты.

Дополнительные данные располагают непосредственно за заглавным дескриптором и должны быть отделены от него.

Допускается указание дополнительных данных также при незаглавных ЛЕ.

4.11.2.4 Для указания языка иноязычных эквивалентов используют латинский код языка по ГОСТ 7.75.

4.11.2.5 В рамках дескрипторной статьи термины располагают в следующем порядке:

- заглавный дескриптор;
- дополнительные данные;
- лексическое примечание;
- аскрипторы или дескрипторы-синонимы;
- вышестоящие дескрипторы;
- нижестоящие дескрипторы;
- ассоциативные дескрипторы;
- дескрипторы, связанные другими видами отношений.

4.11.2.6 Внутри каждой группы ЛЕ, связанных с заглавным дескриптором одним видом парадигматических отношений, должен быть алфавитный порядок расположения. Если незаглавные ЛЕ имеют примечания, предусмотренные 4.5.6.9, то они могут быть сгруппированы по алфавитному порядку примечаний.

4.11.2.7 В дескрипторной статье должны быть приведены все непосредственно связанные с заглавным дескриптором вышестоящие и нижестоящие дескрипторы. Иерархические связи на втором и следующих уровнях могут быть указаны только при необходимости.

4.11.2.8 Для указания связи между ЛЕ тезауруса должны быть использованы ссылки, обозначение и упорядочение которых установлено таблицей ссылок (п. 4.5.1).

Допускается также указывать связи других типов, смысл которых должен быть пояснен во вводной части ИПТ. Обозначение этих ссылок должно отличаться от установленных в настоящем стандарте.

4.11.2.9 Обозначение ссылки в словарной статье может быть приведено только один раз перед первой ЛЕ, связанной с заглавным дескриптором данным отношением.

4.11.2.10 Каждая ЛЕ в словарной статье занимает новую строку. Если она не помещается на строке, то все последующие строки с ее продолжением должны быть сдвинуты относительно начала ЛЕ.

4.11.2.11 При печатном издании ИПТ необходимо соблюдать следующие требования:

- дескрипторы должны быть выделены шрифтом или иным способом, позволяющим легко отличать их от аскрипторов и лексических примечаний;
- реляторы должны быть напечатаны в круглых скобках в конце дескриптора тем же шрифтом, что и остальной текст дескриптора, через пробел (допускается другой вид скобок, отличных от косых);
- лексические примечания должны быть напечатаны после текста дескриптора в косых скобках через пробел или отдельной строкой, имеющей пометку «лп» (лексическое примечание);
- заглавная ЛЕ статьи должна выделяться левым сдвигом на расстояние, равное шести (или более) строчным буквам относительно начала незаглавной ЛЕ.

Примеры

а) Дескрипторная статья

АЛГОРИТМИЧЕСКИЕ ЯЗЫКИ (формализм теории алгоритмов
см. ТЕОРИЯ АЛГОРИТМОВ)
с языки алгоритмические
машиноориентированные языки
проблемноориентированные языки
в ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ
ФОРМАЛЬНЫЕ ЯЗЫКИ
н АВТОКОДЫ
АЛГОЛ
ИПЛ
КОБОЛ
КОМИТ
ПЛ/1
ФОРТРАН
а АЛГОРИТМЫ
ПРОГРАММИРОВАНИЕ
ср искусственные языки

б) Аскрипторные статьи:

1) Языки программирования
см АЛГОРИТМИЧЕСКИЕ ЯЗЫКИ
2) Языки
иа ФОРМАЛЬНЫЕ ЯЗЫКИ
ЕСТЕСТВЕННЫЕ ЯЗЫКИ
3) Алфавитно-цифровые знаки
ик БУКВЫ
ЦИФРЫ

4.11.3 ИПТ представляют в печатном виде, на микроформах и на машиночитаемых носителях.

При представлении ИПТ на машиночитаемых носителях формат записи должен соответствовать ГОСТ 7.47. Обязательным является оформление вводной части в напечатанном виде. Распечатка ИПТ и его фрагментов должна соответствовать требованиям оформления, установленным настоящим стандартом.

Представление ИПТ на видеотерминалах должно соответствовать основным требованиям оформления, установленным настоящим стандартом.

Представление ИПТ на микроформах должно соответствовать требованиям к печатной форме, установленной настоящим стандартом, и требованиям стандартов на микроформы.

5 Депонирование информационно-поискового тезауруса

5.1 ИПТ, разработанные в соответствии с настоящим стандартом, включая их первое издание и последующие переиздания, должны быть задепонированы в соответствующем национальном депозитарном фонде, а также в следующих международных депозитарных фондах:

- ИПТ на английском языке, включая ИПТ, содержащие эквиваленты дескрипторов на английском языке:

Thesaurus Clearinghouse,
The Library,
Faculty of Information Studies,
University of Toronto,
140 St George Street,
TORONTO,
Ontario M5S 1A1,
Canada;

- ИПТ на русском языке, включая ИПТ, содержащие эквиваленты дескрипторов на русском языке:

МТК 191, секретариат,
ВИНИТИ,
ул. Усиевича, 20,
МОСКВА, 125315,
Россия;

- ИПТ на других языках:

Instytut Informacji Naukowej, Technicznej i Ekonomicznej,
Clearinghouse,
ul. Zurawia, 3/5,
00-926 WARSZAWA,
Poland.

5.2 Национальные депозитарии распространяют информацию о составе фонда депонированных ИПТ и предоставляют их разработчикам новых ИПТ с целью заимствования элементов и обеспечения совместимости лингвистического обеспечения различных информационных систем.

5.3 Национальные депозитарии проводят экспертизу представленных ИПТ на соответствие настоящему стандарту и выдают разработчику сертификат соответствия.

ПРИЛОЖЕНИЕ А
(справочное)

Оформление титульного листа информационно-поискового тезауруса

_____
наименование министерства (ведомства)

_____
наименование организации, разработавшей тезаурус

УДК _____ Регистрационный номер _____ _____

Межгосударственный рубрикатор НТИ _____

ТЕЗАУРУС

информационно-поисковый
по

_____
область применения

_____
место и год издания

УДК 002.543.01:006.354 МКС 01.140.20 Т62 ОКСТУ 0007

Ключевые слова: информационно-поисковый тезаурус, лексическая единица, дескриптор, аскриптор, поиск информации, словарная статья, парадигматические отношения

Редактор *Р.Г. Говердовская*
Технический редактор *Н.С. Гришанова*
Корректор *Р.А. Ментова*
Компьютерная верстка *В.И. Грищенко*

Изд. лиц. № 02354 от 14.07.2000. Сдано в набор 02.10.2001. Подписано в печать 06.11.2001. Усл. печ. л. 1,86.
Уч.-изд. л. 1,65. Тираж 489 экз. С 2480. Зак. 1049.

ИПК Издательство стандартов, 107076, Москва, Колодезный пер., 14.
http://www.standards.ru e-mail: info@standards.ru
Набрано в Издательстве на ПЭВМ
Филиал ИПК Издательство стандартов — тип. "Московский печатник", 103062, Москва, Лялин пер., 6.
Плр № 080102

gost.e-ecolog.ru